

AgroBench: Vision-Language Model Benchmark in Agriculture

Supplemental Material

A. Statistics

This section provides detailed statistics of AgroBench. In Figs. 1, 2, 3, 4, 5, and 6, we present the distribution of 682 plant-disease categories, categorized by the cause of the disease. The distribution of 134 pest categories is shown in Fig. 7, and the distribution of 108 weed categories is shown in Fig. 8, both categorized by their order.

B. Experimental Details

B.1. Hyperparameter Settings

We follow the official hyperparameter settings. For closed-source models, we set the temperature hyperparameter to 0.0. For open-source models, we use the default temperature values provided by each implementation. For GPT-4o and GPT-4o mini, we set the detail parameter for visual inputs to low.

B.2. Model Details

GPT-4o [4] and GPT-4o mini [5]. GPT-4o is an extended model in the series of Generative Pre-training Transformers (GPTs). GPT-4o mini is a lightweight model designed for efficiency and speed. While it sacrifices a small degree of accuracy compared to GPT-4o, it remains an efficient option. We use the model gpt-4o-2024-08-06 for GPT-4o and gpt-4o-mini-2024-07-18 for GPT-4o mini.

Gemini 1.5-Pro and Gemini 1.5-Flash [7]. Gemini has been simultaneously trained on a huge amount of multi-modality dataset among image, video, audio, and text data. Gemini 1.5-Flash is the lightweight version of Gemini 1.5-Pro. We use the model gemini-1.5-pro-001 for Gemini 1.5-Pro and gemini-1.5-flash-001 for Gemini 1.5-Flash.

Qwen [1]. Qwen is a vision-language model trained with a huge amount of image-text data and instruction tuning. Qwen has several specific modes, such as coding, additional audio modality, and mathematics. In this paper, we employ Qwen2-72B (QwenVLM-72B) and Qwen2-7B (QwenVLM-7B).

LLaVA [2]. LLaVA takes advantage of trained large language models (LLMs) and instruction tuning with vision models. The LLaVA project has proved that the merged representation is greatly effective for visual reasoning and

dialogue using multi-modal input. In the experiments, we use LLaVA v1.5 with 7 and 13 billion parameters (LLaVA-1.5-{7B, 13B}). We also use LLaVA-Next [3] improved version of LLaVA.

CogVLM [8]. This vision-language model employs trained LLMs and visual encoders and additionally tunes feed-forward layers in order to combine the ability of image-text representations. CogVLM has confirmed question-answering and visual reasoning performance on representative vision and language datasets. We utilize the second version of CogVLM, which contains 19B parameters and has Llama-3 backbones.

Emu [6]. Emu potentially executes both image-to-text and text-to-image across diverse visual, linguistic, and multi-modal tasks. This foundation model enhances the vision and language performance from both recognition and generation learning across multiple modalities. The model size is 37B.

C. Annotation Examples

In this section, we show example annotations for the seven tasks as follows:

- Disease Identification (DID): Fig. 9, Fig. 10
- Disease Management (DMN): Fig. 11, Fig. 12
- Pest Identification (PID): Fig. 13, Fig. 14
- Weed Identification (WID): Fig. 15, Fig. 16
- Crop Management (CMN): Fig. 17, Fig. 18
- Machine Usage (MQA): Fig. 19, Fig. 20
- Traditional Methods (TM): Fig. 21, Fig. 22

As shown in the figures, we provide reasoning annotations for DMN, CMN, MQA, and TM tasks, ensuring explainability. We also list the category for TM in Table. 3 and for MQA in Table. 4.

D. Detailed Analysis

D.1. Context Reliance

We show example cases where GPT-4o can answer questions without input images by guessing the most likely option (Fig. 23 and Fig. 24). Even though we confirmed that our prepared questions do not include crop names or

Metric	DID	PID	WID
F1	27.26	27.09	1.84
Substring	24.57	30.33	9.03

Table 1. Word-based Evaluation Metrics under Free-form Answer Setting

appearance traits, and other options could be correct, the models tend to make conclusions based on estimation.

D.2. CoT Examples

Here, we show output examples of chain-of-thought (CoT) reasoning by GPT-4o. In Fig. 25, the model identifies the pest step by step by observing the image’s appearance, checks all the options, and concludes with the correct answer. In Fig. 26, the model focuses on determining the plant disease species but fails to observe the severity of the disease, leading to an incorrect conclusion.

D.3. Error Examples

Here, we present additional error examples as follows:

- Lack of Knowledge Error: Fig. 27
- Perceptual Error: Fig. 28
- Shortcut Error: Fig. 29
- Reasoning Error: Fig. 30
- Double Answer Error: Fig. 31
- Interpretation Misunderstanding: Fig. 32
- Reject to Answer: Fig. 33

D.4. Free format

We evaluate our dataset under a free-form answer setting, where the model is not constrained to choose from multiple choices. For the identification tasks (DID, PID, WID), we use the following prompt:

```
{question} Respond with the
correct answer only, using a
single noun or short phrase. Do
not include full sentences.
```

We show the result in Table 1. The responses are evaluated using the F1 score and substring match accuracy, which reflect surface-level correctness as well as partial matches.

For the other tasks (DMN, CMN, MQA, TM), the prompt used is:

```
{question} Respond with the
correct answer only.
```

We show the result in Table 2. The answers are evaluated with two semantic similarity metrics: BERTScore (F1) and Sentence-BERT cosine similarity.

Metric	DMN	CMN	MQA	TM
BERTScore (F1)	87.20	88.15	88.16	87.33
Sentence-BERT	0.443	0.557	0.508	0.457

