Information Processing in Agriculture xxx (xxxx) xxx

Contents lists available at ScienceDirect

Information Processing in Agriculture

journal homepage: www.keaipublishing.com/en/journals/information-processing-in-agriculture/

# Robust eggplant disease recognition using a learnable weighted deep ensemble with test-time augmentation

Ebru Ergün [a,*], Hatice Okumus [b]

[a] Department of Electrical and Electronics Engineering, Faculty of Engineering and Architecture, Recep Tayyip Erdogan University, Rize, Turkey
[b] Department of Electrical and Electronics Engineering, Faculty of Engineering, Karadeniz Technical University, Trabzon, Turkey

ARTICLE INFO

*Keywords:*
Artificial intelligence
Deep learning
Eggplant diseases
Ensemble learning
Image classification
Plant disease detection

ABSTRACT

One of the most extensively cultivated vegetables worldwide, eggplant is significantly affected by a wide range of diseases that reduce both yield and quality. Ensuring sustainable crop production and food security therefore requires early and reliable disease diagnosis. The rapid evolution of deep learning architectures and computer vision techniques has established convolutional neural networks as powerful tools for plant disease recognition, often outperforming traditional diagnostic approaches. In this study, three benchmark eggplant image datasets with distinct class structures and imbalance characteristics (6-class Eggplant1, 5-class Eggplant2, and 7-class Eggplant3) were utilized to develop a robust disease classification framework. The proposed methodology introduces a learnable weighted ensemble that adaptively integrates ConvNeXt, DenseNet, and EfficientNet architectures within an end-to-end trainable fusion scheme. The framework is further reinforced by systematic test-time augmentation and cross-validation to enhance inference stability. Experimental evaluation demonstrates that the ensemble model achieves accuracies of 0.9970, 0.9375, and 0.9557 on Eggplant1, Eggplant2, and Eggplant3, respectively, while maintaining balanced performance across heterogeneous class distributions. These results confirm the effectiveness of the adaptive ensemble fusion in capturing complementary feature representations and mitigating inter-class variability across datasets with differing levels of complexity. Overall, this work contributes a methodologically robust and interpretable ensemble-based framework that advances more dependable image-based plant disease diagnosis by addressing practical reliability concerns in precision agriculture applications.

## 1. Introduction

Precision agriculture is undergoing a significant transformation driven by the convergence of computer vision and deep learning (DL), where image-based disease diagnosis has emerged as a key component of sustainable crop management [1]. Modern convolutional neural networks (CNNs), capable of learning hierarchical visual representations from raw images, can detect fine-grained variations in leaf color and morphology, often invisible to the human eye. This capability is particularly critical for eggplant, an economically important crop worldwide [2] that remains highly susceptible to a broad spectrum of pathogens and pests. Despite substantial progress, a persistent gap remains between laboratory-level accuracy and robust performance under real-world field conditions. Challenges such as variable illumination, complex backgrounds, and occlusions often render traditional manual scouting and earlier automated methods unsuitable for large-scale monitoring [2–4]. Consequently, numerous studies have investigated diverse DL architectures—ranging from transfer learning and hybrid feature fusion to lightweight models for edge deployment—aimed at improving generalization and classification accuracy (ACC) in eggplant disease recognition [5]. Haque et al. introduced a two-stream deep fusion framework integrating CNN–Support Vector Machine (SVM) and CNN–Softmax classifiers, supported by an inference model and successfully distinguished nine eggplant diseases using a dataset of 2,284 RGB images, achieving notably higher ACC and fewer false positives than widely used models such as VGG16, Inception V3 and ResNet50 [6]. Saad et al. employed transfer learning using DenseNet201, Xception and ResNet152V2 to diagnose fourteen common eggplant diseases in Bangladesh reporting that DenseNet201 achieved 99.06% ACC and significantly reduced diagnostic errors in farmer-oriented settings [7]. Kursun et al. examined AlexNet feature extraction combined with Random Forest classification on a dataset of 4089 eggplant leaf images

* Corresponding author.
*E-mail address:* ebru.yavuz@erdogan.edu.tr (E. Ergün).

https://doi.org/10.1016/j.inpa.2026.03.010
Received 23 December 2025; Received in revised form 12 March 2026; Accepted 15 March 2026
Available online 18 March 2026

and demonstrated the importance of augmentation-based balancing improving ACC from 53.38% to 74.64% [8]. Siddiqui et al. proposed an advanced ensemble learning framework incorporating ConvNeXt, EfficientNetV2, NFNet and RegNet utilizing more than 19,800 images and robust preprocessing and achieved 99.76% ACC while also deploying a real-time web application [9]. Other studies have explored complementary modalities and diagnostic strategies. Zhang et al. integrated low-cost multispectral imaging with a VGG16-triplet attention model for early detection of Verticillium wilt reaching a precision (PRC) of 86.73% in early infection stages [10]. Kaniyassery et al. analyzed correlations between environmental factors and disease incidence in the Mattu Gulla variety and developed the "Leaf Guard" Android application achieving 98.20% ACC in real-time detection [11]. Meng et al. presented a few-shot learning approach using the MCA-RepVGG-A-B3 network with Randaugment-based feature diversification attaining 89.76% ACC under limited data scenarios [12]. Xie et al. proposed YOLOv5s-BiPCNeXt, a lightweight architecture incorporating MobileNeXt, attention modules and feature recombination achieving up to 99.50% ACC for healthy leaves and over 94.00% ACC for early-stage disease categories at real-time speeds suitable for edge devices [13]. Wang et al. introduced a multimodal data fusion framework integrating image data with sensor information using embedded attention mechanisms, reporting 92.00% ACC and strong localization performance [14]. Gayathri et al. evaluated YOLOv8 with Faster R-CNN, DenseNet201, Xception, ResNet152V2 and VGG16-triplet multispectral imaging achieving classification ACCs exceeding 99.00% in several configurations illustrating the advantage of multi-architecture and multimodal fusion [15].

Further works have expanded to multi crop recognition contexts. Rangarajan et al. evaluated texture features alongside DL models for eggplant and tomato disease identification demonstrating that AlexNet achieved 100.00% ACC for crop identification and above 80.00% for disease classification [16]. Abisha et al. created a hybrid model that combined Radial Basis Feed Forward Neural Networks and deep CNN, enhanced by Gaussian filtering and shearlet-based feature extraction. This model achieved an ACC of up to 93.30% across five important eggplant diseases [17]. In order to provide vital information for resistance breeding efforts, Karthikeyan et al. conducted genetic analyses of phytoplasma-induced little leaf disease and discovered resistant accessions [18]. With accuracies over 99.00% across various spectral configurations, several studies have examined dataset generation and multispectral image analysis for important eggplant illnesses [19]. Hyperspectral imaging research has also contributed non-invasive diagnostic strategies such as grey mold stress detection using PLSR, LS-SVM and BPNN models [20], early identification of Botrytis cinerea infection using visible/NIR reflectance combined with PCA and BP neural networks [21] and disease-level classification using hyperspectral vegetation indices integrated with SVM [22].

Recent studies have expanded intelligent agricultural systems beyond leaf-level disease classification. Raza et al. benchmarked multiple YOLO-family variants for real-time crop growth and weed detection in cotton fields, revealing a clear trade-off between detection accuracy and inference efficiency across model scales. Their results suggested that the main bottleneck was small-object localization under field conditions rather than the network design itself [23]. Complementarily, Bakr et al. compared traditional machine learning, DL, and GPT-based models for crop recommendation on transformed agricultural data, showing that GPT-based approaches can deliver competitive performance while enabling natural language interaction for decision support [24]. In plant disease recognition, dense–inception architectures with attention mechanisms have also demonstrated improved discrimination by emphasizing disease-relevant regions in complex images [25].

Collectively, these studies demonstrate the breadth of methodological innovation in eggplant disease detection, covering RGB-based CNN models, multispectral and hyperspectral sensing, few-shot learning, multimodal fusion, lightweight mobile architectures and ensemble transfer learning. Table 1 provides a detailed summary of the referenced literature including methods, dataset characteristics, class numbers and performance metrics. Despite these advancements, three critical limitations hinder practical deployment: (1) single-architecture sensitivity to domain shifts and imaging variations, (2) insufficient cross-dataset generalization across different agricultural settings and (3) vulnerability to test-time perturbations including illumination changes, occlusions and background clutter. These issues are particularly pronounced in eggplant disease diagnosis where inter-class similarity and intra-class variability demand exceptionally robust feature representations. While ensemble learning offers a compelling solution by integrating multiple architectures that contribute diverse feature representations, conventional ensembles often rely on fixed or heuristic weighting schemes that fail to adapt to specific test conditions.

This paper introduces a novel differentiable ensemble framework that integrates ConvNeXt, DenseNet, and EfficientNet through a learnable adaptive weighting mechanism, further strengthened by systematic TTA. The key innovation lies in an end-to-end trainable ensemble structure that dynamically optimizes the contribution of each backbone network, combined with a multi-view inference strategy designed to improve resilience under real-world agricultural variability. Despite substantial progress in eggplant disease classification using DL, several critical limitations remain. Most prior studies rely on single architectures or fixed-weight ensemble strategies evaluated on a single dataset, thereby restricting their ability to generalize across heterogeneous domains. Systematic cross-dataset evaluation remains scarce, particularly under varying acquisition conditions. Moreover, sensitivity to illumination changes, background complexity, and partial occlusions continues to challenge conventional models, especially when static or heuristic fusion schemes are employed. To address these limitations, we propose an adaptive ensemble framework capable of dynamically balancing complementary feature representations while sustaining stable performance across diverse datasets. The proposed approach is rigorously validated on three heterogeneous eggplant disease datasets (6-class Eggplant1, 5-class Eggplant2, and 7-class Eggplant3), achieving accuracies of 0.9970, 0.9375, and 0.9557, respectively. The main contributions of this study can be summarized as follows:

- A learnable weighted ensemble framework is proposed enabling end-to-end optimization of heterogeneous CNN backbones instead of relying on fixed or heuristic ensemble averaging.
- Test-time augmentation (TTA) is systematically integrated with adaptive ensemble fusion to enhance inference stability and reduce sensitivity to controlled visual variability.
- A multi-dataset evaluation protocol is conducted across three eggplant disease benchmarks with differing class structures allowing assessment of relative cross-dataset robustness under controlled conditions.
- Comprehensive performance analysis beyond accuracy is provided using class-balanced metrics and fold-level stability to demonstrate reliability as well as peak performance.
- Computational efficiency is explicitly analyzed highlighting the trade-offs between accuracy gains and resource requirements for practical deployment in precision agriculture.

## 2. Materials and methods

### 2.1. Datasets

The three datasets used in this study were collected independently and differ in acquisition devices, imaging conditions, target plant organs and class distributions. All datasets were annotated and validated by agricultural experts or plant pathologists following the original protocols ensuring label reliability. Although potential biases such as geographic concentration and limited cultivar diversity remain, these datasets provide heterogeneous and realistic benchmarks for controlled

**Table 1**
Comparative analysis of recent DL approaches for eggplant disease classification.

| Ref. | Method Category | Core Architecture/ Approach | Dataset Size/ Type | Target Diseases | Class Number | Performance | Key Limitations |
|---|---|---|---|---|---|---|---|
| [6] | Hybrid Fusion | CNN-SVM + CNN-Softmax | 2284 images | Multiple pest-, fungal-, bacterial-, and viral-related eggplant diseases. | 9 | 98.90% ACC | Single-model dependence, no TTA |
| [7] | Transfer Learning | DenseNet201 | 2766 images | Diverse insect infestations and fungal/viral foliar disorders. | 14 | 99.06% ACC | Fixed feature extraction, no ensemble |
| [8] | Feature Engineering | AlexNet + Random Forest | 4089 leaf images augmented to 12,000 | Healthy and major fungal, viral, and pest-induced leaf diseases. | 6 | 74.64% ACC | Traditional ML limitations |
| [9] | Ensemble Learning Multimodal | ConvNeXt + EfficientNetV2 + NFNet + RegNet | 1400 initial images and 9800 augmented images | Healthy class with representative fungal, viral, and pest-related conditions. | 7 | 99.76% ACC | Fixed ensemble weights, limited augmentation |
| | | VGG16-triplet + multispectral | 3,116 original image and 10,000 augmented images | Mixed pathological and pest-associated conditions including blights and viral infections. | 10 | 99.71% ACC | Single-model dependence, no TTA |
| [10] | Multimodal | VGG16-triplet + multispectral | 2348 images | Verticillium wilt and healthy class. | 2 | 86.73% PRC | Early-stage detection only |
| [11] | Mobile App | Correlation analysis | – | Leaf and fruit diseases (spot and rot categories). | 2 | 98.20% ACC | Limited validation |
| [12] | Few-Shot Learning | MCA-RepVGG-A-B3 | 1400 images | Healthy and common foliar diseases (fungal, viral, and pest-induced). | 7 | 89.76% ACC | Low-data scenario focus |
| [13] | Lightweight | YOLOv5s-BiPCNeXt | – | Brown spot, powdery mildew, and healthy class. | 3 | 94.9% ACC (brown spot)95.0% ACC (powdery mildew) 99.5% ACC (healthy) | Object detection only |
| [14] | Multimodal Fusion | Attention mechanisms | 6,658 | Fungal and bacterial leaf diseases including wilt and mold types. | 5 | 92.00% ACC | Complex sensor requirements |
| [15] | Comparative Study | DenseNet20, Xception, etc. | 2,766 | Healthy class and representative foliar diseases. | 5 | 99.06% ACC | No novel methodology |
| [16] | Texture Analysis | AlexNet | 135 images | Binary classification: healthy vs. diseased. | 2 | 100.00% ACC | Very small dataset |
| [17] | Hybrid CNN | DCNN + RBFNN | 1500 images | Bacterial, fungal, and viral pathogen classes. | 5 | 93.3% ACC | Computational complexity |
| [18] | Feature Extraction | VGG16 + MSVM | 1088 images | Pest infestation and viral/fungal leaf diseases. | 5 | 99.4% ACC | Traditional classifiers |
| [19] | Hyperspectral | SVM with RBF kernel | – | Binary classification: healthy vs. unhealthy. | 2 | 97.00% ACC | Specialized hardware needed |

evaluation of model robustness under varying acquisition conditions.

