Leveraging a Modified Contrastive Language-Image Pre-training Model to Align Images and Text for Generating Remedy Text for Malus Pumila Lamina Images

Dhanasekaran Menaga
School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, India
menaga.d2018@vitstudent.ac.in

M. Sudha
School of Computer Science Engineering and Information Systems, Vellore Institute of Technology, Vellore, India
msudha@vit.ac.in (corresponding author)

Received: 17 December 2024 | Revised: 5 February 2025 and 22 February 2025 | Accepted: 24 February 2025

 DOI: https://doi.org/10.48084/etasr.9959

ABSTRACT

The increasing threat of leaf diseases to the productivity of precision farming necessitates systematic, logical, and scalable leaf identification methodologies. Conventional plant disease detection approaches are often slow, inefficient, and limited in their applicability, restricting the effective management of leaf diseases. This research work recommends a hybrid multimodal model that uses different modes of activities for leaf disease detection and can integrate image and text data in a single frame to improve the accuracy and proficiency of disease classification. The text data include custom-generated remedy descriptors specifically designed for the proposed model. The latter combines Machine Learning (ML) techniques, such as OTSU thresholding, Gaussian filtering, and modified Contrastive Language-Image Pre-training (mCLIP), to classify diseased leaves and propose suitable remedial actions. The proposed mCLIP model combines image and label data to enhance the effectiveness of multi-class image classification and suitable remedy description generation. Unlike existing multimodal approaches that primarily output text describing image features, the proposed model generates remedy text as the output for specific diseases. This novel approach offers a comprehensive solution for leaf disease detection and renders optimistic results for real-time and automated disease identification in agricultural practices, facilitating early intervention and better crop management. The proposed model obtained an accuracy of 98.1%.

Keywords-OTSU;Gausian filter image segmentation; multimodal; mCLIP

I. INTRODUCTION

The proposed model leverages Deep Learning (DL) techniques to investigate leaf images for optical features, such as discoloration, lesions, and deformities, while integrating contextual descriptions derived from horticultural practices or expert annotations. The combination of these two modalities facilitates the model to extract supplementary data, thereby improving the accuracy of disease detection and the diagnostic capabilities of the model. The proposed methodology employs publicly available leaf disease datasets to evaluate the performance of the system, with experimental analyses demonstrating significant advancements in classification accuracy compared to image-based methods. The introduced model is a unique approach in image enhancement and image classification approaches as it not only detects the disease, but also predicts suitable remedy descriptions for new leaf samples. Its multi-modal approach combines visual data or leaf images with other types of sensory information with the help of environmental or soil data to identify anomalies that may indicate plant diseases.

The multimodal approach [1-4] integrates different inputs, such as images, environmental data, and sensor reading values, to enhance plant disease prediction. This approach intends to improve the precision and efficiency of monitoring the plant health by exploiting advanced DL techniques [1, 2]. This approach improves the diagnostic results by creating authentic training data and reducing overfitting. The research intended to achieve advanced automated crop disease detection in agriculture through extra robust AI models. Plantdet [5] exploits the strengths of different algorithms by combining the excess of models and improves the results accurately in the identification of plant diseases. This approach creates a more reliable diagnostic system for agriculture and addresses the challenges of identifying plant diseases, constituting a scalable and effective solution for precision agriculture. The PDC-VLD model [6] integrates various data types to enhance the accuracy and generalization of disease detection. The approach extracts valuable performance across diverse conditions, making it adaptable to different plant varieties and environments. In [7], the multimodal transformer model was proposed to detect agricultural diseases and facilitate powerful systems. An advanced structure was introduced for detecting different types of leaf diseases and their severity. The models presented in [8, 9] combine Convolutional Neural Networks (CNNs) and transformers with a multi-label approach, effectively handling the complexities of plant disease classification. In [10], Bayesian techniques were employed for classifying tomato leaf diseases combining multiple data modalities in a hybrid learning framework to improve classification accuracy.

In [11], a novel approach was introduced for detecting diseases and pests in pepper plants using a multimodal framework. The framework employed a combination of different modalities for an early detection, enabling better management of pest and disease threats in pepper crops. This resulted in improvisation of crop production through timely interventions. In [12], the use of computer vision and image processing was explored to analyze leaf images and identify symptoms of diseases that can affect healthy leaves. This approach highlights the potential of using technology to improve plant disease management. In [13], DL techniques were utilized to analyze leaf images and accurately detect various plant diseases. The model proposed in [1] exploits visual cues, such as images of plant leaves, along with other relevant details. By integrating various data types, the model enhances the ability to early identify and diagnose plant health issues, contributing to more efficient and precise agricultural practices. In [14], a hybrid approach combining the Xception model and the Random Forest (RF) algorithm was proposed for the classification of multiple plant leaf diseases. In [15], a cascading model was presented for classifying two specific maize diseases, namely, Fusarium stalk rot and charcoal rot. This deep feature extraction technique utilizes processed images of the maize plant to enable accurate disease classification. In [16, 17], a method utilizing biorthogonal wavelet-based entropy features was proposed to identify the diseases in maize leaves. The approach extracts key features from leaf images through wavelet transformation and focuses on entropy to distinguish between healthy and diseased leaves, providing an efficient tool for the early detection and management of maize leaf diseases. The approach presented in [18] aims to provide farmers with a reliable tool for early intervention and effective crop protection and improve disease management in agriculture. PPLC-Net [19] is a combined model optimized with weather data augmentation and a multi-level attention mechanism for its accuracy to be improved. Authors in [24] formulated a model to evaluate and recognize disease symptoms on a broad scale, providing a remote sensing methodology for monitoring the health of trees.

