LeAF: Leveraging Convolutional Neural Networks for Plant Anomaly Detection and Classification for Farmers with Large Language Models for Natural Language Interaction

Aditya Sengupta - Redmond, WA, USA

Introduction

Agriculture has a huge environmental impact, and with climate change, labor shortages, and a rise in global food demand, agriculture needs to become **more efficient and effective**.

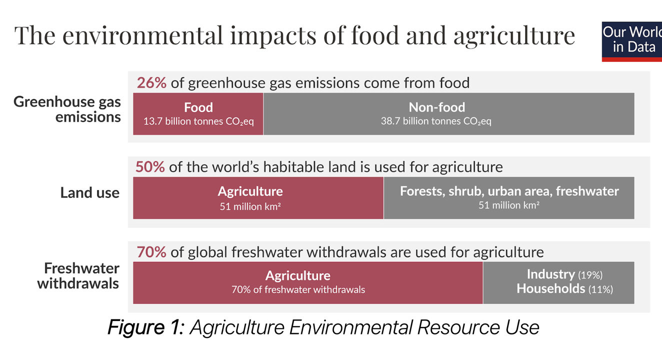


Figure 1: Agriculture Environmental Resource Use

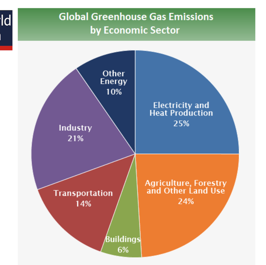


Figure 2: Agriculture Greenhouse Gas Emissions

Figure 3: Farmers spraying pesticides on fields

Farmers struggle to monitor and maintain plant health (plant anomalies such as **pests, weeds, disease**) efficiently and sustainably. **40%** of global crop production lost to plant pests and disease, costing **$220 billion annually.**

With limited time and knowledge, manual surveillance for plant anomalies is difficult for farmers, resulting in excessive and indiscriminate use of fertilizers and pesticides.

Agricultural chemicals emit N_2O and CH_4 (greenhouse gases that are **300x** more potent than CO_2), increase production costs, and harm wildlife and ecosystems. **$60 billion** is spent on pesticides annually. Pesticide residues on food lead to **20,000+** new cases of cancer every year. Contamination in nature results in **80+ million** fish and bird deaths annually.

Need for **low-cost automation to monitor plant anomalies** that is affordable for low-income farmers to manage large farms at scale

Existing Research

Dimension	Existing Work	LeAF
Solution Type	Classification Only	Classification & Bounding Box
Accuracy	30-40%	90%+
Extensibility	Only Specific Plants/Anomalies	All Plants/Anomalies
Model Size	20-100 million parameters	3 million parameters
Natural Language Interaction?	No	Fluent and Knowledgeable
Deployment	No	Yes

Table 1: Comparison of LeAF to existing research on classifying plant anomalies

Current solutions lack comprehensive features such as bounding boxes for localization, have low accuracy rates, and are limited to specific plant anomalies. These solutions rely on large models, making them computationally intensive and unsuitable for low-power devices. Moreover, they lack natural language interaction and explanatory capabilities for farmers. Additionally, they are not end-to-end and do not support robot automation for deployment in real farm scenarios.

Objectives

Research Question

Is it possible to automate plant anomaly detection and classification for farmers, reduce chemical usage, and provide targeted treatment suggestions using Convolutional Neural Networks (CNNs) and Large Language Models (LLMs)?

Engineering Goals

- Lightweight & Accurate
- Multimodal
- Deployable
- Interactable

LeAF aims to use CNNs and LLMs to provide an end-to-end solution for farmers to identify and treat plant anomalies.

