Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.061890

ARTICLE

Mango Disease Detection Using Fused Vision Transformer with ConvNeXt Architecture

Faten S. Alamri[1], Tariq Sadad[2,*], Ahmed S. Almasoud[3], Raja Atif Aurangzeb[4] and Amjad Khan[3]

[1]Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, 11671, Saudi Arabia

[2]Department of Computer Science, University of Engineering & Technology, Mardan, 23200, Pakistan

[3]AIDA Lab, Department of Information Systems, College of Computer & Information Sciences, Prince Sultan University, Riyadh, 12435, Saudi Arabia

[4]Department of Computer Science, International Islamic University, Islamabad, 44000, Pakistan

*Corresponding Author: Tariq Sadad. Email: tariqsadad@gmail.com

Received: 05 December 2024; Accepted: 11 February 2025; Published: 26 March 2025

ABSTRACT: Mango farming significantly contributes to the economy, particularly in developing countries. However, mango trees are susceptible to various diseases caused by fungi, viruses, and bacteria, and diagnosing these diseases at an early stage is crucial to prevent their spread, which can lead to substantial losses. The development of deep learning models for detecting crop diseases is an active area of research in smart agriculture. This study focuses on mango plant diseases and employs the ConvNeXt and Vision Transformer (ViT) architectures. Two datasets were used. The first, MangoLeafBD, contains data for mango leaf diseases such as anthracnose, bacterial canker, gall midge, and powdery mildew. The second, SenMangoFruitDDS, includes data for mango fruit diseases such as Alternaria, Anthracnose, Black Mould Rot, Healthy, and Stem and Rot. Both datasets were obtained from publicly available sources. The proposed model achieved an accuracy of 99.87% on the MangoLeafBD dataset and 98.40% on the MangoFruitDDS dataset. The results demonstrate that ConvNeXt and ViT models can effectively diagnose mango diseases, enabling farmers to identify these conditions more efficiently. The system contributes to increased mango production and minimizes economic losses by reducing the time and effort needed for manual diagnostics. Additionally, the proposed system is integrated into a mobile application that utilizes the model as a backend to detect mango diseases instantly.

KEYWORDS: ConvNeXt model; fusion; mango disease; smart agriculture; vision transformer

1 Introduction

Mango is one of the most popular fruits and holds significant economic importance globally [1]. Mango is highly valued worldwide and is renowned as the “king of fruits” [2]. Global mango production reached 54.83 million tons in 2020, according to the Food and Agriculture Organization of the United Nations [3]. Mangoes are not only a rich source of vitamin A but also contain an array of nutrients, including vitamins, minerals, fiber, prebiotic dietary substances, and antioxidants that promote overall health. Studies have highlighted that carotene-rich fruits like mango can help prevent lung and mouth cancer while protecting against colon, breast, leukemia, and prostate cancer. Despite their nutritional and economic importance, mango crops are highly vulnerable to various diseases and pests, significantly impacting their yield and quality. Diseases such as powdery mildew, mango malformation, bacterial canker, mango dieback, stem-end rot, and mango wilt pose serious threats to mango production [4,5]. The export market is particularly sensitive to bacterial canker due to its detrimental effects on fruit quality. Traditional methods for disease management have relied on chemical treatments and visual inspection by knowledgeable experts. However, these approaches are labor-intensive, prone to errors, and often inaccessible to small-scale farmers in developing nations. Inexperienced farmers, in particular, face challenges in accurately identifying pests and diseases, resulting in poor decision-making and reduced agricultural productivity [6]. Consequently, there is a need for intelligent technological solutions to address these challenges and ensure sustainable agricultural development. Recently, advances in machine learning, deep learning, and Internet-of-Things (IoT)-based approaches have paved the way for early and efficient disease detection in fruits and vegetables [7].

The structure of the paper is as follows: Section 2 reviews the related literature, while the proposed methodology is detailed in Section 3. Section 4 discusses the experimental setup and results, and Section 5 presents a comprehensive discussion. Finally, Section 6 concludes the study and outlines future research directions.

2 Literature Review

Several research works have been conducted on mango leaf and fruit diseases.

2.1 Mango Leaf Diseases

Detection of mango leaf diseases has been a focus in numerous studies, though it presents challenges in practical implementation [8,9]. In [10], the authors explored deep-learning techniques for detecting mango diseases. However, they faced limitations, such as errors in leaf segmentation, issues with real-time processing, and a lack of sufficient training samples. These constraints diminished the applicability and effectiveness of the model, making it less reliable for real-world use. Similarly, the study in [11] developed a machine learning model using Support Vector Machines (SVM) for disease diagnosis, achieving an average diagnostic accuracy of 80% across four disease groups. Despite this, the reliance on image matching produced inconsistent results, especially under varying environmental conditions or image characteristics, which undermines the model's utility in more complex disease scenarios. Multiple models have been proposed to enhance disease detection in mango cultivation. For instance, MobileNetV2 achieved an accuracy of 97%, while hybrid models combining SVM with neural networks, random forests, Inception V3, and Long Short-term Memory (LSTM) reported accuracies ranging between 91% and 92% [12]. However, while these models have shown promising results, Convolutional Neural Networks (CNNs) still fall short of meeting the high demands of agricultural applications, particularly in disease and pest management. In [13], a dataset of 4000 images depicting mango leaf diseases was collected from various orchards in Bangladesh. Specialists carefully labeled the dataset to minimize sampling bias, but its geographic concentration restricts its applicability to other regions or environmental conditions.

Additionally, the manual labeling process is resource-intensive, which could hinder scalability. In [14], the authors utilized a hybrid model combining VGG-16 and MobileNet, employing stacking ensemble learning for disease categorization. This model was trained on a dataset of only 329 images of sunflowers, grouped into five categories. The small dataset raises concerns about the model's ability to generalize to larger or more diverse datasets, and ensemble models are often computationally intensive, making them less suitable for real-time or resource-constrained environments. Recent advancements have helped address some of these limitations. The study in [15] highlighted the use of deep learning for the automated grading and classification of mango fruits, achieving an impressive classification accuracy of 99.2% and grading accuracy of 96.7%. However, these methods primarily focus on grading and classification rather than disease detection. In [16], a novel approach using vein-pattern analysis and Canonical Correlation Analysis (CCA) was employed to segment diseased parts of mango leaves. This model achieved an accuracy of 95.5% for disease detection, but its reliance on vein patterns may limit its effectiveness for diseases that do not show clear vein-related symptoms. Similarly, in [17], the Multi-scale and Multi-pooling Convolutional Neural Network (MSMP-CNN) demonstrated significant potential in disease detection, increasing its accuracy from 95% to 98.5% through pre-training and transfer learning. While promising, these methods require substantial computational resources, which may limit their deployment in resource-limited environments.

2.2 Mango Fruit Diseases

The study on detecting anomalies in mango fruits is still in its early stages, but some valuable contributions have been made. In [18], a system for real-time mango grading was developed, incorporating various measures for assessing maturity, shape, size, and surface defects. The system employed Support Vector Regression (SVR) for maturity prediction, Multiple Attribute Decision Making (MADM) for quality assessment, and fuzzy incremental learning for grading, achieving an accuracy of 87%. While the system showed promise, its accuracy was moderate compared to other systems, indicating room for improvement to make the system more reliable for practical use. In [19], the authors developed a surface defect recognition system using a computer vision algorithm. The system was tested on mango varieties such as Dashehari and Chausa, achieving accuracies of 88.6% and 93.3%, respectively. Although these results are commendable, they are limited to specific mango varieties. They may not generalize well to other varieties or to diverse environmental conditions, which restricts the model's overall applicability. Another study in [20] proposed an image processing algorithm to detect defects and assess maturity using digital images' features such as shape, color, and size. The approach effectively determined defects and maturity but did not address the more complex issues of internal diseases or other hidden anomalies in mango fruits, thus limiting its scope. In [21], the authors used the Bilateral Filtering (BF) technique to reduce image noise, followed by adaptive threshold-based segmentation and feature extraction using the VGG-16 model. Hyperparameter optimization was performed using the Whale Optimization Algorithm (WOA), and classification was carried out using a Quasi-Recurrent Neural Network (QRNN), achieving an accuracy of 99.29%. While this method demonstrated high accuracy, it relies on advanced computational techniques, which may not be feasible for real-time use in resource-constrained environments, limiting its practical deployment in large-scale farming operations.