#### 2.1.1. Eggplant1

The first dataset, hereafter referred to as Eggplant1, was developed by Hasan et al. to facilitate DL research on eggplant leaf diseases [26]. It contains a total of 1,338 high-resolution images all captured in natural daylight using a Canon EOS 1300D DSLR camera as given in Table 2. This dataset is particularly suitable for tasks requiring fine-grained feature extraction because of its outstanding detail provided by its 4000 × 6000 pixel resolution. Image acquisition took place over a five-day period across two agricultural sites in Bangladesh: Changao, Savar, Dhaka and Rayerdia, Kaliganj, Gazipur. The dataset is divided into six disease-related classes, with class labels validated by agricultural experts to ensure reliability. The class distribution is as follows: healthy, insect pest, leaf spot, mosaic virus, small leaf and wilt. The dataset is organized into a structured directory, where each disease class is represented by a separate folder containing JPG images as seen in Fig. 1. Additional resources include a README file describing data collection protocols and a metadata file mapping filenames to class labels.

**Table 2**
Image distribution in the Eggplant1 dataset across the six categories.

| Class Name | Eggplant1 |
|---|---|
| | **Number of original images** |
| Healthy | 267 |
| Insect Pest | 189 |
| Leaf Spot | 232 |
| Mosaic Virus | 189 |
| Small Leaf | 225 |
| Wilt | 236 |
| Total | 1,338 |

#### 2.1.2. Eggplant2

Introduced by Hasan et al. [27], the Eggplant2 dataset captures a different aspect of eggplant health assessment by focusing on fruit-specific diseases rather than foliar symptoms. This collection comprises 1,823 curated images distributed across five clinically significant categories: fruit cracking, healthy, phomopsis blight, shoot and fruit borer and wet rot. Unlike balanced laboratory collections Eggplant2 exhibits a natural class imbalance ranging from 161 images for phomopsis blight to 725 for shoot and fruit borer providing a realistic benchmark for evaluating model robustness under skewed real-world distributions. Systematic field surveys were conducted in the two main brinjal-growing regions of Bangladesh (Bogura and Dhaka) in order to collect data. 2,273 raw images were recorded using 190 smartphone cameras in a variety of field conditions and with varying natural lighting. 1,823 photos that preserved real-world field conditions, such as occlusions, soil residues and natural fruit orientations, were kept after expert validation and quality screening. Where feasible, the original aspect ratios of each image were preserved while standardizing them to 224 × 224 pixels. The dataset's hierarchical organization and representative samples are illustrated in Fig. 2 demonstrating both structural clarity and visual diversity. Table 3 summarizes the class distribution, emphasizing the deliberate imbalance that puts traditional classification techniques to the test. In addition to leaf-based diagnostics, this dataset

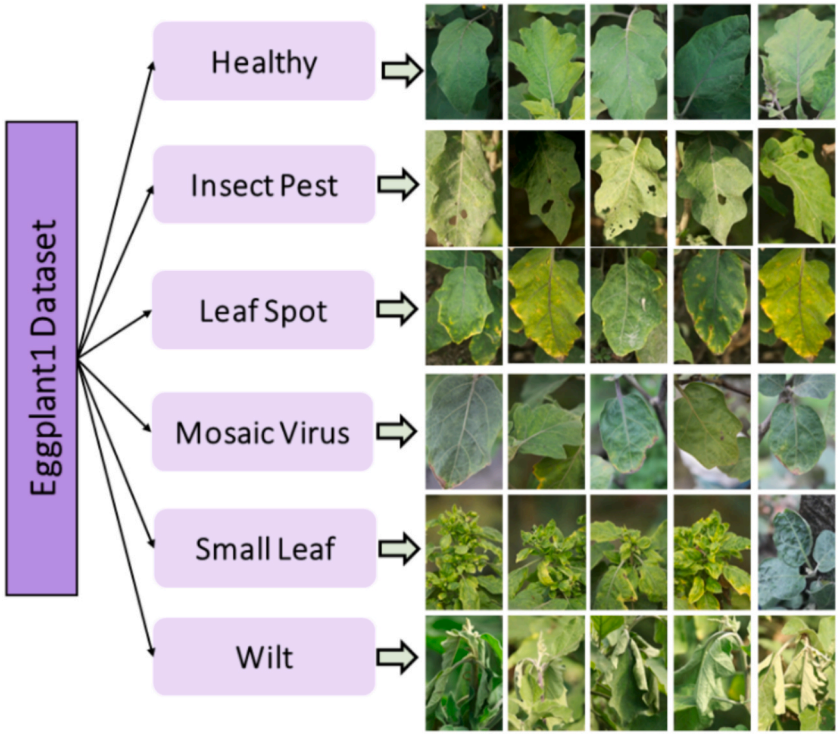


**Fig. 1.** Folder structure and representative samples from the Eggplant1 dataset [26].

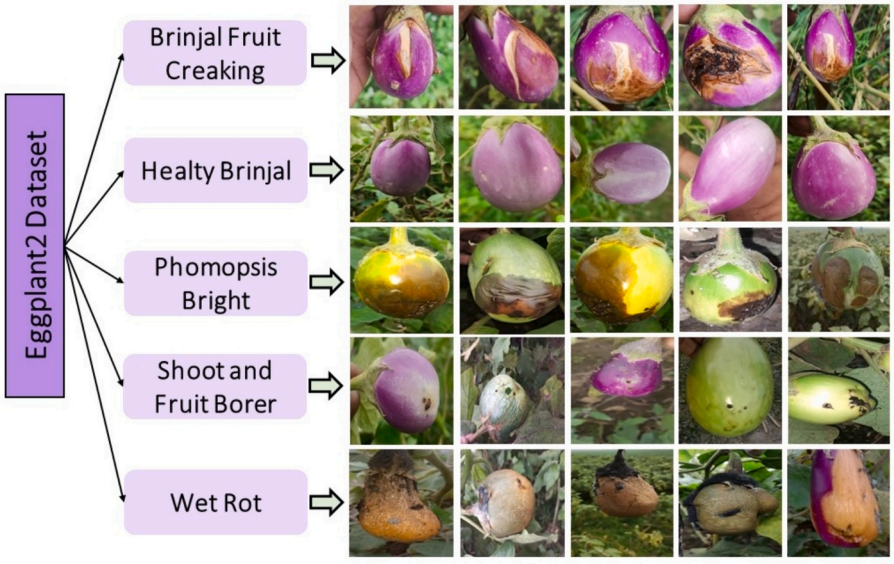


**Fig. 2.** Overview of the folder structure and representative samples in the Eggplant2 dataset [27].

**Table 3**
Image distribution across the five categories in the Eggplant2 dataset.

| Class Name | Eggplant2 |
|---|---|
| | **Number of original images** |
| Wet Rot | 223 |
| Phomopsis Blight | 161 |
| Shoot and Fruit Borer | 725 |
| Fruit Cracking | 200 |
| Healthy | 514 |
| Total | 1,823 |

specifically covers fruit-centric disease manifestation, allowing for a thorough evaluation of plant health.

### 2.1.3. *Eggplant3*

Siddique Chaity et al.'s Eggplant3 dataset [28] provides a carefully balanced standard for classifying eggplant leaf diseases in real world settings. Each of the seven clinically specified categories—healthy, insect pest, leaf spot, mosaic virus, small leaf, white mold and wilt—is equally represented in this collection of 1,400 high-resolution JPG pictures. This intentional balance, 200 images per class, eliminates confounding effects of class imbalance enabling direct evaluation of model discriminative capability rather than bias compensation. Ecological

validity was given top priority throughout data collection and photos were taken in a variety of growing seasons with different lighting conditions (morning, noon and evening) and natural backgrounds, such as soil, mulch and intercropped vegetation. The dataset is unique in that it includes both biotic and abiotic stresses, such as fungal pathogens (white mold), viral infections (mosaic virus), physiological problems (small leaf) and pest damage (insect pest). Agricultural pathologists validated each image twice guaranteeing both annotation consistency and diagnostic accuracy. The dataset's methodical arrangement and visual diversity are demonstrated in Fig. 3 which shows exemplary samples that emphasize intra-class variability and inter-class discriminative characteristics. Table 4's perfectly balanced distribution makes it the appropriate platform for cross-validation procedures, generalization research and benchmarking ensemble algorithms. Eggplant3 is a thorough testbed for assessing model robustness against actual agricultural heterogeneity due to its consistent 224 × 224 pixel resolution and retention of field-acquired metadata.

**Table 4**
Distribution of images across the seven categories in the Eggplant3 dataset.

| Class Name | Eggplant3 |
|---|---|
| | **Number of original images** |
| Healthy | 200 |
| Insect Pest | 200 |
| Leaf Spot | 200 |
| Mosaic Virus | 200 |
| Small Leaf | 200 |
| White Mold | 200 |
| Wilt | 200 |
| Total | 1,400 |

### 2.2. Proposed framework

The proposed framework follows a multi-stage pipeline that systematically combines data preprocessing, deep feature extraction with multiple backbones, weighted ensemble integration, TTA and detailed performance analysis. For each of the three benchmark datasets—Eggplant1, Eggplant2 and Eggplant3—the same workflow is executed under a five-fold cross-validation protocol (5-FCVP) to derive a more generalized and objective performance. The overall structure of the methodology and the flow of information between its components are summarized schematically in Fig. 4 which outlines the main steps from raw image collections to final metric reporting.

In the first stage, all images are organized into class-specific folders according to their annotated disease or condition labels. Afterwards, a stratified 5-FCVP technique is carried out, ensuring that each fold preserves the initial class distribution as closely as possible. For a given fold, four subsets are used for training and one subset is reserved for validation, and this process is repeated so that each subset serves as the validation split exactly once. During training, a rich data augmentation pipeline is employed to improve generalization. After resizing each image to 256 × 256 pixels, a random 224 × 224 crop is applied. To replicate realistic variations in perspective, leaf or fruit orientation and illumination, random horizontal and vertical flipping, mild random rotations and controlled color jittering are used. Finally, each input image underwent normalization based on the ImageNet mean and standard deviation metrics to stabilize optimization. Validation images, in contrast, are only resized to 224 × 224 and normalized, thereby keeping the evaluation process deterministic and comparable across folds.

