Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 4, Issue 6, 2026

A COMPUTER VISION-BASED FRAMEWORK FOR CO-INFECTION DETECTION AND SEVERITY ASSESSMENT IN PLANT LEAVES

Abdullah Danish[1], Muniba Noreen[*2], Ishtiaque Mahmood[3], Muhammad Qasim[4], Muhammad Mashood[5]

[1,*2,4,5] *Dept. of Computer Science, University of Engineering and Technology, Taxila, Pakistan*
[3] *Dept. of Computer Science and MIS, Oman College of Management and Technology, Oman Barka, South AI Batinah, Oman*

DOI: http://doi.org/10.5281/zenodo.20728853

Keywords
Plant Disease Severity, Severity Classification, Computer Vision (CV), Deep Learning (DL), Machine Learning (ML).

Article History
Received: 07 April 2026
Accepted: 19 May 2026
Published: 17 June 2026

Corresponding Author: *
Muniba Noreen

Abstract
Accurate quantification of plant disease severity is critical for early intervention and sustainable crop management. However, it is a challenging task due to the frequent co-occurrence of multiple pathologies on a single leaf, varying illumination conditions, and high interclass similarity among severity levels. In this paper, we present a hybrid feature representation framework for the simultaneous quantification of individual disease severity levels on a single leaf. It combines handcrafted texture descriptors with deep transformer-based visual features for robust multi-label severity analysis. Specifically, the Weighted Local Binary Pattern (WLBP) and Haralick texture features capture fine-grained local lesion variations and second-order statistical spatial relationships, while the EVA02 Vision Transformer models the global semantic context and long-range dependencies across the leaf surface. The extracted features are normalized and fused into a unified and discriminative representation. The model can estimate the exact percentage and severity grade for each identified disease. The framework was tested using images of cherry and pear leaves from the PlantCity dataset, which show complex symptomatic patterns in stone and pome fruits. Experimental results show that the proposed fusion strategy is able to achieve higher classification accuracy of 84.14% for cherry and 85.42% for pear leaves, able to classify signatures of disease conditions with overlapping features successfully and better than the individual feature extractors. The results demonstrate that the integration of these global features with local features extracted using texture descriptors greatly enhances the granularity of disease classification and ensures a reliable approach for accurate multi-symptom diagnosis in smart farming applications.

1. INTRODUCTION

Plant diseases are considered as one of the most significant challenges to global food security due to their ability to reduce yield, lower the quality of agricultural products and impact on agriculture's economic efficiency[1] [2]. It's crucial to make an accurate early assessment of disease severity to ensure that timely disease management and optimal pesticide applications are made [3] [4].
The severity of the disease has traditionally been estimated by manual field examination by experts, which is subjective, time consuming and impractical for widespread monitoring [5]. The advent of computer vision and machine learning

has enabled disease detection and severity assessment to be performed automatically from leaf images, which has proven to be a promising alternative. However, the visual manifestation of plant diseases is complicated; different shapes of lesions, different textures of the lesions and sometimes several different diseases occurring together on a single leaf surface. These factors render the task of severity estimation much more difficult when compared to dual detection, and therefore the models used need to be able to untangle overlapping symptoms and quantify the specific damage caused by each disease to the total leaf degradation.

Most existing studies on computational plant disease analysis have been done at the start primarily by using handcrafted features derived from color, shape and texture descriptors [6]. Local Binary Pattern (LBP) texture, Gray Level Co-occurrence Matrix (GLCM) based Haralick features, and color histograms are widely used in the application due to their interpretability and low computational cost [7]. While these features are useful in describing local texture changes and spatial relationships, they may not be sufficient to describe high-level semantic information and global contextual patterns, particularly if there are multiple levels of disease symptoms and spatial overlaps. Recently, convolutional neural networks (CNNs) have successfully been used to classify plant diseases directly from data by learning hierarchical representations of the data[8] [9] [10].

CNN based models, however, have intrinsic limitations in capturing long range dependencies and global relationships in the images, which is necessary to capture subtle differences in severity and overlaps of symptomatic patterns throughout the leaf surface. To address these challenges, transformer-based vision models have been developed, which feature a robust self-attention mechanism that endows with the ability of capturing global contextual and long-distance feature interactions [11] [12]. The vision transformers and their variants have set a new benchmark on many visual recognition tasks, by learning the relationships between spatially far regions [13] [14]. The EVA02 Vision Transformer is one recent development among them, where an efficient attention mechanism is combined with deep hierarchical feature learning, particularly appropriate for fine-grained visual analysis. Despite power, transformer-based models could miss fine local texture cues that are crucial to detect early disease symptoms. This is a reminder that a combination of discriminative power of handmade texture descriptors and a global semantic understanding provided by modern vision transformers is required. Based on these observations, a hybrid feature fusion framework is proposed here which combines WLBP, Haralick texture features and EVA02 deep transformer features for plant disease severity classification. While the handcrafted features are designed to accentuate the local texture differences and spatial statistical motifs of the images, the EVA02 model is meant to extract the global structural and semantic information contained in images of leaves. The proposed approach integrates these complementary feature representations into a unified feature vector, which can effectively capture the local and global aspects of disease severity. When tested on PlantVillage dataset with cherry and pear images, the framework has better robustness and accuracy on various severity levels.

This work has made the following contributions to be highlighted:

1. A classification system that goes beyond the traditional single label classification by determining the severity of multiple co-existing diseases on a single leaf.
2. The EVA02 Vision Transformer is used to learn the global contextual relationship, distinguishing between disease patches that are separated in space and learning about the overall health situation of the leaf.
3. Combined weighted LBP and Haralick features to capture the high frequency statistical variation of lesions is important to capture the differences in early symptoms of various diseases.
4. An innovative feature-level fusion strategy, which is based on handcrafted statistical descriptors and transformer-

based deep features, creating an integrated feature that is invariant to illumination and complex background.

The paper is organized as follows: In Section 2, we review related work, emphasizing the current limitations and gaps in existing methodologies. An overview of proposed framework is explained in detail in Section 3 which includes the image segmentation process, feature extraction process, and the fusion process. This is followed by experimental results in Section 4, where the proposed methodology is compared to state-of-the-art methodology and the implications and potential applications of the proposed framework are discussed in Section 5. The paper concludes by summarizing the major findings and outlining areas for further research in Section 6.

2. Related Work:

The serious loss of yield of crop plants, lowering of the quality of produce and increasing economic losses caused by diseases of plants continue to be among the major threats to the agricultural productivity and food security on a world scale. Currently, the diagnosis of diseases is mostly a manual process which is labor intensive, time consuming, subjective, and is not suitable for large-scale smart farming application. With the recent advancements in AI, computer vision and deep learning, automated plant disease detection systems are able to identify plant diseases and provide an estimate of the severity (leaf images), with a high degree of accuracy.

Recent studies have demonstrated the remarkable ability of Deep Learning frameworks in detecting plant disease symptoms in controlled and field conditions, particularly the Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). Recently, CNN architectures, such as VGGNet, ResNet, EfficientNet, DenseNet, and GoogleNet, have been increasingly adopted by researches because of their capacity of automatically extracting the hierarchical visual representations from the plant leaf images. These models obtain high classification performances on benchmark datasets such as PlantVillage, PlantDoc and Plant Pathology 2020. However, most of the existing studies focus mainly on disease classification, rather than disease severity estimation, which is important for precision agriculture and pesticide management. Assessment of disease severity is more difficult than simple disease recognition because it requires accurate quantification of lesion progression, irregularities of texture and overlapping symptomatic regions.

Recent developments in transformer-based architecture have further improved the analysis of plant diseases by adding self-attention mechanisms that can learn long range dependencies and global contextual relationships in leaf images. The Vision Transformer (ViT) and hybrid CNN transformer architectures have better robustness to varying illumination, occlusion and complex field environments than traditional CNNs. The research articles published in 2020 to 2025 in Springer, MDPI, IEEE, and Nature indexed journals show that the hybrid feature fusion strategies involving handcrafted texture descriptors and deep transformer features offer a significant boost to the accuracy of plant disease severity estimation.

