Integrating Machine Learning and RAG-Based Chatbot for Mandarin Orange Disease Detection in Hilly Region of Nepal

Ritu Raj Lamsal
lamsal.raj@gmail.com

Medhayantra https://orcid.org/0000-0002-3751-2785

Prahlad Acharya
Dept of Electronics and Computer Engineering, IoE Thapathali Campus, Tribhuvan University

Rijan Pokhrel
Dept of Electronics and Computer Engineering, IoE Thapathali Campus, Tribhuvan University

Sakar Dahal
Dept of Electronics and Computer Engineering, IoE Thapathali Campus, Tribhuvan University

Vishal Sigdel
Dept of Electronics and Computer Engineering, IoE Thapathali Campus, Tribhuvan University

Research Article

Keywords: Citrus Disease Detection, Deep Learning, One-Class SVM, RAG-based Chatbot

Posted Date: February 27th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6105402/v1

Additional Declarations: The authors declare no competing interests.

Abstract

Citrus farming, particularly Mandarin orange cultivation, is a crucial economic activity in Nepal's hilly regions. However, disease detection and management remain major challenges. This study presents an efficient method for identifying and controlling five key citrus diseases affecting the Nepali orange market: black spot, canker, Huanglongbing (HLB), leaf miner, and sooty mold. We employ the MobileNetV2 model for disease prediction and a One-Class SVM model for initial leaf classification. Additionally, we integrate a Llama-3.2-11b-vision RAG-based chatbot, which analyzes leaf images and provides real-time guidance on disease prevention and orchard management. A mobile application has been developed to integrate the chatbot with a user-friendly interface, making it accessible for farmers. Our approach achieves 95.6% accuracy in disease identification and 85.6% accuracy in orange leaf classification. With its intuitive mobile platform and AI-driven chatbot, this system has the potential to transform citrus farming in Nepal by enabling timely interventions and improved disease management.

1. Introduction

Citrus fruits boast a long history reaching back thousands of years to Southeast Asia, especially in the areas of present-day India, Myanmar, and China. These regions, housing wild citrus varieties, were the origin of citrus development, leading to essential species such as citron (Citrus medica), pomelo (Citrus maxima), and mandarin (Citrus reticulata). Throughout time, the natural hybridization of these species established the genetic foundation for most contemporary citrus varieties. Citrus fruits continue to be essential to worldwide agriculture, commerce, and culture. Citrus fruit, mainly mandarin orange cultivation in Nepal's hilly regions, has been a source of economic prosperity and sustainability for many farmers [1]. Nonetheless, diseases such as Citrus Canker, Black Spot, Leaf Miner, Sooty Mold and HLB present difficulties for citrus agriculture, prompting investigations into disease-resistant cultivars and sustainable methods.

2. Related Works

Over the years, technological advancements in orange farming have incorporated AI techniques such as neural networks and deep learning. Li et al. used CNN activation maps for plant leaf disease detection and analyzed recent progress in deep learning for crop disease identification [2]. Qi et al. employed YOLOv5l models to detect citrus HLB, with the Yolov5l-HLB2 model achieving the highest F1 score of 85.19%, making it a valuable tool for preliminary screening before PCR testing [3]. Kaur et al. applied genetic algorithms for feature extraction, achieving 90.40% accuracy in HLB detection using a GA-SVM model [4]. Thomas et al. demonstrated the effectiveness of CNN-generated features with machine learning classifiers, achieving 92.1% accuracy for CBS-infected fruits and 93% for canker-infected leaves using RBFSVM [5]. Jouini et al. proposed CropNet, a hybrid model using RGB imaging and transfer learning, achieving 99.80% accuracy in wheat disease detection [6]. Deng et al. utilized visible spectrum imaging and C-SVC, achieving 91.93% accuracy in low-cost HLB recognition [7]. A multi-model fusion network combining RGB and hyperspectral imaging improved HLB classification, reaching 97.89%

accuracy, outperforming single-modal networks [8]. Dubey and Jalal proposed an adaptive approach using K-Means clustering and SVM, achieving 93% accuracy for apple disease classification [9]. Geetaramani and Pandain developed a Deep CNN model trained on 39 plant leaf classes, achieving 96.46% accuracy [10]. Joeshepin et al. introduced a CNN-LSTM model for citrus disease detection, achieving 96% accuracy, surpassing KNN and SVM methods [11]. Dhiman et al. integrated deep learning with edge computing, optimizing CNN-LSTM models for citrus disease classification, achieving 98.25% accuracy with reduced computational requirements [12]. Sindhu and Indirani incorporated IoT and cloud computing with deep learning, where IPSO-CNN achieved 96.08% accuracy and an AUC of 0.98 [13]. AI and IoT integration in agriculture is a growing global trend, and Nepal is no exception. These technologies are transforming traditional farming practices by enabling real-time monitoring, precision farming, and data-driven decision-making. In Nepal, the adoption of AI and IoT is gradually increasing, offering solutions for challenges such as climate variability, pest control, and resource optimization, ultimately improving productivity and sustainability[14].

This Paper aims to leverage deep learning for efficient citrus disease detection through image analysis, enabling early intervention and management. Additionally, a chatbot will serve as a virtual assistant to guide farmers on disease prevention and control.

3. Methodology

The system architecture, as illustrated in Fig. 1, comprises five primary components. The mobile phone hosts the application through which users provide input. The server, which serves as the backbone of the system, includes essential preprocessing modules and prediction models for disease detection. These models are fine-tuned and trained using datasets of both healthy and diseased leaves. Once a prediction is made, the response is forwarded to the chatbot for additional context before displaying the final result to the user.

The dataset used in this project for training various models was sourced from publicly available repositories and supplemented with locally collected images from an orange orchard from different parts of Nepal namely Sindhuli, Syangja and Dhulikhel. One of the primary datasets, titled *"A Citrus Fruits and Leaves Dataset for Detection and Classification of Citrus Diseases through Machine Learning,"* is accessible on the Data Mendeley website [15]. This dataset includes a wide collection of citrus fruit and leaf images labeled with diseases such as black spot, canker, greening, and scab, along with healthy samples. The dataset serves as a fundamental resource for developing and evaluating machine learning models for citrus disease detection.

For model training, the dataset images were first resized to a resolution of 224 × 224 pixels. A batch size of 32 was used, with the training dataset shuffled to enhance model generalization. The directory names were set to correspond with the class labels of the model. MobileNetV2, a lightweight deep learning model, was chosen as the base model, and its layers were frozen to retain pre-trained weights and reduce computational costs while minimizing the risk of overfitting. The training process involved adding a global average pooling layer, dropout layer, and activation layer to the base model. The hyperparameters were kept consistent with those of the base model, and the Adam optimizer was used for optimization. Binary cross-entropy was used as the loss function, while accuracy was the primary metric for model evaluation. The model was trained for 25 epochs, with an average training time of 95 seconds per epoch. The final validation accuracy achieved was 97.44%, while the test accuracy was 96.98%. Leaf distinction was carried out to ensure that the uploaded image belonged to an orange leaf before disease detection. The image was saved to the server and its file path cached for reuse. The image was then loaded, resized to 224 × 224 pixels, normalized, and converted into a flattened format. A ResNet50 model was used to extract feature vectors from the image, which were then passed to a pre-trained One-Class SVM model. If the output was 1, it confirmed that the image belonged to an orange leaf; otherwise, it was rejected. In cases where the image did not belong to an orange leaf, a JSON response was sent back to the client, whereas valid images were redirected for further disease identification. For disease detection, the preprocessed image was loaded from the cached file path and analyzed using the trained MobileNetV2-based model. The prediction results were generated in array format, converted into JSON format with disease names and their corresponding confidence percentages, and then sent back to the client. The disease detection model followed the same training strategy as mentioned earlier, ensuring high accuracy and reliability.

The chatbot was developed using a fine-tuned pretrained model through a process called Retrieval Augmented Generation (RAG). The dataset for chatbot training was curated from expert-written booklets, online resources, and specialist knowledge on citrus diseases. The data was stored in PDF format and processed using an UnstructuredPDFLoader. The text was segmented into chunks of 1,000 characters, which were then converted into numerical embeddings using the *sentence-transformers/all-MiniLM-L6-v2* model. The chatbot, powered by the *llama-3.2-1b-preview* model, retrieved relevant information by comparing input queries with stored embeddings and generating summarized responses based on the closest match. The mobile application was designed and developed as an intuitive platform for citrus farmers. Using advanced technology, the app enables users to detect plant diseases by analyzing leaf images. Additionally, the integrated chatbot assists farmers in addressing disease-related concerns, offering expert advice and detailed insights. The app aims to serve as a comprehensive tool for citrus farming and disease management, ultimately enhancing productivity and decision-making for farmers.

