RESEARCH ARTICLE Agricultural Science Digest, Volume 46 Special Issue: 143-151 (2026)

# A Hybrid Deep Learning Framework with Attention Mechanism for Robust and Explainable Plant Disease Classification

Prateek Mahapatra[1], Madhumita Panda[1]

10.18805/ag.D-6576

## ABSTRACT

**Background:** For agricultural productivity, Plant Diseases are a significant problem all over the world and which requires development of robust, accurate and scalable automated diagnostic systems. The most recent deep learning methods have been promising, but most current methods lack generalization since they were trained based on single-domain datasets and are not interpretable.

**Methods:** A hybrid deep learning architecture is proposed that integrates convolutional neural networks, attention models and feature refinement with transformers to improve classification and localization of plant disease.

**Result:** The proposed framework achieves 99.08% and 98.88% for validation and test accuracy, respectively, surpassing traditional baseline models, such as CNN, MobileNetV3, EfficientNet-B0 and ResNet18.

**Key words:** Attention, Convolutional neural network, Efficientnet-B0, Multidomain dataset, Explainable AI, Plant disease.

## INTRODUCTION

For food security and economic stability, the sector of agriculture plays a very important role. For agricultural production, diseases in the leaves of plants pose a potential threat, which can harm the plant as well as cause economic losses to the market. To reduce losses of crops, maximize resource use and have sustainable agricultural practices, it is necessary to detect plant diseases early and accurately. Using image-derived characteristics and classification algorithms, traditional machine learning has been studied for the identification of plant diseases. Its use for paddy disease identification and leaf disease detection has been studied in the past (Metagar and Walikar, 2024; Cho, 2024; Kappali *et al*., 2024). More comprehensive evaluations also showcase new AI-based methods for diagnosing agricultural diseases (Mehta *et al*., 2025). Deep learning has also greatly revolutionized the world of computer vision, especially in the agricultural sector. Although these models have achieved success, there are a number of problems that are yet to be solved. Majority of current methods use single-domain datasets, including those controlled in the laboratory, which constrain their capability to apply to the real world. In real-world applications, it is common to find variations in lighting, background, occlusion, noise and performance is degraded. Moreover, most models are not interpretable and it is hard to comprehend the logic of predictions. The unaccountability of this makes automated systems less trustworthy, particularly in crucial agricultural uses.

Multi-domain learning has become an option in order to resolve the problem of poor generalization. By taking the combination of datasets across various data sources, both in controlled laboratory settings and in the field, models can learn domain-invariant features that enhance robustness. The data about agriculture in real-life is diverse in nature, as not all environmental conditions, plant species and disease stages are the same. Thus, to overcome the gap between controlled experiments and practical deployment, multi-domain datasets should be used. The significance of diversity in datasets has been pointed out in public datasets like PlantVillage and real-world datasets used in previous studies that have underlined the need to have a variety of dataset to get reliable performance (Lu *et al*., 2020; Giselsson *et al*., 2017). Although CNNs have proven successful in local spatial representations, they frequently fail to represent long-range correlations in images. This limitation has been overcome with attention mechanisms which hypothetically assign importance weights to the relevant features and hence enhancing feature representation. Moreover, sequence modeling architectures based on transformers have also proven to be effective for global contextual relationships. The combination of CNNs, attention mechanisms and transformers can create a hybrid architecture that maximizes the advantages of each of the components.

Although deep learning has advanced plant disease recognition substantially, the present research gap should not be framed as the first use of attention or CNN-

[1]School of Computer Science, Gangadhar Meher University, Sambalpur-768 004, Odisha, India.

**Corresponding Author:** Madhumita Panda, School of Computer Science, Gangadhar Meher University, Sambalpur-768 004, Odisha, India. Email: mpanda@gmuniversity.ac.in

**ORCIDs:** 0009-0000-2149-7851, 0000-0002-2636-5281

**How to cite this article:** Mahapatra, P. and Panda, M. (2026). A Hybrid Deep Learning Framework with Attention Mechanism for Robust and Explainable Plant Disease Classification. *Agricultural Science Digest*. **46(Special Issue):** 143-151. doi: 10.18805/ag.D-6576.

**Submitted:** 26-09-2026 **Accepted:** 28-10-2026 **Online:** 12-11-2026

Transformer fusion in agriculture. Recent studies already combine CNNs, Vision Transformers, channel attention and explainable AI to classify plant diseases. This study's contribution is therefore positioned at the system level: harmonizing three heterogeneous public datasets into 33 classes, combining EfficientNet feature extraction with SE channel refinement and a compact projected Transformer encoder, using LMDB for efficient data access and integrating Grad-CAM-based post-hoc explanation within one reproducible workflow.