In the second stage, three state-of-the-art convolutional architectures—ConvNeXt-Tiny, DenseNet-121, and EfficientNet-B0—are used as deep feature extractors and classifiers. For each backbone, the original classification head is removed and replaced with a task-specific linear layer whose output dimension matches the number of classes in the corresponding dataset. The convolutional layers are initialized with pretrained weights, while the new classifier layer is trained from scratch. Each backbone is then fine-tuned independently on the augmented training data of the current fold using the AdamW optimizer with a fixed learning rate, weight decay, and a cosine annealing learning rate schedule. In the third stage, the three fine-tuned backbones are merged into a tri-architectural ensemble model. The ensemble is implemented as a learnable weighted combination of the individual model outputs. Specifically, each backbone provides a probability distribution over

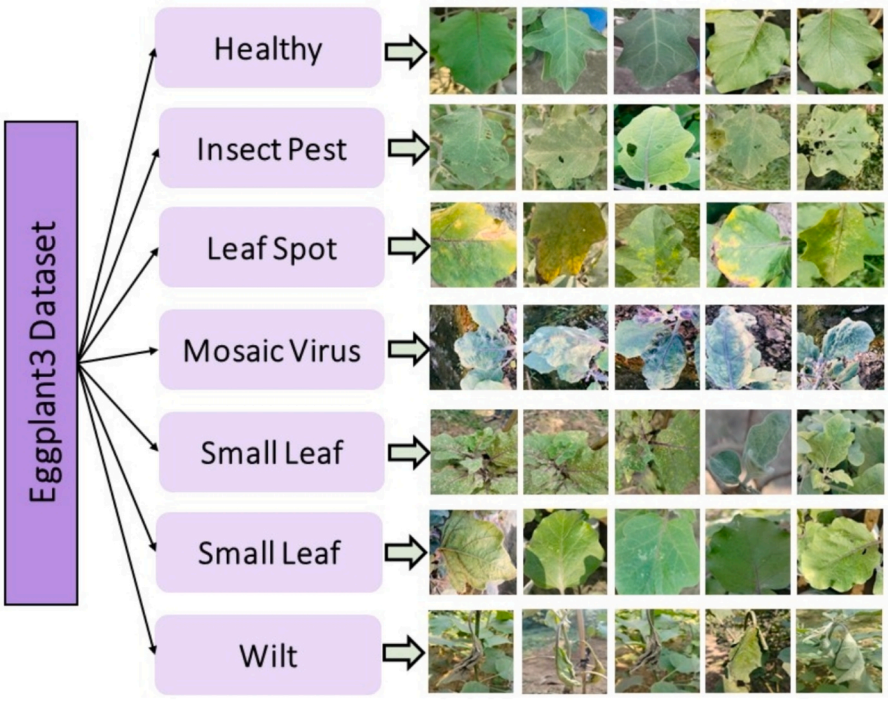


**Fig. 3.** Folder organization and representative images for the Eggplant3 dataset [28].

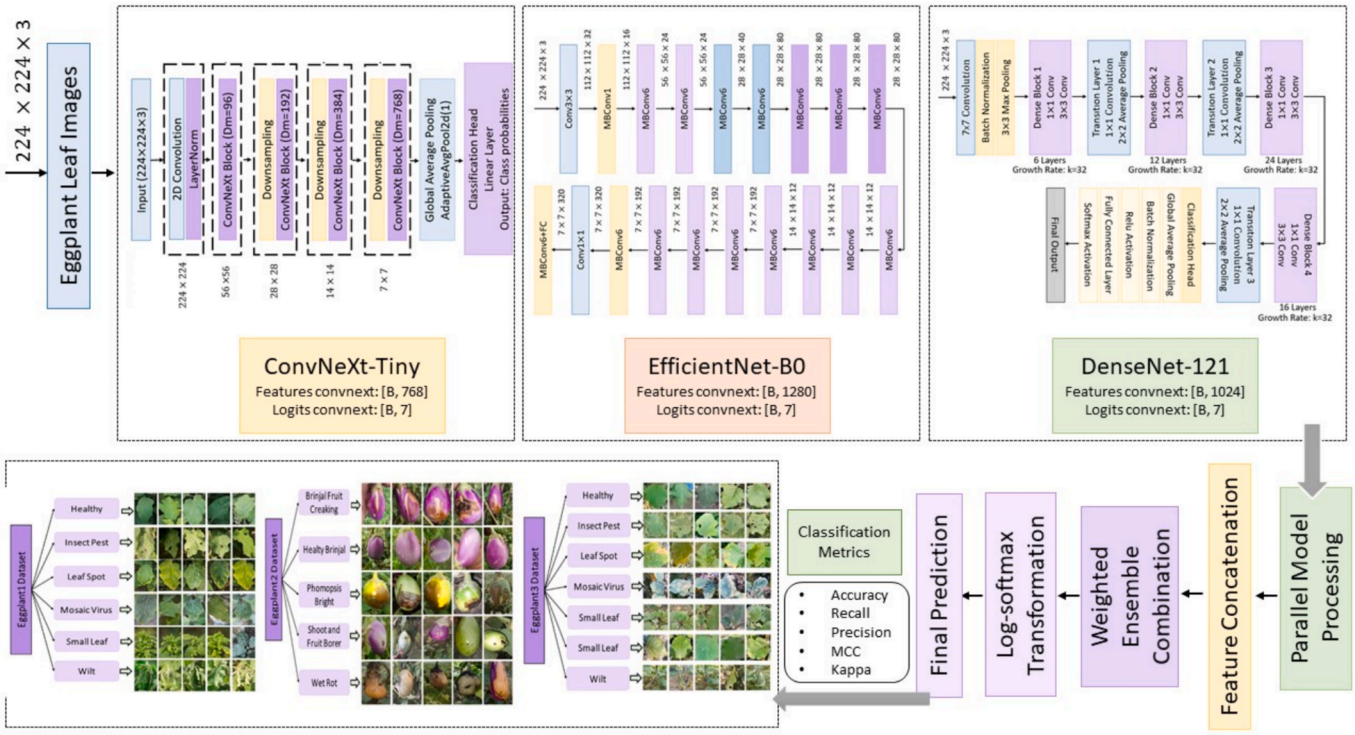


**Fig. 4.** The comprehensive workflow of the proposed methodology.

each class via softmax, and these distributions are combined through a weighted sum, where the weights are learnable parameters normalized by a softmax function to ensure that they form a convex combination. After initializing the ensemble with equal weights, a short fine-tuning phase is conducted in which the ensemble parameters are optimized using the training data within each fold.

To further increase robustness at inference, the ensemble is augmented with a TTA mechanism. For each batch in the validation set, multiple transformed versions of the same input images are generated, including the original view, horizontally flipped images, vertically flipped images and rotations by 90° and 180°. Each variant is passed through the ensemble model and the resulting probability vectors are averaged to obtain the final prediction. In the final stage, predictions obtained from all folds are aggregated and used to compute a comprehensive set of evaluation metrics. For each dataset, ACC, PRC, recall (RCL) and F1-score are calculated to summarize global performance and class-balanced behavior. In addition, the Matthews correlation coefficient (MCC) and Cohen's kappa coefficient are calculated to capture the agreement between predicted and true labels, taking into account both correct and incorrect classifications across all classes.

### 2.3. Deep learning backbones

#### 2.3.1. ConvNeXt

ConvNeXt is a modern CNN that reimagines the classic ResNet architecture by incorporating design principles from Vision Transformers, achieving a balance between efficiency and accuracy in visual recognition tasks [30]. Each ConvNeXt block consists of a depthwise convolution with a large kernel size (7 × 7), followed by Layer Normalization (LN), two pointwise (1 × 1) convolutions, and a GELU activation. Replacing Batch Normalization (BN) with LN, normalizes features channel-wise rather than across the batch. Equation (1) provides LN for an input tensor $x \in \mathbb{R}^{HxWxC}$, where H, W and C stand for height, width and number of channels, respectively [31].

$$LN = \frac{x - \mu}{\sqrt{\sigma^2 - \epsilon}}\gamma + \beta \tag{1}$$

where $\mu$ and σ denote the mean and standard deviation of $x$, $\gamma$ and $\beta$ are learnable affine parameters, and ∈ is a stability constant. This replacement improves training stability and generalization. In this study, the pretrained "ConvNeXt-Tiny" variant was employed. In the model the default classification head was removed and replaced with an adaptive average pooling layer, yielding a feature vector of dimension 768. This feature vector is then used as an input to a linear classifier. Fig. 5 illustrates the internal architecture of a ConvNeXt-Tiny block, highlighting its key components including the large-kernel depthwise convolution, LN, and GELU activation.

#### 2.3.2. DenseNet-121

Dense Convolutional Network (DenseNet) was introduced by Huang et al. to improve feature reuse and alleviate the vanishing gradient problem in very deep networks [32]. Unlike traditional feed-forward CNNs, where each layer connects only to its immediate successor, DenseNet employs dense connectivity meaning that every layer receives as input the concatenation of all feature maps from preceding layers. Formally, the output of the $l^{th}$ layer is expressed as Equation (2) [33].

$$x_l = H_l([x_0, x_1, \cdots x_{l-1}]) \tag{2}$$

where $[x_0, x_1, \cdots x_{l-1}]$ denotes the concatenation of the feature maps produced by layers 0 to $l$-1 and $H_l(.)$ is a composite non-linear transformation. Each transformation typically follows the sequence BN → ReLU → 1 × 1 Convolution → BN → ReLU → 3 × 3 Convolution. DenseNet-121 consists of four dense blocks interleaved with transition layers. Each dense block comprises multiple densely connected layers (6, 12, 24, and 16 layers respectively for DenseNet-121), and each layer contributes $k = 32$ new feature maps, where $k$ is known as the growth rate. If the input to a dense block is $F_0$ feature maps, then after $L$ layers, the output dimension becomes by Equation (3) [34]. The default classifier of the pretrained DenseNet-121 was removed for the

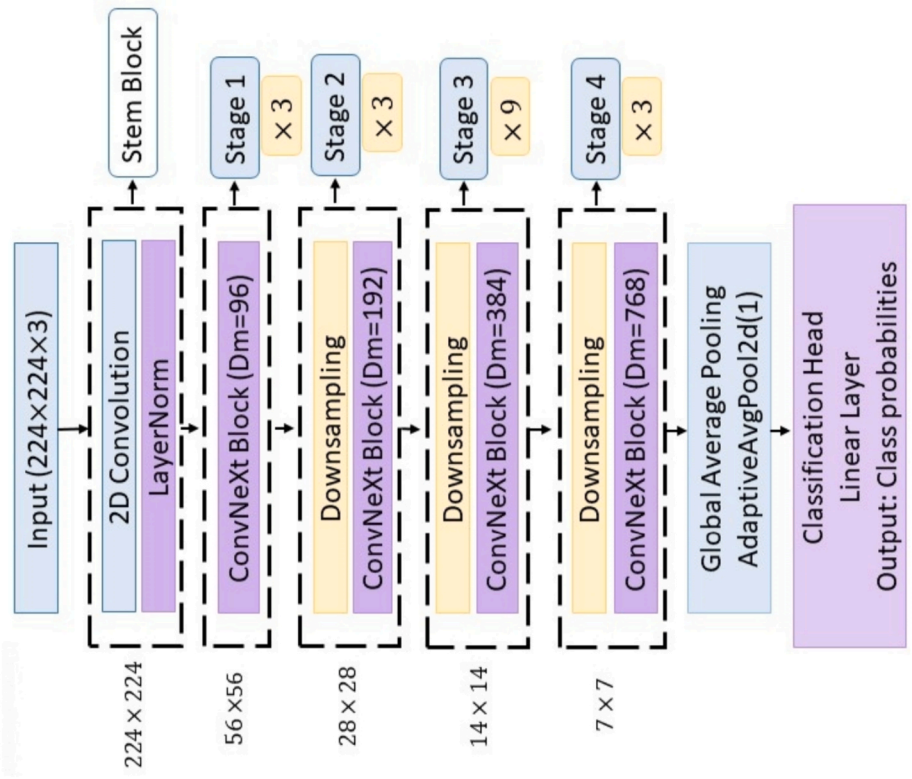


**Fig. 5.** Internal structure of the ConvNeXt-Tiny block.

implementation of this study leaving a 1024-dimensional feature vector after global average pooling. A linear classifier maps this vector into the labels. Fig. 6 visualizes the dense connectivity pattern in DenseNet-121, demonstrating how each layer receives concatenated feature maps from all preceding layers, facilitating maximal feature reuse and gradient flow throughout the network.

$$F_L = F_0 + k.L \tag{3}$$

### 2.3.3. *EfficientNet-B0*

EfficientNet is a family of CNNs which achieves state-of-the-art performance with fewer parameters by equally balancing network depth, width, and resolution through the use of a compound scaling

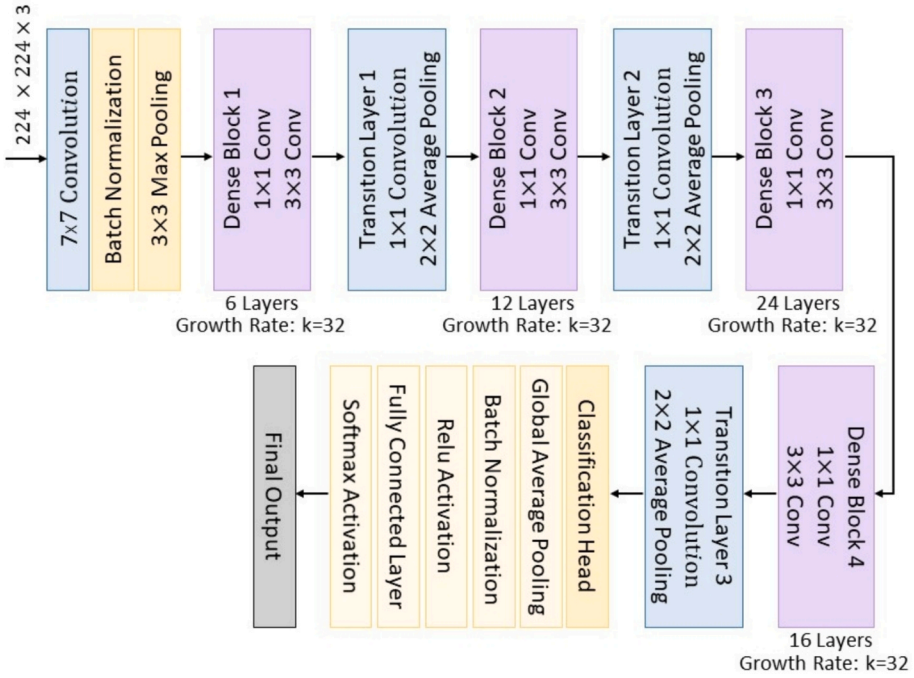


**Fig. 6.** Illustration of the dense connectivity pattern in DenseNet-121.

method [35]. Its backbone is based on Mobile Inverted Bottleneck Convolutions (MBConv), which combine depthwise convolutions, expansion-projection layers, and Swish (SiLU) activation. The general MBConv transformation is given as Equation (4) [36].