4. Results and discussion

4.1. Orange Leaf Distinction and Disease Detection

The One-Class SVM model is first used to differentiate between orange and non-orange leaves before disease classification. The confusion matrix in Fig. 2 indicates an accuracy of 83.65%, with 1,033 true positives and 173 false negatives. The feature space plot in Fig. 3 visually represents how the model separates orange leaves (green points) from non-orange leaves (red points), ensuring accurate classification. The model achieves 92.55% precision, demonstrating reliable classification, though some orange leaves were misclassified. Notably, papaya and guava leaves were correctly placed in the non-orange category, further validating the model's effectiveness.

For disease classification, three pretrained models were evaluated: Inception-v3 from InceptionNet, MobileNetV2 from MobileNet, and EfficientNet-B0 from EfficientNet. Each model was trained separately for over 20 epochs to analyze their individual performance. The training accuracy achieved was 94.05% for Inception-v3, 95% for EfficientNet-B0, and 95.6% for MobileNetV2. While all three models achieved comparable accuracy, MobileNetV2 demonstrated slightly better performance. However, both EfficientNet-B0 and Inception-v3 exhibited signs of overfitting Fig. 4 and Fig. 5, whereas MobileNetV2 maintained a smoother accuracy curve Fig. 9. Additionally, during deployment and testing, MobileNetV2 proved to be faster due to its lightweight architecture and efficiency. As a result, MobileNetV2 was preferred over the other pretrained models.

The ROC curve Fig. 6 shows near-perfect classification across six disease classes—Black Spot, Canker, HLB, Healthy, Leaf Miner, and Sooty Mold—with AUC values of 1.00, confirming strong predictive capability. However, the confusion matrix Fig. 7 reveals occasional misclassifications, such as healthy leaves being labeled as HLB or Black Spot being confused with Canker. The accuracy and loss plots demonstrate steady improvements over 20 epochs, with minimal overfitting. Additionally, feature maps from layer block_15_depthwise_BN. Figure 8 reveal important patterns that contribute to classification.

The classification report Table 1 underscores the model's reliability, with an overall accuracy of 97%. Leaf Miner and Sooty Mold achieve perfect precision and recall, while other disease classes maintain high F1-scores. However, slight misclassifications in HLB and Black Spot suggest that enhancing the dataset could further improve accuracy. In conclusion, combining One-Class SVM for orange leaf detection with MobileNetV2 for disease classification provides a robust and efficient citrus disease identification system. Future improvements, such as expanding the training dataset and fine-tuning the model, could further optimize performance.

As seen by the increased accuracy and decreased loss on both the training and validation sets, these trends show that the model is successfully identifying the underlying patterns in the data. The model appears to be generalizing effectively and not overfitting to the training data, as indicated by the convergence of the training and validation metrics.

Table 1
Classification Report of MobileNetV2

Disease	Precision	Recall	F1-Score	Support
Black Spot	0.91	0.95	0.93	200
Canker	0.98	0.99	0.99	200
HLB	0.95	0.91	0.93	200
Healthy	0.99	0.97	0.98	200
Leaf Miner	1.0	1.0	1.0	200
Sooty Mould	1.0	1.0	1.0	200
Accuracy			0.97	1200
Macro Avg	0.97	0.97	0.97	1200
Weighted Avg	0.97	0.97	0.97	1200

The six classes of orange leaves are clearly distinguished from one another as seen in the Fig. 10. The model has successfully distinguished between the different forms of leaf damage or disease, as shown by the clear clustering of the data points for each class. The model appears to have captured the unique visual features of each condition, as evidenced by the close clustering within classes and the good separation between them. The model's excellent ability to distinguish and identify the many forms of orange leaf damage or diseases found in the dataset is demonstrated by this visualization.

4.2. Chatbot Model

Several chatbot models were tested for integration into the mobile application. Initially, Botpress[17] was tried, but while it had a semantic accuracy of 0.7, its responses were slow and often inaccurate. Next, we tested fine-tuned versions of Alpaca and Gemma2 [18, 19], along with a RAG model using LLaMA-3.2[20]. These models were trained on the same dataset and evaluated using a semantic similarity test, which measures how close the chatbot's response is to the expected answer. The results as seen in Fig. 11 showed that the RAG model outperformed the fine-tuned models in semantic similarity. Additionally, the fine-tuned models were much larger in size and had slower response times. Because of this, the RAG model using LLaMA-3.2-11B-Vision was chosen for the project.

The RAG model was trained on a dataset containing information about citrus diseases in oranges and their prevention, provided in PDF format. The model was tested with 10 random questions, and various evaluation metrics were calculated.

The evaluation of the model revealed that it performed well when responding to direct questions related to citrus diseases. However, for indirect questions, such as "What are its symptoms?", the accuracy was slightly lower, as the model faced challenges in accumulating context. The lowest similarity score was observed for a completely unrelated question, indicating that while the chatbot is highly efficient in handling citrus-related queries, it may struggle with irrelevant ones. Overall, the results confirm that the RAG model is the best choice for this project due to its high accuracy, efficiency, and superior response speed compared to other tested models.

Table 2
Evaluation metrics of chatbot

Metrics	Value
Average Precision	0.95
Average Recall	0.95
Average semantic similarity	0.85

Moreover, the bar graph obtained from semantic similarity score is shown below. For the direct question which are related to citrus diseases, the results are very accurate. For indirect questions such as “What are its symptoms?”, the results are nearly accurate because the model starts to hibernate as the context starts to pile up. The query with lowest similarity score was totally unrelated to the diseases so score is quite low. This shows how efficient the RAG chatbot is for queries related to the citrus diseases.

4.3. Mobile Application UI

5. Future Enhancements

To further enhance the performance and robustness of the citrus disease identification system, several improvements can be explored. First, expanding the training dataset with more diverse and high-quality images, particularly for disease classes with slight misclassifications such as HLB and Black Spot, can improve classification accuracy. Additionally, implementing data augmentation techniques, such as rotation, scaling, and color variations, could help the model generalize better to real-world conditions. Fine-tuning the MobileNetV2 model with transfer learning from larger agricultural datasets may also boost accuracy while maintaining efficiency. Exploring advanced deep learning architectures, such as Vision Transformers (ViTs) or EfficientNet, could provide better feature extraction and improve classification performance. Furthermore, integrating multi-spectral or hyperspectral imaging data could enhance the model's ability to distinguish between similar disease symptoms.On the deployment side, optimizing the model for edge computing and mobile devices by reducing latency and memory footprint will be essential for real-time disease detection in the field. Implementing federated learning approaches could enable continuous model improvement using user-collected data without compromising privacy. Finally, incorporating a real-time decision support system that provides treatment recommendations based on disease severity can enhance the practical impact of the solution. Future work could also involve developing an IoT-based automated monitoring system[1, 15] that integrates disease classification with environmental sensor data to provide predictive analytics for farmers.

6. Conclusion

Citrus farming, particularly Mandarin orange cultivation, plays a vital role in Nepal's hilly regions, but disease detection and management remain persistent challenges. This study presents a comprehensive approach to addressing these issues by integrating machine learning and AI-driven solutions. The One-Class SVM model effectively differentiates between orange and non-orange leaves with 85.6% accuracy, ensuring precise leaf classification before disease identification. The MobileNetV2 model further enhances disease detection, achieving 95.6% accuracy in identifying five major citrus diseases—black spot, canker, Huanglongbing (HLB), leaf miner, and sooty mold. Additionally, the integration of a Llama-3.2-11b-vision RAG-based chatbot provides real-time guidance to farmers, offering disease prevention strategies and orchard management support through a user-friendly mobile application. The combination of these technologies results in a robust and scalable citrus disease identification system with a 97% overall accuracy. The near-perfect AUC values in disease classification confirm the model's predictive strength, while the mobile deployment ensures accessibility for farmers. Although the system demonstrates high efficiency, further improvements, such as expanding the training dataset, incorporating advanced deep learning architectures, and optimizing the model for real-time field applications, could enhance performance. Overall, this research marks a significant step toward intelligent, AI-driven citrus disease management. By enabling timely interventions and improved decision-making, the proposed system has the potential to transform citrus farming in Nepal, ensuring better yield quality and economic sustainability for smallholder farmers.