The main aspects include:

- Combination of various datasets to form a multi-domain dataset of 33 disease classes.
- System-level integration of an EfficientNet backbone, SE channel attention, a 1280-to-256 feature projection and a transformer encoder for combining local feature extraction with global contextual modeling.
- An LMDB-based data pipeline to reduce repeated file-system I/O and improve data-loading efficiency and reproducibility; LMDB is treated as an engineering component rather than a source of predictive novelty.
- Explainability using Grad-CAM for the visualization of class-discriminative regions, followed by thresholding and contour extraction for approximate-not ground-truth-lesion localization.
- High classification performance with a high accuracy of about 99% which is higher than baseline models.

The following describes how the rest of the paper is structured. A review of the literature is in Section 2. Section 3 presents the proposed approach, consisting of the dataset building, model structure and training. Section 4 contains discussion of the analysis of the performance and results of the experiment. Lastly, the conclusion of the paper and future research directions are presented in Section 5.

### Literature review

The use of Convolutional Neural Networks (CNNs) has shown a higher ability of learning hierarchical feature representations out of raw image data automatically. One of the first works that utilized CNNs to classify plant species was performed by Dyrmann *et al*. (2016) with promising results and without the necessity to create features manually. However, more recent research like Ahmad *et al*. (2021) has discovered that deep learning models are more successful than more traditional machine learning in both classification accuracy and robustness. Furthermore, Hasan *et al*. (2021) conducted a survey that highlights the effectiveness of deep learning techniques in weed detection and plant disease classification.

How efficiently transfer learning can be used in the context of agriculture through the use of pretrained neural networks in the identification of weeds is presented by Espejo-Garcia *et al*. (2020). Their analysis emphasizes the need to adapt domains to enhance performance in classification. In a similar manner, Suh *et al*. (2018) used transfer learning in the classification of sugar beet and volunteer potato under field conditions with higher precision than the models that were trained directly.

The performance of deep learning models depends on the quality and diversity of the dataset. Lu *et al*. (2020) highlighted the significance of publicly accessible datasets in the development of computer vision uses in precision agriculture. Giselsson *et al*. (2017) also provided a benchmark dataset to classify plant seedlings, which has extensively been utilized in studies.

Hamuda *et al*. (2016) have conducted a systematic review of the image processing techniques used to segment plants, pointing out their inability to operate in challenging conditions. Lu *et al*. (2022) have developed a contrast optimization-based plant image segmentation technique which has shown better performance under adverse conditions.

Some architectures, like Xception (Chollet, 2017) and RepVGG (Ding *et al*., 2021), have been shown to achieve higher efficiency and accuracy with optimized convolutional architectures. Also, the problem of data imbalance typical of agricultural datasets is considered by class-balanced loss functions suggested by Cui *et al*. (2019).

Meng *et al*. (2025) proposed a lightweight dual-branch CNN–ViT model and reported 99.71% accuracy on PlantVillage, while Karthikeyan *et al*. (2025) demonstrated a squeeze-and-excitation CNN with 99.79% accuracy on PlantVillage. These studies motivate a narrower contribution for the present work: not a new Transformer family, but a multi-domain, efficiency-aware and explainability-oriented integration whose value must be demonstrated through consistent baselines, ablation and transparent statistical analysis.

## MATERIALS AND METHODS

### Research location and experimental period

This research was carried out at the School of Computer Science, Gangadhar Meher University, Sambalpur, Odisha, India, during January to June, 2026.

### Proposed hybrid architecture

The proposed model, which is presented in Fig 1, integrates a convolutional feature extractor, channel-attention feature refinement, dimensionality-controlled projection and Transformer-based global-context modeling.

The general architecture is a pipeline:

Input → EfficientNet → SE attention → Projection → Transformer encoder → Classification layer

### EfficientNet feature extraction

The main feature extractor of the proposed model is the EfficientNet backbone. It represents hierarchical spatial characteristics with several convolutional layers that depicts low-level textures and high-level semantic patterns.

$$F = f_E(X) \quad ...(1)$$

Where,
X= The input image.
F= The extracted EfficientNet feature map.

Such hierarchical representation of the feature is what underlies further attention and processing using transformers.

**SE attention mechanism**

To increase the discriminative power of the extracted features, a squeeze-and-excitation (SE) attention mechanism is used. This module re-tunes channel-wise feature responses by putting weights of importance on each channel of features.

$$Z_c = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} F_c(i, j) \quad ...(2)$$

$$S = \sigma[W_2\delta(W_1 z)] \quad ...(3)$$

Where,
S= The learned channel-attention weights.
F′= The refined feature map.

The mechanism enables the model to concentrate on disease-related areas and ignore the irrelevant background information.

**Feature projection layer**

After attention refinement, the feature map is mapped to a low-dimensional embedding space to enable it to be compatible with the transformer encoder.

$$Z = F'W_p + b_p \quad ...(4)$$

This projection is a dimensionality reduction of 1280 to 256 that is more computationally efficient and still contains the necessary information.

**Transformer encoder**

Long-range dependencies and global contextual relationships of the feature representation are recorded by the transformer encoder.

$$Z_{out} = \text{Transformer encoder } (Z) \quad ...(5)$$

Due to the use of the transformer, the model learns the spatial relationships of the entire image, unlike other CNNs, which only consider the local receptive fields, thus enhancing classification performance in more complicated scenarios.