$$f(x) = W_p.\sigma(\mathrm{DWConv}(W_e.x)) \tag{4}$$

where $W_e$ and $W_p$ denote expansion and projection layers, respectively, σ is the SiLU activation and DWConv is depthwise convolution. EfficientNet-B0 serves as the baseline model of this family. It starts with a standard convolutional layer followed by stacked MBConv blocks with varying kernel sizes ($3 \times 3$ and $5 \times 5$) and expansion ratios. The number of repeated blocks ($s$) differs across stages which allows the model to increase representational capacity without excessive growth in parameters. At the end of the network, a $1 \times 1$ convolution is applied, followed by global average pooling and a fully connected layer. The key innovation of EfficientNet is the compound scaling rule which simultaneously scales depth ($d$), width ($w$), and resolution ($r$) according to Equation (5).

$$d = \alpha^{\phi} \mathrm{w} = \beta^{\phi} \cdot \mathrm{r} = \gamma^{\phi} \tag{5}$$

subject to the constraint to Equation (6).

$$\alpha.\beta^2.\gamma^2 \approx 2\alpha, \beta, \gamma \geq 1 \tag{6}$$

where ϕ is a user-defined scaling coefficient. EfficientNet-B0 corresponds to the case where $\phi = 1$, and larger models (B1–B7) are obtained by increasing $\phi$. In this study, the default classifier was removed, and a 1280-dimensional global average pooled feature vector was retained. This vector was then projected into class logits via a linear layer. Fig. 7 depicts the MBConv block structure used in EfficientNet-B0, illustrating the inverted bottleneck design with depthwise convolution, expansion-projection layers, and the integration of squeeze-and-excitation modules for channel-wise attention.

### 2.4. Weighted ensemble learning framework for multi-architectural integration

The ensemble component of the proposed system is designed to integrate the predictive distributions generated by three complementary deep architectures—ConvNeXt, DenseNet-121and EfficientNet-B0—into a unified probabilistic output. Let $p^{(m)} \in R^K$ denote the softmax probability vector produced by model $m$, where $m \in$ {ConvNeXt, DenseNet-121, EfficientNet-B0} and $K$ represents the total number of classes. Instead of employing fixed, manually assigned weights, the ensemble utilizes trainable fusion parameters $\{w_m\}$. Learnable weight optimization within ensemble architectures reflects a broader trend in machine learning toward integrating analytical modeling principles with data-driven optimization strategies [37]. These parameters are projected onto the probability simplex through a softmax transformation to enforce non-negativity and unit-sum constraints. Accordingly, the normalized ensemble coefficients are defined as Equation (7), [36]. The final ensemble distribution $p^{(ens)}$ is formulated as a convex combination of individual model outputs as Equation (8). Fig. 8 illustrates the architecture of the learnable weighted ensemble module showcasing the parallel processing of three backbone networks, the softmax based weight normalization and the weighted fusion mechanism that generates the final ensemble prediction.

$$\alpha_m = \frac{e^{w_m}}{\sum_j e^{w_j}}, \sum_m \alpha_m = 1 \tag{7}$$

$$p^{(ens)} = \sum_{m=1}^{3} \alpha_m p^{(m)} \tag{8}$$

Ensemble stability was enforced through softmax-normalized fusion weights ensuring bounded and interpretable model contributions during training. By jointly optimizing ensemble weights across cross-validation folds and aggregating predictions via TTA, the framework reduces sensitivity to fold-specific fluctuations and suppresses overfitting to individual backbone biases. This design encourages consistent decision

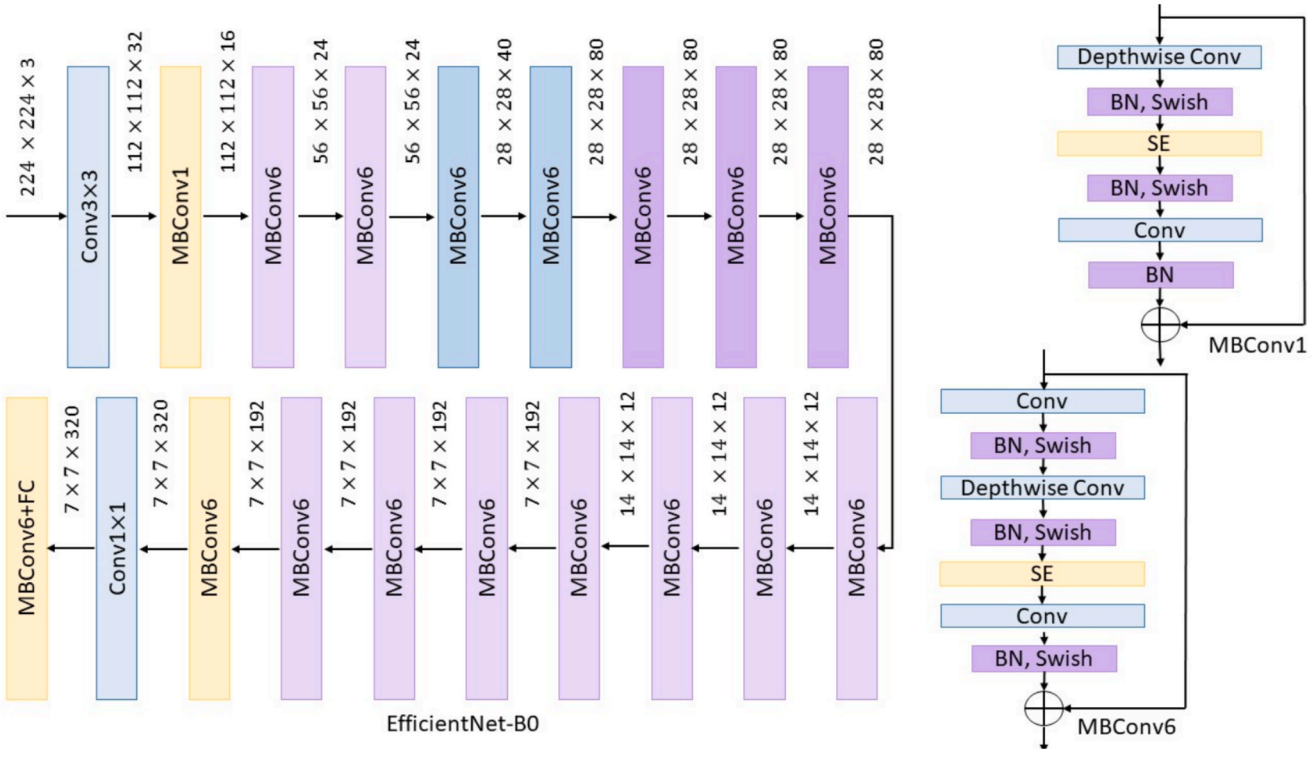


**Fig. 7.** The MBConv structure of EfficientNet-B0.

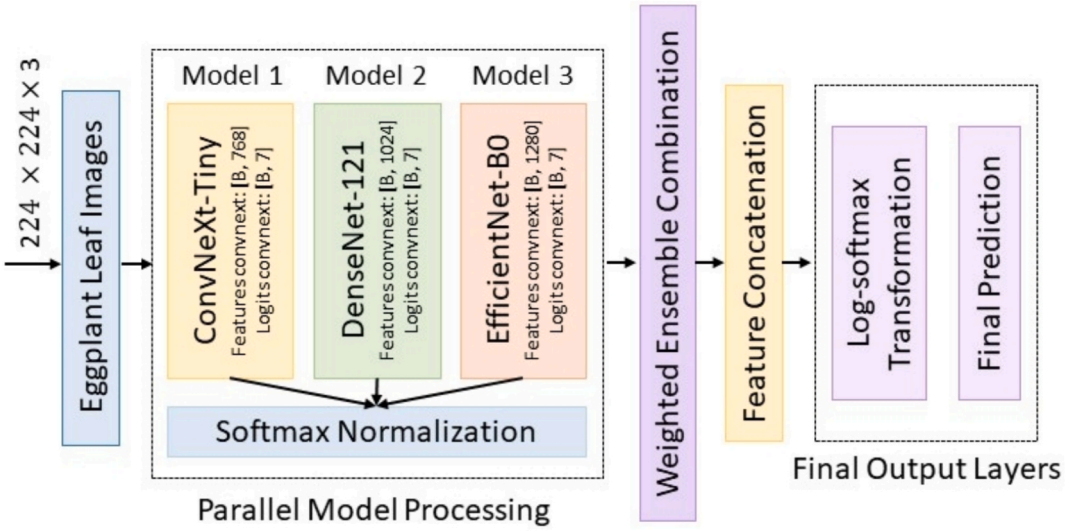


**Fig. 8.** Architecture of the learnable weighted ensemble module.

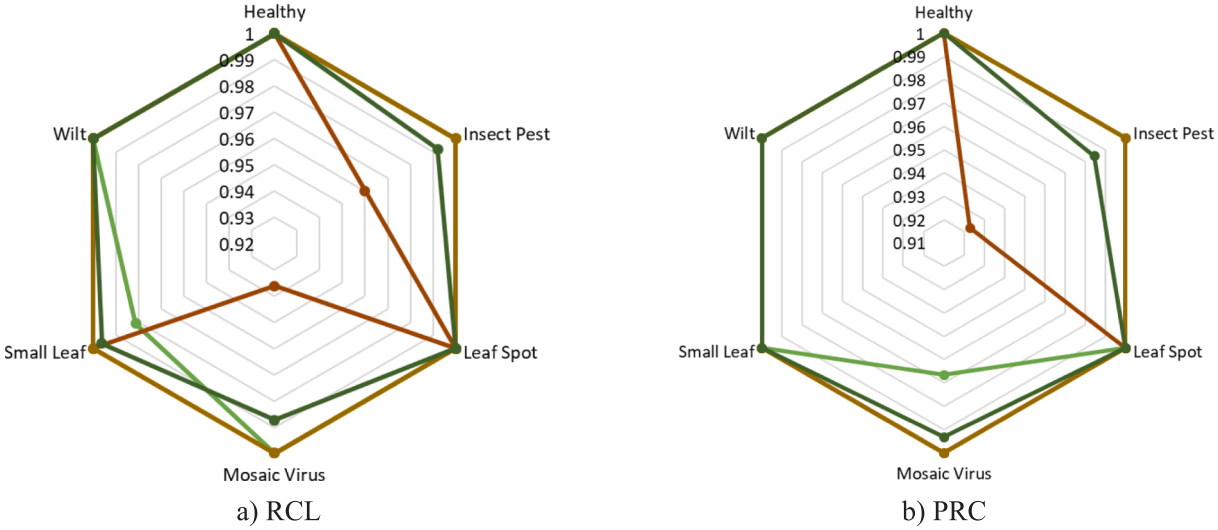


**Fig. 9.** Class-wise RCL (a) and PRC (b) radar visualizations for the Eggplant1 dataset across 5-FCVP.

boundaries rather than reliance on a single dominant model.

