

From Detection to Interpretation: Generating Human-Readable Cabbage Disease Reports Using YOLOv8s and LLaMA 3.1-13B-Instruct

Aulia Akhrian Syahidi*

Data Science and Artificial Intelligence Program, Thai-Nichi International College,

Thai-Nichi Institute of Technology, Bangkok, Thailand

Received: 27 March 2026, Received in revised form: 15 June 2026, Accepted: 16 June 2026

Available online: 30 June 2026

Abstract

Background and Objectives: Recent advances in deep learning have significantly improved plant disease detection, particularly through convolutional neural networks (CNNs) and YOLO-based object detection architectures. In precision agriculture, these approaches have demonstrated strong capabilities for identifying and localizing disease symptoms with high accuracy. However, most existing systems primarily produce technical outputs such as bounding boxes, class labels, and confidence scores, which are difficult for non-expert agricultural users to interpret and utilize for practical decision-making. Consequently, a gap remains between detection performance and real-world usability. Farmers and agricultural practitioners often require contextual explanations, severity interpretation, and actionable recommendations rather than raw detection outputs. At the same time, recent developments in large language models (LLMs) have shown strong potential for transforming structured information into coherent and human-readable narratives. Nevertheless, the integration of object detection and language generation for agricultural disease reporting remains largely unexplored. To address this limitation, this study proposes an end-to-end detection-to-interpretation framework that integrates YOLOv8s and LLaMA 3.1-13B-Instruct to automatically transform cabbage disease detection outputs into human-readable and actionable agricultural reports. The framework aims to bridge the gap between visual disease detection and user-centered interpretation, thereby improving the interpretability, accessibility, and practical usability of artificial intelligence (AI) systems in precision agriculture.

Methodology: The proposed framework consists of three main components: disease detection, structured representation transformation, and natural language generation (NLG). A publicly available cabbage disease dataset containing Black Rot, Downy Mildew, and Healthy classes was adapted for object detection through manual annotation and expert validation. The final dataset comprised 1,500 images and was divided into training, validation, and testing subsets using a 70:15:15 ratio. YOLOv8s was employed as the disease detection model due to its favorable balance between detection accuracy and computational efficiency. The model was trained using resized

images of 640×640 pixels with data augmentation techniques including rotation, scaling, flipping, and brightness adjustment. Detection outputs consisting of bounding boxes, disease labels, and confidence scores were subsequently transformed into a structured representation containing disease type, confidence level, spatial location, and estimated severity. This intermediate representation served as a semantic bridge between computer vision outputs and language generation. LLaMA 3.1-13B-Instruct was then used to generate natural language explanations through prompt-based structured-to-text generation. Few-shot prompting, controlled decoding parameters, and post-processing mechanisms were applied to improve consistency and readability. The framework was evaluated through multiple perspectives, including object detection metrics, text generation metrics, baseline comparisons, ablation analysis, and qualitative expert assessment.

Main Results: Experimental results demonstrate that the proposed framework achieves strong performance across both detection and interpretation tasks. The YOLOv8s detection module achieved a precision of 0.85, recall of 0.83, and mAP@0.5 of 0.87, indicating reliable disease localization and classification performance. Compared with the Faster R-CNN baseline, which achieved an mAP@0.5 of 0.81, YOLOv8s provided superior detection accuracy while maintaining computational efficiency suitable for real-time agricultural deployment. Class-level analysis showed that Black Rot achieved the highest detection performance with an mAP@0.5 of 0.89, while Downy Mildew achieved 0.85 due to the more subtle visual characteristics of its symptoms. For the language generation component, the generated explanations achieved a BLEU score of 0.48, ROUGE-L of 0.61, METEOR of 0.54, and BERTScore of 0.89, demonstrating strong semantic alignment with expert-written references. Baseline comparisons further revealed that the proposed framework outperformed both template-based reporting and a lightweight LLaMA 3.2-3B-Instruct model across all evaluation metrics. Ablation analysis confirmed the importance of the structured representation and prompt engineering components. Removing the structured representation reduced BERTScore from 0.89 to 0.83, while removing few-shot prompting reduced BLEU from 0.48 to 0.42 and resulted in more generic and repetitive outputs. Qualitative evaluation involving three agricultural experts demonstrated that 88% of the generated explanations were rated as accurate, clear, and actionable. Furthermore, the evaluation achieved strong inter-rater agreement with a Cohen's Kappa coefficient of $K = 0.81$, indicating high reliability and consistency among expert assessments.

Conclusions: The proposed YOLOv8s–LLaMA 3.1-13B-Instruct framework effectively bridges visual disease detection and agricultural interpretation by transforming technical detection outputs into understandable, context-aware, and actionable disease reports. The results demonstrate that integrating object detection with LLMs can significantly improve the interpretability and practical value of AI-based plant disease diagnosis systems. Beyond achieving strong detection accuracy and semantic generation quality, the framework supports user-centered

decision-making by providing explanations that are accessible to non-expert agricultural users. The study contributes a novel detection-to-interpretation pipeline, an expert-validated cabbage disease detection dataset, and a structured representation mechanism that connects computer vision and NLG. Nevertheless, the framework currently relies on heuristic-based severity estimation and is limited to a small number of disease categories under constrained environmental conditions. Future research should focus on incorporating lesion segmentation, vision transformer-based architectures, multimodal environmental information, domain-specific language model adaptation, and large-scale field validation to improve robustness, scalability, and applicability in real-world precision agriculture environments.

Keywords: cabbage disease detection; LLaMA 3.1-13B-Instruct; precision agriculture; structured-to-text generation; YOLOv8s

*Corresponding author. E-mail: aulia.sya@tni.ac.th

Introduction

Cabbage is an economically important crop widely cultivated due to its high nutritional value and market demand (Liu *et al.*, 2025). However, cabbage production is highly vulnerable to diseases such as black rot and downy mildew, which can significantly reduce yield and quality, resulting in substantial economic losses (Vega-Álvarez *et al.*, 2021; Ma *et al.*, 2025; Wang *et al.*, 2026). Delayed or inaccurate disease diagnosis may lead to improper treatment, excessive pesticide application, and environmental harm (Ebron *et al.*, 2024). Therefore, accurate and timely disease detection is essential for sustainable agriculture and food security (Szechyńska-Hebda *et al.*, 2025). Traditional disease identification methods mainly rely on manual inspection, which is time-consuming, labor-intensive, and prone to inconsistency, particularly in large-scale agricultural environments (Sambana *et al.*, 2025; Rahman *et al.*, 2025).

Recent advances in deep learning have significantly improved plant disease detection performance, particularly through convolutional neural networks (CNNs) and object detection models (Reya *et al.*, 2022; Girmaw *et al.*, 2024; Sun *et al.*, 2024; Rahman *et al.*, 2025; Kheir *et al.*, 2025; Haris *et al.*, 2025; Agrawal & Sharma, 2021; Magalhães *et al.*, 2021; Sharma *et al.*, 2023; Li *et al.*, 2023; Aleinzi, 2025; Lyu *et al.*, 2025). Among these approaches, YOLO-based architectures have demonstrated strong real-time detection and localization capabilities in agricultural applications (Orchi *et al.*, 2023; Gao *et al.*, 2024; Jeong & Park, 2025; Shen *et al.*, 2025; Joshi *et al.*, 2025). Despite their high detection accuracy, most existing systems primarily generate technical outputs such as

bounding boxes, class labels, and confidence scores, which remain difficult for non-expert users to interpret and utilize for practical decision-making.

This limitation highlights an important gap between detection performance and real-world usability in precision agriculture. Farmers and agricultural practitioners require not only accurate detection results but also clear, contextual, and actionable explanations that support disease management decisions. In this context, structured-to-text generation using large language models (LLMs) provides a promising direction for transforming structured detection outputs into human-readable agricultural reports. Recent studies have demonstrated that LLMs such as LLaMA 3.1-13B-Instruct are capable of generating context-aware and domain-relevant explanations across specialized tasks (Zahid *et al.*, 2024; Balasubramanian *et al.*, 2025; Li *et al.*, 2025a; Lin *et al.*, 2025). However, existing agricultural AI studies still predominantly focus on detection accuracy, while integration between object detection and language generation for interpretable agricultural reporting remains limited.

Although several recent studies have explored the use of LLMs in agriculture, most have focused either on conversational advisory systems or on improving disease classification accuracy. Existing approaches rarely integrate visual object detection outputs with structured-to-text generation to produce comprehensive and actionable disease reports. In particular, previous agricultural AI systems generally stop at reporting disease labels or confidence scores, without transforming these technical outputs into context-aware natural language explanations. Therefore, the proposed framework extends beyond conventional detection pipelines by explicitly bridging object detection and language generation through an intermediate structured representation layer.

