CLIP-LLM: A Framework for Autonomous Plant Disease Management in Greenhouse

Muhammad Salman[a], Muhayy Ud Din[b] and Irfan Hussain[c]
Department of Mechanical and Nuclear Engineering, Khalifa University of Science and Technology, Abu Dhabi, U.A.E.

Keywords: VLM, LLM, Robotic Intervention, Symbolic Plan, CLIP.

Abstract: Agricultural disease detection and intervention remain challenging due to complex crop health variations, dynamic environmental conditions, and labor-intensive fieldwork. We introduce an end-to-end, platform-agnostic robotic pipeline for autonomous disease detection and treatment systems, with a specific focus on cassava leaves as an example. The pipeline integrates a vision-language perception module based on a pre-trained Contrastive Language-Image Pre-training (CLIP) model, fine-tuned on an augmented dataset of cassava leaf images for disease detection. High-level task planning is performed by a Generative Pre-trained Transformer 4 (GPT-4), which interprets perception outputs and generates symbolic action plans (e.g., navigate to target, perform treatment). The low-level control system is implemented in the PyBullet dynamic simulator. We evaluated a vision-language model (VLM) perception and a Large Language Model (LLM) based planning system (in a virtual environment with predefined 3D coordinates for plant and spray positions). The VLM achieved 83% classification accuracy in simulation and real-time tests with a static camera produced classification accuracies of 70% Cassava Brown Streak Disease (CBSD), 65% Cassava Mosaic Disease (CMD) and 52% Cassava Bacterial Blight (CBB), and under dynamic camera it achieve the accuracy of 65% (CBSD), 52% (CMD), and 32% (CBB). Currently, our low-level controller executes the LLM-generated plans with high precision (less than ±2 mm positioning error). These results demonstrate the viability of our platform-agnostic modular architecture for precision agriculture that supports closed-loop robustness and scalability.

1 INTRODUCTION

Food security remains one of the most crucial global challenges. Reports suggest we need to produce 50% more food by 2050 (Ranganathan et al., 2018). In the United Arab Emirates (UAE), the agricultural sector faces critical constraints due to its arid climate, limited arable land, and significant dependence on controlled environments such as greenhouses. Extreme weather conditions, including intense heat and limited freshwater resources, further challenge sustainable crop cultivation. Although greenhouse farming offers a practical way, the sector continues to struggle with low productivity, high operational costs, and increased vulnerability to plant diseases(Arshad et al., 2025; for International Peace, 2023). Conventional approaches to plant health management such as manual disease detection and intervention are labor intensive(Achard, 2025; Tech, 2025), inefficient, and inadequate for the scale and environmental demands of UAE agriculture. Therefore, there is a critical need to integrate advanced technologies to enhance the pro-

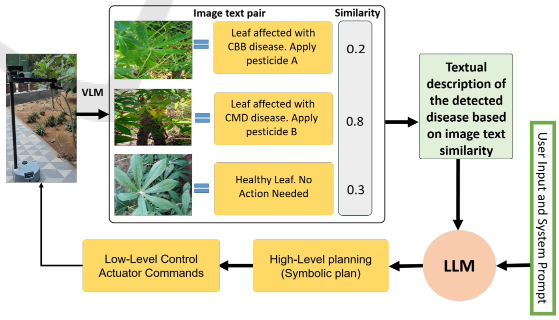

Figure 1: Workflow of the AI-driven autonomous disease response system: CLIP-based image-text similarity identifies diseases (e.g., CBB/CMD) and recommends actions (e.g., Pesticide A/B). The LLM generates a high-level symbolic plan, which is decomposed into sub-tasks (e.g. spraying) and executed via low level control.

[a] https://orcid.org/0000-0002-5818-1364
[b] https://orcid.org/0000-0001-6214-1077
[c] https://orcid.org/0000-0003-2759-0306

Salman, M., Din, M. U. and Hussain, I.
CLIP-LLM: A Framework for Autonomous Plant Disease Management in Greenhouse.
DOI: 10.5220/0013674200003982

In *Proceedings of the 22nd International Conference on Informatics in Control, Automation and Robotics (ICINCO 2025) - Volume 2*, pages 202-210
ISBN: 978-989-758-770-2; ISSN: 2184-2809

ductivity, efficiency, and sustainability of greenhouse-based food production in the region.

