**IAES International Journal of Artificial Intelligence (IJ-AI)**
Vol. 15, No. 6, December 2026, pp. 5692~5702
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i6.pp5692-5702 

# Hierarchical cross-scale convolutional network with dual attention for robust plant disease detection

**Jayashree Ramesh, Kusuma Kumari Boohalli Marigowda**
Department of Studies and Research in Computer Applications, Tumkur University, Tumkur, India

**Article Info**

***Article history:***

Received Sep 3, 2025
Revised Aug 20, 2026
Accepted Sep 19, 2026

***Keywords:***

Attention mechanisms
Cross-scale fusion
Hierarchical convolutional network
Plant disease detection
Transfer learning

**ABSTRACT**

Plant disease detection in the field is difficult because images contain cluttered backgrounds, uneven illumination, and small early-stage lesions, all of which reduce the accuracy of conventional convolutional networks trained on laboratory data. This paper proposes a deep learning framework for robust disease detection built on a hierarchically structured cross-scale convolutional network. The backbone uses a transformer-inspired 3 : 3 : 9 : 3 block arrangement, an inverted bottleneck, and large-kernel depthwise separable convolutions to widen the receptive field at low cost. A parallel dual-focus attention module applies channel and spatial attention independently to avoid the interference of serial designs, a hierarchical feature fusion unit captures cross-scale lesion cues, a pre-trained feature alignment module transfers knowledge to small datasets, and a feature recalibration adapter generates input-aware filter responses for robustness. Across four datasets, the model reaches 92.1% accuracy on rice, 86.0% on sugarcane, 96.8% on tomato, and 93.7% on corn, improving on the baseline by 3.46% to 4.34% across precision, recall, F1-score, Matthew's correlation coefficient, and accuracy. An ablation study confirms that every module contributes, indicating a practical route to real-time in-field diagnosis.

*This is an open access article under the CC BY-SA license.*

***Corresponding Author:***

Jayashree Ramesh
Department of Studies and Research in Computer Applications, Tumkur University
Jnanasiri campus, Bidarakatte, Karnataka, India
Email: jayashreerameshg@gmail.com

## 1. INTRODUCTION

Plant leaf disease is a persistent threat to crop productivity, and reliable detection has become a core concern for precision agriculture. Two closely related technologies drive most of the progress: computer vision supplies the visual features that separate healthy from diseased tissue, and machine learning maps those features to disease categories. Recent pipelines that pair smartphone imaging with convolutional recognition have shown that leaf, fruit, and stem symptoms can be diagnosed in near real time under laboratory conditions, which motivates larger-scale deployment in the field [1]. Untreated infection not only lowers yield but can progressively damage the photosynthetic tissue of the plant; manual assessment by trained agronomists remains the reference standard, yet it is labor-intensive, slow, and unevenly available across regions. An effective automated pipeline must therefore provide both fine-grained localization of the disease and an assessment of its severity, because treatment choice and yield forecasting depend on how the disease is staged [2].

Manual inspection is prone to inter-observer variability and often misses subtle early-stage lesions, which motivates data-driven alternatives. Since the introduction of AlexNet, deep convolutional networks have become the default tool for plant image segmentation, classification, and detection. Comparative reviews report that deeper architectures, combined with saliency-based visualization of the learned features,

tend to outperform hand-crafted descriptors and shallow classifiers on standard leaf-disease benchmarks [3]. A related line of work asks whether model decisions remain trustworthy outside the laboratory; explainable artificial intelligence (XAI) methods have been used to confirm that a classifier attends to diseased tissue rather than background artefacts [4]. Variations in leaf shape, texture, color, and image noise can still confound simple pipelines [5], and traditional methods such as support vector machines or k-means clustering are limited by their reliance on hand-tuned preprocessing. Deep learning that exploits large annotated corpora has therefore become the preferred route to detectors that generalize across crops and growth conditions [6].

Widely used plant-disease collections, including citrus, rice, cassava, and PlantVillage, are mostly captured under controlled conditions, and networks trained on them classify well during training but lose accuracy once deployed in the field. The main reason is the gap between clean laboratory backdrops and real captures that contain overlapping leaves, stems, fruit, soil, and mulch. Segmenting or filtering the background before classification improves field accuracy; lightweight detectors that combine depthwise separable convolutions with feature fusion have recovered small, scattered lesions under cluttered orchard backgrounds [7]. Attention-augmented detectors extend this idea by re-weighting the feature map toward lesion-bearing pixels, with competitive results on tomato viral disease [8]. Where data cannot be centralized across farms, federated learning with an improved detector shares gradients without exchanging raw images [9], and fuzzy deep recurrent formulations have been used to stabilize predictions on noisy crop imagery such as corn leaves [10].

In summary, plant diseases threaten food security most when early signs go unnoticed, when manual diagnosis is slow and expert-dependent, and when field conditions are complicated by changing illumination, occlusion, and overlapping structures. Conventional machine learning struggles with uneven data quality and limited feature extraction, whereas convolutional networks perform better but still degrade under noisy backgrounds and small annotated datasets. This work is driven by the need for a robust, fast, and scalable model that improves feature representation and delivers accurate diagnosis across varied field settings.

The main contributions of this paper are as follows. First, it introduces the hierarchically-structured cross-scale convolutional network (HCS-ConvNet), a backbone that combines transformer-inspired 3 : 3 : 9 : 3 block arrangement with cross-scale feature extraction to capture fine-grained lesion detail at reduced training cost. Second, it proposes a parallel dual-focus attention module (DFAM) that applies channel and spatial attention independently, avoiding the interference of serial designs while adding little computation. Third, it integrates a hierarchical feature fusion unit (HFFU) and a dynamic scale fusion module (DSFM) built on depthwise separable convolutions, together with a pre-trained feature alignment module (PFAM) that transfers ImageNet features to small disease datasets. Fourth, it adds a feature recalibration adapter (FRA) that produces input-aware filter responses without changing the base network. Finally, the framework is evaluated on four crop datasets with an ablation study, a comparison against state-of-the-art architectures, and full reproducibility details.