To address this gap, this study proposes an end-to-end detection-to-interpretation framework integrating YOLOv8s and LLaMA 3.1-13B-Instruct to generate human-readable cabbage disease reports from structured detection outputs. Unlike conventional plant disease detection systems that only provide visual predictions, the proposed framework transforms technical detection information into context-aware explanations containing disease interpretation, severity assessment, and actionable recommendations. In addition, this study addresses practical challenges by converting a classification-based cabbage disease dataset into an object detection dataset through manual annotation and expert validation.

The main contributions of this study are as follows: (1) the development of a detection-to-interpretation pipeline integrating object detection and LLM-based text generation; (2) the creation of an annotated cabbage disease detection dataset through manual annotation and expert validation; (3) the design of a structured representation that bridges visual detection outputs and natural language generation (NLG); and (4) a comprehensive evaluation combining detection metrics, text generation metrics, and qualitative expert assessment.

Methodology

1. Overall Framework

The proposed system is an end-to-end detection-to-interpretation pipeline that transforms cabbage leaf images into human-readable disease reports by integrating YOLOv8s for disease detection and LLaMA 3.1-13B-Instruct for interpretation. The primary objective of the framework is to bridge the gap between technical visual outputs and actionable agricultural insights, thereby improving interpretability and usability for non-expert users. To achieve this, the framework follows a sequential process divided into three core modules: (i) Module I (Object Detection Pipeline) which processes the raw image data to identify diseases; (ii) Module II (Structured Representation Transformation) which converts visual coordinates into structured JSON formats; and (iii) Module III (LLM Interpretation Pipeline) which generates the final natural language explanations. Its modular architecture ensures flexibility, interpretability, and scalability, allowing future extension to broader agricultural applications. The overall system framework is illustrated in Figure 1.

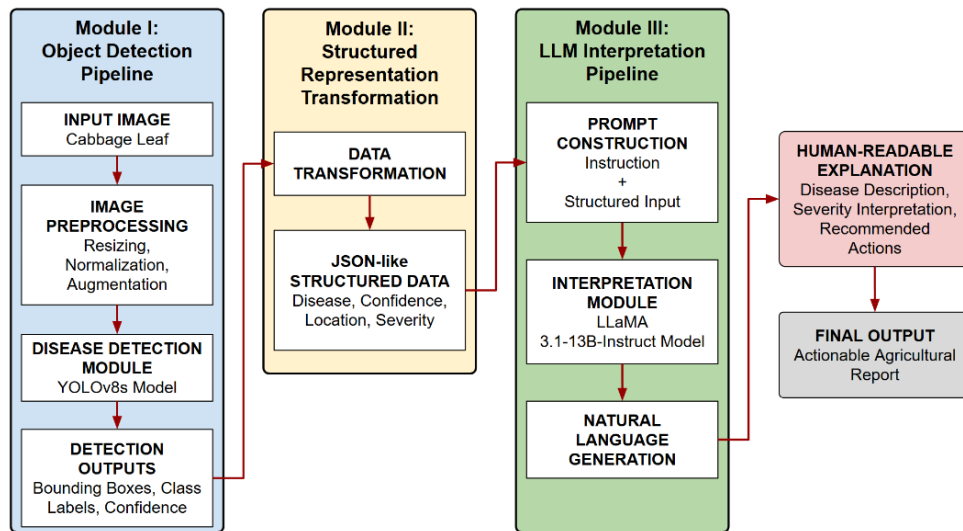


Figure 1 System framework

To provide a clearer formalization of the proposed pipeline, the framework can be represented as a sequential transformation process:

$$I \rightarrow I_{prep} \rightarrow O_{det} \rightarrow D_{json} \rightarrow P_{prompt} \rightarrow E_{text} \rightarrow R$$

Where I denotes the raw input cabbage leaf image, I_{prep} represents the preprocessed image, O_{det} denotes the object detection outputs generated by YOLOv8s, D_{json} represents the transformed structured representation,

P_{prompt} denotes the constructed prompt framework, E_{text} represents the raw textual interpretation from LLaMA 3.1-13B-Instruct, and R is the final human-readable agricultural report.

To improve methodological clarity and reproducibility, the overall system framework is illustrated in Figure 1, while its formal step-by-step execution is detailed in Algorithm 1.

Algorithm 1. End-to-End Detection-to-Interpretation Pipeline

Input:

Cabbage leaf image I

Output:

Human-readable agricultural report R

- 1: $I_{prep} \leftarrow \text{ImagePreprocessing}(I)$ (*Resizing, Normalization, Augmentation*)
 - 2: $O_{det} \leftarrow \text{YOLOv8s}(I_{prep})$
 - 3: Extract features from O_{det} :
 - $B \leftarrow$ bounding boxes
 - $C \leftarrow$ class labels
 - $S_{conf} \leftarrow$ confidence scores
 - 4: $D_{json} \leftarrow \text{TransformStructuredRepresentation}(B, C, S_{conf})$
 - 5: $P_{prompt} \leftarrow \text{ConstructPrompt}(D_{json}, \text{Instruction})$
 - 6: $E_{text} \leftarrow \text{LLaMA 3.1-13B-Instruct}(P_{prompt})$
 - 7: $R \leftarrow \text{NaturalLanguageGeneration}(E_{text})$
 - 8: Return R
-

The algorithm illustrates the sequential transformation from raw visual input into structured semantic representation and finally into natural language agricultural explanations.

The proposed framework consists of two main components: a disease detection module and an interpretation module. The detection module uses YOLOv8s to identify and localize disease symptoms in cabbage leaf images, producing structured outputs in the form of bounding boxes, class labels, and confidence scores. These outputs convert unstructured visual data into structured information for further analysis. The interpretation module utilizes LLaMA 3.1-13B-Instruct to transform these structured outputs into human-readable explanations. It

generates context-aware and actionable reports, including detected diseases, severity levels, affected areas, and recommended treatments, thereby improving interpretability and supporting decision-making.

The overall pipeline includes four stages: image preprocessing, disease detection using YOLOv8s, transformation into structured representation, and text generation using LLaMA 3.1-13B-Instruct. The structured representation acts as a crucial bridge between detection and language generation, encoding key information such as disease type, location, confidence, and estimated severity. When multiple diseases are detected, the system aggregates them into a unified representation. Finally, prompt-based input guides LLaMA 3.1-13B-Instruct to generate clear and practical explanations. The framework is modular, scalable, and user-centric, providing flexibility, interoperability, and practical usability.

In addition to the proposed framework, baseline comparisons and component-level analysis were conducted to better evaluate the contribution of each module. The detection module was compared against Faster R-CNN as a baseline detector, while the text generation module was compared with a template-based reporting approach. Furthermore, ablation analysis was performed by evaluating the framework without structured representation and without few-shot prompting to analyze the impact of intermediate representation and prompt engineering on the quality of generated explanations.

2. Dataset Preparation

This study uses a publicly available cabbage disease dataset from Mendeley Data, created by Ayawli *et al.* (2025). The dataset contains images of cabbage crops affected by various diseases, including Alternaria Leaf Spot, Black Rot, and Downy Mildew, as well as healthy crops. It includes both original images and augmented samples generated using techniques such as rotation, cropping, flipping, scaling, Gaussian noise, Gaussian filtering, and brightness adjustment. The images were collected under diverse environmental conditions to capture realistic variations in lighting, viewing angles, and leaf appearance, thereby enhancing the robustness of model training.

For the purposes of this study, only three classes were selected, namely Black Rot, Downy Mildew, and Healthy leaves. This selection was made to focus on the most common and agriculturally significant cabbage diseases while ensuring a balanced and manageable dataset for both detection and interpretation tasks. The dataset was divided into training, validation, and testing sets using stratified sampling with a ratio of 70:15:15 to preserve class distribution across all subsets. A summary of the dataset is presented in Table 1.

Table 1 Dataset summary

Class	Number of Images
Black Rot	500
Downy Mildew	450
Healthy	550
Total	1,500

Although the dataset contains diverse variations, it was originally designed for classification and lacks spatial information required for object detection. Therefore, additional preprocessing and manual annotation were performed to generate bounding box labels that indicate the location of disease symptoms. This process followed standardized annotation guidelines to ensure consistency and reproducibility.

For diseased samples, bounding boxes were assigned to visible lesion regions corresponding to disease symptoms. In healthy samples, bounding boxes were assigned to the dominant visible cabbage leaf region despite the absence of lesions. This design was intentionally adopted to maintain a unified object detection framework across all classes and to enable the system to explicitly recognize and communicate healthy plant conditions within the same structured representation pipeline. Modeling “Healthy” as a detection class also ensures that the interpretation module can generate affirmative health-related explanations instead of producing empty or undefined outputs. Therefore, the “Healthy” class functions as a contextual detection target representing the absence of visible disease symptoms within the detected leaf region.