To increase predictive stability under spatial transformations and mitigate sensitivity to orientation-induced perturbations, a structured TTA procedure was incorporated [38]. Each test image $x$ was deterministically transformed into the following set: $\{x^{(1)},x^{(2)},x^{(3)},x^{(4)},x^{(5)}\}$= {original, horizontal flip, vertical flip, rotation 90$^o$, rotation 180$^o$}. Let $p^{(i)}$ represent the predicted class-probability distribution associated with the i-th augmented view. The refined prediction obtained from TTA is defined as Equation (9), [39]. Averaging across multiple spatially perturbed representations effectively smooths out view specific fluctuations and enhances the orientation invariance of the inference process. Consequently, the TTA mechanism functions as a post hoc regularization layer improving decision robustness without altering the learned model parameters. Beyond accuracy gains, TTA acts as an implicit regularization mechanism at inference time by reducing prediction variance induced by orientation and spatial perturbations.

$$p^{(TTA)} = \frac{1}{5}\sum_{i=1}^{5} p^{(i)} \tag{9}$$

### 2.5. Training, data augmentation, and evaluation protocol

During the training phase, a structured data preprocessing and augmentation pipeline was applied to improve model robustness and reduce overfitting. All images were first resized to 224 × 224 pixels to ensure a consistent spatial resolution across the dataset. To introduce geometric variability, a Random Affine transformation was applied, parameterized by a random rotation sampled from the interval [−15$^o$, +15$^o$], scaling in the range [0.9,1.1], translations of up to 10% along both spatial axes, and shear distortions up to ±10°. This transformation can be expressed as an Equation (10) where $A$ encapsulates rotation, scale and shear components, and $b$ denotes the translation vector [29].

$$T(x) = A(x) + b \tag{10}$$

To further enhance robustness against noise and local perturbations, a Gaussian blur filter with a kernel size of 3 × 3 was applied, followed by Random Erasing with a probability of 0.2, which randomly masks a rectangular region in the input image. This procedure encourages the network to rely on global and discriminative contextual features rather than localized patterns. Finally, all images were normalized using the standard ImageNet statistics as introduced in Equation (11) where$\mu$ =

[0.485, 0.456, 0.406], $\sigma$=[0.229, 0.224, 0.225]. This normalization ensures stable gradient flow and improves convergence during training. The validation dataset underwent only resizing to 224 × 224 pixels followed by the same ImageNet normalization parameters.

$$x_{norm} = \frac{x - \mu}{\sigma} \tag{11}$$

A 5-FCVP was used in the experimental setup to guarantee an objective and statistically valid performance evaluation. The original class distribution was maintained while the dataset was divided into separate training and validation subsets in each fold. Each backbone model was fine-tuned individually for 5 epochs utilizing the AdamW optimizer with a $10^{-4}$ initial learning rate, weight decay $10^{-4}$ and a Cosine Annealing learning rate scheduler to gradually adapt the step size. The cross-entropy loss function was employed as the optimization objective. After single-model fine-tuning, the ensemble model was further refined for 3 epochs with a reduced learning rate of $10^{-5}$, during which the learnable ensemble weights were updated. Preliminary experiments indicated that performance gains saturated early, while prolonged training led to increased variance across folds. Therefore, a short fine-tuning strategy was adopted to prioritize generalization stability rather than aggressive convergence. To verify the robustness of the proposed ensemble architecture, a variety of performance metrics were considered. ACC was first calculated to represent the fraction of correctly classified samples among all predictions as Equation (12), [40].

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

While accuracy provides a global overview, it can be misleading in imbalanced settings. Therefore, PRC and RCL were also examined. Precision, as defined in Equation (13), measures the model's ability to avoid false positives by calculating the ratio of true positive detections relative to all instances flagged as positive [40].

$$PRC = \frac{TP}{TP + FP} \tag{13}$$

where as RCL, often referred to as sensitivity, measures the proportion of actual positives that were successfully retrieved given in Equation (14), [40].

$$RCL = \frac{TP}{TP + FN} \tag{14}$$

To balance these complementary measures, the F1 score was adopted, defined as the harmonic mean of PRC and RCL given in Equation (15). In this study, the macro-F1 score was reported by averaging F1 values across all classes [41].

$$F1 = \frac{2.PRC.RCL}{PRC + RCL} \tag{15}$$

To further account for both correct and incorrect predictions, the MCC was included, as it incorporates all four elements of the confusion matrix and yields a correlation coefficient between predictions and ground truth as given in Equation (16), [30].

$$MCC = \frac{TP.TN - FP.FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{16}$$

In addition, Cohen's kappa coefficient (κ) was computed to quantify the agreement between predicted and true labels while correcting for the level of agreement expected by chance is given in Equation (17) where $p_o$ is the observed agreement (equivalent to accuracy) and $p_e$ is the expected agreement by random chance [30].

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{17}$$

## 3. Results and discussion

The experimental results obtained from the proposed learnable weighted ensemble framework provide strong evidence for its robustness, generalizability and discriminative capacity across eggplant leaf disease datasets. A comprehensive performance examination was conducted using the Eggplant1, Eggplant2 and Eggplant3 benchmarks, each representing distinct levels of class complexity, distribution imbalance and inter class visual similarity. The results consistently demonstrate that the ensemble architecture substantially enhances predictive ACC, stability and reliability when compared to the performance of individual constituent models. A detailed quantitative summary of the ensemble's performance on each dataset is provided in Table 5 where the superiority of the fusion approach is clearly observable. In Eggplant1, which contains six well-defined classes with relatively balanced distributions, the ensemble achieved an exceptional ACC of 99.70%, surpassing the strongest base learner, ConvNeXt-Tiny and establishing a new performance ceiling for this dataset. As shown in Table 5, the ensemble not only increased ACC but also yielded higher F1-score, PRC, RCL and MCC. The improvement in MCC, rising from 0.9937 for the best individual model to 0.9964 for the ensemble, reflects a notable increase in class separability and a reduction in misclassification uncertainty, even among visually similar disease categories. Although very high ACC values are observed for Eggplant1, model behavior was further examined at the class level to assess potential overfitting and memorization effects. Accordingly, class-wise RCL, PRC, MCC and Cohen's kappa were systematically reported for all disease categories across all folds enabling a detailed characterization of misclassification patterns beyond aggregate metrics. The consistently high RCL and agreement scores observed for minority and visually similar classes indicate that the achieved performance is not dominated by majority-class bias or data leakage, but rather reflects stable and discriminative representation learning under strict fold-wise data partitioning.

Performance trends on the more challenging Eggplant2 dataset, which is characterized by apparent class imbalance and substantial intra-class variability, further emphasize the effectiveness of the ensemble strategy. The ensemble outperformed ConvNeXt-Tiny by 0.83 percentage points and DenseNet-121 by 3.24 percentage points, as shown in Table 5, with an impressive ACC of 93.75%. A significant rise in the overall F1-score and RCL further supports this improvement. Class-wise analyses presented in Figs. 10 and 13 additionally indicate improved behavior for underrepresented classes. Additionally, the ensemble achieved a Cohen's kappa score of 0.9138 and an MCC value of 0.9142, metrics that reflect strong agreement between predicted and true labels under uneven data distributions. These gains highlight the ensemble's ability to minimize the representational bias that individual CNNs tend to develop when exposed to minority-dominant class structures. The ensemble's ability to sustain high performance in scenarios demanding both fine-grained discrimination and large feature generalization is confirmed by the findings from the Eggplant3 dataset, which comprises seven classes with very uniform distributions. With an ACC of 95.57%, the ensemble outperformed all individual models and produced nearly equal values for PRC (0.9564) and RCL (0.9561), as shown in Table 5. This balance implies that the classifier neither overpredicts nor underpredicts certain categories maintaining consistent discriminative confidence across classes. The ensemble's reliability in multi-class classification settings where subtle illness patterns need to be distinguished is further validated by its kappa score of 0.9482 on this dataset.

These numerical results are directly supported by the architectural and methodological design illustrated in Fig. 4, which outlines the full inference pipeline. The ensemble leverages the complementary strengths of its components: ConvNeXt contributes large-kernel depthwise convolutions that emphasize long-range spatial patterns; DenseNet enhances gradient flow and texture preservation through dense feature connectivity; EfficientNet introduces balanced multi-scale representation via compound scaling. The learnable weighting mechanism ensures

**Table 5**
Cross-architectural performance evaluation on Eggplant1, Eggplant2 and Eggplant3.

| | Model | ACC | F1-score | PRC | RCL | MCC | Kappa |
|---|---|---|---|---|---|---|---|
| Eggplant1 | ConvNeXt-Tiny | 0.9948 | 0.9943 | 0.9943 | 0.9945 | 0.9937 | 0.9937 |
| | DenseNet-121 | 0.9761 | 0.9744 | 0.9748 | 0.9760 | 0.9715 | 0.9712 |
| | EfficientNet-B0 | 0.9850 | 0.9844 | 0.9844 | 0.9849 | 0.9820 | 0.9819 |
| | Proposed Ensemble | 0.9970 | 0.9967 | 0.9972 | 0.9963 | 0.9964 | 0.9964 |
| Eggplant2 | ConvNeXt-Tiny | 0.9292 | 0.8969 | 0.8975 | 0.8994 | 0.9030 | 0.9026 |
| | DenseNet-121 | 0.9051 | 0.8600 | 0.8552 | 0.8699 | 0.8697 | 0.8689 |
| | EfficientNet-B0 | 0.9057 | 0.8629 | 0.8656 | 0.8703 | 0.8714 | 0.8704 |
| | Proposed Ensemble | 0.9375 | 0.9081 | 0.9072 | 0.9114 | 0.9142 | 0.9138 |
| Eggplant3 | ConvNeXt-Tiny | 0.9500 | 0.9498 | 0.9504 | 0.9512 | 0.9419 | 0.9415 |
| | DenseNet-121 | 0.9129 | 0.9126 | 0.9130 | 0.9157 | 0.8987 | 0.8981 |
| | EfficientNet-B0 | 0.9200 | 0.9199 | 0.9213 | 0.9219 | 0.9070 | 0.9065 |
| | Proposed Ensemble | 0.9557 | 0.9555 | 0.9564 | 0.9561 | 0.9485 | 0.9482 |

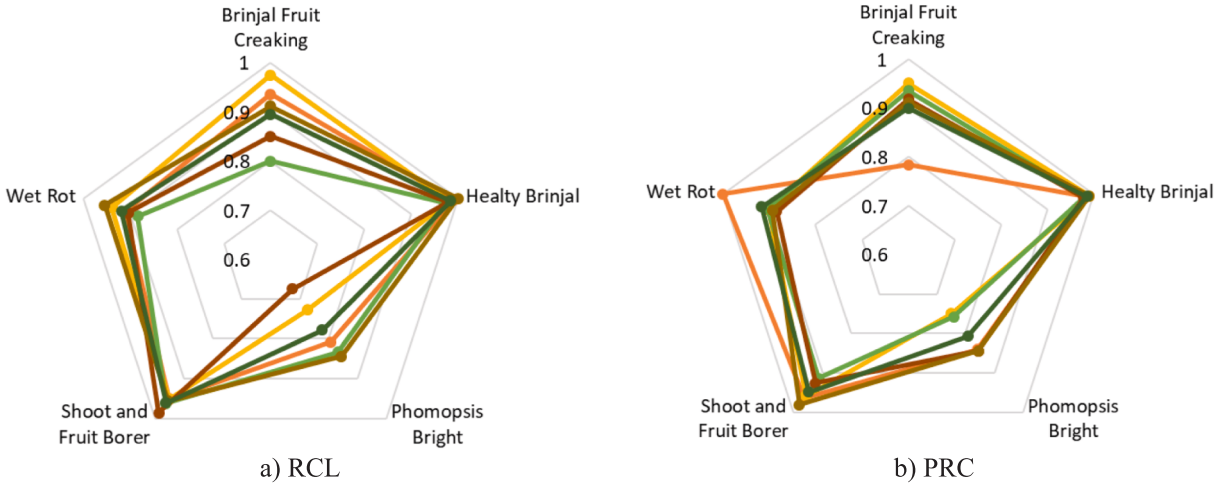


**Fig. 10.** Fold-specific radar charts depicting per-class RCL (a) and PRC (b) on the imbalanced Eggplant2.

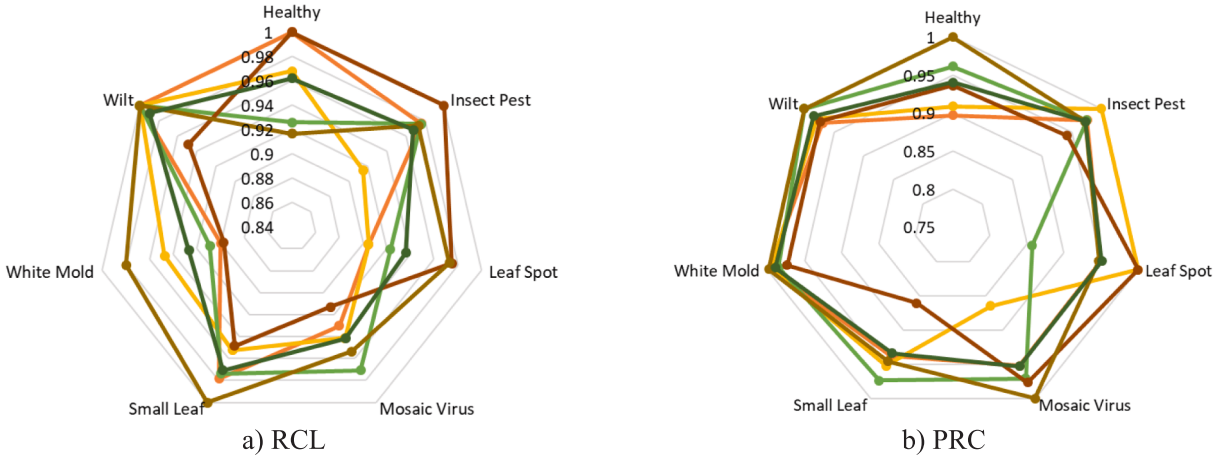


**Fig. 11.** Radar plots of RCL (a) and PRC (b) for the Eggplant3.

that each network's contribution is optimized dynamically enabling the system to integrate global, local and multi-scale cues in a cohesive manner. Through this mechanism, traits that are specifically informative for particular disease classes are amplified while redundancy between models is efficiently reduced.