Table 2. Sentence-based Semantic Similarity under Free-form Answer Setting

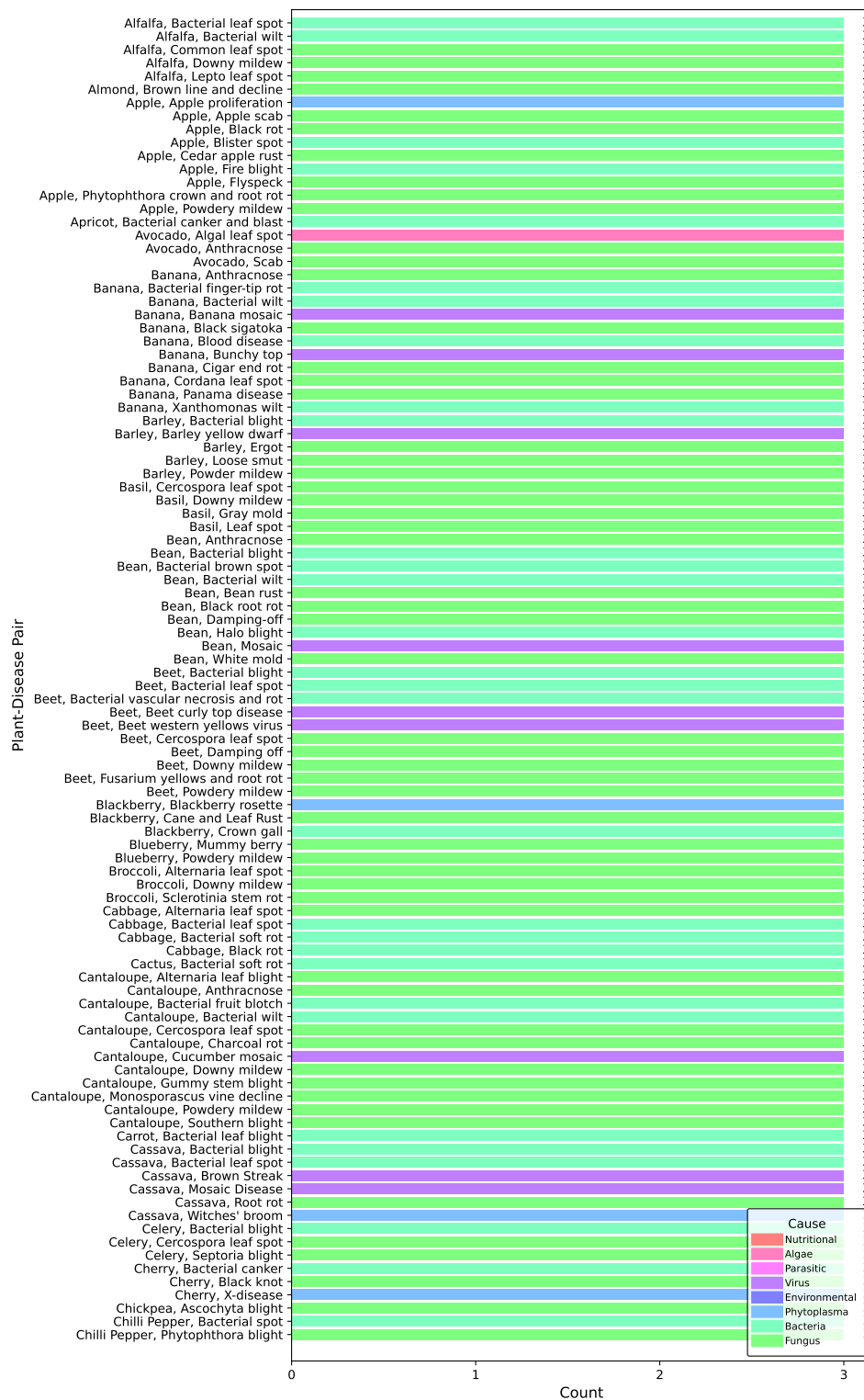


Figure 1. Disease Categories Counts (A).



Figure 2. Disease Categories Counts (B).

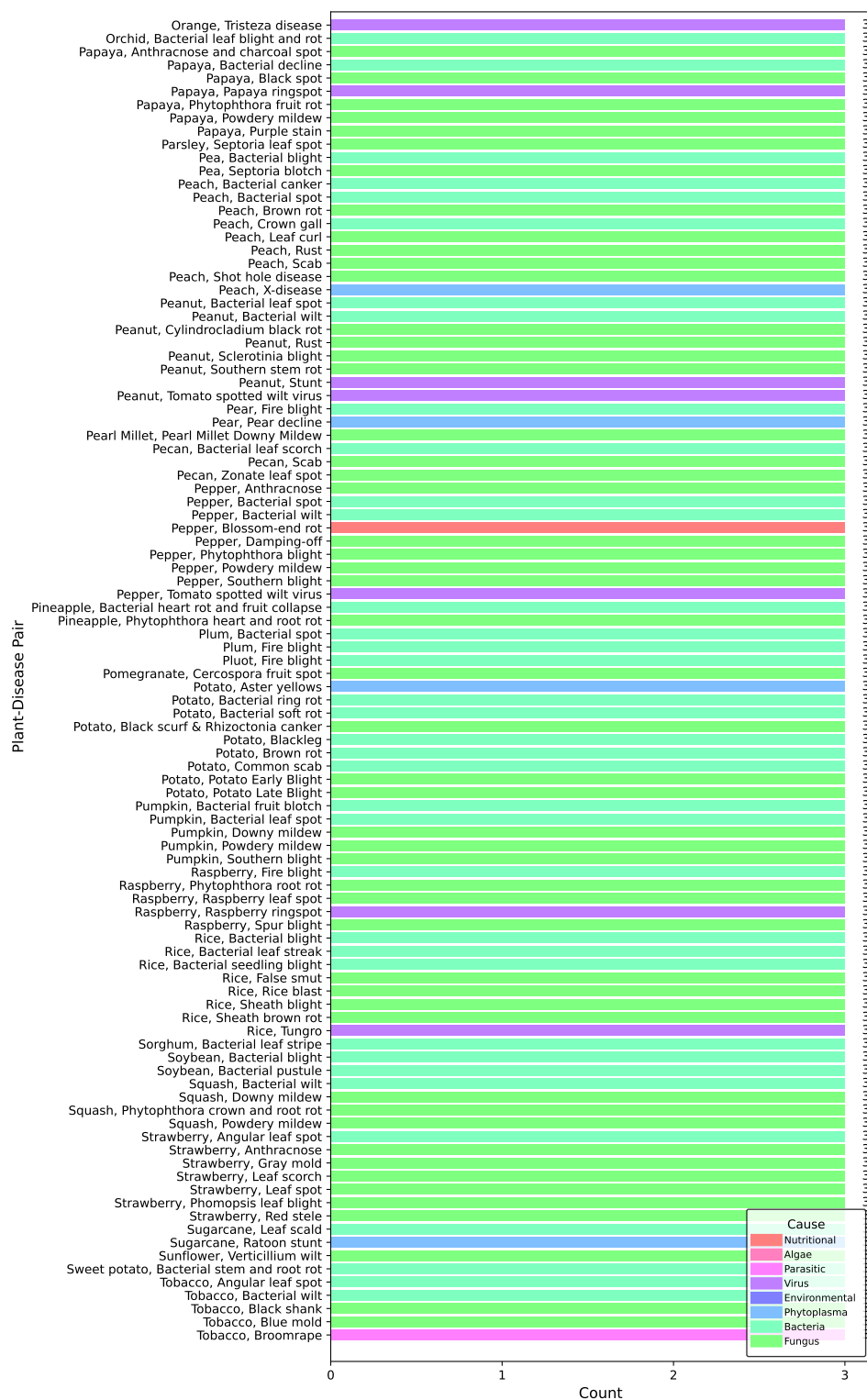


Figure 3. Disease Categories Counts (C).



Figure 4. Disease Categories Counts (D).