To ensure annotation quality, agricultural experts validated the bounding boxes and labels through an iterative review process, reducing bias and improving reliability. In addition to visual annotations, domain experts also created reference textual explanations for each disease, including symptoms, severity, and recommended treatments, which serve as ground truth for evaluating generated outputs.

3. Disease Detection Using YOLOv8s

In this study, the YOLOv8s variant was selected because it provides a balanced trade-off between detection accuracy and computational efficiency (Jiang *et al.*, 2024; Hou *et al.*, 2025; Gan *et al.*, 2025; Qu *et al.*, 2025). Compared to larger variants such as YOLOv8m and YOLOv8l, YOLOv8s requires lower computational resources while still achieving strong localization performance, making it suitable for real-time agricultural deployment.

YOLOv8s operates as a one-stage detector that directly predicts bounding boxes and class probabilities in a single forward pass, enabling efficient object detection with low inference latency. The model was trained on

an annotated dataset with bounding box labels, optimizing localization and classification through a unified loss function.

Input images were resized to 640×640 pixels because this resolution provides a practical balance between detection accuracy and computational efficiency while preserving sufficient visual detail for lesion localization. In addition, 640×640 is widely adopted as a standard training resolution in YOLO-based object detection frameworks, enabling stable optimization and efficient GPU utilization.

The training process was conducted for 100 epochs using the AdamW optimizer with an initial learning rate of 0.001 and a batch size of 16. Training was performed on an NVIDIA RTX 3060 GPU with 12 GB VRAM. Data augmentation techniques including horizontal flipping, rotation, scaling, and brightness adjustment were applied during training to improve robustness and generalization performance.

Model performance was evaluated using precision, recall, and mAP@0.5. Precision measures the correctness of predictions, while recall evaluates the ability to detect all relevant disease regions. The mAP@0.5 metric provides an overall assessment by considering the trade-off between precision and recall. A prediction is considered correct when the Intersection over Union exceeds 0.5 (Wang & Song, 2021).

4. Structured Representation

The raw outputs from YOLOv8s, such as bounding box coordinates, class labels, and confidence scores, are not directly suitable for NLG. Therefore, they are transformed into a structured representation that captures semantically meaningful information in a standardized format. This representation acts as an intermediate layer bridging the detection and interpretation modules.

Each detected instance is converted into key attributes, including disease type, confidence score, spatial location, and severity level. An example of the structured representation is presented below: {"disease": "Black Rot", "confidence": 0.91, "location": "upper leaf region", "severity": "moderate"}. The disease and confidence values are directly obtained from the model, while location is mapped from pixel coordinates into interpretable regions such as upper, middle, or lower leaf areas. Severity is estimated using heuristic rules based on factors such as bounding box size, number of detected regions, lesion area ratio, and lesion distribution. This approach follows common practices in image-based plant disease assessment studies, where the proportion and spatial spread of infected regions are used as indicators of disease severity (Rezaei *et al.*, 2026; Li *et al.*, 2025b; Barbedo, 2019). Let A_d denote the total detected lesion area and A_l denote the estimated leaf area. The severity ratio was calculated using Equation (1).

$$\text{Severity Ratio} = \frac{A_d}{A_l} \quad (1)$$

Where the ratio represents the proportion of infected leaf area relative to the total visible leaf area. Following prior disease severity assessment practices and preliminary expert consultation, severity levels were categorized into mild (< 0.15), moderate ($0.15\text{--}0.35$), and severe (> 0.35). The lower threshold was selected to represent early-stage localized infections with limited lesion spread, while the higher threshold corresponds to extensive visible infection affecting more than one-third of the leaf surface. These thresholds were further refined through iterative validation with agricultural experts to ensure consistency with practical field interpretation.

When multiple disease instances are detected, the system aggregates them into a unified representation by summarizing disease types, averaging confidence scores, and determining overall severity. This ensures concise yet comprehensive information. The structured data is then used as input for the language model through prompt-based generation, enabling coherent, accurate, and context-aware explanations. Overall, this representation transforms low-level outputs into high-level semantic information, improving interpretability, interoperability, and the usefulness of the final results for end users.

5. Text Generation Using LLaMA 3.1-13B-Instruct

In this framework, LLaMA 3.1-13B-Instruct is used to generate natural language explanations from structured detection outputs, aiming to produce human-readable, context-aware, and actionable reports (Zhang *et al.*, 2025; Sorour & Almusallam, 2025). This is achieved through a structured-to-text approach using prompt engineering. Unlike template-based methods, LLaMA 3.1-13B-Instruct enables flexible and context-sensitive generation, allowing explanations to adapt to variations in disease type, severity, and location. The input consists of structured data from the detection module, including disease type, confidence, location, and severity, which is embedded into an instruction-based prompt.

The general prompt template used in this study is formulated as follows:

You are an expert in cabbage diseases. Generate a concise agricultural disease report based on the following structured information. For each detected region, provide a clear and easy-to-understand explanation including the disease type, confidence level, affected leaf region, severity interpretation, and recommended actions suitable for non-expert users. The prompt guides the model to generate outputs containing three key elements: disease description, severity interpretation, and recommended actions. This ensures that the generated explanations are not only descriptive but also actionable for practical agricultural use. The generated output is presented as a concise and coherent narrative that integrates relevant information in a form suitable for non-expert users. The example above illustrates the general instruction-based prompt format used before applying additional few-shot examples shown in Figure 2. To improve reliability and generation consistency, several prompt engineering

strategies were applied, including explicit instruction prompting, few-shot prompting, and controlled decoding parameters. Few-shot prompting was specifically used to guide the model toward generating context-aware and semantically consistent agricultural explanations. In this approach, multiple example pairs of structured inputs and corresponding expert-style outputs were provided within the prompt to establish the expected response pattern. Figure 2 illustrates several examples of input-output pairs (few-shot prompting) used in this study.

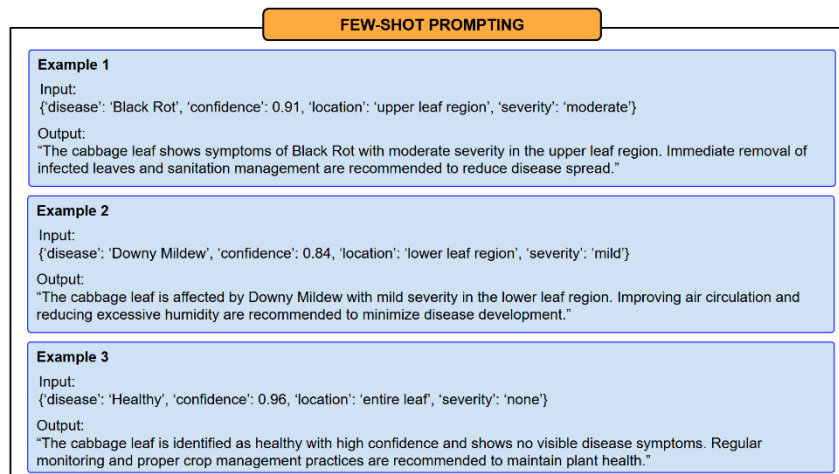


Figure 2 Examples of few-shot prompting used in the study

These examples helped the model learn the desired explanation structure, including disease identification, severity interpretation, spatial context, and actionable recommendations. In addition, controlled decoding parameters were applied with temperature = 0.7, top-p = 0.9, and maximum output length = 256 tokens to balance diversity and coherence. Post-processing steps, including grammatical correction and redundancy removal, were further applied to enhance linguistic quality and readability.

6. Evaluation Methods

To comprehensively assess the performance of the proposed detection-to-interpretation framework, a multi-level evaluation strategy was employed. This evaluation consists of three main components: detection performance evaluation, text generation evaluation, and qualitative expert assessment. This combination ensures that both the technical accuracy of the models and the practical usefulness of the generated outputs are rigorously examined.

6.1 Detection Evaluation

The performance of the YOLOv8s disease detection module was evaluated using precision, recall, and mAP@0.5. Precision measures the proportion of correct positive predictions, reflecting the model's ability to

minimize false positives, while recall evaluates its ability to detect all relevant disease regions. The mAP@0.5 metric provides a comprehensive assessment by combining precision and recall across confidence thresholds. A detection is considered correct if the Intersection over Union exceeds 0.5, and the final score is obtained by averaging precision across all classes. Overall, these metrics ensure a balanced evaluation of both detection accuracy and localization performance.