Declarations

Funding: No external funding

Authors Contribution: RRL conceptualization, methodology, research, data curation, review, PA,VS,SD,RP contributed to methodology, formal analysis, investigation, review, and editing, and data curation. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest: Authors declare no conflict of interest.

References

1. R. R. Lamsal, U. K. Acharya, P. Karthikeyan, P. Otero and A. Ariza (2024). Implementing Internet of Things for Real-Time Monitoring and Regulation of Off-Season Grafting and Post-Harvest Storage in Citrus Cultivation: A Case Study from the Hilly Regions of Nepal. AgriEngineering. 3. 2082-2100.. 10.3390/agriengineering6030122.
2. L. Li, S. Shang and B. Wang, "Plant Disease Detection and Classification by Deep Learning—A Review," IEEE Access, vol. 9, pp. 56683-56698, 2021.
3. R.-Z. Qiu, S.-P. Chen, T. Huang, G.-C. Fan, M.-X. Chi, R.-B. Wang, J. Zhao and Q. Y. Weng, "An automatic identification system for citrus greening disease (Huanglongbing) using a YOLO convolutional neural network," Front. Plant Sci, vol. 13, 20 December 2022.

4. B. Kaur, T. Sharma, B. Goyal and A. Dogra, "A Genetic Algorithm based Feature Optimization method for Citrus HLB Disease Detection using Machine Learning," IEEE Xplore, pp. 1052-1057, 2020.
5. “Automated classification of citrus disease on fruits and leaves using convolutional neural network generated features from hyperspectral images and machine learning classifiers” Journal of Applied Remote Sensing, Vol. 18, Issue 1, 014512 (February 2024). https://doi.org/10.1117/1.JRS.18.014512
6. Jouini, O., Aoueileyine, M. O.-E., Sethom, K., & Yazidi, A. (2024). Wheat Leaf Disease Detection: A Lightweight Approach with Shallow CNN Based Feature Refinement. *AgriEngineering*, *6*(3), 2001-2022. https://doi.org/10.3390/agriengineering6030117
7. X. Deng, Y. Lan, T. Hong and J. Chen, "Citrus greening detection using visible spectrum imaging and C-SVC," Computer and Electronics in Agriculture, November 2016.
8. D. Yang, F. Wang, Y. Hu, Y. Lan and X. Deng, "Citrus Huanglongbing Detection Based on Multi-Modal Feature Fusion Learning," Front. Plant Sci., vol. 12, 23 December 2021.
9. S. R. Dubey and A. S. Jalal, "Adapted Approach for Fruit Disease Identification using Images," International Journal of Computer Vision and Image Processing (IJCVIP), vol. 2, no. 3, pp. 44-58, 2012.
10. G. Geetharamani and J. A. Pandian, "Identification of plant leaf diseases using a nine-layer deep convolutional neural network," Computers & Electrical Engineering, vol. 76, pp. 323-338, June 2019.
11. J. Shermila P, E. A. Devi, A. Victor, and S. O. Manoj, “Automatic detection and classification of disease in citrus fruit and leaves using a customized CNN based model,” Boletin Latinoamericano y del Caribe de Plantas Medicinales y Aromaticas, vol. 23, no. 2, pp. 180–198, Mar. 2024, doi: 10.37360/blacpma.24.23.2.13.
12. Dhiman, Poonam & Kaur, Amandeep & Hamid, Yasir & Alabdulkreem, Eatedal & Elmannai, Hela & Ababneh, Nedal. (2023). Smart Disease Detection System for Citrus Fruits Using Deep Learning with Edge Computing. Sustainability. 15. 4576. 10.3390/su15054576.
13. Sindhu, P., & Indirani, G. (2020). IOT with Cloud based Smart Farming for Citrus Fruit Disease Classification using Optimized Convolutional Neural Networks.
14. Google Scholar https://AI and IoT Integration in Engineering: Use Cases, Challenges, and Opportunities in Nepal, RR Lamsal, M Bhattarai, S Lamsal - International Conference on AI and Engineering (ICAIE 2025) DOI: 10.13140/RG.2.2.30872.51208
15. Rauf, Hafiz Tayyab & Saleem, Basharat & Lali, Muhammad Ikram & Khan, M. & Sharif, Muhammad & Bukhari, Syed. (2019). A Citrus Fruits and Leaves Dataset for Detection and Classification of Citrus Diseases through Machine Learning. Data in Brief. 26. 104340. 10.1016/j.dib.2019.104340.
16. Lamsal, Ritu &. P, Karthikeyan & Otero, Pablo & Quintana, Alfonso. (2023). Design and Implementation of Internet of Things (IoT) Platform Targeted for Smallholder Farmers: From Nepal Perspective. Agriculture. 13. 10.3390/agriculture13101900.
17. Botpress https://botpress.com/ accessed on 12/16/2024
18. Googlecolab https://crfm.stanford.edu/2023/03/13/alpaca.html accessed on 12/16/2024

19. https://blog.google/technology/developers/google-gemma-2/ accessed on 12/16/2024
20. RAG using Llama 3.2 by Meta AI, https://lightning.ai/lightning-ai/studios/rag-using-llama-3-2-by-meta- ai?section=featured , accessed on 12/16/2024