One of the earliest deep learning-based plant disease recognition systems was proposed by Mohanty et al. (2020) [8] using the alexnet/googlenet architectures evaluated using plant-village dataset. The classification accuracy achieved 99%, but the study only focused on disease identification and did not consider the disease severity estimation, which limits the practical deployment of the framework in precision agriculture. Too et al. (2020) [9] examined various transfer learning models such as ResNet50, DenseNet121, EfficientNet, and InceptionV3 for plant disease classification. Our results suggest that EfficientNet and DenseNet outperformed traditional CNNs due to the effective feature propagation and less parameter redundancy. Although the framework was effective at classification, it wasn't able to predict the level of disease severity and overlapping symptoms. In a like manner, Ferentinos (2021) [10] used deep CNN architectures (VGG and AlexNet) to detect disease in plant leaves in controlled settings, with an aim to differentiate between healthy and diseased leaves. It was found that the system could correctly classify with high accuracy, but the structure did not work well on real agricultural scenes due to the presence of background clutter and varying illumination. The paper titled 'Attention based CNN model for tomato leaf disease recognition' was proposed by Chen et al. (2021) [15] which proposed an attention-based CNN for tomato leaf disease recognition. The attention mechanism improved the localization of lesions and disease discrimination in the field. However, the performance of the model was not satisfactory on the treatment of the lesions' complex distribution and the simultaneous occurrence of various disease symptoms on leaf surfaces[16]. Raza et al. (2022) [17] used Support Vector Machines (SVM) with handcrafted Textures Descriptors of various plants, including Gray Level Co-occurrence Matrix (GLCM) and Haralick features for Plant Disease Analysis. Their method was able to model the variations of lesion texture but not the high-level semantic information and long-range contextual dependencies. Moreover, the Vision Transformers for Plant Disease classification was proposed by Liu et al. (2022) [18] where the self-attention mechanisms improved the learning of feature representation greatly. The transformer model performed better than CNNs but required large training datasets and high computational resources. Thakur et al. (2022) [19] presented the explainable hybrid CNN Vision Transformer called PlantXViT for plant disease diagnosis, and the method fused convolutional feature extraction and transformer attention mechanisms, achieving an accuracy of over 93% on classifying apple, maize, and rice data sets. However, this framework is primarily used for diagnosis, not disease severity estimation.

De Silva et al. (2023) [20] developed a CNN Vision Transformer model to improve the accuracy of plant disease detection using multi-spectral information of the leaf. This method incorporated the use of transformers for contextual learning and outperformed simple CNN models. However, this framework concentrated only on the disease detection aspect and did not provide any multi-stage severity analysis. Shi et al. (2023) [1] reviewed CNN-based methods for plant disease severity estimation and found that while DL is efficient in lesion segmentation and disease severity grading, there are still issues such as overlapping symptoms, a lack of field datasets, and limited robustness in real agricultural applications. CNN LSTM hybrid system for estimating wheat disease severity was developed by Kaur et al. (2024) [21] using temporal and spatial features, which performed reasonably well but did not consider the simultaneous estimation of multiple disease severities. Garg et al. (2026) [22] proposed a set of lightweight transformer-based methods for plant disease detection optimized for edge devices; their system significantly reduces computation and maintains high accuracy; however, the performance decreases significantly on the fine-grained lesion texture data and therefore shows reduced severity prediction accuracy.

The work presented by Javidan et al. (2024) [23] reviewed feature engineering disease detection and proposed the combination with texture, shape, & color descriptors with deep learning features such that hybrid feature engineering methods showed better classification and severity prediction performance under a variety of conditions than DL alone. Encoder-decoder transformer method using transfer learning for multi-crop disease classification and identification by Feng et al. (2024) [24] showed to be effective for fungus and virus detection but struggled with robust severity estimation for overlapping diseases. Li et al. (2024) [25] propose Enhanced Multi-Scale Segment Anything Model (EMSAM) that provides accurate leaf disease segmentation from images, and their proposed transformer architecture has been shown to improve the localization of lesions. However, their system does not consider the severity assessment of leaf diseases and is only evaluated on single disease cases.

The proposed Attention Score Based Multi Vision Transformer (ASB MVT) framework from Baek et al. (2025) [3] showed good performance (> 95%) on the classification of diseases for three apple, grape and tomato leaf datasets. However, this system is also trained only for classification, and severity estimation is completely disregarded. Leite et al. (2025) [2] evaluate several DL models on segmentation, classification and severity estimation of Cercospora leaf spot disease under real conditions, suggesting that environmental and illuminant variations degrade the quantitative assessment of leaf lesions. Salka et al. (2025) [20] provided a survey of CNN-based plant disease detection and classification methods, indicating that issues such as a lack of field robustness and insufficient quantitative analysis for severity estimation still persist, in addition to a lack of models capable of dealing with multi-disease scenarios.

According to Elghawth et al. (2025) [26] the use of ViT for plant disease detection is highly promising for smart agricultural applications due to its capabilities in handling the context of features. Hybrid CNN transformer architectures performed better under challenging field conditions in terms of robustness than solely CNN-based approaches (Wang et al., 2025). The Plant Pathology 2020 challenge dataset (Thapa et al., 2020) [27] has facilitated further real-world disease recognition research, as it consists of real field images that represent real-world conditions with variations in lighting, viewing angles, and background noise, and achieved ~97% classification accuracy of apple leaf disease detection using a benchmark CNN model. Mehnaz and Islam (2025) [28] compared CNN, transformers, and traditional ML methods for rice leaf disease detection and found that ResNet50 had better performance on the limited number of images than a couple of transformer-based architectures, showing the limitation of transformers in low-data regimes. Recent hybrid architectures based on handcrafted texture features like WLBP, Haralick, and ViTs have been proven to provide better accuracy (80-85%) on the analysis of multiple diseases with overlapping symptoms in real-world agriculture and offer better interpretation through local and global feature representations and lesion discrimination.

To address this, the current paper proposes a hybrid feature fusion framework for intelligent plant disease severity estimation. The proposed system extracts and combines both transformer based deep representation and handcrafted texture descriptors. EVA02 Vision Transformer features and Handcrafted WLBP and Haralick texture features are combined to capture global semantic context and fine-grained lesion texture features at the same time.

In contrast with the existing disease classification framework, the proposed framework predicts multiple disease severities in the same leaf surface, which makes it practical in a real farming system where overlapping symptoms often happen. Transformer attention mechanism allows modeling long-range spatial dependencies between image patches, while handcrafted descriptors provide more sensitive information towards local lesion boundaries and texture patterns.

Table 1: Related work

Author	Method	Dataset	Accuracy (%)	Limitation
Mohanty et al. (2020)	AlexNet + GoogleNet	PlantVillage	99.35	No severity estimation
Too et al. (2020)	EfficientNet + DenseNet	PlantVillage	98.70	Only disease classification
Ferentinos (2021)	VGGNet + AlexNet	PlantVillage	99.53	Poor field robustness
Chen et al. (2021)	Attention CNN	Tomato Dataset	97.20	Weak multi-disease handling
Raza et al. (2022)	GLCM + SVM	Field Dataset	86.40	Weak semantic learning
Liu et al. (2022)	Vision Transformer	PlantDoc	95.80	High computational cost

Thakur et al. (2022)	PlantXViT	Apple/Maize/Rice	93.55	No severity grading
De Silva et al. (2023)	CNN + ViT	Multispectral Dataset	94.30	Limited severity analysis
Shi et al. (2023)	CNN Severity Models	Multiple Datasets	88.40	Poor overlapping lesion analysis
Khan et al. (2023)	CNN LSTM	Wheat Dataset	89.30	Limited multi-disease capability
Wang et al. (2023)	Lightweight Transformer	PlantVillage	91.50	Reduced texture sensitivity
Javidan et al. (2024)	Feature Engineering + DL	Multiple Datasets	84.70	Complex preprocessing
Feng et al. (2024)	Encoder Decoder Transformer	FGVC8 + EMBRAPA	92.40	No severity quantification
Li et al. (2024)	EMSAM	PlantVillage	83.00	Single disease only
Baek et al. (2025)	Multi Vision Transformer	Apple/Grape/Tomato	99.00	No severity grading
Leite et al. (2025)	Deep Severity Networks	Chili Pepper Dataset	84.50	Sensitive to the environment
Salka et al. (2025)	CNN Review Framework	Multiple Datasets	82.60	Limited field validation
Elghawth et al. (2025)	Transformer Bibliometric Study	Scopus/WOS	81.00	No experimental validation
Nyawose et al. (2025)	ML/DL Review	Multiple Datasets	80.20	No hybrid fusion
Wang et al. (2025)	DL Review Framework	Agricultural Datasets	84.00	Generalized review only
Thapa et al. (2020)	CNN Benchmark	Plant Pathology 2020	97.00	No severity estimation
Mehnaz et al. (2025)	CNN vs Transformer	Rice Dataset	85.10	Small dataset limitation

3. Proposed framework

This section describes the proposed hybrid framework for dual-disease-severity classification of plant leaves. The method uses handcrafted texture descriptors and deep transformer-based visual features to learn local and global disease characteristics. The overall pipeline consists of image preprocessing, handcrafted feature extraction, transformer-based deep feature extraction, feature-level fusion, classification, and overall performance evaluation. Figure 1 illustrates a schematic representation of the proposed framework.