6.2 Text Evaluation

The quality of generated explanations was evaluated using multiple metrics to capture both lexical similarity and semantic alignment with expert references. BLEU measures n-gram overlap and focuses on precision, assessing how closely the generated text matches the reference wording (Gupta *et al.*, 2024; Onan & Alhumyani, 2024; Syahidi & Kiyokawa, 2025). ROUGE-L evaluates the longest common subsequence, emphasizing recall and content coverage at the sentence level (Citarella *et al.*, 2025; Syahidi *et al.*, 2025; Onan & Alhumyani, 2024). METEOR improves upon BLEU by considering synonyms, stemming, and word order, providing a more balanced evaluation of precision, recall, and semantic similarity (Mahyoub *et al.*, 2025; Onan & Alhumyani, 2024). Meanwhile, BERTScore measures semantic similarity using contextual embeddings, enabling deeper evaluation beyond surface-level matching (Tang *et al.*, 2024; Oro *et al.*, 2025). By combining these metrics, the evaluation provides a comprehensive assessment of both textual similarity and semantic correctness in the generated outputs.

6.3 Qualitative Evaluation

In addition to quantitative metrics, a qualitative evaluation was conducted to assess the practical usefulness of the generated explanations. This evaluation involved three agricultural experts with domain knowledge in plant disease management. The experts independently reviewed a set of generated reports and assessed their quality based on three criteria. Accuracy refers to the correctness of the information presented in the generated explanation, including the identification of the disease and the appropriateness of the interpretation. Clarity evaluates how easily the explanation can be understood by non-expert users. This includes the use of simple language, logical structure, and avoidance of ambiguity. Actionability measures the extent to which the generated explanation provides useful and practical recommendations that can support real-world decision-making.

Each explanation was rated using a predefined scoring scale, and the results were aggregated to obtain an overall qualitative performance measure. To assess the consistency of expert judgments, inter-rater agreement was calculated using Cohen's Kappa coefficient. This statistical measure evaluates the level of agreement between raters beyond chance, providing an indication of the reliability of the evaluation process. A high Kappa value indicates strong agreement among experts, confirming the robustness of the qualitative assessment.

Results

1. Detection Performance

The YOLOv8s model was evaluated using precision, recall, and mAP@0.5, achieving scores of 0.85, 0.83, and 0.87, respectively, indicating strong performance in cabbage disease detection. The high precision shows that most detections are accurate with few false positives, while the recall indicates that the model can identify most disease instances. This balance is important to ensure accurate detection without excessive false alarms. The mAP@0.5 score of 0.87 reflects strong overall performance in both localization and classification, demonstrating that the model can accurately detect and classify disease regions, making it suitable for real-world agricultural applications. To better contextualize the detection performance, the proposed YOLOv8s model was benchmarked against Faster R-CNN, which is widely used in agricultural object detection studies due to its strong localization capability. As shown in Table 2, YOLOv8s achieved superior performance with an mAP@0.5 of 0.87 compared to 0.81 obtained by Faster R-CNN, while also providing significantly faster inference suitable for real-time deployment. These results indicate that YOLOv8s offers a more effective balance between detection accuracy and computational efficiency for practical precision agriculture scenarios. Further class-level analysis in Table 2 shows that the model performs better on Black Rot than Downy Mildew, likely due to the more visually distinctive symptoms of Black Rot. In contrast, Downy Mildew often presents subtle texture variations and lower contrast lesion patterns, making detection more challenging.

Table 2 Per-class detection performance

Class	Precision	Recall	mAP@0.5
Black Rot	0.87	0.84	0.89
Downy Mildew	0.83	0.81	0.85
Healthy	0.86	0.85	0.88
Average	0.85	0.83	0.87

Further insights were obtained through confusion matrix analysis, as shown in Table 3. The results reveal that most misclassifications occur between Black Rot and Downy Mildew, indicating that the model occasionally struggles to distinguish diseases with similar visual characteristics. Quantitatively, 10 Black Rot samples were incorrectly classified as Downy Mildew, while 12 Downy Mildew samples were predicted as Black Rot. In contrast, only a small number of healthy samples were misclassified, demonstrating that the model effectively distinguishes diseased leaves from healthy leaves. This is particularly important for minimizing false alarms in practical agricultural applications.

Table 3 Confusion matrix

Actual \ Predicted	Black Rot	Downy Mildew	Healthy
Black Rot	84	10	6
Downy Mildew	12	81	7
Healthy	5	6	89

A more detailed error analysis reveals several major sources of detection errors. Approximately 41% of errors were associated with visually similar early-stage symptoms between Black Rot and Downy Mildew. Another 27% were caused by small lesion regions occupying limited image areas, making localization difficult. Variations in illumination and low-contrast conditions contributed approximately 19% of errors, while the remaining 13% were related to minor annotation inconsistencies and overlapping symptoms. Despite these challenges, the model remains robust due to expert-validated annotations, data augmentation, and the ability of YOLOv8s to capture multi-scale visual features.

Representative detection examples shown in Figure 3 demonstrate accurate localization and classification with high confidence scores across multiple disease conditions. The model successfully identifies disease regions and healthy leaf areas while maintaining relatively stable localization performance under varying visual conditions. Minor localization errors are primarily observed in overlapping lesion regions and low-contrast images.

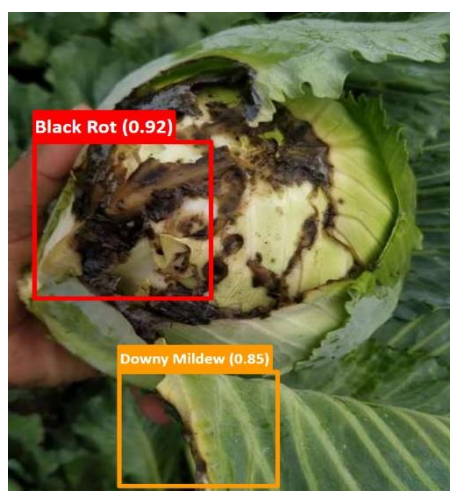


Figure 3 Example detection results using YOLOv8s

Figure 3 presents representative detection outputs produced by the YOLOv8s model on cabbage leaf images. The model successfully identifies disease regions using bounding boxes and class labels with associated confidence scores. The examples demonstrate accurate localization of Black Rot and Downy Mildew symptoms, including cases involving multiple disease instances. Minor detection errors are mainly observed in images containing overlapping symptoms or low-contrast lesion regions.

Overall, these results confirm that the YOLOv8s model provides a robust and reliable foundation for the subsequent interpretation stage. Accurate detection is essential in the proposed framework because errors in this stage may propagate to the language generation module. Therefore, strong detection performance plays a critical role in ensuring the quality, reliability, and trustworthiness of the final human-readable explanations.

2. Text Generation Results and Performance

The performance of the text generation module was evaluated using multiple automatic evaluation metrics to assess both lexical similarity and semantic alignment between the generated explanations and the reference texts created by domain experts. The results are summarized in Table 4.

Table 4 Text generation performance

Metric	Score (Mean $\pm$ Std)
BLEU	0.48 $\pm$ 0.07
ROUGE-L	0.61 $\pm$ 0.06
METEOR	0.54 $\pm$ 0.05
BERTScore	0.89 $\pm$ 0.03

The evaluation shows that the proposed framework achieves moderate to high performance across multiple metrics. A BLEU score of 0.48 indicates moderate n-gram overlap, reflecting expected variability in phrasing for agricultural descriptions. ROUGE-L (0.61) and METEOR (0.54) demonstrate strong sentence-level alignment and semantic flexibility, capturing essential information despite lexical differences. A high BERTScore of 0.89 confirms strong semantic consistency with expert references, emphasizing meaning preservation rather than exact wording. In addition, the low standard deviations across all metrics indicate stable and reliable generation performance.

Table 5 Baseline comparison for text generation

Method/Model	BLEU	ROUGE-L	METEOR	BERTScore
Template-based reporting	0.39	0.52	0.46	0.82
Lightweight LLM (LLaMA 3.2-3B-Instruct)	0.43	0.57	0.50	0.85
Proposed (LLaMA 3.1-13B-Instruct)	0.48	0.61	0.54	0.89

Table 5 presents a systematic comparison between the proposed framework and two baseline approaches. The results indicate that the proposed LLaMA 3.1-13B-Instruct framework consistently outperforms both the template-based reporting system and the lightweight language model across all evaluation metrics. In particular, the higher BERTScore demonstrates stronger semantic alignment with expert-written references, while the improvements in BLEU, ROUGE-L, and METEOR indicate better lexical and contextual quality. These findings confirm the benefit of larger instruction-tuned language models for generating coherent and actionable agricultural disease reports.

