
SAGE: Scalable Agentic Grounded Evaluation for Crop Disease Diagnosis

Abstract

Plant disease diagnosis at scale is limited by the lack of datasets that cover enough crops and diseases. We compile a dataset of approximately 1.1M images spanning 53 crops and 259 disease classes, and complement it with a pipeline that automatically generates structured symptom descriptions with per-field provenance, where every fact traces back to a web source with a verbatim supporting quote. We use this combination to build a diagnostic system where each prediction is made by an autonomous reasoning agent that follows a guided chain of thought: it identifies the anatomical context of the test image, narrows candidates using symptom knowledge, and sequentially compares reference images, producing a full reasoning trace at every step. Adding symptom knowledge improves accuracy by an average of 15.2 percentage points at full reference budget, with consistent gains across all three evaluation crops. The system requires no task-specific training, works for any crop with reference images, and will directly benefit from future improvements in foundation model capabilities.

1 Introduction

Infectious diseases reduce potential crop yields by an average of 40%, with many farmers in developing regions experiencing losses as high as 100% [1]. Accurate diagnosis is a prerequisite for effective treatment, yet it remains bottlenecked by two gaps: no existing dataset covers enough crops and diseases to support diagnosis at scale, and no structured, verifiable symptom knowledge is available alongside the images.

Existing plant disease image datasets have advanced the field but remain limited in scope. PlantVillage [1] covers common crops under controlled conditions; PlantDoc [2] adds real-world images but for only 13 species; LeafNet [3] and CDDM [4] bring scale and multimodal annotations but focus predominantly on leaf images. No existing dataset combines broad crop coverage, multi-organ imagery (leaf, stem, root, seed, ear, head), and structured symptom metadata with source provenance. This gap limits both the training of generalizable models and the development of diagnostic systems that can explain their reasoning in terms of verifiable disease knowledge.

On the methods side, deep learning has achieved high accuracy on controlled benchmarks [5], and vision-language models have shown promise for few-shot disease recognition [6, 7]. Recent work has begun to explore chain-of-thought reasoning [8, 9] and agentic architectures [10] for plant disease diagnosis. However, these approaches either operate on narrow crop sets, require fine-tuning, or lack the ability to produce transparent, step-by-step diagnostic reasoning grounded in traceable knowledge sources.

We address these gaps together. We compile one of the largest plant disease image datasets, complement it with automatically generated symptom descriptions that trace back to their sources, and show how this combination enables an agentic reasoning system to diagnose diseases across crops with full transparency and no task-specific training. Specifically, we present:

1. A multi-crop image dataset spanning 53 crops and 259 disease classes ($\sim 1.1\text{M}$ images), assembled from established benchmarks, expert-curated collections, and community sources, with multi-organ coverage.
2. A source-first disease registry pipeline that, given a crop name, automatically produces structured symptom knowledge with per-field provenance, where every fact traces back to a specific web source with a verbatim supporting quote.
3. An agentic diagnostic pipeline where each prediction is made by an autonomous reasoning agent that produces a full, human-readable reasoning trace showing which references were examined and why.
4. A systematic evaluation across three crops of varying difficulty, multiple reference budgets, knowledge base sources, and model tiers.

2 SAGE: Scalable Agentic Grounded Evaluation

2.1 Problem Setup

Let $\mathcal{C} = \{c_1, \dots, c_n\}$ denote the set of crops and $\mathcal{D} = \{d_1, \dots, d_m\}$ denote the set of known diseases. Each disease $d \in \mathcal{D}$ has:

- Affected organs: $\mathcal{O}_d \subseteq \mathcal{O}$, where \mathcal{O} is the set of all plant organs (leaf, stem, root, etc.)
- Symptom descriptions: $\mathcal{S}_d = \{s_1, s_2, \dots\}$, each with provenance $p(s_i)$ linking to a source quote.

Let $\mathcal{I} = \{I_1, \dots, I_q\}$ denote web images. The diagnostic function is defined as:

$$f_{\text{SAGE}} : I_{\text{test}} \times \text{KB} \times \mathcal{R} \rightarrow (d^*, c, \tau)$$

where d^* is the predicted disease, c is the confidence, and τ is the reasoning trace.

2.2 Phase 1: Knowledge Base Curation

Discovery Agent: For each $(c, d) \in \mathcal{C} \times \mathcal{D}$, issue web queries:

$$\mathcal{W}_{c,d} = \text{WebQuery}(c, d)$$

Extraction Agent: Extract structured facts from sources:

$$\mathcal{F}_w = \{(p, t, \mathcal{O}, \mathcal{S})\}, \quad \forall w \in \mathcal{W}_{c,d}$$

Attach provenance $p(s)$ to each symptom $s \in \mathcal{S}$.

Reconciliation Agent: Resolve conflicts:

$$\text{KB} = \text{Reconcile}\left(\bigcup_w \mathcal{F}_w\right)$$

Filtering & Tagging VLM: For each image $I \in \mathcal{I}$:

$$\hat{o} = \text{DetectOrgan}(I), \quad m(I, d) = \text{MatchSymptoms}(I, \mathcal{S}_d)$$

Keep images where $m(I, d) \geq \theta$ and assign anatomical tags.

2.3 Phase 2: Agentic Inference

Observation:

$$o_{\text{test}} = \text{DetectOrgan}(I_{\text{test}}), \quad \mathcal{S}_{\text{test}} = \text{DescribeSymptoms}(I_{\text{test}})$$

Narrow Candidates:

$$\mathcal{D}_{\text{cand}} = \{d \in \mathcal{D} \mid o_{\text{test}} \in \mathcal{O}_d\}, \quad \mathcal{D}_{\text{rank}} = \text{RankBySymptoms}(\mathcal{D}_{\text{cand}}, \mathcal{S}_{\text{test}})$$

Sequential Comparison Loop: Given reference budget k , for $i = 1 \dots k$:

1. Select candidate disease $d_i = \text{NextCandidate}(\mathcal{D}_{\text{rank}})$
2. Fetch reference image $I_{\text{ref}} = \text{FetchReferenceImage}(\mathcal{R}, d_i, o_{\text{test}})$
3. Compare and reason:

$$r_i = \text{CompareAndReason}(I_{\text{test}}, I_{\text{ref}}, \mathcal{S}_{d_i})$$

Update reasoning trace $\tau \leftarrow \tau \cup \{r_i\}$

Remove rejected candidates until confident.

Prediction:

$$d^* = \arg \max_{d \in \mathcal{D}_{\text{rank}}} \text{Support}(d, \tau), \quad c = \text{Confidence}(d^*, \tau)$$

Algorithm 1 SAGE: Agentic Inference

Require: Test image I_{test} , KB, Reference set \mathcal{R} , Anatomical Index, Reference budget k

Ensure: Predicted disease d^* , confidence c , reasoning trace τ

```

1:  $o_{\text{test}} \leftarrow \text{DetectOrgan}(I_{\text{test}})$ 
2:  $\mathcal{S}_{\text{test}} \leftarrow \text{DescribeSymptoms}(I_{\text{test}})$ 
3:  $\mathcal{D}_{\text{cand}} \leftarrow \{d \in \mathcal{D} \mid o_{\text{test}} \in \mathcal{O}_d\}$ 
4:  $\mathcal{D}_{\text{rank}} \leftarrow \text{RankBySymptoms}(\mathcal{D}_{\text{cand}}, \mathcal{S}_{\text{test}})$ 
5:  $\tau \leftarrow []$ 
6: for  $i = 1$  to  $k$  do
7:   if  $\text{Confident}(\mathcal{D}_{\text{rank}})$  then
8:     break
9:   end if
10:   $d_i \leftarrow \text{NextCandidate}(\mathcal{D}_{\text{rank}})$ 
11:   $I_{\text{ref}} \leftarrow \text{FetchReferenceImage}(\mathcal{R}, d_i, o_{\text{test}})$ 
12:   $r_i \leftarrow \text{CompareAndReason}(I_{\text{test}}, I_{\text{ref}}, \mathcal{S}_{d_i})$ 
13:   $\tau.\text{append}(r_i)$ 
14:  if  $r_i$  rejects  $d_i$  then
15:     $\mathcal{D}_{\text{rank}}.\text{remove}(d_i)$ 
16:  end if
17: end for
18:  $d^* \leftarrow \arg \max_{d \in \mathcal{D}_{\text{rank}}} \text{Support}(d, \tau)$ 
19:  $c \leftarrow \text{Confidence}(d^*, \tau)$ 
20: return  $d^*, c, \tau$ 

```

3 Related Work

The automatic identification of plant diseases from images has been studied extensively using deep learning. Convolutional neural networks trained on labeled image datasets have achieved high accuracy on controlled benchmarks, with architectures such as ResNet, VGG, and DenseNet commonly applied to leaf disease classification [5]. However, these models require large labeled training sets for each new crop or disease and produce opaque predictions: a farmer receives a label but no explanation of what visual features led to it.

The data landscape has evolved in parallel. PlantVillage [1] established the first large-scale open repository of plant disease images, covering 38 classes across common crops. PlantDoc [2] extended this with real-world field images, while PlantWild [7] introduced in-the-wild images paired with text descriptions for multimodal recognition. More recently, LeafNet [3] provides 186K images across 97 classes with a vision-language benchmark, and CDDM [4] offers 137K images with 1M question-answer pairs for conversational diagnosis. AgMMU [11] takes a different approach, constructing a benchmark from real extension expert dialogues that spans identification, symptom description, and management across multiple crops. Despite this progress, existing datasets are predominantly

Table 1: Comparison of plant disease datasets and benchmarks. **Diseases** counts disease-positive classes only (healthy/pest/weed classes excluded). **Task**: CLS classification, Det detection, VQA visual QA, OEQ open-ended QA, MCQ multiple choice. **Symptoms knowledge base**: ✓ if per-disease symptom descriptions are paired with images; parenthetical notes the source. **Demonstrated agentic eval**: ✓ if the dataset is shown to support step-by-step diagnostic reasoning with reference image comparison.

Dataset	Host species	Diseases	Images	Task	Multi-organ	Symptoms knowledge base	Demonstrated agentic eval
PlantVillage [1]	14	26	54,306	CLS	✗	✗	✗
PlantDoc [2]	13	17	2,598	CLS, Det	✗	✗	✗
PlantWild [7]	33	56	18,542	CLS	✗	✓ (LLM-generated)	✗
LeafNet [3]	22	97	186,000	VQA	✗	✓ (limited metadata)	✗
CDDM [4]	16	60	137,000	MCQ	✗	✗	✗
AgroBench [30]	203	682	3,745	MCQ	✓	✗	✗
SAGE (ours)	335	1,251	838,936	CLS	✓	✓ (source-grounded)	✓

4 Dataset

4.1 Image Sources

No single existing plant disease dataset covers enough crops and diseases for a general-purpose diagnostic system. We assembled our dataset from four complementary source categories to maximize coverage (Figure 1).

