International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Interpretable mango leaf disease detection using a hybrid CNN–transformer model with GLCM features

Farhan Bin Jashim [1], Fajle Rabbi Refat [1], Mohammad Hasnatul Karim [1], Farhad Uddin Mahmud [2] and Md Ismail Hossain Siddiqui [3, *]

[1] *Department of Business Administration and Management, International American University, CA 90010, USA.*
[2] *Department of Business Administration in Management Information System, International American University, CA 90010, USA.*
[3] *Department of Engineering/Industrial Management, Westcliff University, Irvine, CA 92614, USA.*

International Journal of Science and Research Archive, 2025, 15(02), 1518–1535

Publication history: Received on 08 April 2025; revised on 27 May 2025; accepted on 29 May 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.15.2.1510

Abstract

Mango leaf diseases significantly hinder crop yield and food security in tropical regions, necessitating accurate and timely diagnostic tools. Traditional visual inspection methods are often subjective, time-consuming, and lack scalability, while existing deep learning approaches struggle with dataset imbalance, generalization limitations, and interpretability issues. To address these challenges, we propose ViX-MangoEFormer, a hybrid model that combines convolutional layers with self-attention mechanisms for robust classification of eight mango leaf conditions. The architecture incorporates MBConv4D and MBConv3D modules to capture both localized textures and global patterns, while GLCM-based statistical features enhance discriminatory power. Additionally, a stacking ensemble (MangoNet-Stack), comprising five pretrained models, is introduced as a comparative benchmark. Both models are trained and validated on a merged dataset of 25,530 images from four public sources, including balanced and imbalanced classes. Grad-CAM-based explainability is natively integrated to offer real-time visual rationales. Experimental results demonstrate that ViX-MangoEFormer achieves an F1 score of 99.78% and MCC of 99.34%, outperforming all baseline models. Furthermore, cross-domain tests reveal strong generalization to morphologically similar crops. A web application has been deployed to deliver real-time predictions with transparent explanations, providing an effective and interpretable solution for precision agriculture.

Keywords: Mango Leaf Disease; Vision Transformer; Explainable AI; Diagnostic Tools; Agricultural Monitoring

1. Introduction

Mango (Mangifera indica) is a globally significant tropical fruit, generating over USD 30 billion annually and sustaining the livelihoods of more than 20 million farmers worldwide [1]. Despite its economic and nutritional importance, mango cultivation is vulnerable to a range of foliar diseases, such as Anthracnose, Powdery Mildew, and Bacterial Canker. These conditions can lead to crop losses of up to 30%, with countries like India incurring annual losses exceeding USD 100 million [2]. On a global scale, plant diseases result in agricultural losses surpassing USD 5 billion per year. In addition to economic setbacks, these issues also threaten food security, particularly in regions where mangoes constitute a key dietary component and income source.

Traditional disease detection methods rely heavily on manual visual inspections by agricultural experts. However, such assessments are inherently subjective and inconsistent, particularly during early infection stages. Diagnostic accuracy can drop below 65%, and a single expert may take up to 30 minutes to examine a tree, limiting daily inspection capacity

* Corresponding author: Md Ismail Hossain Siddiqui.

to roughly 20 trees [3]. Moreover, limited access to trained personnel, especially in rural areas—further exacerbates the problem, contributing to undetected infections and progressive yield decline.

In response, computer vision and deep learning (DL) techniques have been widely explored to automate disease identification. Convolutional Neural Networks (CNNs) and transfer learning approaches have demonstrated high classification accuracy [4]. However, CNN-based models demand extensive labeled datasets and computational resources, which are often unavailable in agricultural contexts. Furthermore, these models typically operate as "black-box" systems, offering limited interpretability, which restricts their acceptance in critical decision-making scenarios.

While ensemble learning methods have been employed to boost performance by aggregating multiple models, they often introduce additional computational overhead [5]. Vision Transformers (ViTs) have emerged as a powerful alternative, offering the ability to model global dependencies and subtle variations in visual features [6]. Yet, their high memory footprint and data-hungry nature hinder widespread adoption in real-time agricultural systems.

Another persistent issue is class imbalance in mango disease datasets. Many existing datasets are either small, region-specific, or synthetically augmented, leading to poor generalization across diverse agroecological conditions[7]. Moreover, most current solutions apply explainability techniques post hoc, limiting their ability to provide immediate, interpretable feedback during inference [8].

To overcome these limitations, we propose a novel deep learning framework called ViX-MangoEFormer (Figure 1). This model integrates the strengths of convolutional and transformer-based architectures using MBConv4D and MBConv3D modules, allowing it to effectively capture both local textures and global context [9]. In parallel, we introduce MangoNet-Stack, a stacking ensemble of five pretrained models, to benchmark performance. Both models incorporate handcrafted texture features derived from the Gray-Level Co-occurrence Matrix (GLCM) and natively embed Grad-CAM to generate real-time visual explanations. The system is deployed as a web-based application, supporting real-time, interpretable diagnosis of eight mango leaf conditions across four diverse datasets.

The key contributions of this study include

- A hybrid vision transformer model that combines MBConv-based convolutional layers and multi-head self-attention for enhanced feature representation.
- A stacking ensemble model integrating diverse pretrained networks for robust benchmarking.
- Integration of handcrafted GLCM texture features with deep-learned embeddings for improved class discrimination.
- Native incorporation of Grad-CAM within a web application to enable transparent and accessible disease prediction.
- Comprehensive evaluation on balanced, imbalanced, and cross-domain datasets, demonstrating generalization and scalability.

The rest of this paper is organized as follows. Section 2 reviews related works on mango leaf disease detection using transfer learning, ensemble learning, Vision Transformers, and explainable AI. Section 3 details the proposed methodology, including dataset preparation, preprocessing, feature extraction, model architectures, and training strategies. Section 4 presents the experimental setup and evaluates model performance across various datasets. Section 5 discusses key findings, model interpretability, limitations, and real-world applicability. Finally, Section 6 concludes the paper and outlines future research directions.

2. Related Work

Deep learning has become central to plant disease classification, particularly in scenarios where early detection is critical. Transfer learning with pretrained CNNs has been widely adopted due to its ability to generalize from limited data. Varma et al. [10] and Swapno et al. [11] utilized InceptionV3 and EfficientNetV2L on a dataset of 4000 mango leaf images, achieving classification accuracies of 99.87% and 96.87%, respectively. Similarly, Rizvee et al. [12] proposed LeafNet on MangoLeafBD with 98.55% accuracy, and Ghosh et al. [13] employed EfficientNet-B0 to attain 99.91% accuracy. However, these studies are largely constrained to region-specific or synthetically augmented datasets, resulting in poor cross-domain generalization [14], [15]. To address overfitting and improve robustness, ensemble learning has been explored. Bezabh et al. [16] combined GoogLeNet, VGG16, and Mask R-CNN, achieving 99.21% accuracy. Gautam et al. [17] proposed an ensemble stacked deep neural network using logistic regression to combine CNN outputs, reaching 98.57% accuracy. Seetha et al. [18] further improved performance by integrating handcrafted

features (LBP, LDP, LOOP) with neural networks optimized via chaotic grey wolf optimization. Ali et al. [19] achieved 99.89% accuracy using an ensemble of EfficientNetB0/B3, DenseNet201, and InceptionResNetV2. Despite their effectiveness, these methods often involve significant computational complexity, posing challenges for real-time applications.

Vision Transformers (ViTs) have demonstrated promising capabilities in capturing global feature dependencies. Islam et al. [20] introduced WaveVisionNet, integrating attention and dilated convolutions, achieving 96.11% accuracy. Salamai et al. [21] proposed a Visual Modulation Network that achieved 99.23% accuracy using patch embeddings and modulator blocks. Ramadan et al. [22] combined ViT-B/16 with CNNs via CycleGAN-based augmentation for binary classification, achieving perfect accuracy. Rani et al. [23] applied ConvNeXtXLarge to a large dataset, surpassing 99% accuracy across pest/pathogen classes. However, ViTs often demand large-scale datasets and high computational resources, limiting their practicality in field settings.