Table 1 summarizes the studies on mango leaf and fruit disease detection, highlighting the methodologies and limitations.

Table 1: Summary of literature on mango disease detection

Study	Methodology	Accuracy	Limitations
[11]	SVM-based model	80%	Inconsistent results due to reliance on image matching
[15]	Deep learning	96.7%	Focused on grading
[18]	SVR, MADM, fuzzy learning	87%	Moderate accuracy, needs improvement
[19]	Deep learning	88.6%–93.3%	Limited to specific varieties, not generalizable
[21]	QRNN	99.29%	High computational complexity limits large-scale use

The need for the study is:

- Current systems achieve only moderate accuracy (e.g., 87%), and improvements are necessary to meet real-world requirements

- Existing models tend to focus on specific mango varieties or geographical locations, limiting their applicability
- Most models only detect surface defects, missing internal diseases and hidden issues.
- The main contribution of this paper is:
- A fused ViT-ConvNeXt architecture that combines convolutional and transformer networks to enhance the detection of subtle disease features in mango leaves and fruits.
- The model is tested on benchmark datasets (MangoLeafBD, SenMangoFruitDDS) and extended to generalization tests with external data to increase its robustness in real-world environments.
- The model minimizes dependency on data augmentation and is designed to work efficiently in resource-constrained environments, making it ideal for smart agriculture applications.

3 Proposed Methodology

In this work, we focus on mango disease detection using the ViT-ConvNeXt architecture, which combines the advanced image processing capabilities of ConvNeXt with the global feature modeling strength of Vision Transformers (ViT) to enable robust and accurate disease identification. The proposed methodology begins with collecting high-quality images of mango leaves and fruits captured using mobile phones under natural orchard conditions. These images are then uploaded to a cloud server that hosts the pre-trained ViT-ConvNeXt model. The model processes the uploaded images by analyzing distinct morphological features to classify diseases affecting mango leaves and fruits as shown in Fig. 1. The results of this analysis are made accessible to farmers and agricultural stakeholders, providing actionable insights for effective disease management. By integrating the complementary strengths of convolutional and transformer-based systems, this approach enhances the efficiency and accuracy of mango disease detection. By employing intelligent technologies, this approach aims to enhance the efficiency and accuracy of mango disease detection, supporting farmers in reducing crop losses and improving yields.

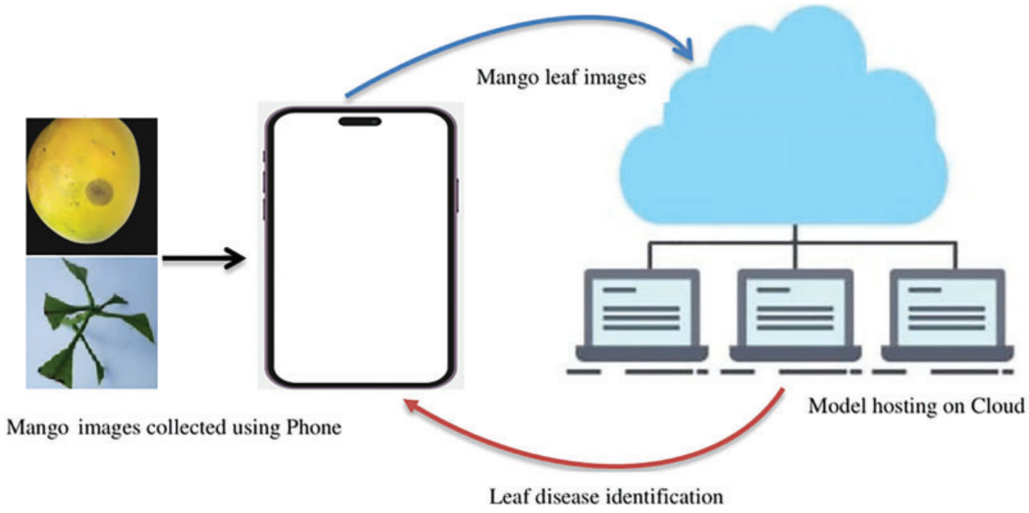

Mango leaf images
Mango images collected using Phone
Model hosting on Cloud
Leaf disease identification

Figure 1: Proposed methodology

3.1 ConvNext Architecture

ConvNeXt is a modern CNN architecture that enhances the traditional CNN framework with innovative features designed for improved performance, efficiency, and scalability [22]. Drawing inspiration from both CNNs and transformer-based models like ViTs [23], ConvNeXt incorporates several key advancements. It uses larger convolutional kernels to capture more detailed spatial information and employs depthwise separable convolutions, which reduce computational costs while maintaining high accuracy. The architecture also introduces Layer Normalization instead of the conventional Batch Normalization, making it less dependent on batch size. It utilizes the Gaussian Error Linear Unit (GELU) activation function for smoother and more stable training. ConvNeXt also adapts ResNet (Residual Network)-style bottleneck blocks with refinements, such as inverted bottlenecks, to enhance efficiency. One of ConvNeXt's strengths is its ability to scale effectively with both depth and width, achieving state-of-the-art results across various benchmarks while requiring fewer computational resources. This scalability, combined with the architecture's straightforward design, makes ConvNeXt both powerful and accessible for a wide range of applications. By integrating insights from transformer models, such as large kernels and advanced normalization techniques, ConvNeXt offers a blend of the best features from CNNs and transformers. As a result, it surpasses traditional CNNs in accuracy and efficiency and demonstrates superior generalization across different datasets and tasks, making it a leading choice for modern image recognition challenges.

Fig. 2 shows a block from the ConvNeXt architecture, an input tensor with 96 channels of depth is used to start the data flow at the left. The input is first subjected to a 7 × 7 depthwise convolution, effectively capturing intricate spatial information. In order to improve training stability and lessen susceptibility to batch sizes, layer normalization (LN), which normalizes the features rather than the batch dimension, is then employed. The feature depth is then increased from 96 to 384 channels using a pointwise convolution with a 1 × 1 kernel, which improves the model's ability to learn intricate representations. The GELU activation function, which processes this convolution's output, offers smoother and more efficient training than conventional ReLU (Rectified Linear Unit) activations. To preserve the model's effectiveness and prepare the tensor for the following step, a second 1 × 1 pointwise convolution lowers the depth back to 96 channels. One of the main characteristics of this block is the residual connection, which directly adds the original input tensor to the second pointwise convolution's output. This skip connection improves gradient flow during training and lessens the vanishing gradient issue. The final output of the block is then passed on to the next layer in the ConvNeXt architecture.

