Journal
of Applied
Horticulture
ISSN: 0972-1045

Journal of Applied Horticulture, 27(3): 553-558, 2025

https://doi.org/10.37855/jah.2025.v27i03.100

Deep learning and GCC-MA based visual system for mango leaf disease analysis with anthracnose and sooty mold as case study

Vaibhav Srivastava[1], M. Srivastava[1] and Shailendra Rajan[2]*

*[1]Amity Institute of Information Technology, Amity University (AIIT), Lucknow Campus, Uttar Pradesh, India. [2]ICAR-Central Institute for Subtropical Horticulture, Rehmankhera, Lucknow, India. *E-mail: srajanlko@gmail.com*

Abstract

Mango leaf diseases threaten yield and quality, underscoring the need for early diagnosis and actionable monitoring. Addressing the absence of an integrated Visual Management System (VMS) in prevailing methodologies, this work introduces an enhanced VMS for real-time mango leaf disease analysis that unifies computer vision and language understanding. The framework employs Soft Root Sign–VGG16–Fast Dropout–ResNet50 (SRS-VGG16-FD-ResNet-50) for visual classification and Generalized Canonical Correlation Multimodal Autoencoders (GCC-MA) for cross-modal feature fusion. On the visual side, mango leaf images were preprocessed, followed by leaf color detection, superpixel segmentation, and feature extraction. In parallel, disease-related question–answer (Q&A) data were curated and preprocessed; keywords are extracted, entity–relation graphs were constructed, and textual features were derived. LeCun Bayesian BERT (LeCunBayBERT) generates embeddings for answers, questions, and keywords to enhance semantic representation. GCC-MA was used to fuse the visual and textual to model crossmodal correlations and finally, the disease classification was done by SRS-VGG16-FD-ResNet-50. With testing, the user is authenticated and provides images with queries via the VMS interface, and the system provides the predicted diagnosis as well as the natural language responses. The proposed framework outperforms existing methodologies for mango leaf disease detection, achieving higher accuracy (98.45%) while supporting interpretable, real-time decision support for orchard management. As a case study, we demonstrate the system's effectiveness on anthracnose and sooty mold, illustrating end-to-end detection and interpretation.

Key words: Mango leaf disease, anthracnose, sooty mold, visual management system (VMS), radial kriging + CLAHE, multimodal autoencoders, SRS-VGG16-FD-ResNet-50

Introduction

Mango, one of the significant tropical fruits, containing vitamins, fiber, and antioxidants, serves as a livelihood for millions of people across the globe but is reported to face great losses due to diseases (Rajan, 2021). Plant diseases are still one of the major constraints on global agricultural productivity, causing substantial declines in yield and quality in several crops. Among them, leaf diseases (anthracnose and sooty mold) play an important role and are detrimental, as they can reduce photosynthetic activity, leading to a decline in fruit yield and quality (Arauz, 2000). In order for these diseases to be effectively managed, fast and accurate diagnoses must be made. The phenology of mango is becoming altered by climate change and this is offering pests and diseases a new opportunity to spread. AI can help look for changes and hazards that farmers need to address while managing disease in climate-sensitive fruit crops like mangoes (Srivastava *et al.*, 2024).

Over the last decade, artificial intelligence has emerged as a promising approach for automated disease detection across all stages, including data acquisition, preprocessing, feature extraction, and classification. Conventional machine learning techniques such as Random Forest, Support Vector Machine and Naïve Bayes, and deep learning models like Convolutional Neural Networks, Recurrent Neural Networks and Gated Recurrent Units have been found to be highly promising. Advancements in ensemble learning methods have also increased the prediction accuracy and decreased bias. Yet, few of the current reports recognize the requirement for a complete Visual Management System (VMS) that combines effective image analysis and decision-making support to guide practical implementation on-farms.

Deep learning-based automated disease diagnosis is highly promising, but existing techniques are mainly ineffective in terms of fine-grained dif ferentiation and in handling low-resolution field data. Many existing systems are unable to distinguish between diseases that resemble each other, and this demonstrates the significance of having models with the ability to detect subtle textural and boundary cues (Gautam *et al.*, 2024). Recent ensemble-based models have shown a tendency to degrade when dealing with low-resolution images, which underscores the need to incorporate super-resolution or resolution-invariant attributes to successfully deploy the model in the field (Bezabh *et al.*, 2024).

To mitigate this gap, a study was needed to develop a GCC-MA and SRS-VGG16-FD-ResNet-50 based VMS for predicting mango leaf diseases. Super-resolution to handle low-quality field images, distilled VGG16 representation for fine-grained lesion description and a ResNet-50 classifier with global cross-correlation mixed attention to better understand the context may be adopted as a solution. This approach can help in implementing visual question answering for query-based queries about leaf health, pathogens and environmental conditions. The proposed VMS enhances the management of diseases in mango production through reliable early disease detection and decision support that is comprehensible.