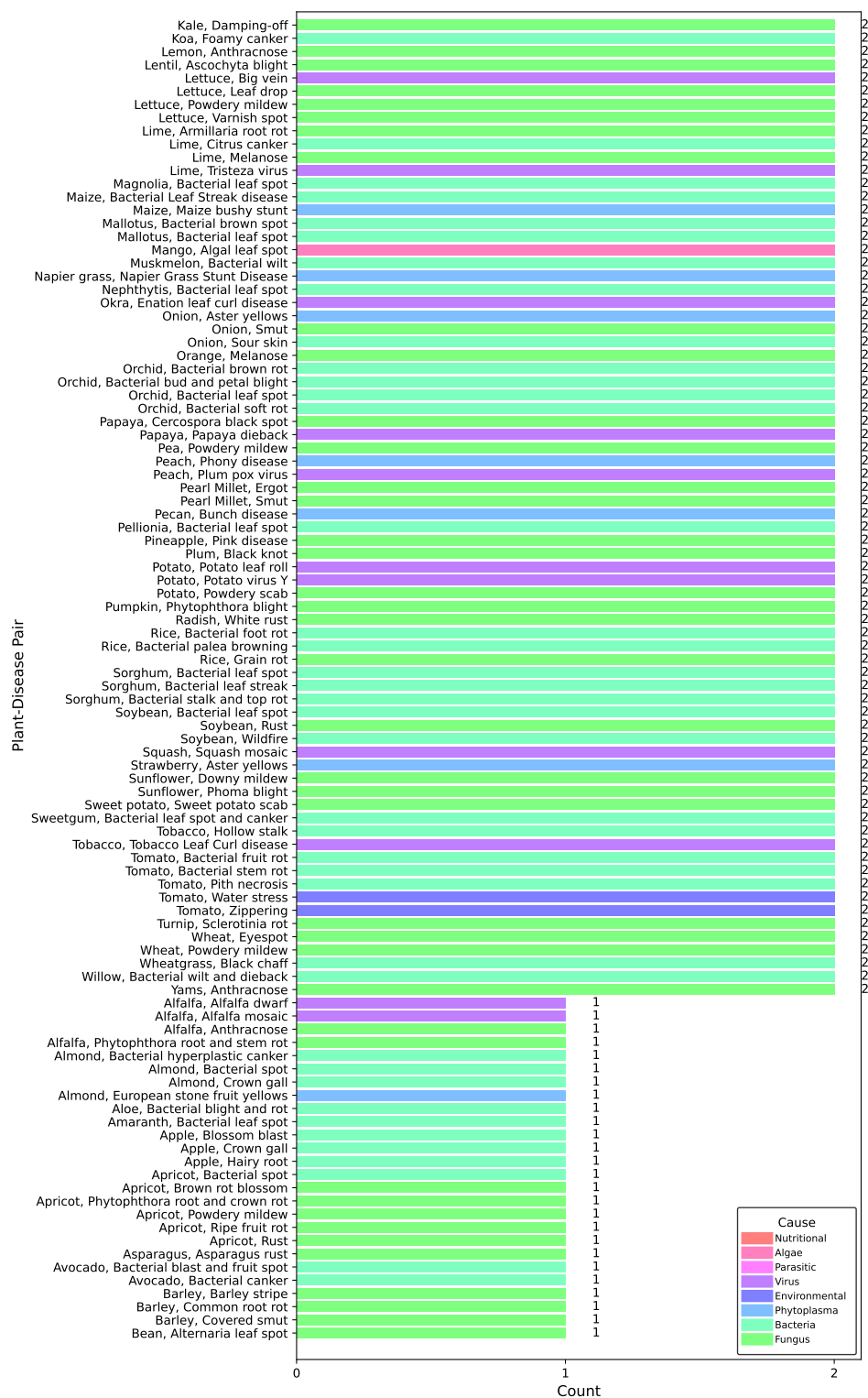


Figure 5. Disease Categories Counts (E).



Figure 6. Disease Categories Counts (F).