Table I provides a summary of the reviewed studies.

TABLE I. LITERATURE REVIEW SUMMARY

Ref.	Year	Methodology	Applications
[17]	2023	Feature maps	Plant disease detection
[20]	2023	DL	Early diagnosis
[21]	2023	DL	Detection of blight diseases
[22]	2023	CNN	Rice disease classification
[23]	2023	CNN	Coffee disease diagnosis
[24]	2023	Hyperspectral imagery	Vascular plant pathogens
[25]	2023	IP	Wheat stripe rust disease
[26]	2023	Feature extraction	Olive disease diagnosis
[27]	2024	Feature selection techniques	Weed removal
[28]	2023	SHO	Rice disease classification
[29]	2023	YOLO v5	Tomato disease classification
[30]	2024	IP	Potato plant disease detection
[31]	2024	Optimized Edge AI	Real-time tomato leaf disease identification.
[32]	2024	MRFO	Maize leaf disease detection
[33]	2023	Feature aggregation	Nutrient deficiencies in chili plants

II. PROPOSED SYSTEM ALGORITHM

The proposed hybrid multi model's sequential approach is explained in the following algorithm.

```
Input : Apple leaf image dataset I [0… n].
Prepared custom remedy text descriptions
T [0… m]
Target label for the disease leaf image
classification L [0, 1, 2…, and 12], Total
number of labels =13
Develop CustomDataset (I,T,L)
Apply Gaussian filter approach to remove
the noise from the image set I
Apply the OTSU thresholding approach to
separate the background and foreground
features from the Image Set I.
Define a custom CLIP model that combines
image/text features with label embeddings
combined_features=concat(image_features,la
bel_embeds), dim=1
combined_features=concat(text_features,lab
el_embeds), dim=1
output=fc(combined_features)
num_labels=len(dataset.data['label'].uniqu
e())
label_embedding=LabelEmbedding(num_labels)
optimizer=Adam(modified_clip.parameters(),
lr=1e-5)
Extract image and text features
Normalize image and text features
```

$$\text{image features} = \frac{\text{image\_features}}{\|\text{image\_features}\|^2}$$

$$\text{text features} = \frac{\text{tex\_features}}{\|\text{text\_features}\|^2}$$

$$\text{Label features} = \frac{\text{Label\_features}}{\|\text{Label\_features}\|^2}$$

```
Calculate Contrastive Loss (CL) of image,
text, and label from similarity scores
Compute overall loss for both text and
image
print the selected text for the image S
(image)
```

III. PROPOSED SYSTEM ARCHITECTURE

The mCLIP is a new approach for the text, image, and label embedding for effective multi-class disease classification with remedy text generation.

A. *Conventional Contrastive Language-Image Pre-training*

The conventional CLIP model allows only embedding of text and images. The conventional CLIP model uses ResNet or a Vision Transformer model for encoding images. Similarly for text encoding, the BERT or GPT model can be used. Both encoders can map the text and image in a shared environment and train the conventional CLIP model utilizing contrastive learning. The latter maximizes the similarity between images and matching text and minimizes the similarity of non-matching images.

B. *Modified Contrastive Language-Image Pre-training*

The proposed mCLIP, incorporates label encoding with the conventional CLIP model. The labels are custom-generated labels, trained along with the original CLIP features. This alteration includes a label embedding layer to the conventional CLIP model. The proposed mCLIP allows a separate embedding procedure for the class labels and concatenates with the text and image features before loading into the classification layer.

The multimodal model consists of several phases, namely, text and image preprocessing phase, feature extraction phase, image segmentation and noise reduction phase, text and image embedding vector generation, contrastive learning approach, similarity computation, and image classification. The output of the mCLIP model is fed into the classification layer that maps the text/image with label encoding to the final leaf disease classification. Figure 2 explains the proposed mCLIP system's architecture. The step by step procedure of the mCLIP model is explained below.

1) *Preprocessing Step*

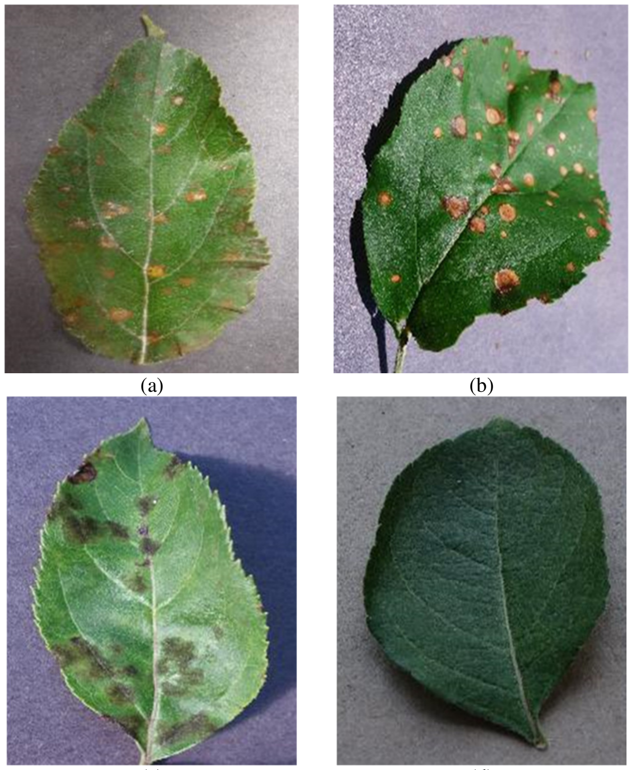
Resizing and normalizing are carried out in preprocessing. During this phase, the images are reduced to the default size of 224 × 224. Some sample input images are displayed in Figure 1.