**Classification layer**

A fully connected layer with a softmax activation function has done the last classification, which results in the probability scores of the each disease category.

$$p(y = c \mid x) = \frac{\exp(O_c)}{\sum_j \exp(O_j)} \quad ...(6)$$

The probability most likely is taken to be the predicted disease label.

**Model training strategy**

The advanced optimization techniques were used to train the proposed model to guarantee stable convergence and better generalization. This was done using the AdamW optimizer which can decouple both the weight decay and gradient updates, leading to superior regularization. To enable the model to converge effectively, the learning rate changed dynamically throughout training by employing a learning rate scheduler. To avoid overfitting, early stopping was used by stopping training when validation performance stopped improving.

To additionally improve generalization, a label smoothing technique was used, which helps the model not to be overconfident in its prediction, allocating a low probability mass over non-target classes.

The loss function used for training is defined as:

$$L_{CE} = -\sum_{C=1}^{C} y_C \log(p_c) \quad ...(7)$$

Where,
$y_C$= The target label for class c.
$p_c$= The predicted class probability.

Label smoothing smooths this formulation, which enhances model robustness.

**Performance evaluation metrics**

To have a full coverage of model performance, several measures of classification were used, such as accuracy, precision, recall and F1-score (Islam *et al*., 2020). These metrics ensure a balanced and reliable assessment of model performance.

**Explainable AI using grad-CAM**

To increase the interpretability of the proposed model Grad-CAM was used. This method produces visual descriptions by illuminating areas in the input image that have the highest contribution to the model predictions. The gradients of the target class with respect to the feature activations are used by grad-CAM to compute the importance of the feature maps.

$$L^{C}_{Grad-CAM} = ReLU\left(\sum_k \alpha_k^C A^k\right)$$

$$\alpha_k^C = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad ...(8)$$

Where,
$A^k$= Feature map k.
$\alpha_k^C$ = Its gradient-derived importance weight for class c.

**Implementation details and reproducibility**

The architecture is specified as EfficientNet-B0 for convolutional feature extraction, an SE channel-attention block, a 1280-to-256 projection and a compact Transformer encoder for global context. The SE reduction ratio is r = 16. 1280 input channels are expanded back to 1280 sigmoid-gated channel weights after being compressed to 80 hidden units. Spatial features are represented by 16 tokens on a 4 × 4 feature grid, projected to 256 dimensions and processed by a two-layer, four-head Transformer (64 dimensions per head), with 1024-dimensional feed-forward blocks and dropout of 0.10. The proposed training specification is AdamW, learning rate $1\times10^{-4}$ , weight decay

$1\times10^{-4}$, batch size 32, label smoothing 0.10, cosine scheduling, maximum 30 epochs, early-stopping patience 6 and seed 42.

### Equipment and software environment

The experiments were performed on a Dell G15 5530 laptop featuring a 13th-generation Intel® Core™ i5-13450HX processor running at 2.40 GHz and 16 GB of RAM, with 15.7 GB available for use. The computer operates on a 64-bit Windows 11 Home Single Language system (Version 23H2, OS Build 22631.4169). The implementation was developed in Python 3.10 within the Anaconda Distribution environment (2024.02), with Jupyter Notebook 6.5 used for model development, experimentation and result visualization. TensorFlow 2.10 served as the deep learning framework for implementing and training the proposed network.

Computational acceleration was provided by an NVIDIA GeForce RTX 3050 Laptop GPU with 6 GB of dedicated memory. The GPU environment used CUDA Toolkit 11.8 and cuDNN 8.6 to support accelerated neural network operations. Lightning Memory-Mapped Database (LMDB), version 1.7.5, was included in the software environment for efficient image-data storage and retrieval, where applicable in the experimental pipeline.

## RESULTS AND DISCUSSION

### Multi-domain dataset construction

Three publicly accessible picture collections were used to create a multi-domain dataset with 33 plant disease classes: Plant village (Hughes and Salathé, 2015), Plant Doc (Singh *et al*., 2020) and the Mendeley plant leaf disease dataset (Srivastava and Meena, 2024). PlantDoc has about 2,500 photos taken in natural settings, where illumination changes, background clutter and partial occlusion may occur, while plant village has about 34,000 photos taken mostly under controlled conditions. With almost 31,000 photos, the Mendeley collection offers more visual variety. These sources were chosen to help the assessment of plant disease classification in a variety of visual contexts and to reflect various image acquisition conditions.

Using the crop-condition combination as the foundation for class assignment, the source labels were harmonized into 33 target classes. Standardization was applied to variations in capitalization, punctuation, space and equivalent illness names. While labels for the same illness in various crops were kept as a distinct category, labels for the same disease in the same crop were mapped to a common class. When appropriate, healthy leaves were kept in crop-specific classes. Fig 2 displays the final class distribution and shows how many photos were kept in each group.