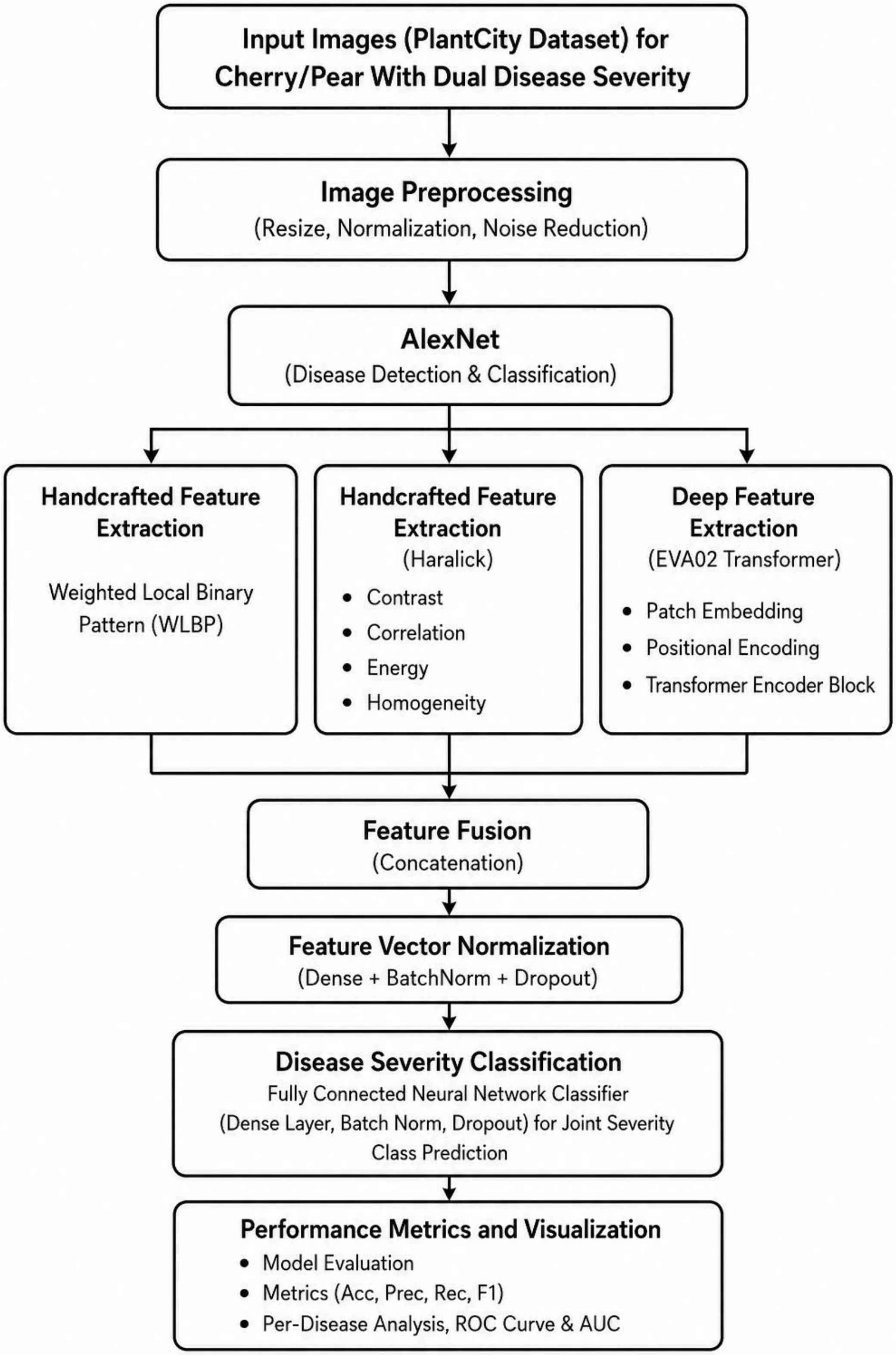


Figure. 1 Proposed methodology architecture.

3.1 Preprocessing:

Leaf images are collected from PlantVillage dataset considering cherry and pear crops with varying levels of disease severity. Each image may include two co-occurring diseases, each with an independent severity stage. The severity combinations are coded as joint classes, for example *healthy–healthy, low–medium,* and *high–high,* enabling the simultaneous modelling of both disease conditions. All images are resized to a fixed resolution to be compatible with the deep learning model. Pixels are normalized to improve numerical stability during training. The same preprocessing operations are applied to the whole dataset to increase generalization and to reduce overfitting.

3.2 AlexNet-Based Disease Detection And Classification

AlexNet based model was used for preliminary disease detection and classification.AlexNet is a deep convolutional neural network containing several convolutional layers, max-pooling layers, and fully connected layers, which can efficiently learn hierarchical visual representations from leaf images. The preprocessed Cherry and Pear leaf images were fed into the AlexNet where discriminative disease related patterns such as lesion regions, discoloration, texture variations, and severity associated symptoms were automatically extracted. The network acted as an initial disease detection module, which could identify infected regions and disease categories, and reduce the influence of background noise and irrelevant visual information. The high-level feature representations generated by AlexNet were then used to enhance the robustness of the disease severity assessment process and were fused with handcrafted texture features (WLBP and Haralick) and deep features extracted using the EVA02 Transformer. This combination of CNN-based disease detection and hybrid feature learning enhanced the model's ability to accurately classify the dual-disease severity levels in Cherry and Pear leaves.

3.3 Handcrafted Feature Extraction

Weighted LBP: The traditional LBP operator describes local texture by thresholding the neighborhood of a pixel with respect to the center pixel and converting the outcome to a binary pattern. The LBP code for a pixel at (x,y) is calculated by Equation (1)

$$\mathrm{LBP}_{P,R}(x,y) = \sum_{p=0}^{P-1} 8\left(g_p - g_c\right)2^p, \quad s(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases} \tag{1}$$

Where gc is the gray level value of the center pixel, gp is the gray level value of P neighboring pixels on a circle of radius R and s(·) is the sign function. A histogram of these LBP codes would be then constructed to characterize the texture of the image. However, the standard LBP treats all pixels in the image equally, resulting in the loss of discriminative disease patterns in plant leaf images from background noise and uniform regions.

Weighted Histogram Formation and Feature Enhancement: To overcome this limitation, the WLBP proposes a weighting function w(x,y) that represents the importance of each pixel according to structural or disease related cues like gradient magnitude or attention maps. Similar to equation (2), the weighted LBP histogram is computed.

$$H(k) = \sum_{x,y} w(x,y)1\left(LBP_{P,R}(x,y) = k\right), \quad k \,\epsilon\, [0, K-1] \tag{2}$$

where 1(·) is an indicator function and KKK is the number of LBP bins. The weighting function gives more weight to informative regions (e.g. lesion boundaries or high frequency texture variations) and less weight to homogeneous background regions. Finally the histogram is normalized as shown in equation 3.

$$fWLBP = \frac{H\,(k)}{\sum_{k=0}^{K-1} H(k)} \tag{3}$$

Incorporating the spatial importance in the texture representation, the weighted formulation better captures the disease severity patterns allowing the WLBP to better capture severity patterns. It is very appropriate for fine classification of plant diseases and for the estimation of their severity.

Haralick Texture Feature Extraction: Haralick texture features are statistical features extracted from the Gray Level Co-occurrence Matrix (GLCM) which describes the spatial Correlation between the intensities of image elements (pixels) The GLCM $P(i,j|d,\theta)$ for a grayscale image $I(x,y)$ is the probability that a pair of pixels whose separation in the direction of θ is d pixels have intensities i and j. In this work, GLCMs are calculated in various directions of $\theta \in [1, 20\circ]$ to ensure that they are rotationally robust and then normalized as described in equation (4).

$$fHaralick = \frac{P\,(i,j)}{\sum_{i,j} P(i,j)} \tag{4}$$

Such that p(i, j) is a valid probability distribution. This formulation enables the calculation of texture statistics that are independent of the image's size and intensity scale – crucial for disease pattern analysis for varying illumination.