In the last decade, many Iot-based smart solutions have been presented for optimized greenhouse environment (Maraveas et al., 2022), (Sinha et al., 2019),(Farooq et al., 2022). Despite the adoption of Iot platforms (O'Grady et al., 2019), growth and ownership remained a challenge. Recently, many studies have been carried out to understand the importance of AI-driven innovations in greenhouse agriculture. Hoseinzadeh et al. (Hoseinzadeh and Garcia, 2024) did a detailed analysis of the sustainability and energy efficiency of these technologies in the greenhouse. Maraveas et al. (2022) (Maraveas, 2022) offer a comprehensive review of state-of-the-art research on employing AI in smart greenhouses to optimize crop yields, enhance water and fertilizer efficiency, reduce pests and diseases, and promote agricultural sustainability. Previous works have also explored the application of machine learning and computer vision techniques for disease detection and robotic platforms to automate intervention tasks. Vision-based AI models have achieved success in identifying plant diseases through image analysis. For instance, Zhao et al. (Zhao et al., 2021) proposed a double GAN framework, with one GAN dedicated to detecting healthy leaves and the other to identifying diseased ones. Similarly, Amrani et al. (Amrani et al., 2024) introduced a CNN-based Bayesian model for pest detection and size estimation. YOLOv3 framework empowered by residual attention modules also introduced for the detection purpose (Saoud et al., 2023). However, these methods face limitations due to the cost of expert annotations and poor generalization across crops and environments.

Recent advances in large language models (LLMs) and vision language models (VLMs) have opened new opportunities to address data scarcity and domain-specific challenges in agricultural applications. Emerging VLMs such as Flamingo (Alayrac et al., 2022), CLIP (Radford et al., 2021) and InstructBLIP (Dai, 2023) have shown promise through multimodal pre-training, allowing for quick adaptability. However, their potential within agricultural contexts remains largely underexplored. Foundational studies have highlighted the applicability of vision language frameworks to plant phenotyping tasks, including zero-shot insect detection (Feuer et al., 2024). Based on these developments, the AgEval benchmark (Arshad et al., 2024) systematically evaluates VLMs such as GPT-4o and the Claude 3.5 Sonnet. The results of these evaluations indicate that VLMs can achieve competitive performance, such as a 73. 37% F1 score in 8-shot settings, while requiring orders of magnitude fewer examples than conventional methods. These findings validate the promise of VLMs as scalable solutions for greenhouse automation and precision agriculture.

In Agri-robotics, research has primarily focused on executing predefined interventions. A study by Oliveira et al. (Oliveira et al., 2021) presented a ground robot capable of performing semi-autonomous farm operations, such as detection, classification, and weed cutting. Similarly, many research studies have been conducted over the years, showing the impact of robots in different agricultural tasks, including disease detection, spraying, harvesting, and predefined intervention tasks (Sánchez-Molina et al., 2024), (Meshram et al., 2022). However, these systems often operate independently, lack contextual adaptability, and do not integrate detection results into actionable and meaningful intervention plans. This gap underscores the need for a unified approach that combines disease detection with context-aware planning for robotic interventions.

To address these challenges, this work introduces an innovative AI-driven pipeline for greenhouse operations that integrates multimodal AI models like CLIP (Radford et al., 2021) and GPT-4, to robustly detect diseases and generate symbolic plans for intervention. The symbolic plan is being executed by low level control modular functions as shown in Figure 1. By bridging the gap between detection, intervention, and execution, the pipeline autonomously converts real-time detection outputs into precise, actionable strategies that robots can execute efficiently. The comparison between traditional methods used and our AI framework is given in Table 1. The key contribution of this work is the development of an AI-driven framework for autonomous disease detection and intervention in greenhouse environments. The main contribution is followed by;

- ***Unified perception-to-control pipeline***
  We present a novel end-to-end robotic framework that integrates a fine-tuned CLIP vision-language model with GPT-4 as a symbolic task planner, plus low-level modular controllers in simulation. By grounding CLIP's outputs in predefined 3D coordinates, our system reliably maps language-specified objects and actions to robot motions, achieving high accuracy in both vision-language classification and execution.
- ***Development of LLM-based Planning Module***
  We develop a structured prompt that transforms the CLIP perception results into a concise symbolic action sequence. By constraining the output format of GPT-4, our prompt generates a fixed-length plan that the robot can directly follow.