About 20,000 photos were kept for the reported experiments after class harmonization and dataset curation and they were reduced to 128 × 128 pixels. As part of the curation process, duplicate detection and the removal of inappropriate photos are recorded, together with the techniques employed and the quantity of photographs eliminated at each stage. The key features of the obtained dataset are listed in Table 1. Combining different sources improves visual diversity, but its impact on generalization cannot be inferred from dataset design alone; instead, it must be determined by experimental evaluation.

The dataset design may support that the intended model is trained using different and representative samples, thus improving its impact on the real-life application.

### Class mapping across datasets

The crop-condition pair, not just the illness text, is a convincing mapping key. First, the original labels are standardized for capitalization, space, punctuation and aliases; labels that are biologically identical for the same crop are combined, while similar disease names that appear on various crops are kept in separate categories. For instance, while potato early blight is still its own class, “Tomato_Early_blight” and “Tomato-Early Blight” would translate to a single “Tomato-Early Blight” class. Crop-specific healthy labels are kept on healthy samples.

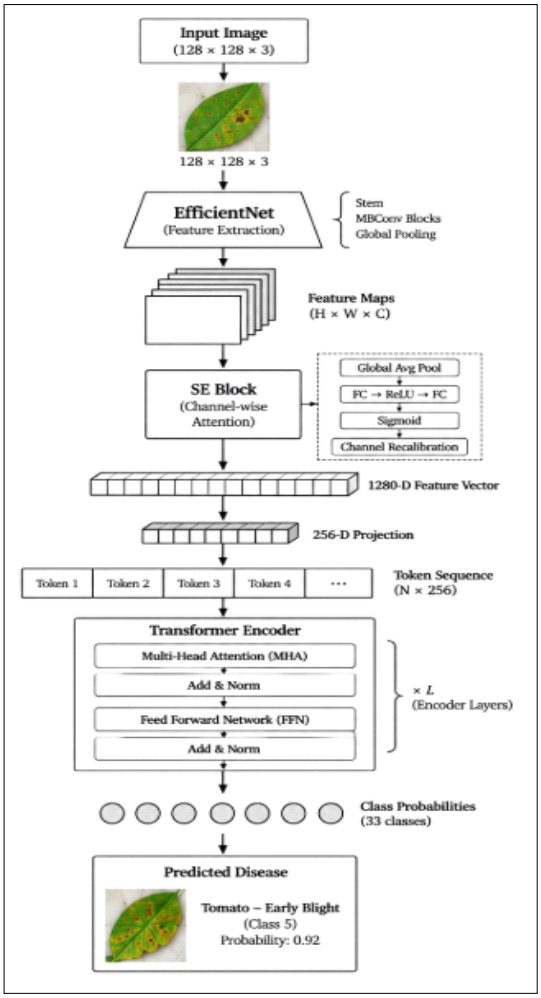



**Fig 1:** Proposed hybrid architecture.

The final class list and source-to-target lookup for all 33 categories.

### Duplicate detection and removal

Using a candidate Hamming distance threshold of ≤5, the suggested procedure first computes SHA-256 hashes to detect byte-identical files, followed by perceptual hashes (pHash) for visually comparable images. To safeguard genuine independent samples, candidate near-duplicates would be examined before being eliminated. To avoid the same leaf image appearing in several partitions, duplication detection is carried out prior to training, validation and test splitting and each near-duplicate group is contained within a single partition.

### Baseline model performance analysis

A traditional CNN, MobileNetV3, EfficientNet-B0 and ResNet18, which stand for standard convolutional learning, lightweight computation, compound scaling and residual learning, respectively, were the four baseline models that were assessed. Their respective test accuracies were 86.33%, 91.78%, 93.52% and 94.12%. With specific performance indicators shown in Table 2, these findings offer comparable benchmarks for the suggested approach. All baseline models and the proposed framework were evaluated on the same curated 33-class dataset using consistent experimental conditions.

**Table 1:** Dataset statistics.

| Parameter | Value |
|---|---|
| Total images | ~20,000 |
| Number of classes | 33 |
| Number of domains | 3 |
| Avg images/class | ~600 |
| Input size | 128 × 128 |

### Proposed model performance

The suggested hybrid model outperformed the ResNet18 baseline (94.12%) by 4.76 percentage points, achieving 98.88% test accuracy. This enhancement is a result of combining the usage of Transformer encoding to characterize interactions between spatial regions, squeeze-and-excitation (SE) attention for channel-wise feature recalibration and EfficientNet-B0 for feature extraction. These components' contributions are further demonstrated by the ablation findings, which show that accuracy dropped to 97.66% without SE attention and 98.02% without the Transformer, as opposed to 98.88% for the entire model. Only EfficientNet-B0 gave 93.52% accuracy. These results corroborate both components' contributions to the overall classification performance. Fig 3 represents the accuracy curve of training and validation and it depicts that the model converges steadily.