Figures

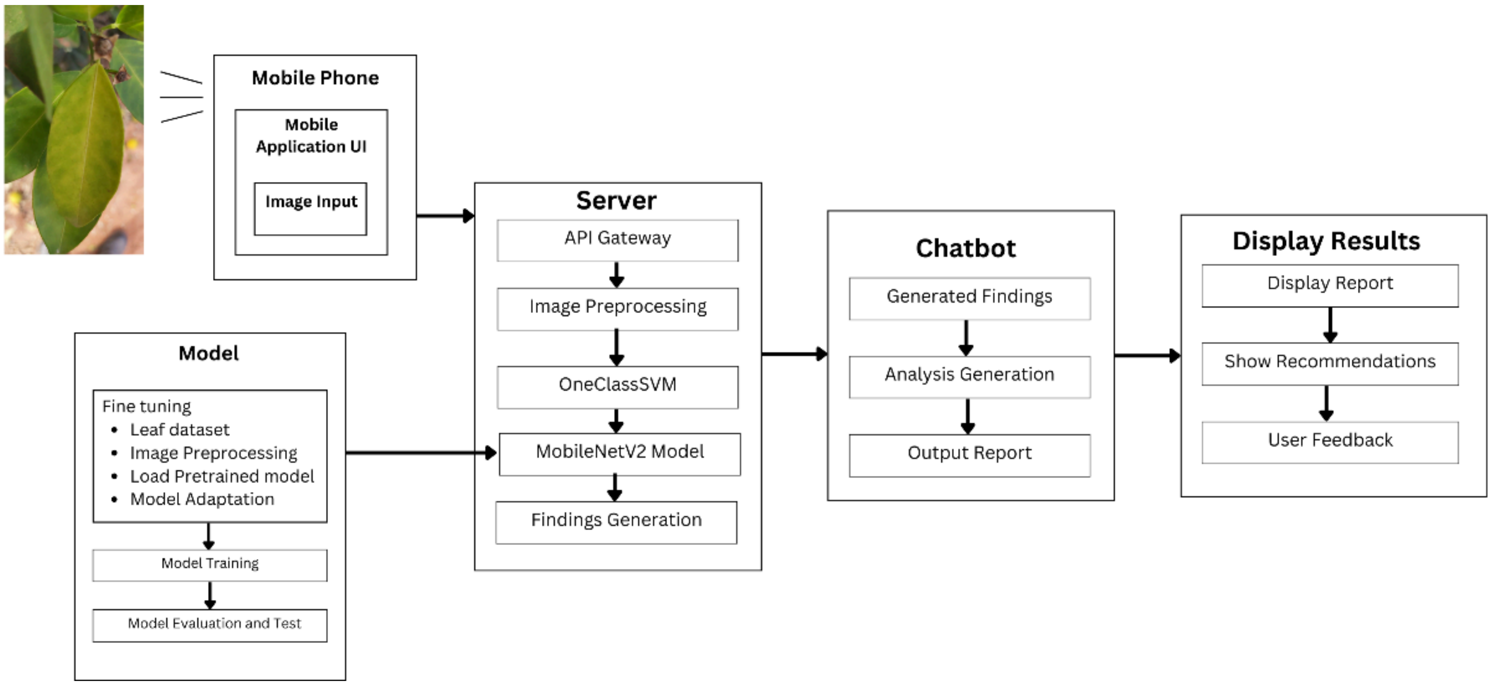

Figure 1

Overall System Architecture

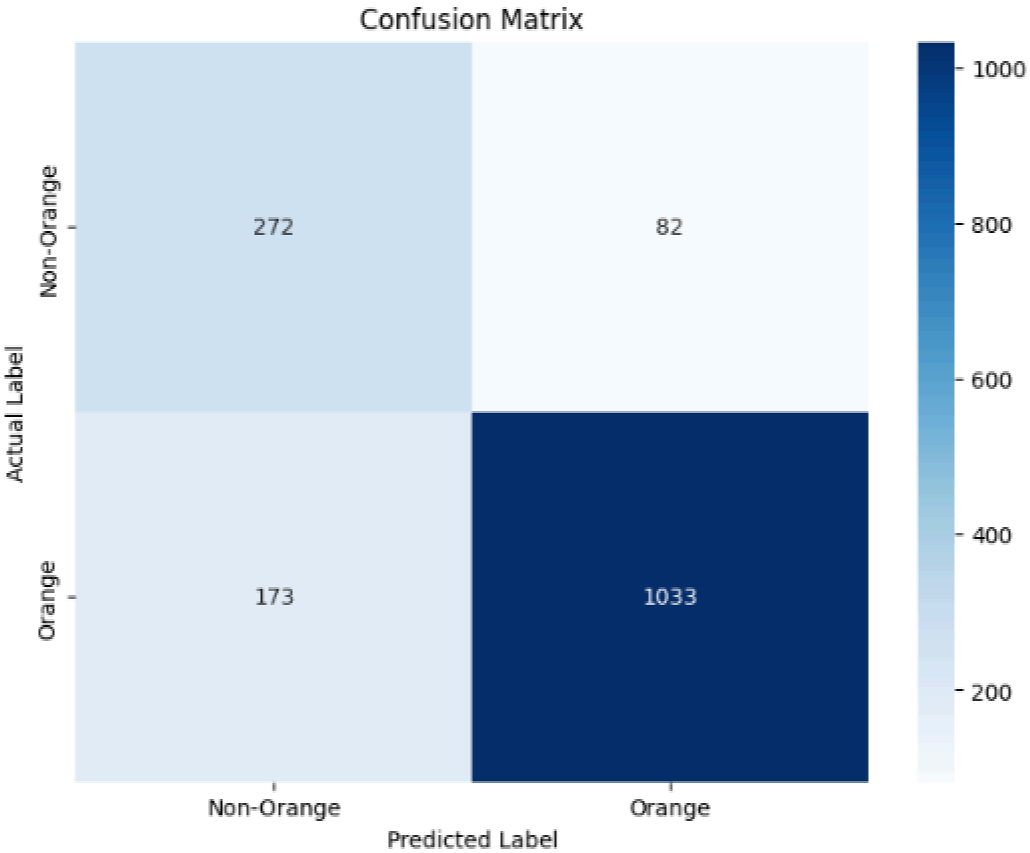

Figure 2

Confusion Matrix of One-Class SVM

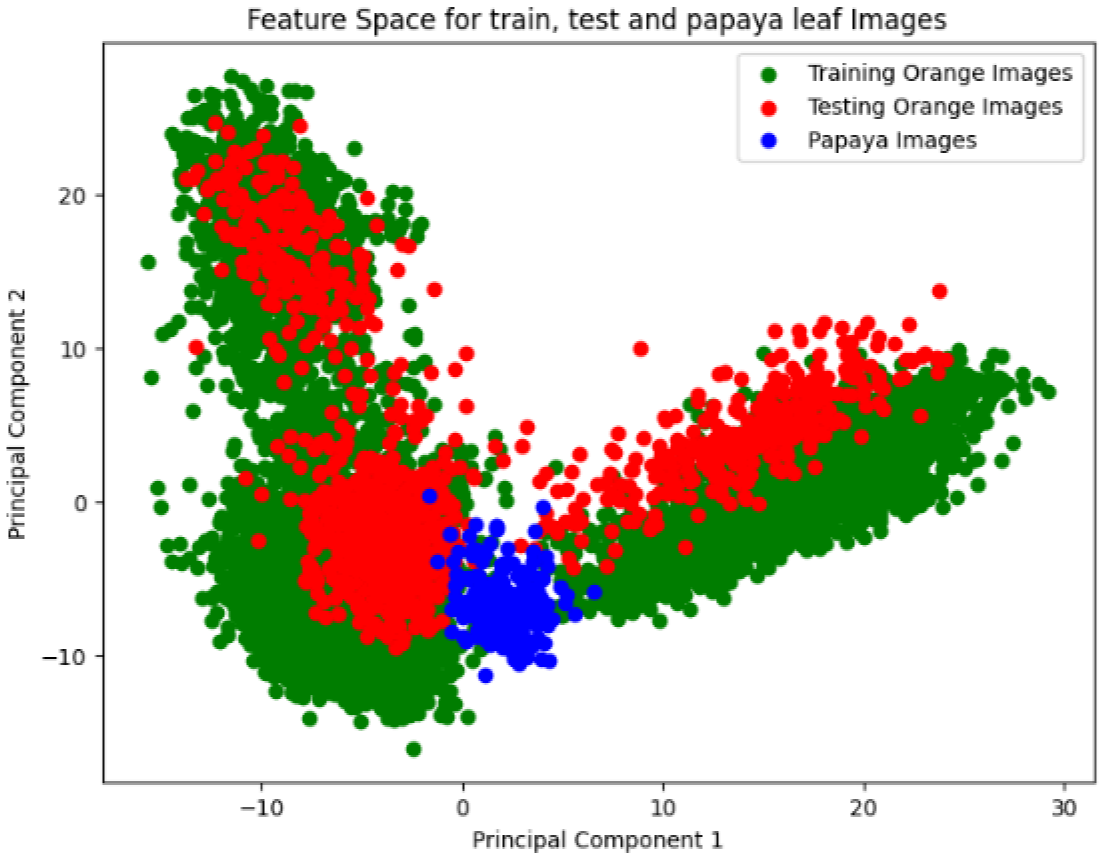