In addition to aggregate performance indicators, the reliability and stability of the proposed framework were further investigated through class-wise and fold-wise visualizations presented in Figs. 9–14. These analyses summarize RCL, PRC, MCC and Cohen's kappa distributions across individual classes enabling an explicit evaluation of error tendencies and minority-class behavior. Collectively, the visualizations indicate that the ensemble exhibits highly stable performance across the 5-FCVP, with only limited variation between folds. The fold-specific color coding—orange (Fold 1), yellow (Fold 2), light green (Fold 3), brown (Fold 4), army green (Fold 5), and dark green (macro average)—facilitates visual comparison and highlights the consistency of the proposed approach. This stability, as illustrated in Figs. 9–14, indicates strong generalization and low sensitivity to partitioning differences, attributes expected from a model intended for real-world deployment where data shifts frequently occur. Particularly in Eggplant2, where

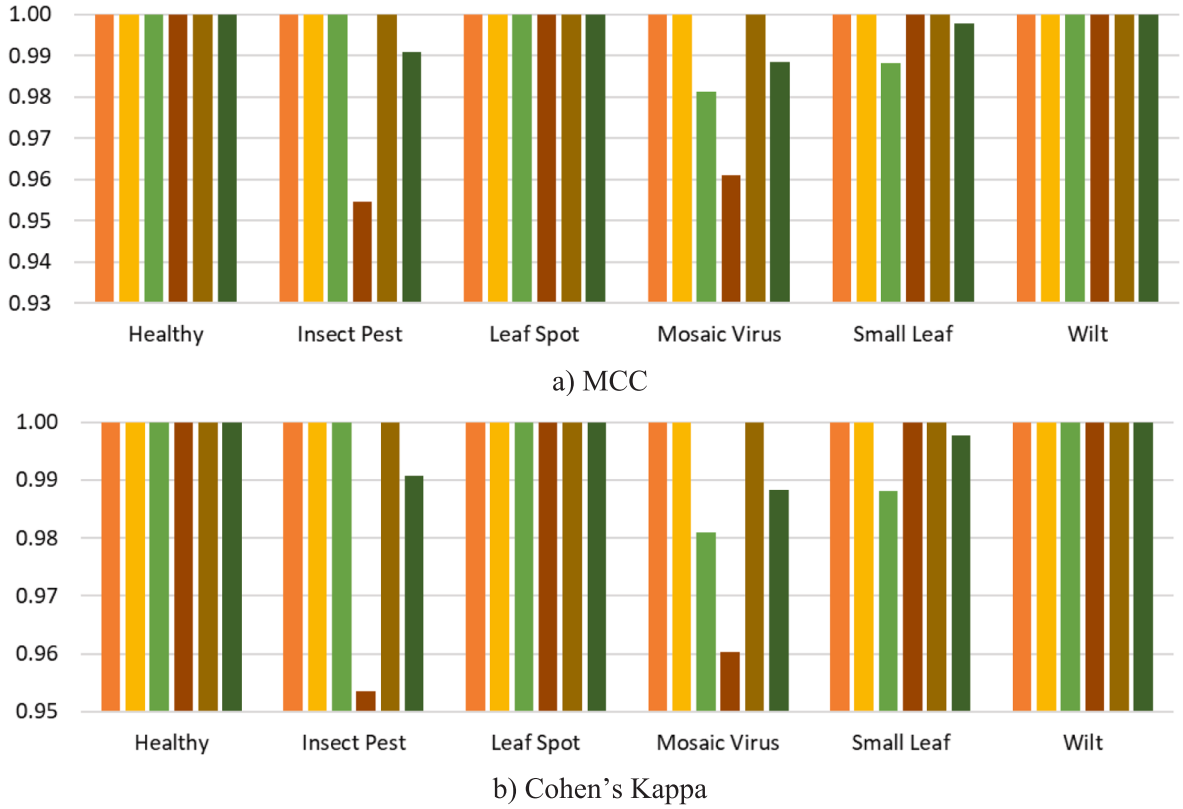


**Fig. 12.** Per-class MCC (a) and Cohen's Kappa (b) for Eggplant1.

individual models exhibit noticeable performance fluctuations, the ensemble significantly reduces inter-fold variability, especially in RCL for underrepresented disease classes.

In addition to quantitative stability analyses, representative misclassified samples are provided in Fig. 15 to qualitatively illustrate the failure modes of the proposed framework. Most misclassifications occur between visually similar disease categories or under challenging conditions such as partial occlusion, non-uniform illumination, background clutter, and early-stage symptom manifestation, highlighting the intrinsic difficulty of fine-grained disease discrimination in real agricultural environments.

Although no explicitly staged stress-test experiments were conducted, real-world challenges such as varying illumination, background clutter, partial occlusion, and leaf overlapping are inherently present in the field-acquired datasets used in this study, particularly Eggplant2 and Eggplant3. The robustness of the proposed framework against these factors is reflected through consistent class-wise RCL and agreement metrics across folds, as well as through the variance-reducing effect of TTA (Table 6), which mitigates sensitivity to illumination and orientation changes.

TTA constitutes a key robustness-enhancing component of the proposed framework. To explicitly isolate its contribution, an ablation-style analysis was conducted by comparing ensemble performance with and without TTA across all three datasets, while single-backbone models serve as implicit baselines for assessing the effect of learnable ensemble weighting. As summarized in Table 6, incorporating TTA consistently improves mean ACC and, in most cases, reduces fold-level performance variance, with the most pronounced gains observed on the imbalanced Eggplant2 dataset, where ACC increased by 0.72% and variance was reduced by more than one-third. Similar, though smaller, improvements are observed on Eggplant1 and Eggplant3 indicating that TTA enhances prediction stability by aggregating evidence from multiple augmented views particularly under illumination changes, orientation variability and natural noise artifacts.

To assess whether the reported performance improvements are statistically consistent, fold-wise performance behavior was analyzed using the 5-FCVP results. Due to the limited sample size inherent to cross-validation ($n = 5$), conventional hypothesis testing yields limited statistical power and may not provide reliable significance estimates. Instead, the analysis focuses on consistency trends and variance behavior across folds. The results indicate that the proposed framework exhibits stable improvements across all folds, particularly on the imbalanced Eggplant2 dataset. For Eggplant1 and Eggplant3, the near-saturated baseline performance introduces a ceiling effect, constraining the magnitude of statistically observable differences despite consistent fold-wise behavior.

To contextualize the proposed approach within existing research, a comparative analysis with contemporary state-of-the-art methods is presented in Table 7. It is crucial to recall that Table 7′s comparison is intended to be contextual rather than a strictly controlled head-to-head benchmark. The referenced studies differ substantially in terms of task formulation (pure image classification versus object detection), number of target classes, and experimental settings. In particular, object detection frameworks such as YOLO-based pipelines optimize localization-aware objectives and are therefore not directly comparable to leaf-level multi-class classification models using ACC as a primary metric. Accordingly, Table 7 is included to position the proposed method within the broader spectrum of intelligent agricultural vision research, while rigorous and fair performance evaluation is ensured through controlled intra-dataset comparisons under identical protocols, as reported in Tables 5. The results demonstrate that the proposed ensemble framework either matches or surpasses previously reported accuracies across single and multi-dataset studies. Compared to Siddiqui et al.'s [9] ensemble, the proposed system achieves competitive performance (99.70% vs. 99.76%) on Eggplant1 while outperforming all reported methods across Eggplant2 and Eggplant3. The cross-dataset evaluation presented here,

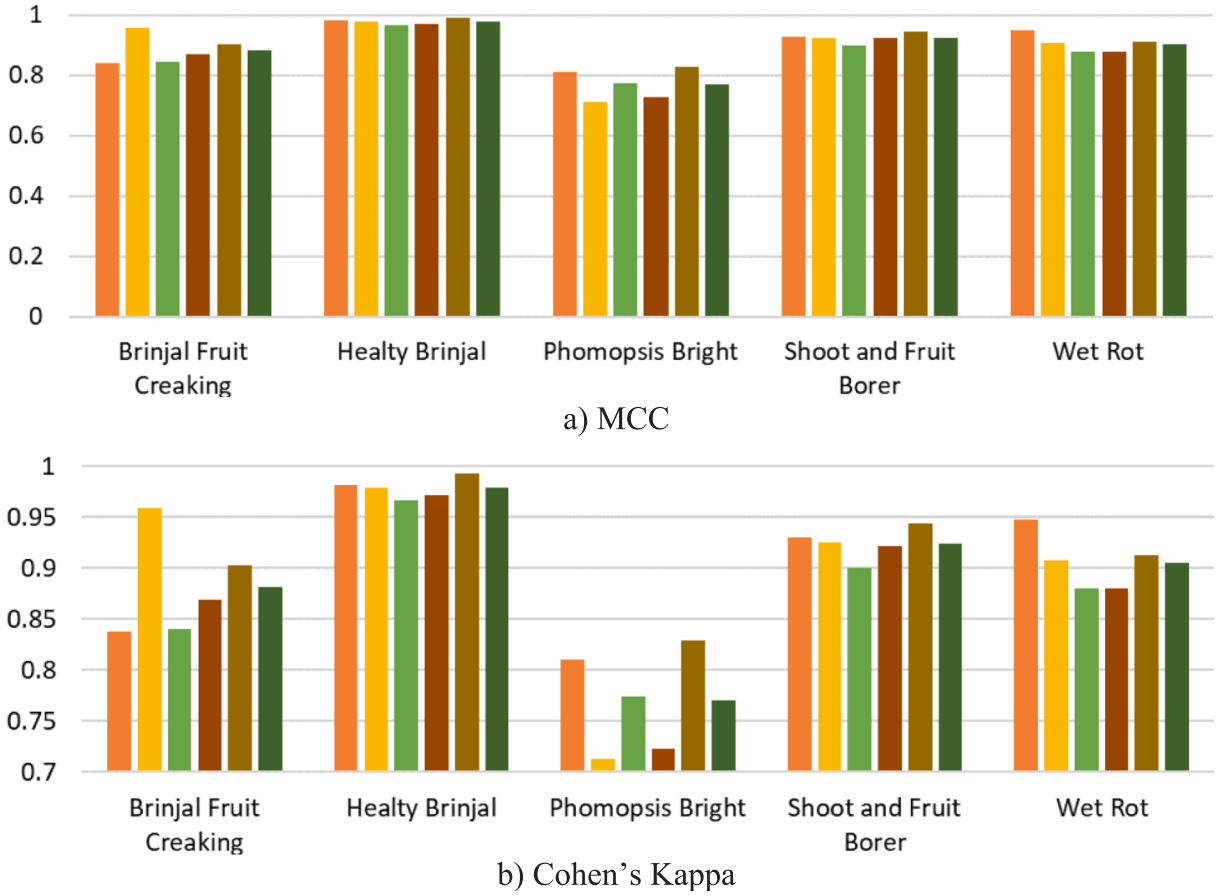


**Fig. 13.** Statistical robustness evaluation on Eggplant2 through bar graphs of MCC (a) and Kappa (b).

which examines three heterogeneous benchmarks, establishes a more stringent and comprehensive testing protocol than the single-dataset focus characteristic of earlier work. The gains observed over Haque et al. (+0.80%) [6] and Saad et al. (+0.64%) [7] highlight improvements derived not only from architectural differences but also from methodological enhancements such as learnable weighting and augmented inference. Additionally, the proposed method demonstrates superior robustness compared to object detection-focused pipelines such as YOLOv5s-BiPCNeXt [13], indicating that targeted classification frameworks can achieve improved discriminative performance under heterogeneous dataset conditions.

Table 8 provides a summary of the system's computational efficiency, including parameters, Giga Floating Point Operations (GFLOPs), training times, inference latency, model throughput, and GPU memory use. With 41.9 million parameters and 7.77 GFLOPs per input, the ensemble exceeds the resource needs of individual models while still being feasible for contemporary GPUs. With an average inference delay of 38.6 ms per image, training takes about 798 s per epoch, resulting in a throughput of 25.9 images per second. These values suggest that the ensemble is computationally feasible for real-time or near-real-time agricultural monitoring systems. The low variance in inference latency, supported by a coefficient of variation below 3.00%, confirms the system's consistency under repeated execution—an essential property for embedded or field-level deployments.