2) *Feature Extraction*

The OTSU thresholding approach is deployed to extract features from the images. This thresholding approach calculates the local and global features and threshold values and divides the image into foreground and background images.

3) *Embedded Vector Generation*

Custom remedy texts are generated based on the examination of leaf diseases and the latter's severity. The remedy text is tokenized into words, and the BERT model is used to create the text embedding vector. In the image segmentation phase, the OTSU method is employed to extract the image features. The text- and image-embedded vectors are mapped together, resulting in the creation of a text-image embedded vector for the CLIP architecture.



(a) (b) (c) (d)

Fig. 1. Apple leaf samples: (a) Cedar rust, (b) black rot, (c) scab, (d) healthy.

4) *Modified Contrastive Learning Approach*

The proposed mCLIP model adds a layer on the conventional CLIP model and uses the cross-entropy loss for the training phase. The mCLIP model is trained to predict the labels by combining image and text features with the label embedding. The model tries to reduce the incorrect alignment between the image–text pair with the help of class labels.

5) *Similarity Computation*

In the mCLIP model, the images and texts are processed through the respective encoders and their features are extracted. The extracted features are normalized with the L2 normalization approach. The label embedding layer converts the labels into the dense embedding of default size vectors. The default size for the label embedding is 64. In this phase, the similarities between the text and image pairs are computed in the mCLIP model by the dot product approach.

6) Image Classification

The mCLIP model uses the zero shot classification to match the similarity of the image pairs. The former is a completely enhanced version of the conventional CLIP, because it computes the similarity between text-image pairs and label embedding. In contrast, the conventional CLIP model only computes the similarity between image-text pairs.

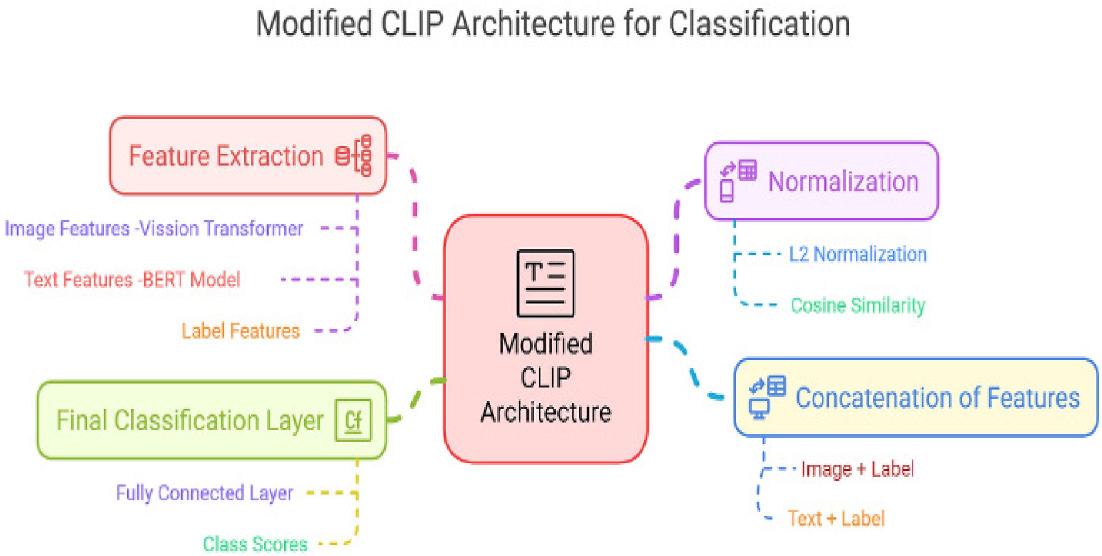


Fig. 2. Architecture of the proposed mCLIP system.

IV. DATASET DESCRIPTION

The proposed system used apple leaf images from the plant village dataset [34]. It contained 3171 leaf images of diseased and healthy leaves. The detailed description of selected image dataset is explained in Table II. The number of leaves existing in the plant village dataset for each of the three categories is mentioned along with the generated text description. Totally, 13 labels were assigned to the images, based on the disease severity.

TABLE II. APPLE LEAF DISEASE CATEGORIES AND LABEL TYPES

	Leaf disease category	**Custom remedy description**	**No. of images**
1	Black Rot	Prune the leaves on light black spot	124
2		Pesticides are applied on heavy black spots	124
3		Proper lighting is needed for shape changes on leaf	124
4		Dissolve baking soda spray on leaf bugs	125
5	Cedar Rust	Watering the leaves to the leaf twigs	54
6		Leaf curl is prevented by adding the pesticides	55
7		Inhibit the fungicides on brownish spots	55
8		Avoid over watering for discoloring to yellowish	54
9	Scab	Pesticides on leaf spots	127
10		Spray phosporous liquid on mottled leaf	126
11		Apply fungicides for black mold	128
12		Clean the leaf regularly to remove the black spot	127
13	Healthy	Healthy	1645

V. RESULTS AND DISCUSSION

The proposed multimodal mCLIP system was evaluated in apple leaf image classification. A total of 2536 images were selected for the training phase and the remaining 635 images were considered for the testing phase. The performance of the proposed model was compared with those of CLIP, Bootstrapping Language-Image Pretraining (BLIP), and Vision Transformer (ViT). The current paper analyzed those models' performance on the leaf disease detection. The output analysis with the other models concentrates on predicting one among n outputs. The multimodal models enable having multiple outputs for a single input. The introduced approach outputs a suitable remedy text for the given image along with disease classification. The proposed model framework consists of OTSU, Gaussian filter, and the mCLIP model to detect the leaf diseases from images. The former is a novel approach in the multi model image-text pair classification.