To further evaluate the effectiveness of the proposed language generation module, Table 5 presents a systematic comparison between the proposed framework and two baseline approaches: (1) a conventional template-based reporting pipeline and (2) a lightweight instruction-tuned language model, LLaMA 3.2-3B-Instruct. The template-based approach generated rigid and repetitive outputs with limited contextual flexibility, achieving lower BLEU (0.39) and BERTScore (0.82) compared with the proposed framework. Meanwhile, LLaMA 3.2-3B-Instruct was evaluated using the same structured input format, identical prompt template, few-shot examples, decoding parameters (temperature = 0.7, top-p = 0.9), and maximum output length (256 tokens). Although computationally more efficient, the smaller model produced less coherent explanations with weaker domain-specific recommendations and reduced semantic consistency than LLaMA 3.1-13B-Instruct. Overall, the results demonstrate that the proposed framework consistently outperforms both baseline approaches across all evaluation metrics. In addition, an ablation study was performed to analyze the contribution of several key components in the proposed pipeline. Removing the structured representation layer reduced BERTScore from 0.89 to 0.83, indicating that the intermediate semantic representation significantly improves contextual consistency. Similarly, removing few-shot prompting resulted in more generic and repetitive explanations, reducing BLEU from 0.48 to 0.42. These findings demonstrate that both the structured representation and prompt engineering strategies contribute substantially to the quality of the generated explanations. More specifically, when the structured representation module was removed, raw detection outputs were directly inserted into the prompt without semantic organization. This often weakened the contextual relationships among disease type, spatial location, and severity level, resulting in lower semantic alignment with expert references. Likewise, removing few-shot prompting eliminated exemplar

guidance for the language model, leading to more generic, repetitive, and less domain-specific explanations. These observations suggest that the structured representation and few-shot prompting components play complementary roles in improving explanation quality, contextual coherence, and semantic consistency (Berhe *et al.*, 2026).

Qualitatively, the generated explanations effectively integrate disease type, severity, spatial context, and recommended actions into concise and informative narratives aligned with agricultural domain knowledge. Minor phrasing variations and occasional redundancy slightly affect BLEU scores but do not substantially reduce interpretability or practical usefulness. Overall, these results demonstrate that the integration of LLaMA 3.1-13B-Instruct within the proposed framework produces high-quality, human-readable explanations that support actionable decision-making in precision agriculture.

3. Qualitative Evaluation

In addition to quantitative evaluation, a qualitative assessment was conducted to evaluate the practical usefulness of the generated explanations in real-world agricultural contexts. A total of 100 randomly selected generated explanations were independently reviewed by three agricultural experts. This evaluation assessed the outputs based on three key criteria, namely accuracy, clarity, and actionability. To ensure consistency and objectivity in the evaluation process, a standardized scoring rubric was developed using a 5-point Likert scale (See Table 6).

Table 6 Expert evaluation rubric (5-point scale)

Score	Accuracy	Clarity	Actionability
5	Completely correct disease identification and interpretation; no errors	Very clear, easy to understand, no ambiguity	Provides highly specific, practical, and immediately applicable actions
4	Minor inaccuracies but overall correct interpretation	Clear with slight ambiguity but still easily understandable	Provides relevant and useful actions with minor generalization
3	Partially correct but includes noticeable inaccuracies	Moderately clear but requires some interpretation by the user	Provides general advice but lacks specificity or detail
2	Mostly incorrect or misleading interpretation	Difficult to understand due to unclear or poorly structured language	Provides vague or limited actions with low practical usefulness
1	Completely incorrect or irrelevant explanation	Very unclear and confusing	No actionable recommendation or completely irrelevant advice

Experts independently scored outputs on accuracy, clarity, and actionability, with the final assessment averaged across criteria. Explanations scoring 4 or higher were considered high-quality, and 88% met this standard, indicating that the generated outputs are accurate, understandable, and actionable for non-expert users such as farmers. Most explanations correctly reflected disease type, interpreted severity appropriately, and provided practical recommendations aligned with agricultural guidelines.

Inter-rater agreement analysis produced a Cohen's Kappa coefficient of $K = 0.81$, indicating strong agreement among the experts beyond chance. According to commonly accepted interpretation guidelines, a Kappa value above 0.80 represents almost perfect agreement, confirming the reliability and consistency of the qualitative evaluation process. This result suggests that the generated explanations were consistently interpreted as accurate and useful by multiple evaluators with agricultural expertise.

The remaining 12% of outputs exhibited minor issues, including slight inaccuracies in severity interpretation, overly general recommendations, and occasional redundancy in phrasing. More specifically, approximately 5% of the evaluated outputs contained generalized recommendations lacking disease-specific detail, while 4% exhibited repetitive sentence structures. Another 3% showed slight inconsistencies between detected severity and generated interpretation. These findings indicate opportunities for future improvements in severity estimation, prompt engineering, and linguistic post-processing.

Overall, the qualitative evaluation demonstrates that the YOLOv8s-LLaMA 3.1-13B-Instruct framework produces interpretable, actionable, and reliable outputs, effectively bridging the gap between disease detection and user-centered interpretation in precision agriculture.

4. End-to-End System Demonstration

In addition to quantitative and qualitative evaluations, this study further demonstrates the effectiveness of the proposed framework through an end-to-end system analysis. This demonstration highlights how detection outputs generated by YOLOv8s are transformed into structured representations and subsequently converted into human-readable explanations using LLaMA 3.1-13B-Instruct. This process provides a comprehensive view of how the system operates in practical scenarios and validates its real-world applicability.

4.1 From Detection Output to Structured Representation

The YOLOv8s model produces detection outputs in the form of bounding boxes, class labels, and confidence scores. However, these outputs are not directly interpretable by end users. Therefore, the detection results are transformed into a structured format that captures semantically meaningful attributes, including disease type, confidence level, spatial location, and severity.

In this process, bounding box coordinates are mapped into interpretable leaf regions, such as upper leaf region, lower leaf region, or center area. Additionally, a severity level is inferred using heuristic rules based on factors such as bounding box size and the number of detected regions. An example of detection output is presented in Table 7.

Table 7 Example of detection results

Disease	Confidence	Bounding Box (x1, y1, x2, y2)
Black Rot	0.94	(45, 60, 180, 200)
Downy Mildew	0.81	(200, 50, 330, 180)

The corresponding structured representation is generated as follows:

```
[{"disease": "Black Rot", "confidence": 0.94, "location": "upper leaf region", "severity": "moderate"},
{"disease": "Downy Mildew", "confidence": 0.81, "location": "center leaf region", "severity": "mild"}]
```

This structured representation ensures that the detection outputs are both machine-readable and semantically meaningful, enabling effective integration with the language generation module.

4.2 Text Generation Process Using LLaMA 3.1-13B-Instruct

The structured data is then used as input for LLaMA 3.1-13B-Instruct through a prompt-based mechanism. The prompt is designed to instruct the model to generate explanations that are clear, informative, and actionable for agricultural users.

The prompt template used in this study is defined as follows:

You are an expert in cabbage diseases. Generate a concise agricultural disease report based on the following structured information. For each detected region, provide a clear and easy-to-understand explanation including the disease type, confidence level, affected leaf region, severity interpretation, and recommended actions suitable for non-expert users.

Based on this structured input, the model generates coherent explanations that integrate all relevant attributes into a single narrative.

4.3 Example of Generated Outputs and Expert Validation

To further validate the effectiveness of the system, several generated outputs are compared with expert evaluations, as shown in Table 8.

Table 8 Example of generated explanations and expert evaluation

Structured Input	Generated Explanation (LLaMA 3.1-13B-Instruct)	Expert Comment
Black Rot, 0.91, upper leaf region, moderate	The cabbage leaf shows symptoms of Black Rot with moderate severity, primarily located in the upper leaf region. This disease typically causes yellowing that progresses to dark necrotic tissue. It is recommended to remove infected leaves and improve field sanitation to prevent further spread.	Accurate and actionable
Downy Mildew, 0.83, lower leaf region, mild	The system detected Downy Mildew with mild severity in the lower leaf region. This condition appears as pale spots and may develop further under humid conditions. Improving air circulation and avoiding excessive moisture are recommended.	Accurate and practical
Healthy, 0.95, entire leaf, none	The cabbage leaf is identified as healthy with high confidence, showing no visible disease symptoms. Regular monitoring and proper crop management practices are recommended to maintain plant health.	Accurate and informative

The results demonstrate that the generated explanations effectively incorporate all structured attributes, including severity, and align well with expert knowledge.

Discussion

This study presents a comprehensive detection-to-interpretation framework that integrates YOLOv8s for disease detection and LLaMA 3.1-13B-Instruct for generating human-readable explanations. The experimental results demonstrate strong performance across detection accuracy, text generation quality, and qualitative expert evaluation. This section discusses these findings in depth, focusing on their implications, strengths, limitations, and broader contributions to precision agriculture.