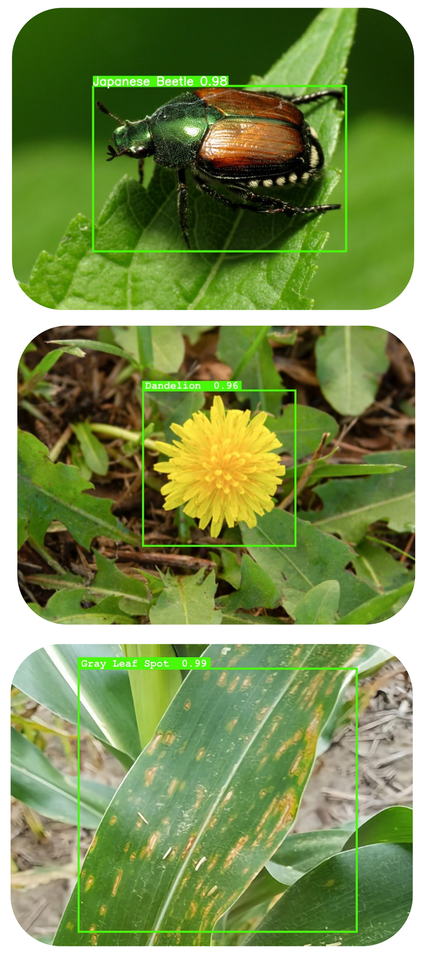


Figure 4: Sample detections & classifications for pests, weeds, & disease

Dataset

iNaturalist

I collected **10,000** images from iNaturalist (initial focus on pests). I also did some data filtering and cleaning based on image quality, research grade, number of validations.

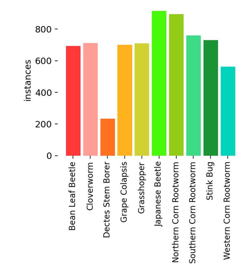


Figure 5: Dataset Class Image Comparison

I chose the **10 most harmful agricultural pests** in US for Corn and Soybean, the most grown crops. Dataset comprises about 1,000 images per class with an **80%-10%-10% train, validation, and test split**.

I used GroundingDINO to annotate the images with insect bounding boxes and trained a ResNet-18 classification model.

The image initially goes to GroundingDINO for bounding box, then cropped image goes to ResNet-18 for classification.

