PLANT SCIENCE TODAY
Vol x(x): xx–xx
https://doi.org/10.14719/pst.10613

eISSN 2348-1900

RESEARCH ARTICLE

Legume leaf disease classification via attention-enhanced Swin transformer with feature pyramid fusion (AE-SwinFPF)

Vimal Singh[1,2*], Anuradha Chug[1] & Amit Prakash Singh[1]

[1]University School of Information, Communication & Technology Guru Gobind Singh Indraprastha University Dwarka, Sector-16 C 110 078, New Delhi, India

[2]School of Computing Science & Engineering Galgotias University, Sector 17-A, Greater Noida 203 201, Uttar Pradesh, India

*Correspondence email - 26.vimal@gmail.com

Received: 13 July 2025; Accepted: 18 August 2025; Available online: Version 1.0: 22 September 2025

Cite this article: Vimal S, Anuradha C, Amit PS. Legume leaf disease classification via attention-enhanced Swin transformer with feature pyramid fusion (AE-SwinFPF). Plant Science Today (Early Access). https:/doi.org/10.14719/pst.10613

Abstract

Legume crops are vital to global agriculture due to their high nutritional value, ability to fix atmospheric nitrogen and role in promoting sustainable farming practices. However, these crops are susceptible to various foliar diseases that adversely impact productivity and crop quality. Accurate and early disease identification is essential for effective disease control and yield protection. This study proposed an Attention-Enhanced Swin Transformer integrated with Feature Pyramid Fusion (AE-SwinFPF) to effectively capture multi-scale spatial and semantic features for legume leaf disease classification. For enhanced interpretability, Grad-CAM visualizes the model's focus on disease-relevant regions in the output, providing insights into the decision-making process. The proposed model was evaluated on publicly available legume crop leaf image datasets comprising peas, beans and black gram, achieving classification accuracies of 97.16 %, 98.50 % and 99.99 % respectively. It also consistently yielded high precision, recall and F1-score, demonstrating its reliable and effective performance across all three legume types. Comparative analysis against several baseline Convolutional Neural Network (CNN) models and previously published methods revealed consistent improvements in classification performance. The integration of hierarchical attention with interpretable feature fusion highlights the model's effectiveness and reliability for real-world deployment. However, further validation across diverse crop types and field conditions is recommended to ensure broader applicability.

Keywords: bean leaf; black gram leaf; leaf disease classification; legume crops; pea leaf; swin transformers

Introduction

Legume crops are a cornerstone of global agriculture, playing a vital role in food security, nutrition and sustainable farming systems (1). Major examples include chickpeas, lentils, soybeans, beans, peas, pigeon peas, black grams and cowpeas, which are widely cultivated across diverse agro-climatic regions due to their adaptability and resilience (2). These crops contribute significantly to the human diet by providing a rich source of plant-based protein, dietary fiber, vitamins (e.g. folate) and essential micronutrients such as iron and zinc (3). In many developing countries, legumes are staple foods, often forming the primary source of affordable nutrition for low-income populations (4). Beyond their nutritional value, legumes offer important agroecological benefits. One of the most critical attributes of legumes is their symbiotic relationship with nitrogen -fixing bacteria (Rhizobia), which enables them to convert atmospheric nitrogen into a usable form in the soil (5). This biological nitrogen fixation reduces dependency on synthetic fertilizers, thereby lowering production costs and mitigating environmental impacts such as groundwater contamination and greenhouse gas emissions. Additionally, legumes contribute to crop rotation systems by improving soil structure and health, thereby supporting long-term agricultural productivity (6). Global demand for legumes has increased steadily due to shifting dietary preferences towards plant-based proteins and the recognition of legumes' role in sustainable diets and climate-smart agriculture (7). As the world confronts the dual challenges of increasing population and climate change, legume crops stand out as both nutritionally essential and environmentally beneficial (8).

Despite their value, the productivity of legume crops is often compromised by various biotic stresses, most notably foliar diseases. These diseases, which affect the leaves of the plant, significantly impact photosynthesis, nutrient transport and overall plant vitality (9). Common foliar diseases in legumes include rusts (*Uromyces* spp.), blights (*Xanthomonas axonopodis* pv. *phaseoli*), anthracnose (*Colletotrichum* spp.), powdery mildew (*Erysiphe* spp.) and leaf spots (*Cercospora* spp.). Pathogens responsible for these conditions may be fungal, bacterial or viral in origin and are often favoured by specific climatic conditions such as high humidity, rainfall and temperature extremes (10). Foliar diseases in legumes can lead to substantial yield losses, sometimes up to 60 % in severe cases. For example, rust diseases in chickpeas and common beans can spread rapidly under conducive environmental conditions, affecting large areas within days (11). Similarly, anthracnose and Ascochyta blight have been known to devastate lentil and pea crops, especially in regions lacking timely disease monitoring and intervention systems (12). In smallholder farming contexts where access to plant protection technologies is limited, the economic burden of such diseases can be particularly severe, affecting both food security and household income. The economic impact is not limited to reduced yields (13). Quality degradation, loss of marketable produce and increased dependency on chemical fungicides also contribute to reduced profitability and environmental sustainability. Given the scale and speed at which foliar diseases can spread, early and accurate detection is critical for effective management and disease containment (14).