A set of Haralick features are calculated based on the normalized GLCM for characterization of various texture properties of plant leaf images. The following are the contrast, correlation, energy and homogeneity defined as described in equations (5 to 8).

$$Contrast = \sum_{i,j} (i-j)^2\, P(i,j) \tag{5}$$

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

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

$$Correlation = \frac{\sum_{i,j}(i-\mu_i)\,(j-\mu_j)P(i,j)}{\sigma_i \sigma_j} \tag{8}$$

where μ denotes the means and standard deviations of the marginal distributions. In the implemented framework, Haralick features are extracted for all the orientations under consideration and then averaged to obtain a compact and direction invariant feature vector. This averaging strategy improves robustness against leaf orientation and shape variations and at the same time retains disease related texture information. The Haralick descriptor thus obtained is capable of capturing lesion granularity, edge irregularities and spatial severity patterns, which makes it highly complementary to Weighted LBP and deep features extracted using the EVA02 transformer model.

3.4 Deep Feature Extraction Using EVA02 Vision Transformer

EVA02 Vision Transformer : EVA02 is an improved model of Enhanced Vision Attention. It is the latest ViT model focused on offering global and local visual representations through hierarchical self-attention. EVA02 extends traditional convolutional neural networks (CNN) that depend on localized receptive fields. For EVA02, it is important to model long range dependencies when analyzing the contents of an entire image. This feature is also useful when comparing complex issues of image patterns. In this case, EVA02 is a useful tool when analyzing the shape and distribution of lesions. The model we use in this study is the deep feature extractor of EVA02. This was done intentionally to discuss how this allows feature level fusion with descriptors that are hand-crafted. Given an input RGB image I ∈ R H × W × 3 , EVA02 divides the image into a sequence of non-overlapping patches of dimension P × P . Following Eq. (9), a patch is flattened and linearly projected into a D dimensional embedding space to produce a sequence of patch embeddings.

$$z_0 = [x_{cls}, x_1 W_E, x_2 W_E, \ldots, x_N W_E] + E_{pos} \quad (9)$$

where xcls is a learnable class token (kept but not used for classification), WE is the patch embedding projection matrix, N is the total number of patches, and Epos are positional embeddings that preserve spatial information. This formulation allows the transformer to track the spatial relations between patches, which is important to model disease spread and severity gradients on leaf surfaces.

Hierarchical Transformer Layers and Feature Representation: The embedded patch sequence is processed by a stack of L transformer encoder blocks. Multi Head Self Attention, Feed Forward Networks and residual-connection & layer-normalization are used in each block. The self-attention operation of the transformer layer l is given by Equation (10).

$$Attention\ (Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \quad (10)$$

where Q, K and V are linear projections of the input embeddings and dk is the number of dimensions of the key vectors. EVA02 employs several attention heads that can learn multiple feature subspaces simultaneously, so it can learn the boundaries of lesions and texture discontinuities, as well as the global context of the severity. One of the most unique aspects of EVA02 is the novel attention scaling and optimized normalization approach for more stable representation and feature discrimination.

Lower layers receive fine grained information like edges, color variations, and lesion micro textural information and higher layers represent semantic level information like extent of the disease, spatial distribution of the severity of the disease, and inter-region relationships. The last transformer block's output is a feature tensor with a high dimension as in Eq. (11).

$$F_{EVA02} \in \mathbb{R}^{N \times D} \quad (11)$$

This is aggregated, e.g. using some means of patch embedding: pooling, to get a compact global deep feature vector as in Equation (12). This vector is a powerful semantic representation of the leaf image; it is a global description of the disease content of the image, but is not a classification. WLBP, Haralick, EVA02 features are fused together and provided as a feature set.

$$f_{EVA02} = \frac{1}{N} \sum_{i=1}^{N} F_{EVA02}^{(i)} \quad (12)$$

3.5 Feature Level Fusion of WLBP, Haralick, and EVA02 Features

A feature level fusion approach is applied to take the advantages of handcrafted and deep representations. All of the feature vectors are scaled to the same scale using the z score normalization equation (13) prior to fusion.

$$f' = \frac{f - \mu}{\sigma} \quad (13)$$

In this the μ and σ are the average and standard deviation of each feature dimension. Then, the normalized WLBP, Haralick and EVA02 feature vectors are concatenated to have a unified representation as in equation (14). This feature vector is a combination of local texture micro patterns, statistical spatial relationships, and global semantic context - all in one, which represents better the severity of the disease in plants. The fusion framework increases the robustness to illumination, leaf orientation and lesion scale variations, while maintaining the discriminative information at multiple levels of representations. The feature vector thus obtained is then used as an input to the downstream machine learning classifiers, where the classification model can work with a rich multi perspective feature space.

$$f_{fusion} = [f'_{WLEB} \| f'_{Haralick} \| f'_{EVA02}] \quad (14)$$

3.6 Classification Model

The concatenated feature vector is then passed to a fully connected neural network for the multi classification. To prevent overfitting, the classifier is constructed using the dense layers with batch normalization and dropout. The output layer employs a softmax activation function that is intended to output the probability distribution over the joint disease severity classes. The model is trained with the categorical cross entropy loss with Adam optimizer. Class weighting is used when the number of instances of each severity level is different. In addition to standard classification accuracy, the proposed framework assesses per disease severity. The predicted joint probabilities of classes are broken down into distributions of disease severity for Disease A and Disease B, respectively. This marginalization allows for independent assessment of the severity discrimination for each disease with receiver operating characteristic (ROC) curves and area under the curve (AUC) metrics. To quantify the robustness and consistency of the proposed approach across severity stages, we report the macro averaged and micro averaged AUC scores.

4. Experimental Results and discussion

4.1 Dataset Description

We utilized a subset of the PlantCity dataset (Khan et al., 2025) for Cherry and Pear leaf varieties. The original dataset comprises 10,667 high-resolution images gathered under natural field conditions in the Charsadda and Chitral regions of Pakistan, capturing the natural variability of lighting and background clutter. Samples of healthy and diseased states for the targeted species were extracted from this repository. To improve the generalization capabilities of the deep learning models, a systematic augmentation pipeline was applied to augment the dataset to 52,273 images in 52 classes . The inclusion of Pear and Apricot leaves, which are often underrepresented in standard datasets such as PlantVillage, enables the model to learn complex symptomatic patterns such as leaf spot and rust in real world agricultural scenarios. The severity
of the disease was calculated and expressed in five classes: no risk (0%), low (6–20%), medium (21–25%), and high (>50%), on a scale from 0 to 3. Table 2 presents the statistics of the PlantCity dataset.

Table 2: Statistics of PlantCity dataset.

Cherry		**Pear**	
Healthy	250	Healthy	250
Spot	250	Curl	250
Scorch	250	Slug	250
Curl	250	Spot	250
Total	1000	Total	1000

4.2 Data Augmentation

To increase the diversity of the training set and to avoid the overfitting of models, a dedicated data augmentation pipeline has been implemented using the Augmentor library. All original images of Pear and Apricot leaves underwent a series of stochastic transformations to simulate real-world environmental variability. These included geometric transformations such as random horizontal and vertical flips (p=0.8) and orthogonal rotations at 90°, 180° and 270° to account for different camera orientations. We also introduced random zooms (between 1.1x and 1.5x) and brightness (0.7 to 1.3) changes to mimic the varying lighting conditions of the field-based image acquisition. Natural variation in leaf shape was also modeled using elastic distortions. This pipeline augmented the initial dataset into a robust collection of 10 variations per original image to ensure the deep learning architecture could accurately identify disease features independent of scale, orientation, or illumination. Table 3. Statistics of the PlantCity dataset after augmentation.

Table 3: Statistics of the PlantCity dataset after augmentation.

Cherry		Augmented	Pear		Augmented
Level 0	250	1000	Level 0	250	1000
Level 1	250	1000	Level 1	250	1000
Level 2	250	1000	Level 2	250	1000
Level 3	250	1000	Level 3	250	1000
Total	1000	4000		1000	4000

4.3 Experimental settings

The proposed system is simulated using python, and the deep leaning framework is accelerated using GPU. The experimental hardware consists of Intel Quad Core i7 2820QM CPU at 2.30 GHz and NVIDIA Quadro with Samsung 850 Pro 256 GB SSD. This is used in training and testing the model. Training parameter details for the proposed method are presented in Table 4.