Methodology

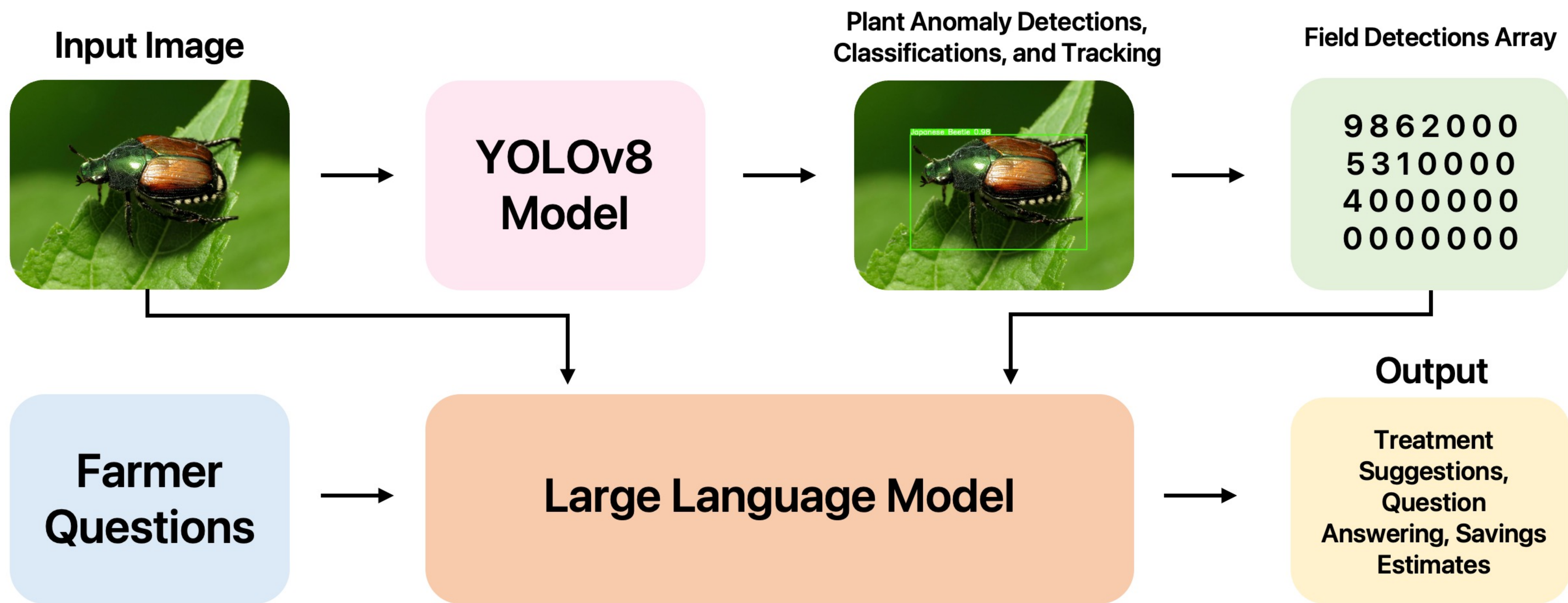


Figure 6: LeAF model pipeline and structure (Image → Classification and Detection Model → Object Tracking → Field Array → Large Language Model → Output Treatment Suggestions)

Training & Evaluation

I iteratively trained, tested, and tuned hyperparameters to increase accuracy from **~65% initially to ~91% finally**.

The model performed well on test images and was able to successfully detect and classify small and camouflaged pests. I analyzed performance on the following cases:

Similar Background Color (Bean Leaf Beetle, Southern CRW), **Small and Blending with Background** (Northern CRW), **Extreme Camouflage** (Grasshopper), and **different orientations/angles**.

Prediction was taking ~3 seconds per image on an Nvidia T4 GPU, with majority time spent on GroundingDINO model.

This created an opportunity for a lightweight model to run on a robot with low compute and fast enough to keep up with video feed.

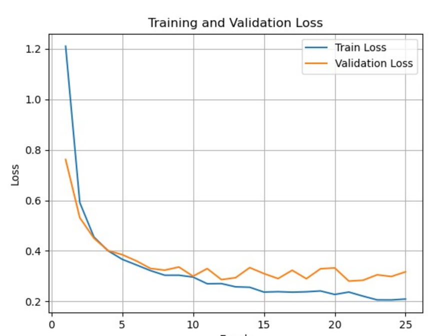


Figure 9: Training and Validation Loss (Lower is better, decreases with more epochs.)

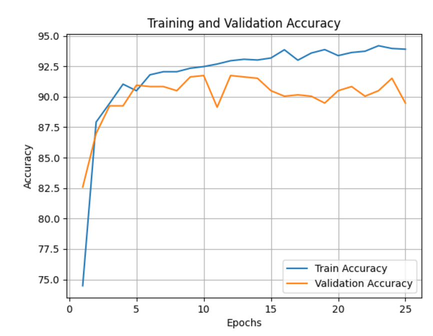


Figure 10: Training and Validation Accuracy (Higher is better, increases with more epochs.)

Tuning Technique	Validation Accuracy
Original Accuracy	65.71%
Dataset Filtering and Augmentation	81.29%
Adam Optimizer	88.97%
Learning Rate 0.001	90.96%
Batch Size 32	91.29%
Train 10 Epochs (Last Layer)	91.43%

Table 2: Hyperparameter tuning steps to increase accuracy.