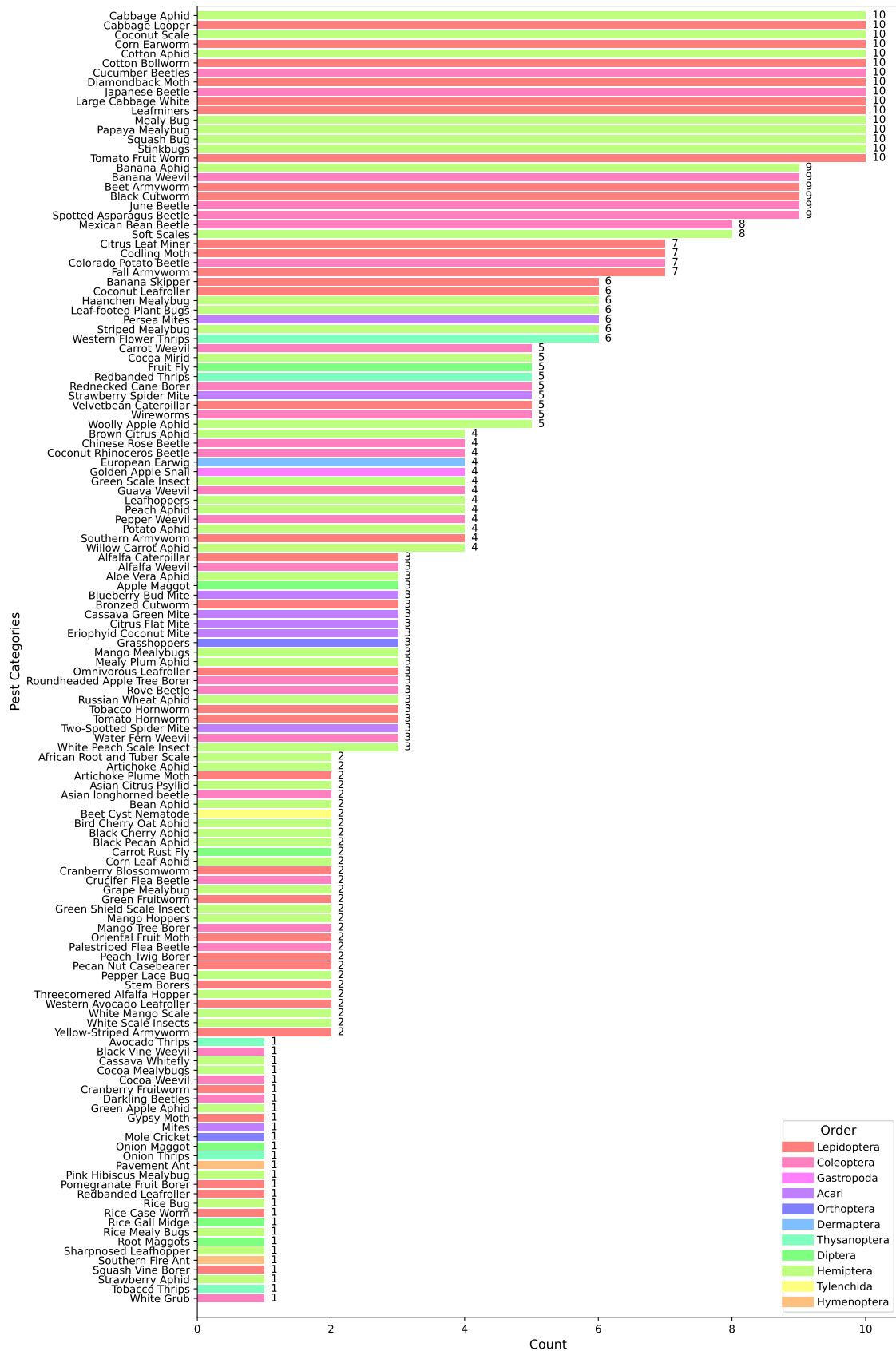


Figure 7. Pest Counts.

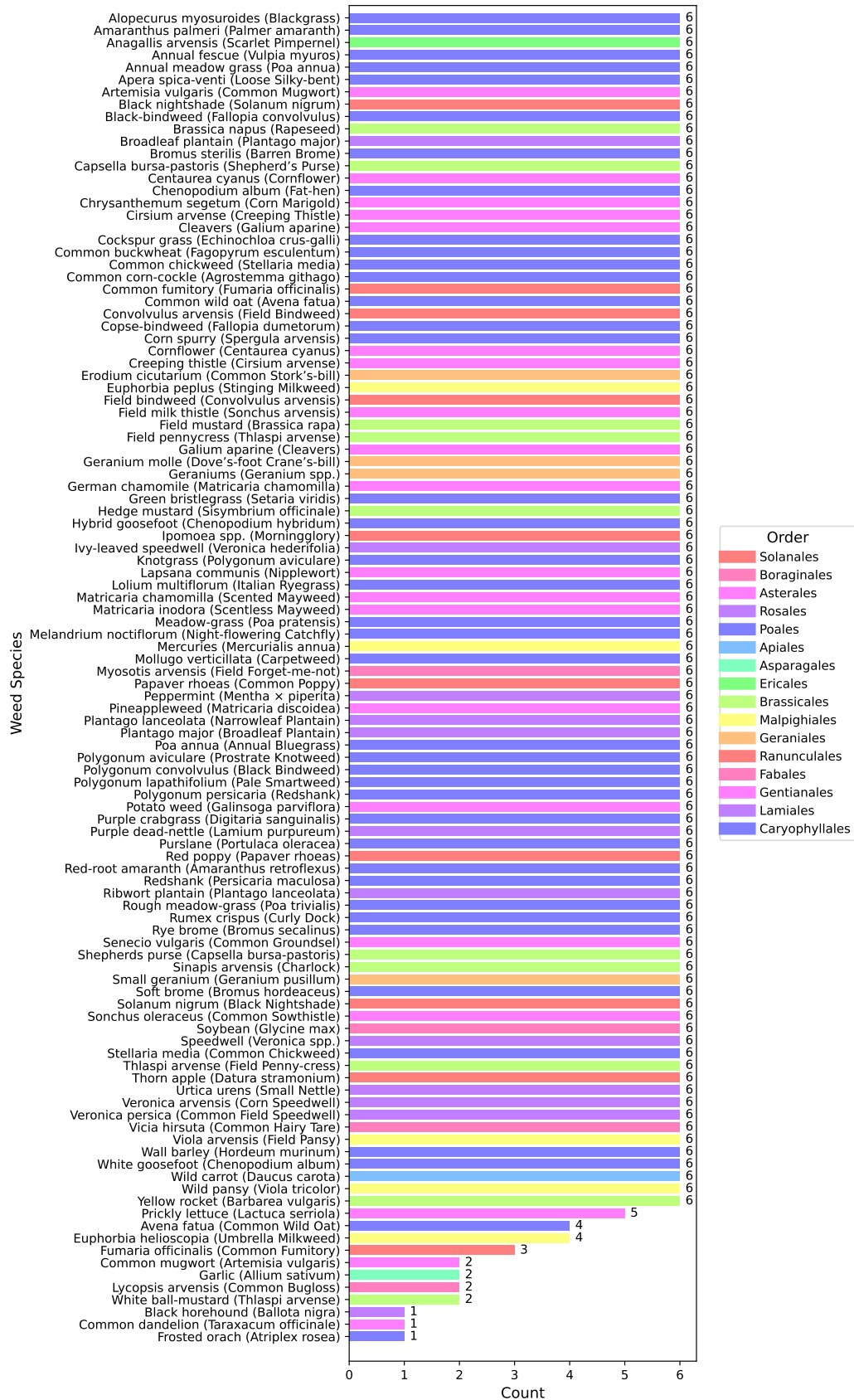


Figure 8. Weed Counts.

**Question**

What disease is affecting the plant?

Options

- A. Bacterial pocket
- B. Yellow wilt
- C. Bacterial vascular necrosis and rot
- D. Beet western yellows virus
- E. Fusarium yellows and root rot

Answer B. Yellow wilt

Plant Beet

Disease Yellow wilt

Cause Fungus

Figure 9. Disease Identification Task (DID) Example (A).

**Question**

What disease is affecting the plant?

Options

- A. Cedar apple rust
- B. Apple proliferation
- C. Blister spot
- D. Apple scab
- E. Flyspeck

Answer D. Apple scab

Plant Apple

Disease Apple scab

Cause Fungus

Figure 10. Disease Identification Task (DID) Example (B).



Question

What is one recommended practice for managing irrigation when dealing with the disease shown in the image?

Options

- A. Completely stop irrigation at hull split to conserve water.
- B. Allow water to pool around the base of the trees to improve soil saturation.
- C. Keep irrigation levels constant throughout the growing season.
- D. Reduce irrigation at hull split to manage disease and stress.
- E. Increase irrigation at hull split to promote faster growth.

Answer D. Reduce irrigation at hull split to manage disease and stress.

Reasoning

Reducing irrigation at hull split is recommended to manage disease and plant stress effectively. Additionally, maintaining optimal water management during this stage ensures that the plant remains healthy while reducing the risk of over-irrigation, which could exacerbate the disease or lead to other stress-related issues.

Plant Almond

Disease Hull rot

Cause Fungus

Figure 11. Disease Management Task (DMN) Example (A).



Question

What planting technique helps avoid the introduction of the disease shown in the image?

Options

- A. Harvest cuttings from plants showing early wilting symptoms.
- B. Use cuttings taken only from the upper part of the plant.
- C. Reuse planting slips left over from previous growing season.
- D. Use the entire plant, including tubers, for propagation.
- E. Collect planting material from below-ground tubers directly.