1. Analysis of Detection Performance

The detection results indicate that YOLOv8s achieves strong overall performance, with a precision of 0.85, recall of 0.83, and mAP@0.5 of 0.87. These results suggest that the model is capable of accurately identifying and localizing cabbage disease symptoms while maintaining a balanced trade-off between false positives and false negatives. In practical agricultural settings, this balance is particularly important. High precision reduces the risk of

unnecessary interventions such as excessive pesticide application, while high recall ensures that most disease instances are detected and not overlooked.

Compared with the Faster R-CNN baseline, YOLOv8s achieved higher detection performance, improving mAP@0.5 from 0.81 to 0.87 while also providing lower inference latency suitable for real-time deployment. This finding indicates that one-stage detectors such as YOLOv8s can provide a more effective balance between computational efficiency and localization accuracy for agricultural disease detection tasks. The obtained performance is also comparable to or slightly higher than several recent YOLO-based plant disease detection studies reported in the literature, suggesting that the proposed framework achieves competitive performance despite using a relatively limited dataset size and focusing on an end-to-end interpretation-oriented pipeline rather than detection alone.

A more detailed class-level analysis reveals that the model performs better on Black Rot compared to Downy Mildew. Specifically, Black Rot achieved an mAP@0.5 of 0.89, while Downy Mildew achieved 0.85. This difference highlights an important characteristic of vision-based disease detection systems, namely their sensitivity to visual distinctiveness. Black Rot typically presents clearer and more localized visual patterns, which are easier for convolutional models to learn. In contrast, Downy Mildew often exhibits diffuse and lower-contrast symptoms, making it inherently more difficult to detect. This observation aligns with known challenges in plant pathology, where subtle or early-stage symptoms are more difficult to identify, even for human experts.

The confusion matrix analysis further strengthens this interpretation. Quantitatively, 10 Black Rot samples were incorrectly classified as Downy Mildew, while 12 Downy Mildew samples were predicted as Black Rot. In contrast, the Healthy class achieved the highest number of correct predictions, with 89 correctly classified samples and only 11 total misclassifications. These findings indicate that the primary challenge lies in differentiating visually similar disease classes rather than distinguishing diseased leaves from healthy leaves. From a practical perspective, the low confusion involving healthy samples is particularly important because it minimizes unnecessary warnings and reduces the likelihood of false-positive agricultural interventions.

Error analysis reveals several major contributing factors to detection failures. Approximately 41% of errors were associated with visually similar early-stage symptoms between Black Rot and Downy Mildew. Another 27% resulted from small lesion regions occupying limited image areas, making localization more difficult. Variations in illumination and low-contrast conditions contributed approximately 19% of errors, while the remaining 13% were related to minor annotation inconsistencies and overlapping symptoms. These findings suggest that future improvements may benefit from more diverse environmental training data, lesion segmentation techniques, or attention-based architectures capable of capturing subtle disease characteristics more effectively.

Representative detection examples shown in Figure 3 demonstrate accurate localization and classification with high confidence scores across multiple disease conditions. The model successfully identifies disease regions and healthy leaf areas while maintaining relatively stable localization performance under varying visual conditions. Minor localization errors are primarily observed in overlapping lesion regions and low-contrast images.

Overall, the detection module provides a robust foundation for the proposed framework. This is particularly important because errors at this stage can propagate to the interpretation module. Therefore, the strong detection performance directly contributes to the reliability, consistency, and trustworthiness of the final generated explanations.

2. Analysis of Text Generation Performance

The text generation results show that the proposed framework achieves moderate lexical similarity but strong semantic alignment with expert references. Although the BLEU score of 0.48 indicates that exact word matching is limited, this is acceptable in structured-to-text generation tasks where multiple sentence variations may convey the same meaning. Higher ROUGE-L (0.61), METEOR (0.54), and especially BERTScore (0.89) confirm that the generated texts successfully preserve essential semantic information, domain knowledge, and contextual relevance rather than relying solely on rigid lexical overlap. The low standard deviations across metrics further indicate stable and consistent generation performance, which is important for maintaining user trust in decision-support systems.

Compared with the template-based reporting baseline, the proposed LLaMA 3.1-13B-Instruct framework produces more flexible, context-aware, and semantically rich explanations. The template-based approach achieved lower BLEU (0.39) and BERTScore (0.82), primarily because it relies on fixed sentence structures that cannot adapt effectively to variations in disease severity, location, or contextual information. Similarly, the alternative lightweight language model generated less coherent explanations with weaker domain-specific recommendations, highlighting the importance of instruction-tuned LLMs for agricultural interpretation tasks.

The ablation study further demonstrates the importance of the structured representation and prompt engineering components within the proposed framework. Removing the structured representation layer reduced BERTScore from 0.89 to 0.83, indicating that the intermediate semantic representation significantly improves contextual consistency and information preservation. Likewise, removing few-shot prompting reduced BLEU from 0.48 to 0.42 and resulted in more repetitive and generic outputs. These findings confirm that both the structured intermediate representation and prompt design play critical roles in producing coherent and contextually appropriate agricultural explanations.

Functionally, the generated explanations effectively integrate multiple structured input attributes, including disease type, severity, confidence level, spatial location, and recommended actions. Minor issues such as occasional redundancy and slight phrasing variations are still observed but do not substantially reduce interpretability or practical usefulness. These limitations may be further mitigated through domain-specific fine-tuning, improved prompt optimization, or more advanced post-processing strategies.

Overall, the results demonstrate that LLaMA 3.1-13B-Instruct can successfully transform structured disease detection outputs into informative, human-readable explanations that align closely with expert knowledge while maintaining semantic accuracy and practical usability.

3. Insights from Qualitative Evaluation

The qualitative evaluation demonstrates the practical effectiveness of the proposed framework, with 88% of generated explanations rated as accurate, clear, and actionable by agricultural experts. These findings indicate that the structured representation successfully preserves essential information from the YOLOv8s detection outputs, enabling LLaMA 3.1-13B-Instruct to produce factually correct, concise, and understandable explanations. Transforming technical detection outputs into accessible language is particularly valuable for non-expert users, including farmers and agricultural practitioners.

In addition, the generated recommendations were generally aligned with agricultural management guidelines, supporting informed decision-making and practical disease mitigation strategies. The strong inter-rater agreement (Cohen's $K = 0.81$) further confirms the reliability and consistency of the expert evaluation process. According to standard interpretation criteria, a Kappa value above 0.80 indicates almost perfect agreement, suggesting that the generated explanations were consistently perceived as useful and accurate across multiple evaluators.

Analysis of the lower-rated outputs reveals several remaining limitations. Approximately 5% of explanations contained recommendations that were overly general and lacked disease-specific detail, while 4% exhibited repetitive phrasing patterns. Another 3% showed slight inconsistencies between the detected severity level and the generated interpretation. These findings suggest that future improvements should focus on more precise severity estimation, enhanced prompt engineering strategies, and refinement of linguistic post-processing mechanisms to improve fluency and contextual specificity.

Overall, the qualitative evaluation demonstrates that the proposed framework effectively bridges the gap between technical disease detection and user-centered agricultural interpretation, providing reliable and actionable outputs suitable for precision agriculture applications.

4. End-to-End System Implications

The end-to-end system demonstrates the effectiveness of integrating disease detection, structured representation, and language generation into a unified pipeline capable of transforming raw image inputs into meaningful agricultural reports. The structured representation layer serves as a crucial intermediate component that enables effective communication between the visual detection module and the language generation module while improving interpretability and modularity.

The framework is also capable of handling multiple disease instances within a single image and generating coherent explanations that integrate all relevant attributes into a unified narrative. This capability is particularly important in real-world agricultural environments where multiple symptoms may coexist under varying environmental conditions.

More broadly, the proposed framework shifts agricultural AI from detection-centric systems toward interpretation-oriented decision support. Rather than providing only technical outputs such as bounding boxes and confidence scores, the system translates visual predictions into actionable insights that can be directly utilized by non-expert users. This contributes to improving accessibility, usability, and practical adoption of AI technologies in precision agriculture.

Conclusions

This study presented an end-to-end detection-to-interpretation framework that integrates YOLOv8s and LLaMA 3.1-13B-Instruct to transform cabbage disease detection outputs into human-readable and actionable agricultural reports. The experimental results demonstrate that the proposed framework achieves strong detection performance, with YOLOv8s obtaining high precision, recall, and mAP@0.5 scores while outperforming the Faster R-CNN baseline in both accuracy and real-time suitability. In addition, the language generation module produced explanations with strong semantic alignment to expert references, supported by high BERTScore performance and positive qualitative evaluation results from agricultural experts.