We start with **established benchmarks**: PlantVillage [1], the most widely used plant disease image repository covering common crops under controlled imaging conditions, and PlantDoc [2], which extends this with real-world field images. We supplement these with **recent large-scale datasets**: LeafNet [3], a collection designed for vision-language understanding of plant diseases, and PlantWild [7], which focuses on in-the-wild disease recognition under uncontrolled conditions.

For agriculturally important crops, we include **expert-curated data** with multi-organ coverage (leaf, stem, root, seed, ear, and head), going beyond the leaf-only focus of most public datasets. The soybean stress identification dataset [5] contributes images with expert-verified disease labels, and the mango evaluation subset draws from MangoLeafBD [15]. For crops not well-represented in the above sources, we incorporate **community-contributed datasets** covering rice [16], wheat rust [17], banana [18], bean [19], lettuce [20], cucumber [21], durian [22], eggplant [23], cotton [24], pumpkin [25], rose [26], strawberry [27], sugarcane [28], and chickpea [29].

The combined dataset spans 53 crops and 259 disease classes, totaling approximately 1.1 million images. For the agentic evaluation in this work, we select three crops that represent different scales of difficulty: Soybean (25 disease classes), Corn (30 classes), and Mango (4 classes). Evaluation on additional crops is planned but limited by the compute cost of agentic inference. Table 1 positions SAGE against existing plant disease datasets and benchmarks. Several related works are cited but excluded from the comparison because they pursue different goals: AgMMU [11] benchmarks broader agricultural QA from extension dialogues, and Agri-LLaVA [13] and AgroGPT [12] release instruction-tuned VLMs trained on synthetic conversations.

4.2 Splits and Image Filtering

Because images are aggregated from many sources, the same disease often appears under several different class names (e.g., “Leaf_Rust”, “leaf rust”, and “rust_disease” across three sources). Before any automated filtering, a domain expert reviews the union of class labels across sources and merges variants that refer to the same disease, producing a single canonical class per disease. After this dedupe step, raw image collections still contain noise: mislabeled images, images showing healthy tissue, or images where the disease presentation does not match the documented symptoms. We filter images using the disease registry as ground truth. For each image in a class, a vision-language model receives the image alongside the KB symptom description for that class and determines whether the image is consistent with the description. Matched images are kept; mismatches are rejected.

During filtering, each matched image is also tagged with its *anatomical context*: the primary plant organ visible (leaf, stem, root, seed, pod, or whole plant). This produces an anatomical index that

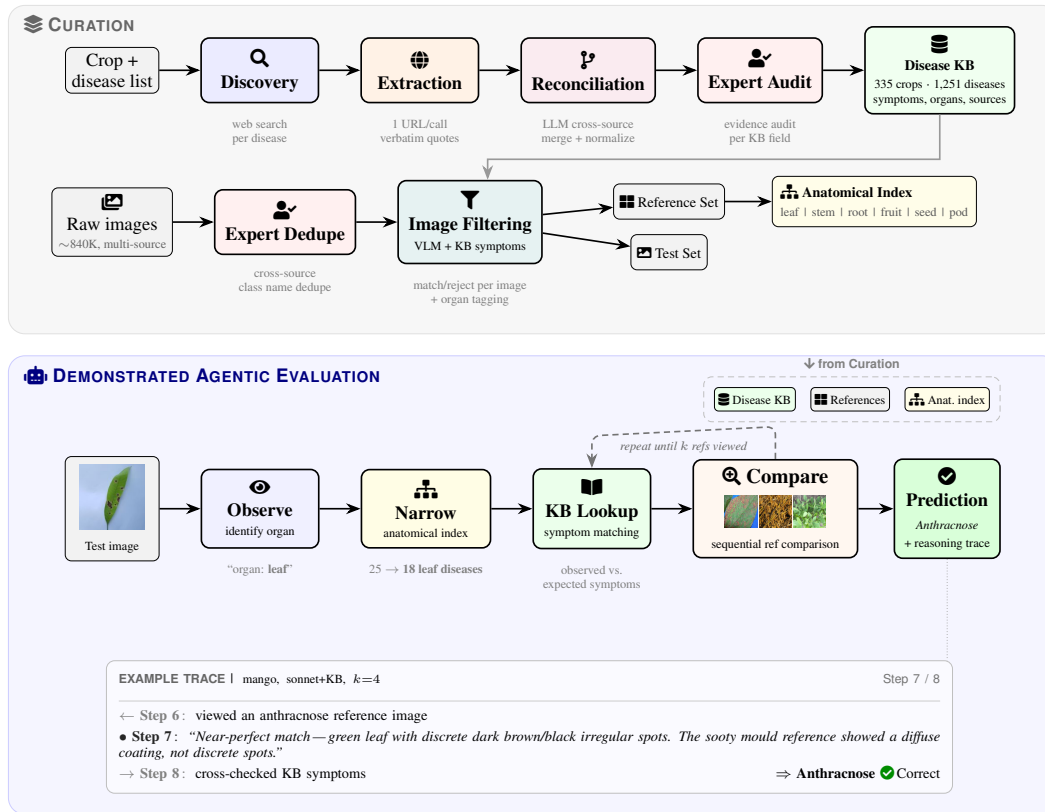


Figure 2: System overview. **Curation** (top): two parallel tracks meet at image filtering. The KB track turns web pages into a source-cited disease KB (335 crops, 1,251 diseases) via automated discovery, extraction, and reconciliation, followed by an *expert evidence audit* of every cited verbatim quote. The image track collects $\sim 840\text{K}$ raw images from multiple sources and applies an *expert dedupe* of class-name variants across sources before any filtering. The deduped images are then filtered against the audited KB and tagged by organ, producing reference/test splits and an anatomical index. **Demonstrated agentic evaluation** (bottom): given a test image and the curated inputs, a VLM agent observes the affected organ, narrows candidates via the anatomical index, consults KB symptoms, and sequentially compares reference images in an open-ended loop bounded by reference budget k , producing a prediction with a full step-by-step trace; the strip shows one real step from such a trace.

maps each organ to the set of diseases that affect it. The agent uses this index at inference to narrow candidate classes based on what it observes in the test image.

The anatomical index is a key output of this step: it maps each organ type to the set of diseases that affect it, enabling the agent to narrow candidates at inference based on the organ visible in the test image. Figure 2 (Curation, top row) illustrates this filtering and tagging process.

Matched images are split into reference and test sets. Reference images are organized by class and organ; test images are held out for evaluation. For the three evaluation crops, this yields Soybean (25 classes, 74 test images), Corn (30 classes, 88 test images), and Mango (4 classes, 40 test images).

5 Disease Registry Pipeline

The diagnostic agent (Section 6) needs structured disease knowledge to reason over images at inference. Querying a language model directly for disease symptoms risks hallucination and produces facts that cannot be verified. Our pipeline instead follows a source-first principle: it fetches real documents from the web, extracts only what is explicitly stated in those documents, and traces every fact back to its source with a verbatim quote. Given a crop name, three automated stages followed by an expert evidence audit (Figure 2) produce a fully cited disease registry.

The pipeline begins by issuing targeted web search queries for each disease in the crop, using the Claude CLI’s web search capability to retrieve authoritative pages such as university extension factsheets, CABI datasheets, and APS disease compendia. Each retrieved page is then passed to the language model with an extraction prompt that requires the model to identify all diseases discussed on the page and extract structured fields: pathogen name, disease type, affected plant organs, and visual symptom descriptions. The model is required to provide a verbatim quote from the source text supporting each extracted field and is explicitly prevented from filling in information from its own knowledge. The raw extractions from multiple sources are then reconciled into a canonical registry: naming inconsistencies are resolved through LLM-based matching, conflicting claims across sources are tracked, and each field in the final registry retains its provenance. A domain expert then audits the reconciled registry field by field, checking that the verbatim quote cited for each symptom actually supports the claim and discarding fields whose evidence is weak or off-topic. Only registries that pass this audit are released to the agent. Once generated for a crop, the reviewed registry is reused across all evaluation runs.

6 Agentic Diagnostic Pipeline

Each test image is diagnosed by an autonomous reasoning agent that operates in a multi-turn loop. The agent receives three inputs: (1) the test image to classify, (2) individually labeled reference images organized by disease class, and (3) optionally, the symptom descriptions from the disease registry (Section 5) together with an anatomical index that maps plant organs to the diseases that affect them. It also receives the full list of candidate disease classes and a reference budget k that limits how many reference images it may examine.

The agent follows a guided chain of thought. It first reads the test image and identifies the anatomical context (leaf, stem, root, seed) and visible symptoms: color changes, lesion shape, texture, and spatial patterns. If the knowledge base is available, the agent consults the anatomical index to narrow candidates to only those diseases that affect the observed organ, then reviews symptom descriptions to further prioritize. It then views reference images one at a time, comparing the test image against each labeled example and reasoning about what matches and what does not before selecting the next reference to examine. This sequential, deliberative comparison ensures the agent explores multiple candidates rather than confirming an initial guess. The agent concludes by producing a structured prediction with a confidence score and a brief explanation of its reasoning.

The full sequence of the agent’s actions and reasoning is captured as a trace. Each trace records which reference images the agent chose to view, the text it produced between views, and its final prediction. These traces are the primary vehicle for explainability: they allow a user to see not just what the system predicted but how it arrived at that prediction, which references it found most informative, and what alternative diagnoses it considered and ruled out. This stands in contrast to single-pass classification, where the model produces a label with no visibility into the decision process. The trace below illustrates this process on a soybean test image at $k=8$ with internet KB. Additional traces are provided in Appendix E.

7 Experiments

We evaluate the diagnostic pipeline on three crops chosen to represent different scales of difficulty: Soybean (25 disease classes), Corn (30 classes), and Mango (4 classes). For each crop, we use 3 test images per class (10 for Mango), selected with a fixed random seed for reproducibility. We vary three axes: the reference budget k (0, 1, 4, 8, 16), the knowledge base source (none or internet), and the model tier (Haiku, Sonnet, Opus). The baseline is the agent at $k=0$ without a knowledge base, where the model receives only the test image and the list of candidate class names with no reference images or symptom descriptions.