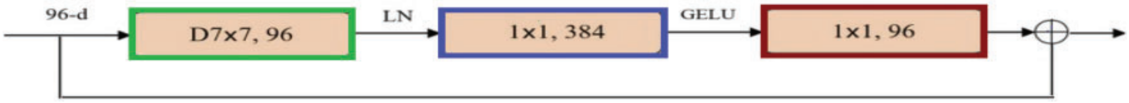


Figure 2: ConvNext architecture

3.2 Fused ViT ConvNext Architecture

The architecture of the FusionModel is made up of two state-of-the-art vision models, the ViT and ConvNeXt model as demonstrated in Fig. 3. The purpose of the ViT part is to process the input images, which are divided into patches, using transformer strokes to extract features. It consists of an embedding layer with patches embeddings, a twelve-layer encoder composed of self-attention and feed forward networks, several layer normalizations around the layers, and a pooler to collect the outputs. On the other hand, the ConvNeXt part deals with images in a hierarchical processing manner using convolutional layers. Its first stem is a convolutional layer that shrinks the size and a following LayerNorm. The ConvNeXt stages possess several modules referred to as "ConvNeXtBlock" in each stage with depth-wise convolutions, LayerNorm and MLP (Multilayer Perceptron) with Global Response Norm (GRN), respectively. These stages perform operations to reduce the size of the features and increase the number of channels while maintaining a flow of gradients by means of skip connections. As such, this construction signifies an architecture that integrates the global self-attention property of ViT that captures long-range dependencies and the local feature learning structure of ConvNeXt, achieving excellent performance in visual representations learning.

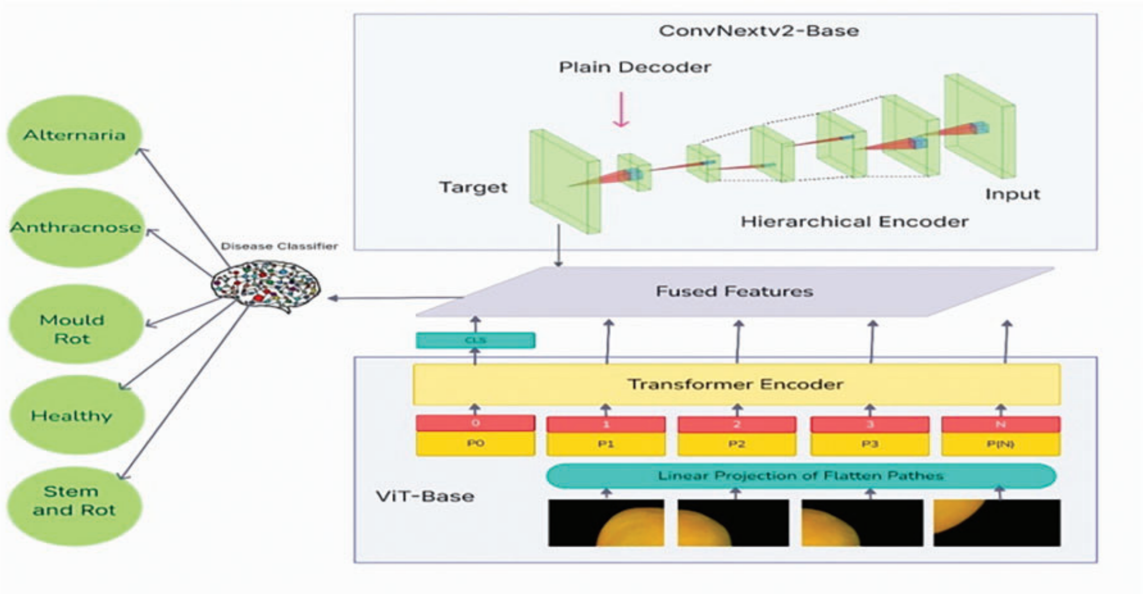


Figure 3: Fused ViT ConvNext architcture

3.3 Dataset

We used two datasets to demonstrate the efficacy of the proposed model.

3.3.1 MangoLeafBD

The MangoLeafBD dataset comprises 4000 high-quality digital images in PNG format, each sized at 240 × 320 pixels and containing RGB (Red, Green, Blue) color channels [13]. These images were collected before winter 2021 with assistance from agricultural specialists, who identified trees affected by various mango leaf diseases. The dataset encompasses eight categories: seven representing diseased leaves and one for healthy leaves, each containing 500 images. This makes the dataset inherently balanced and eliminates the need for additional balancing techniques during model training. The labeling process was conducted manually with expert guidance to ensure accurate categorization of each image. To improve the dataset's quality, a rigorous preprocessing pipeline was implemented. This involved resizing images to a uniform resolution, discarding low-quality images, and applying data augmentation techniques such as zooming and rotation. These steps enhanced the diversity and robustness of the dataset, ensuring that it is well-suited for training machine learning models. The disease categories covered in the dataset include Powdery Mildew, which appears as white, powdery growths on leaves, flowers, and fruits; Honeydew, characterized by sticky insect secretions that promote sooty mold growth; Anthracnose, marked by black necrotic patches often along the edges of leaves; Bacterial Canker, caused by Pseudomonas mangifera and resulting in water-soaked spots evolving into cankers; Cutting Weevil Disease, which creates scissor-like cuts on leaves; Dieback, where twigs dry out and break off with leaves turning brown and falling; and Gall Midge Disease, identified by pimple-like bumps that may lead to leaf loss and reduced fruit yield. The healthy leaf category includes images of leaves free from visible disease symptoms. Figs. 4–6 in the study provide visual examples of each category, showcasing the distinct characteristics of the diseases and healthy leaves.

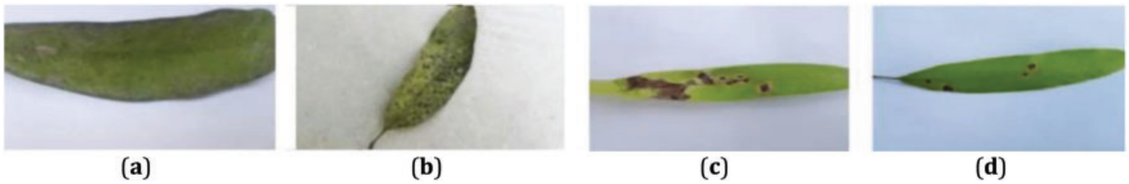


Figure 4: (**a**) Powdery mildew (**b**) Sooty mold (**c**) Anthracnose (**d**) Anthracnose

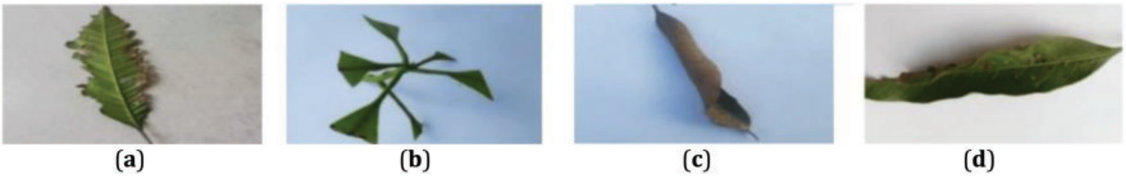


Figure 5: (**a**) Bacterial canker (**b**) Cutting weevil (**c**) Dieback (**d**) Gall midge

Figure 6: Healthy leaves

3.3.2 SenMangoFruitDDS

The SenMangoFruitDDS dataset comprises 838 labeled images in 224 × 224 JPG format [24]. This dataset captures five phenotypic classes, including four representing diseases caused by fungal infections and one class for healthy mango fruits as illustrated in Fig. 7. Images were taken under natural orchard conditions in Senegal using a mobile phone camera. The labels were determined with the assistance of agricultural experts, ensuring accurate identification of diseases such as Alternaria, Anthracnose, Black Mould Rot, and Stem End Rot. The phenotypic traits associated with these diseases, such as necrotic lesions, enzymatic degradation, and tissue breakdown, were visually distinguished and documented. The dataset were balanced to the extent possible, ensuring an even distribution of images across the various classes. Augmentation techniques, including rotation, cropping, and flipping, were employed to mitigate the risk of overfitting and improve the diversity of the training data.