## 2. RELATED WORK

Recent detectors combine attention, cross-scale fusion, and transfer learning. Liu *et al.* [7] proposed dynamic attention convolutional (DAC)-path aggregation path planning you only look once efficient (PPYOLOE), a lightweight real-time detector that couples dynamic depthwise separable convolutions with residual connectivity to recover small pest and lesion targets under cluttered apple-orchard backgrounds. Qi *et al.* [8] embedded a visual attention mechanism into you only look once version 5 (YOLOv5) to localize small, scattered symptoms on tomato leaves. Deng *et al.* [9] handled multi-disease and pest detection with an improved faster region-based convolutional neural network (Faster R-CNN) trained under a federated regime, while Irianto and Findley [10] used a fuzzy deep recurrent model to reduce noise sensitivity for corn leaf disease. Chang *et al.* [11] injected cross-scale edge cues into vision transformers to strengthen boundary representation, and Xiao *et al.* [12] combined a lightweight residual network with channel-spatial attention to cut parameters while retaining accuracy. Gao *et al.* [13] followed with a dual-branch efficient channel-attention architecture, Pandi *et al.* [14] used a dilated convolutional network with global average pooling for rice, and Liu *et al.* [15] applied EfficientNet with attention and cross-scale fusion to cassava. Bezabh *et al.* [16] concatenated several backbones for pepper disease in small-sample settings, and Gutiérrez *et al.* [17] compared convolutional architectures on tomato and found that deeper ImageNet-pretrained backbones performed best.

Survey and dataset studies give useful context. Bischoff *et al.* [18] mapped machine learning applications in agriculture and noted the dominance of RGB imagery and the concentration of effort on disease detection and yield prediction. Lu *et al.* [19] reviewed convolutional designs for leaf disease and observed that many rely on laboratory images and generalize poorly to the field, and Ali *et al.* [20] surveyed non-destructive detection techniques across imaging modalities. Moupojou *et al.* [21] released FieldPlant, a

field-captured dataset that exposes classifiers to cluttered backgrounds and variable illumination, and reported a marked accuracy drop when models move from laboratory to field images. Cross-domain studies also inform the present design: Huang *et al.* [22] showed that a cascaded hierarchical convolutional neural network (CNN) reduces confusion between similar classes, Wang *et al.* [23] found that combining attention with a ResNet backbone improves fine-grained classification, and Anandhi and Sathiamoorthy [24] paired metaheuristic tuning with deep multimodal fusion for rice segmentation on small datasets.

Despite this progress, several gaps remain. Most attention-based models apply channel and spatial attention serially, which lets channel weighting distort spatial reasoning; few cross-scale approaches include transfer learning tailored to small agricultural datasets; input-aware feature recalibration that adapts to changing field conditions is rarely used; and few frameworks address cluttered backgrounds, small dispersed lesions, and data scarcity together. These gaps motivate the unified design proposed in this work.

## 3. METHOD

### 3.1. Network overview

HCS-ConvNet is built to extract features efficiently while keeping training cost low. It has three parts: an input stem that extracts low-level features, four intermediate stages formed by stacked HCS blocks, and a classification head [22]. The blocks are distributed across the four stages in a non-uniform 3 : 3 : 9 : 3 pattern, which reduces to a 1 : 1 : 3 : 1 proportion and places most of the network's depth in the third stage. In place of a conventional convolutional stem, the model uses a patchify operation with a 4 × 4 kernel and stride 4 that divides the input into non-overlapping patches, in the manner of a vision transformer [2], [11]. Every block adopts an inverted-bottleneck layout, so the channel dimension is first reduced and then expanded around the feature-extraction step, which helps preserve compact but informative features. Depthwise separable convolutions lower the computational load and let the network model longer-range dependencies, layer normalization replaces batch normalization, and activation functions are applied sparingly to keep training stable. Spatial resolution is reduced between stages by a 2 × 2 stride-2 convolution preceded by layer normalization, and the usual 3 × 3 kernels are replaced by 7 × 7 kernels to widen the receptive field.

The four stages are deliberately not given equal depth. Most blocks sit in the third stage, where the resolution is moderate and mid-level lesion textures, the cues that best separate one disease from another, are most discriminative. Concentrating depth there rather than spreading it evenly builds richer representations at the scale that matters while keeping the shallow, high-resolution stages inexpensive; this is why the 3 : 3 : 9 : 3 (1 : 1 : 3 : 1) arrangement is adopted. To confirm the choice is not arbitrary, it was compared against a uniform 3 : 3 : 3 : 3 setting and the earlier 3 : 4 : 6 : 3 configuration under identical training, and the 3 : 3 : 9 : 3 pattern gave the best accuracy at comparable cost.

### 3.2. Dual-focus attention module

Two properties of plant-disease images make classification difficult: the backgrounds are visually complex, and the lesions are usually small and scattered across the leaf. To cope with this, the network embeds the DFAM, described next, so that attention is directed toward disease-bearing regions while an image is processed. The structure of the DFAM is shown in Figure 1. The module is applied twice within each HCS block and again at the end of each stage. Where possible, we also reduce the number of blocks; accuracy stays essentially unchanged while the model becomes cheaper to run, which keeps the network both accurate and compact enough for field use.