The study further demonstrates that the structured representation layer effectively bridges visual detection and NLG by converting technical outputs into semantically meaningful information suitable for interpretation. Qualitative evaluation confirmed that most generated explanations were accurate, clear, and actionable, indicating the practical usefulness of the framework for non-expert agricultural users. Overall, the proposed system contributes toward more interpretable, user-centered, and decision-supportive AI systems in precision agriculture by extending disease detection into actionable agricultural interpretation. Nevertheless, the current study is constrained by the limited number of disease classes, environmental variability represented in the dataset, and the use of heuristic

rules for severity estimation. These limitations may affect the generalizability of the framework and motivate future work toward broader disease coverage and more robust severity assessment methods.

Limitations and Future Work

Despite its promising performance, several limitations remain. The current severity estimation mechanism is heuristic-based and may be less reliable in cases involving overlapping lesions, irregular symptom patterns, or subtle early-stage infections. In addition, the dataset is limited to several cabbage disease categories collected under constrained environmental conditions, which may affect generalization to broader agricultural scenarios. The language generation module also occasionally produces redundant or overly generalized recommendations.

Future work will focus on several concrete technical improvements. First, pixel-level lesion segmentation and vision transformer-based attention mechanisms will be explored to improve fine-grained disease localization and severity estimation. Second, temporal and multi-modal agricultural data, including environmental conditions such as humidity and temperature, may be incorporated to support context-aware disease interpretation and progression analysis. Third, domain-specific fine-tuning and retrieval-augmented generation techniques will be investigated to improve recommendation specificity and reduce linguistic redundancy.

From a practical perspective, the proposed framework has strong potential for deployment in real-world agricultural environments, including mobile-based crop monitoring applications, greenhouse decision-support systems, and edge-AI-assisted smart farming platforms. Future studies should therefore include large-scale field validation, real-time deployment experiments on resource-constrained devices, and user-centered evaluations involving farmers and agricultural extension officers to further assess usability, reliability, and operational impact in precision agriculture.

References

- Agrawal, N. K., & Sharma, A. K. (2021). Detection of diseases in plants using convolutional neural networks. In Proc. 10th Int. Conf. SMART (pp. 215-217). IEEE. <https://doi.org/10.1109/SMART52563.2021.9676299>
- Aleinzi, M. (2025). Enhancing precision agriculture with YOLOv8: a deep learning approach to potato disease identification. *Int. J. Adv. Comput. Sci. Appl.*, 16(4), 1156-1166.
<https://dx.doi.org/10.14569/IJACSA.2025.01604110>
- Ayawli, B. B. K., Kopri, G., Dawson, J. K., Bediako-Kyeremeh, B., Buabeng, S. A., Deveh, D., Ayawli, E. I. B., Badu, E., & Lamusah, D. (2025). Image dataset of Cabbage Crop diseases. Mendeley Data, V1.
<https://data.mendeley.com/datasets/sjmgzhwrxv/1>

- Balasubramanian, J. B., Adams, D., Roxanis, L., Gonzalez, A. B. d., Coulson, P., Almeida, J. S., & García-Closas, M. (2025). Leveraging large language models for structured information extraction from pathology reports. *J. Pathol. Inform.*, 1(1), 100521, 1-28. <https://doi.org/10.1016/j.jpi.2025.100521>
- Barbedo, J. G. A. (2019). Plant disease identification from individual lesions and spots using deep learning. *Biosyst. Eng.*, 180(4), 96-107. <https://doi.org/10.1016/j.biosystemseng.2019.02.002>
- Berhe, T. S., Gebreab, S., Musamih, A., AlKhader, W., Salah, K., Ellahham, S., & Jayaraman, R. (2026). Few-shot prompting strategies for improving large language model-based cardiovascular disease risk prediction. *Array*, 30(2026), 100819, 1-17. <https://doi.org/10.1016/j.array.2026.100819>
- Citarella, A. A., Barbella, M., Ciobanu, M. G., Marco, F. D., Biasi, L. D., & Tortora, G. (2025). Assessing the effectiveness of ROUGE as unbiased metric in extractive vs. abstractive summarization techniques. *J. Comput. Sci.*, 87(1), 102571, 1-17. <https://doi.org/10.1016/j.jocs.2025.102571>
- Ebron, J. G., Apuada, G. R. B., Parra, D. L. R., & Virtucio, J. P. F. T. (2024). Classification of leafy diseases of brassica rapa (Chinese cabbage) using selected parameters in an uncontrolled environment via image processing. *Chem. Eng. Trans.*, 113(1), 451-456. <https://doi.org/10.3303/CET24113076>
- Gan, B., Pu, G., Xing, W., Wang, L., & Liang, S. (2025). Enhanced YOLOv8 with lightweight and efficient detection head for detecting rice leaf diseases. *Sci. Rep.*, 15(1), 22179, 1-13. <https://doi.org/10.1038/s41598-025-06843-8>
- Gao, X., Wang, G., Qi, J., Wang, Q., Xiang, M., Song, K., & Zhou, Z. (2024). Improved YOLOv7 for sustainable agriculture significantly improves precision rate for Chinese cabbage (*brassica pekinensis* rupe.) seedling belt (CCSB) detection. *Sustainability*, 16(11), 4759, 1-20. <https://doi.org/10.3390/su16114759>
- Girmaw, D. W., Salau, A. O., Mamo, B. S., & Molla, T. L. (2024). A novel deep learning model for cabbage leaf disease detection and classification. *Discover Appl. Sci.*, 6(1), 521, 1-20. <https://doi.org/10.1007/s42452-024-06233-1>
- Gupta, A., Rastogi, A., Malhotra, H., & Rangarajan, K. (2024). Comparative evaluation of large language models for translating radiology reports into Hindi. *Indian J. Radiol. Imaging*, 35(1), 88-96. <https://doi.org/10.1055/s-0044-1789618>

- Haris, M., Tripathi, A., Pandey, S. S., Tiwari, H., & Ansari, A. A. (2025). A deep learning-based model for plant disease detection. *J. Neonatal Surg.*, 14(5s), 913-925. <https://doi.org/10.52783/jns.v14.3655>
- Hou, S., Pang, Y., Wang, J., Hou, J., & Wang, B. (2025). RS-YOLO: A highly accurate real-time detection model for small-target pest. *Smart Agric. Technol.*, 12(1), 101212, 1-13. <https://doi.org/10.1016/j.atech.2025.101212>
- Jeong, G-S., & Park, J-H. (2025). Early-stage cabbage monitoring using UAV multispectral imagery and a modified faster R-CNN: enhanced object delineation with ExG-based dissolve. *Korean J. Remote Sens.*, 41(1), 73-86. <https://doi.org/10.7780/kjrs.2025.41.1.7>
- Jiang, P., Qi, A., Zhong, J., Luo, Y., Hu, W., Shi, Y., & Liu, T. (2024). Field cabbage detection and positioning system based on improved YOLOv8n. *Plant Methods*, 20(1), 96, 1-15. <https://doi.org/10.1186/s13007-024-01226-y>
- Joshi, K., Kumar, S., Kumar, V., Nandal, R., Kumar, Y., Tuteja, N., Gill, R., & Gill, S. S. (2025). AgriCure: A web application based layered augmentation-enhanced YOLOv8 for disease and nutrient deficiency detection in bitter melon leaves. *Curr. Plant Biol.*, 43(1), 100517, 1-14. <https://doi.org/10.1016/j.cpb.2025.100517>
- Kheir, A. M. S., Koubaa, A., Kolluru, V., Mungara, S., & Feike, T. (2025). Smart plant disease diagnosis using multiple deep learning and web application integration. *J. Agric. Food Res.*, 21(1), 101948, 1-13. <https://doi.org/10.1016/j.jafr.2025.101948>
- Li, D., Gupta, K., Bhaduri, M., Sathiadoss, P., Bhatnagar, S., & Chong, J. (2025a). Comparative diagnostic accuracy of GPT-4o and LLaMA 3-70b: Proprietary vs. open-source large language models in radiology. *Clin. Imaging*, 118(1), 110382, 1-15. <https://doi.org/10.1016/j.clinimag.2024.110382>
- Li, K., Song, Y., Zhu, X., & Zhang, L. (2025b). A severity estimation method for lightweight cucumber leaf disease based on DM-BiSeNet. *Inf. Process. Agric.*, 12(1), 68-79. <https://doi.org/10.1016/j.inpa.2024.03.003>
- Li, J., Li, J., Zhao, X., Su, X., & Wu, W. (2023). Lightweight detection networks for tea bud on complex agricultural environment via improved YOLOv4. *Comput. Electron. Agric.*, 211(1), 107955. <https://doi.org/10.1016/j.compag.2023.107955>