Model Metrics and Formulas

$$logloss = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} * \log(pij)$$

$$accuracy = \frac{TP + TN}{Total} = 91.43\%$$

$$precision = \frac{TP}{TP + FP} = 90.94\%$$

$$recall = \frac{TP}{TP + FN} = 90.46\%$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} precision_i * recalli = 0.823$$

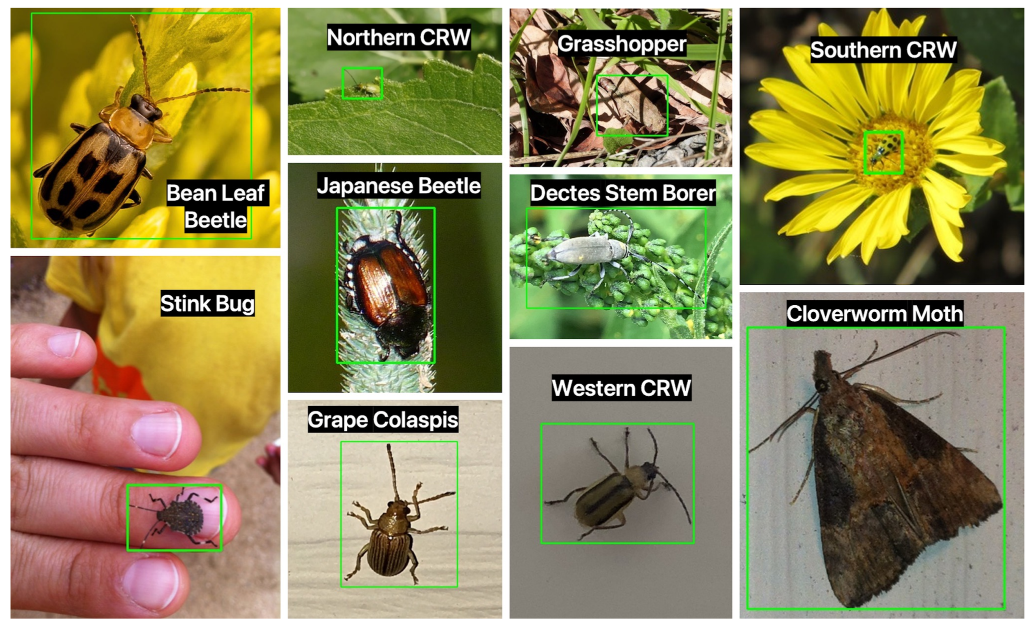


Figure 7: Sample detections and classifications from the initial model pipeline with GroundingDINO & ResNet-18

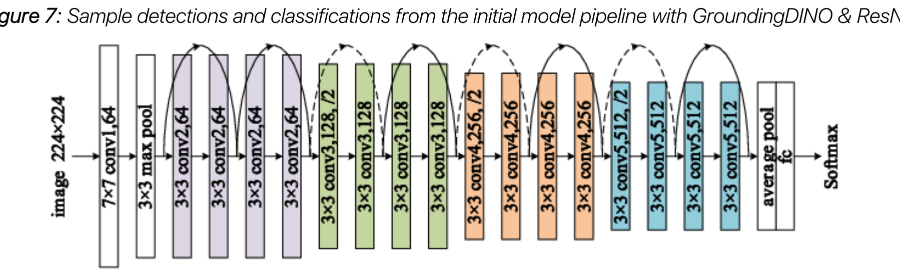


Figure 8: ResNet-18 model structure with convolutional and pooling layers

Distilled Model for Mobile Devices

Prediction time of 3 seconds per image even on a high-performance GPU meant that the solution needed to use a smaller model in order to **run on a video feed on edge devices** (like robots that have low compute.

I decided to train a model with **less parameters** (student model) on the dataset annotated by the initial model pipeline (teacher model). This process is known as model distillation. The goal was to get **similar accuracy with a smaller size**.