Real World Challenges: Classifying these diseases poses significant challenges due to overlapping phenotypic traits, such as similar discoloration and lesion patterns. Disease symptoms vary based on the stage of infection, environmental factors, and fruit maturity, further complicating accurate differentiation. Additionally, fungal pathogens often induce visually similar damage, making it difficult to distinguish between infections solely through visual inspection. Variability in imaging conditions, including lighting and angle, adds another layer of complexity, necessitating robust image processing and classification techniques to achieve reliable results.

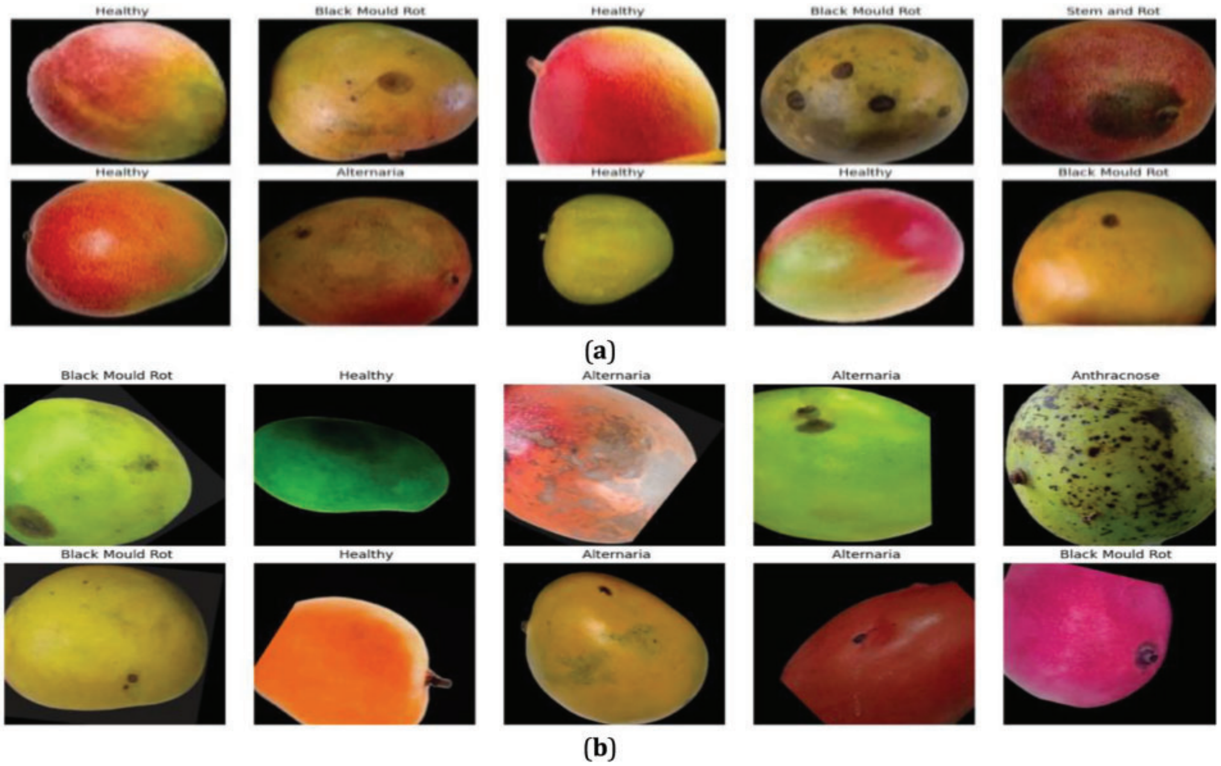


Figure 7: Mango fruit disease images before & after augmentation ((**a**) & (**b**)). (**a**): Dataset Before Augmentation (838 Samples); (**b**): Dataset After Augmentation (5000 Samples)

3.4 Implementation Details

This section describes the implementation details of the proposed model. The data preparation phase involves augmentation techniques, including random horizontal and vertical flips, to increase training data diversity and improve the model's ability to recognize objects in various orientations. Images are resized to 224 × 224 pixels and processed in batches of 32, aligning with the model's input requirements. The architecture consists of two parallel branches. The ConvNeXt branch processes input images through three successive blocks and a final layer, which extract spatial hierarchies and combine local features.

Meanwhile, the ViT branch begins with a patch embedding layer that divides the image into patches and processes them through two linear transformer blocks, leveraging attention mechanisms to capture long-range dependencies and global context. The outputs of the two branches are merged in a Fusion Layer, where ConvNeXt's spatial hierarchies and ViT's global context are integrated to create a comprehensive feature representation as illustrated in Fig. 8. The fused features are then passed through subsequent layers, starting with a Global Average Pooling (GAP) layer, which condenses feature maps into a single vector. This is followed by a Batch Normalization layer to stabilize and accelerate training, and two dense layers: 512 neurons, GeLU (Gaussian Error Linear Unit) activation, and a 0.4 dropout rate, and the second with 256 neurons and a 0.3 dropout rate, to enhance robustness. The final classification uses a dense output layer with a Softmax activation function, producing a probability distribution for the target classes. The model is trained using the Adam optimizer with an initial learning rate of 0.001, which decays over time for stable convergence. The loss function employed is Sparse Categorical Cross-Entropy, suitable for multi-class classification tasks, with accuracy as the primary evaluation metric in Table 2.

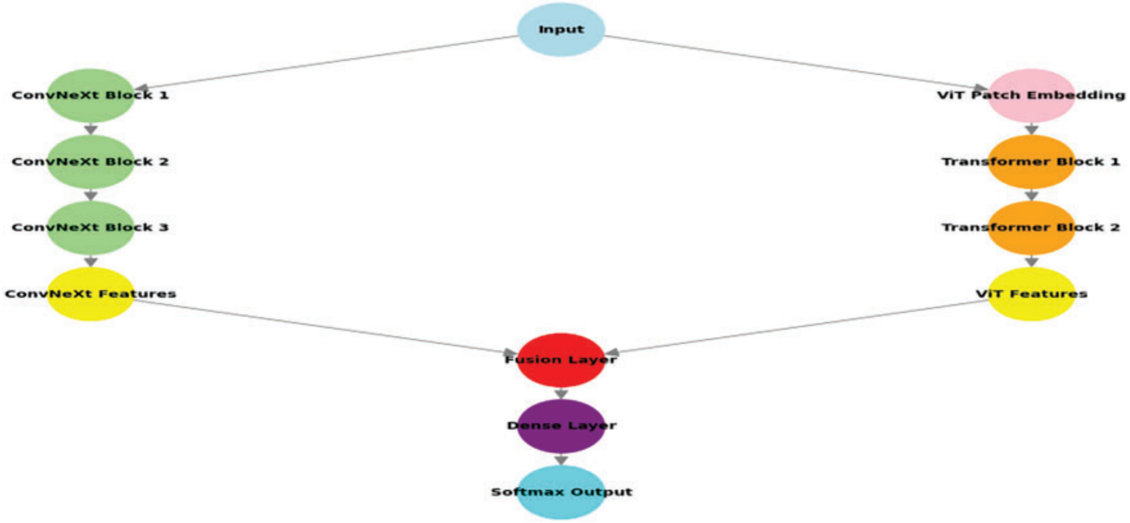


Figure 8: Model layout

Pseudo code of the proposed model is also provided below for more clarity of the proposed model.

```
I.   Load Datasets
     • $D_{MangoLeafBD} \leftarrow$ Load Dataset (MangoLeafBD)
     • $D_{SenMangoFruitDDS} \leftarrow$ LoadDataset (SenMangoFruitDDS)
II.  Define Model Architectures
     • ConvNeXt model:
         ■ $M_{ConvNeXt}$ = DefineModel(ConvNeXt)
     • ViT model
         ■ $M_{ViT}$ = DefineModel (ViT)
III. Fuse Models
     • Extract features from both models
         ■ $F_{ConvNeXt}$ = ExtractFeatures ($M_{ConvNeXt}$, $D_{train}$)
         ■ $F_{ViT}$ = ExtractFeatures ($M_{ViT}$, $D_{train}$)
     • Fuse features
         ■ $F_{fused}$ = Fuse ($F_{ConvNeXt}$, $F_{ViT}$)
IV.  Train Fusion Model
     • Train fusion model using fused features
         ■ $M_{Fusion}$ = Train (FusionModel, $F_{fused}$)
```