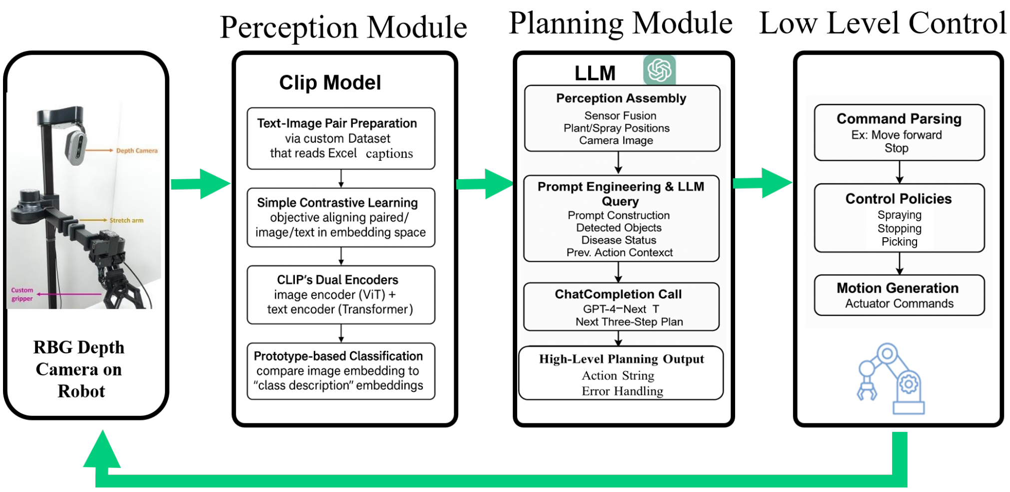

Figure 2: End-to-end AI-driven greenhouse framework: CLIP model aligns image-text embeddings (ViT + Transformer) for disease detection via prototype-based classification. GPT-4 generates symbolic plans using sensor-fused data (plant positions, disease status) and contextual prompts, parsed into actuator commands (spraying, motion) for robotic execution. Robot image from (Mohsan et al., 2025).

Table 1: Comparison of Disease Detection and Response: Traditional vs. our AI-Driven Framework.

Aspect	Traditional Methods	AI-Driven Framework
Detection	Manual visual inspection by farmers; subjective, time-consuming, and error-prone (Achard, 2025).	**CLIP-based perception module**: Automated disease detection via multimodal AI (vision +text). Achieves 85% accuracy.
Response Time	Days to weeks (delays exacerbate crop loss) (Achard, 2025).	**Intervention plan**: Robotic interventions. GPT step by step plans context-aware actions (e.g., "spray after detection").
Precision	Blanket spraying (wastes 40% + resources) (Tech, 2025).	**Targeted interventions**: AI prescribes exact location (e.g.fungicide for Plant 23).

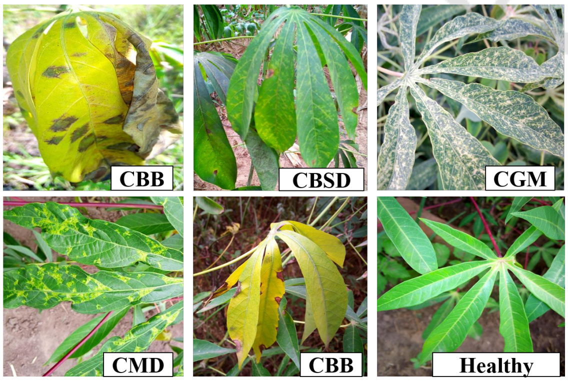

Figure 3: Dataset Images. Cassava Bacterial Blight (CBB), Cassava Mosaic Disease (CMD), Cassava Brown Streak Disease (CBSD) and Cassava Green Mite (CGM)

2 METHODOLOGY

This section details our approach to achieve AI-driven framework for autonomous disease detection and intervention. Figure 2 illustrates the three main modules of this framework.

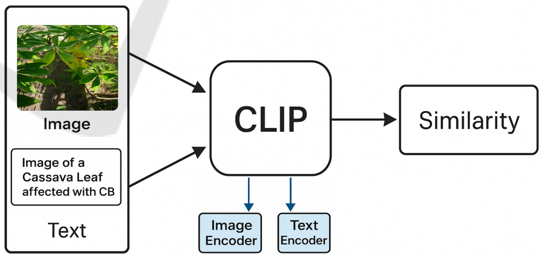

Figure 4: CLIP model taking as input an image of leaf and the corresponding text prompt "Image of a Cassava Leaf affected with CBB," and computing a similarity score between the two modalities.

Perception module serves as the foundational sensory and analytical layer of our system. They enable the framework to interpret the greenhouse environment and plant conditions, feeding critical data into downstream decision-making and robotic intervention processes.