Interpretability remains essential for deploying AI in agriculture. Patel et al. [24] used Grad-CAM within a DenseNet121-VGG19 hybrid model, while Raval et al. [25] incorporated LIME with MobileNetV3 and EfficientNetV2B3, achieving 99.89% accuracy. Chang et al. [26] developed Edge Feature Guidance (EFG) modules with Swin Transformer to improve transparency. Yet, these approaches typically apply XAI techniques post hoc, lacking real-time feedback during inference. Deployment-focused systems are gaining attention. Hossain et al. [27] optimized DeiT for mobile deployment with 99.75% accuracy. Tools like DeepCrop LEAFLE, and MVGG1 offer browser-based diagnostics but are often limited by dataset quality, narrow class scopes, and lack of interpretability [28].

Existing research is hindered by dataset imbalance, high model complexity, and limited explainability. Our proposed ViX-MangoEFormer addresses these challenges through a hybrid convolution-transformer model enhanced with GLCM features and integrated Grad-CAM visualization. Additionally, the MangoNet-Stack ensemble serves as a strong benchmark. Both models are deployed in a real-time web interface, offering practical and interpretable solutions for mango disease recognition.

3. Methodology

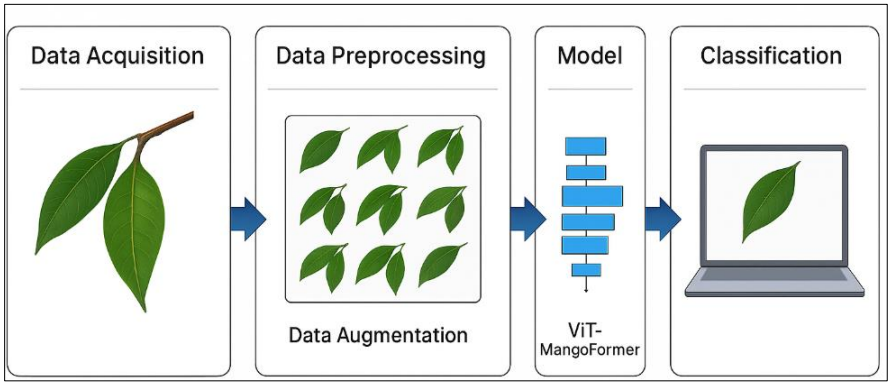

Figure 1 Proposed methodology

3.1. Data Description

To develop and evaluate the proposed mango leaf disease classification models, four datasets were used—three publicly available and one merged. Each dataset contains images categorized into eight classes: seven disease types and one healthy class, with varying distributions and collection environments [29]. The MangoLeafBD dataset (D1) [30] includes 4000 images obtained from 1800 unique samples, augmented through zooming and rotation. Images were collected from four mango orchards in Bangladesh using mobile devices under controlled lighting conditions. All samples were manually labeled by agricultural experts. The Mango Leaf Disease Dataset (D2) [31] comprises 12,730 images, derived from 2336 field-collected samples and further expanded through augmentation. Images span eight classes and were captured in natural environments across Pabna and Savar, Bangladesh. The MLD24 dataset (D3) [32] offers a balanced distribution of 6400 images, with 800 samples per class. Captured in Kushtia using an iPhone SE, images were resized to 240×240 pixels for uniformity and computational efficiency. To increase dataset diversity and address class imbalance, a unified dataset (D4) was created by merging D1, D2, and D3. It comprises 25,530 images, with varied disease severities, lighting conditions, and collection locations (Dhaka, Kushtia, Rajbari, Nilphamari, Savar, and Pabna). The class distribution in D4 includes 3834 images of Bacterial Canker, 3533 of Gall Midge, 3049 of Anthracnose, 2883

of Cutting Weevil, 2625 of Sooty Mould, 2580 of Die Back, 2550 of Healthy, and 2076 of Powdery Mildew. This comprehensive dataset enables the development of models that generalize well across diverse agroecological contexts.

3.2. Data Preprocessing

To ensure consistent input quality and enhance model performance, several preprocessing steps were applied: resizing, noise reduction, normalization, and data augmentation [33]. Each technique was selected to improve generalization, reduce bias, and accommodate environmental variability common in agricultural imaging scenarios [34].

3.2.1. Resizing

All images were resized to 224×224 pixels using bilinear interpolation, which computes pixel intensity at a given location using the weighted average of the four nearest pixel values [35], [36]. Let $I(x, y)$ denote the intensity at position (x, y); the resized intensity $I'(x', y')$ is computed as Eq. 1, where w_{ij} are the interpolation weights and (x_i, y_j) are the neighboring pixel coordinates. The selected size ensures compatibility with pretrained models and balances detail retention with memory efficiency.

$$I'(x', y') = \sum_{i=1}^{2} \sum_{j=1}^{2} w_{ij} \cdot I(x_i, y_j) \tag{1}$$

3.2.2. Noise Reduction

To suppress background artifacts and preserve critical leaf features, anisotropic diffusion filtering was employed [37], [38]. This method preserves edges while smoothing uniform regions and is governed by Eq. 2.

$$\frac{\partial I}{\partial t} = \nabla \cdot (c(x, y, t) \cdot \nabla I) \tag{2}$$

Here, ∇ is the gradient operator, $c(x, y, t)$ is the diffusion coefficient, and t is the iteration index. This edge-aware smoothing technique is more suitable than Gaussian or median filtering for plant disease classification.

3.2.3. Normalization

To ensure stable training and faster convergence, min-max normalization was applied to scale pixel intensity values to the range ([0,1]) [39]. This technique prevents gradients from becoming unstable during backpropagation and ensures that all input features contribute proportionally to the learning process [40]. Given an image ($I(x, y)$), where (x) and (y) represent pixel coordinates, the normalized intensity ($I'(x, y)$) is computed using:

$$I'(x, y) = \frac{I(x, y) - I_{\min}}{I_{\max} - I_{\min}} \tag{3}$$

3.2.4. Data Augmentation

To enhance model generalization and address class imbalance, various augmentation techniques were applied (See Table 1). These transformations simulate real-world variability in leaf orientation, lighting, and scale, ensuring the model can handle diverse conditions during inference. Random rotation within [−45, +45] was used to introduce orientation invariance, while horizontal and vertical flipping accounted for different viewpoints [40]. Brightness adjustment, with scaling factors between 0.8 and 1.2, helped the model adapt to varying illumination conditions. Scaling transformations [0.5,1.5] introduced robustness to leaf size and camera distance, and Gaussian noise with zero mean and variance of 0.01 simulated sensor imperfections. These augmentations were selectively applied across all datasets to balance the training distribution [37]. Post-augmentation, class sizes were expanded to 1000 in D1, 2500 in D2, 1500 in D3, and 4000 in D4. This improved data diversity, reduced model overfitting, and ensured fair representation of minority classes during training [41].

Table 1 Augmentation Parameters

Technique	Parameter	Value
Rotation	Angle (θ)	$[-45°], +45°]$
Flipping	Direction	Horizontal, Vertical
Brightness	Scale Factor (β)	[0.8, 1.2]
Scaling	Factors (s_x, s_y)	[0.5, 1.5]
Gaussian Noise	μ , σ^2	$\mu = 0, \sigma^2 = 0.01$

3.3. GLCM-Based Texture Feature Extraction

In plant disease classification, early symptoms often manifest as subtle textural variations, such as small lesions, mottling, or surface roughness [35]. These fine-grained patterns are critical for distinguishing visually similar disease classes but may not be effectively captured by deep features alone [33], [38]. To address this, we incorporate handcrafted statistical descriptors using the Gray-Level Co-occurrence Matrix (GLCM), a widely used method for quantifying texture based on the spatial relationship between pixel intensities. The GLCM ($P(i,j \mid d,\theta)$) is computed by counting how often pairs of pixel values (i) and (j) occur at a specified offset (d) and direction (θ) as Eq. 4, where (I(x, y)) is the pixel intensity at position ((x, y)), and ((Δx, Δy)) defines the relative position determined by (d) and (θ).

$$P(i,j \mid d,\theta) = \sum_{x=1}^{M}\sum_{y=1}^{N} \delta(I(x,y) = i \wedge I(x+\Delta x, y+\Delta y) = j) \tag{4}$$

From the normalized GLCM, four second-order texture features are extracted using Eq. (5-8). Contrast measures intensity variation, Energy evaluates texture uniformity, Homogeneity assesses proximity to the diagonal, and Correlation quantifies linear dependencies between pixel pairs, where ($\mu_x, \mu_y, \sigma_x, \sigma_y$) are the means and standard deviations of the marginal distributions of(p(i, j)).

$$\text{Contrast} = \sum_{i,j} |i-j|^2 \cdot p(i,j) \tag{5}$$

$$\text{Energy} = \sum_{i,j} p(i,j)^2 \tag{6}$$

$$\text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \tag{7}$$

$$\text{Correlation} = \sum_{i,j} \frac{(i-\mu_x)(j-\mu_y)\cdot p(i,j)}{\sigma_x \sigma_y} \tag{8}$$

These handcrafted features are concatenated with the deep features extracted by the ViX-MangoEFormer and MangoNet-Stack models during feature fusion [33]. This hybrid representation combines domain-specific interpretability with the abstract discriminative power of deep networks. The use of GLCM improves robustness to lighting variation and enhances sensitivity to texture patterns—critical for accurately distinguishing between morphologically similar leaf diseases.