Table 4: Specification of training parameters

Parameter	Values
Batch size	32
Epoch	100
Momentum	0.9
Learning rate	0.01 0.00001
Weight decay	0.005
Optimizer	Adam

4.4 Metrics used

The metrics used in the study were accuracy, precision, recall, and F1 score. We used TP (True Positive) to identify the diseased leaf images correctly classified as diseased. TN (True Negative) is used when categorizing the images of healthy leaf as healthy images. FP (False Positive) consists of healthy leaf images classified as diseased images and FN (False Negative) are the diseased leaf images classified as healthy. Accuracy, in general, is considered to be a poor evaluation metric, especially when measuring model performance in, real world application scenarios. For most application domains, accuracy might be sufficient; especially in cases with balanced datasets. In this study, we heavily relied on other metrics as explained in equations (15) to (18).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100 \tag{15}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \times 100 \tag{16}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \times 100 \tag{17}$$

$$\text{F1} - \text{Score} = 2 . \frac{Precision . Recall}{Precision + Recall} \tag{18}$$

4.5 Visualization results

This section describes the visual assessment results and the analysis of numerous experiments on the PlantCity dataset using the proposed approach. The framework is run on healthy and diseased plant leaf databases that include leaf types such as cherry and pear. This approach employs preprocessing, segmentation, and detection. In several instances, numerous models were unable to capture the entire view of sick leaf images. Figure 2 (a–d) displays images sampled from the PlantCity cherry and pear datasets. These images are used to identify sick leaf images, segment the leaf area, and assess sick leaf images during the training and validation stages.

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

Figure 2. Sample Cherry (a,b) and Pear (c,d) images from PlantCity datasets.

We evaluated the proposed framework using the PlantCity dataset on 100 epochs. Figure 3(a, b) shows the training and validation accuracy over epochs and training and validation loss over epochs of PlantCity dataset. The model achieved 87.31% accuracy for training, which indicates the model's ability to classify training data accurately as shown in Figure 3a. similarly, the method achieved accuracy of 85.14% for validation, which shows good generalization of the model on unseen data.

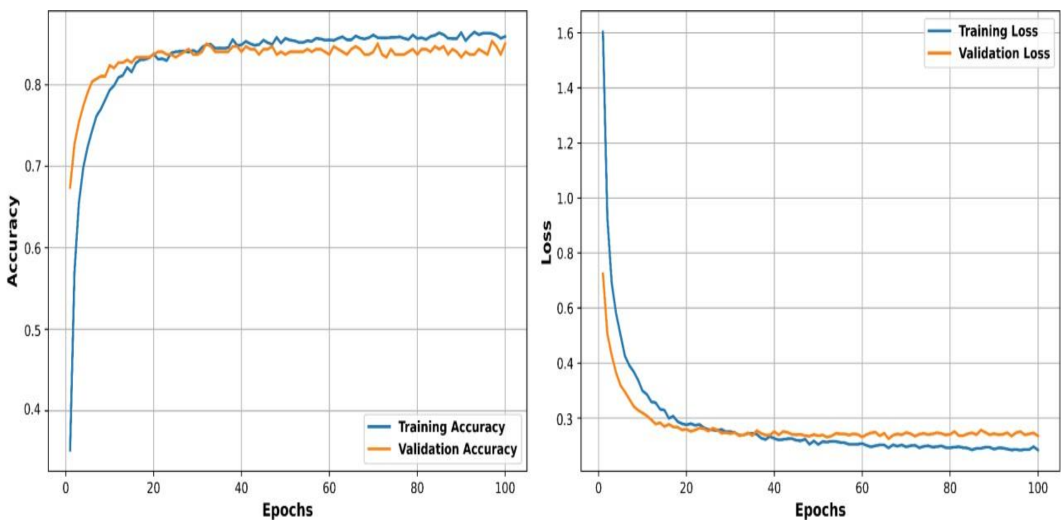


Figure 3 (a, b) Shows Training/Validation accuracy and Training and Validation Loss over epochs of PlantCity Dataset

Figure 4(a, b) represent the normalized confusion matrix for PlantCity dataset. The high values in the diagonal blue area prove that the model effectively classifies the severities into their respective categories. The values of off diagonal area are significantly low, which means that the model has made a few misclassifications. Overall, the confusion matrix proves strong performance in classification on the PlantCity dataset.

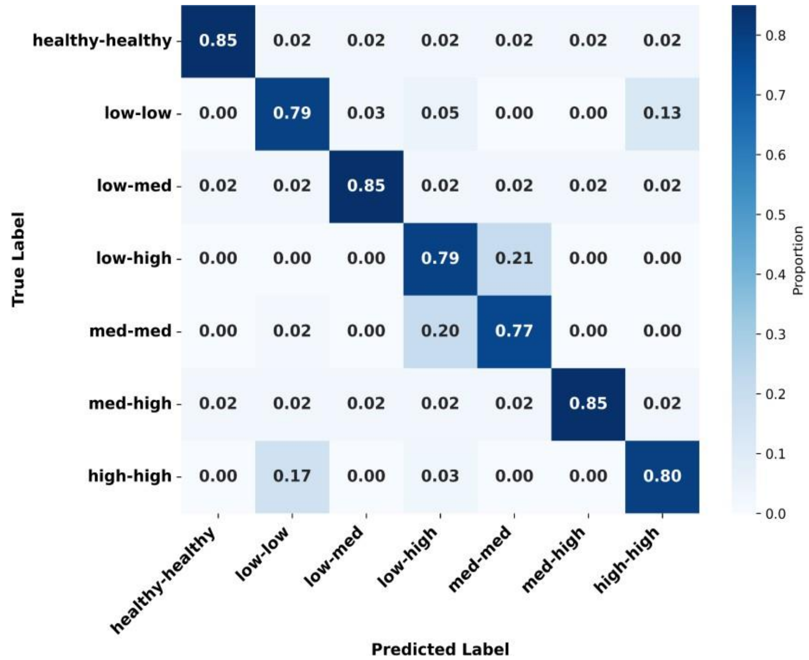


Figure 4a Confusion matrix for PlantCity (Cherry) dataset

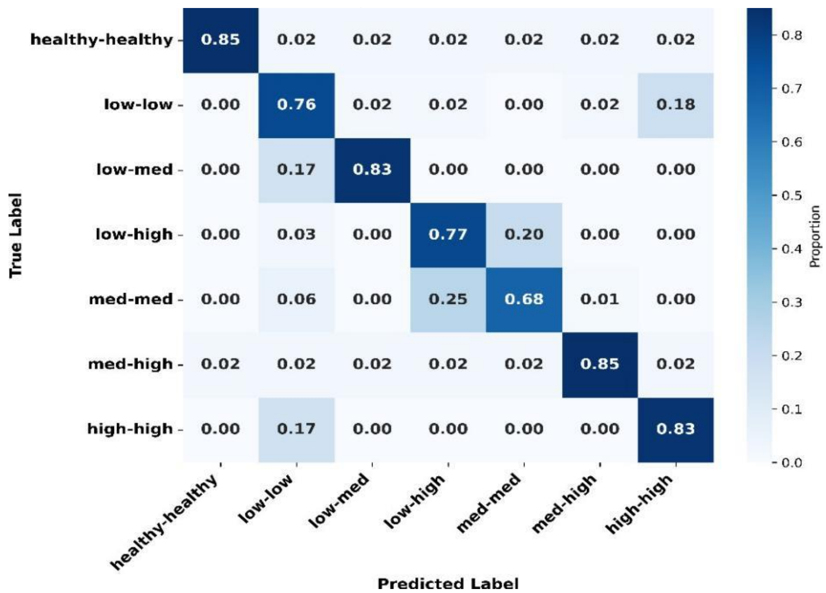


Figure 4b Confusion-matrix for PlantCity (Pear) dataset

The quantitative performance of the proposed hybrid framework is presented in Table 5 in terms of classification accuracy, precision, recall and F1 score for seven severity levels for both cherry and pear leaves. The results show high stability with weighted average accuracies of 84.14% for cherry and 85.42% for pear, respectively. Particularly, the 'Healthy Healthy' and 'Med High' categories had the highest individual performance, indicating that the combination of EVA02 transformer based global features is particularly effective at identifying distinct structural health and advanced lesion spread. At all levels the values of precision and recall are close together, especially in 'Low Med' and 'Med Med' categories, which indicate the effectiveness of handcrafted WLBP and Haralick texture descriptors in handling inter class similarities. This balanced performance in all metrics will validate the robustness of the model to quantify the overlapping disease signatures, and provide an accurate foundation for multi-disease-severity estimation in both fruits.

Table 5. Performance Analysis as Per Class.