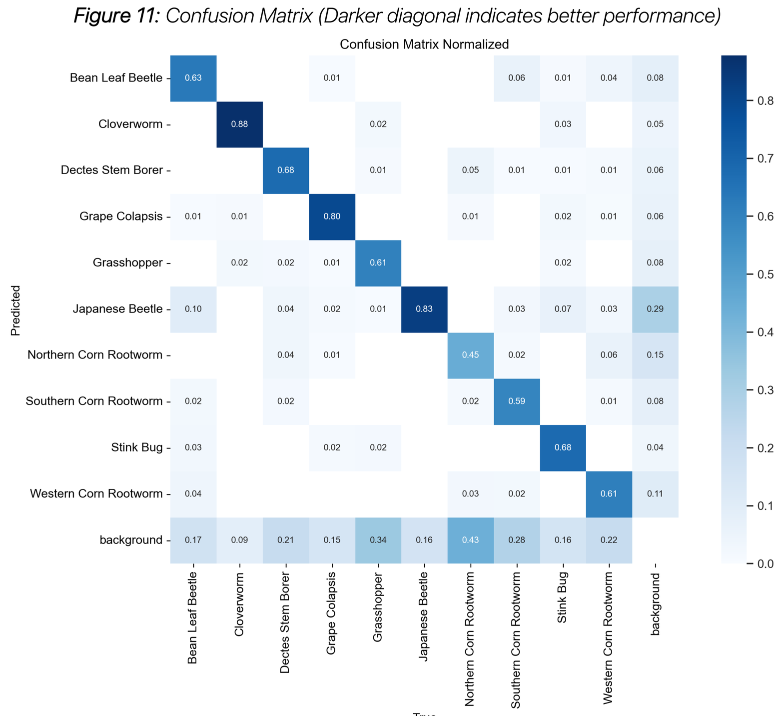
For this, I used **YOLOv8n (Nano)**, a smaller model (distilled from the larger model pipeline) that provided both detection and classification output and is capable of running on low compute devices.

I trained the YOLOv8n model with bounding box data from GroundingDINO coupled with my trained ResNet-18 for class labels.

ResNet-18 has **11 million** parameters and GroundingDINO has **172 million** parameters, while YOLOv8n only has **3.2 million** parameters.

So, the resulting custom-trained model had similar accuracy at only **0.17%** of parameters as original pipeline and took only **~30msec** for prediction per image **(100x improvement)**, allowing it to run on video captured by robot.

Figure 11: Confusion Matrix (Darker diagonal indicates better performance)



Distilled Model Metrics

$$Accuracy = 95.90\%$$

$$mAP = 0.815$$

$$Precision = 93.80\%$$

$$Recall = 81.30\%$$

Figure 12: Sample detections and classifications from the custom-trained and distilled YOLOv8n model