Figure 3 demonstrates the confusion matrix of the mCLIP model. It shows that mCLIP mostly predicts the samples correctly. The performance of the other models was poor, as can be seen in Figures 4-6, which depict the confusion matrices of ViT, BLIP, and conventional CLIP models. The mCLIP model obtained 98.1% overall test accuracy, while the other multimodal models were not able to secure more than 90%. This can be attributed to the label embedding layer introduced in the mCLIP model. Table III outlines the computed class-wise precision, recall, and F1-score results of the proposed mCLIP model and the other multimodla models. Figures 7-9 show the chart representation of class-wise precision, recall, F1-score, and accuracy from the obtained results of the considered models. The other multi-models failed to recognize classes 1, 6, and 7. The class label encoder helps the mCLIP model in the training phase to recognize similar images in the testing phase. This label embedding is a very effective approach utilized by the proposed model.

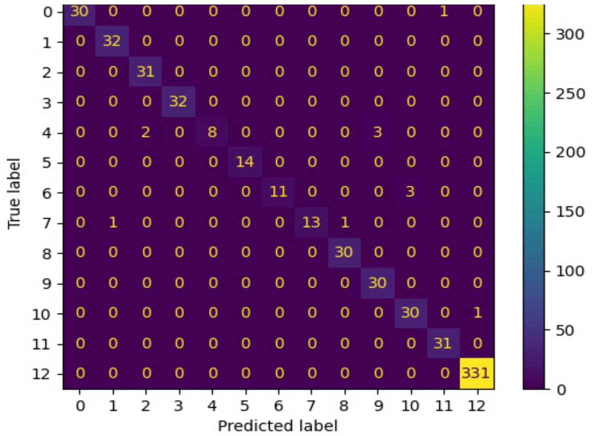


Fig. 3. Confusion matrix of the proposed mCLIP model.

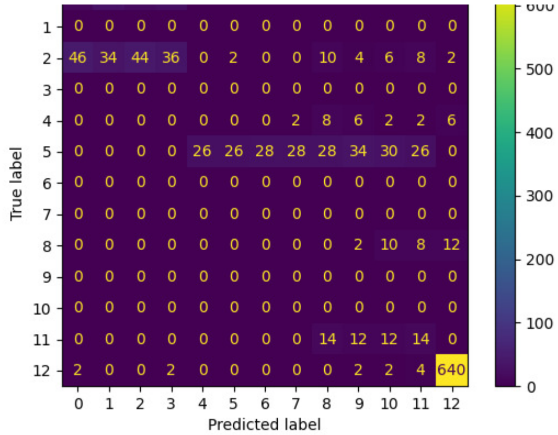


Fig. 5. Confusion matrix of the BLIP model.

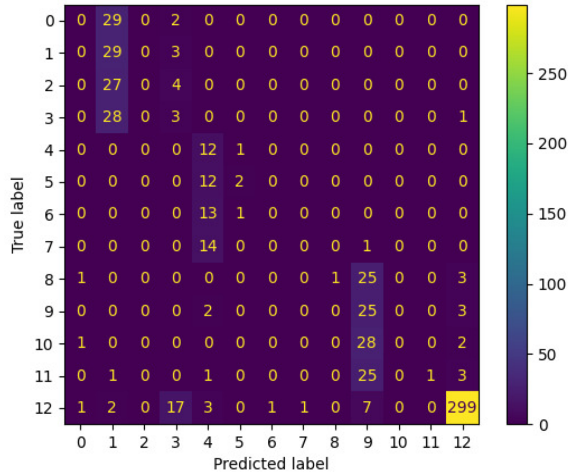


Fig. 4. Confusion matrix of the ViT model.

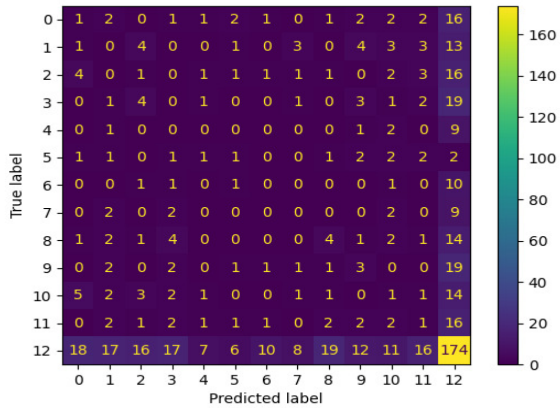


Fig. 6. Confusion matrix of the conventional CLIP model.

TABLE III. PERFORMANCE METRICS FOR BLIP, CLIP, VIT, AND mCLIP

	ViT			BLIP			CLIP			mCLIP		
Class	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
0	0.1	1	0.19	0.23	0.16	0.18	0.03	0.03	0.03	1	0.97	0.98
1	0	0	0	0	0	0	0	0	0	0.97	1	0.98
2	0	0	0	0.71	0.23	0.35	0.03	0.03	0.03	0.94	1	0.97
3	0	0	0	0	0	0	0	0	0	1	1	1
4	0	0	0	0	0	0	0	0	0	1	0.62	0.76
5	0	0	0	0.93	0.12	0.2	0.07	0.07	0.07	1	1	1
6	0	0	0	0	0	0	0	0	0	1	0.79	0.88
7	0	0	0	0	0	0	0	0	0	1	0.87	0.93
8	0	0	0	0	0	0	0.13	0.13	0.13	0.97	1	0.98
9	0	0	0	0	0	0	0.1	0.1	0.1	0.91	1	0.95
10	0	0	0	0	0	0	0.03	0.03	0.03	0.91	0.97	0.94
11	0	0	0	0.23	0.27	0.25	0.03	0.03	0.03	0.97	1	0.98
12	1	1	1	0.97	0.98	0.97	0.53	0.53	0.53	1	1	1
Macro avg	0.08	0.15	0.09	0.24	0.13	0.15	0.07	0.07	0.07	0.97	0.94	0.95
Weighted avg	0.53	0.57	0.53	0.79	0.58	0.61	0.29	0.29	0.29	0.98	0.98	0.98