This research was primarily aimed at addressing the gap by providing a single Visual Management System (VMS) that could be used to diagnose and monitor mango leaf diseases in real-time using anthracnose and sooty mold as case study. To realize this objective, the proposed VMS structure (Fig. 1 and Fig. 2) was employed to unite sophisticated computer vision and natural language understanding to obtain a complete system of orchard management.

Materials and methods

Data collection: The experimental workflow began with the acquisition of the Mango Leaf Disease Dataset from Kaggle (2024), which comprises 2,000 RGB images of both healthy and diseased leaves, including specific cases of anthracnose and sooty mold. RGB images of healthy and diseased mango leaves were paired with disease-related Q&A text to enable visual diagnosis and textual interpretation, following multimodal decision-support practices in biomedicine and vision–language systems (Ngiam *et al.*, 2011; Perera *et al.*, 2020).

Image preprocessing: To be able to see the leaves under various lighting conditions, Gaussian Mixture Models were applied to remove the backgrounds. An RKri-CLAHE based on CLAHE was employed to enhance the local contrast and retain structures of low-quality images. Radial kriging interpolation was used to minimize the boundary artifact caused by bilinear interpolation. SSIM and MSE/PSNR were used to check the perceptual quality (Zhang *et al.*, 2015; Singh and Agarwal, 2017; Wang *et al.*, 2004). Disease segmentation: SLIC was adapted with a Charniak-Kantorovich distance (CKSLIC) to enhance the adherence of the boundaries and homogeneity of the pixels, without affecting the performance of SLIC in CIELAB space (Achanta *et al.*, 2012).

Extracting features: CIELAB was used to measure leaf color by a* (green↔red) and b* (blue↔yellow). The supplementary information available through color histograms, shape descriptors, and GLCM texture features was able to characterize the disease (Haralick *et al.*, 1973; Robertson, 1977).

Text preprocessing and representation: Text was cleaned and standardized. NER and relation extraction formed entity–relation graphs linking diseases, symptoms, and agents. A Bayesian-initialized transformer (LeCunBayBERT) was motivated by evidence that informative priors improve stability and calibration in neural networks (Perera *et al.*, 2020; Blundell *et al.*, 2015).

Classification and dense captioning: Features were concatenated as one input to a GCC-MA-based SRS-VGG16-FD-ResNet-50 model for dense labeling, which returned the disease labels and descriptive symptom text. This process is akin to encoder-decoder captioning pipelines featuring a CNN backbone (Wang *et al.*, 2016; Jang, 2018; Hindarto, 2024)

Performance evaluation: Comparisons against ResNet, EfficientNetB0, VGG16, and LSTM used accuracy, recall, F1, sensitivity, specificity, FPR/FNR, and training time. The performance of the system was validated using two distinct categories of metrics. The quality of image enhancement was measured in terms of Mean Squared Error (MSE), Root Mean Squared Error (RMSE), peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) (Wang *et al.*, 2004). Silhouette Score and Jaccard Index, also referred to as Intersection over Union (IoU) were used to assess segmentation accuracy in detecting diseased regions (MathWorks, 2025).

Results

According to the structural design that was presented in the methodology, the proposed Visual Management System (VMS) effectively incorporated multimodal data streams to diagnose mango leaf diseases, including anthracnose and sooty mold. The VMS enabled the end-to-end functionality, including image acquisition and enhancement, semantic interpretation, and disease identification by combining visual and textual branches of a single pipeline (Fig. 1 and 2).

Image processing and preprocessing: The suggested VMS used a formal image preprocessing pipeline that was written in Python. The first improvement step included noise elimination,