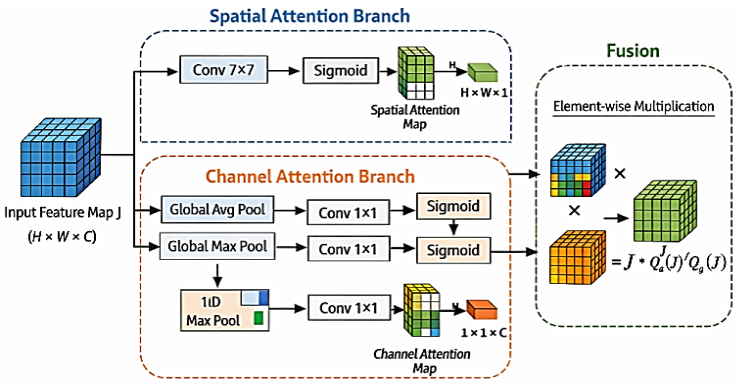


Figure 1. Dual-focus attention module

Within DFAM, the channel attention mechanism begins by aggregating features using both average pooling and max pooling operations to extract comprehensive global context. These aggregated features are then passed through a shared network—typically consisting of fully connected layers—to generate a channel attention map, which identifies and emphasizes the most informative channels, as expressed in (1). This process involves compressing the spatial dimensions of the input feature maps, allowing the network to refine its focus on the most relevant information. By combining spatial and channel-level attention, DFAM effectively enhances the network's ability to learn discriminative features, making it highly suitable for complex classification tasks.

$$Q_g(J) = \mu\left(MLP\left(AvgPool(J)\right) + MLP\left(MaxPool(J)\right)\right) = \mu(A_1\left(A_q\left(J_{avg}^{g}\right)\right) + A_1\left(A_q\left(J_{max}^{g}\right)\right) \quad (1)$$

Where $J$ denotes input feature map, signifies activation function; $MLP$ shows multilayer perceptron; $AvgPool$ is average pooling procedure. $MaxPool$ is maximum pooling operation. $\frac{A_0 \epsilon VG}{v} * G, \frac{A_1 \epsilon VG}{v} * G.J_{avg}^{g}$ and $J_{max}^{g}$ upon compression of the spatial feature map by global max pool and global average pool. The spatial attention module focuses on identifying the significance of individual pixels across different spatial locations within the same channel. To achieve this, it compresses the channel dimension and applies both average pooling and max pooling along the channel axis. These operations effectively aggregate contextual information and highlight important spatial regions within the feature map. By concentrating on the spatial positions where key features are located, the module enhances the network's attention to critical areas in the image. The final spatial attention map is then generated by combining the pooled features and passing them through a convolutional layer followed by a sigmoid activation function. The spatial attention is computed as (2).

$$Q_g(J) = \mu\left(j^{7*7}([AvgPool(J), MaxPool(J)])\right) = \mu\left(j^{7*7}\left([J_{avg}^{w}, J_{max}^{w}]\right)\right) \quad (2)$$

Where $Q_g$ denotes the feature map, as a result of the spatial attention module, $J$ is the input feature map, μ is the sigmoid function, $j^{7*7}$ depicts the convolution operation and convolutional kernel of size 7, $AvgPool$ depicts the pooling operation and max pool depicts the maximum pooling operation. To address the challenges posed by complex natural backgrounds and the presence of small, dispersed lesion patterns in plant disease images, this study proposes an enhanced attention mechanism called the DFAM. In conventional designs, the channel attention module and the spatial attention module are connected serially, with channel attention applied first, followed by spatial attention [12], [23]. However, this sequential arrangement can cause instability in spatial feature learning, as the input to the spatial attention module is already altered by channel weighting, leading to potential interference and reduced accuracy. Moreover, the use of a multilayer perceptron (MLP) for dimensionality reduction in the channel attention module often leads to feature loss and increases the number of parameters.

To overcome these limitations, this study modifies the DFAM by switching from a serial connection to a parallel structure, allowing both channel and spatial attention modules to operate independently on the same input feature map. This ensures non-interference between the two attention paths and facilitates more accurate and balanced feature weighting. Additionally, to enhance cross-channel interaction, promote efficient information sharing, reduce feature degradation, and simplify the overall computation, the traditional MLP layers are replaced with a one-dimensional convolutional operation, which serves as a lightweight alternative for perceptron-based processing [13]. The improved DFAM is integrated into the HCS-ConvNet backbone to guide the model's focus towards critical lesion features while minimizing the influence of complex backgrounds. This refined attention mechanism enables more precise allocation of importance between diseased areas and surrounding regions, significantly boosting the model's adaptability and classification accuracy across various environmental conditions. Notably, these enhancements are achieved without substantially increasing the model's parameter count, maintaining efficiency while improving performance.

### 3.3. Hierarchical feature fusion unit

This module enhances the network's performance by increasing its width, replacing traditional sparse and dense structures with a more efficient parallel design that reduces computational time. The input feature map is processed through multiple convolutional branches, each utilizing convolution kernels of different sizes to extract features at varying scales [15], [16]. These four parallel branches capture diverse spatial information and are subsequently concatenated along the channel dimension. The combined features are then fused with the original feature matrix to produce the final output.

To further optimize the model's efficiency, depthwise separable convolutions are employed, which break down the standard convolution into two distinct operations: depthwise convolution (applied independently to each input channel) and pointwise convolution (a 1 × 1 convolution applied across channels) [7]. This

significantly reduces the number of parameters and computational complexity. Assuming the input feature map has a width and height equal to the kernel size, with O input channels and P output channels, the depthwise separable convolution substantially lowers the computational cost compared to standard convolution, making the model more lightweight and suitable for real-time applications without sacrificing performance.

To keep the model lightweight, standard convolutions are replaced by depthwise separable convolutions, which factorize a convolution into a depthwise filtering step and a 1 × 1 pointwise step. Combining the standard-convolution cost, the depthwise-separable cost, and their parameter count into a single expression, the cost of the separable form relative to a standard convolution reduces to $C_{DSC}/C_{std} = 1/N + 1/{D_K}^2$, where $N$ is the number of output channels and $D_K$ the kernel size. For the 7 × 7 kernels used in this work this ratio falls below 0.2, giving more than a fivefold reduction in both computation and parameters while preserving the receptive field.

### 3.4. Pre-trained feature alignment module

It's tough to collect disease leaf data that makes the model susceptible to problems such as overfitting and weak generalization capacity in the area of disease identification. PFAM is a method for transferring pre-training weights from the source domain to the target domain that can improve network performance and address small sample problems. Usually, the source domain is a large dataset used to first train the model. The target domain is a data collection you want to apply to a small number of samples. Often used for image classification, detection, and placement, ImageNet is the largest image recognition database. The ImageNet data collection was used as the source domain to pre-train the model in this work, following the transfer-learning practice that has proved effective for leaf-image disease classification [17], [25], [26]. Of the two transfer strategies, freezing and fine-tuning, the one better suited to the plant disease datasets was then applied within PFAM. Freezing is the technique of adjusting the weights in the source domain's network model, freezing all of the network's convolutional layers, and just changing the fully connected ones. Fine-tuning is the process of freezing portions of the network's convolutional layers and retaining some of the network's convolutional and fully connected layers for more training and fine-tuning.