Table 2: Parameters for ConvNext and ViT architecture

Parameter	Description	Parameter	Description
Horizontal and vertical flip	True	Loss function	Sparse Categorical Cross-Entropy
Batch size	32	Dropout	0.45
Input shape	(224, 224, 3)	Dense layer	Softmax
Optimizer	Adam	Activation function	GELU
Learning rate	0.001	Regularization	L1 & L2 regularization

4 Experimental Results

This section discusses the experimental results obtained through the proposed model from both the datasets.

4.1 Results of MangoLeafBD

During training on the MangoLeafBD dataset, the model adjusts its weights to minimize the loss function, and its performance is monitored using the EarlyStopping callback. This stops the training if the validation loss does not improve for 5 consecutive epochs, preventing overfitting and ensuring that the best model is saved. After training, the model is ready to make predictions on new images, outputting a probability distribution for each class and selecting the class with the highest probability as its prediction. Combining data augmentation, transfer learning, careful regularization, and optimization strategies, this approach ensures that the model is both effective and generalizes well to new data. The graph in Fig. 9 illustrates the progression of training and validation accuracy across 13 epochs of model training. The training accuracy is depicted in red, while the validation accuracy is shown in blue. The model demonstrates strong performance from the first epoch, with training accuracy starting at 97.50% and validation accuracy at 98.25%. Both metrics show a consistent upward trend as the epochs progress, indicating that the model is learning effectively from the data. By the third epoch, the validation accuracy surpasses 99%, and this trend continues, peaking at 99.87% by the 13th epoch. The training accuracy increases steadily, reaching 99.03% by the 13th epoch. The close alignment between the training and validation accuracy curves suggests that the model generalizes well, meaning it performs equally well on unseen data as on the training data. The absence of a significant gap between the two lines indicates that the model is not overfitting, thanks to regularization techniques and dropout layers.

The confusion matrix illustrates in Fig. 10 to validate the performance of the classification model on the mango leaf disease dataset. The proposed model demonstrates excellent accuracy across most disease classes. Specifically, it correctly identified 100% of the cases for Anthracnose, Bacterial Canker, Cutting Weevil, Die Back, Gall Midge, Healthy, and Sooty Mould, indicating no false positives or false negatives in these categories. However, for the Powdery Mildew class, the model correctly predicted 96% of the cases, with a slight misclassification rate of 4%, where these cases were possibly confused with another disease. Despite this minor error, the model shows a strong ability to accurately classify different mango leaf diseases, with near-perfect performance in almost all categories.

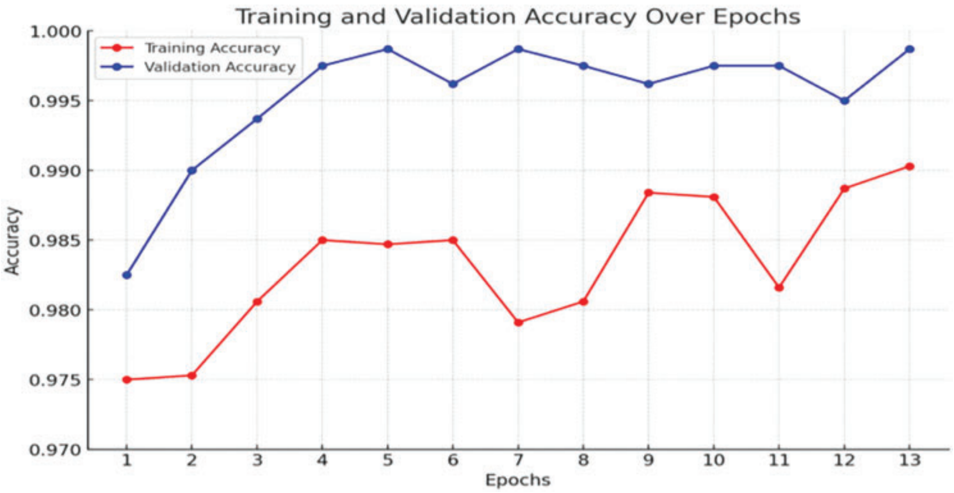
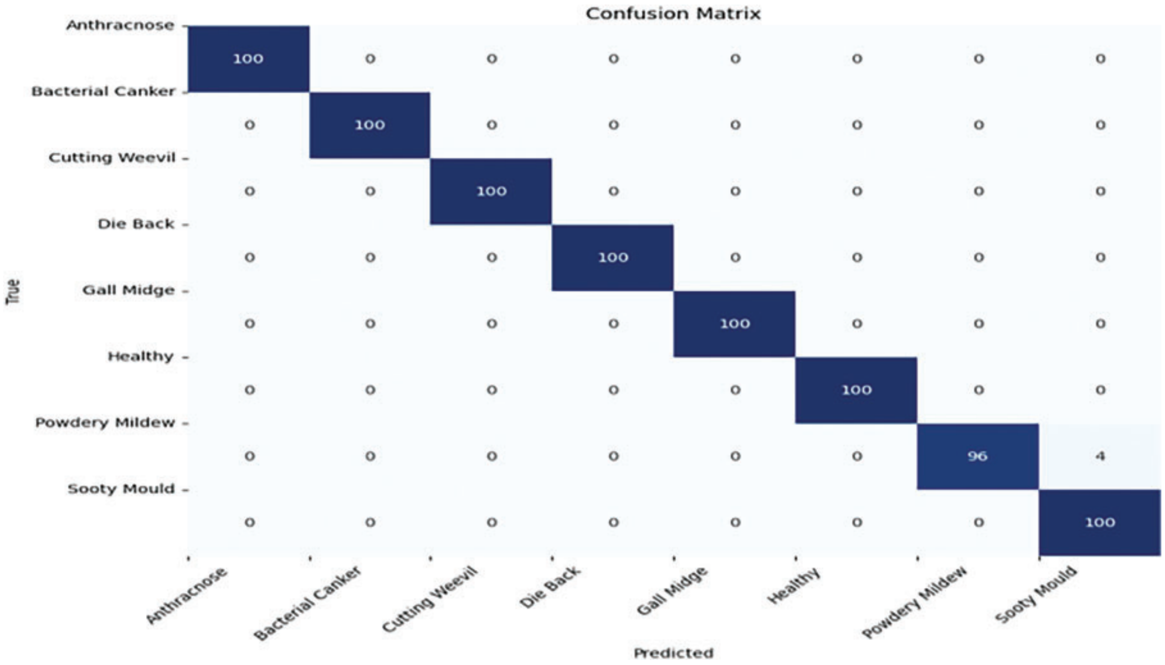