Table 2 presents the main results. The $k=0$ column isolates the KB contribution: without any reference images, the agent relies entirely on the test image and symptom descriptions. On Soybean, the KB alone adds 13.5 percentage points (31.1% to 44.6%), and on Corn 9.1 points (42.0% to 51.1%). As the reference budget increases, accuracy improves consistently, with Corn reaching 64.8% at $k=16$ with KB (Figure 3).

Model: sonnet
KB source: internet
Reference budget (k): 8
Prediction: Frogeye_leaf_spot
Ground truth: Frogeye_leaf_spot
Outcome: **Correct**
Confidence: 0.93



Test image

Step 1: Observe test image

Step 2: The test image shows a soybean leaf with small, circular lesions having gray/tan centers and distinct reddish-brown margins. This pattern is very characteristic of Frogeye leaf spot. Let me check the part index and confirm with reference images.

Step 3: View reference: part_index

Step 4: Leaf is the affected part. Top candidates: Frogeye_leaf_spot (gray centers, reddish-brown borders), Septoria_brown_spot, Phyllosticta_leaf_spot. Let me check references.

Step 5: View reference: Soybean_Dise_6

Step 6: Strong match - circular lesions with reddish-brown borders. Now checking 7 more different classes.

Step 7: View reference: Soybean_Dise_12

Step 8: View reference: Soybean_Dise_18

Step 9: View reference: Soybean_Dise_17

Step 10: View reference: Soybean_Dise_74

Step 11: View reference: Soybean_Dise_93

Step 12: View reference: Soybean_Dise_137

Step 13: View reference: Soybean_Dise_62

Step 14: View reference: Soybean_Dise_39

Step 15: All 8 views used. The analysis is clear:

- Test image: Small circular lesions (5-10mm), gray/tan centers, distinct reddish-brown margins, scattered on green leaf - Frogeye_leaf_spot reference: Near-identical match - same circular morphology, same gray centers with reddish-brown borders - All other candidates: Don't match - Septoria shows irregular dark lesions with yellowing; Rust shows tiny numerous specks; Cercospora shows diffuse bronze discoloration; Bacterial Blight shows angular yellow-haloed lesions; SDS shows interveinal chlorosis; Downy mildew shows pale yellow patches; Bacterial Pustule shows confluent dark blotches

Figure 4 shows the cost-accuracy tradeoff, averaging accuracy across all three crops. Small dots show per-image API costs, while large bubbles mark aggregate means at each reference budget. Increasing k improves accuracy but at growing cost, with diminishing returns beyond $k=8$. Model quality provides a complementary lever: at $k=8$, Opus achieves higher mean accuracy than Sonnet at $k=16$ for comparable cost, while Haiku is the cheapest option but substantially less accurate.

Figure 5 shows the confusion matrices for Soybean. In the baseline (left), Sudden_death_syndrome is heavily over-predicted (14 false positives), absorbing predictions from many other classes. With the full pipeline (right), this drops to 3 false positives as the agent uses KB symptoms and reference comparisons to distinguish visually similar diseases. Corn and Mango confusion matrices are in Appendix D.

Table 2: Diagnostic accuracy across crops, methods, and reference budgets k (Sonnet model, 3 test images per class). Parentheses in the Crop column denote number of disease classes. The baseline is Agent (no KB) at $k=0$ (model receives only the test image and class names). Values show accuracy % with improvement over baseline in parentheses. Best per crop- k in **bold**. Bottom rows show mean improvement (pp) over baseline, averaged across crops.

Crop	Method	$k = 0$	$k = 1$	$k = 4$	$k = 8$	$k = 16$
Soybean (25)	Agent (no KB)	31.1% (+0.0)	40.5% (+9.5)	37.8% (+6.8)	45.9% (+14.9)	48.6% (+17.6)
	Agent + internet KB	44.6% (+13.5)	41.9% (+10.8)	48.6% (+17.6)	48.6% (+17.6)	51.4% (+20.3)
Corn (30)	Agent (no KB)	42.0% (+0.0)	43.2% (+1.1)	46.6% (+4.5)	52.3% (+10.2)	54.5% (+12.5)
	Agent + internet KB	51.1% (+9.1)	52.3% (+10.2)	60.2% (+18.2)	60.2% (+18.2)	64.8% (+22.7)
Mango (4)	Agent (no KB)	92.5% (+0.0)	85.0% (-7.5)	92.5% (+0.0)	92.5% (+0.0)	95.0% (+2.5)
	Agent + internet KB	87.5% (-5.0)	97.5% (+5.0)	92.5% (+0.0)	97.5% (+5.0)	95.0% (+2.5)
<i>Mean Δ (pp)</i>	<i>Agent (no KB)</i>	+0.0	+1.0	+3.8	+8.4	+10.9
	<i>Agent + internet KB</i>	+5.9	+8.7	+11.9	+13.6	+15.2

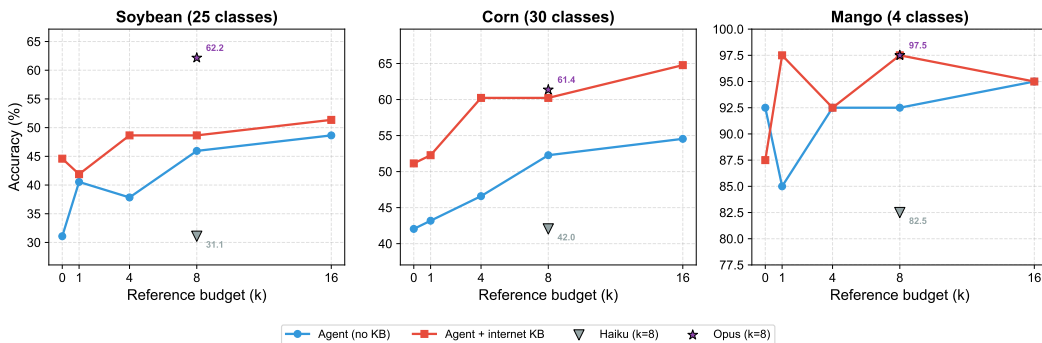


Figure 3: Diagnostic accuracy vs. reference budget k across three crops. Each panel shows the agent with no KB (blue) and with internet KB (red).

8 Discussion and Conclusion

The agent consistently improves over the baseline across all three crops. The knowledge base provides the largest gains at low reference budgets, where symptom descriptions and anatomical context guide the agent to examine the most relevant references first. On Corn, adding the KB improves accuracy by 9.1 to 10.3 percentage points across all k values. On Soybean, the KB contributes 13.5 points at $k=0$ but the gap narrows at higher k as visual evidence becomes sufficient. On Mango, with only 4 classes, visual comparison alone reaches over 90% and the KB provides marginal additional value.

Model quality is the single most impactful factor: upgrading the underlying model yields larger gains than adding a knowledge base or increasing the reference budget. This means the system improves automatically as foundation models improve, without retraining or architectural changes.

We also compare against a single-pass few-shot baseline (Appendix Table 5), where k labeled images are provided in a single context with no reasoning loop. The agent outperforms few-shot at matched k values because the guided chain of thought allows it to selectively compare references rather than processing all examples in a single pass. However, few-shot is a black box: it produces a label with no visibility into the decision process. The agent’s reasoning traces remain a practical advantage regardless of accuracy, allowing users to verify the diagnosis, understand what visual features drove it, and identify failure points.

The evaluation uses 3 test images per class (10 for Mango), which introduces per-class variance. Extending evaluation to all 53 crops is planned but constrained by compute cost. KB quality depends on available web sources, which varies across crops. The agentic approach is also more expensive per image than single-pass alternatives, making it better suited for scenarios where interpretability matters than for high-throughput screening.

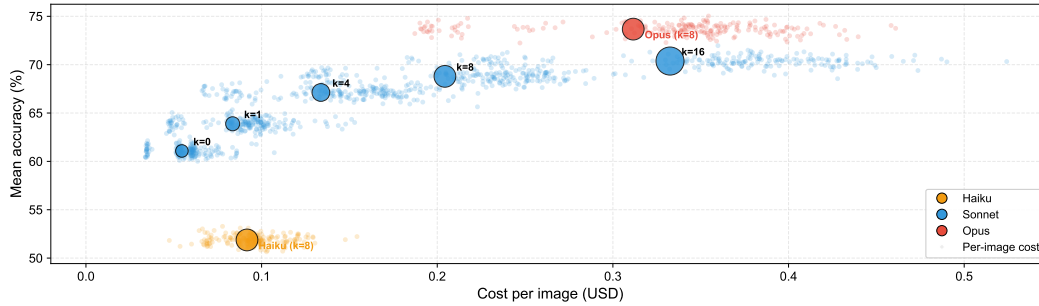


Figure 4: Cost-accuracy tradeoff (mean accuracy across all three crops, internet KB). Small dots show individual per-image API costs; large bubbles show aggregate means with bubble size proportional to reference budget k . Haiku and Opus shown at $k=8$.

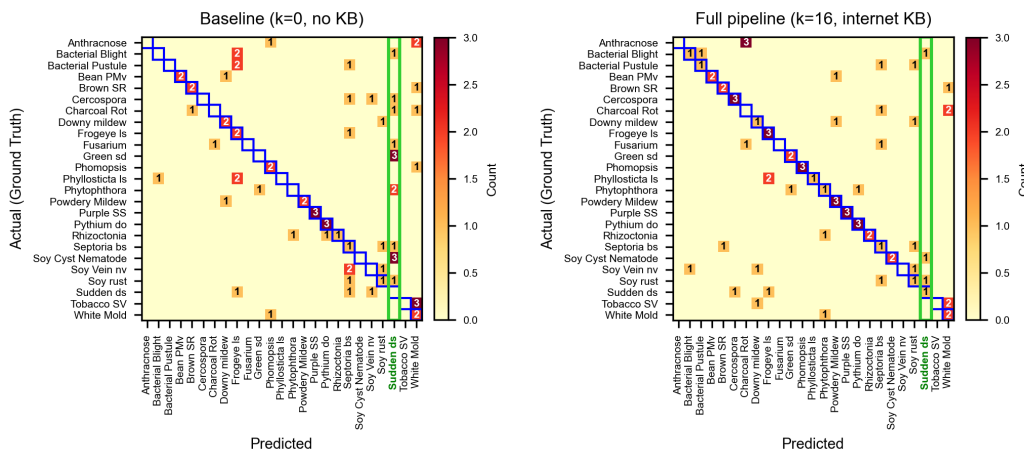


Figure 5: Confusion matrices for Soybean (Sonnet, 25 classes). Left: baseline, $k=0$, no KB (31.1%). Right: full pipeline, $k=16$, internet KB (51.4%). The highlighted Sudden_death_syndrome column shows over-prediction reduced from 14 to 3 false positives.