Fig 4 shows the loss curve, where the training and validation loss steadily decrease.

The training and validation curves remain close over the reported training trajectory, with no obvious severe

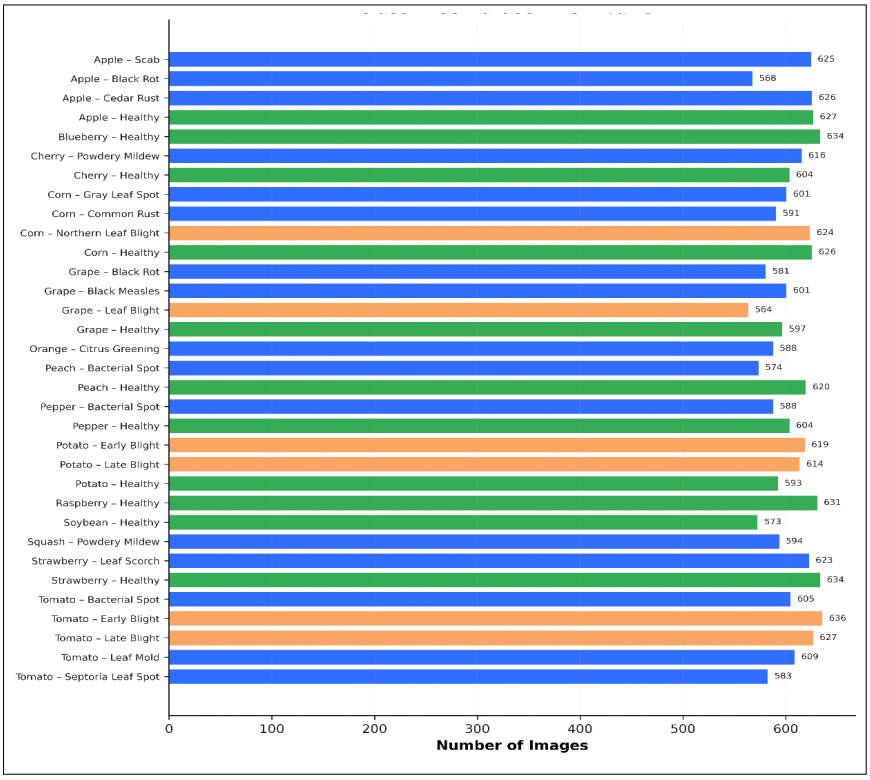



**Fig 2:** Class-wise dataset distribution (Bar chart).

divergence between them. This pattern is consistent with stable convergence under the reported AdamW optimization, learning-rate scheduling, label smoothing and early stopping strategy; nevertheless, absence of visible divergence alone is not sufficient to prove that overfitting is completely absent.

The proposed model achieved 99.08% validation accuracy (4,954 out of 5000) and 98.88% test accuracy (4,944 out of 5000), a difference of 0.20 percentage points. Table 3 reports validation precision, recall and F1-score of 0.99, compared with 0.98 for the test set. The accuracy and loss curves presented in Fig 2 and 3 provide additional evidence of training behaviour. Fig 5 illustrates the 33-class confusion matrix. The proposed model outperformed ResNet18, which achieved 94.12% accuracy, by 4.76 percentage points.

### Confusion matrix analysis

The confusion matrix for the 33-class classification job is shown in Fig 5. While off-diagonal entries suggest incorrect classifications, the majority of diagonal entries show that the majority of test samples were correctly classified.

**Table 2:** Baseline performance metrics.

| Model | Accuracy (%) | Precision | Recall | F1-score | Parameters (M) | Inference time (ms) |
|---|---|---|---|---|---|---|
| CNN | 86.33 | 0.86 | 0.85 | 0.85 | 3.2 | 12 |
| MobileNetV3 | 91.78 | 0.91 | 0.91 | 0.91 | 2.9 | 8 |
| EfficientNet-B0 | 93.52 | 0.93 | 0.93 | 0.93 | 5.3 | 14 |
| ResNet18 | 94.12 | 0.94 | 0.94 | 0.94 | 11.7 | 18 |