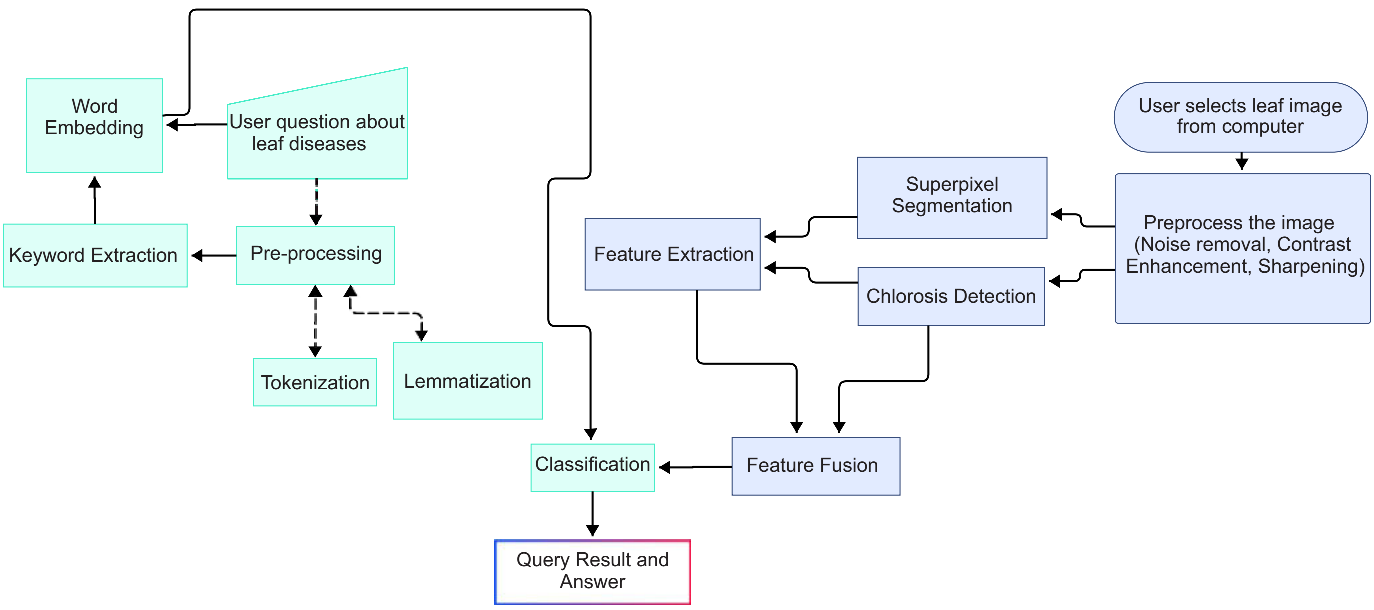

Fig.1. Proposed model for identifying the leaf diseases

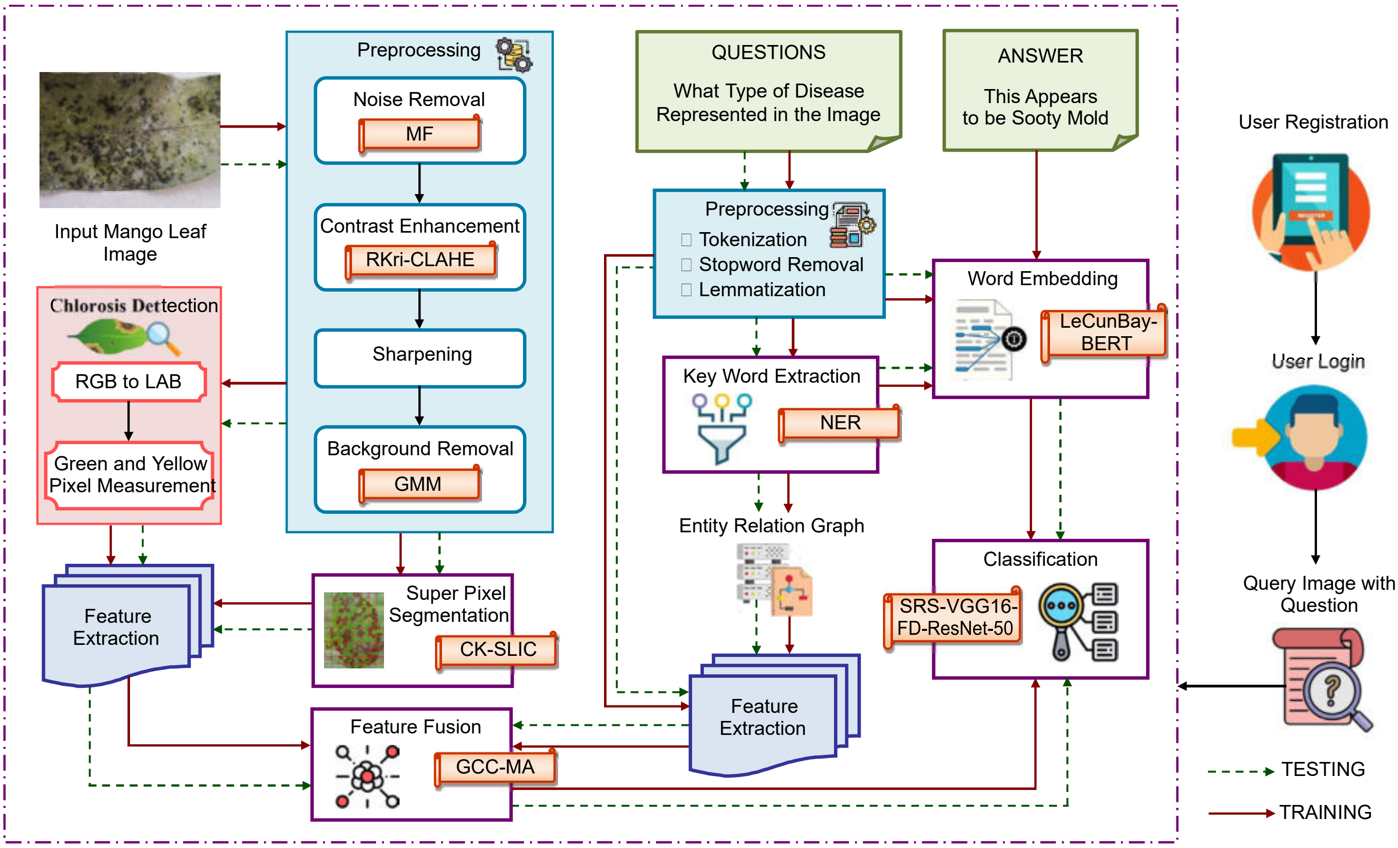

Fig. 2. Structural design of the proposed architecture

contrast, and sharpening (Fig. 3), which guaranteed that the raw images were ready enough to be segmented and classified further. The enhanced level of clarity in lesions enabled the correct identification of minute discolorations typical of anthracnose, which are usually hard to detect in circumstances of natural light. After the process of enhancement, a Gaussian Mixture Model-based classifier was used to remove background pixels and the image data were transformed to LAB color space which is a necessary step of the process of successful color discrimination in diseased regions. The system determined good differentiation between healthy and infected areas by measuring the intensity of green and yellow pixels. To retain essential information for accurate mango leaf disease analysis, the diseased areas were segmented from using the proposed Churnet-Kantorovich Simple Linear Iterative Clustering (CK-SLIC). These procedures contributed to achieving the target diagnostic accuracy and confirmed the preprocessing reliability of the VMS framework.

Description of the dataset and training behaviour: To train and validate a model, the framework used the Mango Leaf Disease Dataset, which can be found on Kaggle (data set link in the reference section). The data set consisted of 2000 high-resolution leaf images, of which about 800 were unique and the rest were enhanced by zooming and rotation of the images. The pictures were used to depict eight classes, two types of disease, anthracnose and sooty mold, and healthy samples. There were 200 images in each of the classes, which allowed equal learning. An 80:20 training-test split provided 1600 and 400 images, respectively, which was sufficient to have diversity in both subsets. The preprocessed and segmented samples (Fig. 3) revealed clear lesion contrast and improved texture homogeneity compared to the original images, validating the preprocessing strategies used.