Figure 9: Training and validation graph for MangoLeafBD



True \ Predicted	Anthracnose	Bacterial Canker	Cutting Weevil	Die Back	Gall Midge	Healthy	Powdery Mildew	Sooty Mould
Anthracnose	100	0	0	0	0	0	0	0
Bacterial Canker	0	100	0	0	0	0	0	0
Cutting Weevil	0	0	100	0	0	0	0	0
Die Back	0	0	0	100	0	0	0	0
Gall Midge	0	0	0	0	100	0	0	0
Healthy	0	0	0	0	0	100	0	0
Powdery Mildew	0	0	0	0	0	0	96	4
Sooty Mould	0	0	0	0	0	0	0	100

Figure 10: Confusion matrix obtained from MangoLeafBD

4.2 Results of SenMangoFruitDDS

The analysis of SenMangoFruitDDS of the proposed model yields insightful results, as visualized in the Fig. 10 using the t-Distributed Stochastic Neighbor Embedding (t-SNE) plots. In the t-SNE visualization of the ConvNeXt model as shown in Fig. 11a, data points representing different mango leaf disease classes form distinct clusters, demonstrating the model's effective feature extraction and separation capabilities. The Alternaria class, depicted in blue, forms a distinct cluster, indicating that the ConvNeXt model successfully identifies and differentiates features specific to Alternaria. The Anthracnose class, shown in orange, is slightly more dispersed but still mostly forms a coherent group, illustrating good feature extraction by the model. Black Mould Rot, represented in red, shows a tight cluster, signifying clear differentiation by the ConvNeXt model. The Healthy class, in green, forms tightly clustered points, indicating the model's high accuracy in recognizing healthy leaves. Lastly, the Stem and Rot class, shown in purple, is also effectively distinguished by the model, with data points forming a separate cluster. These results demonstrate the ConvNeXt model's robust performance in distinguishing between various classes of mango leaf diseases. Fig. 11b depicts the t-SNE plot for the ViT model, illustrating the feature space it has learned. The Alternaria class, shown in blue, forms a visible cluster, although there is slightly more overlap with other classes than the ConvNeXt model. The Anthracnose class, represented in orange, has more scattered points, indicating less precise feature separation.

In contrast, the Black Mould Rot class, depicted in red, is well-clustered, demonstrating good recognition by the ViT model. The Healthy class, shown in green, forms a distinct, tight cluster, similar to the results from ConvNeXt, indicating accurate recognition of healthy leaves. Lastly, the Stem and Rot class, represented in purple, has points grouped, showing effective.

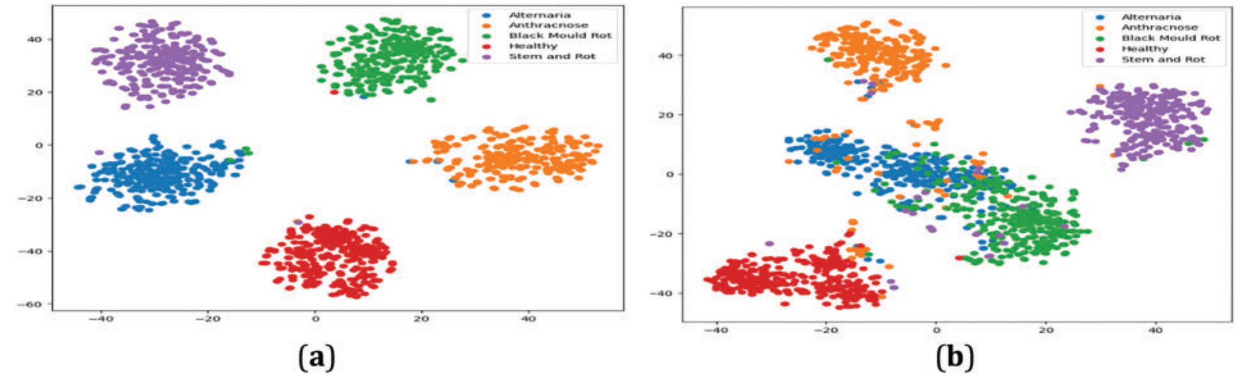


Figure 11: (**a**) t-SNE using ConvNeXt (**b**) t-SNE using ViT

Fig. 12 presents the confusion matrix for the proposed model's classification performance on the SenMangoFruitDDS dataset. The matrix displays the number of instances correctly and incorrectly classified for each disease class. For Alternaria, the model correctly identified 205 instances, with only three misclassified as Anthracnose, six as Black Mould Rot and one as Stem and Rot, showing good performance in distinguishing this class. In the case of Anthracnose, 199 instances were correctly classified, but three were misclassified as Alternaria and two as Black Mould Rot, indicating some overlap between these classes. The Black Mould Rot class showed strong performance, with 190 correctly classified instances, only three misclassified as Alternaria and one as Stem and Rot. The model accurately identified Healthy leaves, with 187 correct classifications and no misclassifications into other classes. Similarly, for Stem, anorrectly identified 199 instances, with just 1 for Stem and Rot misclassified as Black Mould Rot.

4.3 Generalization Experiment

To provide more insights into how well the model generalizes beyond the training and validation sets, we conducted additional experiments to evaluate its performance on completely unseen data. Specifically, we tested the model on the Mango Leaf Disease Identification Dataset (MLDID) [25], which consists of images from five distinct classes: healthy leaves, gall midge, dieback, bacterial cancer, and anthracnose disease. In these experiments, the model achieved an accuracy of 94.8%. The dataset includes images captured under diverse environmental conditions, further highlighting the model's robustness. However, minor misclassifications in certain disease categories suggest opportunities for future improvement through ensemble models or advanced segmentation techniques [26,27].

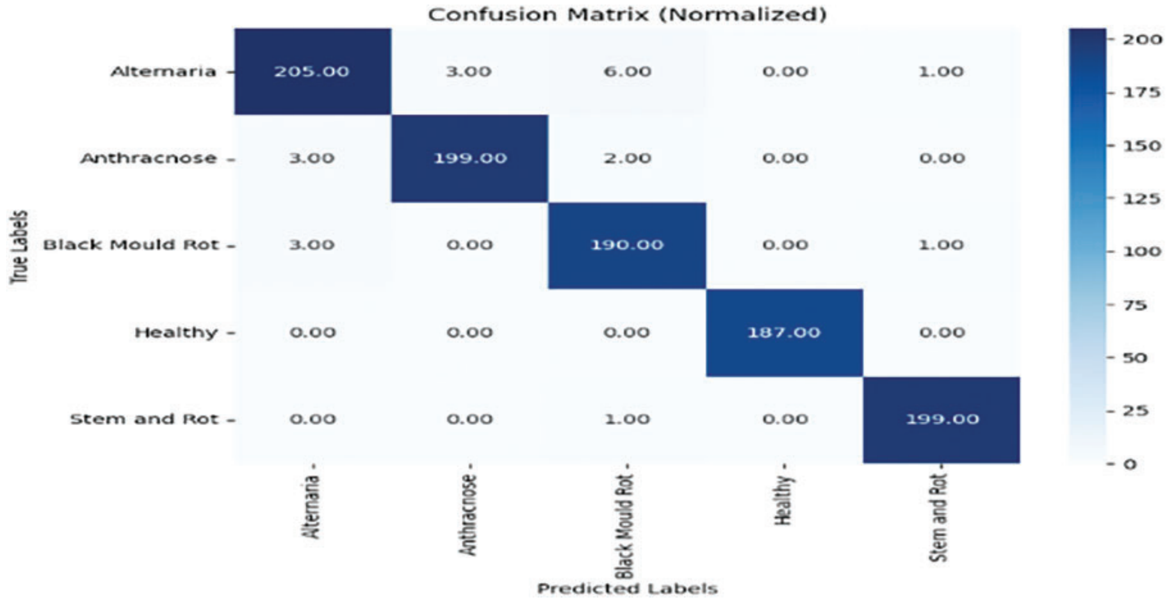


Figure 12: Confusion matrix obtained from SenMangoFruitDDS

5 Discussion

A significant improvement in the classification of mango fruit and leaf diseases is provided by the Fused ViT-ConvNeXt architecture, owing to its unique combination of convolutional and transformer-based systems. The hybrid balances ConvNeXt's ability to extract local features with ViT's strength in modeling the global context. This enables the architecture to detect small and unusual local features, such as disease-specific spots or lesions, while considering features' overall shape and spatial distribution. This capability is crucial for distinguishing diseases with similar manifestations. One of the main merits of this architecture is its high-level classification performance, even with limited data augmentation. Unlike previous models that relied heavily on extensive augmentation, the Fused ViT-ConvNeXt achieves robust results with minimal augmentation, reducing susceptibility to overfitting and ensuring reliability in real-world applications.