Figure 10 presents some classification outputs and the accompanying remedy text descriptions. Figure 11 and Table IV exhibit the overall performance of mCLIP, BLIP, and CLIP multi models. The obtained accuracy of the proposed mCLIP surpasses those of the other models. The proposed hybrid multimodal system obtained overall precision of 97.3%, recall of 98.5%, and F1-score of 97.66%. Precision, recall, and F1-score are key metrics utilized to evaluate classification models, especially for imbalanced datasets. High precision indicates that the model is good at identifying positive instances with few false positives. High recall means that the model captures most of the actual positive instances, with few false negatives. F1-score is a good metric when the dataset is imbalanced, as it considers both false positives and false negatives. When Precision is high but recall is low, the model is cautious in making positive predictions. Subsequently, when recall is high but precision is low, the model is more lenient in predicting positives.

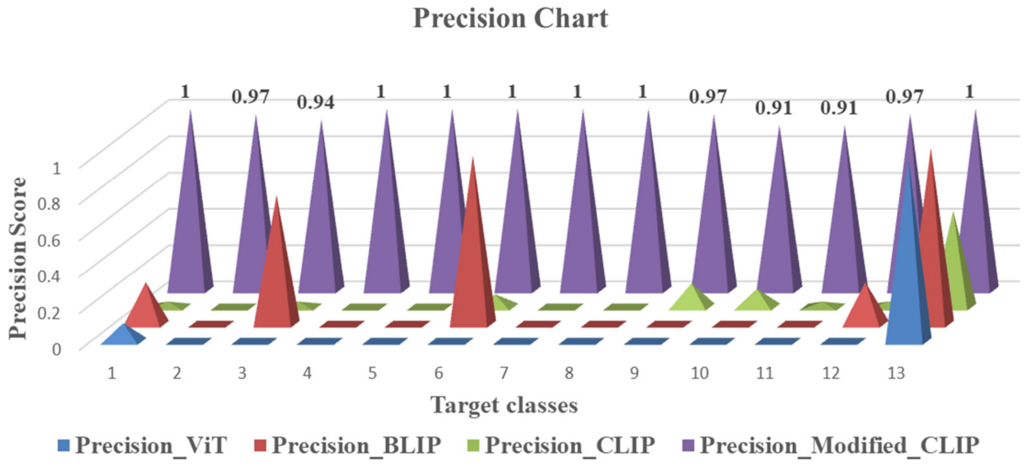


Fig. 7. Obtained precision results of ViT, mCLIP, BLIP, and CLIP models.

Fig. 8. Obtained recall results of ViT, mCLIP, BLIP, and CLIP models.

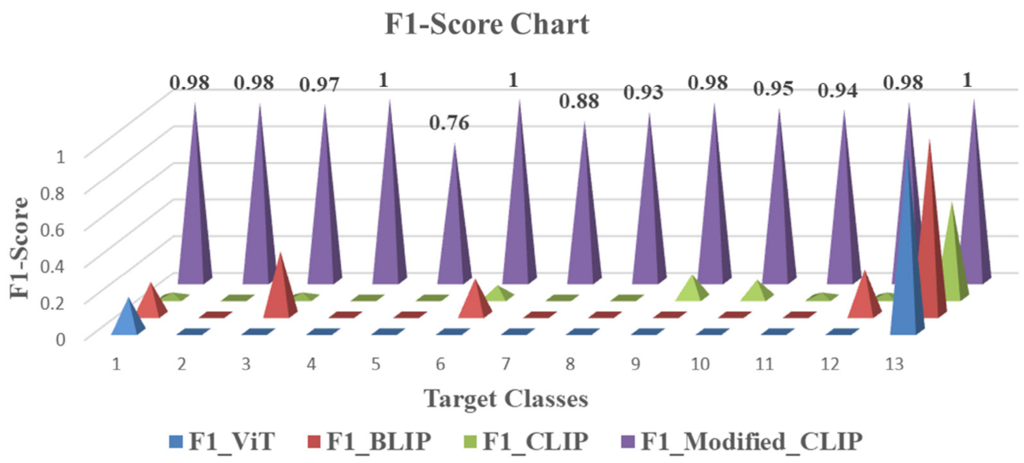


Fig. 9. Obtained F1-score results of ViT, mCLIP, BLIP, and CLIP models.