Performance validation and comparative evaluation of classification efficiency: Classification performance was strengthened by the hybrid SRS-VGG16-FD-ResNet-50 model, which combined a super-resolution module (SRS), feature distillation (FD), and hybrid deep learning backbones. The model's performance was compared with conventional architectures such as ResNet, EfficientNetB0, VGG16, and LSTM.

As shown in Figures 4 and 5, the proposed model achieved an accuracy of 98.45%, precision of 99.18%, recall of 98.11%, F-measure of 98.73%, sensitivity of 98.11%, and specificity of 98.93%, with low FNR (0.086) and FPR (0.08889). Traditional techniques, by contrast, averaged 92.62% accuracy and lower recall and precision scores, indicating that they frequently misclassified subtle or overlapping lesion categories. This performance gap underscores how the combination of FD regularization and SRS significantly improved discriminative power and robustness, particularly under varying light or noise conditions.

The suggested model (Fig. 2) outperformed all explored models—ResNet, EfficientNetB0, VGG16 and LSTM—based on key evaluation metrics. Best performance model is ResNet with the highest accuracy here (96.5%), recall (96.96%) and F1-Score (96.89%). But the proposed model beats those numbers, so it classifies better. It also exhibits that it has better sensitivity, specificity as well as lower FPR and FNR rates, which proves that our model is far more reliable in detecting as well as dismissing the cases accurately.

Training time comparison: The training efficiency (Table 1) further established the practicality of the proposed system. The