3.4. Model Training Execution Process

The training pipeline consists of two parallel strategies: individual deep learning models and a stacked ensemble architecture called MangoNet-Stack. For the individual models—EfficientNetB8, and CapsuleNet—each network was initialized with pretrained weights from ImageNet and fine-tuned on the mango leaf datasets. These models were selected for their complementary strengths in feature extraction: EfficientNetB8 balances depth and resolution efficiently, and CapsuleNet captures spatial hierarchies using dynamic routes [29]. The outputs from each model were subjected to global pooling to reduce dimensionality while retaining salient features. The MangoNet-Stack ensemble model operates in two stages. In Level 0, all base models process the same input image and produce feature maps. These

features are globally pooled and concatenated into a unified feature vector. In Level 1, this combined feature vector is passed to a Flatten layer that acts as the meta-learner, enabling end-to-end training without the need for an external classifier such as logistic regression or SVM. This architecture enables the ensemble to learn complementary representations across different model families. Categorical cross-entropy was used as the loss function, and optimization was performed using the AdamW optimizer with early stopping and model checkpointing [28], [36]. To avoid overfitting, regularization techniques such as dropout and weight decay were applied. All models were trained on a workstation with an NVIDIA RTX 3090 GPU (24 GB VRAM), ensuring sufficient computational capacity to support ensemble-level training and evaluation.

3.5. Proposed Model Architecture: ViX-MangoEFormer

To address the limitations of conventional CNNs and transformer-based models in agricultural disease detection, we propose ViX-MangoEFormer, a hybrid architecture that integrates convolutional efficiency with global context modeling. This model is specifically designed to capture both local lesion textures and long-range dependencies found in mango leaf disease patterns [9]. The architecture begins with a Convolutional Stem that downsamples the input image of size ($B \times 3 \times H \times W$), where (B) is the batch size and ($H \times W$) denotes the image resolution (Figure 2). This stem reduces spatial dimensions while preserving essential low-level features, preparing the input for hierarchical feature extraction.

The first two stages utilize MBConv4D blocks, a variant of mobile inverted bottleneck convolutions, to extract localized features efficiently. These stages progressively reduce the spatial dimensions to ($H/4$) and ($H/8$), while increasing the number of channels to (C_1) and (C_2) , respectively. Each MBConv4D block includes depthwise separable convolutions followed by pointwise convolutions, batch normalization, and GeLU activation to maintain non-linearity and stability. In Stage 3 and Stage 4, the model transitions to MBConv3D blocks with Multi-Head Self-Attention (MHSA). These layers are responsible for capturing global interactions between distant image regions. Attention maps are computed using query (Q), key (K), and value (V) projections, followed by scaled dot-product attention. This mechanism enables the model to identify spatially distributed disease patterns that may not be visible through local operations alone. Outputs from each attention block are normalized and passed through linear projection layers to refine global embeddings. The final feature maps are projected via fully connected layers to the classification space, corresponding to the eight mango leaf disease classes. By combining MBConv4D (local focus) and MHSA-enabled MBConv3D (global reasoning), ViX-MangoEFormer effectively balances precision and efficiency. This design makes it suitable for deployment in real-time agricultural systems with limited computational resources. The model's robustness is further enhanced through the integration of GLCM-based texture features, which are fused with learned deep features at the bottleneck layer. This fusion improves sensitivity to fine-grained lesion characteristics. To facilitate model interpretability, Grad-CAM is natively embedded into the architecture, enabling real-time visual explanations of classification outcomes—crucial for gaining user trust in practical settings.

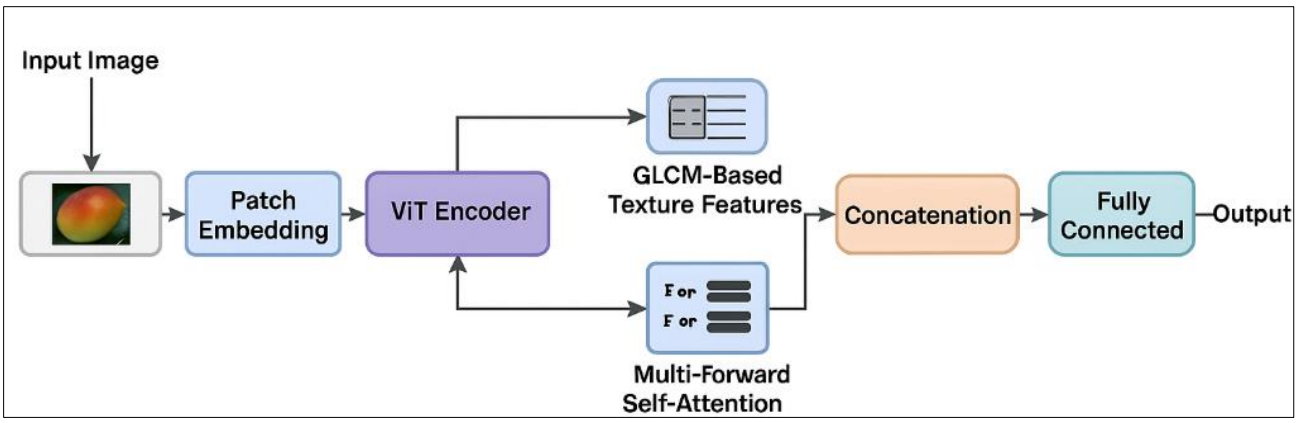

Figure 2 Proposed Model Architecture

3.6. Training Parameters

To optimize model performance and ensure stable convergence, all experiments were conducted using carefully selected hyperparameters. The training was implemented in Python 3.10 using PyTorch with CUDA 11.7 support, and executed on a high-performance workstation equipped with an NVIDIA RTX 3090 GPU (24 GB VRAM), Intel Core i9 processor, and 64 GB RAM. The training process was carried out for 50 epochs, selected from a candidate set {20, 30, 50, 100} based on early convergence patterns observed during initial trials [8]. To prevent overfitting, multiple

regularization strategies were employed, including dropout layers, weight decay, and early stopping, which terminated training if validation loss failed to improve. The AdamW optimizer was used due to its ability to combine adaptive learning rate adjustment with decoupled weight decay, offering better generalization than Adam or SGD in transformer-based and CNN models. The initial learning rate was set to 0.001, providing a stable balance between convergence speed and performance. Although schedulers such as ReduceLROnPlateau were evaluated, they were excluded as no significant performance gains were observed. A batch size of 10 was chosen empirically from the set {8, 10, 16, 32}, balancing memory usage and gradient estimation stability.