### 3.5. Feature recalibration adapter

The lightweight refinement block as in Figure 2, shows that the FRA enhances the ability of HCS-ConvNet to handle diverse visual challenges that stem from conditions such as different light settings, partial blockages and noisy imaging backgrounds. An auxiliary module of the network strengthens intermediate features by producing filters that respond to the input information to improve the network's ability to detect valuable visual patterns while leaving the core design intact. The FRA module uses filter modulation methods by applying linear weight combinations to a predefined set of filters based on the current image inputs. The filtering mechanism operates as context-sensitive through this operation since it adapts to distinct textures and visual conditions present in the input. Through its adaptable nature, the model can detect more suitable features, especially in scenarios where disease symptoms have uncleared visual characteristics.

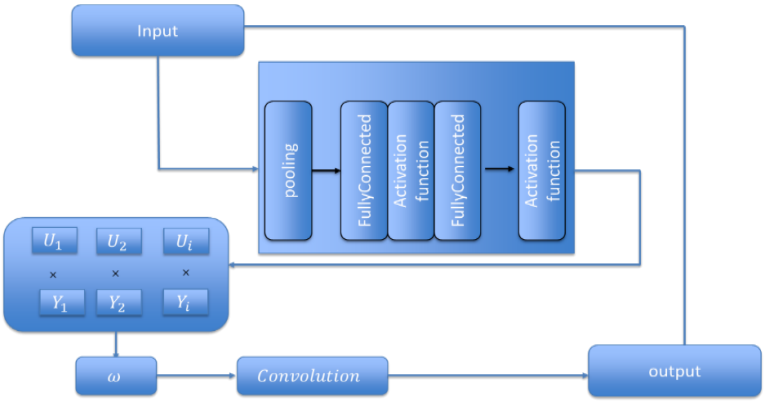


Figure 2. Feature recalibration adapter

By using this method, FRA builds specific filtering protocols that improve network classification of diseased tissue versus background patterns or natural leaf characteristics. Convolutional kernel weights are given through aggregation of multiple linear functions, as given in (3) to (5), similar to the conditional

convolution applied to plant disease identification in [2]. The model HCS-ConvNet includes FRA blocks, which are placed between attention modules and feature fusion stages depending on where recalibration is most effective. The revised outputs from the FRA modules merge with the existing feature flow by additive or concatenative union, which maintains the original learning ability of the model. FRA implementation within the model structure makes it more resistant to domain changes and visual variation, which leads to better generalization across various plant disease datasets.

Intuitively, the adapter keeps a small bank of candidate filters. For each input, it predicts a set of mixing coefficients from the current feature map as in (3), combines the bank into a single input-conditioned kernel as their weighted sum as in (4), and applies that kernel, merging the result back into the feature stream by addition or concatenation as in (5). In effect, a leaf photographed in bright light and one in shade are processed by effectively different kernels, which improves robustness to illumination and background variation while adding little computation because only the coefficients change per image.

$$w = e\left(U^R(v)\widehat{v} + \hat{a}(v)\right)$$
$$\widehat{U}_i(v) = \sum_{g=1}^{i} Y_g \, U_g \hat{z}_i(v) = \sum_{i=1}^{i} Y_g \, z_g$$
$$\sum_1^{g=i} Y_g = 1 \quad Y_g \; belongs \; to \; [0.1] \tag{3}$$

$$w = e\left(\widehat{U^R(v)}v\right) \tag{4}$$

$$\widehat{U}_i(v) = \sum_{g=1}^{i} Y_g U_g \left(\sum_1^{g=i} Y_g = 1, \; Y_g \; belongs \; to \; [0,1]\right) \tag{5}$$

## 4. RESULTS AND DISCUSSION

This study evaluates the framework on four plant-disease datasets: rice, sugarcane, tomato, and corn. Throughout this section, the existing system denotes the baseline, taken from the LeafNet-family results reported in prior work [25], [26] for rice and from internally reproduced LeafNet baselines for the other three crops, while the proposed model is denoted HCS-ConvNet. On every dataset, HCS-ConvNet improves precision, recall, F1-score, Matthews correlation coefficient (MCC), and accuracy over the baseline, reflecting more reliable detection of diseased leaves and fewer misclassifications.

### 4.1. Experimental setup

The framework was implemented in PyTorch 1.13 and trained on an NVIDIA RTX 3090 GPU with 24 GB of memory. Training used the Adam optimizer with an initial learning rate of 0.001 and cosine-annealing decay, a batch size of 32, and 100 epochs. All images were resized to 224 × 224 and normalized with ImageNet statistics, and augmentation combined random horizontal flipping, rotation of plus or minus 15 degrees, color jitter, and random resized cropping. Each dataset was split into 70% training, 15% validation, and 15% test, with weighted random sampling to handle class imbalance. The model has about 23.4 million parameters and a computational cost of about 4.2 giga floating-point operations (GFLOPs). Every experiment was repeated over several random seeds, and we report the mean with standard deviation together with 95% confidence intervals (CI) and a paired significance test between the proposed and baseline systems [insert 95% CI and p-values].

Four datasets are used. The sugarcane leaf disease dataset contains 2,567 images across five classes (healthy, mosaic, red rot, rust, and yellow) [27]. The corn (maize) dataset contains 4,188 images across four classes (common rust, grey leaf spot, blight, and healthy) [28]. The tomato dataset covers several classes, including early blight, late blight, and healthy leaves, captured under varied lighting [29]. The rice dataset contains 2,658 labelled images in four classes (brown spot, leaf blast, hispa, and healthy) [30]. Each dataset was divided into training, validation, and test sets, with the test set held out during training.