Figure 3

Feature Space with Train, Test and Additional Image

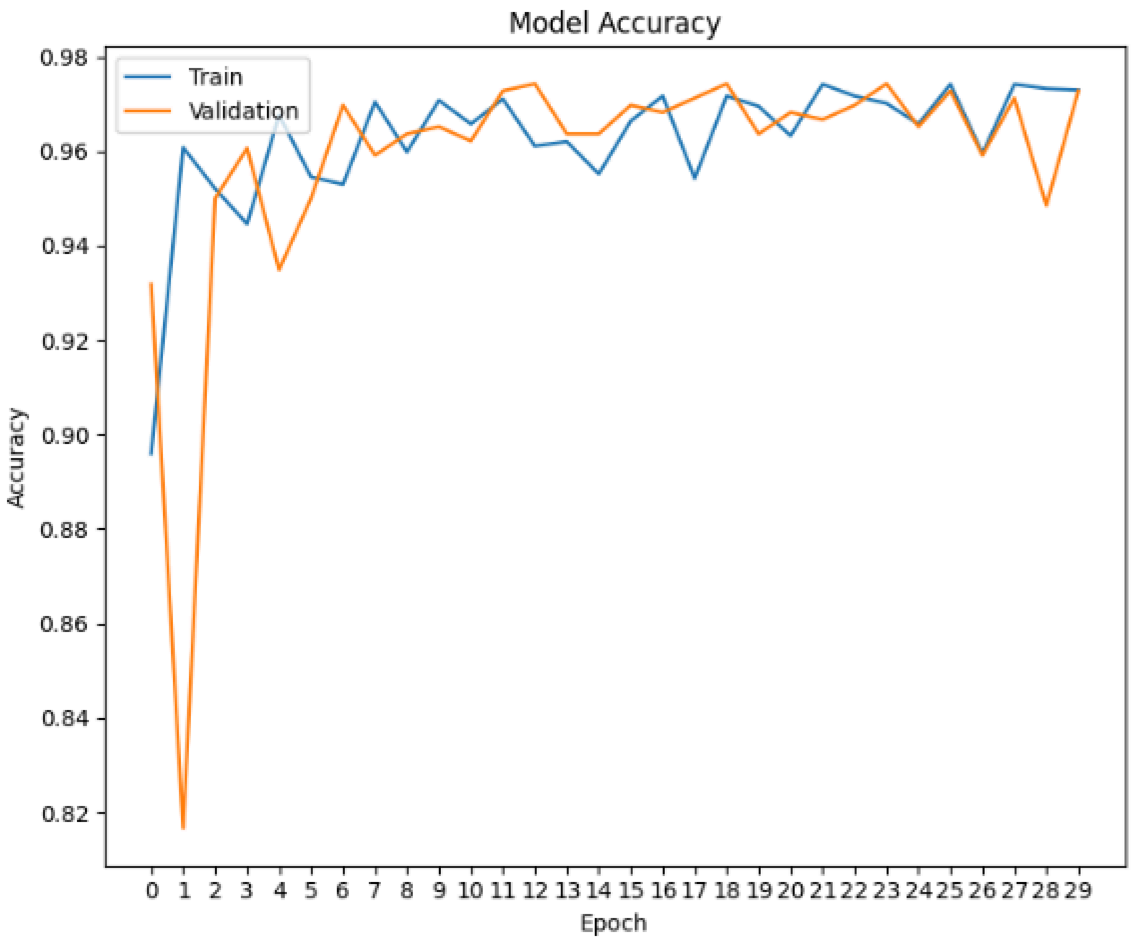

Figure 4

Accuracy graph of Inceptionv3

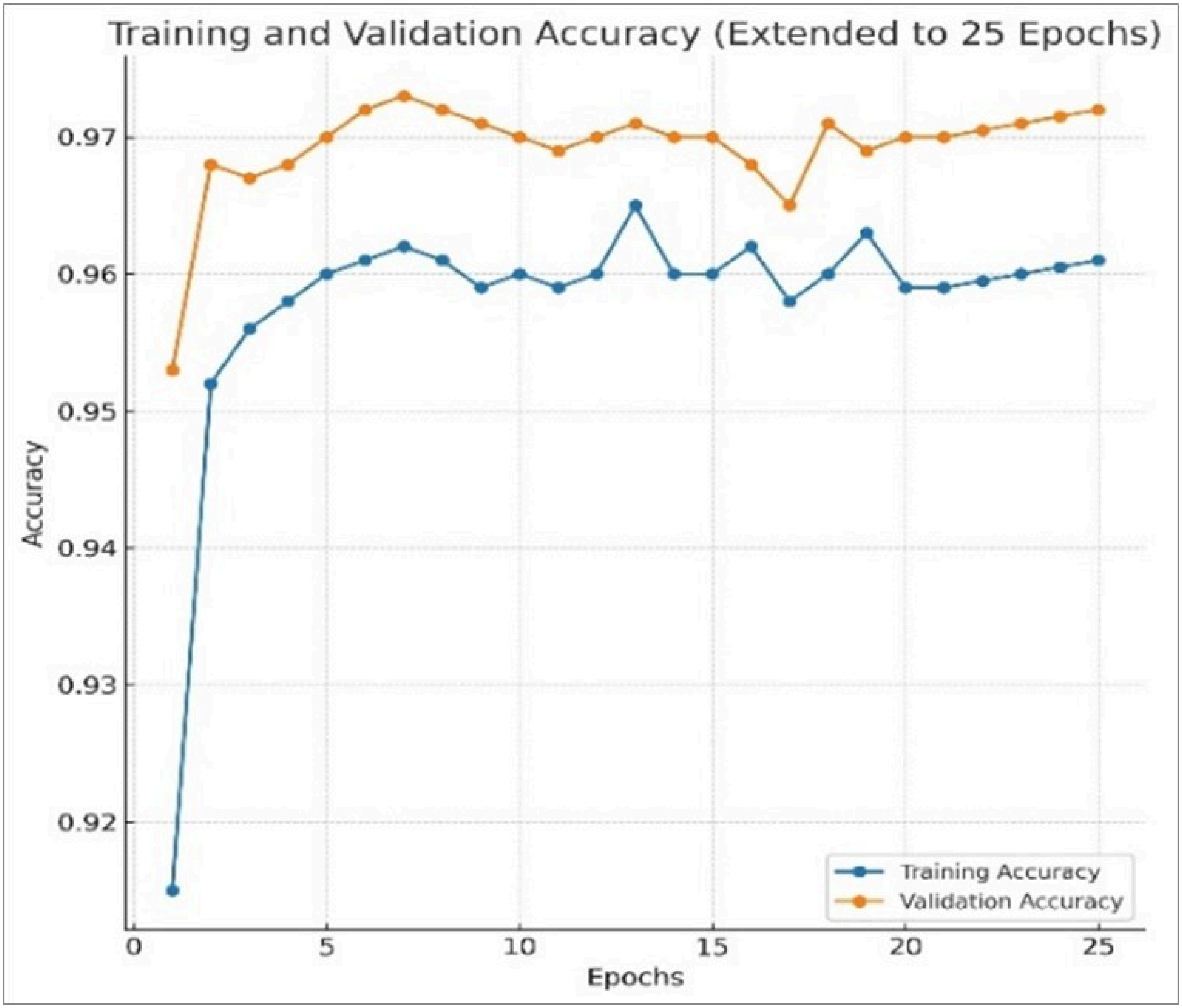

Figure 5

Accuracy graph of EfficientNet-B0