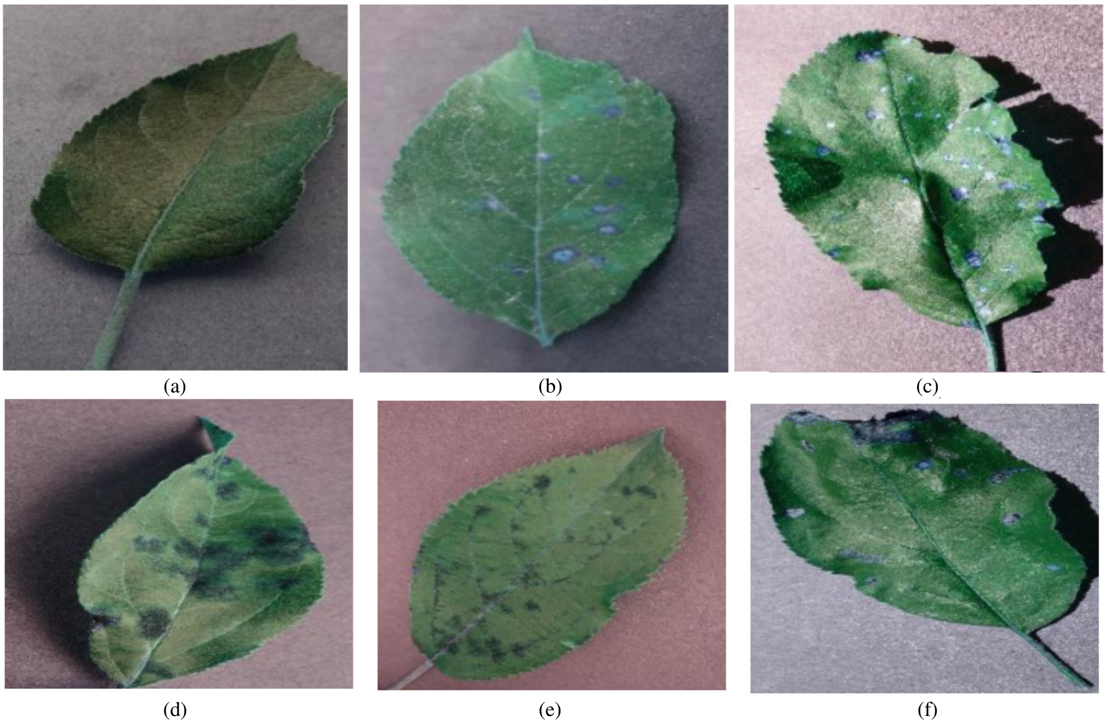


Fig. 10. Proposed system's sample outputs of remedy text: (a) Healthy, (b) avoid over watering for discoloring to yellowish, (c) proper lighting is needed for shape changes on leaf, (d) healthy, (e) avoid over watering for discoloring to yellowish, (f) proper lighting is needed for shape changes on leaf.

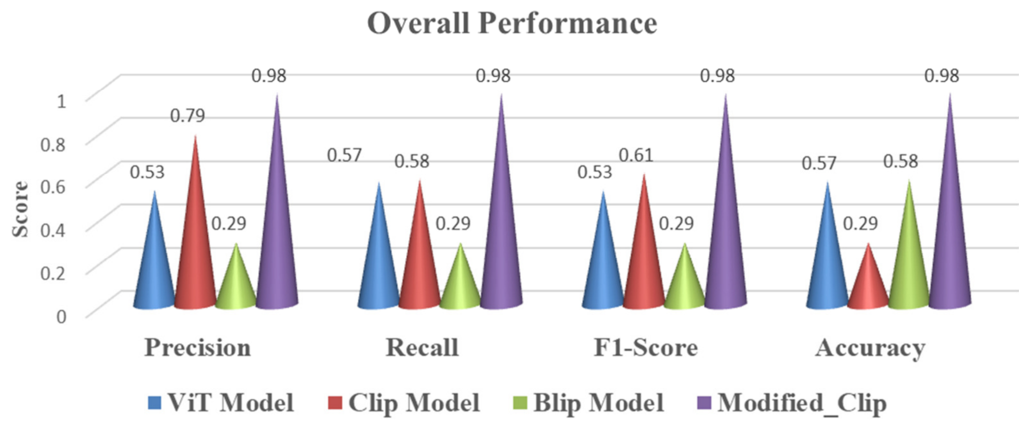


Fig. 11. Overall performance comparison.

TABLE IV. OVERALL PERFORMANCE COMPARISON

Model	Precision	Recall	F1-Score	Accuracy
ViT	0.53	0.57	0.53	0.57
CLIP	0.79	0.58	0.61	0.29
BLIP	0.29	0.29	0.29	0.58
mClip	0.98	0.98	0.98	0.98

VI. CONCLUSION

The proposed framework consists of an OTSU, a Gaussian filter, and a modified Contrastive Language-Image Pre-training (mCLIP) model to detect diseases in leaf images. The proposed model is a novel approach in the multimodal image-text pair classification. To date, the existing methods focus on predicting one output among n outputs. However, multimdal models can predict multiple outputs from a single input. The presented model is a new approach to multi-modeling and is able to show a suitable remedy text for certain diseased leaf images. The proposed approach generates the suitable treatment tests for the given images and also classifies the type of disease found in the leaf images. The introduced mCLIP approach is explicitly designed for multi-output and multi-class classification that integrates images, text, and class labels for richer feature representation. This label embedding layer makes the mCLIP model a fine-grained classification and categorization. Unlike Vision Transformer (ViT) and Bootstrapping Language-Image Pretraining (BLIP), mCLIP utilizes both methods in combination with labels. Due to the embedding layer, the mCLIP provides a stronger supervised learning framework model.

The proposed mCLIP approach has certain limitations. When the number of classes is high, the computational complexity may increase, requiring a suitable optimization technique to address this issue. Therefore, mCLIP must be evaluated for its suitability in VQA systems. Future research directions of the proposed mCLIP may include object classification within images, multi-lingual context understanding with fewshot learning, and developing mCLIP with semi-supervised models.