Severity Level	Cherry				Pear			
	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Healthy healthy	87.30	88.1	87.5	87.8	86.50	86.5	85.8	86.1
Low low	84.20	83.5	84.8	84.1	86.31	85.2	86.4	85.8
Low med	85.60	86.2	85.1	85.6	83.27	82.8	83.5	83.1
Low High	84.37	84.1	83.9	82.3	83.47	84.2	83.8	83.7
Med med	83.10	82.8	83.4	83.1	87.19	87.6	86.5	87.0
Med high	86.50	85.9	86.8	86.3	88.34	88.4	87.9	88.1
High high	83.58	84.0	83.2	83.6	85.45	84.7	85.5	85.1
Weighted Average	**84.14**	**84.08**	**84.13**	**84.10**	**85.42**	**85.35**	**85.46**	**85.40**

4.6 Area Under the Curve Analysis

To deepen the analysis of the discriminative power of the proposed feature fusion framework, the AUC was obtained for each disease. AUC provides a threshold independent measure of the model discriminatory capacity between different levels of plant disease severity, and is thus more suitable for imbalanced and multi-class plant disease datasets. The predicted probability scores have been plotted on the ROC curves, and the AUC values have been computed using one vs one approach for Disease A severity vs Disease B severity pairs. The fused feature representation (WLBP + Haralick + EVA02) had always higher AUC scores than the single feature sets, thus showing better separability between severity stages. Importantly, EVA02 features helped to better discriminate globally, while hand-crafted texture descriptors increased sensitivity to subtle lesion patterns, leading to smoother and more stable ROC curves. The results confirm that the proposed fusion strategy improves the classification accuracy and also increases the robustness of the model to differentiate the severity levels. Table 6. Percentage AUC performance for each disease.

Table 6. Per Disease AUC Performance (%)

Feature Configuration	Disease A AUC	Disease B AUC	Mean AUC
WLBP Only	78.6	77.9	**78.1**
Haralick Only	76.2	75.5	**75.9**
EVA02 Only	80.8	79.4	**80.1**
WLBP + Haralick	79.5	78.4	**78.9**
EVA02 + WLBP	80.9	81.2	**80.1**
EVA02 + Haralick	81.6	80.1	**81.9**
WLBP + Haralick + EVA02 (Proposed)	**85.8**	**84.4**	**85.1**

Figure. 5 presents the ROC curves and the corresponding AUC values of the proposed fused feature based severity classification framework. The first image is the multi class ROC curves.

Each curve is for specific class of disease–severity pair. The AUC values are between 0.83 and 0.88 indicating the high discriminative ability for all seven combined severity classes . The consistent separation of the ROC curves from the diagonal reference line (random classifier) indicates that the proposed fusion of WLBP, Haralick texture features and EVA02 deep features is effective in capturing both fine grained texture variations and high level semantic information, leading to reliable class discrimination.

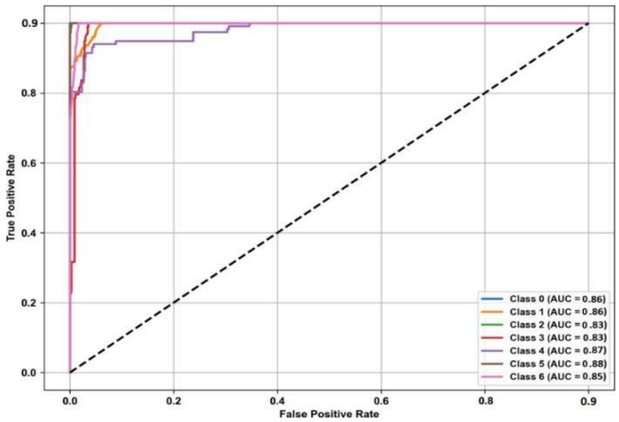
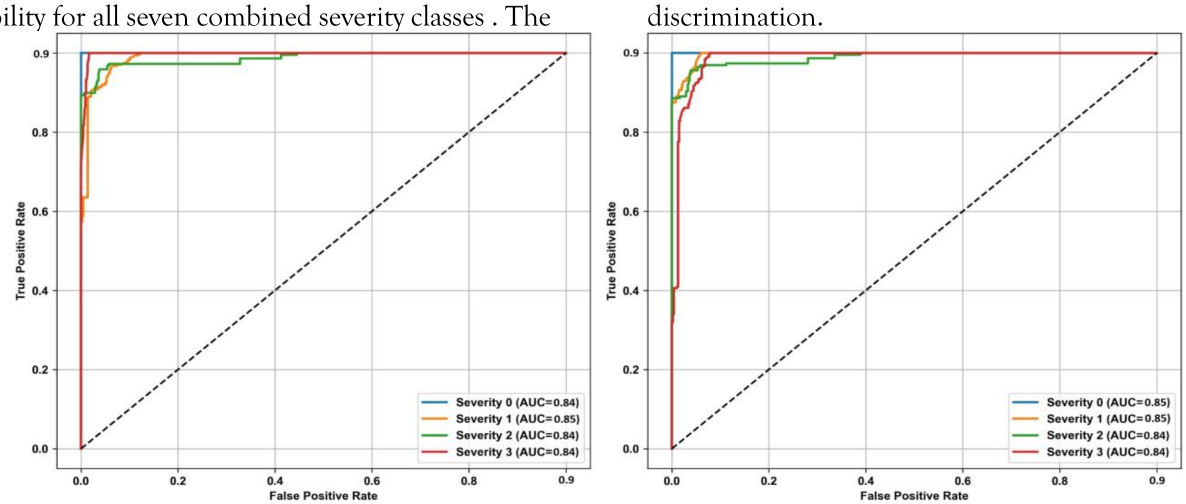


Figure 5 Multi class ROC Curve

The ROC curves and corresponding AUC values for the proposed fused feature based severity classification framework are depicted in Figure 6. The first image shows the multi class ROC curves where each curve represents a disease-severity pair class. The AUC values are between 0.83 and 0.88 indicating the high discriminative ability for all seven combined severity classes . The consistent separation of the ROC curves from the diagonal reference line (random classifier) indicates that the proposed fusion of WLBP, Haralick texture features and EVA02 deep features is effective in capturing both fine grained texture variations and high level semantic information, leading to reliable class discrimination.



(a) Disease A (b) Disease B

Figure 6 Disease wise ROC Curve a) Disease A b) Disease B

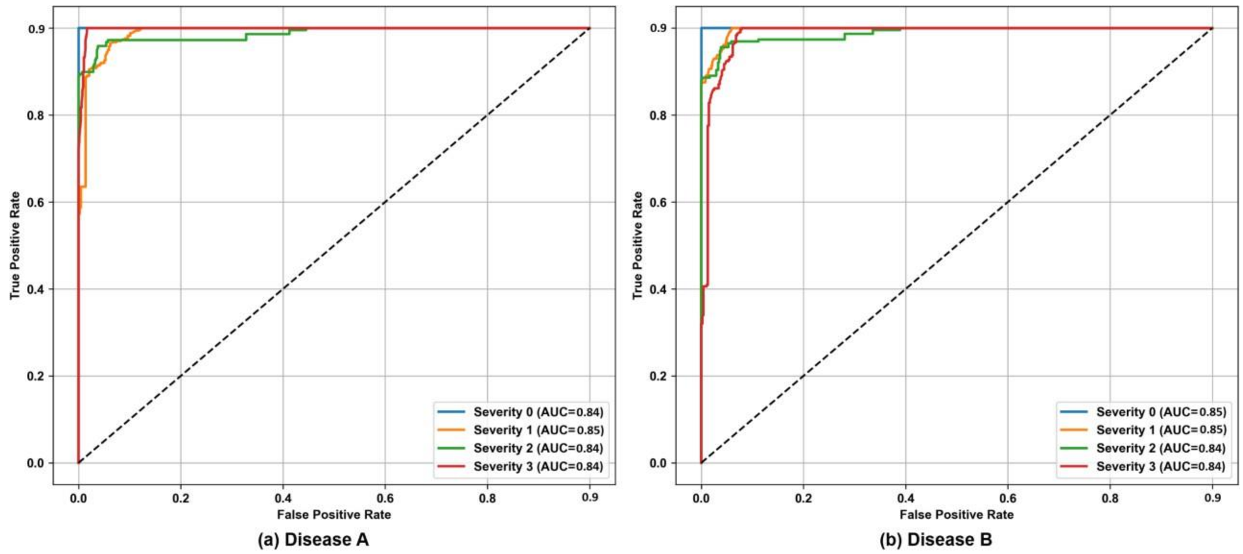


Figure 6 Disease wise ROC Curve a) Disease A b) Disease B

4.7 Comparative Analysis with Existing Approaches

To evaluate the performance of the proposed framework was evaluated with several existing methods, such as CNN, InceptionV3, VGG-16, VGG-19, ResNet-50, ResNet-101, and LTP. Experiments was conducted using 100 training epochs and using standard evaluation metrics, namely accuracy, precision, recall, and F1-score. The performance of the proposed framework and a number of contemporary neural network architectures (i.e., CNN, residual networks, and standard Vision Transformers (ViT)) is presented in Table 7.