Answer B. Use cuttings taken only from the upper part of the plant.

Reasoning

Above-ground cuttings reduce the chance of transferring pathogens from tubers or infected basal tissue, helping prevent black rot.

Plant Sweet potato

Disease Black rot

Cause Fungus

Figure 12. Disease Management Task (DMN) Example (B).

**Question**

What pest is affecting the plant?

Options

- A. Omnivorous Leafroller
- B. Coconut Scale
- C. Eriophyid Coconut Mite
- D. Coconut Leafroller
- E. Coconut Rhinoceros Beetle

Answer D. Coconut Leafroller

Species Coconut Leafroller

Order Lepidoptera

Figure 13. **Pest Identification Task (PID) Example (A).**

**Question**

What pest is affecting the plant?

Options

- A. Cocoa Weevil
- B. Squash Vine Borer
- C. Cassava Green Mite
- D. Cocoa Mealybugs
- E. Cocoa Mirid

Answer E. Cocoa Mirid

Species Cocoa Mirid

Order Hemiptera

Figure 14. **Pest Identification Task (PID) Example (B).**

**Question**

What species of weed does the highlighted bounding box represent in the image?

Options

- A. Morningglory (*Ipomoea* spp.)
- B. *Polygonum convolvulus* (Black Bindweed)
- C. *Plantago lanceolata* (Narrowleaf Plantain)
- D. Rough meadow-grass (*Poa trivialis*)
- E. *Convolvulus arvensis* (Field Bindweed)

Answer E. *Convolvulus arvensis* (Field Bindweed)

Species *Convolvulus arvensis* (Field Bindweed)

Order Solanales

Figure 15. **Weed Identification Task (WID) Example (A).**

**Question**

What species of weed does the highlighted bounding box represent in the image?

Options

- A. *Polygonum aviculare* (Prostrate Knotweed)
- B. Common dandelion (*Taraxacum officinale*)
- C. *Apera spica-venti* (Loose Silky-bent)
- D. *Galium aparine* (Cleavers)
- E. *Melandrium noctiflorum* (Night-flowering Catchfly)

Answer D. *Apera spica-venti* (Loose Silky-bent)

Species *Apera spica-venti* (Loose Silky-bent)

Order Poales

Figure 16. **Weed Identification Task (WID) Example (B).**

**Question**

Is the crop shown in the image ready for harvest based on its current condition?

Options

- A. Yes, the plants are fully green, which indicates freshness for harvest.
- B. No, the pods need to dry out further before harvesting.
- C. No, the seed pods are still green, so the crop is not ready for harvest.
- D. No, the leaves and stems are still green.
- E. Yes, the leaves and stems have turned yellow.

Answer E. Yes, the leaves and stems have turned yellow.

Reasoning

Soybeans are ready for harvest when the leaves and stems have turned yellow, and the seed pods are brown to black.

Crop Soybean

Figure 17. **Crop Management Task (CMN) Example (A).**

**Question**

What soil conditions are ideal for the plant shown in the image to grow well?

Options

- A. Clay soil with pH 5.1–6.3
- B. Neutral soil with pH 6.5–7.1
- C. Alkaline soil with pH 7.5–8.1
- D. Highly acidic soil with pH 4.5–5.8
- E. Sandy soil with pH 6.0–6.5

Answer D. Highly acidic soil with pH 4.5–5.8

Reasoning

Blueberry plants thrive in well-drained, highly acidic soil with a pH between 4.5 and 5.8.

Crop Blueberry

Figure 18. **Crop Management Task (CMN) Example (B).**

**Question**

What machine was used to achieve the current state of the field shown in the image?

Options

- A. Disc plough
- B. Rotary tiller
- C. Spring-tooth harrow
- D. Cambridge roller
- E. Subsoiler

Answer D. Cambridge roller

Reasoning

The Cambridge roller is designed to press down soil and break up large clumps, creating a smooth and compacted surface. This outcome is distinct from the effects of harrows, ploughs, or subsoilers, which focus more on loosening or turning the soil.

Machine Name Cambridge roller

Figure 19. Machine Usage QA Task (MQA) Example (A).

**Question**

What is the purpose of the equipment shown in the image?

Options

- A. To add a fertilizer spreader for crop growth
- B. To attach additional equipment for planting
- C. To increase the tilling depth in hard soil
- D. To adjust the center of gravity and improve balance
- E. To mark and guide the position for precise planting

Answer D. To adjust the center of gravity and improve balance

Reasoning

The balance weight on the front of the tiller is designed to shift the center of gravity, providing better stability and control when operating the machine.

Machine Name Tiller

Figure 20. Machine Usage QA Task (MQA) Example (B).

**Question**

At what stage of growth is the method shown in the image typically performed?

Options

- A. During the early growth stages after the rice seeds sprout
- B. When the rice plants start bearing grains
- C. After the rice grains have matured
- D. Before the rice plants reach full height
- E. Right before the rice is ready for harvest

Answer A. During the early growth stages after the rice seeds sprout

Reasoning

The method shown in the image is duck rice farming, implemented during the early stages to allow ducks to help control weeds and pests naturally, without damaging the rice grains.

Category Duck rice farming

Figure 21. **Traditional Methods (TM) Example (A).**

**Question**

What is the agricultural method shown in the picture?

Options

- A. Irrigation canals
- B. Terrace farming
- C. Chinampa
- D. Slash-and-burn agriculture
- E. Hydroponics

Answer C. Chinampa

Reasoning

The chinampa method involves creating small, rectangular plots of land on shallow lake beds for agriculture. These plots, often referred to as 'floating gardens,' provide fertile soil and retain moisture, making them ideal for crop cultivation in wetland areas.

Category Chinampa

Figure 22. **Traditional Methods QA (TM) Example (B).**

Class	Tools / Techniques
System	Agroforestry, Chinampa, Crop rotation, Duck rice method, Floating Agriculture, Milpa System, Rice-fish system, Shade-grown coffee, Three Sisters method
Landscape	Andes' steep, Below Sea Level Farming, Moray Circle Terrace, Polder, Terrace Farming, Xinghua Duotian
Irrigation	Bisse d'Ayent, Clay Pot Irrigation, Drainage system, Furrow irrigation, Irrigation canal, Level Basin Irrigation, Noria, Pattern tile drainage, Qanat, Rainwater Harvesting Pit, Reservoir irrigation, Subak, Turpan Karez water system, Waterladder pump
Soil	Composting, Controlled burning, Erosion barriers, Hugelkultur, Jhum Cultivation, Key-line design, Ridge and Furrow, Slash-and-burn, Sloping Agricultural Land Technology, Zai pits
Processing	Basket press, Cotton gin, Flail, Grape-treading, Matcha Stone Mill, Metate, Mortar and pestle, Rice Sieving, Sheaf, Stook, Threshing machine, Treadmill, Watermill, Winnowing
Tool	Digging stick, Foot plough, Hoe, Horse-ploughing, Machete, Ox-ploughing, Pitchfork, Rice Field Marker, Sickle, Wheelbarrow, Yoke
Storage	Hasa Drying, High-floored storage, Karausu, Mangoku, Para-para drying rack, Tomi, Traditional fruit drying
Pest Practice	Botanical Pest Control, Scarecrows, Shishi-odoshi, Stork-friendly farming
Hand Planting, Tea-picking	

Table 3. Categorized List of Traditional Methods (TM)



Question

Which environmental condition increases the susceptibility of the plants to the disease shown in the image?