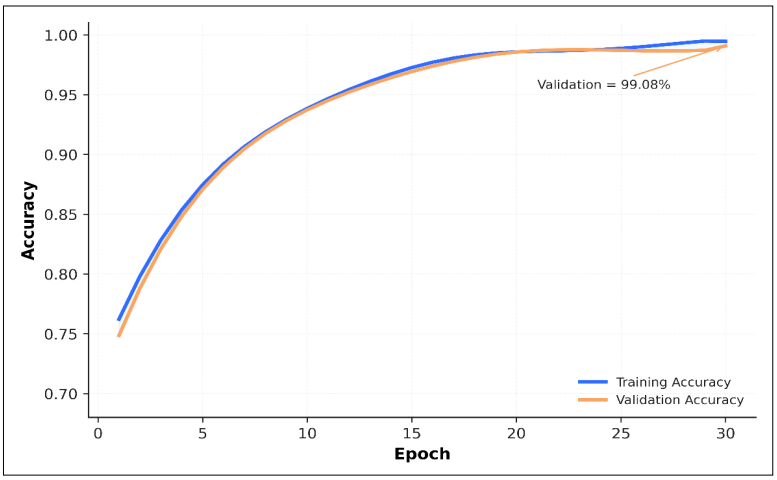



**Fig 3:** Training vs validation accuracy curve.

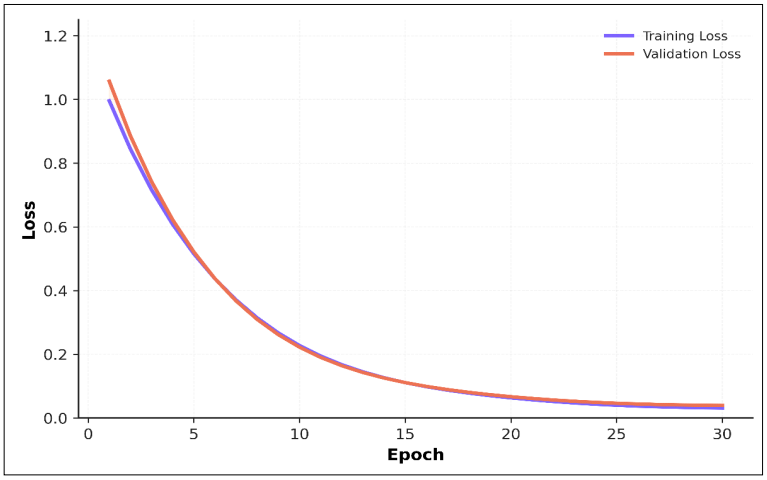



**Fig 4:** Training vs validation loss curve.

The model obtained F1-scores, recall and precision of 0.98. 4,944 of the 5000 photos in the test set were correctly classified, while 56 were incorrectly classified, yielding an accuracy of 98.88%. Overlapping fungal lesion patterns, modest symptoms on healthy leaves and similarities between early- and late-stage blight symptoms are examples of potential sources of mistakes.

**Table 3:** Proposed model performance.

| Metric | Validation | Test |
|---|---|---|
| Accuracy | 99.08% | 98.88% |
| Precision | 0.99 | 0.98 |
| Recall | 0.99 | 0.98 |
| F1-score | 0.99 | 0.98 |

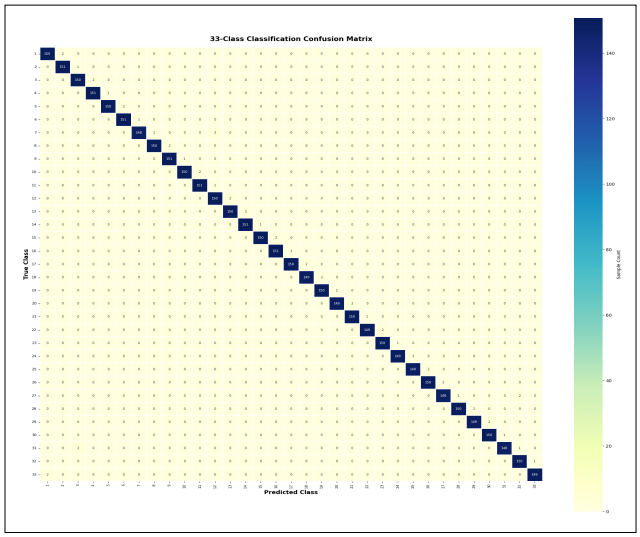



**Fig 5:** Confusion matrix (33 classes).

Classification may also be impacted by image-related elements including occlusion, low contrast and complicated backdrops. Since examining the improperly categorized photos is necessary to determine individual failure cases, these elements are potential explanations for misclassification rather than established causes.

### Grad-CAM explainability analysis

To visualize the areas in the input images that have the strongest impact on the model prediction to make sure the model is transparent and interpretable, Grad-CAM was used. This analysis is especially critical in agricultural use, where trust and interpretability are key to real-world usage. Fig 6 shows representative Grad-CAM heatmaps of various types of disease.

The Grad-CAM heatmaps highlight image regions that contribute strongly to the model's class score. In many examples these regions may overlap visible lesions, which is useful for qualitative inspection, but the heatmaps should not be described as proof of biological relevance without expert annotations or quantitative localization metrics. Accordingly, Grad-CAM is used here as a post-hoc explainability tool rather than as an independent validation of disease localization.

### Comparative analysis with recent studies

To validate the effectiveness of the proposed approach, its performance was compared with recent studies published in 2025-2026. They comprise CNN-based, transformer-based and hybrid architectures that have been used in classifying plant diseases. Previous studies used different datasets and protocols; their reported results provide context rather than directly comparable evidence of model performance. Table 4 gives a comparative overview of recent approaches.