3.7. Evaluation Metrics

To comprehensively evaluate the performance of the proposed models across multiclass and imbalanced datasets, four core metrics were utilized: F1 Score, Precision–Recall Area Under the Curve (PR AUC), Matthews Correlation Coefficient (MCC), and Specificity (Eq. 9-13). Each metric was calculated using micro-averaging, which aggregates true positives, false positives, false negatives, and true negatives across all classes, offering a global evaluation that is less sensitive to class imbalance. The F1 Score is the harmonic means of precision and recall, balancing false positives and false negatives [42].

$$Precision_{\text{micro}} = \frac{\sum_c \text{TP}_c}{\sum_c (\text{TP}_c + \text{FP}_c)} \tag{9}$$

$$\text{Recall}_{\text{micro}} = \frac{\sum_c \text{TP}_c}{\sum_c (\text{TP}_c + \text{FN}_c)} \tag{10}$$

$$\text{F1}_{\text{micro}} = \frac{2 \cdot \text{Precision}_{\text{micro}} \cdot \text{Recall}_{\text{micro}}}{\text{Precision}_{\text{micro}} + \text{Recall}_{\text{micro}}} \tag{11}$$

The PR AUC was computed using a one-vs-rest strategy per class and aggregated with micro-averaging. This metric is essential for class-imbalanced settings, indicating the model's precision across varying recall thresholds. Matthews Correlation Coefficient (MCC) provides a balanced assessment of classification quality by incorporating all elements of the confusion matrix.

$$\text{MCC}_{\text{micro}} = \frac{\text{TP} \cdot \text{TN} - \text{FP} \cdot \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \tag{12}$$

Specificity or the true negative rate measures the model's ability to correctly identify negative samples:

$$\text{Specificity}_{\text{micro}} = \frac{\sum_c \text{TN}_c}{\sum_c (\text{TN}_c + \text{FP}_c)} \tag{13}$$

4. Results analysis

4.1. Performance Comparison of Experimental Models

Table 2 compares the performance of four models on augmented datasets D1 and D4 using F1 Score, MCC, PR AUC, and Specificity. ViX-MangoEFormer outperforms all baselines, achieving the highest scores across all metrics, with F1 scores above 99.6% and MCC exceeding 99%. Its strong performance reflects the effectiveness of combining MBConv layers, attention mechanisms, and GLCM features for capturing both local and global patterns. MangoNet-Stack follows closely, maintaining high scores across datasets, particularly in F1 and PR AUC. Its ensemble strategy effectively integrates diverse feature representations, enhancing generalization. However, it shows slightly reduced MCC and specificity compared to ViX-MangoEFormer. EfficientNetB8 and CapsuleNet perform moderately well on D1 but degrade in D4, highlighting limited robustness to dataset diversity. Their lower specificity and MCC indicate challenges in managing class imbalance and complex features. ViX-MangoEFormer demonstrates superior accuracy and generalization, making it well-suited for real-world deployment. MangoNet-Stack remains a strong benchmark, while single CNNs show limitations under varied conditions.

Table 2 Performance on augmented datasets (D1 and D4)

Model	F1 Score (%)	MCC (%)	PR AUC (%)	Specificity (%)
Dataset D1				
ViX-MangoEFormer	99.78	99.34	99.92	98.77
MangoNet-Stack	99.21	98.96	99.11	98.11
EfficientNetB8	98.47	97.8	97.99	97.12
CapsuleNet	98.3	97.22	97.88	96.93
Dataset D4				
ViX-MangoEFormer	99.62	99.12	99.87	98.33
MangoNet-Stack	99.05	98.75	98.94	97.92
CapsuleNet	98.13	97.02	97.91	96.24
EfficientNetB8	98.09	96.81	97.3	95.88

Table 3 presents model performance on augmented datasets D2 and D3. On D2, ViX-MangoEFormer achieves the highest F1 Score (99.13%) and MCC (97.02%), showing strong generalization without augmentation. CapsuleNet closely follows, with slightly better specificity (99.61%) but lower MCC. MangoNet-Stack performs moderately, while EfficientNetB8 records the lowest scores, indicating limited robustness in imbalanced, unaugmented conditions. In D3, MangoNet-Stack outperforms others with an F1 Score of 98.87% and MCC of 96.98%, while ViX-MangoEFormer maintains strong results. CapsuleNet shows decent performance but lower specificity (93.42%), and EfficientNetB8 again underperforms. ViX-MangoEFormer remains the most consistent across datasets, while MangoNet-Stack shows strength on more balanced data. These results highlight the importance of hybrid and ensemble designs in scenarios lacking augmentation.

Table 3 Performance on augmented datasets (D2 and D3)

Model	F1 Score (%)	MCC (%)	PR AUC (%)	Specificity (%)
Dataset D2				
ViX-MangoEFormer	99.13	97.02	99.17	99.32
CapsuleNet	98.84	96.45	99.71	99.61
MangoNet-Stack	97.21	95.8	97.25	94.22
EfficientNetB8	96.17	93.79	95.04	93.67
Dataset D3				
ViX-MangoEFormer	98.87	96.98	98.12	97.49
MangoNet-Stack	97.59	96.6	97.8	95.11
CapsuleNet	96.02	94.83	96.09	93.42
EfficientNetB8	95.7	93.81	95.71	92.71

Table 4 summarizes the performance of four models on the non-augmented D2 and D3 datasets. On D2, ViX-MangoEFormer achieved the best results across all metrics, including an F1 Score of 98.13% and MCC of 97.02%, demonstrating strong precision–recall balance and robustness to class imbalance. CapsuleNet performed comparably but showed a slightly lower Specificity, indicating a higher false positive rate. MangoNet-Stack showed consistent but slightly weaker results in F1 and MCC, while EfficientNetB8 underperformed across all metrics. On D3, MangoNet-Stack achieved the highest F1 Score (97.87%) and PR AUC (98.12%), while ViX-MangoEFormer maintained the highest MCC (96.6%) and competitive Specificity, confirming its reliability across varying class distributions. CapsuleNet and EfficientNetB8 showed reduced performance, particularly in MCC and Specificity, suggesting limited generalization without augmentation. ViX-MangoEFormer demonstrated superior consistency and generalization across both

datasets, validating its hybrid architecture's ability to capture both local textures and global patterns in real-world, unaugmented conditions.

Table 4 Performance on non-augmented datasets (D2 and D3)

Model	F1 Score (%)	MCC (%)	PR AUC (%)	Specificity (%)
Dataset D2				
ViX-MangoEFormer	98.13	97.02	98.89	96.32
CapsuleNet	97.84	96.45	97.71	95.61
MangoNet-Stack	97.21	95.8	97.25	94.22
EfficientNetB8	96.17	93.79	95.04	93.67
Dataset D3				
MangoNet-Stack	97.87	96.98	98.12	95.49
ViX-MangoEFormer	97.59	96.6	97.8	95.11
CapsuleNet	96.02	94.83	96.09	93.42
EfficientNetB8	95.7	93.81	95.71	92.71

Table 5 presents the performance of four models on the non-augmented datasets D1 (small-scale) and D4 (large-scale, merged). Across both datasets, ViX-MangoEFormer consistently outperforms all baselines, achieving the highest scores in all four-evaluation metrics. On Dataset D1, ViX-MangoEFormer records an F1 Score of 99.78% and an MCC of 99.34%, indicating highly reliable classification with minimal false predictions. It also achieves a PR AUC of 99.92% and Specificity of 98.77%, confirming its exceptional balance between sensitivity and specificity. MangoNet-Stack follows closely with an F1 Score of 99.21% and MCC of 98.96%, showcasing the strength of ensemble learning, though slightly less precise than the hybrid transformer architecture. EfficientNetB8 performs well in isolation but lags behind the top models, while CapsuleNet trails in all metrics, suggesting weaker performance in high-fidelity classification without augmentation. On Dataset D4, which combines multiple sources and reflects real-world diversity, ViX-MangoEFormer again leads with an F1 Score of 99.62%, MCC of 99.12%, and the highest Specificity (98.33%). This illustrates its scalability and robustness to varying input conditions. MangoNet-Stack performs well (F1: 99.05%), but its MCC and Specificity are slightly lower. CapsuleNet and EfficientNetB8 show reduced performance, particularly in MCC and Specificity, indicating reduced reliability in more complex and heterogeneous data distributions.