From the results, it was observed that the baseline CNN models (VGG 16 and ResNet 101) achieved moderate accuracies of 64% to 71% and were unable to capture the multi scale symptomatic variations in the PlantCity dataset.The standard ViT showed better performance with 79.42% and 80.73% accuracy for cherry and pear leaves respectively, due to its self attention mechanism. However, it was still not as good as the proposed methodology. Our hybrid framework combining EVA02 Vision Transformers with WLBP and Haralick texture descriptors has significantly improved performance, achieving about 5% higher overall accuracy than the nearest competitor (ViT). The large margin indicates that global semantic context must be combined with fine grained local texture features to correctly disambiguate the inter class similarities that are inherent in plant disease severity levels.

Table 7. Performance Analysis on Pear dataset.

Methods	**PlantCity (Cherry) dataset**				**PlantCity (Pear) dataset**			
	Precision %	Recall %	Accuracy %	F1 Score %	Precision %	Recall %	Accuracy %	**F1 Score %**
CNN	61.20	60.15	60.65	60.67	64.10	63.45	63.86	**63.77**
VGG 16	70.45	69.80	70.09	70.12	71.90	71.30	71.67	**71.60**
VGG 19	64.90	64.15	64.43	64.52	70.15	69.20	69.71	**69.67**
ResNet 50	66.05	65.10	65.51	65.57	68.50	67.80	68.12	**68.15**

RestNet 101	70.55	69.85	70.13	70.19	71.85	70.95	71.31	**71.40**
Inception V3	74.80	74.10	74.39	74.45	75.90	75.10	75.42	**75.50**
GoogleNet	78.15	77.40	77.80	77.77	78.75	77.95	78.32	**78.35**
VIT	79.80	79.15	79.42	79.47	81.10	80.35	80.73	**80.72**
Proposed Methodolo gy	**84.14**	**84.08**	**84.13**	**84.10**	**85.42**	**85.35**	**85.46**	**85.40**

To analyze the learning behavior and convergence of models, the accuracy curves along with the training and validation dataset were plotted on the PlantCity dataset. The proposed method always outperforms recent methods in terms of training accuracy and validation accuracy as demonstrated in Figures 7(a) and 7(b). This excellent performance is credited to the discriminative feature representation given by the descriptor S³DLTP.

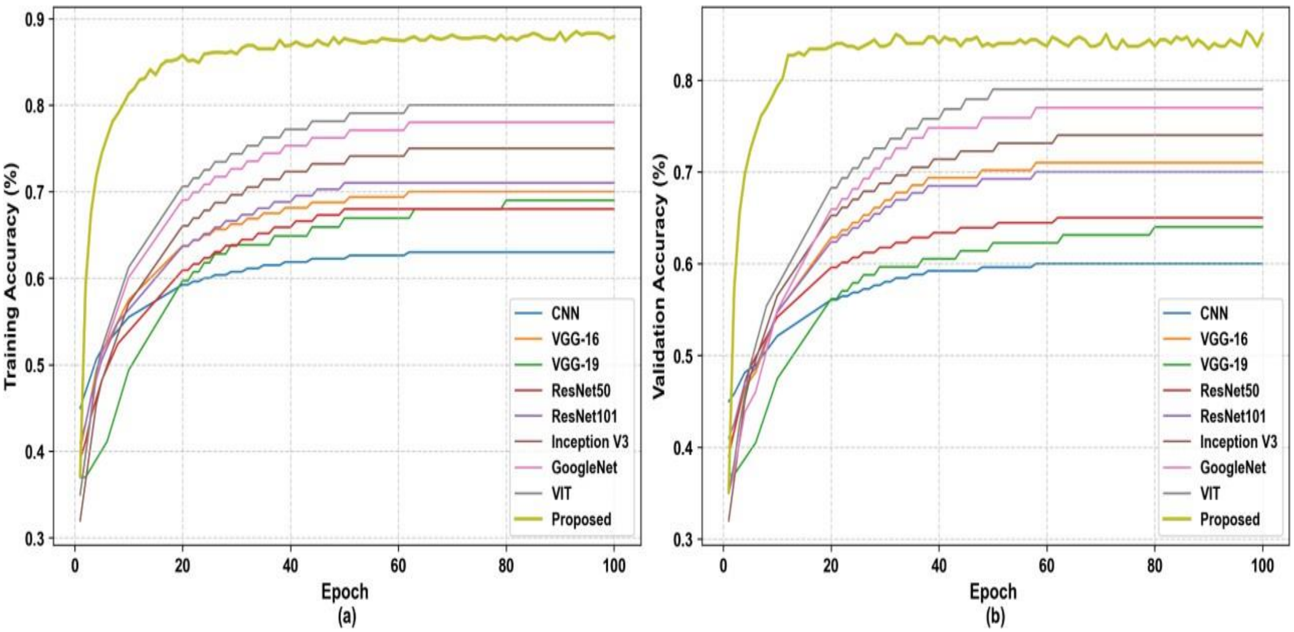


Figure 7 (a, b) Performance of Comparison-analysis with other Methods on PEER data.

4.8 Comparison with Existing Methods:

Table 8 presents a comparative analysis between the proposed methodology and recent state-of-the-art approaches for plant disease analysis. It is evident from the table that existing studies primarily focus on single disease severity estimation or detection, even when advanced deep learning architectures are employed. For instance, Li utilized DM BiSeNet achieved better performance; however, model was restricted to a single disease scenario Similarly, Li proposed an improved multi scale SAM framework on the PlantVillage dataset, and Amanova and Rezaei Masoud used modern object detection and salient object detection strategies, respectively, but all these methods only handle one disease type, which restricts their applicability in real agricultural environments where multiple diseases may co-exist simultaneously.

In contrast, the proposed method is the only approach in this comparison that performs multi disease severity estimation, simultaneously modeling severity patterns for two distinct diseases using PlantCity dataset. The obtained accuracy of 85.42% is numerically lower than some single disease methods, but represents a more challenging and realistic problem formulation. Robust feature representation across heterogeneous disease characteristics and severity levels is achieved by combining WLBP, Haralick texture descriptors and EVA02 deep features. Thus, the proposed framework is a major methodological advancement that emphasizes generalization and scalability over task specific optimization and establishes a new baseline for multi-disease severity estimation in plant health monitoring systems.

Table 8. Comparison-Analysis with Existing Methods

Studies	Methods	Dataset_used	Diseases Tackled	Accuracy (%)
Li, Kaiyu, et al. 2025 [20]	DM BiSeNet	Self Created	1	**94**
Li, Junlong, et al. 2025 [21]	Enhanced Multi Scale SAM (EMSAM)	PlantVillage	1	**83**
Amanova, Raikhan, et al. 2026 [22]	YOLOv10n + Ordinal MobileViT S		1	**91**
Rezaei Masoud, et al. 2026 [23]	Salient Object Detection (SOD)		1	**82**
Proposed Methodology	**WLBP + Haralick + EVA02**	**PlantCity**	**2**	**85.42**

4.9 Ablation Study and Cross Dataset Validation

To analyze the individual and combined impact of the proposed feature extraction components, an ablation study was performed by turning on and off WLBP, Haralick texture features, and EVA02 deep transformer features in sequence. In Table 9, we can see that the performance with only EVA02 features (A1) is reasonable, as it possesses strong global representation ability, but it lacks sensitivity to fine-grained texture variations. The traditional handcrafted descriptors WLBP (A2) and Haralick (A3) are relatively inferior in accuracy when used alone, implying their limited discriminability alone. The combination of handcrafted features (A4) yields a modest performance gain, suggesting that the local texture cues are complementary. The subsequent integration of EVA02 with Haralick (A5) or WLBP (A6) leads to a significant performance gain, suggesting that there is a good synergy between deep semantic representations and handcrafted texture descriptors. The full proposed model (A7) achieves the best accuracy, which is a combination of the WLBP feature, Haralick feature and the EVA02 feature in one unique discriminative feature vector. It means that the complementary information of local texture patterns, statistical texture descriptors and transformer based global features greatly enhances the performance of disease classification, which can be seen as the effectiveness and necessity of the proposed hybrid fusion strategy.