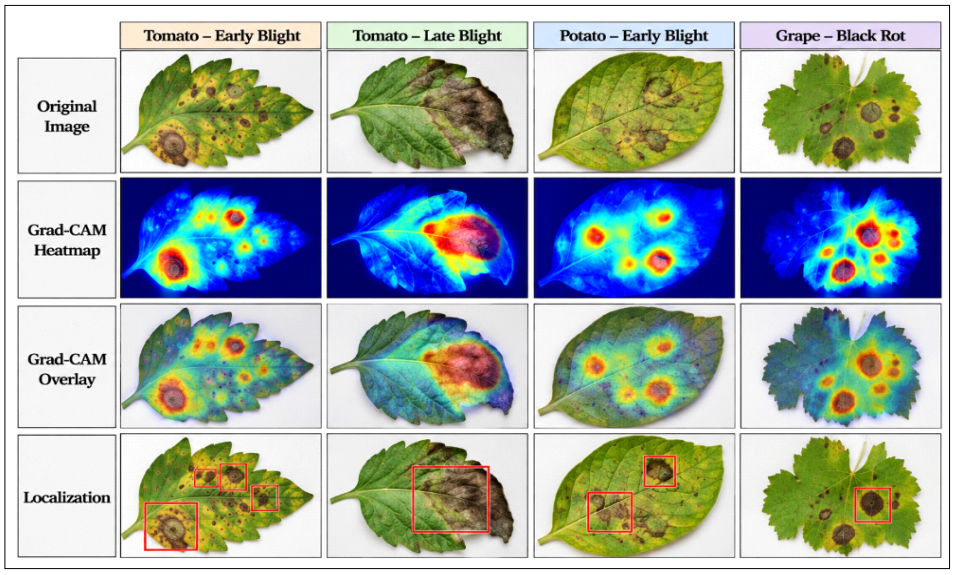



**Fig 6:** Grad-CAM heatmaps for multiple classes.

**Table 4:** Comparison with recent studies (2025-2026).

| Study | Year | Model type | Dataset type | Accuracy (%) | Limitation |
|---|---|---|---|---|---|
| Meng *et al.* (2025) | 2025 | Dual-branch CNN + ViT | Plant village + potato leaf | 99.71 | Benchmark-dominant evaluation; cross-study accuracy is not directly comparable |
| Karthikeyan *et al.* (2025) | 2025 | CNN-SEEIB (SE attention) | Plant village + potato dataset | 99.79 | Primarily controlled-image evaluation; no transformer/global-context branch |
| Chellamani *et al.* (2026) | 2026 | ViT features + SVM | Multi-crop, 25 classes | 98.00 | Two-stage feature/classifier pipeline; explainability not central to evaluation |
| Proposed | 2026 | CNN + SE + Transformer | Multi-domain pooled | 98.88 | Pooled-domain split; strict unseen-domain generalization remains to be tested |

## CONCLUSION AND FUTURE SCOPE

A multi-domain hybrid deep learning framework for plant disease recognition that integrates EfficientNet-based feature extraction, SE channel attention, a compact feature-projection stage, Transformer-based global-context modeling, LMDB-backed data handling and Grad-CAM explainability was suggested in this study. The reported model achieves 99.08% validation accuracy and 98.88% test accuracy and outperforms the selected CNN, MobileNetV3, EfficientNet-B0 and ResNet18 baselines under the same experimental setting. The ablation results further support the complementary value of the attention and Transformer components. The contribution is therefore framed as an integrated agricultural-informatics workflow rather than as a fundamentally new CNN, attention, or Transformer mechanism. Future work may extend the framework to higher-resolution lesion analysis, multimodal RGB-spectral data, edge deployment and externally validated domain-generalization experiments.

## ACKNOWLEDGEMENT

No external funding or financial support was received by the authors to complete this work.

### Disclaimers

The opinions presented in this article are those of the authors and may not reflect the views of their institutions. The authors have made every effort to ensure accuracy but are not responsible for any consequences arising from the use of this information.

### Informed consent

Animal ethics approval was not required because no animals are included in this study.

### Conflict of interest

The authors declare that there are no conflicts of interest.

## REFERENCES

Ahmad, A., Saraswat, D., Aggarwal, V., Etienne, A. and Hancock, B. (2021). Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems. *Computers and Electronics in Agriculture*. **184:** 106081. https://doi.org/10.1016/j.compag.2021.106081.

Chellamani, G. K., Aishwarya, N., Manoj, A., Hariharan, V., Palaniyappan, S. and SK, S. (2026). A modular vision transformer and machine learning framework for scalable multi-crop plant disease classification. *Scientific Reports*. https://doi.org/10.1038/s41598-026-65004-7.

Cho, O.H. (2024). An evaluation of various machine learning approaches for detecting leaf diseases in agriculture. *Legume Research*. **47(4):** 619-627. doi: 10.18805/LRF-787.

Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE. (pp. 1251-1258). https://openaccess.thecvf.com/content_cvpr_2017/html/Chollet_Xception_Deep_Learning_CVPR_2017_paper.html

Cui, Y., Jia, M., Lin, T.Y., Song, Y. and Belongie, S. (2019). Class-Balanced Loss Based on Effective Number of Samples. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE. (pp. 9268-9277). https://openaccess.thecvf.com/content_CVPR_2019/html/Cui_Class-Balanced_Loss_Based_on_Effective_Number_of_ Samples_ CVPR_2019_paper.html.

Ding, X., Zhang, X., Ma, N., Han, J., Ding, G. and Sun, J. (2021). RepVGG: Making VGG-Style ConvNets Great Again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE. (pp. 13733-13742). https://openaccess.thecvf.com/content/CVPR2021/html/Ding_RepVGG_Making_VGG-Style_ConvNets_Great_Again_CVPR_2021_paper.html.

Dyrmann, M., Karstoft, H. and Midtiby, H.S. (2016). Plant species classification using deep convolutional neural network. *Biosystems Engineering*. **151:** 72-80. https://doi.org/10.1016/j.biosystemseng.2016.08.024.

Espejo-Garcia, B., Mylonas, N., Athanasakos, L. and Fountas, S. (2020). Improving weeds identification with a repository of agricultural pre-trained deep neural networks. *Computers and Electronics in Agriculture*. **175:** 105593. https://doi.org/10.1016/j.compag.2020.105593.

Giselsson, T.M., Jørgensen, R.N., Jensen, P.K., Dyrmann, M. and Midtiby, H.S. (2017). A public image database for benchmark of plant seedling classification algorithms. *arXiv*. https://arxiv.org/abs/1711.05458.

Hamuda, E., Glavin, M. and Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and Electronics in Agriculture*. **125:** 184-199. https://doi.org/10.1016/j.compag.2016.04.024.

Hasan, A.M., Sohel, F., Diepeveen, D., Laga, H. and Jones, M.G. (2021). A survey of deep learning techniques for weed detection from images. *Computers and Electronics in Agriculture.* **184:** 106067. https://doi.org/10.1016/j.compag.2021.106067.

Hughes, D. and Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv*. https://arxiv.org/abs/1511.08060.

Islam, R., Imran, S., Ashikuzzaman, M. and Khan, M.M.A. (2020). Detection and classification of brain tumor based on multilevel segmentation with convolutional neural network. *Journal of Biomedical Science and Engineering.* **13(4):** 45-53. https://doi.org/10.4236/jbise.2020.134004.

Kappali, H.R., Sadyojatha, K.M. and Prashanthi, S.K. (2024). Computer vision and machine learning in paddy diseases identification and classification: A review. *Indian Journal of Agricultural Research.* **58(2):** 183-187. doi: 10.18805/IJARe.A-6061.

Karthikeyan, S., Charan, R., Narayanan, S. and Jani Anbarasi, L. (2025). Enhanced plant disease classification with attention-based convolutional neural network using squeeze and excitation mechanism. *Frontiers in Artificial Intelligence.* **8:** 1640549. https://doi.org/10.3389/frai.2025.1640549.

Lu, Y. and Young, S. (2020). A survey of public datasets for computer vision tasks in precision agriculture. *Computers and Electronics in Agriculture.* **178:** 105760. https://doi.org/10.1016/j.compag.2020.105760.

Lu, Y., Young, S., Wang, H. and Wijewardane, N. (2022). Robust plant segmentation of color images based on image contrast optimization. *Computers and Electronics in Agriculture.* **193:** 106711. https://doi.org/10.1016/j.compag.2022.106711.

Mehta, A.R., Kumar, P., Prem, G., Aggarwal, S. and Kumar, R. (2025). AI-powered innovations in agriculture: A systematic review on plant disease detection and classification. *Indian Journal of Agricultural Research.* **59(9):** 1321-1330. doi: 10.18805/IJARe.A-6371.

Meng, Q., Guo, J., Zhang, H., Zhou, Y. and Zhang, X. (2025). A dual-branch model combining convolution and vision transformer for crop disease classification. *Plos One.* **20(4):** e0321753. https://doi.org/10.1371/journal.pone.0321753.

Metagar, S.M. and Walikar, G.A. (2024). Machine learning models for plant disease prediction and detection: A review. *Agricultural Science Digest.* **44(4):** 591-602. doi: 10.18805/ag.D-5893.

Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S. and Batra, N. (2020). PlantDoc: A Dataset for Visual Plant Disease Detection. In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD. Association for Computing Machinery. (pp. 249-253). https://doi.org/10.1145/3371158.3371196.

Srivastava, M. and Meena, J. (2024). Plant leaf disease detection and classification using modified transfer learning models. *Multimedia Tools and Applications*. **83(13):** 38411-38441.

Suh, H.K., Ijsselmuiden, J., Hofstee, J.W. and van Henten, E.J. (2018). Transfer learning for the classification of sugar beet and volunteer potato under field conditions. *Biosystems Engineering.* **174:** 50-65. https://doi.org/10.1016/j.biosystemseng.2018.06.017.