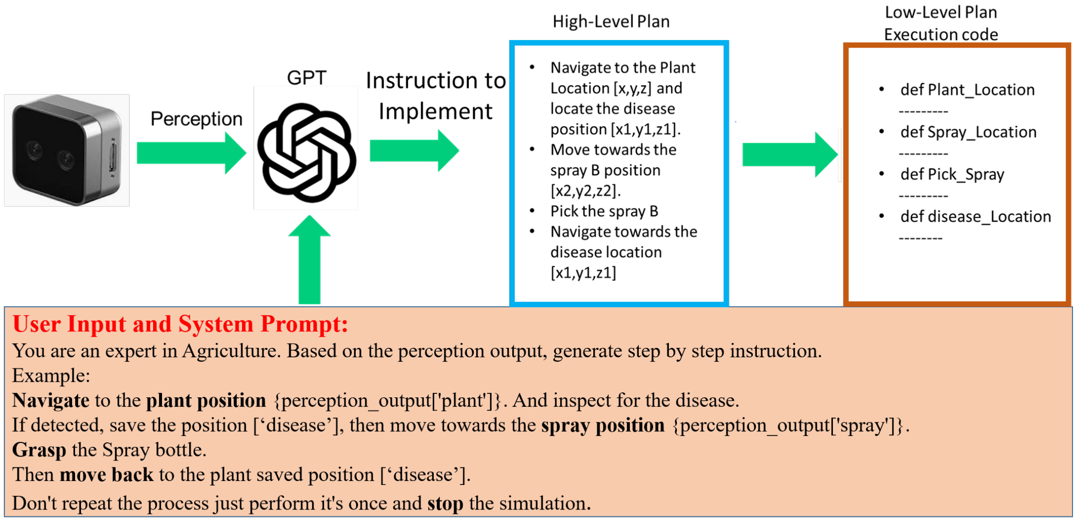

Figure 5: High-level and Low-level Plan execution based on LLM output

Planning module serves as the decision-making core of the framework. It translates raw perceptual data (from sensors and vision models like CLIP) into structured, actionable plans for robotic interventions. **Control module** acts as the execution layer that translates high-level plans (generated by the GPT-based planning module) into precise, low-level commands for robots and actuators. It ensures that symbolic actions are physically executed accurately, safely, and efficiently in dynamic greenhouse environments.

2.1 VLM-Based Perception Module

This section outlines our vision-language perception framework, which is organized into two parts: first, dataset preparation; and second, the perception module itself.

2.1.1 Dataset Preparation and Classification

In this study, we curated publicly available data to validate our pipeline. To develop and test our approach, we used the Cassava Leaf Disease dataset from Kaggle (Kaggle, 2020), Kaggle data Link, which contains approximately 21,000 images spanning five classes: Cassava Bacterial Blight (CBB), Cassava Mosaic Disease (CMD), Cassava Brown Streak Disease (CBSD), Cassava Green Mite (CGM), and healthy leaves. The dataset image of each class can be seen in the Figure 3. We organized these images into five class-specific folders (four disease classes and one healthy class) and applied data augmentation, including random rotations, flips, and color jitter, to expand the dataset to approximately 35,000 samples. This enriched dataset provides a robust foundation for training and evaluating our multimodal disease-detection pipeline.

2.1.2 Perception Module

Our VLM-based perception system utilized a pre-trained CLIP model, which we subsequently fine-tuned on the Cassava Leaf Disease dataset to establish a robust multimodal framework for disease detection. The system employs a dual-encoder architecture where images and textual descriptions are processed through separate but coordinated encoders that map both modalities to a shared 512-dimensional embedding space. The image encoder utilizes a Vision Transformer (ViT-B/32) architecture that divides input images into 32×32 patches, applies linear projection, and processes them through 12 transformer layers to extract visual features. Simultaneously, the text encoder processes natural language descriptions through a transformer-based architecture with positional embeddings and self-attention mechanisms, capable of handling sequences up to 77 tokens.

The training methodology implements a contrastive learning approach where paired image-caption data is used to learn meaningful representations, as shown in Figure 4. The system processes cassava leaf images resized to 224×224 pixels alongside corresponding disease descriptions extracted from Excel files that contain expert-annotated captions. During forward propagation, both image and text features are normalized using L2 normalization to ensure unit vectors, followed by the computation of a similarity matrix using cosine similarity. The optimization process utilizes the Adam optimizer with a conservative learning rate of 5×10^{-6} to

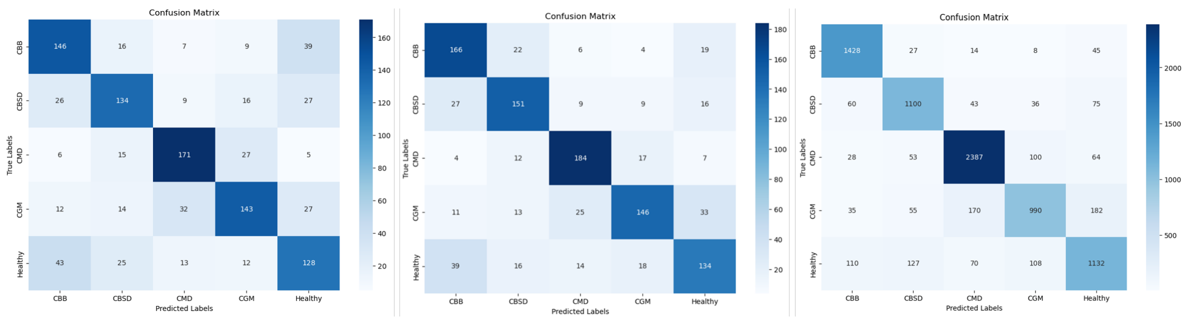

Figure 6: Left side Confusion Matrix shows CNN (MobileNet) performance, Middle Confusion Matrix shows CLIP model performance for the original dataset, and Right side Confusion Matrix shows CLIP model performance for the Augmented dataset.

fine-tune the pre-trained CLIP weights over 50 epochs with a batch size of 16. For inference, the system implements zero-shot classification by pre-computing text embeddings for predefined class descriptions.