REFERENCES

[1] J. Dong, Y. Yao, A. Fuentes, Y. Jeong, S. Yoon, and D. S. Park, "Visual information guided multi-modal model for plant disease anomaly detection," *Smart Agricultural Technology*, vol. 9, Dec. 2024, Art. no. 100568, https://doi.org/10.1016/j.atech.2024.100568.

[2] S. Pandey, A. Somaddar, S. Jain, S. Perti, and U. Pathak, "A Multimodal Deep Learning Approach for Advanced Plant Disease Prediction," in *1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0*, Wardha, India, Nov. 2023, vol. 1, pp. 1–6, https://doi.org/10.1109/IDICAIEI58380.2023.10406344.

[3] H. Lee, Y.-S. Park, S. Yang, H. Lee, T.-J. Park, and D. Yeo, "A Deep Learning-Based Crop Disease Diagnosis Method Using Multimodal Mixup Augmentation," *Applied Sciences*, vol. 14, no. 10, Jan. 2024, Art. no. 4322, https://doi.org/10.3390/app14104322.

[4] A. Musa, M. Hassan, M. Hamada, and F. Aliyu, "Low-Power Deep Learning Model for Plant Disease Detection for Smart-Hydroponics Using Knowledge Distillation Techniques," *Journal of Low Power Electronics and Applications*, vol. 12, no. 2, Jun. 2022, Art. no. 24, https://doi.org/10.3390/jlpea12020024.

[5] Md. S. H. Shovon, S. J. Mozumder, O. K. Pal, M. F. Mridha, N. Asai, and J. Shin, "PlantDet: A Robust Multi-Model Ensemble Method Based on Deep Learning For Plant Disease Detection," *IEEE Access*, vol. 11, pp. 34846–34859, Jan. 2023, https://doi.org/10.1109/ACCESS.2023.3264835.

[6] J. Li *et al.*, "A Multi-Modal Open Object Detection Model for Tomato Leaf Diseases with Strong Generalization Performance Using PDC-VLD," *Plant Phenomics*, vol. 6, Aug. 2024, Art. no. 0220, https://doi.org/10.34133/plantphenomics.0220.

[7] Y. Lu *et al.*, "Application of Multimodal Transformer Model in Intelligent Agricultural Disease Detection and Question-Answering Systems," *Plants*, vol. 13, no. 7, Jan. 2024, Art. no. 972, https://doi.org/10.3390/plants13070972.

[8] M. Jung *et al.*, "Construction of deep learning-based disease detection model in plants," *Scientific Reports*, vol. 13, no. 1, May 2023, Art. no. 7331, https://doi.org/10.1038/s41598-023-34549-2.

[9] B. Yang *et al.*, "A novel plant type, leaf disease and severity identification framework using CNN and transformer with multi-label method," *Scientific Reports*, vol. 14, no. 1, May 2024, Art. no. 11664, https://doi.org/10.1038/s41598-024-62452-x.

[10] B. Khan *et al.*, "Bayesian optimized multimodal deep hybrid learning approach for tomato leaf disease classification," *Scientific Reports*, vol. 14, no. 1, Sep. 2024, Art. no. 21525, https://doi.org/10.1038/s41598-024-72237-x.

[11] J. Liu and X. Wang, "A multimodal framework for pepper diseases and pests detection," *Scientific Reports*, vol. 14, no. 1, Nov. 2024, Art. no. 28973, https://doi.org/10.1038/s41598-024-80675-w.

[12] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, "Plant leaf disease detection using computer vision and machine learning algorithms," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 305–310, Jun. 2022, https://doi.org/10.1016/j.gltp.2022.03.016.

[13] Y. M. Abd Algani, O. J. Marquez Caro, L. M. Robladillo Bravo, C. Kaur, M. S. Al Ansari, and B. Kiran Bala, "Leaf disease identification and classification using optimized deep learning," *Measurement: Sensors*, vol. 25, Feb. 2023, Art. no. 100643, https://doi.org/10.1016/j.measen.2022.100643.

[14] A. Abisha and B. Bharathi, "A Hybrid Feature Extraction and Classification using Xception-RF for Multiclass Disease Classification in Plant Leaves," *Applied Artificial Intelligence*, vol. 37, no. 1, Dec. 2023, Art. no. 2176614, https://doi.org/10.1080/08839514.2023.2176614.

[15] A. Dash, P. K. Sethy, S. G. K. Patro, and A. O. Salau, "Deep feature extraction based cascading model for the classification of Fusarium stalk rot and charcoal rot disease in maize plant," *Informatics in Medicine Unlocked*, vol. 42, Jan. 2023, Art. no. 101363, https://doi.org/10.1016/j.imu.2023.101363.

[16] G. Heller *et al.*, "Multisource neural network feature map fusion: An efficient strategy to detect plant diseases," *Intelligent Systems with Applications*, vol. 19, Sep. 2023, Art. no. 200264, https://doi.org/10.1016/j.iswa.2023.200264.

[17] B. Mazumder, M. S. I. Khan, and K. M. Mohi Uddin, "Biorthogonal wavelet based entropy feature extraction for identification of maize leaf diseases," *Journal of Agriculture and Food Research*, vol. 14, Dec. 2023, Art. no. 100756, https://doi.org/10.1016/j.jafr.2023.100756.

[18] M. V. Shewale and R. D. Daruwala, "High performance deep learning architecture for early detection and classification of plant leaf disease," *Journal of Agriculture and Food Research*, vol. 14, Dec. 2023, Art. no. 100675, https://doi.org/10.1016/j.jafr.2023.100675.