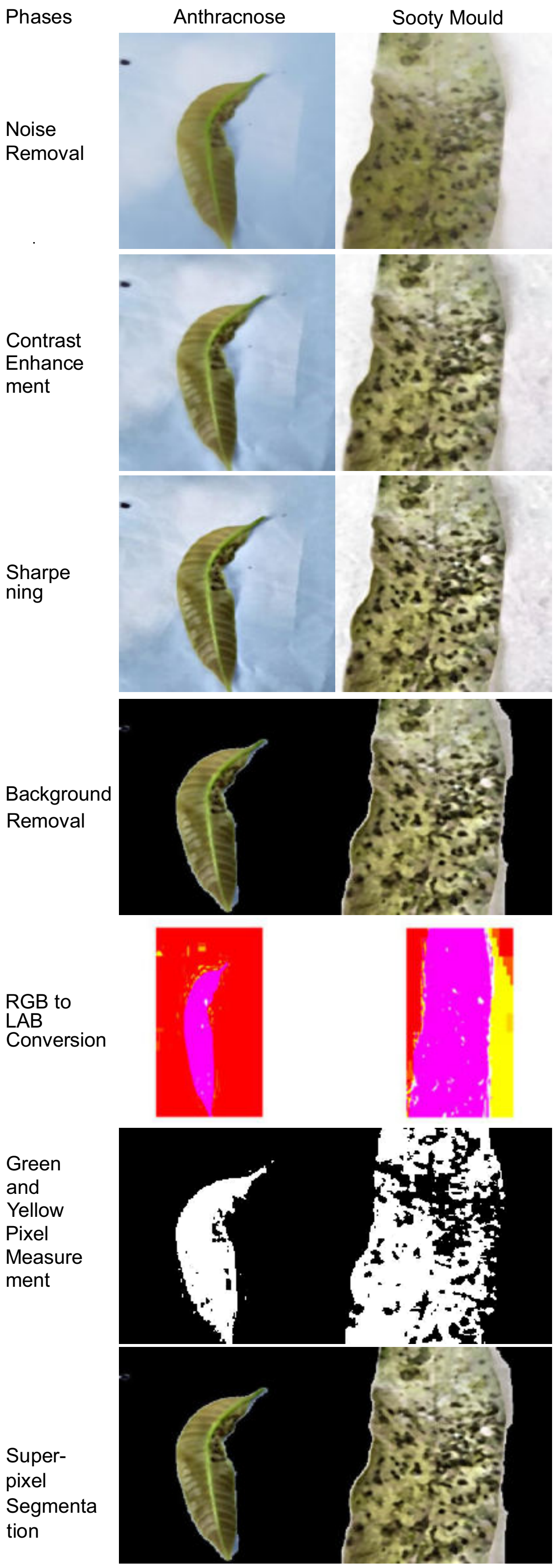

Fig. 3. Workflow of the image processing pipeline for anthracnose and sooty mold detection

SRS-VGG16-FD-ResNet-50 model required 3,672,000 ms of training time—considerably lower than ResNet (3,852,000 ms), EfficientNetB0 (4,140,000 ms), VGG16 (4,536,000 ms), and LSTM (4,951,440 ms). This reduction demonstrates the advantage of streamlined feature distillation, which eliminates redundant backpropagation steps and stabilizes convergence. The use of SRS activation enhanced

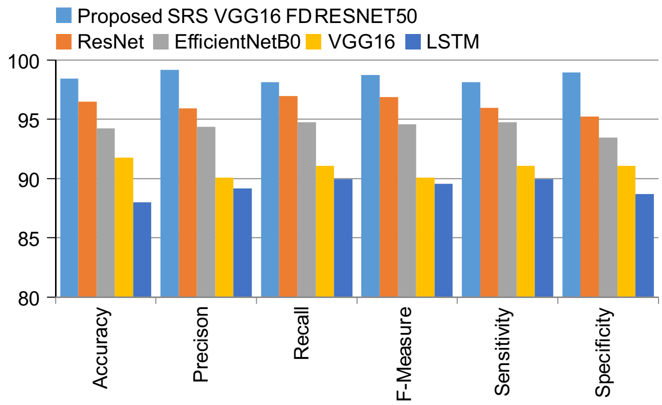

Fig. 4. Comparative analysis of classification performance (Accuracy, Precision, Recall, F-Measure, Sensitivity, and Specificity) between the Proposed SRS VGG16 FDRESNET50 model and existing architectures like ResNet, EfficientNetB0, VGG16, and LSTM

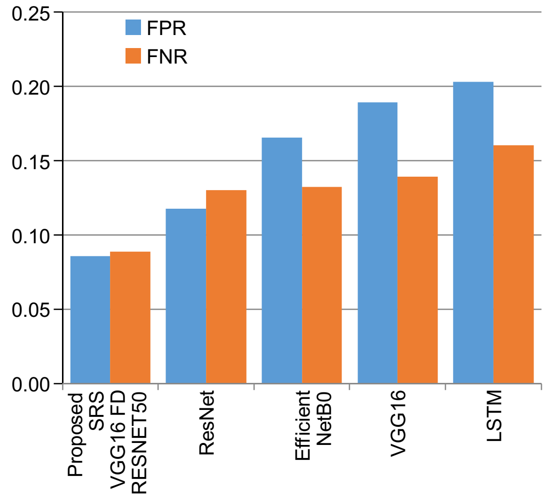

Fig. 5. Comparative evaluation of error rates (FPR and FNR) for the Proposed SRS VGG16 FD RESNET50 model against ResNet, EfficientNetB0, VGG16 and LSTM

Table 1. Training Time Analysis

Methodologies	Training time (ms)
Proposed SRS-VGG16-FD-ResNet-50	3672000
ResNet	3852000
EfficientNetB0	4140000
VGG16	4536000
LSTM	4951440

generalization while mitigating overfitting during training, a key limitation of prior CNN-based disease recognition systems (Bezabh *et al.*, 2024).