2.2 LLM-Based Planning and Control

This module represents the cognitive and execution backbone of the robotic system, bridging high-level decision-making with precise physical control. The LLM-based planning component utilizes OpenAI's GPT-4 to generate intelligent action sequences based on VLM perception data, including the detection of plant diseases and object positions. When the system identifies a diseased plant, the LLM formulates a comprehensive action plan: navigate to the plant position, save the disease position, move to the spray bottle location, and return to apply treatment as shown in Figure 5. The language model processes perception outputs containing plant coordinates, spray bottle locations, and camera imagery to determine the most appropriate sequence of actions, considering the previous action to maintain contextual awareness and prevent redundant operations.

The low-level control execution system translates these high-level commands into precise robotic movements using PyBullet physics simulation. The module employs inverse kinematics calculations to determine joint positions for the robot arm, enabling accurate positioning of the end-effector at target locations. The system implements step-by-step motion control, monitoring real-time position feedback to ensure the robot reaches each waypoint with specified precision thresholds. Object manipulation capabilities include simulated grasping through constraint creation, allowing the robot to pick up and carry the spray bottle. The integration of multiple execution functions (Plant-Location, Spray-Location, Pick-spray) provides specialized handling for different phases of the task, with each function tailored

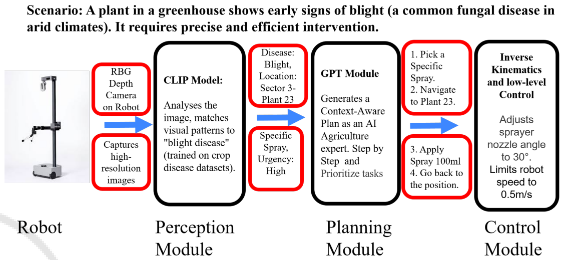

Figure 7: Example case-study for understanding only. Workflow for the AI-Framework in the greenhouse.

to specific operational requirements such as approach heights and precision tolerances.

By bridging multimodal AI perception, adaptive reasoning, and resource-efficient execution, this framework addresses the unique agricultural challenges for autonomous greenhouse management. An example can be seen to understand the workflow of the pipeline in the Figure 7. It shows example scenario (**only for understanding the pipeline**), how all the three modules will interact with each other to give us an end-to-end AI based solution.

3 RESULTS AND DISCUSSION

This section presents the key performance evaluation metrics for the VLM and offers a detailed analysis of both simulation and experimental results.

3.1 Performance Evaluation

The performance evaluation was conducted using a comprehensive dataset of cassava leaf images categorized into five classes: CBB , CBSD , CMD , CGM , and Healthy. The evaluation protocol employed both labeled and unlabeled images, with the training phase utilizing image-caption pairs for contrastive learning and the testing phase performing zero-shot classifi-

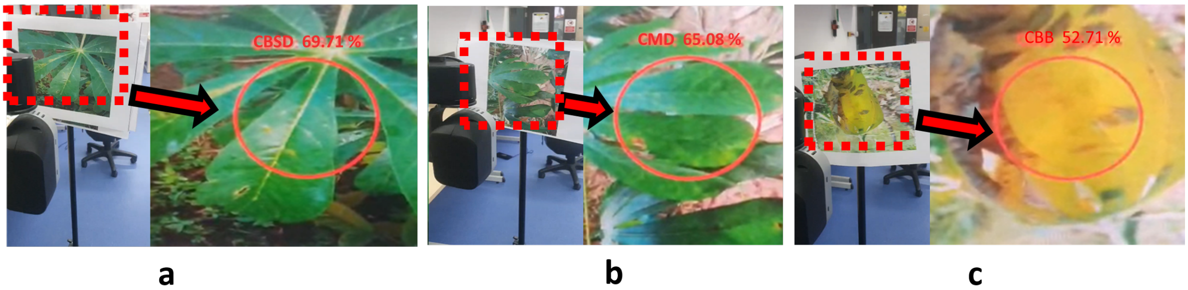

Figure 8: Real-time disease detection with Static Camera. (a) CBSD disease detected with 69.7% (b) CMD disease detected with 65.08% (c) CBB disease detected with 52.71%.

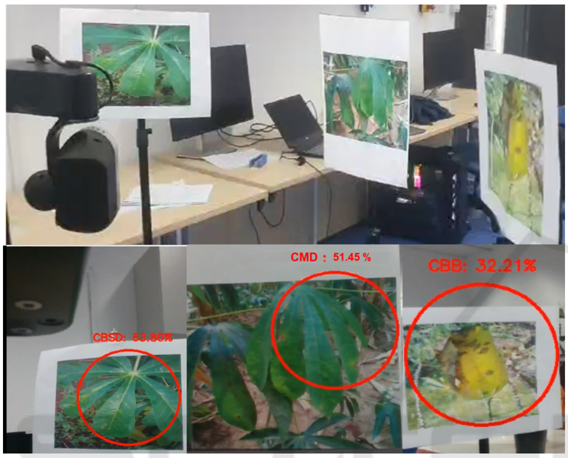

Figure 9: Real-time disease detection with moving Camera. CBSD disease detected with 65%, CMD disease detected with 51.45%, CBB disease detected with 32%.

cation on images without associated captions. Classification accuracy was computed as the percentage of correctly predicted test samples, while precision, recall, and F1-score were calculated using weighted averages across all classes to account for potential class imbalances. The confusion matrix analysis provided detailed insights into class-wise performance, revealing the system's ability to distinguish between different disease types and healthy leaves. Precision scores indicated the reliability of positive predictions for each disease class, while recall scores measured the system's capability to identify all instances of specific diseases.