Table 5 Performance on non-augmented datasets (D1 and D4)

Model	F1 Score (%)	MCC (%)	PR AUC (%)	Specificity (%)
Dataset D1				
ViX-MangoEFormer	99.78	99.34	99.92	98.77
MangoNet-Stack	99.21	98.96	99.11	98.11
EfficientNetB8	98.47	97.8	97.99	97.12
CapsuleNet	98.3	97.22	97.88	96.93
Dataset D4				
ViX-MangoEFormer	99.62	99.12	99.87	98.33
MangoNet-Stack	99.05	98.75	98.94	97.92
CapsuleNet	98.13	97.02	97.91	96.24
EfficientNetB8	98.09	96.81	97.3	95.88

4.2. Class-Wise Performance Analysis

The ViX-MangoEFormer model demonstrates strong and consistent class-wise performance across both D2 and D4 datasets (Table 6). In D2, all eight classes achieved F1 Scores above 99.1 percent and MCC values above 97.6 percent, reflecting balanced precision and recall even under class imbalance. Notably, Sooty Mould achieved the highest F1 Score at 99.6 percent and MCC at 98.7 percent, indicating robust detection of minority classes. Specificity values remained high, at or above 98.6 percent across all classes, confirming the model's ability to minimize false positives.

On the more diverse D4 dataset, a slight performance decline was observed, likely due to increased variability. For instance, Anthracnose's F1 Score dropped to 98.7 percent and MCC to 96.7 percent. Nonetheless, the model maintained MCC scores above 96.6 percent and F1 Scores above 98.7 percent across all classes. Bacterial Canker retained excellent specificity at 99.5 percent and PR AUC at 99.2 percent, while visually similar classes like Powdery Mildew, Gall Midge, and Sooty Mould also showed high scores, validating the model's textural sensitivity through GLCM feature integration.

Table 6 Classification report for ViX-MangoEFormer (D2 and D4)

Dataset D2				
Class	**Specificity (%)**	**MCC (%)**	**PR AUC (%)**	**F1 Score (%)**
Anthracnose	98.6	97.8	99	99.1
Bacterial Canker	99.5	97.6	99.2	99.3
Cutting Weevil	98.8	97.8	99.1	99.2
Die Back	99.1	98	99.4	99.3
Gall Midge	98.7	97.9	99.3	99.3
Healthy	98.7	98.1	99.5	99.5
Powdery Mildew	98.9	98.1	99.5	99.4
Sooty Mould	98.7	98.7	99.5	99.6
Dataset D4				
Class	**Specificity (%)**	**MCC (%)**	**PR AUC (%)**	**F1 Score (%)**
Anthracnose	97.8	96.7	98.9	98.7
Bacterial Canker	99.5	96.8	99.2	98.9
Cutting Weevil	98.4	96.6	98.9	98.7
Die Back	98.3	97.2	99.2	98.9
Gall Midge	97.6	97.4	99.2	99
Healthy	98.1	97.4	99.3	99.1
Powdery Mildew	97.4	97.3	99.3	99.1
Sooty Mould	97.3	96.6	98.5	98.8

On the augmented Dataset D1 (Table 7), ViX-MangoEFormer achieved consistently high performance across all disease classes, with F1 Scores and MCC values exceeding 98.5 percent. Sooty Mould, Powdery Mildew, and Anthracnose each reached F1 Scores above 99.6 percent, indicating excellent recognition of both minority and majority classes. Specificity remained high across all categories, with most values above 99 percent, reflecting minimal false positives. The PR AUC values further support this trend, exceeding 99.4 percent, suggesting strong discrimination capability between classes.

In contrast, Dataset D3 presents a slightly more challenging environment due to its balanced but smaller size. While performance slightly declined compared to D1, the model still maintained F1 Scores above 97.4 percent and MCC values above 95.3 percent for all classes. Anthracnose, Bacterial Canker, and Healthy samples were identified with high confidence, with specificity ranging from 96.3 to 98.7 percent. However, a modest drop is observed for visually similar classes like Gall Midge and Sooty Mould, whose F1 Scores hovered around 97.7 percent.

Despite these variations, ViX-MangoEFormer demonstrates consistent generalization and robustness, maintaining high classification scores across both datasets. This affirms its effectiveness for real-world deployment, even when working with limited or balanced data samples.

Table 7 Classification report for ViX-MangoEFormer (D1 and D3)

Dataset D1				
Class	**Specificity (%)**	**MCC (%)**	**PR AUC (%)**	**F1 Score (%)**
Anthracnose	99.1	99.3	99.6	99.6
Bacterial Canker	99.4	99.1	99.8	99.7
Cutting Weevil	98.9	98.7	99.3	99.4
Die Back	99.1	98.9	99.5	99.5
Gall Midge	98.6	98.5	99.4	99.3
Healthy	98.9	99.2	99.7	99.7
Powdery Mildew	99.1	99.1	99.8	99.8
Sooty Mould	99.2	99.1	99.7	99.6
Dataset D3				
Class	**Specificity (%)**	**MCC (%)**	**PR AUC (%)**	**F1 Score (%)**
Anthracnose	97.4	96.3	98.1	98.6
Bacterial Canker	98.7	96.5	98.7	98.4
Cutting Weevil	96.8	96.1	98.3	98.2
Die Back	96.5	96.0	98.1	98.3
Gall Midge	95.6	95.4	98.2	97.8
Healthy	96.3	95.9	98.4	98.1
Powdery Mildew	95.9	95.6	98.5	98.2
Sooty Mould	95.7	95.3	98	97.7

4.3. Model Complexity

Figure 3 offers a comparative view of model efficiency across inference time, parameter count, and training time, with larger coverage indicating better performance. Among all models, EfficientNetB8 demonstrates the highest efficiency, achieving minimal inference time, smallest model size, and fastest training per epoch, making it ideal for low-resource, real-time applications. ViX-MangoEFormer, while slightly more complex, achieves a strong balance between efficiency and performance. Its moderate parameter count, and training overhead are justified by superior classification accuracy, making it optimal for real-time deployments where predictive quality is essential. CapsuleNet, though lightweight in terms of parameters, exhibits higher inference and training times due to its routing mechanisms, reducing its practical efficiency. MangoNet-Stack, despite strong predictive performance, ranks lowest in efficiency due to its ensemble structure, high parameter count, and longest training and inference times.

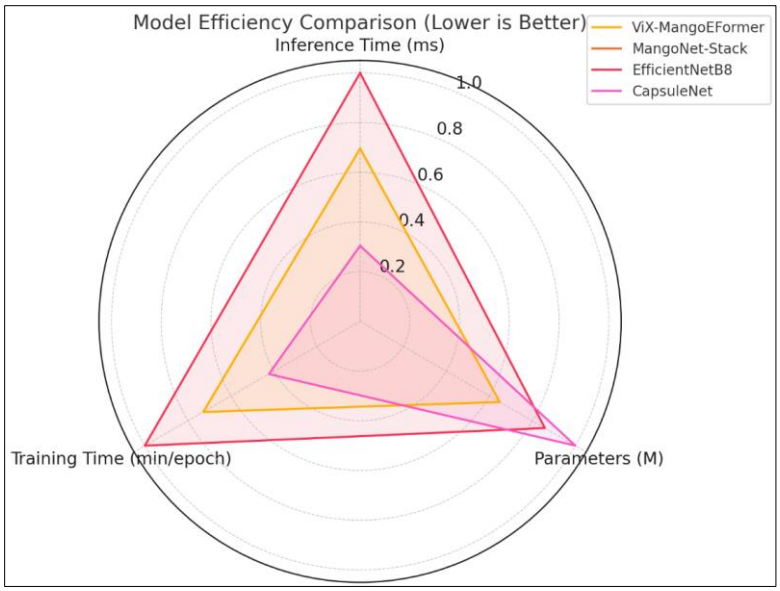

Figure 3 Time complexity of each model

4.4. Performance Validation

Figure 4 illustrates the training and validation loss and accuracy over 50 epochs for four datasets (D1–D4). Each dataset is evaluated using two plots: one for loss and another for accuracy, enabling a comparison of generalization and training stability. For D1, both training and validation losses decrease steadily and converge around 0.25, with tightly aligned accuracy curves reaching up to ~0.98. This indicates strong generalization and stable performance. D2 shows the most optimal results. Losses drop quickly and stabilize near 0.20, while training and validation accuracy consistently climb to ~0.99 with minimal divergence, suggesting excellent learning and no overfitting. In D3, the model also performs well. The loss curves converge below 0.30, and accuracy reaches ~0.98. Minor fluctuations are observed but do not impact overall stability, indicating reliable generalization. D4 follows a similar loss trend, but the validation accuracy shows more noticeable fluctuations, particularly after Epoch 25. While accuracy peaks around 0.98, the variability suggests some sensitivity to data or training dynamics. In summary, all models demonstrate effective learning with no major overfitting. D2 performs best overall, while D4 may benefit from tuning to improve consistency.