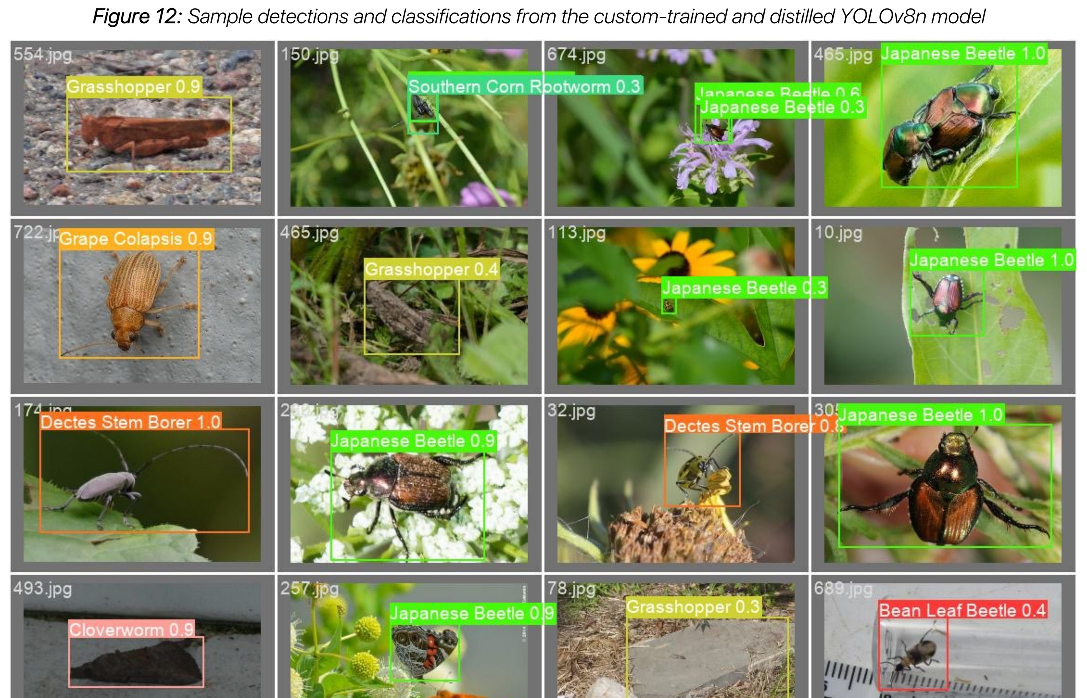


Figure 13: Box, Class, and Object Loss (Lower is better, decreases with more epochs)

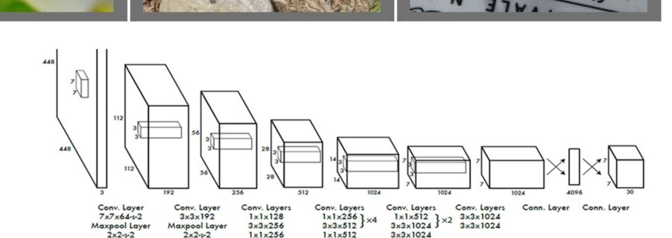


Figure 14: YOLOv8 model structure with convolutional and pooling layers

Field Mapping

The YOLOv8 model runs on the **robot camera feed** as it traverses the field.

On top of this, I implemented an **object tracking algorithm** to count detected pests and compute how many have been detected at field level.

I also added a class to **detect plant stems**. With this, we count the number of pests between plant stem detections to get **plant-specific anomaly counts**. These counts are stored in 2D-matrix (**Field Detections Array**) where each element represents the number of pests on that plant in the field.

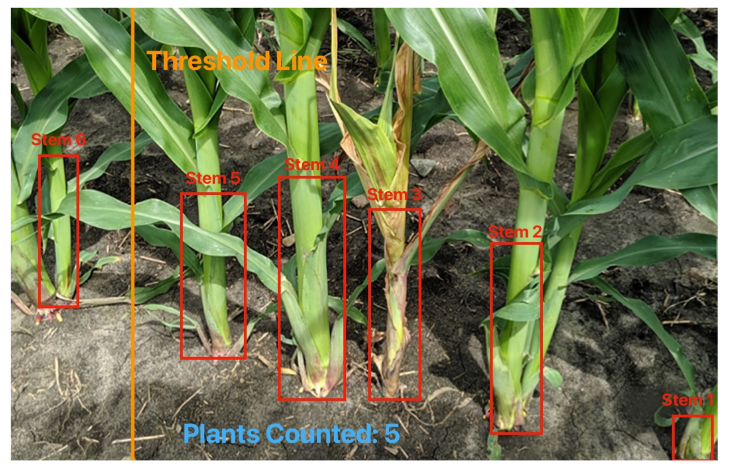


Figure 15: Sample Plant Stem Detection Counts

This allows mapping at field level to find infestation clusters. Farmers see the map and can focus treatment.

Large Language Model & Web Tool

I wanted to add natural language and multimodal interaction with the farmer. For this, I fine-tuned the **Google Gemini** Large Language Model (LLM) with sample field array prompts, farmer questions, pest images, and example outputs I created. This enables the farmer to obtain analysis of infestations based on the field map and treatment suggestions. The farmer can also ask questions to the LLM for more insights.

I built this into a web tool to provide personalized estimates to the farmer on their financial, chemical, and emissions savings based on farm-specific usage data.