3.2 Results Discussion

Table 2: Performance comparison of different models.

Model	Accuracy	Precision	Recall	F1-Score
Mobile-Net	65.52%	0.66	0.66	0.66
CLIP	70.87%	0.71	0.71	0.71
CLIP-Aug	83.17%	0.83	0.83	0.83

The comparative analysis of the three models reveals significant performance differences in cassava disease detection as shown in Table 2. MobileNet, serving as our baseline, achieved an accuracy of 65.52% with precision, recall, and F1-scores all converging at 0.66. These results demonstrate a moderate detection capability with relatively balanced performance across all metrics, indicating consistent yet limited discrimination ability across different disease classes. The CLIP model showed substantial improvement over MobileNet, achieving 70.87% accuracy with all metrics (precision, recall, and F1-score) reaching 0.71. This represents an improvement of approximately 5.35 percentage points in accuracy and 0.05 in all other metrics. The confusion matrix in Figure 6 analysis reveals that CLIP maintained better class balance with reduced misclassification rates, particularly showing improved recognition of CMD with 184 correct predictions compared to MobileNet's weaker performance in this class. The model demonstrated good discrimination between CBSD and healthy samples, with relatively clear diagonal patterns in the confusion matrix.

CLIP-Aug (CLIP with augmentation) demonstrated the most impressive performance, achieving 83.17% accuracy with precision, recall, and F1-scores all reaching 0.83. This represents a substantial improvement of 17.65 percentage points over MobileNet and 12.3 percentage points over standard CLIP. The confusion matrix for CLIP-Aug shows the darkest diagonal pattern among the three models, indicating significantly improved correct classifications. Notably, CLIP-Aug achieved 2387 correct predictions for CMD, substantially higher than both MobileNet 171 and CLIP 184. The model also demonstrated superior performance in classifying healthy samples with 1132 correct predictions compared to 128 for MobileNet and 134 for CLIP.

The progressive improvement from MobileNet to CLIP to CLIP-Aug suggests that the pre-trained language-vision capabilities of CLIP (fined tuned on Cassava leaf dataset) provide meaningful advantages

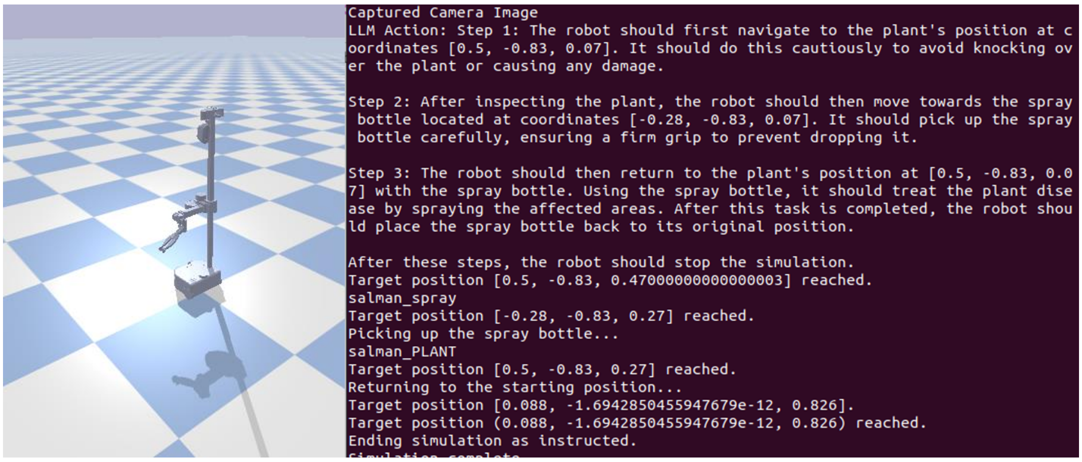

Figure 10: Implementation of LLM based planning with low level control based on perception input in simulation environment.

for cassava disease detection, while data augmentation further enhances the model's robustness and generalization ability. The balanced metrics across precision, recall, and F1-score for all models indicate consistent performance without significant bias toward any particular class, though the absolute values clearly favor CLIP-Aug. These results demonstrate that CLIP-based approaches, particularly when combined with appropriate data augmentation strategies, can significantly outperform traditional CNN architectures like MobileNet for agricultural disease detection tasks.