We presented a system that combines a large-scale disease image dataset, an automated source-cited knowledge base pipeline, and an agentic diagnostic method that reasons transparently over both. The approach requires no task-specific training, generalizes to any crop with reference images, and improves directly as foundation models advance. Extending the evaluation to all 53 crops and improving KB coverage for underrepresented crops are natural next steps.

References

- [1] David P. Hughes and Marcel Salathé. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*, 2015.
- [2] Davinder Singh, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawat, and Nipun Batra. Plantdoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, pages 249–253. Association for Computing Machinery, 2020.
- [3] Khang Nguyen Quoc, Phuong D Dao, and Luyi-Da Quach. Leafnet: A large-scale dataset and comprehensive benchmark for foundational vision-language understanding of plant diseases. *arXiv preprint arXiv:2602.13662*, 2026.
- [4] Xiang Liu, Zhaoxiang Liu, Huan Hu, Zezhou Chen, Kohou Wang, Kai Wang, and Shiguo Lian. A multimodal benchmark dataset and model for crop disease diagnosis. In *European Conference on Computer Vision*, pages 157–170. Springer, 2024.
- [5] Sambuddha Ghosal, David Blystone, Asheesh K Singh, Baskar Ganapathysubramanian, Arti Singh, and Soumik Sarkar. An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences*, 115(18):4613–4618, 2018.
- [6] Muhammad Arbab Arshad, Talukder Zaki Jubery, Tirtho Roy, Rim Nassiri, Asheesh K Singh, Arti Singh, Chinmay Hegde, Baskar Ganapathysubramanian, Aditya Balu, Adarsh Krishnamurthy, et al. Leveraging vision language models for specialized agricultural tasks. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6320–6329. IEEE, 2025.
- [7] Tianqi Wei, Zhi Chen, Zi Huang, and Xin Yu. Benchmarking in-the-wild multimodal disease recognition and a versatile baseline. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 1593–1601, 2024.
- [8] Jiandong Pan, Renhai Zhong, Fulin Xia, Jingfeng Huang, Linchao Zhu, Yi Yang, and Tao Lin. Chatleafdisease: a chain-of-thought prompting approach for crop disease classification using large language models. *Plant Phenomics*, page 100094, 2025.
- [9] Kunpeng Zhang, Li Ma, Beibei Cui, Xin Li, Boqiang Zhang, and Na Xie. Visual large language model for wheat disease diagnosis in the wild. *Computers and Electronics in Agriculture*, 227: 109587, 2024.
- [10] Lufu Qin, Xingcai Wu, Xinyu Dong, Huan Wang, Tingwei Yang, and Qi Wang. Pdd-agent: Multimodal large language model-driven ai agent for enhanced plant disease diagnosis. In *2025 IEEE International Conference on Image Processing (ICIP)*, pages 1271–1276. IEEE, 2025.
- [11] Aruna Gauba, Irene Pi, Yunze Man, Ziqi Pang, Vikram S Adve, and Yu-Xiong Wang. Agmmu: A comprehensive agricultural multimodal understanding and reasoning benchmark. *arXiv preprint arXiv:2504.10568*, 2025.
- [12] Muhammad Awais, Ali Husain Salem Abdulla Alharthi, Amandeep Kumar, Hisham Cholakkal, and Rao Muhammad Anwer. Agrogpt: Efficient agricultural vision-language model with expert tuning. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5687–5696. IEEE, 2025.
- [13] Liqiong Wang, Teng Jin, Jinyu Yang, Ales Leonardis, Fangyi Wang, and Feng Zheng. Agri-llava: Knowledge-infused large multimodal assistant on agricultural pests and diseases. *arXiv preprint arXiv:2412.02158*, 2024.
- [14] Hossein Zaremehrjerdi, Shreyan Ganguly, Ashlyn Rairdin, Elizabeth Tranel, Benjamin Feuer, Juan Ignacio Di Salvo, Srikanth Panthulugiri, Hernan Torres Pacin, Victoria Moser, Sarah Jones, et al. Towards large reasoning models for agriculture. *arXiv preprint arXiv:2505.19259*, 2025.
- [15] Sarder Iftekhar Ahmed, Muhammad Ibrahim, Md Nadim, Md Mizanur Rahman, Maria Mehjabin Shejunti, Taskeed Jabid, and Md Sawkat Ali. Mangoleafbd: A comprehensive image dataset to classify diseased and healthy mango leaves. *Data in Brief*, 47:108941, 2023.

- [16] Isaac Ritharson. Severity-based rice disease classification. <https://www.kaggle.com/datasets/isaacritarson/severity-based-rice-leaf-diseases-dataset>, 2021. Kaggle.
- [17] Tolga Hayit. Yellowrust19: Yellow rust disease in wheat. <https://www.kaggle.com/datasets/tolgahayit/yellowrust19-yellow-rust-disease-in-wheat>, 2020. Kaggle.
- [18] Gimril Lozarita. Banana leaf disease dataset v1.1. <https://www.kaggle.com/datasets/gimrillozarita/banana-leaf-disease-dataset-v1-1>, 2022. Kaggle.
- [19] Marquis03. Bean leaf lesions classification dataset. <https://www.kaggle.com/datasets/marquis03/bean-leaf-lesions-classification>, 2023. Kaggle.
- [20] Ashish Jena. Lettuce diseases dataset. <https://www.kaggle.com/datasets/ashishjstar/lettuce-diseases>, 2024. Kaggle.
- [21] Karim Negm. Cucumber plant diseases dataset. <https://www.kaggle.com/datasets/kareem3egm/cucumber-plant-diseases-dataset>, 2020. Kaggle.
- [22] Cthng123. Durian leaf disease dataset. <https://www.kaggle.com/datasets/cthng123/durian-leaf-disease-dataset>, 2025. Kaggle.
- [23] Kamalmoha. Eggplant disease recognition dataset. <https://www.kaggle.com/datasets/kamalmoha/eggplant-disease-recognition-dataset>, 2023. Kaggle.
- [24] Shuvo Kumar Basak. Cotton disease multi transformation dataset. <https://www.kaggle.com/datasets/shuvokumarbasak2030/cotton-disease-multi-transformation-dataset>, 2026. Kaggle.
- [25] Shuvo Kumar Basak. Pumpkin leaf disease multi transformation dataset. <https://www.kaggle.com/datasets/shuvokumarbasak2030/pumpkin-leaf-disease-multi-transformation-dataset>, 2024. Kaggle.
- [26] Shuvo Kumar Basak. Rose leaf disease multi transformation dataset. <https://www.kaggle.com/datasets/shuvokumarbasak2030/rose-leaf-disease-multi-transformation-dataset>, 2026. Kaggle.
- [27] Usman Afzaal. Strawberry disease detection dataset. <https://www.kaggle.com/datasets/usmanafzaal/strawberry-disease-detection-dataset>, 2021. Kaggle.
- [28] Betty D. Puspasari. Sugarleaf-idn: Sugarcane leaf diseases dataset. <https://www.kaggle.com/datasets/bettydpuspasari/sugarleafidn>, 2026. Kaggle.
- [29] Tolga Hayit. Fusarium wilt disease in chickpea dataset. <https://www.kaggle.com/datasets/tolgahayit/fusarium-wilt-disease-in-chickpea-dataset>, 2022. Kaggle.
- [30] Risa Shinoda, Nakamasa Inoue, Hirokatsu Kataoka, Masaki Onishi, and Yoshitaka Ushiku. AgroBench: Vision-language model benchmark in agriculture. In *arXiv preprint arXiv:2507.20519*, 2025.

A Dataset Licenses

Table 3: Dataset license summary: hosting and distribution rights. Each row lists the license and citation requirements for datasets included in our collection.

Dataset	License	Notes & Citation
SBRD — Rice Leaf Diseases	World Bank ToU	World Bank Dataset Terms of Use apply; review before redistribution. Cite: Ethy, P.K., Barpanda et al. (2020); Ritharson, Kaggle, 2021.

Continued on next page

Table 3 — continued

Dataset	License	Notes & Citation
MangoLeafBD — Mango Leaf Disease	CC BY-NC 3.0	Attribution required; non-commercial use only. Ahmed et al., <i>Data in Brief</i> 47, 2023. DOI: 10.1016/j.dib.2023.108941.
Bean Leaf Lesions Classification	Apache 2.0	Attribution required; include original license text. Marquis03, Kaggle.
YellowRust19 — Yellow Rust in Wheat	Unknown (Kaggle ToS)	Listed as “Unknown” on Kaggle; contact author before redistribution. Cite: T. Hayit, <i>YellowRust19</i> , Kaggle, 2020.
Banana Leaf Disease Dataset v1.1	No explicit license	Compiled from three Kaggle datasets (rayhanarlistya, shifatearman, sujaykapadnis); cite all original creators.
Lettuce Diseases Dataset	Apache 2.0	Attribution required; include LICENSE and NOTICE file. Ashish Jena, Kaggle.
Durian Leaf Disease Dataset	CC BY 4.0	Attribution required. Cthng123, Kaggle.
Eggplant Disease Recognition Dataset	CC BY 4.0	Canonical Mendeley source (DOI: 10.17632/r3tb5mzn4d.2) confirmed CC BY 4.0. Mafi & Ava, 2023.
Cotton Disease Multi Transformation	MIT	Attribution required; include original MIT license text. Shuvo Kumar Basak, Kaggle.
Pumpkin Leaf Disease Multi Transformation	MIT	Attribution required; include original MIT license text. Shuvo Kumar Basak, Kaggle.
Rose Leaf Disease Multi Transformation	MIT	Attribution required; include original MIT license text. Shuvo Kumar Basak, Kaggle.
Strawberry Disease Detection Dataset	Unknown (permissive)	Listed as “Unknown” on Kaggle; description grants permission to use, modify, and augment. Cite: Afzaal et al., <i>Sensors</i> 21, 6565 (2021).
SugarLeafIDN — Sugarcane Leaf Diseases	Apache 2.0	Attribution required; include LICENSE and NOTICE file. Betty D. Puspasari, Kaggle.
FUSARIUM22 — Fusarium Wilt in Chickpea PlantDoc Dataset	Unknown (Kaggle ToS) CC BY 4.0	Listed as “Unknown” on Kaggle; contact author before redistribution. Cite: Hayit et al., <i>Eur. J. Plant Pathol.</i> 168 (2024). Attribution required. Singh et al., <i>ACM IKDD CoDS & COMAD</i> , 2020. DOI: 10.1145/3371158.3371196.
PlantVillage Dataset	CC BY-SA 3.0	Attribution + ShareAlike required; derivatives must carry the same license. Hughes & Salathé, arXiv:1511.08060, 2015.
LeafNet Dataset	CC BY 4.0	Confirmed CC BY 4.0 on HuggingFace (enalis/LeafNet); ~70% of data is public (training split only). Quoc et al., arXiv:2602.13662, 2026.
Soybean Stress Identification (PNAS)	CC BY 4.0	Attribution required. Ghosal et al., <i>PNAS</i> 115(18), 4613–4618, 2018. DOI: 10.1073/pnas.1716999115.
PlantWild v2 Dataset	CC BY-NC-ND 4.0	No commercial use; no derivatives permitted. Wei et al., <i>ACM MM</i> , 2024. DOI: 10.1145/3664647.3680599.
CDDM — Crop Disease Domain Multimodal	CC BY-NC-ND 4.0	No commercial use; no derivatives permitted. Liu et al., arXiv:2503.06973, 2025.