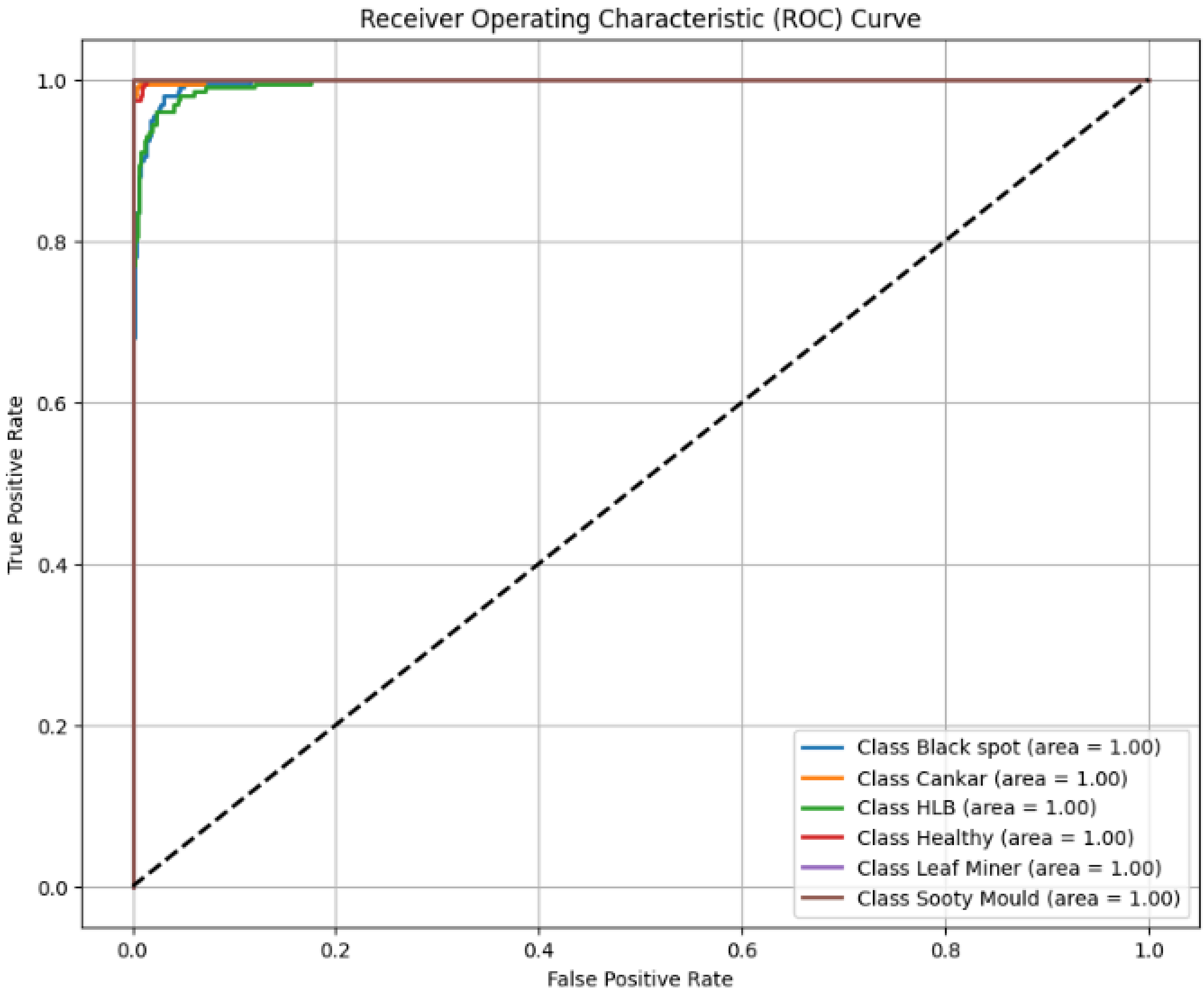

Figure 6

ROC curve of MobileNetV2

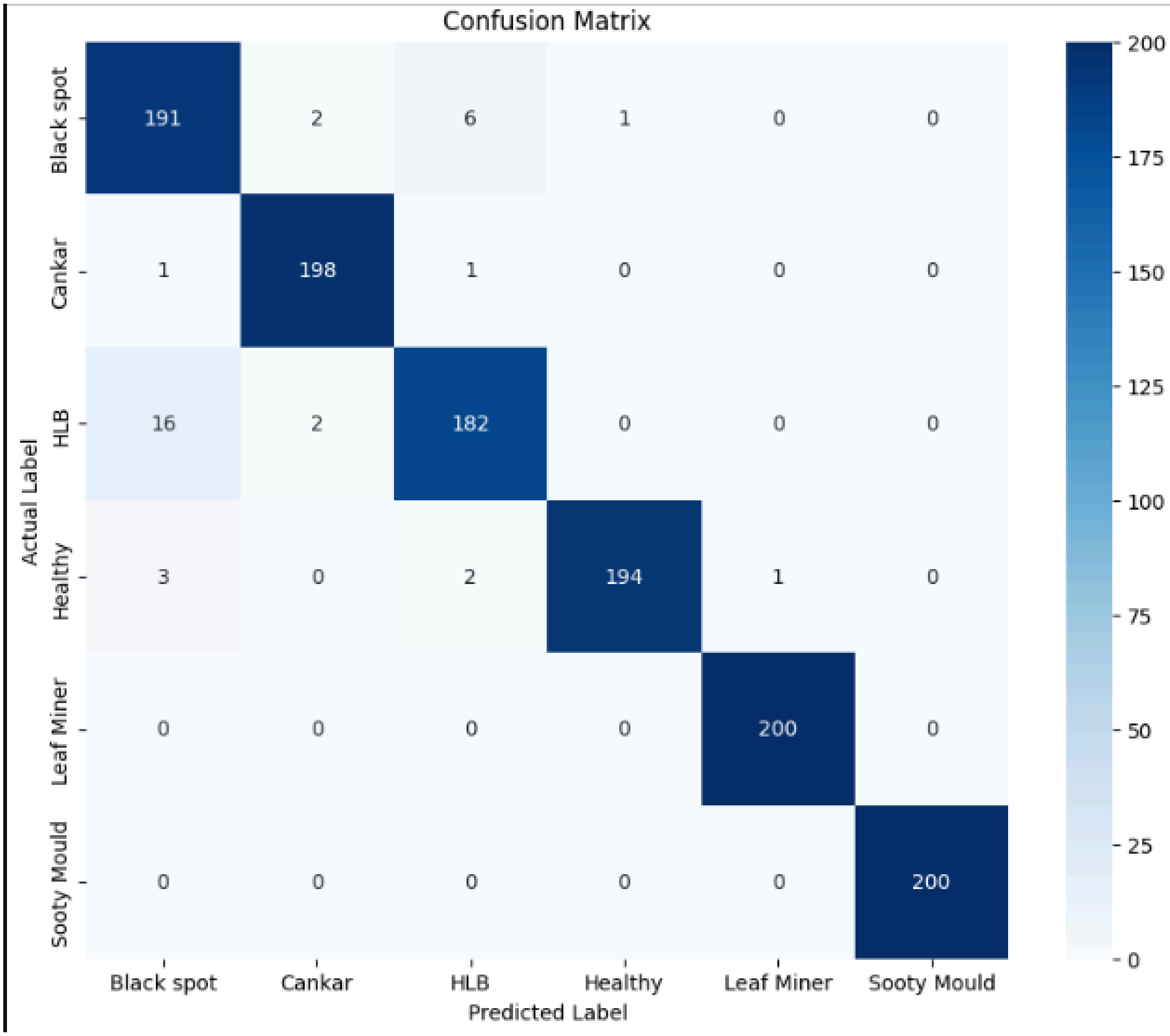

Figure 7

Confusion Matrix of MobileNetV2

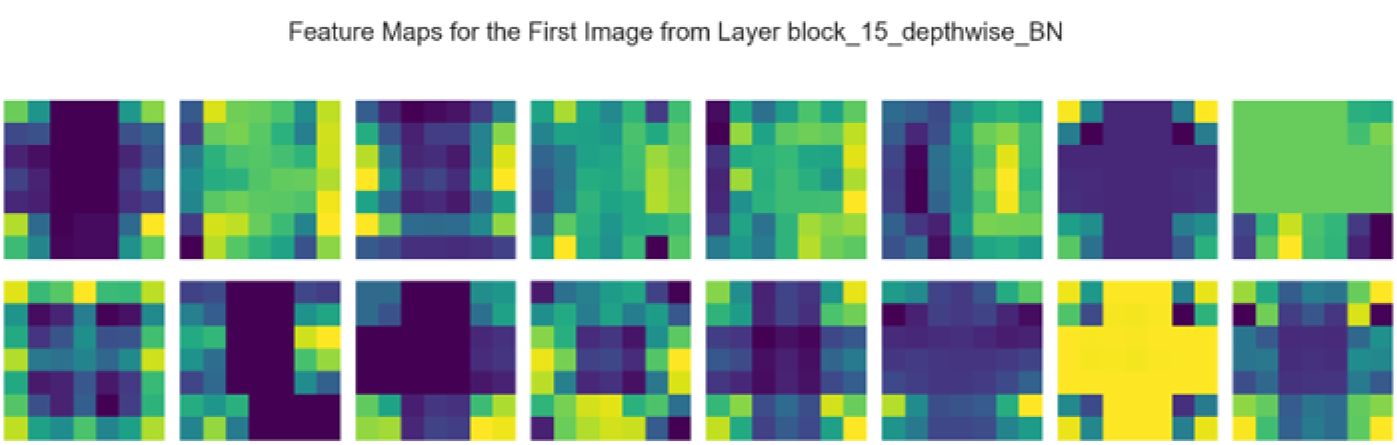

Figure 8