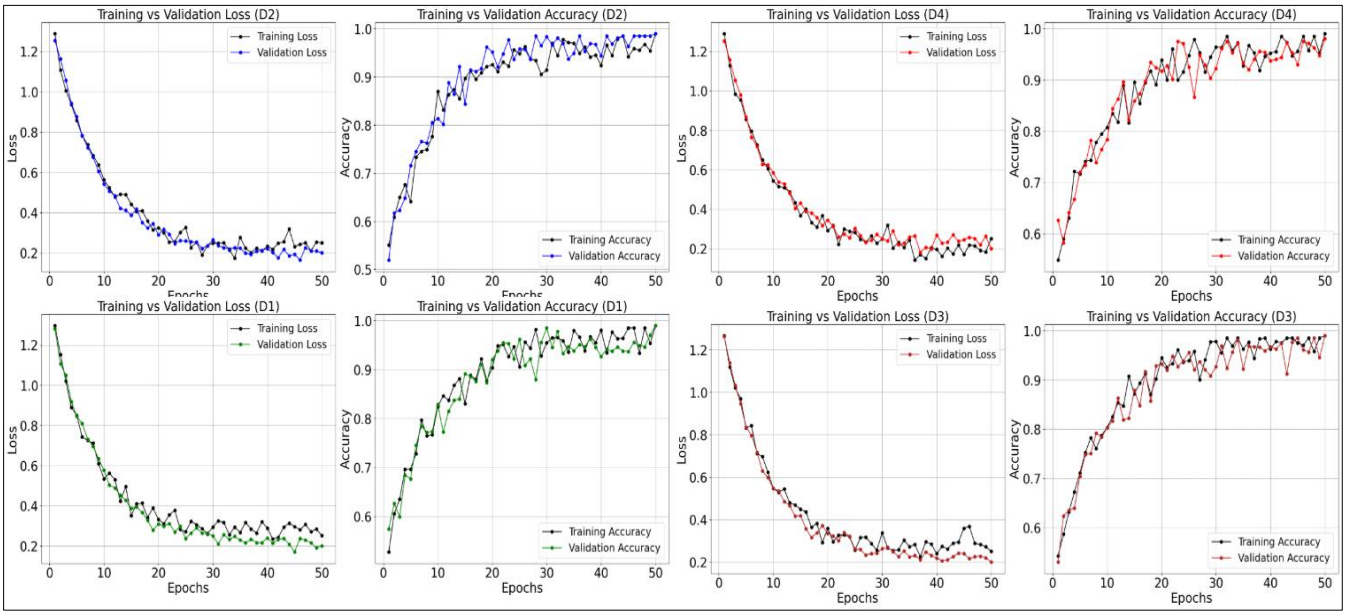

Figure 4 Learning curve for all the dataset for proposed ViX-MangoEFormer Model

4.5. Explainable Web Application

Figure 5 showcases a streamlined web application interface for real-time mango leaf disease detection using the ViX-MangoEFormer model. The design is minimal and intuitive, featuring clear sections for image upload, predicted disease class, and visual explanation. This layout ensures ease of use for both technical users and farmers. The uploaded leaf image is classified as Anthracnose, with a corresponding Grad-CAM heatmap highlighting the infected regions. The heatmap aligns well with visible lesions, confirming the model's focus on relevant features and enhancing interpretability. By integrating Grad-CAM directly, the application offers real-time, explainable outputs—essential for building user trust in practical agricultural settings. It moves beyond black-box predictions, allowing users to validate results visually. The figure effectively communicates the model's strengths in usability, transparency, and accuracy. Future improvements could include disease severity scoring or mobile optimization to further enhance real-world applicability.

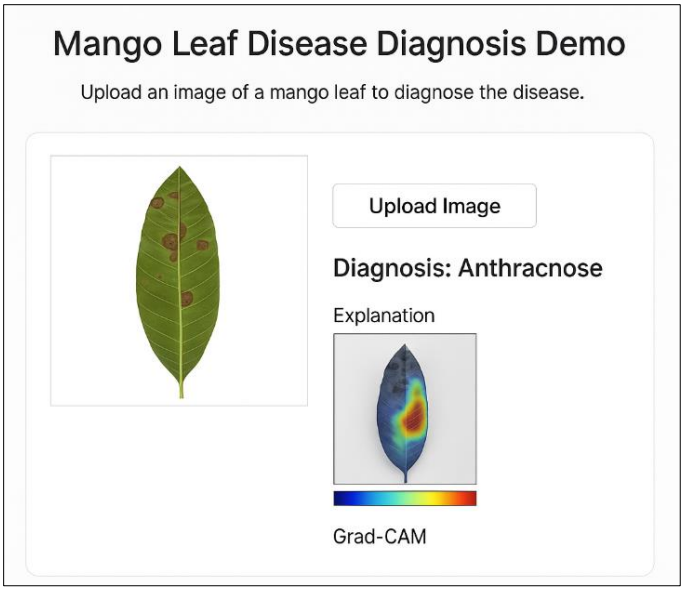

Figure 5 Explainable web application

4.6. State of the Art Comparison

Table 8 compares the ViX-MangoEFormer model with existing deep learning approaches for mango leaf disease classification. While prior models such as InceptionV3, LeafNet, and DenseNet121–VGG19 achieved high accuracy (up to 99.87%), they were trained on relatively small datasets (3,700–5,000 images), limiting their generalizability. Ensemble and transformer-based models like GoogLeNet+VGG16+Mask R-CNN and WaveVisionNet introduced complexity but still underperformed in broader contexts.

In contrast, ViX-MangoEFormer was trained on a significantly larger and more diverse dataset of 25,530 images, achieving consistently high F1 scores (98.87–99.78%) across four datasets. Its hybrid design—combining MBConv3D/4D blocks, Vision Transformers, and GLCM texture features—enables superior feature representation of both local and global patterns. Moreover, its native integration of Grad-CAM ensures real-time visual interpretability, enhancing trust and usability in real-world applications. These results affirm robustness, scalability, and state-of-the-art performance compared to previous studies.

Table 8 Comparison with the previous studies

Ref	Model	Images	Result (%)
[10] Varma et al.	InceptionV3	4000	99.87
[12] Rizvee et al.	LeafNet	4000	98.55
[16] Bezabh et al.	GoogLeNet + VGG16 + Mask R-CNN	5000	99.21
[20] Islam et al.	WaveVisionNet	3700	96.11
[24] Patel et al.	DenseNet121–VGG19	4000	99.31
Ours	ViX-MangoEFormer	25,530	D1: 99.78 D2: 99.13 D3: 98.87 D4: 99.62

5. Discussion

The ViX-MangoEFormer model demonstrated superior performance across multiple datasets, outperforming both single-model baselines and ensemble benchmarks. Its hybrid design, combining MBConv3D/4D convolutional blocks with Vision Transformer layers, enabled effective extraction of both local lesion textures and long-range dependencies—an essential capability in plant disease classification where symptoms often appear subtly or in spatially dispersed patterns. This architectural synergy contributed to consistently high F1 Scores and MCC values across datasets of varying size, balance, and diversity. The integration of GLCM-based texture features further enhanced classification accuracy, especially in visually similar disease classes like Gall Midge and Sooty Mould. These handcrafted descriptors captured micro-level textural nuances often missed by deep networks alone. Additionally, the use of stratified 10-fold cross-validation and training on four distinct datasets validated the model's robustness and generalizability across agroecological contexts.