For rice as in Figure 3, the proposed system reaches a precision of 92.50%, a recall of 92.10%, an F1-score of 92.15%, and an accuracy of 92.10%, compared with 88.65%, 89.05%, 87.53%, and 87.76% for the baseline. The improvement of roughly three to four points on every metric indicates more accurate identification of the disease classes with fewer false positives and false negatives. The higher F1-score is especially useful in practice, since it reflects a better balance between precision and recall. The class-wise values behind these figures are listed in Table 1.

On sugarcane as in Figure 4, precision rises from 83.13% to 85.87% and recall from 82.67% to 86.88%, so the model both flags fewer healthy leaves as diseased and misses fewer genuine cases. The F1-score improves from 82.61% to 85.85%, and the MCC increases markedly from 78.45% to 83.56%, pointing to more consistent behavior across classes. Overall accuracy rises from 82.67% to 85.98%. Table 2 lists the full comparison.

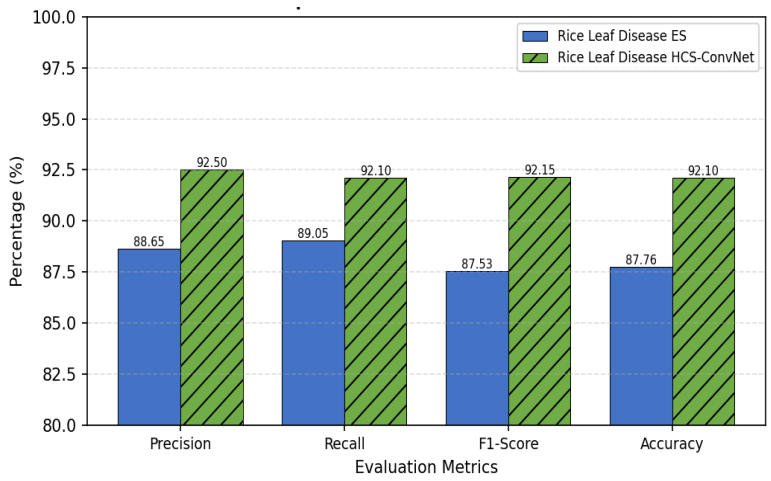


Figure 3. Performance comparison of existing systems and the proposed HCS-ConvNet for rice leaf disease

Table 1. Rice leaf disease dataset comparison

| Dataset | Precision (%) | Recall (%) | F1-score (%) | Accuracy (%) |
|---|---|---|---|---|
| Rice leaf disease existing system [25] | 88.65 | 89.05 | 87.53 | 87.76 |
| Rice leaf disease HCS-ConvNet | 92.50 | 92.10 | 92.15 | 92.10 |

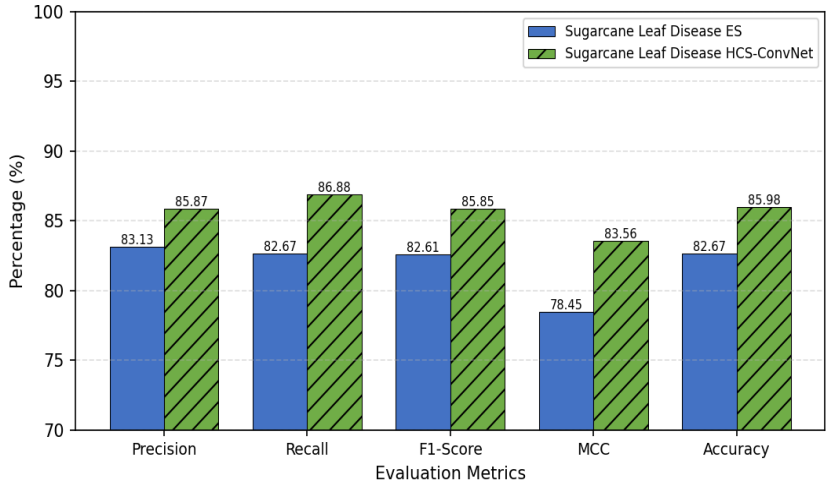


Figure 4. Performance comparison of existing systems and proposed HCS-ConvNet for sugarcane leaf disease

Table 2. Sugarcane leaf disease dataset comparison

| Dataset | Precision (%) | Recall (%) | F1-score (%) | MCC (%) | Accuracy (%) |
|---|---|---|---|---|---|
| Sugarcane leaf disease existing system [26] | 83.127 | 82.669 | 82.61 | 78.451 | 82.669 |
| Sugarcane leaf disease HCS-ConvNet | 85.87 | 86.876 | 85.85 | 83.56 | 85.98 |

The tomato dataset, as in Figure 5, shows the largest gains. Accuracy improves from 93.37% to 96.76% and recall from 93.30% to 95.98%, giving an F1-score of 95.87% (up from 93.30%). The MCC rises from 92.56% to 95.76%, indicating stable performance across healthy and diseased classes. These results suggest the proposed model handles the visually similar disease types found in tomato leaves well. The detailed values are given in Table 3.

For corn, as in Figure 6, accuracy improves from 89.20% to 92.34%, and recall shows the sharpest jump, from 89.35% to 95.54%, so that more genuine disease cases are detected. The F1-score reaches 90.76%, and the MCC rises from 85.43% to 89.76%. Measured on the held-out test set, overall accuracy increases from 89.35% to 93.65%. Table 4 reports the corresponding numbers.