Feature Map of Layer block

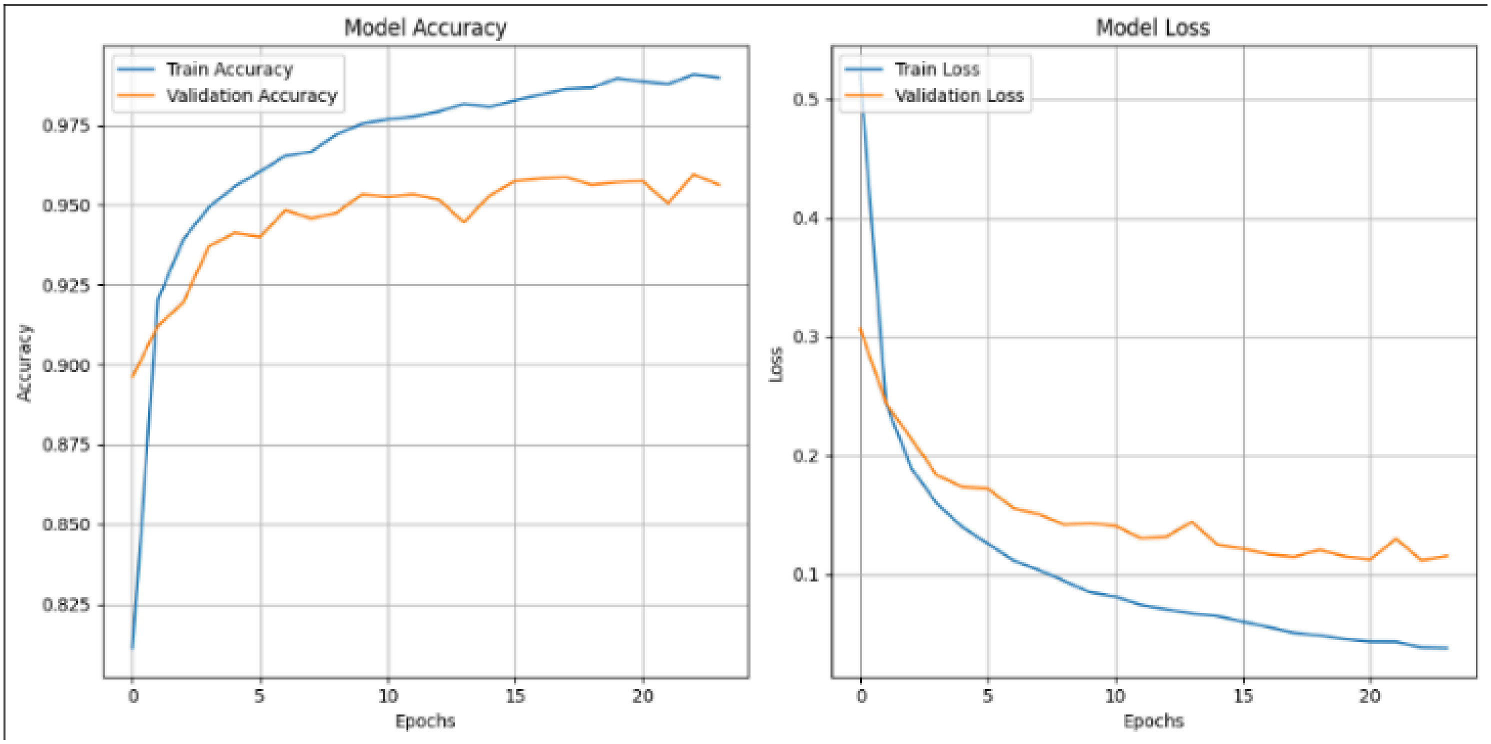

Figure 9

Accuracy and Loss plots of MobileNetV2

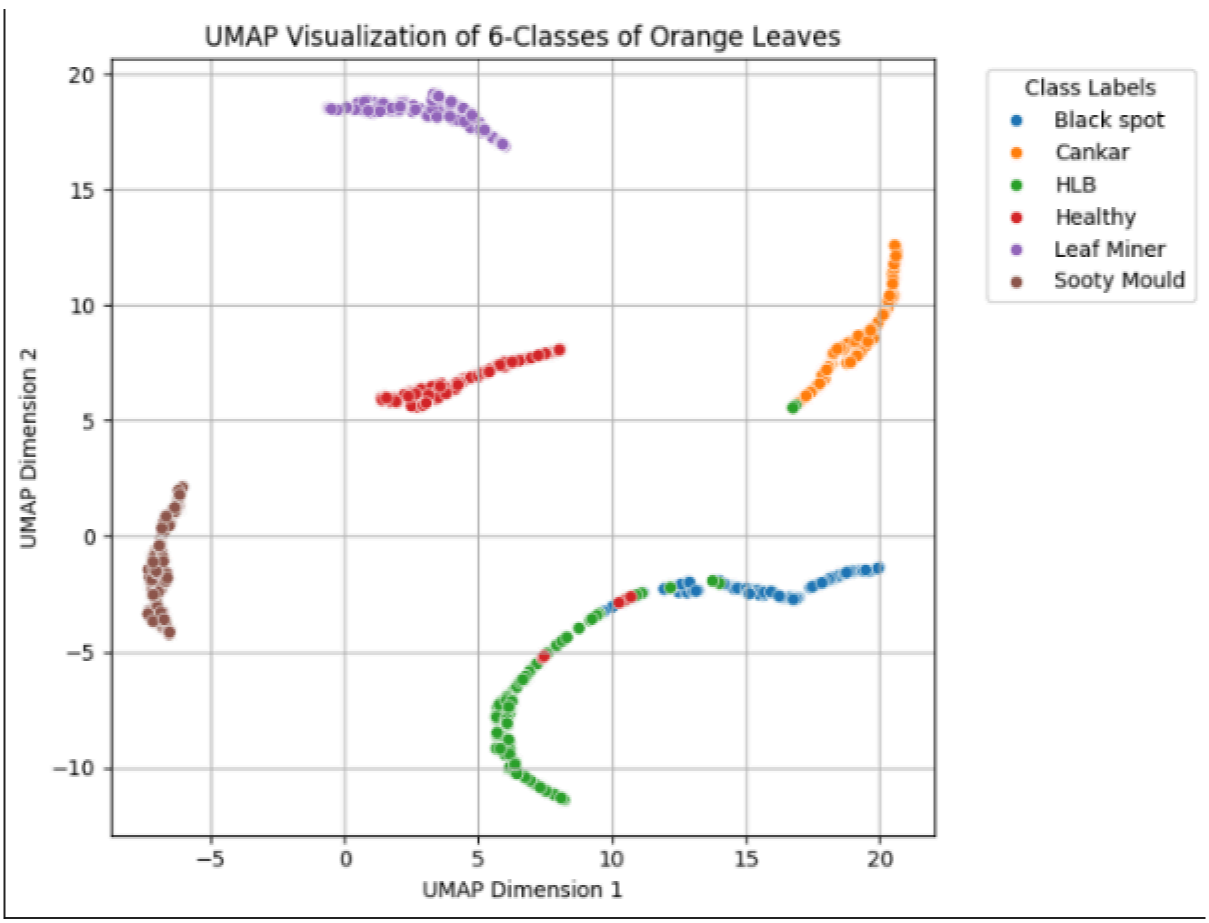

Figure 10

UMAP of MobileNetV2

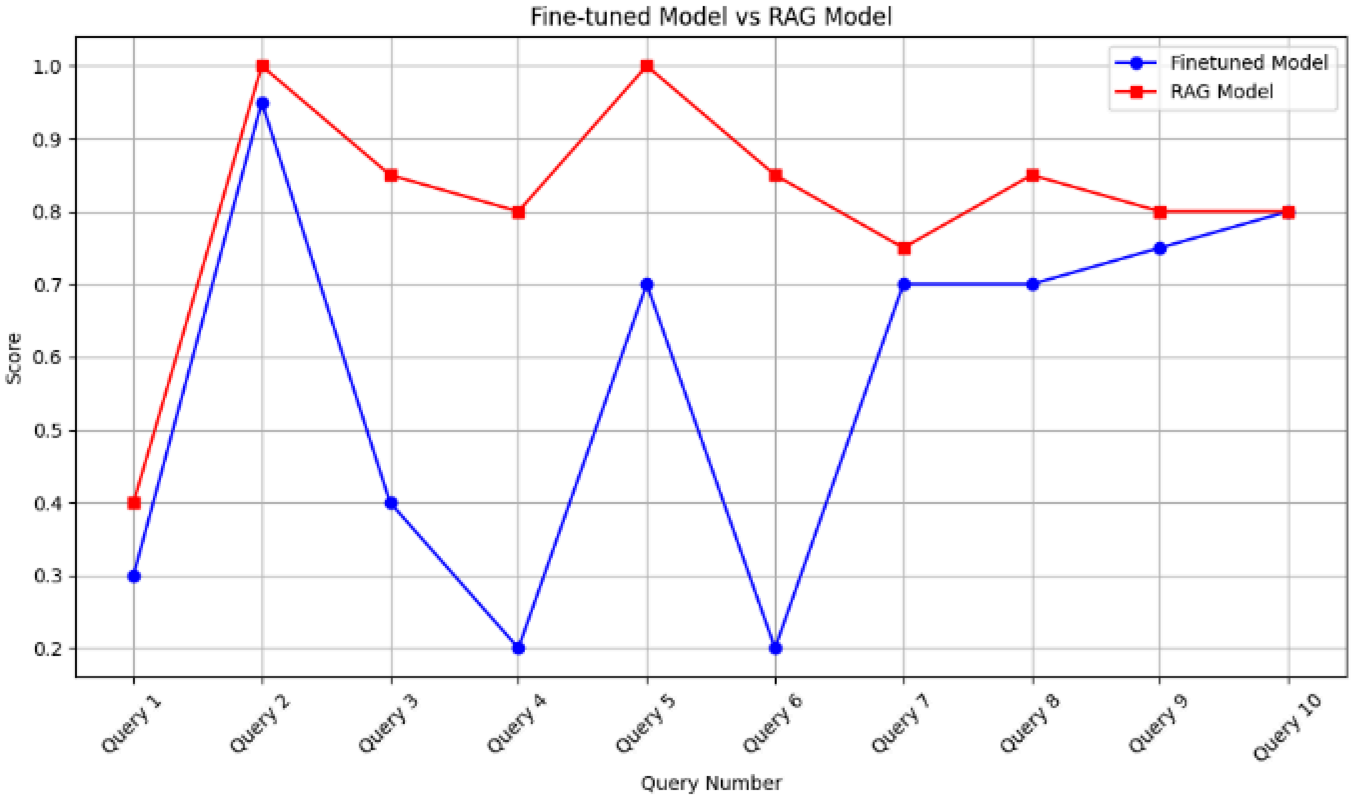