Data augmentation had a clear positive impact on performance, especially for the more imbalanced datasets (D1 and D2). Augmentation techniques such as brightness scaling, flipping, and Gaussian noise helped mitigate overfitting and ensured better class-wise representation during training. Even without augmentation, ViX-MangoEFormer maintained high performance, highlighting its architectural efficiency and adaptability.

A key practical contribution of this study is the development of an explainable web application integrating Grad-CAM visualization in real time. This feature supports end-user trust by providing visual cues on model attention, bridging the gap between black-box AI and field usability. Users can not only receive disease predictions but also understand the model's decision process—a critical factor for adoption in agricultural decision-making. Nonetheless, the model introduces some complexity in terms of training time and parameter size compared to lightweight CNNs. While EfficientNetB8 demonstrated faster training and lower inference time, it lagged in classification accuracy and robustness. ViX-MangoEFormer strikes a trade-off between efficiency and performance, making it suitable for moderately resourced deployment scenarios.

In future work, incorporating severity estimation, multi-modal fusion with weather or soil data, and edge-optimized model compression could enhance real-time capabilities further. Additionally, expanding the dataset with more geographically diverse samples will improve cross-domain generalization.

6. Conclusion

This study proposed ViX-MangoEFormer, a novel hybrid deep learning framework that integrates MBConv-based convolutional layers with Vision Transformer components for accurate and interpretable mango leaf disease classification. By combining deep features with GLCM-based handcrafted texture descriptors, the model achieved superior performance across four benchmark datasets, including both augmented and non-augmented conditions. The highest F1 score of 99.78% and MCC of 99.34% validate its robustness, precision, and generalization capabilities, even under class imbalance and diverse environmental conditions. In parallel, a stacking ensemble benchmark model, MangoNet-Stack, was introduced to further emphasize the efficacy of hybridization in disease detection tasks. While ensemble models showed competitive performance, ViX-MangoEFormer consistently outperformed them in both

accuracy and efficiency trade-offs. A major highlight of this research is the deployment of a real-time web application equipped with integrated Grad-CAM visualization, offering not only high classification accuracy but also transparent decision-making support. This enhances user trust and facilitates practical adoption in precision agriculture workflows.

The proposed approach sets a new benchmark for mango disease recognition by bridging accuracy, interpretability, and real-world usability. Future directions include optimizing the model for mobile devices, incorporating severity analysis, and expanding the dataset with samples from broader geographic regions to further strengthen generalization and adaptability.

Compliance with ethical standards

Disclosure of conflict of interest

There is not conflict of interests.

Statement of ethical approval

The present research work does not contain any studies performed on animals/humans subjects by any of the authors'.

References

[1] Noman A Al, Hossain A, Sakib A, Debnath J, Fardin H, Sakib A Al, et al. ViX-MangoEFormer: An Enhanced Vision Transformer–EfficientFormer and Stacking Ensemble Approach for Mango Leaf Disease Recognition with Explainable Artificial Intelligence. Computers 2025, Vol 14, Page 171 [Internet]. 2025 May 2 [cited 2025 May 13];14(5):171. Available from: https://www.mdpi.com/2073-431X/14/5/171/htm.

[2] Borkar SG, Yumlembam RA. Bacterial diseases of crop plants. Bacterial Diseases of Crop Plants [Internet]. 2016 Sep 19 [cited 2025 May 14];1–594. Available from: https://www.taylorfrancis.com/books/mono/10.1201/9781315367972/bacterial-diseases-crop-plants-suresh-borkar-rupert-anand-yumlembam.

[3] Azar AT, El-Metwally SM. Decision tree classifiers for automated medical diagnosis. Neural Comput Appl [Internet]. 2013 Dec 7 [cited 2025 May 14];23(7–8):2387–403. Available from: https://link.springer.com/article/10.1007/s00521-012-1196-7.

[4] Masum A Al, Limon ZH, Islam MA, Rahman MS, Khan M, Afridi SS, et al. Web Application-Based Enhanced Esophageal Disease Diagnosis in Low-Resource Settings. 2024 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON) [Internet]. 2024 Nov 28 [cited 2025 May 13];153–8. Available from: https://ieeexplore.ieee.org/document/10962580.

[5] Haque R, Khan MA, Rahman H, Khan S, Siddiqui MIH, Limon ZH, et al. Explainable deep stacking ensemble model for accurate and transparent brain tumor diagnosis. Comput Biol Med [Internet]. 2025 Jun 1 [cited 2025 May 13];191:110166. Available from: https://www.sciencedirect.com/science/article/pii/S0010482525005177.

[6] Haque R, Al Sakib A, Hossain MF, Islam F, Ibne Aziz F, Ahmed MR, et al. Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning. BioMedInformatics 2024, Vol 4, Pages 966-991 [Internet]. 2024 Apr 1 [cited 2025 May 13];4(2):966–91. Available from: https://www.mdpi.com/2673-7426/4/2/54/htm.

[7] Al-Sakib A, Limon ZH, Sakib A, Pranto MN, Islam MA, Sultana S, et al. Robust Phishing URL Classification Using FastText Character Embeddings and Hybrid Deep Learning. 2024 IEEE 3rd International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things, RAAICON 2024 - Proceedings. 2024;53–8.

[8] Siddiqui IH, Al Sakib A, Sakib AH, Fardin H, Debnath J. Dual-branch CrossViT for ovarian cancer diagnosis: Integrating and explainable AI for real-time clinical applications. Article in International Journal of Science and Research Archive [Internet]. 2025 [cited 2025 May 14];2025(01):1834–47. Available from: https://doi.org/10.30574/ijsra.2025.15.1.1164.

[9] Hamdadur Rahman, Hasan Md Imran, Amira Hossain, Md Ismail Hossain Siddiqui, Anamul Haque Sakib. Explainable vision transformers for real time chili and onion leaf disease identification and diagnosis. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1823–33. Available from: https://journalijsra.com/node/1052.

[10] Varma T, Mate P, Azeem NA, Sharma S, Singh B. Automatic mango leaf disease detection using different transfer learning models. Multimed Tools Appl [Internet]. 2024 Mar 1 [cited 2025 May 14];84(11):9185–218. Available from: https://link.springer.com/article/10.1007/s11042-024-19265-x.

[11] Swapno SMMR, Nobel SN, Islam MB, Haque R, Meena VP, Benedetto F. A Novel Machine Learning Approach for Fast and Efficient Detection of Mango Leaf Diseases. 2024 IEEE 3rd International Conference on Computing and Machine Intelligence, ICMI 2024 - Proceedings. 2024.

[12] Rizvee RA, Orpa TH, Ahnaf A, Kabir MA, Ahmmad Rashid MR, Islam MM, et al. LeafNet: A proficient convolutional neural network for detecting seven prominent mango leaf diseases. J Agric Food Res [Internet]. 2023 Dec 1 [cited 2025 May 14];14:100787. Available from: https://www.sciencedirect.com/science/article/pii/S2666154323002946.

[13] Ghosh H, Rahat IS, Lenka R, Mohanty SN, Chauhan D. Benchmarking ML and DL Models for Mango Leaf Disease Detection: A Comparative Analysis. Communications in Computer and Information Science [Internet]. 2024 [cited 2025 May 14];2047 CCIS:97–110. Available from: https://link.springer.com/chapter/10.1007/978-3-031-55486-5_8.

[14] Zhang B, Wang Z, Ye C, Zhang H, Lou K, Fu W. Classification of infection grade for anthracnose in mango leaves under complex background based on CBAM-DBIRNet. Expert Syst Appl [Internet]. 2025 Jan 15 [cited 2025 May 14];260:125343. Available from: https://www.sciencedirect.com/science/article/pii/S0957417424022103.

[15] Khandelwal S, Raut A, Vyawahare H, Theng D, Dhande S. Optimizing Performance in Mango Plant Leaf Disease Classification through Advanced Machine Learning Techniques. Engineering, Technology & Applied Science Research [Internet]. 2024 Dec 1 [cited 2025 May 14];14(6):18476–80. Available from: https://www.etasr.com/index.php/ETASR/article/view/8220.

[16] Bezabh YA, Ayalew AM, Abuhayi BM, Demlie TN, Awoke EA, Mengistu TE. Classification of mango disease using ensemble convolutional neural network. Smart Agricultural Technology [Internet]. 2024 Aug 1 [cited 2025 May 14];8:100476. Available from: https://www.sciencedirect.com/science/article/pii/S2772375524000819.