[19] G. Dai, J. Fan, Z. Tian, and C. Wang, "PPLC-Net:Neural network-based plant disease identification model supported by weather data augmentation and multi-level attention mechanism," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 5, May 2023, Art. no. 101555, https://doi.org/10.1016/j.jksuci.2023.101555.

[20] S. Pudumalar and S. Muthuramalingam, "Hydra: An ensemble deep learning recognition model for plant diseases," *Journal of Engineering Research*, vol. 12, no. 4, pp. 781–792, Dec. 2024, https://doi.org/10.1016/j.jer.2023.09.033.

[21] A. O. Anim-Ayeko, C. Schillaci, and A. Lipani, "Automatic blight disease detection in potato (*Solanum tuberosum* L.) and tomato (*Solanum lycopersicum*, L. 1753) plants using deep learning," *Smart Agricultural Technology*, vol. 4, Aug. 2023, Art. no. 100178, https://doi.org/10.1016/j.atech.2023.100178.

[22] S. P. Singh, K. Pritamdas, K. J. Devi, and S. D. Devi, "Custom Convolutional Neural Network for Detection and Classification of Rice Plant Diseases," *Procedia Computer Science*, vol. 218, pp. 2026–2040, Jan. 2023, https://doi.org/10.1016/j.procs.2023.01.179.

[23] B. M. Abuhayi and A. A. Mossa, "Coffee disease classification using Convolutional Neural Network based on feature concatenation," *Informatics in Medicine Unlocked*, vol. 39, Jan. 2023, Art. no. 101245, https://doi.org/10.1016/j.imu.2023.101245.

[24] T. Poblete *et al.*, "Detection of symptoms induced by vascular plant pathogens in tree crops using high-resolution satellite data: Modelling and assessment with airborne hyperspectral imagery," *Remote Sensing of Environment*, vol. 295, Sep. 2023, Art. no. 113698, https://doi.org/10.1016/j.rse.2023.113698.

[25] R. Mumtaz, M. H. Maqsood, I. ul Haq, U. Shafi, Z. Mahmood, and M. Mumtaz, "Integrated digital image processing techniques and deep learning approaches for wheat stripe rust disease detection and grading," *Decision Analytics Journal*, vol. 8, Sep. 2023, Art. no. 100305, https://doi.org/10.1016/j.dajour.2023.100305.

[26] H. H. Alshammari and H. Alkhiri, "Optimized recurrent neural network mechanism for olive leaf disease diagnosis based on wavelet transform," *Alexandria Engineering Journal*, vol. 78, pp. 149–161, Sep. 2023, https://doi.org/10.1016/j.aej.2023.07.037.

[27] L. L. Janneh, Y. Zhang, M. Hydara, and Z. Cui, "Deep learning-based hybrid feature selection for the semantic segmentation of crops and weeds," *ICT Express*, vol. 10, no. 1, pp. 118–124, Feb. 2024, https://doi.org/10.1016/j.icte.2023.07.008.

[28] D. R. F. R. Anandhi and S. Sathiamoorthy, "Enhanced Sea Horse Optimization with Deep Learning-based Multimodal Fusion Technique for Rice Plant Disease Segmentation and Classification," *Engineering,*

Technology & Applied Science Research, vol. 13, no. 5, pp. 11959–11964, Oct. 2023, https://doi.org/10.48084/etasr.6324.

[29] R. Rajamohanan and B. C. Latha, "An Optimized YOLO v5 Model for Tomato Leaf Disease Classification with Field Dataset," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12033–12038, Dec. 2023, https://doi.org/10.48084/etasr.6377.

[30] A. Abbas, U. Maqsood, S. U. Rehman, K. Mahmood, T. AlSaedi, and M. Kundi, "An Artificial Intelligence Framework for Disease Detection in Potato Plants," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12628–12635, Feb. 2024, https://doi.org/10.48084/etasr.6456.

[31] A. Gatla *et al.*, "Optimizing Edge AI for Tomato Leaf Disease Identification," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 16061–16068, Aug. 2024, https://doi.org/10.48084/etasr.7802.

[32] S. Vimalkumar and R. Latha, "Maize Leaf Disease Detection using Manta-Ray Foraging Optimization with Deep Learning Model," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17068–17074, Oct. 2024, https://doi.org/10.48084/etasr.7821.

[33] D. Rahadiyan, S. Hartati, Wahyono, and A. P. Nugroho, "Feature aggregation for nutrient deficiency identification in chili based on machine learning," *Artificial Intelligence in Agriculture*, vol. 8, pp. 77–90, Jun. 2023, https://doi.org/10.1016/j.aiia.2023.04.001.

[34] "PlantVillage Dataset." https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset.

AUTHORS PROFILE

Dhanasekaran Menaga received the B.Tech. degree in Information Technology from C. Abdul Hakeem College of Engineering & Technology in 2015 and an M.E. degree in Computer Science and Engineering from C. Abdul Hakeem College of Engineering & Technology in 2019. She is currently pursuing a Ph.D. at VIT University, Vellore, Tamil Nadu, India. Her research interests lie in the domain of Machine Learning.

M. Sudha is currently a Senior Associate Professor in the Department of Information Technology at VIT Vellore. She received her Ph.D. from VIT in 2016. She has more than 22 years of teaching experience. Her research interests span in the domain of Machine Learning, Healthcare Analytics, and Cloud Security. She has explored the presence and implications of self-similarity and analyzed the efficacy of Machine Learning approaches in weather forecast modeling. She holds a patent derived from her research. She has published several research articles, monographs, and books in reputed journals and conferences.