B Registry Schema

Field	Example
Crop (host)	Soybean
Disease	White mold (Sclerotinia Stem Rot)
Pathogen	<i>Sclerotinia sclerotiorum</i>
Pathogen type	Fungal
Affected organs	Stem, seed
Visual symptoms	lesion shape, color, texture, ...

Table 4: Registry schema. Each field carries per-field provenance: the extracted value, the source URL, and a verbatim supporting quote.

Table 5: Single-pass few-shot classification accuracy. The model receives k randomly sampled labeled images in a single API call with no reasoning loop (Sonnet, 3 test images per class). Baseline is Agent (no KB) at $k=0$. Best per crop- k in **bold**.

Crop	$k = 0$	$k = 1$	$k = 4$	$k = 8$	$k = 16$
Soybean	41.9% (+10.8)	41.9% (+10.8)	40.5% (+9.5)	47.3% (+16.2)	48.6% (+17.6)
Corn	42.0% (+0.0)	38.6% (-3.4)	48.9% (+6.8)	50.0% (+8.0)	47.7% (+5.7)
Mango	92.5% (+0.0)	95.0% (+2.5)	92.5% (+0.0)	92.5% (+0.0)	95.0% (+2.5)
<i>Mean Δ (pp)</i>	+3.6	+3.3	+5.4	+8.1	+8.6

C Few-Shot Comparison

Table 5 compares the agentic pipeline against a single-pass few-shot baseline, as discussed in Section 8.

D Confusion Matrices: Corn and Mango

TODO: Improve placement and sizing of appendix confusion matrices. Consider regenerating with highlights for key classes, matching the format used for Soybean in the main text.

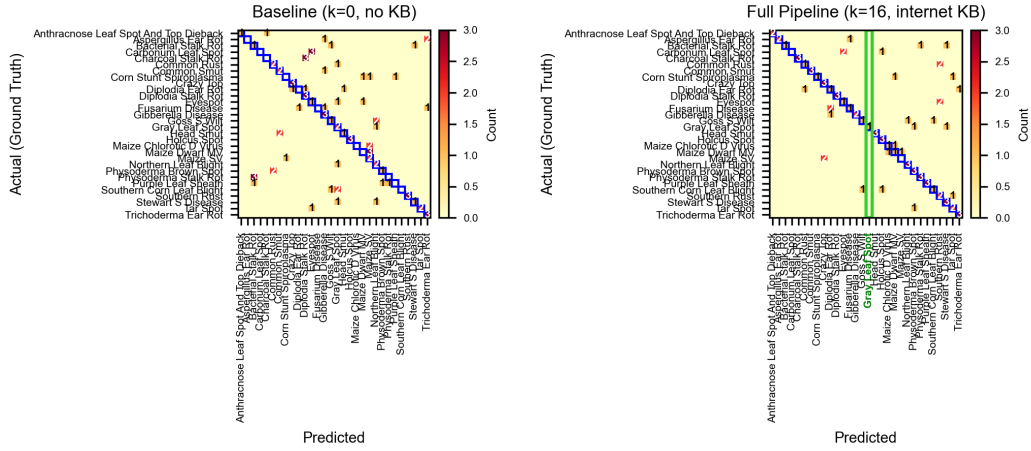


Figure 6: Confusion matrices for Corn (Sonnet, 30 classes). Left: baseline, $k=0$, no KB (42.0%). Right: full pipeline, $k=16$, internet KB (64.8%).

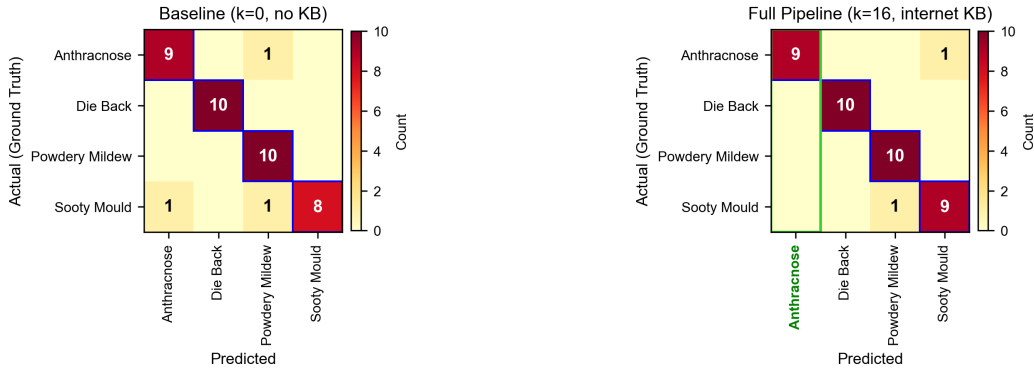



Figure 7: Confusion matrices for Mango (Sonnet, 4 classes). Left: baseline, $k=0$, no KB (92.5%). Right: full pipeline, $k=16$, internet KB (95.0%).

E Extended Reasoning Traces

The following traces show the agent's step-by-step reasoning across different crops, KB sources, models, and reference budgets. Each trace includes the configuration, the agent's observations and comparisons, and the final prediction with outcome.

Reasoning Trace: Soybean Diseases — Bacterial_Blight__Soybean_Dise_125

Model:	sonnet
KB source:	internet
Reference budget (k):	4
Prediction:	Bacterial_Blight
Ground truth:	Bacterial_Blight
Outcome:	Correct
Confidence:	0.82



Test image

Step 1: Observe test image

Step 2: The test image shows a soybean leaf with dark brown/black irregular lesions surrounded by yellow/chlorotic halos. The lesions appear angular and concentrated along the mid-vein area. The green tissue between lesions remains intact.

Top candidates: Bacterial_Blight (angular lesions + yellow halos), Septoria_brown_spot (dark irregular lesions), Soybean_Vein_necrosis_virus (lesions along veins with yellow tissue).

Now I'll view one reference from each top candidate.

Step 3: View reference: Soybean_Dise_17

Step 4: This Bacterial_Blight reference matches well - dark irregular lesions with yellow-green halos. Let me check Septoria_brown_spot next.

Step 5: View reference: Soybean_Dise_12

Step 6: Septoria shows extensive yellowing with small dark spots - different from the test image which has a mostly green leaf with distinct dark lesions and yellow halos. Let me check Soybean_Vein_necrosis_virus.

Step 7: View reference: Soybean_Dise_126

Step 8: Vein necrosis shows diffuse yellow spots without clear dark lesion centers - doesn't match. I need one more class view. The dark lesions with yellow halos in the test image best match Bacterial_Blight so far. Let me check Frogeye_leaf_spot as my 4th class.

Step 9: View reference: Soybean_Dise_6

Step 10: Analysis summary:

- Test image: Green soybean leaf with dark brown/black irregular lesions surrounded by yellow-green halos, concentrated along the mid-vein.
- Bacterial_Blight reference: Nearly identical - dark angular lesions with yellow-green halos on green leaf. Best match.
- Septoria_brown_spot: Extensive yellowing with tiny dark spots - doesn't match (test leaf is mostly green).
- Soybean_Vein_necrosis_virus: Diffuse yellowish spots without distinct dark centers - doesn't match.
- Frogeye_leaf_spot: Small circular lesions with holes, reddish-brown borders - doesn't match the larger irregular pattern.

The yellow-green halos + dark angular/irregular lesions on an otherwise green leaf is the hallmark of Bacterial_Blight.

Reasoning Trace: Soybean Diseases — Anthracnose__Soybean_Dise_10

Model: sonnet
KB source: internet
Reference budget (k): 4
Prediction: Charcoal_Rot
Ground truth: Anthracnose
Outcome: **Incorrect**
Confidence: 0.88



Test image

- Step 1:** Observe test image
- Step 2:** Key observations from the test image: - Plant part: Stem (3 stems shown close-up) - Key features: Bleached/pale white-tan stems with numerous small black specks scattered across the surface - Texture: Fluffy/hairy white mycelial growth visible on stems - Black structures: Many small black bodies embedded in/on the white bleached stem tissue
The white bleached stems + black sclerotia embedded in white mycelium is classic White Mold. Let me verify with reference images, then also check Charcoal Rot (which has black microsclerotia under the epidermis) as a candidate.
- Step 3:** View reference: Soybean_Dise_339
- Step 4:** This White Mold reference shows early-stage large sclerotia but the stem is still green. Let me check Charcoal Rot to compare the black speck pattern.
- Step 5:** View reference: Soybean_Dise_108
- Step 6:** Charcoal Rot shows a pale stem but without the dense black specks pattern. Now let me check the other White Mold stem image.
- Step 7:** View reference: Soybean_Dise_56
- Step 8:** Both White Mold references show bleached stems with black structures, but the test image has very dense, numerous tiny black specks — more consistent with microsclerotia. Let me use my last view on Charcoal Rot to compare.
- Step 9:** View reference: Soybean_Dise_49
- Step 10:** This is a strong match. The Charcoal Rot reference (Soybean_Dise_49) shows exactly the same pattern as the test image: - Multiple stems shown together - Pale/bleached tan-gray stems - Dense, numerous tiny black specks scattered across the stem surface (microsclerotia giving the "charcoal-sprinkled" appearance) - The internal split stem shows white pith
This matches the test image almost perfectly — three stems displayed, pale/bleached with countless tiny black dots (microsclerotia) across the surface. White Mold sclerotia are larger and fewer; what's visible here is the characteristic microsclerotia of Charcoal Rot.