- Lin, K-H., Kao, T-H., Wang, L-C., Kuo, C-T., Chen, P. C-H., Chu, Y-C., & Yeh, Y-C. (2025). Benchmarking large language models GPT-4o, llama 3.1, and qwen 2.5 for cancer genetic variant classification. *npj Precis. Oncol.*, 9(1), 141, 1-10. <https://doi.org/10.1038/s41698-025-00935-4>
- Liu, M., Yang, B., Wang, X., Xu, X., Ding, W., He, P., & Zhou, W. (2025). Co-benefits of a customized nutrient management approach tailored to smallholder farming for cabbage (*Brassica oleracea* L.). *Eur. J. Agron.*, 163(127453), 1-20. <https://doi.org/10.1016/j.eja.2024.127453>
- Lyu, Y., Kuswidiyanto, L. W., Wang, P., Noh, H-H., Jung, H-Y., & Han, X. (2025). One-dimensional convolutional neural network for automated kimchi cabbage downy mildew detection using aerial hyperspectral images. *Remote Sens.*, 17(9), 1626, 1-23. <https://doi.org/10.3390/rs17091626>
- Ma, H., Kong, C., Deng, S., Zhao, T., Ji, J., Wang, Y., Zhang, Y., Zhuang, M., Yang, L., Lebedeva, M., Taranov, V., Artemyeva, A. M., Hu, Z., Yu, J., Fang, Z., & Lv, H. (2025). Resistance screening of cabbage to black rot and inheritance pattern analysis. *Sci. Hortic.*, 345(1), 114129, 1-7. <https://doi.org/10.1016/j.scienta.2025.114129>
- Magalhães, S. A., Castro, L., Moreira, G., Santos, F. N. d., Cunha, M., Dias, J., & Moreira, A. P. (2021). Evaluating the single-shot multibox detector and YOLO deep learning models for the detection of tomatoes in a greenhouse. *Sensors*, 21(10), 3569, 1-24. <https://doi.org/10.3390/s21103569>
- Mahyoub, M., Wang, Y., & Khasawneh, M. T. (2025). GPT-4o in radiology: in-context learning based automatic generation of radiology impressions. *Nat. Lang. Process. J.*, 11(1), 100145, 1-10. <https://doi.org/10.1016/j.nlp.2025.100145>
- Onan, A., & Alhumyani, H. (2024). Knowledge-enhanced transformer graph summarization (KETGS): integrating entity and discourse relations for advanced extractive text summarization. *Mathematics*, 12(23), 3638, 1-25. <https://doi.org/10.3390/math12233638>
- Orchi, H., Sadik, M., Khaldoun, M., & Sabir, E. (2023). Real-time detection of crop leaf diseases using enhanced YOLOv8 algorithm. In *Proc. Int. Wireless Commun. Mobile Comput.* (pp. 1690-1696). IEEE. <https://doi.org/10.1109/IWCMC58020.2023.10182573>

- Oro, E., Granata, F. M., & Ruffolo, M. (2025). A Comprehensive Evaluation of Embedding Models and LLMs for IR and QA Across English and Italian. *Big Data Cogn. Comput.*, 9(5), 141, 1-41.
<https://doi.org/10.3390/bdcc9050141>
- Qu, J., Li, G., Li, Y., Zhang, X., Yan, X., & Jin, H. (2025). SPW-YOLOv8: a high-precision and efficient deep learning model for rice leaf disease detection. *Meas. Sci. Technol.*, 36(9), 095701, 1-10.
<https://doi.org/10.1088/1361-6501/ae01c4>
- Rahman, K. N., Banik, S. C., Islam, R., & Fahim, A. A. (2025). A real time monitoring system for accurate plant leaves disease detection using deep learning. *Crop Des.*, 4(1), 100092, 1-16.
<https://doi.org/10.1016/j.crope.2024.100092>
- Reya, S. S., Malek, M. D. A., & Debnath, A. (2022). Deep learning approaches for cabbage disease classification. *In Proc. Int. Conf. Recent Progr. Sci. Eng. Technol.* (pp. 1-5). IEEE.
<https://doi.org/10.1109/ICRPSET57982.2022.10188553>
- Rezaei, M., Diepeveen, D., Laga, H., & Sohel, F. (2026). Automatic pixel-level annotation for plant disease severity estimation. *Comput. Electron. Agric.*, 241(2026), 111316, 1-17.
<https://doi.org/10.1016/j.compag.2025.111316>
- Sambana, B., Nnadi, H. S., Wajid, M. A., Fidelia, N. O., Camacho-Zuñiga, C., Ajuzie, H. D., & Onyema, E. M. (2025). An efficient plant disease detection using transfer learning approach. *Sci. Rep.*, 15(1), 19082, 1-15. <https://doi.org/10.1038/s41598-025-02271-w>
- Sharma, R., Kukreja, V., & Bordoloi, D. (2023). Deep learning meets agriculture: a faster RCNN based approach to pepper leaf blight disease detection and multi-classification. *In Proc. 4th Int. Conf. Emerg. Technol.* (pp. 1-5). IEEE. <https://doi.org/10.1109/INCET57972.2023.10170692>
- Shen, Y., Yang, Z., Khan, Z., Liu, H., Chen, W., & Duan, S. (2025). Optimization of improved YOLOv8 for precision tomato leaf disease detection in sustainable agriculture. *Sensors*, 25(5), 1398, 1-23.
<https://doi.org/10.3390/s25051398>

- Sorour, S. E., & Almusallam, N. (2025). L3D-RAG: leveraging LLaMA 3.1 and DeepSeek with retrieval-augmented generation (RAG) for advanced reddit community analysis. *Alexandria Eng. J.*, 131(1), 125-152.
<https://doi.org/10.1016/j.aej.2025.09.070>
- Sun, H., Liu, T., Wang, J., Zhai, D., & Yu, J. (2024). Evaluation of two deep learning-based approaches for detecting weeds growing in cabbage. *Pest Manag. Sci.*, 80(6), 2817-2826.
<https://doi.org/10.1002/ps.7990>
- Syahidi, A. A., & Kiyokawa, K. (2025). Automatic text generation in Banjar language using GPT-4 for low-resource language preservation. *In Proc. 7th Symp. Comput. Informatics* (pp. 238-243). IEEE.
<https://doi.org/10.1109/ISCI65687.2025.11167521>
- Syahidi, A. A., Kiyokawa, K., & Nuchitprasitchai, S. (2025). A fine-tuned GPT-4-based question answering system for e-government services using a custom-built dataset. *In Proc. 7th Symp. Comput. Informatics* (pp. 232-237). IEEE. <https://doi.org/10.1109/ISCI65687.2025.11166787>
- Szechyńska-Hebda, M., Hołownicki, R., Doruchowski, G., Sas, K., Puławska, J., Jarecka-Boncela, A., Ptaszek, M., & Włodarek, A. (2025). Application of hyperspectral imaging for early detection of pathogen-induced stress in cabbage as case study. *Agronomy*, 15(7), 1516, 1-30.
<https://doi.org/10.3390/agronomy15071516>
- Tang, G., Yousuf, O., & Jin, Z. (2024). Improving BERTScore for Machine Translation Evaluation Through Contrastive Learning. *IEEE Access*, 12(1), 77739, 1-11. <https://doi.org/10.1109/ACCESS.2024.3406993>
- Vega-Álvarez, C., Francisco, M., & Soengas, P. (2021). Black rot disease decreases young brassica oleracea plants' biomass but has no effect in adult plants. *Agronomy*, 11(3), 569, 1-12.
<https://doi.org/10.3390/agronomy11030569>
- Wang, S., Ma, S., Wang, Y., Yang, X., Jiao, C., Yang, C., Wang, X., Wang, X., Xiang, W., & Zhao, J. (2026). Fungal diversity, pathogenicity, and fungicide sensitivity in leaf diseases of Chinese cabbage from Heilongjiang Province, China. *Physiol. Mol. Plant Pathol.*, 141(1), 102991, 1-14.
<https://doi.org/10.1016/j.pmpp.2025.102991>

Wang, X., & Song, J. (2021). ICIoU: improved loss based on complete intersection over union for bounding box regression. *IEEE Access*, 9(1), 105686, 1-10. <https://doi.org/10.1109/ACCESS.2021.3100414>

Zahid, I. A., Joudar, S. S., Albahri, A. S., Albahri, O. S., Alamoodi, A. H., Santamaria, J., & Alzubaidi, L. (2024). Unmasking large language models by means of OpenAI GPT-4 and Google AI: A deep instruction-based analysis. *Intell. Syst. Appl.*, 23(1), 200431, 1-18. <https://doi.org/10.1016/j.iswa.2024.200431>

Zhang, Y., Lin, S., Xiong, Y., Li, N., Zhong, L., Ding, L., & Hu, Q. (2025). Fine-tuning large language models for interdisciplinary environmental challenges. *Environ. Sci. Ecotechnol.*, 27(1), 100608, 1-11. <https://doi.org/10.1016/j.esa.2025.100608>