Options

- A. Shady conditions with low humidity.
- B. Cool, dry conditions throughout the season.
- C. Windy conditions with high sunlight exposure.
- D. Low soil potassium levels.
- E. High soil nitrogen content.

Ground truth: E **Labels:** High soil nitrogen content.

GPT-4o: E

Figure 23. **Example case of GPT-4o answering without an input image (DMN).** Even when the model cannot determine the crop or disease based on the text, it guesses the answer.



Question

When is the crop shown in the image ready for harvest in the home garden?

Options

- A. As soon as the leaves are large enough to use.
- B. When the plant reaches a height of 50 cm.
- C. Once the plant has developed a strong root system.,
- D. When the leaves begin to turn yellow.
- E. After consistent watering and adequate sunlight.

Ground truth: A **Labels:** As soon as the leaves are large enough to use.

GPT-4o: A

Figure 24. **Example case of GPT-4o answering without an input image (CMN).** Even when the model cannot determine the crop based on the text, it guesses the answer.

Category	Machinery
Harvester	Aquatic Weed Harvester, Bean Harvester, Beet Harvester, Carrot Harvester, Coffee Bean Harvester, Combine, Corn Harvester, Daikon Radish Harvester, Forage Harvester, Grape Harvester, Harvester, Haulm Topper, Mechanical Tree Shaker, Onion Harvester, Peanut Harvester, Potato Harvester, Pumpkin Picking Machine, Reaper-Binder, Stripper, Sugarcane Harvester, Swather, Tomato Harvester, Yam Harvester
Seeder	Autonomous Seeder, Broadcast Seeder, Seeding Machine, Semi-Automatic Seeder
Transplanter	Cabbage Transplanter, Transplanter
Planter	Planter, Potato Planter
Handling	Bale Gripper, Bale Sledge, Claw Machine, Packing Machine, Sorting Machine
Transport	Chaser Bin, Conveyor, Dump Trailer, Electric Cart, Gravity Wagon, Mother Bin
Tillage	Cultivator, Disc Harrow, Drag Harrow, Excavator, Motorized Plough, Reversible Plough, Spading Machine, Spring-Tooth Harrow, Strip-Till Implement, Tiller
Soil Preparation	Cambridge Roller, Destoner, Land Imprinter, Potato Bed Former, Puddling Machine, Ridge Maker, Ridging Hiller, Rotary Hiller, Smooth Roller
Spreader	De-Icing Agent Spreader, Fertilizer Spraying Robot, Fertilizer Spreader, Manure Spreader, Silage Spreader, Sprayer
Irrigation	Irrigation Machine, Sprinkler Irrigation, Water Wheel
Chopper	Green Manure Chopper, Leaf Chopper, Root Cutting Machine, Shredder, Straw Shredder
Baling	Roll Baler, Tree Baler
Weeder	Tine Weeder, Weeder Machine
Processing	Beet Cleaner Loader, Fanner, Rice Huller, Rice Milling Machine, Sugarcane Press, Threshing Machine
Pruner	Grape Pruning Machine
Rock Removal	Rock Windrower, Stone Picker
Power Equipment	PTO
Autonomous System	Smart Robotic Farmer, Unmanned Helicopter
Facility Equipment	Multi-Tunnel, Vertical Farming System
Planting Support	Border Coating Machine, Seed-Counting Machine, Seedling Machine
Mower	Flail Mower, Mower

Table 4. Categorized List of Machine QA (MQA)



Question

What pest is affecting the plant?

Options

- A. Apple Maggot
- B. Bronzed Cutworm
- C. Stinkbugs
- D. Black Cutworm
- E. Oriental Fruit Moth

Ground truth: B **Labels:** Bronzed Cutworm

GPT-4o (without CoT): D ❌

GPT-4o (with Zero-shot CoT)

To identify the pest affecting the plant, let's analyze the image step by step:

1. ****Examine the Insect****: The image shows a moth with a brownish color and distinct wing patterns.
2. ****Compare with Options****:
 - ****Apple Maggot****: Typically a fly, not a moth.
 - ****Bronzed Cutworm****: A type of moth, often brownish, which matches the image.
 - ****Stinkbugs****: These are not moths; they have a shield-like shape.
 - ****Black Cutworm****: Also a moth, but usually darker in color.
 - ****Oriental Fruit Moth****: A moth, but generally smaller and with different wing patterns.
3. ****Conclusion****: The insect in the image resembles the Bronzed Cutworm due to its color and wing pattern. Therefore, the pest affecting the plant is likely the Bronzed Cutworm (B). ✅

Figure 25. **Example case of zero-shot CoT success.** The model identifies the pest step by step by observing the image's appearance, checks all the options, and concludes with the correct answer.



Question

Which of the following is the most effective management practice for the disease shown in the image?

Options

- A. Prune infected branches to stop disease spread.
- B. Cover the infected areas to prevent further spread.
- C. Maintain a humid climate to support plant growth.
- D. Keep the temperature warm throughout the year.
- E. Removal of infected trees from the orchard.

Ground truth: E **Labels:** Removal of infected trees from the orchard.

GPT-4o (without CoT): E ✓

GPT-4o (with One-shot CoT)

To determine the most effective management practice for the disease shown in the second image, let's analyze the options:

1. ****Identify the Disease****: The second image shows symptoms consistent with a common plant disease affecting trees, such as fire blight or a similar bacterial or fungal infection.
2. ****Evaluate Options****:
 - ****A. Prune infected branches to stop disease spread****: Pruning is a common method to control the spread of diseases in trees by removing infected parts.
 - ****B. Cover the infected areas to prevent further spread****: Covering might not be effective for diseases that spread internally or through vectors.
 - ****C. Maintain a humid climate to support plant growth****: Humidity can often exacerbate fungal and bacterial diseases.
 - ****D. Keep the temperature warm throughout the year****: Temperature control is not typically a direct method for managing tree diseases.
 - ****E. Removal of infected trees from the orchard****: This is a drastic measure often used when the disease is severe and cannot be controlled by other means.
3. ****Select the Best Option****:
 - ****A. Prune infected branches to stop disease spread**** is often the first line of defense to manage and control the spread of diseases in trees. It is less drastic than removing entire trees and can effectively manage the disease if caught early.