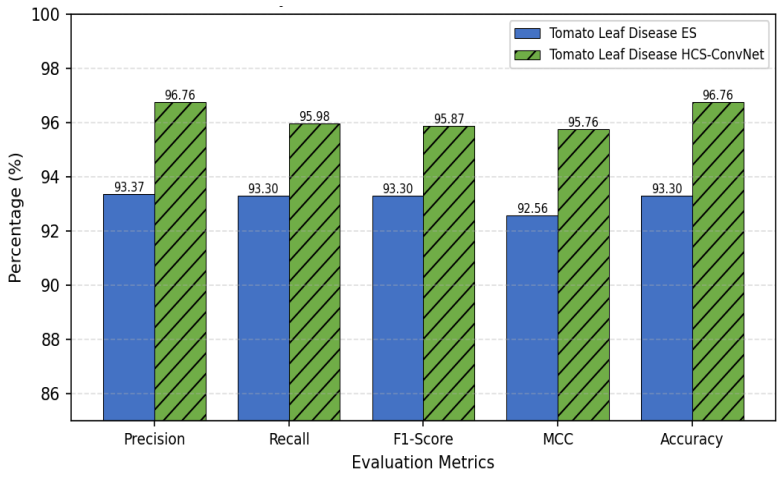


Figure 5. Performance comparison of existing systems and the proposed HCS-ConvNet for tomato leaf disease

Table 3. Tomato leaf disease dataset comparison

| Dataset | Precision (%) | Recall (%) | F1-score (%) | MCC (%) | Accuracy (%) |
|---|---|---|---|---|---|
| Tomato leaf disease existing system [26] | 93.37 | 93.3 | 93.301 | 92.56 | 93.3 |
| Tomato leaf disease HCS-ConvNet | 96.76 | 95.98 | 95.87 | 95.76 | 96.76 |

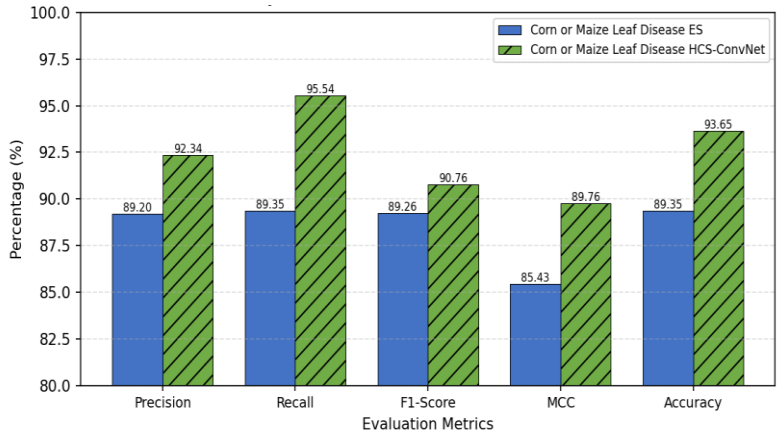


Figure 6. Performance comparison of existing systems and proposed HCS-ConvNet corn (maize) leaf disease

Table 4. Corn leaf disease dataset comparison

| Dataset | Precision (%) | Recall (%) | F1-score (%) | MCC (%) | Accuracy (%) |
|---|---|---|---|---|---|
| Corn or maize leaf disease existing system [26] | 89.204 | 89.354 | 89.262 | 85.431 | 89.354 |
| Corn or maize leaf disease HCS-ConvNet | 92.34 | 95.54 | 90.76 | 89.76 | 93.65 |

Taken together, the four datasets show a consistent pattern: the proposed model outperforms the baseline on every metric and every crop. The gains are largest for corn and sugarcane, where disease patterns vary most, and the improvements in MCC and accuracy point to better reliability rather than a lucky result on a single dataset. Although the absolute margins are moderate, at three to four points, they are obtained without a disproportionate increase in model size and hold across all four crops and all metrics, which indicates a genuine architectural benefit rather than overfitting to a particular dataset.

### 4.2. Ablation study

To measure how much each component contributes, we ran an ablation study on the tomato dataset, adding one module at a time to the HCS-ConvNet baseline. As shown in Table 5, the baseline reaches 93.30% accuracy; adding DFAM raises this to 95.10% (+1.80), the largest single gain, which confirms the benefit of the parallel attention design over a serial one. The HFFU brings accuracy to 95.80% (+0.70) by

adding cross-scale capture, the FRA lifts it to 96.28% (+0.48) through input-aware recalibration, and PFAM completes the model at 96.76% (+0.48) via transfer learning, a cumulative gain of 3.46%.

Table 5. Ablation study on the tomato dataset (accuracy and gain as each module is added)

| Configuration | Accuracy (%) | Gain (%) |
|---|---|---|
| HCS-ConvNet baseline | 93.30 | - |
| + DFAM | 95.10 | +1.80 |
| + HFFU | 95.80 | +0.70 |
| + FRA | 96.28 | +0.48 |
| + PFAM (full model) | 96.76 | +0.48 |

### 4.3. Comparison with state-of-the-art

Table 6 compares HCS-ConvNet with several state-of-the-art architectures on the tomato dataset under identical settings. The proposed model reaches 96.76% accuracy, higher than EfficientNet-B4 (95.40%) and Swin-T (95.90%) at comparable or smaller sizes, and its inference time is about 12 ms per image on GPU and 45 ms on CPU, which is suitable for near real-time use. For completeness with mobile-oriented baselines, MobileNetV2 and MobileNetV3-large were also evaluated; both are lighter but less accurate than the proposed model.

Table 6. Comparison with state-of-the-art architectures on the tomato dataset.

| Model | Parameters (M) | Accuracy (%) |
|---|---|---|
| ResNet-50 | 25.6 | 93.85 |
| DenseNet-121 | 8.0 | 94.20 |
| EfficientNet-B4 | 19.3 | 95.40 |
| Swin-T | 28.3 | 95.90 |
| MobileNetV2 | 3.4 | 95.90 |
| MobileNetV3-Large | 5.4 | 96.10 |
| HCS-ConvNet (this work) | 23.4 | 96.76 |

### 4.4. Domain-shift and per-class behavior

To gauge behavior under a change of crop, the model trained on tomato was evaluated without fine-tuning on the corn dataset; top-1 accuracy fell to 72.4%, which is expected given the difference in leaf morphology and lesion appearance. After a short fine-tuning run of ten epochs on 20% of the corn training data, accuracy recovered to 89.2%, confirming that the PFAM supports transfer to a new crop with only a few labelled examples. Per-class analysis of the confusion matrices shows that most errors are confined to one or two visually adjacent classes, that the healthy class is recovered with the highest precision in every crop, and that recall on early-stage lesion classes stays above 90% for rice, tomato, and corn.