Taken together, the results presented across Tables 5–8 and Figs. 4 and 9–14 demonstrate that the proposed ensemble framework establishes a highly reliable and generalizable solution for eggplant leaf disease classification. The combination of architectural complementarity, learnable fusion, cross-validation rigor and augmentative inference culminates in a model that not only achieves strong comparative performance but also maintains stability across varying dataset conditions making it a strong candidate for real-world agricultural applications.

The findings of this study underscore the effectiveness of multi-architectural ensemble learning in addressing the inherent variability of eggplant leaf disease datasets. The integration of ConvNeXt-Tiny, DenseNet-121 and EfficientNet-B0 brings together complementary inductive biases—hierarchical spatial modeling, dense feature reuse and compound-scaled representational efficiency—which together improves the model's capacity to detect tiny visual cues unique to a disease. Learnable weight fusion systematically aligns the strengths of heterogeneous CNNs into a unified decision space that maintains accuracy even under class imbalance and inter-dataset diversity as evidenced by the consistent improvements seen across all metrics and datasets (Tables 5).

TTA further contributes to this robustness by mitigating distributional shifts at inference time. As detailed in Table 6, TTA elevates mean ACC and can reduce performance variance, particularly on the imbalanced Eggplant2 dataset. This indicates that augmentative inference is especially beneficial when class priors are skewed or when disease symptoms manifest with substantial visual heterogeneity. Nevertheless, TTA introduces additional computational overhead, and its effectiveness varies depending on the dataset's intrinsic complexity—highlighting both its advantages and its operational limitations. When benchmarked against prior research (Table 7) the proposed method exhibits competitive or superior performance across nearly all comparative settings. While several earlier studies report high ACC on isolated datasets, few demonstrate strong cross-dataset generalization, a capability that emerges clearly in our results. The inclusion of three datasets with distinct class structures provides a broader and more realistic evaluation landscape positioning our model closer to real world agricultural deployment scenarios.

From an application standpoint, the reported ACC, stability and

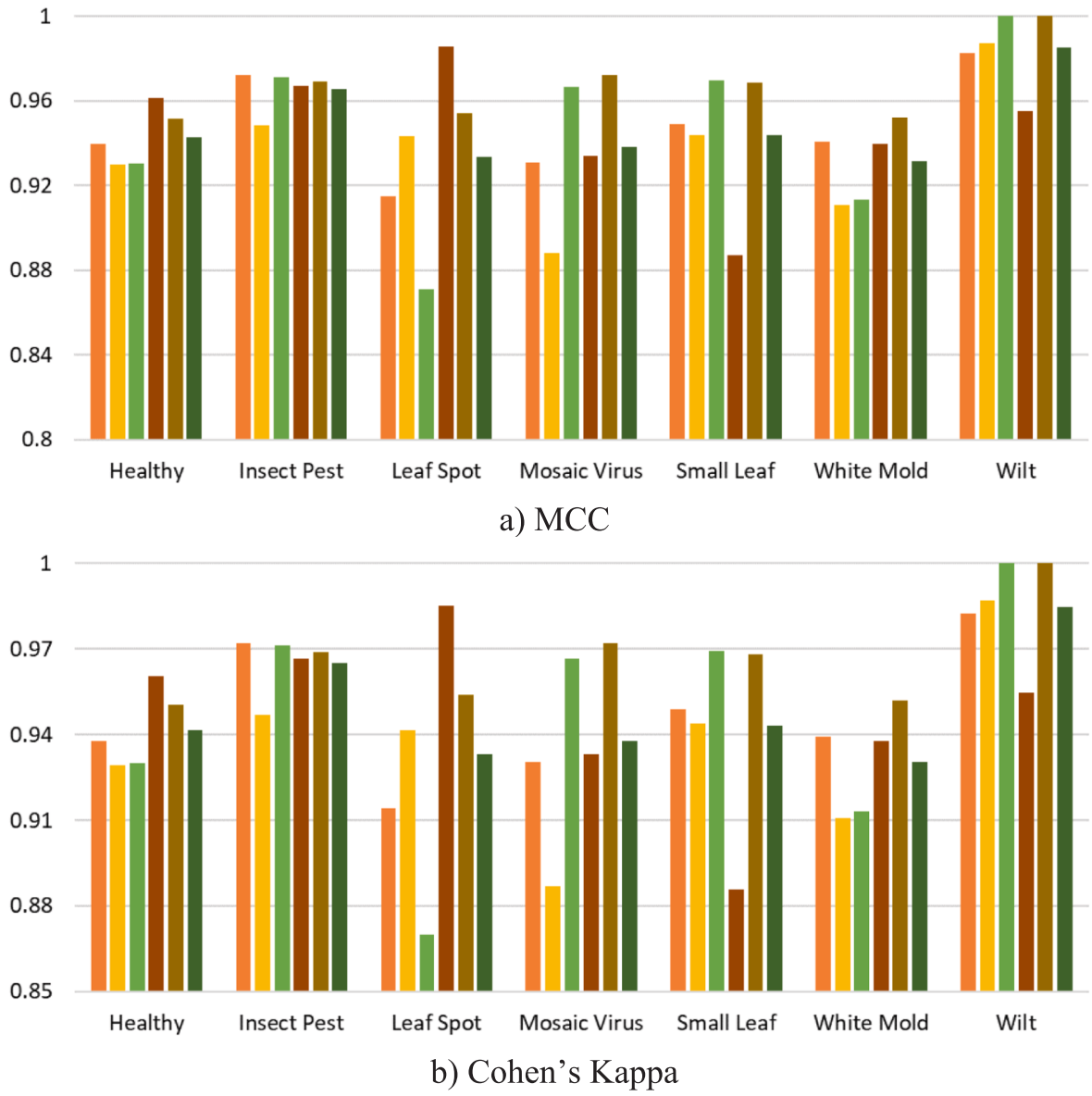


**Fig. 14.** Comparison of fold-wise MCC (a) and Cohen's Kappa (b) for Eggplant3.

computational feasibility collectively support the adoption of this framework in precision agriculture. The inference speed of 25.9 images per second (Table 8) surpasses typical sensor rates used in greenhouse and open-field monitoring, enabling real-time or near-real-time disease detection. However, future deployment should consider environmental factors such as illumination variability, occlusion and camera noise which were not explicitly modeled in this study. While a full factorial ablation of all architectural components was beyond the scope of this study, the presented analyses provide clear empirical evidence for the individual contributions of learnable ensemble weighting and TTA, with more exhaustive ablation reserved for future work.

Qualitative inspection of misclassified samples indicates that residual errors predominantly arise from high inter-class visual similarity and ambiguous early-stage symptoms, suggesting that future work may benefit from incorporating lesion-level localization or multimodal cues to further improve interpretability and robustness.

## 4. Conclusion

This study presented a learnable weighted ensemble architecture that integrates three complementary CNN backbones—ConvNeXt-Tiny, DenseNet-121, and EfficientNet-B0—combined with systematic TTA. The proposed design moves beyond conventional ensemble strategies that rely on static or heuristic weighting by introducing a differentiable fusion mechanism that dynamically optimizes the contribution of each backbone during training. Through this adaptive integration, the framework effectively exploits the complementary representational strengths of the participating architectures, enabling the model to capture both global spatial patterns and fine-grained disease characteristics present in complex agricultural imagery. Extensive experiments conducted under a stratified 5-FCVP across three heterogeneous eggplant datasets demonstrated the effectiveness and stability of the proposed approach. The ensemble consistently outperformed its individual backbone networks across all evaluation metrics, including ACC, PRC, RCL, F1-score, MCC, and Cohen's kappa. Notably, the proposed model achieved a peak classification accuracy of 99.70% on the Eggplant1 dataset while also maintaining strong and balanced performance on the more challenging Eggplant2 and Eggplant3 datasets. These results confirm that adaptive ensemble fusion significantly improves discriminative capacity and decision stability, particularly in scenarios

**Fig. 15.** Representative misclassified samples from Eggplant1-3.

**Table 6**
Ablation-style analysis of TTA within the proposed ensemble framework.

| Evaluation Dimension | Eggplant1 | Eggplant2 | Eggplant3 | TTA Benefit |
|---|---|---|---|---|
| Mean ACC | 99.55% → 99.70% | 93.03% → 93.75% | 95.14% → 95.57% | ↑ All datasets |
| Performance Variance | 0.0046 → 0.0090 | 0.0192 → 0.0124 | 0.0108 → 0.0088 | ↓ 2/3 datasets |
| Fold-to-Fold Consistency | Moderate | High | Moderate | Enhanced stability |
| Worst-Case Performance | 98.88% → 97.75% | 92.05% → 90.96% | 94.64% → 94.29% | Context-dependent |
| Best-Case Performance | 100% → 100% | 95.33% → 95.33% | 97.14% → 96.07% | Maintained |

characterized by class imbalance, intra-class variability, and visually similar disease symptoms.

A key contribution of this work lies in demonstrating that learnable ensemble weighting combined with augmented inference can substantially enhance robustness in plant disease recognition tasks. The differentiable weighting mechanism enables the system to automatically prioritize the most informative feature representations extracted by each backbone network, thereby reducing the influence of redundant or noisy features. At the same time, the integration of TTA strengthens prediction reliability by aggregating information from multiple spatially transformed views of the same input image. This dual mechanism—adaptive model fusion and multi-view inference—collectively improves the model's resilience to practical challenges such as illumination variability, background clutter, and orientation changes that frequently occur in real agricultural environments. Another important contribution of this study is the adoption of a multi-dataset evaluation strategy. While many previous studies report high accuracy on a single dataset, such evaluations often fail to capture generalization behavior across different acquisition conditions. By validating the proposed architecture on three independent eggplant datasets with distinct class structures, imaging conditions, and distribution characteristics, this work provides stronger evidence for the generalization capability of the proposed framework. The consistent performance observed across these datasets indicates that the model is capable of maintaining reliable diagnostic behavior even when confronted with heterogeneous visual conditions. From an application perspective, the proposed framework offers promising potential for deployment in precision agriculture systems. Reliable automated detection of eggplant diseases can support early intervention strategies, reduce unnecessary pesticide use, and ultimately contribute to more sustainable crop management practices. In addition, the reported computational performance demonstrates that the ensemble model remains computationally feasible for modern GPU-based agricultural monitoring systems, achieving stable inference latency while maintaining high diagnostic accuracy.

Despite these promising results, several directions remain open for further research. Future work will focus on improving cross-domain robustness through domain adaptation and cross-dataset training

**Table 7**
Contextual comparison with representative state-of-the-art methods in eggplant disease analysis.

| Ref | Method | Key Innovation | Test Dataset | Classes | Best Reported ACC | Our Method | Improvement (Δ) |
|---|---|---|---|---|---|---|---|
| [6] | Two-stream CNN fusion | CNN-SVM + CNN-Softmax hybrid | Single | 9 | 98.90% | 99.70% (Eggplant1) | +0.80% |
| [7] | DenseNet201 transfer learning | Single architecture optimization | Single | 14 | 99.06% | 99.70% (Eggplant1) | +0.64% |
| [9] | Ensemble transfer learning | Multi-architecture ensemble | Two | 7, 10 | 99.76%, 99.71% | 99.70%, 93.75%, 95.57% | Comparable on Eggplant1, superior on Eggplant2& Eggplant3 |
| [13] | YOLOv5s-BiPCNeXt | Lightweight real-time detection | Single | 3 | 94.9–99.5%* | 95.57% (Eggplant3) | +0.70–3.93% |
| [15] | Multi-model comparison | Architecture benchmarking | Single | 5 | 99.06% | 93.75% (Eggplant2) | Context-dependent |
| [18] | VGG16 + MSVM | Traditional ML + deep features | Single | 6 | 99.40% | 95.57% (Eggplant3) | Lower on this benchmark; evaluated across three additional datasets |

*[13] reports variable ACC per class: 94.9% (brown spot), 95.0% (powdery mildew), 99.5% (healthy).