Therefore, the most effective management practice for the disease shown in the image is ****A. Prune infected branches to stop disease spread****. ✗

Figure 26. **Example of one-shot CoT failure.** The model focuses on determining the plant disease species but fails to observe the severity of the disease, leading to an incorrect conclusion.



Question

When is the best time to plant the seeds of the plant shown in the image for a harvest?

Options

- A. In mid-Summer for a late harvest.
- B. 2 months before the first frost.
- C. As soon as the soil is workable in the Spring.
- D. 1 month before the last frost of the Spring.
- E. After the first frost.

Ground truth: B **Labels:** 2 months before the first frost.

GPT-4o (with Zero-shot CoT)

The plant in the image is a rutabaga. Let's consider the options:

1. **Rutabagas** are cool-season crops that thrive in cooler temperatures.
2. **Option A: In mid-Summer for a late harvest.**
 - This is a common practice for rutabagas, as they are often planted in mid-summer for a fall harvest.
3. **Option B: 2 months before the first frost.**
 - This could work, as rutabagas need time to mature before the frost, but it might be cutting it close depending on the growing season.
4. **Option C: As soon as the soil is workable in the Spring.**
 - Rutabagas can be planted in early spring, but they are more commonly planted for fall harvests.
5. **Option D: 1 month before the last frost of the Spring.**
 - This might be too early for a fall harvest, as rutabagas prefer cooler weather as they mature.
6. **Option E: After the first frost.**
 - This is not suitable, as rutabagas need time to grow before the frost.

Considering these points, **Option A: In mid-Summer for a late harvest** is the best choice for planting rutabagas to ensure they mature in cooler weather."

(Successfully determines that the plant is Rutabagas. However, the model does not know the optimal timing for planting the seeds.)

Figure 27. **Example of Lack of knowledge Error (CMN).** The model successfully determines that the plant is Rutabagas. However, the model does not know the optimal timing for planting the seeds.



Question

What disease is affecting the plant?

Options

- A. Botrytis leaf blight
- B. Bacterial leaf streak and bulb rot
- C. Pink root
- D. Purple blotch
- E. Enterobacter bulb decay

Ground truth: B **Labels:** Bacterial leaf streak and bulb rot

GPT-4o (with Zero-shot CoT)

To identify the disease affecting the plant, let's consider the characteristics of each option:

- A. **Botrytis leaf blight**: Typically causes grayish mold and blight on leaves.
- B. **Bacterial leaf streak and bulb rot**: Often shows as water-soaked streaks on leaves and can lead to bulb rot.
- C. **Pink root**: Causes pink to purple discoloration of roots, not typically affecting leaves directly.
- D. **Purple blotch**: Characterized by purple lesions with concentric rings on leaves.
- E. **Enterobacter bulb decay**: Primarily affects bulbs, causing decay, not usually visible on leaves.

The image shows a lesion with a distinct pattern on the leaf. The description of **Purple blotch** (D) matches the symptoms seen in the image, with purple lesions and concentric rings.

Therefore, the disease affecting the plant is likely **D. Purple blotch**.

(The image does not show purple lesions. The affected areas are yellow and brown in color.)

Figure 28. **Example of Perceptual Error (DID).** The image shows the bacterial leaf streak and bulb rot, and the correct answer is B. The image does not show purple lesions. The affected areas are yellow and brown.



Question

What species of weed does the highlighted bounding box represent in the image?

Options

- A. Sinapis arvensis (Charlock)
- B. Common wild oat (Avena fatua)
- C. Chenopodium album (Fat-hen)
- D. Purple crabgrass (Digitaria sanguinalis)
- E. Wall barley (Hordeum murinum)

Ground truth: B **Labels:** Common wild oat (Avena fatua)

GPT-4o (with Zero-shot CoT)

To identify the species of weed in the highlighted bounding box, let's analyze the characteristics visible in the image:

1. **Leaf Shape and Arrangement**: The plant has long, narrow leaves that are typical of grasses.
2. **Growth Habit**: The plant appears to be growing in a spreading manner, which is common for certain grass species.
3. **Color and Texture**: The leaves are green and have a smooth texture.

Now, let's compare these observations with the options provided:

- A. **Sinapis arvensis (Charlock)**: This is a broadleaf plant with yellow flowers, not a grass.
- B. **Common wild oat (Avena fatua)**: This is a grass with long, narrow leaves, similar to the plant in the image.
- C. **Chenopodium album (Fat-hen)**: This is a broadleaf plant, not a grass.
- D. **Purple crabgrass (Digitaria sanguinalis)**: This is a grass with a spreading growth habit and narrow leaves, matching the plant in the image.
- E. **Wall barley (Hordeum murinum)**: This is a grass, but it typically has a more upright growth habit.

Based on the characteristics observed, the plant in the image is most likely **D. Purple crabgrass (Digitaria sanguinalis)**.

(Successfully guesses that B and D are close to the image; however, it concludes that D is correct without further observation or reasoning.)

Figure 29. **Example of Perceptual Error (DID).** The image shows the Common wild oat (Avena fatua), and the correct answer is B. The model successfully guesses that B and D are close to the image; however, it concludes that D is correct without further observation or reasoning.



Question

What disease is affecting the plant?

Options

- A. Varnish spot
- B. Leaf drop
- C. Head rot
- D. Big vein
- E. Marginal leaf blight

Ground truth: A **Labels:** Varnish spot

GPT-4o (with Zero-shot CoT)

To identify the disease affecting the plant, let's analyze the symptoms visible in the image:

1. ****Varnish Spot****: Typically appears as small, dark, varnish-like spots on the leaves.
2. ****Leaf Drop****: Involves leaves dropping off the plant, often without significant discoloration.
3. ****Head Rot****: Characterized by decay and rot in the head of the plant, often with dark, mushy areas.
4. ****Big Vein****: Causes enlarged veins and chlorosis in the leaves.
5. ****Marginal Leaf Blight****: Shows as browning or blight along the edges of the leaves.

The image shows dark, decayed areas on the head of the plant, which is consistent with ****Head Rot**** (C).

(Successfully reasoning A and C leads to dark areas, but by ignoring the reasoning, the model concludes that C is the correct answer.)

Figure 30. **Example of Reasoning Error (DID)**. Successfully reasoning A and C leads to dark areas, but by ignoring the reasoning, the model concludes that C is the correct answer.)



Question

Why is it important to perform this machine?

Options

- A. To prepare the field by transplanting rice seedlings uniformly.
- B. To enhance soil fertility by mixing in organic matter.
- C. To prevent water from leaking out.
- D. To promote faster drainage of excess water from the field before planting.
- E. To aerate the soil and increase the amount of oxygen available to plant roots.

Ground truth: C **Labels:** To prevent water from leaking out.