Reasoning Trace: Soybean Diseases — Bacterial_Blight__Soybean_Dise_125

Model: sonnet
KB source: no KB
Reference budget (k): 4
Prediction: Bacterial_Blight
Ground truth: Bacterial_Blight
Outcome: **Correct**
Confidence: 0.87



Test image

- Step 1:** Observe test image

- Step 2:** The test image shows a soybean leaf with dark brown/purplish-black lesions surrounded by yellow halos, distributed along the leaf midrib area. There's also one hole. The yellow halo + dark center pattern strongly suggests bacterial or fungal leaf spotting.
Top candidates: Bacterial_Blight (angular dark lesions + yellow halos), Septoria_brown_spot, Cercospora, Frogeye_leaf_spot.
Now viewing one reference from each of my 4 top candidates:
- Step 3:** View reference: Soybean_Dise_17
- Step 4:** This Bacterial_Blight reference matches very well - dark brown/purple lesions with yellow halos. Now checking Septoria and Cercospora.
- Step 5:** View reference: Soybean_Dise_12
- Step 6:** View reference: Soybean_Dise_33
- Step 7:** The Cercospora reference shows seed discoloration - completely different from the test image. The Bacterial_Blight reference is a clear match.
Analysis: - Test image: soybean leaf with dark brown/purple lesions + prominent yellow halos, angular distribution along midrib - Bacterial_Blight reference: identical pattern - dark lesions with yellow halos on green leaf tissue - Septoria_brown_spot: more extensive yellowing, smaller spots scattered throughout - Cercospora: seed discoloration, not leaf lesions
The yellow halo + dark angular lesion pattern is the defining feature of Bacterial_Blight.

Reasoning Trace: Soybean Diseases — Anthracnose__Soybean_Dise_10

Model: sonnet
KB source: no KB
Reference budget (k): 4
Prediction: Phomopsis
Ground truth: Anthracnose
Outcome: **Incorrect**
Confidence: 0.62



Test image

- Step 1:** Observe test image
- Step 2:** The test image shows soybean stems with pale/bleached (cream-white) coloration and scattered small black dots across the stem surface. These are likely pycnidia or microsclerotia. Candidates: Phomopsis (bleached stems + black pycnidia), Charcoal_Rot (gray stem + black microsclerotia), Anthracnose (dark lesions + black acervuli).
Now viewing one reference from each top candidate:
- Step 3:** View reference: Soybean_Dise_14
- Step 4:** This Phomopsis image shows a pod. Let me check Charcoal_Rot and Anthracnose references.
- Step 5:** View reference: Soybean_Dise_10
- Step 6:** Let me use my last view on another Charcoal_Rot reference showing the stem close-up:
- Step 7:** View reference: Soybean_Dise_52
- Step 8:** Visual analysis of the test image: - 3 soybean stems shown close-up - Stems are pale/cream/bleached white - Numerous small black dots scattered across the stem surface - Top stem is more brownish (earlier stage?), middle and bottom stems are distinctly bleached
Reasoning: The characteristic pale/cream bleached stems with scattered black pycnidia (fruiting bodies) visible on the surface is classic Phomopsis stem blight. The bleaching pattern and black dot distribution match this disease. Charcoal_Rot was also considered (microsclerotia + silvery-gray stem), but Charcoal_Rot typically shows more grayish discoloration rather than this cream/white bleaching. The reference Charcoal_Rot images also showed field-level wilting symptoms rather than close-up stem bleaching with black dots.

Model: opus
KB source: internet
Reference budget (k): 8
Prediction: Bacterial_Blight
Ground truth: Bacterial_Blight
Outcome: **Correct**
Confidence: 0.92



Test image

- Step 1:** Observe test image
- Step 2:** View reference: part_index
- Step 3:** The test image shows a soybean leaf with dark brown irregular lesions concentrated in a band, surrounded by yellowish-green halos. Some tissue has fallen out leaving holes, giving a ragged appearance. Lesions appear somewhat angular.
This is a leaf disease. Top candidates based on symptoms: Bacterial_Blight, Septoria_brown_spot, Soybean_Vein_necrosis_virus, Bacterial_Pustule. Let me view references across 8 classes.
- Step 4:** View reference: Soybean_Dise_17
- Step 5:** Bacterial_Blight reference shows angular brown spots with yellow halos, merged dead patches. Very similar pattern to test image. Let me check other candidates.
- Step 6:** View reference: Soybean_Dise_12
- Step 7:** Septoria brown spot shows numerous small dark spots with extensive leaf yellowing. Different pattern from the test image. Let me continue.
- Step 8:** View reference: Soybean_Dise_126
- Step 9:** SVN shows yellow spots along veins, different from the test image's angular merged lesions with halos. Let me check more.
- Step 10:** View reference: Soybean_Dise_11
- Step 11:** Bacterial Pustule reference shows a close-up of severely affected leaf tissue with extensive brown/orange discoloration. Very different from the test image. Let me check more classes.
- Step 12:** View reference: Soybean_Dise_6
- Step 13:** Frogeye leaf spot shows discrete circular gray spots with reddish-brown borders, scattered across the leaf. Very different from the test image's merged angular lesions. Let me continue.
- Step 14:** View reference: Soybean_Dise_137
- Step 15:** Downy mildew shows light yellow spots spread across the leaf surface. Very different from the test image. Let me check two more classes.
- Step 16:** View reference: Soybean_Dise_74
- Step 17:** Soybean rust shows small tan to brick-red spots scattered across the leaf. Very different from the test image. One more class to check.
- Step 18:** View reference: Soybean_Dise_18
- Step 19:** Phyllosticta shows large tan lesions with dark margins, different pattern from the test image. I've now viewed 8 references across 8 different classes. The evidence strongly points to Bacterial_Blight:
 - Angular, vein-limited brown lesions matching the description perfectly - Yellowish-green halos surrounding the darkened spots - Merged dead patches where spots have coalesced
 - Tissue falling out leaving holes, creating the characteristic "ragged appearance" - The Bacterial_Blight reference image showed the most similar pattern
 No other class matched: Septoria had extensive yellowing with small scattered spots, Frogeye had discrete circular gray-centered spots, Downy mildew/Soybean rust had very different spot patterns, and Phyllosticta had larger irregular tan lesions.

Model: opus
KB source: internet
Reference budget (k): 8
Prediction: Charcoal_Rot
Ground truth: Anthracnose
Outcome: **Incorrect**
Confidence: 0.92



Test image

- Step 1:** Observe test image
- Step 2:** View reference: part_index
- Step 3:** The test image shows soybean stems with numerous small black specks/dots scattered across bleached/tan stem surfaces. This narrows candidates to stem diseases.
 Top candidates based on symptoms: - Charcoal_Rot: microsclerotia giving "charcoal-sprinkled" appearance - Anthracnose: black acervuli with setae on stems
 Let me view references systematically, one at a time.
- Step 4:** View reference: Soybean_Dise_108
- Step 5:** Charcoal rot reference shows gray streaks but not the dense black specks pattern in the test image. Let me check Anthracnose.
- Step 6:** View reference: Soybean_Dise_17
- Step 7:** Anthracnose shows dark lesions on stems with pods attached. Let me check the other Charcoal Rot stem reference.
- Step 8:** View reference: Soybean_Dise_49
- Step 9:** This Charcoal Rot reference is a strong match - small black specks (microsclerotia) scattered on gray/tan stems, very similar to the test image. Let me check other stem disease candidates for comparison.
- Step 10:** View reference: Soybean_Dise_339
- Step 11:** White Mold shows fluffy white mycelium and large black sclerotia - doesn't match. Let me check Anthracnose stem more closely and other candidates.
- Step 12:** View reference: Soybean_Dise_67
- Step 13:** Anthracnose on green stems with pods - different stage and appearance. Let me check Phytophthora stem and Rhizoctonia stem.
- Step 14:** View reference: Soybean_Dise_57
- Step 15:** Phytophthora shows dark chocolate-brown stem discoloration on living plants - doesn't match. Let me check Rhizoctonia and Fusarium stems.
- Step 16:** View reference: Soybean_Dise_18
- Step 17:** Rhizoctonia shows sunken lesions at the soil line on seedlings - doesn't match.
- Step 18:** View reference: Soybean_Dise_54
- Step 19:** Fusarium shows discolored roots/stems on seedlings - doesn't match.
 All 8 reference views used. The evidence strongly points to Charcoal Rot:
 - Test image: Mature, dry soybean stems with dense, small black specks scattered across a gray/tan stem surface - Charcoal Rot reference (Dise_49): Nearly identical pattern of small black microsclerotia on gray stems - Symptom description match: "small black specks (microsclerotia) beneath the epidermis giving a charcoal-sprinkled appearance" matches perfectly - Other candidates eliminated: Anthracnose (different fruiting body pattern), White Mold (fluffy mycelium + large sclerotia), Phytophthora (chocolate-brown discoloration on living stems), Rhizoctonia (soil-line lesions), Fusarium (seedling stage)

Reasoning Trace: Corn Diseases — Anthracnose_Leaf_Spot_And_Top_Dieback__Corn_Disease_47

Model: sonnet
KB source: internet
Reference budget (k): 4
Prediction: Anthracnose_Leaf_Spot_And_Top_Dieback
Ground truth: Anthracnose_Leaf_Spot_And_Top_Dieback
Outcome: **Correct**
Confidence: 0.42



Test image

- Step 1:** Observe test image
- Step 2:** View reference: part_index
- Step 3:** The test image shows corn stalks with very dark blackish/bluish-black discoloration, particularly on the rind, with yellowish areas near nodes. This is a stem specimen. Stem candidates: Anthracnose, Bacterial_Stalk_Rot, Charcoal_Stalk_Rot, Diplodia_Stalk_Rot, Physoderma_Stalk_Rot, Purple_Leaf_Sheath.
The dark blackish nodes are striking. Top candidates: Physoderma_Stalk_Rot (black nodes), Anthracnose (black shiny stalk), Charcoal_Stalk_Rot (charred interior), Diplodia_Stalk_Rot. Let me view one reference per top candidate.
- Step 4:** View reference: Corn_Disease_17
- Step 5:** View reference: Corn_Disease_12
- Step 6:** View reference: Corn_Disease_10
- Step 7:** View reference: Corn_Disease_2
- Step 8:** Test image analysis: - Affected part: stem (stalk sections cut and laid on wooden planks) - Key features: extensive deep blue-black discoloration across the entire stalk rind, yellowish-tan coloring at nodes, multiple stalk sections
Reference comparison: 1. Physoderma_Stalk_Rot — reference shows localized dark spots on green stalk. Doesn't match the extensive rind blackening. 2. Charcoal_Stalk_Rot — reference shows dark shredded interior with charred pith. Interior charring doesn't match the exterior rind blackening in the test image. 3. Anthracnose (stem) — reference shows brownish-black patches on rind with some rotted sections. Closer match — Anthracnose stalk rot is known for extensive black discoloration of the stalk rind. 4. Purple_Leaf_Sheath — reference shows localized dark lesion on otherwise green stalk. Doesn't match.
The deep, extensive blue-black coloration of the entire stalk exterior rind, with yellowish nodes, is most consistent with Anthracnose stalk rot. The Anthracnose stalk phase causes characteristic shiny black lesions across the lower stalk rind — which aligns with what's visible.