Figure 11

Chatbot Evaluation Line Graph

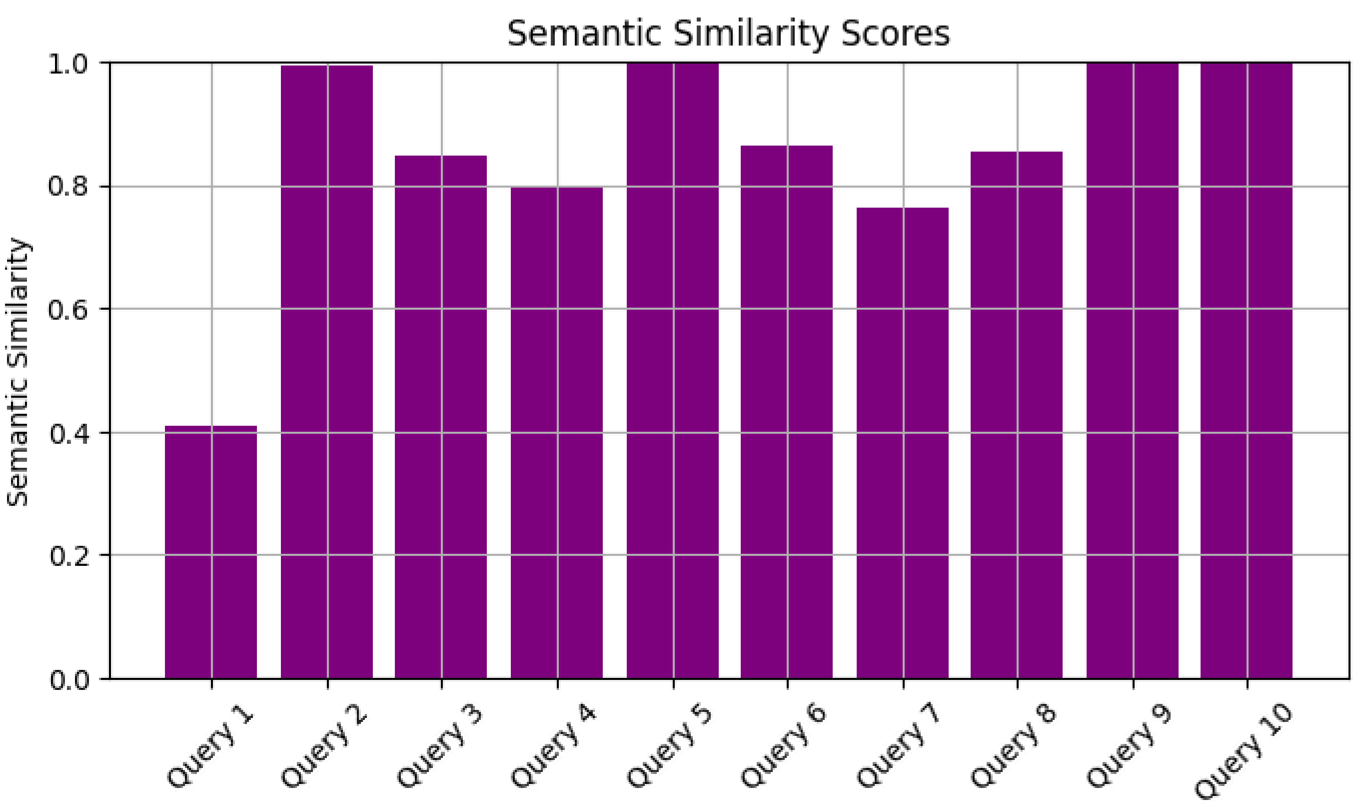

Figure 12

Bar chart showing Semantic Similarity Score

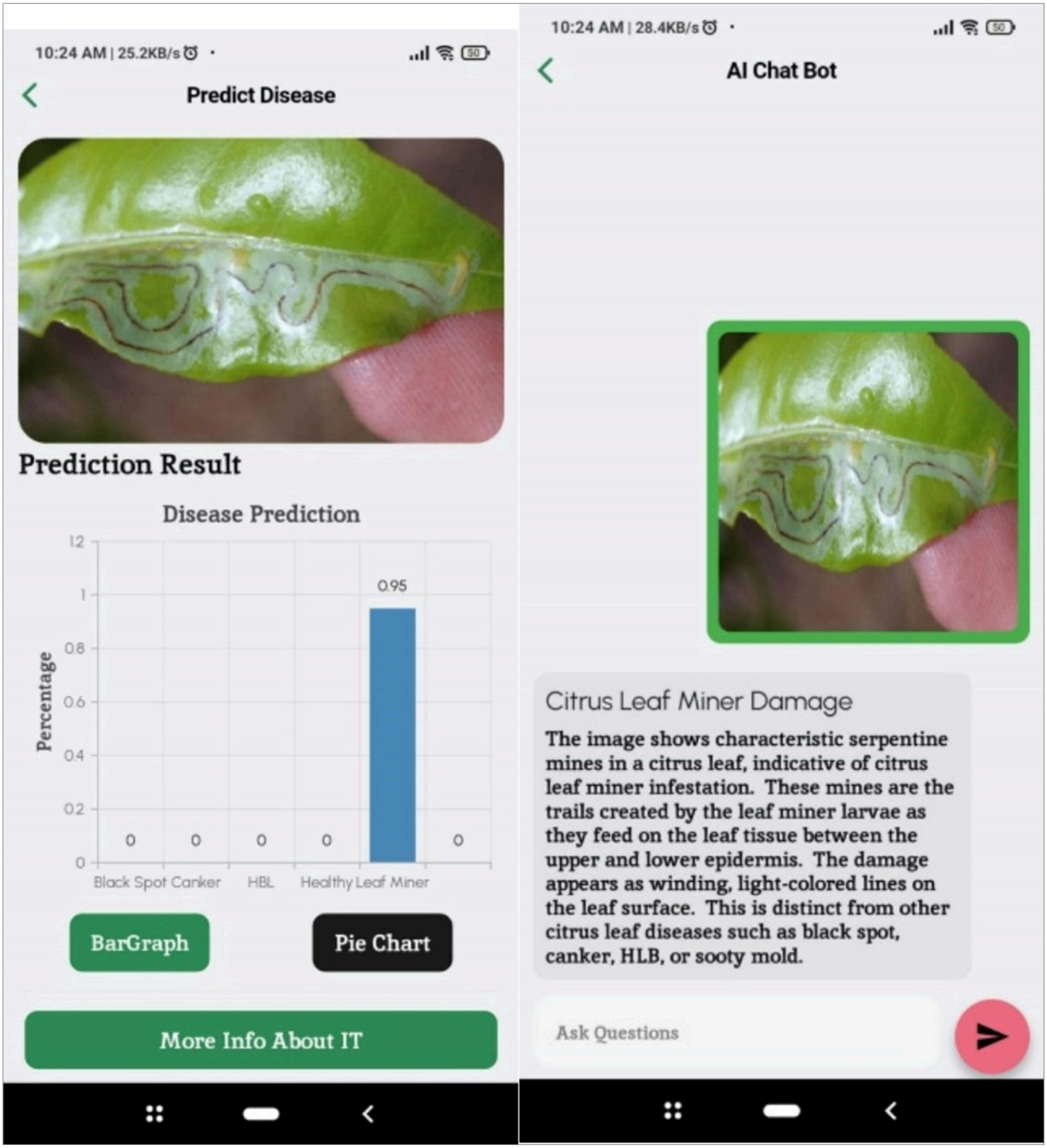

Figure 13

Mobile application UI