GPT-4o (with Zero-shot CoT)

To determine the purpose of the machine in the image, let's analyze the options step by step:

- A. ****To prepare the field by transplanting rice seedlings uniformly.****
 - This option involves planting rice seedlings, which typically requires a rice transplanter. The machine in the image appears to be more focused on soil preparation or management rather than planting.
 - B. ****To enhance soil fertility by mixing in organic matter.****
 - This could be a function of a plow or tiller, which mixes organic matter into the soil. However, the machine in the image seems to be creating a trench or furrow, which is not primarily for mixing organic matter.
 - C. ****To prevent water from leaking out.****
 - The machine appears to be creating a trench or furrow, which can be used to manage water flow and prevent water from leaking out of a field. This is a likely purpose.
 - D. ****To promote faster drainage of excess water from the field before planting.****
 - Creating trenches or furrows can also help with drainage, allowing excess water to flow out of the field. This is another possible purpose.
 - E. ****To aerate the soil and increase the amount of oxygen available to plant roots.****
 - Aerating the soil is typically done with a different type of equipment, such as an aerator, which creates small holes in the soil. The machine in the image is not designed for this purpose.
- Based on the image and the options, the most likely purposes of the machine are C and D, as it seems to be creating trenches or furrows to manage water flow. *(The model stops thinking and provides both C and D as answers.)*

Figure 31. **Example of Double Answer Error (MQA).** The image shows the border coating machine with the correct answer: C. However, the model stops thinking and provides both C and D as answers, even though it should continue reasoning to determine a single correct answer.



Question

What pest is affecting the plant?

Options

- A. Alfalfa Weevil
- B. Banana Weevil
- C. Pepper Weevil
- D. Water Fern Weevil
- E. Guava Weevil

Ground truth: D **Labels:** Water Fern Weevil

GPT-4o (with Zero-shot CoT)

To identify the pest affecting the plant, let's analyze the characteristics of the weevil in the image:

1. **Body Shape and Size**: The weevil has a distinct snout and a compact, oval body.
2. **Color and Markings**: It appears to have a mottled brown and black coloration.

Now, let's consider the options:

- A. **Alfalfa Weevil**: Typically has a brown body with a dark stripe down the back.
- B. **Banana Weevil**: Generally dark brown to black and more elongated.
- C. **Pepper Weevil**: Smaller, with a more uniform dark color.
- D. **Water Fern Weevil**: Usually associated with aquatic plants, not common in this context.

(Even though the image was taken in a lab environment, the model considers this option incorrect because the background is not water.)

- E. **Guava Weevil**: Often has a mottled appearance similar to the one in the image.

Based on the appearance, the weevil in the image most closely resembles the **Guava Weevil** (E).

Figure 32. **Example of Interpretation Misunderstanding (PID)**. The image shows the Water Fern Weevil, and the correct answer is D. However, even though the image was taken in a lab environment, the model considers this option incorrect because the background is not water.

References

- [1] Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, Jianxin Ma, Rui Men, Xingzhang Ren, Xuancheng Ren, Chuanqi Tan, Sinan Tan, Jianhong Tu, Peng Wang, Shijie Wang, Wei Wang, Shengguang Wu, Benfeng Xu, Jin Xu, An Yang, Hao Yang, Jian Yang, Shusheng Yang, Yang Yao, Bowen Yu, Hongyi Yuan, Zheng Yuan, Jianwei Zhang, Xingxuan Zhang, Yichang Zhang, Zhenru Zhang, Chang Zhou, Jingren Zhou, Xiaohuan Zhou, and Tianhang Zhu. Qwen technical report. *arXiv preprint arXiv:2309.16609*, 2023. [1](#)
- [2] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. In *NeurIPS*, 2023. [1](#)
- [3] Haotian Liu, Chunyuan Li, Yuheng Li, Bo Li, Yuanhan Zhang, Sheng Shen, and Yong Jae Lee. Llava-next: Improved reasoning, ocr, and world knowledge, 2024. [1](#)
- [4] OpenAI. Gpt-4o, 2024. Accessed: 2024-11-13. [1](#)
- [5] OpenAI. Gpt-4o mini, 2024. Accessed: 2024-11-13. [1](#)
- [6] Quan Sun, Qiyang Yu, Yufeng Cui, Fan Zhang, Xiaosong Zhang, Yueze Wang, Hongcheng Gao, Jingjing Liu, Tiejun Huang, and Xinlong Wang. Generative pretraining in multimodality. *arXiv preprint arXiv:2307.05222*, 2023. [1](#)
- [7] Google Gemini Team. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context, 2024. [1](#)
- [8] Weihan Wang, Qingsong Lv, Wenmeng Yu, Wenyi Hong, Ji Qi, Yan Wang, Junhui Ji, Zhuoyi Yang, Lei Zhao, Xixuan Song, Jiazheng Xu, Bin Xu, Juanzi Li, Yuxiao Dong, Ming Ding, and Jie Tang. Cogvlm: Visual expert for pretrained language models. In *NeurIPS*, 2024. [1](#)



Question

What is a recommended method to prevent the spread of the disease shown in the image?

Options

- A. Dig a trench around diseased plants to retain run-off rain water.
- B. Water diseased plants regularly to reduce the spread of disease.
- C. Continue growing crops around diseased plants without any changes.
- D. Compost the diseased plants to reuse as organic material.
- E. Move diseased plants to a different location.

Ground truth: A **Labels:** Water Dig a trench around diseased plants to retain run-off rain water.

GPT-4o (with Zero-shot CoT)

To determine the best method to prevent the spread of the disease shown in the image, let's analyze each option:

- A. ****Dig a trench around diseased plants to retain run-off rain water.****
 - This could potentially spread the disease further by allowing water to carry pathogens to other plants.
- B. ****Water diseased plants regularly to reduce the spread of disease.****
 - Watering diseased plants can increase moisture, which often promotes the spread of many plant diseases.
- C. ****Continue growing crops around diseased plants without any changes.****
 - This would likely allow the disease to spread to healthy plants.
- D. ****Compost the diseased plants to reuse as organic material.****
 - Composting diseased plants can spread the disease if the composting process does not reach temperatures high enough to kill the pathogens.
- E. ****Move diseased plants to a different location.****
 - Moving diseased plants can spread the disease to new areas.

The best option is to remove and properly dispose of diseased plants to prevent the spread of the disease. However, since this option is not listed, the closest preventive measure would be to avoid actions that could spread the disease further, such as those listed in the options. Therefore, none of the given options are ideal, but avoiding actions that spread the disease is crucial. *(Concludes that there is no answer and stops reasoning instead of reconsidering)*

Figure 33. **Example of Reject to Answer (DMN).** The image shows the blood disease of the banana, and the correct answer is A. However, the model concludes that there is no answer and stops reasoning instead of reconsidering.