Reasoning Trace: Corn Diseases — Aspergillus_Ear_Rot__Corn_Disease_19

Model: sonnet
KB source: internet
Reference budget (k): 4
Prediction: Fusarium_Disease
Ground truth: Aspergillus_Ear_Rot
Outcome: **Incorrect**
Confidence: 0.50



Test image

- Step 1:** Observe test image
- Step 2:** View reference: part_index

Step 3: The test image shows a corn ear with damage concentrated at the tip — brownish/reddish-pink rotted kernels at the tip, some whitish material mixed in, healthy yellow kernels below. This is a seed/pod part. Top candidates: Fusarium_Disease, Gibberella_Disease, Aspergillus_Ear_Rot, Diplodia_Ear_Rot.

Step 4: View reference: Corn_Disease_25

Step 5: View reference: Corn_Disease_18

Step 6: View reference: Corn_Disease_15

Step 7: View reference: Corn_Disease_27

Step 8: Analysis:

- Test image: Corn ear with tip rot — rotted/shriveled brownish-reddish-pink kernels concentrated at the tip, whitish material in the damaged area, insect damage visible, healthy yellow kernels below. - Fusarium_Disease ref: Whitish-pinkish mold scattered mid-ear — partial match (tip location fits, color fits partially). - Gibberella_Disease ref: Vivid red/pink mold starting at tip — the red color is more intense than the test image's brownish-pink. - Aspergillus_Ear_Rot ref: Olive-green powdery mold — no match. - Diplodia_Ear_Rot ref: White mycelium from base, bleached husks — no match.

Reasoning: The test image shows tip-concentrated rot with brownish-pinkish kernels and some whitish mold material. Fusarium is described as "whitish to pink-colored fungal growth on kernels, often at the ear tip" with association with insect damage entry points. The whitish material visible + tip location + pinkish-brownish coloration fits Fusarium better than the vivid red of Gibberella. Aspergillus and Diplodia are clearly ruled out.

Reasoning Trace: Corn Diseases — Anthracnose_Leaf_Spot_And_Top_Dieback_Corn_Disease_47

Model: sonnet
KB source: no KB
Reference budget (k): 8
Prediction: Anthracnose_Leaf_Spot_And_Top_Dieback
Ground truth: Anthracnose_Leaf_Spot_And_Top_Dieback
Outcome: **Correct**
Confidence: 0.52



Test image

Step 1: Observe test image

Step 2: The test image shows corn stalks laid out, with dark/black discoloration on the outer surface and yellow-tan internal tissue. This looks like a stalk rot disease. Key features: dark blackened exterior, yellowing/tan mixed tissue, stalks split open showing internal decay.

Candidates: Charcoal_Stalk_Rot, Diplodia_Stalk_Rot, Gibberella_Disease, Bacterial_Stalk_Rot, Physoderma_Stalk_Rot.

Let me view reference images one at a time.

Step 3: View reference: Corn_Disease_12

Step 4: View reference: Corn_Disease_14

Step 5: View reference: Corn_Disease_17

Step 6: View reference: Corn_Disease_12

Step 7: View reference: Corn_Disease_16

Step 8: View reference: Corn_Disease_32

Step 9: View reference: Corn_Disease_10

Step 10: View reference: Corn_Disease_52

Step 11: I've now viewed 8 reference images. Let me assess:

Test image features: Two corn stalks laid out on a wooden surface, showing heavy dark/black exterior discoloration with yellow-tan patches, one stalk split open revealing fibrous internal tissue. Classic stalk rot presentation.

Reference comparison: - Charcoal_Stalk_Rot (Corn_Disease_12, 32): Shows dark internal pith with gray-black sclerotia — internal focus, doesn't match the heavy black exterior
 - Diplodia_Stalk_Rot (Corn_Disease_14): Dry, tan/brown stalk breaking at node — no black exterior - Physoderma_Stalk_Rot (Corn_Disease_17): Dark spots on a green stalk — very different - Gibberella_Disease (Corn_Disease_12): Ear rot with pink/white mold — completely different - Bacterial_Stalk_Rot (Corn_Disease_16): Tan/brown leaf sheath discoloration — different - Anthracnose_Leaf_Spot_And_Top_Dieback (Corn_Disease_10): Shows stalks with brown/dark exterior patches including black areas at nodes — closer match. Anthracnose stalk rot is specifically known for causing shiny black discoloration of the outer stalk rind

The heavy black/dark exterior discoloration on the stalks is most consistent with Anthracnose stalk rot, which is part of the Anthracnose_Leaf_Spot_And_Top_Dieback complex. The Corn_Disease_10 reference showed similar dark stalk exterior features.

Reasoning Trace: Corn Diseases — Aspergillus_Ear_Rot_Corn_Disease_19

Model: sonnet
KB source: no KB
Reference budget (k): 8
Prediction: Gibberella_Disease
Ground truth: Aspergillus_Ear_Rot
Outcome: **Incorrect**
Confidence: 0.82



Test image

- Step 1:** Observe test image
- Step 2:** The test image shows a corn ear with rot concentrated at the tip. The affected kernels are brown/pinkish with some white-gray mold growth. The rest of the ear has healthy yellow kernels. This pattern suggests an ear rot disease — likely Gibberella (pink/red tip rot) or Fusarium.
 Let me check reference images for top candidates.
- Step 3:** View reference: Corn_Disease_12
- Step 4:** View reference: Corn_Disease_25
- Step 5:** View reference: Corn_Disease_27
- Step 6:** View reference: Corn_Disease_15
- Step 7:** View reference: Corn_Disease_42
- Step 8:** View reference: Corn_Disease_2
- Step 9:** View reference: Corn_Disease_42
- Step 10:** View reference: Corn_Disease_64
- Step 11:** Analysis summary:
 - Test image: Corn ear with tip-concentrated rot. Kernels at the tip are brownish/pinkish-discolored with white-gray mold. Lower ear has healthy yellow kernels.
 Reference comparisons: - Gibberella_Disease (ref 12): Very strong match — same tip rot pattern, brownish-pink kernels, white mold growth concentrated at tip. - Gibberella_Disease (ref 42): Confirms tip-to-base progression pattern. - Gibberella_Disease (ref 64): Shows pink discoloration on husks, consistent with the disease. - Fusarium_Disease: Scattered patches across the ear, not tip-focused — does not match. - Diplodia_Ear_Rot: Different presentation (dried/bleached), no pink coloration. - Aspergillus_Ear_Rot: Distinctive olive-green mold — clearly different. - Trichoderma_Ear_Rot: Blue-green mold all over — clearly different.
 The tip-origin rot with brownish-pink kernel discoloration and white mold is the hallmark of Gibberella ear rot. The visual match to the Gibberella reference (Corn_Disease_12.jpg) is strong.

Model: opus
KB source: internet
Reference budget (k): 8
Prediction: Anthracnose_Leaf_Spot_And_Top_Dieback
Ground truth: Anthracnose_Leaf_Spot_And_Top_Dieback
Outcome: **Correct**
Confidence: 0.95



Test image

- Step 1:** Observe test image
- Step 2:** View reference: part_index
- Step 3:** The test image shows corn stalks with dark black discoloration, particularly at the nodes. The stalks appear broken/split, exposing internal tissue. This is a stem disease.
Stem candidates: Anthracnose, Bacterial_Stalk_Rot, Charcoal_Stalk_Rot, Common_Smut, Diplodia_Stalk_Rot, Physoderma_Brown_Spot, Physoderma_Stalk_Rot, Purple_Leaf_Sheath, among others.
Top suspects based on symptoms: Physoderma_Stalk_Rot (black nodes, snapping), Charcoal_Stalk_Rot, Anthracnose (stem), Purple_Leaf_Sheath. Let me view references one at a time.
- Step 4:** View reference: Corn_Disease_17
- Step 5:** Physoderma Stalk Rot reference shows a green stalk with dark spots, quite different from the test image's dry, heavily blackened stalks. Let me check Charcoal_Stalk_Rot next.
- Step 6:** View reference: Corn_Disease_12
- Step 7:** Charcoal Stalk Rot shows shredded pith with dark interior. The external dark discoloration is somewhat similar. Let me check Anthracnose stem, Purple_Leaf_Sheath, and Diplodia_Stalk_Rot.
- Step 8:** View reference: Corn_Disease_10
- Step 9:** Anthracnose stem shows shiny black areas on the stalk exterior, similar to the test image. Let me check Purple_Leaf_Sheath and Diplodia_Stalk_Rot.
- Step 10:** View reference: Corn_Disease_11
- Step 11:** Purple_Leaf_Sheath shows a green, living plant with brown/dark lesions on the sheath. Very different from the test image's dry, dead stalks. Let me check Diplodia_Stalk_Rot.
- Step 12:** View reference: Corn_Disease_14
- Step 13:** Diplodia Stalk Rot shows tan/brown dry stalks, not the heavy black discoloration in the test image. Let me check Bacterial_Stalk_Rot and Common_Smut stem.
- Step 14:** View reference: Corn_Disease_16
- Step 15:** Bacterial Stalk Rot shows wilted, brown leaves on a still-standing plant. Not similar. Let me check Common_Smut stem and a second Anthracnose stem reference.
- Step 16:** View reference: Corn_Disease_176
- Step 17:** Common Smut shows galls, completely different. Let me view the second Anthracnose stem reference, which was the closest match so far.
- Step 18:** View reference: Corn_Disease_60
- Step 19:** This is a very strong match. The Anthracnose stem reference (Corn_Disease_60.jpg) shows the exact same presentation as the test image: dry corn stalks laid on wooden boards over rocks, with shiny black discoloration on the rind interspersed with tan/yellow areas, and exposed lighter internal tissue where the stalk is broken. The background, stalk appearance, and black lesion pattern are nearly identical.