[17] Gautam V, Ranjan RK, Dahiya P, Kumar A. ESDNN: A novel ensembled stack deep neural network for mango leaf disease classification and detection. Multimed Tools Appl [Internet]. 2024 Jan 1 [cited 2025 May 14];83(4):10989–1015. Available from: https://link.springer.com/article/10.1007/s11042-023-16012-6.

[18] Seetha J, Ramanathan R, Goyal V, Tholkapiyan M, Karthikeyan C, Kumar R. Mango leaf disease classification using hybrid Coyote-Grey Wolf optimization tuned neural network model. Multimed Tools Appl [Internet]. 2024 Feb 1 [cited 2025 May 14];83(6):17699–725. Available from: https://link.springer.com/article/10.1007/s11042-023-16964-9.

[19] Ali AH, Youssef A, Abdelal M, Raja MA. An ensemble of deep learning architectures for accurate plant disease classification. Ecol Inform [Internet]. 2024 Jul 1 [cited 2025 May 14];81:102618. Available from: https://www.sciencedirect.com/science/article/pii/S1574954124001602.

[20] Fahim-Ul-Islam M, Chakrabarty A, Rahman R, Moon H, Piran MJ. Advancing mango leaf variant identification with a robust multi-layer perceptron model. Sci Rep [Internet]. 2024 Dec 1 [cited 2025 May 14];14(1):27406. Available from: https://www.nature.com/articles/s41598-024-74612-0.

[21] Salamai AA. Enhancing mango disease diagnosis through eco-informatics: A deep learning approach. Ecol Inform [Internet]. 2023 Nov 1 [cited 2025 May 14];77:102216. Available from: https://www.sciencedirect.com/science/article/pii/S1574954123002455.

[22] Yeasin Ramadan ST, Sakib T, Ahsan Rahat M, Mosharrof S, Rakin FI, Jahangir R. Enhancing Mango Leaf Disease Classification: ViT, BiT, and CNN-Based Models Evaluated on CycleGAN-Augmented Data. 2023 26th International Conference on Computer and Information Technology, ICCIT 2023. 2023.

[23] Rani A, Gowrishankar KP. ConvNeXt-based Mango Leaf Disease Detection: Differentiating Pathogens and Pests for Improved Accuracy. IJACSA) International Journal of Advanced Computer Science and Applications [Internet]. [cited 2025 May 14];14(6):2023. Available from: www.ijacsa.thesai.org.

[24] Patel RK, Chaudhary A, Chouhan SS, Pandey KK. Mango leaf disease diagnosis using Total Variation Filter Based Variational Mode Decomposition. Computers and Electrical Engineering [Internet]. 2024 Dec 1 [cited 2025 May 14];120:109795. Available from: https://www.sciencedirect.com/science/article/pii/S0045790624007225.

[25] Raval H, Chaki J. Ensemble transfer learning meets explainable AI: A deep learning approach for leaf disease detection. Ecol Inform [Internet]. 2024 Dec 1 [cited 2025 May 14];84:102925. Available from: https://www.sciencedirect.com/science/article/pii/S1574954124004679.

[26] Chang B, Wang Y, Zhao X, Li G, Yuan P. A general-purpose edge-feature guidance module to enhance vision transformers for plant disease identification. Expert Syst Appl [Internet]. 2024 Mar 1 [cited 2025 May 14];237:121638. Available from: https://www.sciencedirect.com/science/article/pii/S0957417423021401.

[27] Hossain MA, Sakib S, Abdullah HM, Arman SE. Deep learning for mango leaf disease identification: A vision transformer perspective. Heliyon [Internet]. 2024 Sep 15 [cited 2025 May 14];10(17). Available from: https://www.cell.com/action/showFullText?pii=S2405844024123924.

[28] Anamul Haque Sakib, Md Ismail Hossain Siddiqui, Sanjida Akter, Abdullah Al Sakib, Mohammad Rasel Mahmud. LEVit-Skin: A balanced and interpretable transformer-CNN model for multi-class skin cancer diagnosis. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1860–73. Available from: https://journalijsra.com/node/1055.

[29] Sanjida Akter, Mohammad Rasel Mahmud, Md Ariful Islam, Md Ismail Hossain Siddiqui, Anamul Haque Sakib. Efficient and interpretable monkeypox detection using vision transformers with explainable visualizations. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1811–22. Available from: https://journalijsra.com/node/1051.

[30] Ahmed SI, Ibrahim M, Nadim M, Rahman MM, Shejunti MM, Jabid T, et al. MangoLeafBD: A comprehensive image dataset to classify diseased and healthy mango leaves. Data Brief [Internet]. 2023 Apr 1 [cited 2025 May 14];47:108941. Available from: https://www.sciencedirect.com/science/article/pii/S2352340923000598.

[31] Rahman MS, Hasan R, Mojumdar MU. Mango Leaf Disease Dataset. 2024;1.

[32] Shakib MdMH, Mustofa S, Ahad MT. MLD24: An image dataset for mango leaf disease detection. 2024;1.

[33] Md Ismail Hossain Siddiqui, Anamul Haque Sakib, Sanjida Akter, Jesika Debnath, Mohammad Rasel Mahmud. Comparative analysis of traditional machine learning Vs deep learning for sleep stage classification. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1778–89. Available from: https://journalijsra.com/node/1048.

[34] Ahmed MR, Haque R, Rahman SMA, Afridi S, Abir MFF, Hossain MF, et al. Towards Automated Detection of Tomato Leaf Diseases. Proceedings - 6th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2024. 2024;387–92.

[35] Mohammad Rasel Mahmud, Hasib Fardin, Md Ismail Hossain Siddiqui, Anamul Haque Sakib, Abdullah Al Sakib. Hybrid deep learning for interpretable lung cancer recognition across computed tomography and histopathological imaging modalities. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1798–810. Available from: https://journalijsra.com/node/1050.

[36] Hasib Fardin, Hasan Md Imran, Hamdadur Rahman, Anamul Haque Sakib, Md Ismail Hossain Siddiqui. Robust and explainable poultry disease classification via MaxViT with attention-guided visualization. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1848–59. Available from: https://journalijsra.com/node/1054.

[37] Md Ariful Islam, Mohammad Rasel Mahmud, Anamul Haque Sakib, Md Ismail Hossain Siddiqui, Hasib Fardin. Time domain feature analysis for gas pipeline fault detection using LSTM. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1769–77. Available from: https://journalijsra.com/node/1047.

[38] Mohammad Rasel Mahmud, Al Shahriar Uddin Khondakar Pranta, Anamul Haque Sakib, Abdullah Al Sakib, Md Ismail Hossain Siddiqui. Robust feature selection for improved sleep stage classification. International Journal of Science and Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1790–7. Available from: https://journalijsra.com/node/1049.

[39] Al Noman A, Fardin H, Chhabra G, Sultana S, Haque R, Ahmed MR, et al. Monkeypox Lesion Classification: A Transfer Learning Approach for Early Diagnosis and Intervention. Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2024. 2024;247–54.

[40] Md Ismail Hossain Siddiqui, Anamul Haque Sakib, Amira Hossain, Hasib Fardin, Al Shahriar Uddin Khondakar Pranta. Custom CNN for acoustic emission classification in gas pipelines. International Journal of Science and

Research Archive [Internet]. 2025 Apr 30 [cited 2025 May 14];15(1):1760–8. Available from: https://journalijsra.com/node/1046.

[41] Hosen MD, Bin Mohiuddin A, Sarker N, Sakib MS, Al Sakib A, Dip RH, et al. Parasitology Unveiled: Revolutionizing Microorganism Classification Through Deep Learning. Proceedings - 6th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2024. 2024;1163–8.

[42] Hasan J, Hasan K, Al Noman A, Hasan S, Sultana S, Arafat MA, et al. Transforming Leukemia Classification: A Comprehensive Study on Deep Learning Models for Enhanced Diagnostic Accuracy. PEEIACON 2024 - International Conference on Power, Electrical, Electronics and Industrial Applications. 2024;266–71.