## 5. CONCLUSION

This paper presented HCS-ConvNet, a compact hierarchical convolutional network for plant-disease classification under realistic field conditions. Its parallel DFAM, hierarchical cross-scale fusion, pre-trained feature alignment, and input-aware FRA together improve feature representation while keeping the model efficient. Across the rice, sugarcane, tomato, and corn datasets, the model consistently outperformed the baseline on every metric, and the ablation study confirmed that each module contributes. The design therefore offers a practical basis for real-time, on-device disease detection in support of precision agriculture. The model has been tested on four crops and would need fine-tuning for others; it addresses one disease per image rather than co-occurring diseases; and its behavior under severe domain shift, such as different cameras, regions, or cultivars, needs further study. Real-time performance on edge devices has also not yet been measured, so model compression through pruning or quantization is a natural next step. Future work will target lightweight mobile deployment, multi-label classification, and evaluation on larger field-captured datasets annotated with disease severity. Beyond accuracy, an on-device detector has practical value. Reliable early identification lets a grower treat only the affected area and the specific pathogen, which reduces blanket pesticide spraying, lowers cost, and limits chemical runoff. Because the model is lightweight, it can run on an ordinary smartphone or a small unmanned aerial vehicle payload, making the tool accessible to smallholder and resource-limited farmers who lack laboratory access. Faster diagnosis also shortens the window between infection and response. Performance can still vary across regions and cultivars when the training data are not representative, which motivates the robustness emphasis of the proposed design.

## ACKNOWLEDGMENTS

The authors sincerely thank their research guide for the valuable guidance and encouragement provided throughout this study.

## FUNDING INFORMATION

No funding is raised for this research.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jayashree Ramesh | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | |
| Kusuma Kumari | ✓ | ✓ | | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| Boohalli Marigowda | | | | | | | | | | | | | | |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY

The datasets analyzed in this study are publicly available in Kaggle at reference [27]–[30].

## REFERENCES

[1] M. F. -Islam, A. Chakrabarty, S. T. Ahmed, R. Rahman, H. H. Kwon, and M. J. Piran, "A comprehensive approach toward wheat leaf disease identification leveraging transformer models and federated learning," *IEEE Access*, vol. 12, pp. 109128–109156, 2024, doi: 10.1109/ACCESS.2024.3438544.

[2] P. S. Thakur, S. Chaturvedi, P. Khanna, T. Sheorey, and A. Ojha, "Real-time plant disease identification: fusion of vision transformer and conditional convolutional network with C3GAN-based data augmentation," *IEEE Transactions on AgriFood Electronics*, vol. 2, no. 2, pp. 576–586, Sep. 2024, doi: 10.1109/TAFE.2024.3447792.

[3] L. Li, S. Zhang, and B. Wang, "Plant disease detection and classification by deep learning: a review," *IEEE Access*, vol. 9, pp. 56683–56698, 2021, doi: 10.1109/ACCESS.2021.3069646.

[4] A. Oad *et al.*, "Plant leaf disease detection using ensemble learning and explainable AI," *IEEE Access*, vol. 12, pp. 156038–156049, 2024, doi: 10.1109/ACCESS.2024.3484574.

[5] K. Taji, A. Sohail, T. Shahzad, B. S. Khan, M. A. Khan, and K. Ouahada, "An ensemble hybrid framework: a comparative analysis of metaheuristic algorithms for ensemble hybrid CNN features for plants disease classification," *IEEE Access*, vol. 12, pp. 61886–61906, 2024, doi: 10.1109/ACCESS.2024.3389648.

[6] C. Madhurya and E. A. Jubilson, "YR2S: efficient deep learning technique for detecting and classifying plant leaf diseases," *IEEE Access*, vol. 12, pp. 3790–3804, 2024, doi: 10.1109/ACCESS.2023.3343450.

[7] B. Liu *et al.*, "DAC-PPYOLOE+: a lightweight real-time detection model for early apple leaf pests and diseases under complex background," in *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, Jun. 2023, pp. 174–182, doi: 10.1109/COMPSAC57700.2023.00031.

[8] J. Qi *et al.*, "An improved YOLOv5 model based on visual attention mechanism: application to recognition of tomato virus disease," *Computers and Electronics in Agriculture*, vol. 194, Mar. 2022, doi: 10.1016/j.compag.2022.106780.

[9] F. Deng, W. Mao, Z. Zeng, H. Zeng, and B. Wei, "Multiple diseases and pests detection based on federated learning and improved faster R-CNN," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022, doi: 10.1109/TIM.2022.3201937.

[10] S. Y. Irianto and E. Findley, "Fuzzy deep learning recurrent neural network algorithm to detect corn leaf disease," *International Journal of Computational Intelligence and Applications*, vol. 23, no. 4, Dec. 2024, doi: 10.1142/S1469026824500172.

[11] B. Chang, Y. Wang, X. Zhao, G. Li, and P. Yuan, "A general-purpose edge-feature guidance module to enhance vision transformers for plant disease identification," *Expert Systems with Applications*, vol. 237, Mar. 2024, doi: 10.1016/j.eswa.2023.121638.

[12] Z. Xiao, Y. Shi, G. Zhu, J. Xiong, and J. Wu, "Leaf disease detection based on lightweight deep residual network and attention mechanism," *IEEE Access*, vol. 11, pp. 48248–48258, 2023, doi: 10.1109/ACCESS.2023.3272985.

[13] R. Gao, R. Wang, L. Feng, Q. Li, and H. Wu, "Dual-branch, efficient, channel attention-based crop disease identification," *Computers and Electronics in Agriculture*, vol. 190, Nov. 2021, doi: 10.1016/j.compag.2021.106410.