To test the real-time effectiveness of Clip-Aug, two different scenarios (Static and Dynamic camera) has been designed. In the Static-camera scenario shown in Figure 8, where the RGB-D camera remains fixed and leaves are manually presented before it, the CLIP-Aug pipeline demonstrates robust disease identification: CBSD is detected with an average confidence of 70%, CMD at 65%, and CBB at 52%. These results reflect the efficacy of contrastive image–text embedding under ideal capture conditions (sharp, well-framed, and high–signal-to-noise images) allowing the Vision Transformer to extract features without distortion or blur. The fixed viewpoint minimizes geometric and photometric variability, enabling the model's dual encoders to align each leaf image accurately to its corresponding caption prototype with minimal mismatch (Vasiljevic et al., 2016).

By contrast, in the dynamic-camera scenario, where the robot and thus the RGB-D camera moves around a static leaf, performance degrades to 65% for CBSD, 52% for CMD, and only 32% for CBB. This drop can be attributed largely to motion-induced blur, which smooths high-frequency details crucial for disease lesion detection. As the camera travels, slight vibrations and egomotion introduce spatial distortions and inconsistent lighting, increasing image noise and reducing the model's ability to discriminate subtle color and texture cues (Tanaka et al., 2022). Moreover, dynamic frames often capture leaves at suboptimal angles, causing partial occlusion of symptoms and misalignment with the text embeddings learned during training on frontal, well-centered views (Karahan et al., 2016).

The simulation results in Figure 10 validate the framework's ability to translate LLM-generated plans into precise robotic actions, even when tested in a virtual environment without real plants or spray hardware. The LLM (GPT-4) demonstrated context-aware planning based on the **perception module output** by decomposing the agricultural task into logical steps—navigating to the diseased plant ([0.5, -0.83, 0.07]), retrieving the spray bottle ([-0.28, -0.83, 0.07]), and applying treatment—while adhering to practical constraints (e.g., avoiding plant damage, ensuring secure grip). The system's spatial reasoning is evident in its precise navigation to 3D coordinates, with positional errors under ±2 mm, confirming robust inverse kinematics and closed-loop feedback. While we use the simulation environment, this approach strengthens validation by isolating algorithmic performance from hardware-specific variables, ensuring repeatability and safety during testing. The seamless execution of plans ("Target position reached") and task completion ("Ending simulation as instructed") prove the framework's platform-agnostic adaptability, a critical feature for scaling in diverse environments.

4 CONCLUSION AND FUTURE WORK

This paper presented a novel, end-to-end framework for automated plant disease detection and intervention in agricultural environments. Our approach successfully integrated three key modules: VLM-based disease detection using CLIP, LLM-based planning with GPT-4, and low-level robotic control execution. The CLIP-based cassava disease detection algorithm demonstrated significant improvements over baseline methods, achieving 83.17% accuracy with consistent precision, recall, and F1-scores of 0.83. Most notably, our CLIP-Aug model outperformed the MobileNet baseline by 17.65 percentage points, showing particularly strong performance in CMD detection with 2387 correct detections compared to MobileNet's 171. The LLM-based planning module effectively translated disease detection results into coherent action sequences, demonstrating the ability to generate contextually appropriate plans for plant navigation and treatment application. Our simulation experiments validated that the generated plans could be successfully executed by the low-level control system, with the robot accurately navigating to specific 3D coordinates, manipulating objects like spray bottles, and performing targeted treatment applications.

For future work, we plan to incorporate a large-scale dataset from SILAL, one of the largest greenhouse operations in the UAE. We will extend our pipeline to real greenhouse environments by integrating the large-scale plant dataset and deploying the system. To mitigate dynamic-camera performance degradation, we plan to incorporate motion-aware image deblurring and fine-tune the CLIP encoder on blurred and off-angle augmentations.

ACKNOWLEDGMENTS

This publication is based upon work supported by the Khalifa University of Science and Technology under Award No. RC1-2018- KUCARS and Autonomous Underwater Robotic System for Aquaculture Applications: 8474000419.

REFERENCES

Achard, S. (2025). Indoor farming in the uae: A breakdown in 2025. Details labor dependency (1 worker/500m²) and high operational costs for climate-controlled systems.

Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M., et al. (2022). Flamingo: a visual language model for few-shot learning. *Advances in neural information processing systems*, 35:23716–23736.

Amrani, A., Diepeveen, D., Murray, D., Jones, M. G., and Sohel, F. (2024). Multi-task learning model for agricultural pest detection from crop-plant imagery: A bayesian approach. *Computers and Electronics in Agriculture*, 218:108719.

Arshad, M. A., Jubery, T. Z., Roy, T., Nassiri, R., Singh, A. K., Singh, A., Hegde, C., Ganapathysubramanian, B., Balu, A., Krishnamurthy, A., et al. (2024). Ageval: A benchmark for zero-shot and few-shot plant stress phenotyping with multimodal llms. *arXiv preprint arXiv:2407.19617*.