**Table 8**
Experimentally measured computational performance.

| Performance Metric | ConvNeXt-Tiny | DenseNet-121 | EfficientNet-B0 | Proposed Ensemble | Measurement Protocol |
|---|---|---|---|---|---|
| Model Parameters | 28.6 M | 8.0 M | 5.3 M | 41.9 M | sum(p.numel()) |
| Theoretical FLOPs* | 4.47 G | 2.90 G | 0.40 G | 7.77 G | fvcore (224 × 224) |
| Training Time/epoch | 254 ± 18 s | 318 ± 22 s | 228 ± 16 s | 798 ± 42 s | 5-FCVP |
| Inference Latency | 12.8 ± 1.0 ms | 15.2 ± 1.2 ms | 10.6 ± 0.8 ms | 38.6 ± 2.8 ms | CUDA events |
| GPU Memory Peak | 1.84 GB | 1.62 GB | 1.41 GB | 4.92 GB | torch.cuda.max_memory_allocated() |
| Throughput | 78.1 img/s | 65.8 img/s | 94.3 img/s | 25.9 img/s | 1000 / latency |
| Training Speedup | 1.00 × (baseline) | 0.80 × | 1.11 × | 0.32 × | Relative to ConvNeXt-Tiny |
| ACC /Time Ratio | 3.71%/s | 2.98%/s | 4.19%/s | 2.48%/s | ACC / latency |

strategies that enable models trained on laboratory-level datasets to transfer more effectively to field environments. In addition, incorporating transformer-based architectures and multimodal sensing modalities—such as hyperspectral imaging or environmental sensor data—may further enhance diagnostic precision and interpretability under more complex agricultural scenarios. Exploring lesion-level localization mechanisms and explainable artificial intelligence techniques may also provide deeper insights into the visual features driving disease classification decisions.

## 5. Declaration of generative AI in scientific writing

The authors declare that no generative artificial intelligence tools were used to generate the scientific content, results, or conclusions of this manuscript. Generative AI tools were used only for language editing and grammar refinement, and the authors take full responsibility for the accuracy, originality, and integrity of the work.

## CRediT authorship contribution statement

**Ebru Ergün:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation. **Hatice Okumus:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation.

## Funding

This research was financially supported by the Recep Tayyip Erdogan University Development Foundation (Grant number: 02026003016184). The author also gratefully acknowledges the support provided by Recep Tayyip Erdoğan University through the Scientific Research Projects Coordination Unit (BAP) under project code FBA-2025-2236.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] Yu X, Liu S, Wang C, Jiao B, Huang C, Liu B, et al. Detection of fungal disease in citrus fruit based on hyperspectral imaging. Inform Process Agri 2025. https://doi.org/10.1016/j.inpa.2025.02.006.
[2] Zhang H, Ren G. Intelligent leaf disease diagnosis: image algorithms using Swin Transformer and federated learning. Vis Comput 2025;41(7):4815–38. https://doi.org/10.1007/s00371-024-03692-w.
[3] Lee S, Yun CM. A deep learning model for predicting risks of crop pests and diseases from sequential environmental data. Plant Methods 2023;19(1):145. https://doi.org/10.1186/s13007-023-01122-x.
[4] Huang X, Xu D, Chen Y, Zhang Q, Feng P, Ma Y, et al. EConv-ViT: a strongly generalized apple leaf disease classification model based on the fusion of ConvNeXt and transformer. Inform Process Agri 2025. https://doi.org/10.1016/j.inpa.2025.03.001.
[5] Ergün E. High precision banana variety identification using vision transformer based feature extraction and support vector machine. Sci Rep 2025;15(1):10366. https://doi.org/10.1038/s41598-025-95466-0.
[6] Haque MR, Sohel F. Deep network with score level fusion and inference-based transfer learning to recognize leaf blight and fruit rot diseases of eggplant. Agriculture 2022;12(8):1160. https://doi.org/10.3390/agriculture12081160.
[7] Saad IH, Islam MM, Himel IK, Mia MJ. An automated approach for eggplant disease recognition using transfer learning. Bull Electr Eng Inform 2022;11(5):2789–98. https://doi.org/10.11591/eei.v11i5.3575.
[8] Kursun R, Koklu M. Classification of eggplant diseases using feature extraction with AlexNet and Random Forest. In: Proceedings of the International Conference on Artificial Intelligence and Data Science; 2025. p. 45–52.
[9] Siddiqui MIH, Limon ZH, Khan S, Khan MA, Rahman H, et al. Eggplant disease diagnosis using a robust ensemble of transfer learning architectures. In: Proceedings of the 2025 International Conference on Electrical, Computer and Communication Engineering (ECCE); 2025. p. 1–6. https://doi.org/10.1109/ECCE64574.2025.11013918.
[10] Zhang Y, Zhang D, Zhang Y, Cheng F, Zhao X, et al. Early detection of verticillium wilt in eggplant leaves by fusing five image channels: a deep learning approach. Plant Methods 2024;20(1):173.
[11] Kaniyassery A, Goyal A, Thorat SA, Rao MR, Chandrashekar HK, et al. Association of meteorological variables with leaf spot and fruit rot disease incidence in eggplant and YOLOv8-based disease classification. Eco Inform 2024;83:102809. https://doi.org/10.1016/j.ecoinf.2024.102809.
[12] Meng D, Ma D, Ma M. Applying few-shot transfer learning with RandAugment and MCA-RepVGG-A-B3 network for eggplant leaf disease classification. In: Proceedings of the 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON); 2024. p. 1–7. https://doi.org/10.1109/NMITCON62075.2024.10698826.
[13] Xie Z, Li C, Yang Z, Zhang Z, Jiang J, Guo H. YOLOv5s-BiPCNeXt, a lightweight model for detecting disease in eggplant leaves. Plants 2024;13(16):2303. https://doi.org/10.3390/plants13162303.
[14] Wang X, Yan F, Li B, Yu B, Zhou X, et al. A multimodal data fusion and embedding attention mechanism-based method for eggplant disease detection. Plants 2025;14(5):786. https://doi.org/10.3390/plants14050786.
[15] Gayathri R, Umadevi K. Early identification of eggplant diseases using cutting-edge deep learning models. In: Proceedings of the 2025 3rd International Conference on Data Science and Information System (ICDSIS); 2025. p. 1–7. https://doi.org/10.1109/ICDSIS65355.2025.11070910.
[16] Rangarajan, A. K., Purushothaman, R., Prabhakar, M., Szczepański, C.: Crop identification and disease classification using traditional machine learning and deep learning approaches. Journal of Engineering Research, vol. 11, no. 1B (2023). https://doi.org/10.36909/jer.11941.
[17] Abisha S, Mutawa AM, Murugappan M, Krishnan S. Brinjal leaf diseases detection based on discrete Shearlet transform and Deep Convolutional Neural Network. PLoS One 2023;18(4):e0284021. https://doi.org/10.1371/journal.pone.0284021.
[18] Karthikeyan M, Yogiraj GP, Elaiyabharathi T, Jesu BAJ, Johnson I, et al. Comprehensive analysis of little leaf disease incidence and resistance in eggplant. BMC Plant Biol 2024;24(1):576. https://doi.org/10.1186/s12870-024-05257-4.
[19] Krishnaswamy Rangarajan A, Purushothaman R. Disease classification in eggplant using pre-trained VGG16 and MSVM. Sci Rep 2020;10(1):2322. https://doi.org/10.1038/s41598-020-59108-x.
[20] Xie C, He Y. Spectrum and image texture features analysis for early blight disease detection on eggplant leaves. Sensors 2016;16(5):676. https://doi.org/10.3390/s16050676.
[21] Xie C, Feng L, Feng B, Li X, Liu F, He Y. Relevance of hyperspectral image feature to catalase activity in eggplant leaves with grey mold disease. Trans Chin Soc Agri Eng 2012;28(18):177–84.
[22] Wu D, Feng L, Zhang C, He Y. Early detection of Botrytis cinerea on eggplant leaves based on visible and near-infrared spectroscopy. Trans ASABE 2008;51(3):1133–9. https://doi.org/10.13031/2013.24504.
[23] Raza H, Abu Bakr M, Khan SD, Batool H, Ullah H, Ullah M. Benchmarking YOLO models for crop growth and weed detection in cotton fields. AgriEngineering 2025;7(11):1–375. https://doi.org/10.3390/agriengineering7110375.
[24] Bakr MA, Khan AJ, Khan SD, Zafar MH, Ullah M, Ullah H. Evaluation of learning-based models for crop recommendation in smart agriculture. Information 2025;16(8):622–32. https://doi.org/10.3390/info16080632.
[25] Khan SD, Basalamah S, Naseer A. Classification of plant diseases in images using dense-inception architecture with attention modules. Multimed Tools Appl 2025;84(19):20607–32. https://doi.org/10.1007/s11042-024-19860-y.
[26] Hasan R, Hossain Sanzit S, Hosen MM, Hasan F, Topu MMH, Islam M. High-resolution eggplant leaf image dataset for plant disease classification and detection. Mendeley Data 2025;6. https://doi.org/10.17632/ss63ftnjnh.6.

[27] Hasan MZ, Bitto AK, Bijoy MHI. BrinjalFruitX: a field-collected image dataset for machine learning and deep learning-based disease identification in brinjal fruits. Mendeley Data 2025;1. https://doi.org/10.17632/ngc58fsxgd.1.
[28] Siddique Chaity S, Raj Saha J, Hasan R. A curated dataset of eggplant leaves: healthy specimens and six major disease conditions. Mendeley Data 2025;1. https://doi.org/10.17632/jv6tnm4t5c.1.
[29] Kim HK, Kim JD. Region-based shape descriptor invariant to rotation, scale and translation. Signal Process Image Commun 2000;16(1–2):87–93. https://doi.org/10.1016/S0923-5965(00)00018-7.
[30] Ergün E. Attention-enhanced hybrid deep learning model for robust mango leaf disease classification via ConvNeXt and vision transformer fusion. Front Plant Sci 2025;16:1638520. https://doi.org/10.3389/fpls.2025.1638520.
[31] Chun B. NeXtSRGAN: enhancing super-resolution GAN with ConvNeXt discriminator for superior realism. Vis Comput 2025;41(10):7141–67. https://doi.org/10.1007/s00371-024-03797-2.
[32] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.: Q. Densely connected convolutional networks. In IEEE conference on computer vision and pattern recognition, pp. 4700–4708 (2017). https://doi.org/10.48550/arXiv.1608.06993.
[33] Putra IPGYP, Dewi NWJK, Lesmana PSW, Suryawan IGT, Putra PSU. Comparison of ResNet-50 and DenseNet-121 architectures in classifying diabetic retinopathy. Indonesian Journal of Data and Science 2025;6(1):64–72. https://doi.org/10.56705/ijodas.v6i1.232.
[34] Kansal K, Chandra TB, Singh A. ResNet-50 vs. EfficientNet-B0: multi-centric classification of various lung abnormalities using deep learning. Procedia Comput Sci 2024;235:70–80. https://doi.org/10.1016/j.procs.2024.04.007.
[35] Yustanti W. Implementation of EfficientNet-B0 CNN model for web-based strawberry plant disease detection. J Emerg Inform Syst Bus Intell 2025;6(3):346–54. https://doi.org/10.26740/jeisbi.v6n3.72957.
[36] San KK, Win HH, Chaw KEE. Enhancing hybrid course recommendation with weighted voting ensemble learning. J Fut Artif Intell Technol 2025;1(4):337–47. https://doi.org/10.62411/faith.3048-3719-55.
[37] Altunkaya AN, Ozkat EC, Avci M. Analytical-to-AI pipeline: Modeling and optimization of entropy generation in pulsating non-Newtonian heat flow. Comput Math Appl 2026;205:195–211. https://doi.org/10.1016/j.camwa.2025.12.021.
[38] Feng CM, He Y, Zou J, Khan S, Xiong H, et al. Diffusion-enhanced test-time adaptation with text and image augmentation. Int J Comput Vis 2025. https://doi.org/10.1007/s11263-025-02412-8.
[39] Enomoto S, Busto MR, Eda T. Dynamic test-time augmentation via differentiable functions. IEEE Access 2024;12:123456–67. https://doi.org/10.1109/ACCESS.2024.3477533.
[40] Ergün E, Aydemir O, Korkmaz OE. A novel scrolling text reading paradigm for improving the performance of multiclass and hybrid brain computer interface systems. PLoS One 2025;20(5):e0322711. https://doi.org/10.1371/journal.pone.0334784.
[41] Ergün E. SwinFishNet: a Swin Transformer-based approach for automatic fish species classification using transfer learning. PLoS One 2025;20(5):e0322711. https://doi.org/10.1371/journal.pone.0322711.

**Ebru Ergün**, born in 1991 in Trabzon, Turkey, earned her B.Sc. in Electrical and Electronic Engineering from Karadeniz Technical University in 2014. She pursued her M.Sc. at the same institution, completing it in 2017, followed by a Ph.D. program in Electrical and Electronic Engineering. Since 2015, she has served as a research assistant at Recep Tayyip Erdoğan University. Her research interests include biomedical engineering, brain-computer interfaces, and machine learning.

**Hatice Okumuş** received the B.Sc. and M.Sc. degrees from the Department of Electrical and Electronics Engineering, Karadeniz Technical University, Trabzon, Turkey, in 2014 and 2016, respectively. She completed her Ph.D. in the same department in 2022. Her research interests include machine learning, signal processing, power system modeling and analysis, transmission and distribution, protection systems.