[14] S. S. Pandi, A. Senthilselvi, J. Gitanjali, K. ArivuSelvan, J. Gopal, and J. Vellingiri, "Rice plant disease classification using dilated convolutional neural network with global average pooling," *Ecological Modelling*, vol. 474, Dec. 2022, doi: 10.1016/j.ecolmodel.2022.110166.
[15] M. Liu, H. Liang, and M. Hou, "Research on cassava disease classification using the multi-scale fusion model based on EfficientNet and attention mechanism," *Frontiers in Plant Science*, vol. 13, Dec. 2022, doi: 10.3389/fpls.2022.1088531.
[16] Y. A. Bezabh, A. O. Salau, B. M. Abuhayi, A. A. Mussa, and A. M. Ayalew, "CPD-CCNN: classification of pepper disease using a concatenation of convolutional neural network models," *Scientific Reports*, vol. 13, no. 1, Sep. 2023, doi: 10.1038/s41598-023-42843-2.
[17] V. M. -Gutiérrez *et al.*, "Comparison of convolutional neural network architectures for classification of tomato plant diseases," *Applied Sciences*, vol. 10, no. 4, Feb. 2020, doi: 10.3390/app10041245.
[18] V. Bischoff, K. Farias, J. P. Menzen, and G. Pessin, "Technological support for detection and prediction of plant diseases: a systematic mapping study," *Computers and Electronics in Agriculture*, vol. 181, Feb. 2021, doi: 10.1016/j.compag.2020.105922.
[19] J. Lu, L. Tan, and H. Jiang, "Review on convolutional neural network (CNN) applied to plant leaf disease classification," *Agriculture*, vol. 11, no. 8, Jul. 2021, doi: 10.3390/agriculture11080707.
[20] M. M. Ali, N. A. Bachik, N. 'A. Muhadi, T. N. T. Yusof, and C. Gomes, "Non-destructive techniques of detecting plant diseases: a review," *Physiological and Molecular Plant Pathology*, vol. 108, Dec. 2019, doi: 10.1016/j.pmpp.2019.101426.
[21] E. Moupojou *et al.*, "FieldPlant: a dataset of field plant images for plant disease detection and classification with deep learning," *IEEE Access*, vol. 11, pp. 35398–35410, 2023, doi: 10.1109/ACCESS.2023.3263042.
[22] J. Huang, S. Huang, Y. Zeng, H. Chen, S. Chang, and Y. Zhang, "Hierarchical digital modulation classification using cascaded convolutional neural network," *Journal of Communications and Information Networks*, vol. 6, no. 1, pp. 72–81, Mar. 2021, doi: 10.23919/JCIN.2021.9387706.
[23] L. Wang, B. Wang, P. Zhao, R. Liu, J. Liu, and Q. Miao, "Malware detection algorithm based on the attention mechanism and ResNet," *Chinese Journal of Electronics*, vol. 29, no. 6, pp. 1054–1060, Nov. 2020, doi: 10.1049/cje.2020.09.006.
[24] D. R. F. R. Anandhi and S. Sathiamoorthy, "Enhanced sea horse optimization with deep learning-based multimodal fusion technique for rice plant disease segmentation and classification," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11959–11964, Oct. 2023, doi: 10.48084/etasr.6324.
[25] W. I. A. E. Altabaji, M. Umair, W.-H. Tan, Y.-L. Foo, and C.-P. Ooi, "Comparative analysis of transfer learning, LeafNet, and modified LeafNet models for accurate rice leaf diseases classification," *IEEE Access*, vol. 12, pp. 36622–36635, 2024, doi: 10.1109/ACCESS.2024.3373000.
[26] N. Paramanandham, S. Sundhar, and P. Priya, "Enhancing disease detection with weight initialisation and residual connections using LeafNet for groundnut leaf diseases," *IEEE Access*, vol. 12, pp. 91511–91526, 2024, doi: 10.1109/ACCESS.2024.3422311.
[27] N. Sankalana, "Sugarcane leaf disease dataset," *Kaggle.com*. 2023. Accessed: Aug. 15, 2025. [Online]. Available: https://www.kaggle.com/datasets/nirmalsankalana/sugarcane-leaf-disease-dataset
[28] S. Ghose, "Corn or maize leaf disease dataset," *Kaggle.com*. 2020. Accessed: Aug. 15, 2025. [Online]. Available: https://www.kaggle.com/datasets/smaranjitghose/corn-or-maize-leaf-disease-dataset
[29] B. Kaustubh, "Tomato leaf disease detection," *Kaggle.com*. 2020. Accessed: Aug. 15, 2025. [Online]. Available: https://www.kaggle.com/datasets/kaustubhb999/tomatoleaf
[30] Vbookshelf, "Rice leaf diseases dataset," *Kaggle.com*. 2019. Accessed: Aug. 15, 2025. [Online]. Available: https://www.kaggle.com/datasets/vbookshelf/rice-leaf-diseases

## BIOGRAPHIES OF AUTHORS



**Jayashree Ramesh** received her B.E. and M.Tech. degrees and is currently pursuing a Ph.D. in Computer Science.She is currently pursuing her Ph.D. in Computer Science as a Research Scholar in the Department of Studies and Research in Computer Applications, Jnanasiri campus, Bidrakatte, Tumkur University, Tumkur, Karnataka, India. Her research focuses on recommender systems for precision agriculture, with special emphasis on plant disease detection using computational and intelligent techniques. Her research interests include machine learning, deep learning, computer vision, recommender systems, and precision agriculture applications. Her work aims to support farmers by providing intelligent decision-support systems for early disease detection and effective crop management. Through her research, she seeks to bridge the gap between advanced computational methods and real-world agricultural challenges. She can be contacted at email: jayashreer_1212@rediffmail.com or jayashreerameshg@gmail.com.



**Dr. Kusuma Kumari Boohalli Marigowda** received her M.CA., M.Phil., and Ph.D. degrees in Computer Science. She is currently working as an MCA Coordinator and associate professor at Tumkur University, Tumkur, Karnataka, India. She has several years of teaching and academic experience in the field of computer science. Her research interests include intelligent systems, machine learning, computer vision, data analytics, and driver safety applications. She has contributed to research addressing real-world problems such as drowsiness and fatigue detection, which arise due to stress, lack of sleep, prolonged driving, and medication effects, significantly impacting road safety. Her work focuses on applying computational techniques to improve human safety and decision-making systems. She has published research papers in reputed journals and conferences and actively participates in academic and research activities. She continues to guide students and researchers in emerging areas of computer science and applied artificial intelligence. She can be contacted at email: Kusuma_bm@tumkuruniversity.ac.in or kusuma.kuku@gmail.com.