Word embedding and semantic understanding: Textual interpretation effectiveness was primarily tested through LeCunBay-BERT, which outperformed BERT, Word2Vec, TF-IDF, and CBOW in transforming natural language queries into vector representations. As indicated in Fig. 6, LeCunBay-BERT obtained a higher precision of 97.12, recall of 97.54, and F-measure of 97.87 which is higher than the rest of the algorithms in maintaining contextual semantics. This is due to improvement of a LeCun-Bayesian (LCB) weight initialisation method that accelerated and reduced generalisation errors. In this way, the framework would be able to match the complex textual queries with the visual evidence appropriately, necessary for multimodal reasoning in VMS. The mango disease classification systems in the past were usually characterized by inadequate text-image correlation, resulting in inconsistent or irrelevant findings (Rizvee *et al.*, 2023). The suggested LeCunBay-BERT module directly overcomes this deficiency by integrating text-query semantics into a domain-informed latent space, aligned to vision feature maps. This has enabled users to get answers based on context and not based on categorical diagnoses.

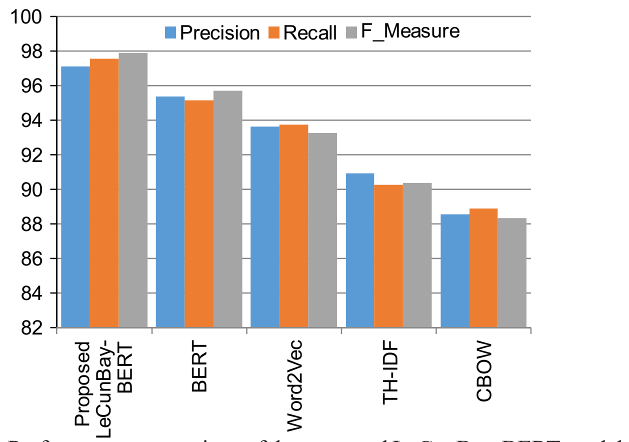

Fig. 6. Performance comparison of the proposed LeCunBay-BERT model with standard NLP techniques (BERT, Word2Vec, TF-IDF, and CBOW) based on Precision, Recall, and F_Measure.

Superpixel segmentation: CK-SLIC segmentation module increased the accuracy of the boundaries of disease regions as compared to the conventional segmentation algorithms, such as SLIC, Linear Spectral Clustering (LSC), Felzenszwalb Segmentation (FS), and SEEDS. As shown in Fig. 7, CK-SLIC had a Silhouette Score of 0.975 and a Jaccard Index of 0.985, both higher than those of any other approach. A local feature discriminatory power was to be improved through the integration of a cosine kernel function, which reduced noise sensitivity. In comparison with the traditional distance-based clustering, CK-SLIC was highly precise even when lesion edges were in contact with the non-diseased tissue. This accuracy is critical in diseases like anthracnose, where necrotic margins blend gradually with green tissue (Gautam *et al.*, 2024).

Contrast enhancement and image quality analysis: The RKri-CLAHE approach yielded substantial improvements in image contrast and structure retention compared to conventional histogram equalization methods such as CLAHE, DPHE, HE, and AHE. As shown in Table 2, RKri-CLAHE achieved a PSNR of 28.4145 dB and SSIM of 0.9095, whereas the next-best method (CLAHE) achieved only 20.4145 dB and 0.8295, respectively. These indicators affirm that RKri-CLAHE enhanced detail visibility without amplifying noise or distortion.

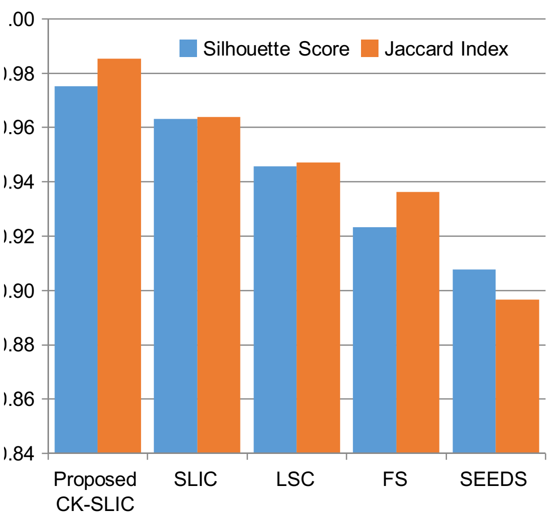

Fig. 7. Segmentation quality comparison using Silhouette Score and Jaccard Index for the Proposed CK-SLIC method against SLIC, LSC, FS, and SEEDS.