Model: opus
KB source: internet
Reference budget (k): 8
Prediction: Fusarium_Disease
Ground truth: Aspergillus_Ear_Rot
Outcome: **Incorrect**
Confidence: 0.55



Test image

- Step 1:** Observe test image
- Step 2:** View reference: part_index
- Step 3:** The test image shows a corn ear with damage at the tip. Kernels at the tip are discolored, swollen, and some appear damaged with whitish material between them. The rest of the ear looks healthy. This is a seed/ear disease.
Candidates from the seed/pod categories: Aspergillus_Ear_Rot, Common_Smut, Diplodia_Ear_Rot, Fusarium_Disease, Gibberella_Disease, Trichoderma_Ear_Rot, Head_Smut. Let me view reference images one per class, starting with top candidates.
- Step 4:** View reference: Corn_Disease_25
- Step 5:** Fusarium reference shows whitish mold on scattered kernels mid-ear. The test image has damage concentrated at the tip with swollen kernels. Let me check other candidates.
- Step 6:** View reference: Corn_Disease_218
- Step 7:** Common Smut shows large silvery-white galls. Not what's in the test image. Let me check more candidates.
- Step 8:** View reference: Corn_Disease_15
- Step 9:** Aspergillus shows olive-green powdery mold. Not matching. Let me check Diplodia, Gibberella, Trichoderma, and Head_Smut.
- Step 10:** View reference: Corn_Disease_41
- Step 11:** Diplodia shows cross-section with grayish-brown rotted kernels. Not matching the test image pattern. Let me check Gibberella.
- Step 12:** View reference: Corn_Disease_18
- Step 13:** Gibberella shows pinkish mold at ear tip. The test image doesn't have pink coloration. Let me check Trichoderma and Head_Smut.
- Step 14:** View reference: Corn_Disease_26
- Step 15:** Trichoderma shows dark blue-green conidia. Not matching. Let me check Head_Smut.
- Step 16:** View reference: Corn_Disease_139
- Step 17:** None of these are strong matches. The test image shows tip damage with whitish material, which aligns best with Fusarium description ("whitish to pink fungal growth at ear tip"). Let me use my last view on another Fusarium reference.
- Step 18:** View reference: Corn_Disease_42
- Step 19:** This Fusarium reference shows whitish-pink mold on kernels with healthy kernels nearby, similar to the test image. The test image shows: - Damaged, discolored kernels concentrated at the ear tip - Whitish material between kernels - Rest of ear is healthy yellow - No green, pink, or dark colored mold
This matches Fusarium_Disease: "whitish to pink fungal growth at ear tip" with "kernels can also be infected without visible symptoms" for the scattered pattern.

Reasoning Trace: Mango Leaf Disease — Anthracnose_Mango_Leaf_105

Model: sonnet
KB source: internet
Reference budget (k): 4
Prediction: Anthracnose
Ground truth: Anthracnose
Outcome: **Correct**
Confidence: 0.93

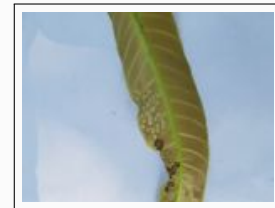


Test image

- Step 1:** Observe test image
- Step 2:** The test image shows a mango leaf with discrete dark brown/black spots — irregular, somewhat sunken-looking lesions on the leaf surface. This matches Anthracnose (irregular brown-black spots on leaves) more than the others. Let me check reference images, one per class.
- Step 3:** View reference: Mango_Leaf_14
- Step 4:** That anthracnose reference shows a very yellowed/brown leaf — different from the test image. Let me check one from each remaining class.
- Step 5:** View reference: Mango_Leaf_288
- Step 6:** View reference: Mango_Leaf_383
- Step 7:** This anthracnose reference (Mango_Leaf_383) is a near-perfect match to the test image — green leaf with discrete dark brown/black irregular spots scattered across the surface. The sooty mould reference showed a more diffuse dark coating with a greyish cast, not discrete spots.
- Summary of evidence: - Test image: green mango leaf, discrete dark brown-black irregular lesions scattered on surface - Anthracnose ref 383: identical pattern — green leaf, dark irregular spots, same morphology and distribution - Sooty mould ref: diffuse dark coating/encrustation, not discrete spots - Anthracnose KB: "oval or irregular brown to deep brown sunken spots on leaves" — matches exactly

Reasoning Trace: Mango Leaf Disease — Anthracnose_Mango_Leaf_289

Model: sonnet
KB source: internet
Reference budget (k): 4
Prediction: Powdery_Mildew
Ground truth: Anthracnose
Outcome: **Incorrect**
Confidence: 0.90



Test image

- Step 1:** Observe test image
- Step 2:** The test image shows a mango leaf with white, powdery-looking patches on the surface. This strongly suggests Powdery_Mildew. Let me check one reference per class.
- Step 3:** View reference: Mango_Leaf_14
- Step 4:** View reference: Mango_Leaf_14
- Step 5:** View reference: Mango_Leaf_147
- Step 6:** View reference: Mango_Leaf_166
- Step 7:** Clear differentiation across all four classes:
- Test image: Green leaf with white/cream powdery patches concentrated on the surface
- Powdery_Mildew ref: White/grey powdery deposits on a green leaf — matches the test

image well - Anthracnose ref: Brown withered leaf with dark necrotic lesions — does not match - Sooty_Mould ref: Black soot-like deposits — opposite color (black vs white) - Die_Back ref: Completely brown, dried/curled leaf — does not match

The white powdery deposits on an otherwise green leaf are the defining feature of Powdery_Mildew. No brown lesions (Anthracnose), no black encrustations (Sooty_Mould), no wilting/browning (Die_Back).

Reasoning Trace: Mango Leaf Disease — Anthracnose_Mango_Leaf_105

Model: sonnet
KB source: no KB
Reference budget (k): 8
Prediction: Anthracnose
Ground truth: Anthracnose
Outcome: **Correct**
Confidence: 0.95



Test image

- Step 1:** Observe test image
- Step 2:** Test image: A green mango leaf with several discrete, dark brown/black circular to irregular lesions scattered across the leaf surface. The spots appear sunken or necrotic with defined edges. No powdery coating, no tip dieback, no black sooty film. This pattern strongly suggests Anthracnose.
 Let me view one reference from each class before deciding.
- Step 3:** View reference: Mango_Leaf_14
- Step 4:** This Anthracnose reference shows a mostly yellowed/brown leaf with extensive necrosis - different from the test image's discrete spots. Let me check other classes.
- Step 5:** View reference: Mango_Leaf_14
- Step 6:** View reference: Mango_Leaf_14
- Step 7:** View reference: Mango_Leaf_147
- Step 8:** Die_Back is fully dead/shriveled. Powdery_Mildew shows diffuse white coating. Sooty_Mould shows dark smudgy patches. The test image has discrete dark spots on a green leaf. Let me check more Anthracnose references.
- Step 9:** View reference: Mango_Leaf_24
- Step 10:** This matches the test image closely - green leaf with discrete dark brown/black lesions. Let me view one more Anthracnose and one more Sooty_Mould to be thorough.
- Step 11:** View reference: Mango_Leaf_383
- Step 12:** Excellent match. Let me use my final view on Sooty_Mould to rule it out.
- Step 13:** View reference: Mango_Leaf_164
- Step 14:** Sooty_Mould shows a diffuse dark film/coating along the leaf surface, not discrete spots. This confirms the test image is not Sooty_Mould.
 Summary: - Test image: green mango leaf with discrete, well-defined dark brown/black necrotic spots scattered across it - Anthracnose (Mango_Leaf_24, Mango_Leaf_383): green leaves with identical discrete dark lesions — strong visual match - Die_Back: fully shriveled/brown dead tissue — no match - Powdery_Mildew: diffuse white powdery coating — no match - Sooty_Mould: dark smudgy film over leaf surface — no match

Model: sonnet
KB source: no KB
Reference budget (k): 8
Prediction: Sooty_Mould
Ground truth: Anthracnose
Outcome: **Incorrect**
Confidence: 0.72



Test image

- Step 1:** Observe test image
- Step 2:** The test image shows a mango leaf with small, dark brown/black scattered spots and lesions across the leaf surface. The leaf is green with irregular dark spots, some appearing as small necrotic lesions with irregular margins. No white powdery coating, no sooty black coating, no tip dieback. This pattern suggests Anthracnose.
Let me view one reference per class before deciding.
- Step 3:** View reference: Mango_Leaf_14
- Step 4:** This Anthracnose reference shows a yellowed/tan leaf with severe tissue collapse. The test image looks greener with small scattered spots. Let me check Die_Back.
- Step 5:** View reference: Mango_Leaf_14
- Step 6:** Die_Back shows a completely brown/dead, shriveled leaf - very different from the test image. Let me check Powdery_Mildew.
- Step 7:** View reference: Mango_Leaf_14
- Step 8:** Powdery Mildew shows a grey-white powdery coating across the leaf surface - not matching the test image. Let me check Sooty_Mould.
- Step 9:** View reference: Mango_Leaf_147
- Step 10:** Sooty Mould shows dark scattered patches on a green leaf - somewhat similar to test image. Let me view another Anthracnose reference to compare more carefully.
- Step 11:** View reference: Mango_Leaf_24
- Step 12:** Anthracnose_24 shows larger coalescing dark brown lesions on green leaf. The test image spots are smaller and more scattered. Let me check another Sooty Mould reference.
- Step 13:** View reference: Mango_Leaf_164
- Step 14:** This Sooty Mould (164) looks very similar to the test image - wavy green leaf with small scattered dark spots. Let me check one more Sooty Mould and one more Anthracnose.
- Step 15:** View reference: Mango_Leaf_288
- Step 16:** View reference: Mango_Leaf_383
- Step 17:** Now I have a clear comparison:
 - Anthracnose_383: Larger, well-defined dark brown lesions with clear margins on a green leaf - Sooty_Mould_164: Wavy green leaf with small scattered dark spots - very close to the test image - Sooty_Mould_288: Green leaf with small scattered dark spots and some grey patches
 The test image (wavy green leaf, small scattered dark spots) most closely matches Sooty Mould references, particularly Mango_Leaf_164. The spots in the test image are smaller and more numerous than typical Anthracnose lesions which tend to be larger with distinct margins.