The capability of a single model to detect diseases in both mango leaves and fruits is influenced by factors such as dataset composition, feature diversity, and model design. In this study, the datasets for mango leaf diseases (e.g., MangoLeafBD) and mango fruit diseases (e.g., SenMangoFruitDDS) were curated separately, each focusing on the unique visual features specific to the respective plant structures. Mango leaves and fruits exhibit significantly different morphological and pathological characteristics. Leaf diseases often present as patterns such as necrotic spots, discoloration, or structural deformations, while fruit diseases typically manifest as lesions, rotting, or soft patches caused by fungal invasion. These differences necessitate distinct feature extraction and learning capabilities, making it challenging to train a single unified model to effectively handle both diseases without compromising accuracy. To address this challenge, the same Fused ViT-ConvNeXt model was trained separately for detecting diseases in mango leaves and fruits. While the architecture remained consistent, the datasets for MangoLeafBD and SenMangoFruitDDS were processed and utilized independently. This independent training approach allowed the model to optimize its performance for the unique traits of each dataset. By isolating the learning tasks, the model could better handle the specific challenges of each category, such as morphological diversity, environmental factors, and variations in pathological symptoms. The performance metrics presented in Table 3 further highlight the model's effectiveness. For the MangoLeafBD dataset, the model achieves precision, recall, and F1-score values of 99.5%, showcasing its exceptional ability to accurately detect and classify leaf diseases. For the SenMangoFruitDDS dataset, the model achieves 98.0% across all metrics, demonstrating strong generalization to diverse fruit disease conditions. The slight reduction compared to MangoLeafBD reflects the inherent challenges of fruit disease classification, such as overlapping symptoms and environmental variations. These results validate the Fused ViT-ConvNeXt architecture as a highly effective solution for smart agriculture, capable of providing reliable and efficient disease detection across multiple datasets.

Table 3: Performance measures

Dataset	Precision	Recall	F1-score
MangoLeafBD	99.5	99.5	99.5
SenMangoFruitDDS	98	98	98

The proposed model for detecting mango leaf disease is compared with the most recent cutting-edge techniques to show the efficacy of the model as presented in Table 4. Utilizing Lightweight ConvNet architecture, the study in [28] achieved a 98.00% accuracy rate. In contrast, visual Modulation Networks in [1] improve feature representation and attain a greater accuracy of 99.23%. Using a conventional deep convolutional neural network (Deep CNN), the approach in [29] achieves 99.55% accuracy, marginally surpassing DenseNet in [30], which achieves 99.44%. All of these techniques are outperformed by the proposed ConvNeXt model, which has a maximum accuracy of 99.87%. This indicates how well Fused ViT-ConvNeXt sophisticated design detects mango disease, making it a promising solution for enhancing smart agricultural practices.

Table 4: Comparison with recent work for mango leaf disease

Ref.	Method	Dataset	Accuracy
[28]	Lightweight ConvNet	MangoLeafBD	98.00
[1]	Visual Modulation Networks		99.23
[29]	Deep CNN		99.55
[30]	DenseNet		99.44
Proposed	Proposed Fused ViT-ConvNeXt		99.87

Similarly, Table 5 compares the proposed model's performance with recent works in mango fruit disease classification, highlighting differences in methods, datasets, and accuracy. The Ref. [31] used ResNet50 and achieved 98.20% accuracy on the SenMangoFruitDDS dataset comprising 37,432 samples. However, this method relied on extensive data augmentation, which could introduce bias into the test set and inflate accuracy metrics. Similarly, the lightweight convolutional neural network (LCNN) in the same study also used the SenMangoFruitDDS dataset with identical augmentation practices, achieving a lower accuracy of 95.25%, likely due to its simpler architecture. In contrast, the proposed method, which fuses and ConvNeXt architectures, demonstrated superior performance with an accuracy of 98.40%. Notably, this was achieved using a significantly smaller dataset of 7500 samples, approximately seven times less augmentation compared to previous works, thereby avoiding the risk of overfitting or test set bias. Despite limited data and augmentation, the proposed model's high accuracy highlights its ability to effectively extract and generalize features, outperforming both ResNet50 and LCNN while ensuring scalability and efficiency. This demonstrates the robustness of the Fused ViT-ConvNeXt architecture in tackling mango fruit disease classification.

Table 5: Comparison with recent work for mango fruit disease

Ref.	Method	Dataset/total samples		Accuracy
[31]	ResNet50	SenMangoFruitDDS	37,432 (More Augmentation can lead to biased test set)	98.20
[31]	LCNN		37,432 (More Augmentation can lead to biased test set)	95.25
Proposed	Proposed Fused ViT-ConvNeXt		7500 (7 Timeless Augmentation)	98.40

6 Conclusion

In conclusion, this study demonstrates the effectiveness of deep learning models, specifically ConvNeXt and ViT, in diagnosing mango diseases with high accuracy. The proposed fusion model effectively utilizes the strengths of convolutional and transformer-based systems to achieve exceptional performance in classifying mango plant diseases. The model achieved a remarkable accuracy of 99.87% on the MangoLeafBD dataset and 98.40% on the SenMangoFruitDDS dataset, indicating its robustness and generalizability across different types of disease data. These results underscore the model's capability to enhance both the efficiency and accuracy of disease diagnosis in mango farming. Farmers can benefit from instant disease detection by integrating this model into a mobile application, which provides timely and actionable insights. This can significantly reduce the reliance on traditional, labor-intensive methods while empowering farmers especially those in small-scale and resource-limited settings with tools to make quicker and better-informed decisions. Consequently, this contributes to minimizing crop losses, improving yield quality, and boosting overall agricultural productivity.

One of the primary limitations of the study is the challenge of accurately diagnosing mango fruit diseases due to variations in fruit size, shape, and surface texture, which can impact disease detection accuracy. Future work could address this limitation by incorporating segmentation techniques to better isolate and detect diseased regions, improving the accuracy of fruit disease diagnosis. Moreover, future research could expand the scope to include more diverse datasets, capturing various environmental conditions, disease stages, and crop types. This would further enhance the model's generalizability and robustness in real-world applications.

Acknowledgement: This research was supported by Princess Nourah bint Abdulrahman University , Riyadh, Saudi Arabia and Researchers Supporting Project. The authors would like to acknowledge the support of AIDA Lab CCIS Prince Sultan University, Riyadh Saudi Arabia for APC support.