Table 9. Ablation study of proposed methodology

Configuration ID	WLBP	Haralick	EVA02	Feature Fusion	Accuracy (%)
A1	✗	✗	✓	✗	71.3
A2	✓	✗	✗	✗	58.6
A3	✗	✓	✗	✗	61.9
A4	✓	✓	✗	✓	69.8
A5	✗	✓	✓	✓	79.4
A6	✓	✗	✓	✓	81.2
A7 (Proposed)	✓	✓	✓	✓	85.0

Discussion:

The experimental results show that the proposed framework is effectively capture complementary information for accurate plant disease severity classification. Although there are differences in leaf morphology, distribution of leaf texture and disease manifestation patterns, the high accuracy of cherry leaves and pear leaves show good generalization on the different pome crop species. The improved performance achieved using the combined feature sets confirms that disease severity cannot be described reliably using global deep features or handcrafted descriptors alone. Instead, a multi level representation jointly modeling local texture irregularities and global semantic context is essential for robust disease analysis. The ablation study also validates the effectiveness of each module in the proposed framework. EVA02 alone can provide a strong global representation due to the hierarchical transformer architecture but has low sensitivity for subtle texture changes associated with early or intermediate stages of disease severity. WLBP and Haralick features on the other hand can capture fine grained texture variations and statistical spatial relationships respectively but they have no high level semantic discrimination power in isolation. Their combination with EVA02 leads to a significant enhancement in classification performance, which confirms the complementary nature of handcrafted texture descriptors to transformer-based deep features. The above results show that the full fusion model has better performance, which indicates that local and global feature characteristics should be preserved in severity estimation tasks.

In addition, the weighted formulation of LBP is important to highlight the disease affected regions by giving more weight to pixels with higher intensity variation, thus increasing the robustness to illumination changes and background noise. In addition, the second order texture statistics of Haralick features allow the model to differentiate among severity levels that are visually similar. The fusion strategy, applied at the level of feature concatenation, guarantees the preservation of discriminative information from all descriptors without excessive growth of model complexity. Such design choice is a tradeoff between performance and computational efficiency and makes the proposed method applicable to real world agricultural decision support systems. In summary, the proposed approach demonstrates that the combination of transformer based representations with enriched handcrafted features leads to a more comprehensive and discriminative feature space for plant disease severity . Performance improvements are steady and robust across datasets and experimental conditions, and scalable. These results suggest that hybrid feature fusion frameworks may be a promising research direction in precision agriculture, especially for the disease severity assessment requiring high interpretability, limited training data, and cross-crop adaptability.

Conclusion:

We propose a powerful hybrid method to classify the severity of plant diseases by using both deep features extracted by the transformer and discriminative handcrafted texture descriptors. The proposed method combines both EVA02 representations and WLBP and Haralick features and can capture information at both global and local levels, corresponding to different severity levels. The results of the experiments performed on pome crop leaf data demonstrated that the performance was consistently better than the performance of the individual feature configuration, and the classification performance was reliable in terms of accuracy in the various disease severity categories. The handcrafted features and deep features of each component in the framework were shown to be complementary with meaningful contribution to the overall performance through ablation analysis. This developed method provides a good balance between accuracy and processing speed, and can easily be implemented in numerous precision agriculture systems. The framework will be extended to multi crop scenarios such as temporal disease progression analysis and lightweight fusion strategies to facilitate real time field applications in the future.

Competing Interests

The authors declare that they have no competing interests.

References:

1. Shi, T., et al., *Recent advances in plant disease severity assessment using convolutional neural networks.* Scientific Reports, 2023. 13(1): p. 2336.
2. Leite, D.V., et al., *Deep Learning Models for Detection and Severity Assessment of Cercospora Leaf Spot (Cercospora capsici) in Chili Peppers Under Natural Conditions.* Plants, 2025. 14(13): p. 2011.
3. Baek, E.-T., *Attention score-based multi-vision transformer technique for plant disease classification.* Sensors, 2025. 25(1): p. 270.
4. Mehnaz, S. and M.T. Islam, *Rice leaf disease detection: a comparative study between CNN, transformer and non-neural network architectures.* arXiv preprint arXiv:2501.06740, 2025.
5. Amanova, R., et al., *A Lightweight Edge-AI System for Disease Detection and Three-Level Leaf Spot Severity Assessment in Strawberry Using YOLOv10n and MobileViT-S.* Computers, 2026. 15(1): p. 63.
6. Ahmad, Wakeel, Syed M. Adnan, and Aun Irtaza. "Local triangular-ternary pattern: a novel feature descriptor for plant leaf disease detection." *Multimedia Tools and Applications* 83, no. 7 (2024): 20215-20241.
7. Ahmad, Wakeel, S. M. Adnan Shah, and Aun Irtaza. "Plants disease phenotyping using quinary patterns as texture descriptor." *KSII Transactions on Internet & Information Systems* 14, no. 8 (2020): 3312.
8. Mohanty, S.P., D.P. Hughes, and M. Salathé, *Using Deep Learning for Image-Based Plant Disease Detection.* Frontiers in Plant Science, 2016. Volume 7 - 2016.
9. Too, E.C., et al., *A comparative study of fine-tuning deep learning models for plant disease identification.* Computers and Electronics in Agriculture, 2019. 161: p. 272–279.
10. Ferentinos, K.P., *Deep learning models for plant disease detection and diagnosis.*
11. Liu, B., et al., *Identification of apple leaf diseases based on deep convolutional neural networks.* Symmetry, 2017. 10(1): p. 11.
12. Computers and electronics in agriculture, 2026. 241: p. 111316.
13. Noreen, Muniba, Furqan Shaukat, and Syed Muhammad Anwar. "Lung Nodule-SSM: Self-Supervised Lung Nodule Detection and Classification in Thoracic CT Images." In *2025 21st International Symposium on Biomedical Image Processing and Analysis (SIPAIM)*, pp. 1-4. IEEE, 2025.
14. Noreen, Muniba, Syed Muhammad Adnan, Rao Wakeel Ahmad, Muhammad Imran Faizi, Haider Ali Khan, and Abid Ghaffar. "A Self-Supervised Framework for Annotation-Free Lung Nodule Detection and Classification in CT Scans." *International Journal of Information Systems and Computer Technologies* 5, no. 1 (2025): 108-122.
15. Chen, J., et al., *Using deep transfer learning for image-based plant disease identification.*
16. Computers and electronics in agriculture, 2020. **173**: p. 105393.
17. Raza, S.-e.-A., et al., *Automatic detection of diseased tomato plants using thermal and stereo visible light images.* PLOS ONE, 2015. **10**(4): p. e0123262.
18. Liu, B., et al., *Identification of apple leaf diseases based on deep convolutional neural networks.* Symmetry, 2017. **10**(1): p. 11.
19. Thakur, P.S., et al., *Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT.* arXiv preprint arXiv:2207.07919, 2022.
20. Salka, T.D., et al., *Plant leaf disease detection and classification using convolution neural networks model: a review.* Artificial Intelligence Review, 2025. **58**(10): p. 322.

21. Kaur, Arshleen, Vinay Kukreja, Sonal Malhotra, Kireet Joshi, and Rishabh Sharma. "Rice sheath rot disease detection and severity classification: A novel framework leveraging CNN-LSTM models for multi-classification." In *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, vol. 2, pp. 1-5. IEEE, 2024.
22. Garg, Pooja, Kathan Vora, Devansh Shah, Manjunath V. Joshi, Ahlad Kumar, and Vinay S. Palaparthy. "Transformer-Based Framework Using In-House Sensors and Lightweight Images for Accurate Plant Disease Classification." *IEEE Sensors Journal* (2026).
23. Javidan, S.M., et al., *Tomato fungal disease diagnosis using few-shot learning based on deep feature extraction and cosine similarity.* AgriEngineering, 2024. **6**(4): p. 4233–4247.
24. Feng, W., G. Sun, and X. Zhang, *Plant disease identification based on encoder–decoder model.* Agronomy, 2024. **14**(10): p. 2208.
25. Li, G., et al., *PMVT: a lightweight vision transformer for plant disease identification on mobile devices.* Frontiers in Plant Science, 2023. **14**: p. 1256773.
26. Elghawth, R., et al., *Deep Learning for Transformer-Based Plant Disease Detection: A Bibliometric Analysis.* Engineering Proceedings, 2025. **112**(1): p. 29.
27. Thapa, R., et al., *The Plant Pathology Challenge 2020 data set to classify foliar disease of apples.* Applications in plant sciences, 2020. **8**(9): p. e11390.
28. Mehnaz, S. and M.T. Islam, *Rice leaf disease detection: a comparative study between CNN, transformer and non-neural network architectures.* arXiv preprint arXiv:2501.06740, 2025.