Table 2. PSNR and SSIM analysis

Methodologies	PSNR (dB)	SSIM
Proposed RKri-CLAHE	28.4145	0.9095
CLAHE	20.4145	0.8295
DPHE	17.304	0.7383
HE	14.2427	0.6707
AHE	9.998	0.4427

Besides, Figure 8 proves the idea that the proposed improvement algorithm obtained the minimal MSE (5.8879) and RMSE (2.42649) values. These findings demonstrate the importance of how the RK (Rician Kullback-based) regularization reduced the under- and over-enhancement challenges that were prevalent in the previous studies (Alberto *et al.*, 2023). The technique avoided false edges to preserve the morphology of the lesions and venation and enhance the reliability of feature extraction in the future.

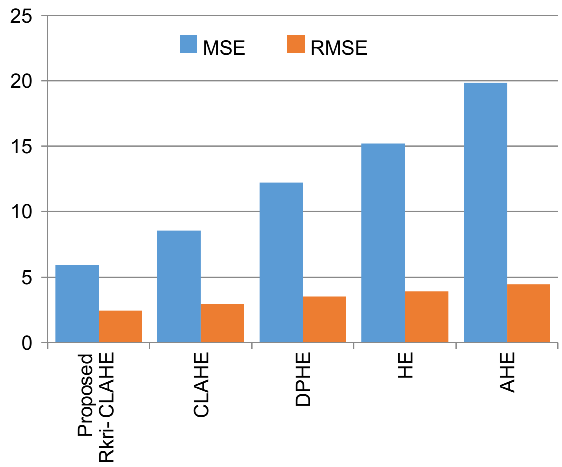

Fig. 8. Analysis of image enhancement performance measuring MSE and RMSE for the Proposed Rkri-CLAHE technique compared to CLAHE, DPHE, HE, and AHE

Discussion

Recent CNN-based mango disease classification systems have a moderate accuracy, but they are not always robust in the field and do not work well in non-uniform light and blurred images (Bezabh *et al.*, 2024). Moreover, the majority of the past architectures do not combine textual analysis, which restricts their applicability in field-deployed decision support environments. The current research shows that a cross-modal framework that involves the use of super-resolution imaging, attention-guided feature extraction, and semantic reasoning is capable of attaining not only high accuracy but also realistic decision reliability. Other more recent systems like the hybrid ensemble+SVM models have achieved praiseworthy accuracy, yet they demand too many computing resources and take too much time to train (Jain and Jaidka, 2023). Likewise, the architectures based on linear discriminant analysis (LDA) offered early anthracnose detection, but was still computationally expensive and not as scalable to multi-disease conditions (Alberto *et al.*, 2023).

The performance of SRS-VGG16-FD-ResNet-50 is thus a significant move towards light but accurate disease classification. The feature distillation mechanism assisted in preserving fine-grained data that is easily lost in the deep structures like the surface textural features and subtle chlorotic areas that signify sooty mold. In addition, the Global Cross-Correlation Mixed Attention (GCC-MA) module that was included enabled long-range cross-channel integration by assisting the system to differentiate between visually similar classes of leaf damage. This difference is particularly significant in the differentiation between fungus and abiotic stress symptoms, conditions that are similar to each other in visual perception but different in etiology. Although the proposed VMS shows a major advancement in the mango disease diagnosis, there are some limitations.

The generic applicability of the current implementation to other significant mango leaf diseases, including powdery mildew, bacterial black spot, or algal leaf spot was limited by its training on only two classes of disease, anthracnose and sooty mold. Also, the fact that the model is based on the idea of quantification of green and yellow pixels creates some ambiguity in the process of distinguishing between chlorosis due to disease and nutrient deficiencies or natural senescence. Other inputs of multispectral or hyperspectral imaging might be used to address this drawback. Furthermore, although the background removal based on the Gaussian Mixture Model performed well in a controlled lighting environment, its functionality under extreme field conditions (*e.g.* strong shadows or non-homogeneous backgrounds) should be validated further. Likewise, RKri-CLAHE can be optimized to broader illumination dynamics.

The VMS provided a full time real time diagnostic pipeline, where the user can be registered, upload the images, enter the questions and the disease response is generated immediately. By so doing, it reflected a primary objective of precision horticulture decision support based on interpretable outputs. It can also be adapted to other types of fruit crops that may have different visual patterns of potential attack by the fungi or bacteria, since the system is designed as an integrated architecture. The visual and textual reasoning as attained by LeCunBay-BERT and SRS-VGG16-FD-ResNet-50 makes this VMS an extremely versatile basis of cross-domain agricultural analytic systems. The final classification accuracy of 98.45%, together with superior PSNR, SSIM, and Jaccard Index metrics, confirms its standing as both highly efficient and interpretable.