Funding Statement: This research was funded by Princess Nourah bint Abdulrahman University and Researchers Supporting Project number (PNURSP2025R346), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Faten S. Alamri and Amjad Khan; data collection: Raja Atif, Faten S. Alamri, and Ahmed S. Almasoud; analysis and

interpretation of results: Tariq Sadad and Ahmed S. Almasoud; draft manuscript preparation: Amjad Khan, Tariq Sadad, and Raja Atif. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in Mendeley Data at https://data.mendeley.com/datasets/hxsnvwty3r/1 (accessed on 15 November 2024) and https://data.mendeley.com/datasets/jvszp9cbpw/4 (accessed on 17 November 2024).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Ali Salamai A. Enhancing mango disease diagnosis through eco-informatics: a deep learning approach. Ecol Inform. 2023;77(12):102216. doi:10.1016/j.ecoinf.2023.102216.
2. Shah M, Naik S. Identification of artificially ripen mango using aroma and texture features. In: 2021 International Conference on Intelligent Technologies (CONIT); 2021 Jun 25–27; Hubli, India.
3. Liu B, Xin Q, Zhang M, Chen J, Lu Q, Zhou X, et al. Research progress on mango post-harvest ripening physiology and the regulatory technologies. Foods. 2023;12(1):173. doi:10.3390/foods12010173.
4. Kumar S, Gupta B, Chhabra M, Garg L. Comparative evaluation of mango fruit diseases classification using machine learning. In: 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP); 2024 May 25–26; Sonipat, India.
5. Dofuor AK, Quartey NK, Osabutey AF, Antwi-Agyakwa AK, Asante K, Boateng BO, et al. Mango anthracnose disease: the current situation and direction for future research. Front Microbiol. 2023;14:1168203. doi:10.3389/fmicb.2023.1168203.
6. Gupta S, Tripathi AK. Fruit and vegetable disease detection and classification: recent trends, challenges, and future opportunities. Eng Appl Artif Intell. 2024;133:108260. doi:10.1016/j.engappai.2024.108260.
7. Xu X, Yang G, Wang Y, Shang Y, Hua Z, Wang Z, et al. Plant leaf disease identification by parameter-efficient transformer with adapter. Eng Appl Artif Intell. 2024;138(1):109466. doi:10.1016/j.engappai.2024.109466.
8. Veling PS, Kalelkar RS, Ajgaonkar LV, Mestry NV, Gawade NN. Mango disease detection using image processing. Int J Res Appl Sci Eng Technol. 2019;7(4):3717–26. doi:10.22214/ijraset.2019.4624.
9. Kusrini K, Suputa S, Setyanto A, Agastya IMA, Priantoro H, Chandramouli K, et al. Data augmentation for automated pest classification in Mango farms. Comput Electron Agric. 2020;179(1):105842. doi:10.1016/j.compag.2020.105842.
10. Faye D, Diop I, Dione D. Mango diseases classification solutions using machine learning or deep learning: a review. J Comput Commun. 2022;10(12):16–28. doi:10.4236/jcc.2022.1012002.
11. Mia MR, Roy S, Das SK, Rahman MA. Mango leaf diseases recognition using neural network and support vector machine. Iran J Comput Sci. 2020;3(3):185–93. doi:10.1007/s42044-020-00057-z.
12. Aldossary M, Alharbi HA, Anwar Ul Hassan C. Internet of Things (IoT)-enabled machine learning models for efficient monitoring of smart agriculture. IEEE Access. 2024;12(2):75718–34. doi:10.1109/ACCESS.2024.3404651.
13. 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. 2023;47(6):108941. doi:10.1016/j.dib.2023.108941.
14. Malik A, Vaidya G, Jagota V, Eswaran S, Sirohi A, Batra I, et al. Design and evaluation of a hybrid technique for detecting sunflower leaf disease using deep learning approach. J Food Qual. 2022;2022:9211700. doi:10.1155/2022/9211700.
15. Rizwan Iqbal HM, Hakim A. Classification and grading of harvested mangoes using convolutional neural network. Int J Fruit Sci. 2022;22(1):95–109. doi:10.1080/15538362.2021.2023069.
16. Saleem R, Shah JH, Sharif M, Yasmin M, Yong HS, Cha J. Mango leaf disease recognition and classification using novel segmentation and vein pattern technique. Appl Sci. 2021;11(24):11901. doi:10.3390/app112411901.

17. Chen YC, Wang JC, Lee MH, Liu AC, Jiang JA. Enhanced detection of mango leaf diseases in field environments using MSMP-CNN and transfer learning. Comput Electron Agric. 2024;227(2):109636. doi:10.1016/j.compag.2024.109636.
18. Nandi CS, Tudu B, Koley C. A machine vision technique for grading of harvested mangoes based on maturity and quality. IEEE Sens J. 2016;16(16):6387–96. doi:10.1109/JSEN.2016.2580221.
19. Patel KK, Kar A, Khan MA. Common external defect detection of mangoes using color computer vision. J Inst Eng Ind Ser A. 2019;100(4):559–68. doi:10.1007/s40030-019-00396-6.
20. Sahu D, Potdar RM. Defect identification and maturity detection of mango fruits using image analysis. Am J Artif Intell. 2017;1(1):5–14.
21. Kalaivani R, Saravanan A. Automated detection and classification of mango fruit diseases using a novel WOA-QRNN technique on infected mango fruit images through transfer learning. Int J Intell Syst Appl Eng. 2023;12(3):3755–63.
22. Ramos L, Casas E, Romero C, Rivas-Echeverría F, Morocho-Cayamcela ME. A study of ConvNeXt architectures for enhanced image captioning. IEEE Access. 2024;12:13711–28. doi:10.1109/ACCESS.2024.3356551.
23. Ali AM, Benjdira B, Koubaa A, El-Shafai W, Khan Z, Boulila W. Vision transformers in image restoration: a survey. Sensors. 2023;23(5):2385. doi:10.3390/s23052385.
24. Faye D, Diop I, Mbaye N, Diedhiou MM, Dione D. SenMangoFruitDDS. [cited 2025 Jan 1]. Available from: https://SenMangoFruitDDS-MendeleyData.
25. Rahman MM, Kowser KA, Islam M, Arefin MN, Shah A, et al. Mango leaf disease identification dataset: (MLDID); [cited 2025 Jan 1] Available from: https://MangoLeafDiseaseIdentificationDataset:(MLDID)-MendeleyData.
26. Khan MA, Akram T, Sharif M, Alhaisoni M, Saba T, Nawaz N. A probabilistic segmentation and entropy-rank correlation-based feature selection approach for the recognition of fruit diseases. EURASIP J Image Video Process. 2021;2021(1):14. doi:10.1186/s13640-021-00558-2.
27. Khan MA, Akram T, Sharif M, Awais M, Javed K, Ali H, et al. CCDF: automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features. Comput Electron Agric. 2018;155(1):220–36. doi:10.1016/j.compag.2018.10.013.
28. Mahbub NI, Naznin F, Hasan MI, Shifat SMR, Hossain MA, Islam MZ. Detect Bangladeshi mango leaf diseases using lightweight convolutional neural network. In: 2023 International Conference on Electrical, Computer and Communication Engineering; 2023 Feb 23–25; Chittagong, Bangladesh.
29. 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. 2023;14(6):100787. doi:10.1016/j.jafr.2023.100787.
30. Mahmud BU, Al Mamun A, Hossen MJ, Hong GY, Jahan B. Light-weight deep learning model for accelerating the classification of mango-leaf disease. Emerg Sci J. 2024;8(1):28–42. doi:10.28991/ESJ-2024-08-01-03.
31. Faye D, Diop I, Mbaye N, Dione D, Diedhiou MM. MangoFruitDDS: a standard mango fruit diseases dataset made in Africa. In: The International Conference on Advanced Research in Technologies, Information, Innovation and Sustainability; 2023 Oct 18–20; Madrid, Spain. Cham, Switzerland: Springer.