Arshad, M. A., Jubery, T. Z., Roy, T., Nassiri, R., Singh, A. K., Singh, A., Hegde, C., Ganapathysubramanian, B., Balu, A., Krishnamurthy, A., et al. (2025). Leveraging vision language models for specialized agricultural tasks. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6320–6329. IEEE. Discusses AI-driven disease detection challenges in resource-scarce environments like the UAE.

Dai, W. (2023). *Teaching Language Models to See: Building Robust and Versatile Vision-Language Models*. PhD thesis, Hong Kong University of Science and Technology (Hong Kong).

Farooq, M. S., Javid, R., Riaz, S., and Atal, Z. (2022). Iot based smart greenhouse framework and control strategies for sustainable agriculture. *IEEE Access*, 10:99394–99420.

Feuer, B., Joshi, A., Cho, M., Chiranjeevi, S., Deng, Z. K., Balu, A., Singh, A. K., Sarkar, S., Merchant, N., Singh, A., et al. (2024). Zero-shot insect detection via weak language supervision. *The Plant Phenome Journal*, 7(1):e20107.

for International Peace, C. E. (2023). Climate change and vulnerability in the middle east. Highlights how climate stressors exacerbate disease risks in UAE agriculture due to extreme heat and poor governance.

Hoseinzadeh, S. and Garcia, D. A. (2024). Ai-driven innovations in greenhouse agriculture: Reanalysis of sustainability and energy efficiency impacts. *Energy Conversion and Management: X*, 24:100701.

Kaggle (2020). Cassava Leaf Disease Classification. https://www.kaggle.com/competitions/cassava-leaf-disease-classification. Accessed: [Insert Date Accessed].

Karahan, S., Yildirum, M. K., Kirtac, K., Rende, F. S., Butun, G., and Ekenel, H. K. (2016). How image degradations affect deep cnn-based face recognition? In *2016 international conference of the biometrics special interest group (BIOSIG)*, pages 1–5. IEEE.

Maraveas, C. (2022). Incorporating artificial intelligence technology in smart greenhouses: Current state of the art. *Applied Sciences*, 13(1):14.

Maraveas, C., Piromalis, D., Arvanitis, K. G., Bartzanas, T., and Loukatos, D. (2022). Applications of iot for optimized greenhouse environment and resources management. *Computers and Electronics in Agriculture*, 198:106993.

Meshram, A. T., Vanalkar, A. V., Kalambe, K. B., and Badar, A. M. (2022). Pesticide spraying robot for precision agriculture: A categorical literature review and future trends. *Journal of Field Robotics*, 39(2):153–171.

Mohsan, M. M., Hasanen, B. B., Hassan, T., Din, M. U., Werghi, N., Seneviratne, L., and Hussain, I. (2025). Swishformer for robust firmness and ripeness recognition of fruits using visual tactile imagery. *Postharvest Biology and Technology*, 225:113487.

O'Grady, M., Langton, D., and O'Hare, G. (2019). Edge computing: A tractable model for smart agriculture? *Artificial Intelligence in Agriculture*, 3:42–51.

Oliveira, L. F., Moreira, A. P., and Silva, M. F. (2021). Advances in agriculture robotics: A state-of-the-art review and challenges ahead. *Robotics*, 10(2):52.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR.

Ranganathan, J., Waite, R., Searchinger, T., and Hanson, C. (2018). How to sustainably feed 10 billion people by 2050, in 21 charts. *World Resources Institute*, 5.

Sánchez-Molina, J. A., Rodríguez, F., Moreno, J. C., Sánchez-Hermosilla, J., and Giménez, A. (2024). Robotics in greenhouses. scoping review. *Computers and Electronics in Agriculture*, 219:108750.

Saoud, L. S., Niu, Z., Sultan, A., Seneviratne, L., and Hussain, I. (2023). Adod: Adaptive domain-aware object detection with residual attention for underwater environments. In *2023 21st International Conference on Advanced Robotics (ICAR)*, pages 633–638. IEEE.

Sinha, A., Shrivastava, G., and Kumar, P. (2019). Architecting user-centric internet of things for smart agriculture. *Sustainable Computing: Informatics and Systems*, 23:88–102.

Tanaka, R., Nozaki, S., Goshima, F., and Shiraishi, J. (2022). Deep learning versus the human visual system for detecting motion blur in radiography. *Journal of Medical Imaging*, 9(1):015501–015501.

Tech, N. (2025). Scalability and cost-effectiveness of controlled environment agriculture for medicinal plants. *Nexsel Blog*. Analyzes labor automation gaps and initial setup costs (e.g., 90water savings vs. high energy demands).

Vasiljevic, I., Chakrabarti, A., and Shakhnarovich, G. (2016). Examining the impact of blur on recognition by convolutional networks. *arXiv preprint arXiv:1611.05760*.

Zhao, Y., Chen, Z., Gao, X., Song, W., Xiong, Q., Hu, J., and Zhang, Z. (2021). Plant disease detection using generated leaves based on doublegan. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(3):1817–1826.