In conclusion, the proposed VMS system transcends limitations seen in past vision-only models by successfully incorporating cross-modal analysis and knowledge-driven image interpretation. The development of RKri-CLAHE for adaptive contrast enhancement, CK-SLIC for precise lesion segmentation, and multimodal fusion via LeCunBay-BERT and GCC-MA architecture represents a coherent, field-ready framework that enhances both the accuracy and trustworthiness of mango disease diagnosis. This approach offers a strong foundation for reducing yield loss through early, explainable detection and real-time monitoring—paving the way for more resilient digital orchard management systems.

References

Achanta, R., A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Süsstrunk, 2012. SLIC Superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(11): 2274-2282.

Alberto, L. R., C.E. C. Ardila and F.A. P. Ortiz, 2023. A computer vision system for early detection of anthracnose in sugar mango (Mangifera indica) based on UV-A illumination. *Inf. Process. Agric.*, 10(2): 204-215.

Arauz, L. F., 2000. Mango anthracnose: Economic impact and current options for integrated management. *Plant Dis.*, 84(6): 600-611.

Bezabh, Y. A., A.M. Ayalew, B.M. Abuhayi, T.N. Demlie, E.A. Awoke and T.E. Mengistu, 2024. Classification of mango disease using ensemble convolutional neural network. *Smart Agric. Technol.*, 8: 100476.

Blundell, C., J. Cornebise, K. Kavukcuoglu and D. Wierstra, 2015. Weight Uncertainty in Neural Networks. In *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, (PMLR Vol. 37, pp. 1613-1622).

Gautam, V., R.K. Ranjan, P. Dahiya and A. Kumar, 2024. ESDNN: A novel ensembled stack deep neural network for mango leaf disease classification and detection. *Multimed. Tools Appl.*, 83(4): 1-27.

Haralick, R. M., K. Shanmugam and I. Dinstein, 1973. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.*, SMC3(6): 610-621.

Hindarto, D., 2024. Use of ResNet50 neural network in diagnosing diseases mango leaves. *J. Comput. Netw. Archit. High Perform. Comput.*, 6(1): 220-230.

Jain, S. and P. Jaidka, 2023. Mango leaf disease classification using deep learning hybrid model. In *2023 Int. Conf. Power Instrum. Energy Control (PIECON)*, (pp. 1-7). IEEE.

Jang, H. 2018. Image Captioning Using Deep Learning. https://hankyujang.github.io/files/course_project/2018_05_ImageCaptioningUsingDeepLearning.pdf

Kaggle, 2024. Kaggle Dataset, https://www.kaggle.com/datasets/aryashah2k/mango-leaf-disease-dataset

MathWorks, 2025. Jaccard Similarity Coefficient for Image Segmentation. https://www.mathworks.com/help/images/ref/jaccard.html

Ngiam, J., A. Khosla, M. Kim, J. Nam, H. Lee and A.Y. Ng, 2011. Multimodal Deep Learning. In *Proc. 28th Int. Conf. Mach. Learn. (ICML)*.

Perera, N., D. Bollegala and F. Coenen, 2020. Named Entity Recognition and Relation Detection for Biomedical Information Extraction. *PeerJ Comput. Sci.*, 6: e190.

Rajan, S. 2021. Mango: The King of Fruits. In: *The Mango Genome*, (pp. 1-12). Springer, Cham.

Rizvee, R. A., T.H. Orpa, A. Ahnaf, M.A. Kabir, M.R. Ahmmad Rashid, M.M. Islam, M. Islam, T. Jabid and M.S. Ali, 2023. LeafNet: A proficient convolutional neural network for detecting seven prominent mango leaf diseases. *J. Agric. Food Res.*, 14: 100787.

Robertson, A.R. 1977. The CIE 1976 color-difference formulae. *Color Research Application,* 2(1): 7-11.

Singh, A. and R. Agarwal, 2017. Efficient Medical Image Enhancement Using CLAHE Technique. *Int. J. Comput. Appl.*, 167(5): 22-26.

Srivastava, V., M. Srivastava, S. Rajan, P. Sagar, P.K. Mishra and V.K. Singh, 2024. Leveraging a web-based tool for phenological data management in climate-sensitive fruit crops like mango. *J. Appl. Hortic.*, 26(1): 112-116.

Wang, W., X. Yan, H. Lee and K. Livescu, 2016. Deep Variational Canonical Correlation Analysis. *arXiv preprint*.

Wang, Z., A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4): 600-612.

Zhang, Y., Z. Li, C. Liu and Y. Guo, 2015. Image background reconstruction by Gaussian mixture based model. *Proc. IMechE Part H: J. Eng. Med.*, 229(2): 124-135.

Received: April, 2025 ; Revised: May, 2025 ; Accepted: June, 2025