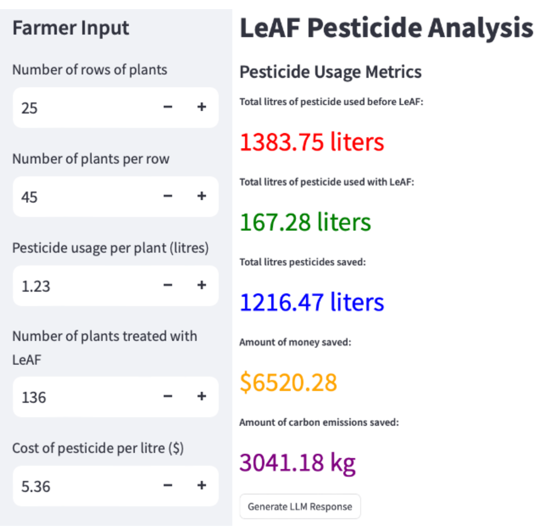


Figure 16: Web tool to provide LeAF savings estimates

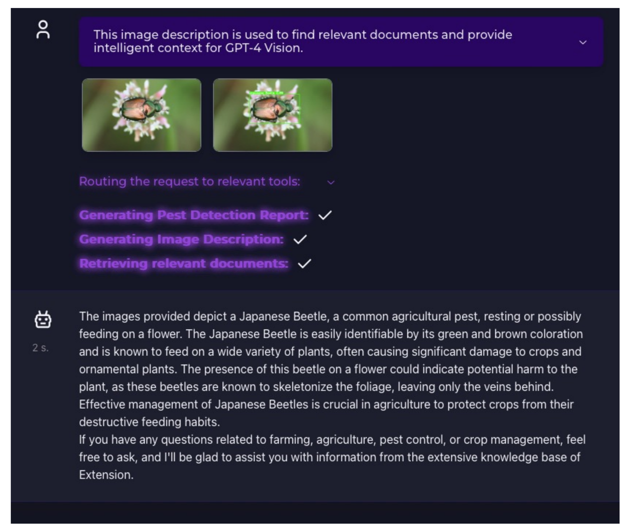


Figure 17: Sample Multimodal Farmer LLM Interaction

BRANCH Deployment in Farms

I have deployed LeAF using my custom-made **BRANCH Robot (Budget-friendly Robot for Agricultural Nonintrusive Crop Photography)**.

Existing commercial robots like EarthSense TerraSentia cost $5,000+ and require complex robot infrastructure (inaccessible to most farmers). I made BRANCH as an accessible way to deploy LeAF **under $500.**

BRANCH is a rugged robot drivetrain with tripod and brain with sensors for manual and auto navigation. Farmers can **attach their phone** on which the YOLOv8n model runs in the **LeAF mobile app**.

Figure 18: Custom-built BRANCH Robot

Figure 19: Deployment with Farmers

I am currently working with Washington State farms **(Nelson Farms – Farmington, WA and Dancing Crow Farm – Carnation, WA)** and have partnered with **SnoValley Tilth**, a nonprofit organization of **80+ farmers** and meet with farmers weekly. I am also working with the **WSDA** and **USDA** to expand across the USA and eventually globally.

Conclusion & Impact

LeAF utilizes recent AI/ML advancements like Convolutional Neural Networks and Large Language Models to offer an **end-to-end solution** for **monitoring plant anomalies** with plant anomaly detection and classification, **field mapping**, suggestions for **treatment**, detailed **savings estimates**, and **BRANCH robot-based deployment**. LeAF is effective and efficient with **high accuracy** (95+%). LeAF aims to **reduce costs, chemical usage, and environmental impact** for farmers while maximizing crop health, yield, and productivity.

LeAF has the potential to **revolutionize agriculture** and **empower farmers to efficiently produce food that has no negative impact on the environment.**

Future Research

Can this structure be expanded to 3,000+ pests in a single model?

Can Aerial/Drone imagery be leveraged on dense/rough terrain farms?

Can data from soil moisture/humidity sensors correlate with and predict future pest infestations?