CNN have long dominated the landscape of plant disease image classification. Models like ResNet, MobileNet and InceptionNet have achieved commendable performance on curated datasets (15). However, CNNs inherently suffer from limited receptive fields, constraining their ability to capture global contextual features in complex field images. This makes them less reliable when symptoms are small, irregular or dispersed, especially under variable environmental conditions such as lighting, occlusion and background clutter (12). Moreover, CNN typically requires careful hyperparameter tuning and its dependence on local feature extraction limits generalization in diverse agroecological zones. The emergence of transformer-based architecture, originally introduced for NLP tasks, has revolutionized computer vision. Vision Transformers (ViT) and their variants model an image as a sequence of patches and use self-attention to capture long-range spatial dependencies across the image (16). While ViT models have demonstrated exceptional accuracy in standard datasets, their demand for large-scale annotated data and heavy computational costs restricts real-world agricultural adoption.

To overcome the limitations of conventional models, the Swin Transformer, a hierarchical Vision Transformer using shifted windows, was developed to enable efficient and scalable feature extraction with strong spatial awareness (17). While it has shown success in medical and industrial imaging, its use in plant disease detection remains limited. To enhance its performance for real-field agricultural datasets, this study proposes a fine-tuned AE-SwinFPF. These enhancements help the model focus more effectively on disease-affected regions and better capture multi-scale features. As a result, the improved architecture demonstrates enhanced accuracy, interpretability and robustness, offering significant potential for real-time, explainable plant disease classification in agricultural settings.

The proposed model was rigorously evaluated on three independent legume leaf image datasets, pea leaf, bean leaf and black gram leaf. It achieved state-of-the-art classification performance, registering validation accuracies of 97.16 % on the Pea Leaf dataset, 98.50 % on the Beans Leaf dataset and a near-perfect accuracy of 99.99 % on the Black Gram Leaf dataset. These outcomes notably surpass the performance of different baseline CNN models and various past studies on the same legume datasets. The incorporation of multi-scale attention fusion significantly enhanced the model's ability to distinguish fine-grained disease patterns, particularly in cases involving subtle lesions and complex background noise. Additionally, the shifted window mechanism of the Swin Transformer enabled more effective aggregation of contextual information across image patches.

Beyond high predictive accuracy, the model offers enhanced interpretability through the integration of Grad-CAM-based attention map visualizations. This explainability feature enables agricultural stakeholders to comprehend and trust the model's decision-making process, thereby facilitating transparency and supporting the model's deployment in real-world agricultural settings.

The primary contributions of this study include the following:

1.Novel architecture

Introduces AE-SwinFPF attention-enhanced Swin Transformer architecture incorporating a Feature Pyramid Fusion module to effectively capture hierarchical multi-scale features.

2.Comprehensive data coverage

The approach utilizes three dedicated legume-leaf image sets-beans, peas and black gram, so the findings generalize across multiple crops.

3.Built-in interpretability

Grad-CAM is embedded in the validation pipeline, producing heat maps that reveal the regions the model relies on when classifying disease symptoms.

4.State-of-the-art accuracy

The system reaches validation accuracies of 97.16 % on pea leaves, 98.50 % on bean leaves and 99.99 % on black-gram leaves, surpassing earlier deep learning baselines reported for the same datasets.

5.Real-world readiness

The model is lightweight, explainable and highly precise, making it practical for on-device use in precision agriculture and continuous disease surveillance applications.

The remainder of this paper is structured as follows: the Literature Survey section reviews existing work relevant to legumes disease classification; Materials and Methods outlines the proposed model architecture and implementation details; Results and Discussion presents the experimental setup, evaluation metrics and analysis of outcomes; finally, Conclusions and Suggestions summarize the key findings, discuss limitations and propose directions for future research.

Literature survey

Recent advancements in computer vision and artificial intelligence have significantly enriched plant disease diagnostics, particularly in legumes such as beans and soybeans. Traditional CNN-based systems, while effective for localized feature extraction, often exhibit limitations in capturing long-range dependencies and contextual understanding, which are crucial for distinguishing visually similar disease symptoms in real-world conditions. Addressing these challenges, research indicates that a developed an explainable architecture by integrating a Pyramid Vision Transformer (PVT) with a custom group context aware depthwise shuffle network (GCADSN) to enhance both classification performance and interpretability (18). The model leverages both global and local feature extraction and incorporates Grad-CAM-based interpretability, demonstrating superior accuracy (97.66 %) on the iBean dataset. This dual-branch architecture marks an evolution toward interpretable and robust agricultural diagnostics. Similarly, research indicates that SoyaTrans, a Transformer-based model designed to capture nuanced differences in soybean leaf diseases. The architecture utilizes hierarchical attention blocks, allowing for fine-grained pattern recognition with high classification accuracy (98.00 %). This emphasis on attention mechanisms not only enhances performance but also introduces transparency through visualization techniques, bridging the gap between model decision -making and user trust. Focusing on edge deployments, research indicates that BeanWatchNet, a lightweight CNN-based system optimized for mobile and embedded devices, enables real-time disease detection without reliance on constant internet connectivity or high-compute infrastructure (20). By combining adversarial training and Out-of-Distribution (OOD) detection, BeanWatchNet achieved 90 % accuracy in three and four class settings. In a separate evaluation, YOLO v8 attained a mAP @ 50 of 87.6 %, indicating its effectiveness for object detection tasks in this context. These findings underscore the feasibility of deploying AI-based plant disease diagnostics in low-resource environments.

In contrast, research indicates that interpretability on the bean leaf dataset by leveraging CLAHE-enhanced images in conjunction with an EfficientNetB0 backbone, using Grad-CAM to highlight disease-relevant regions such as lesions and achieved 98.44 % test accuracy, illustrating how explainable preprocessing can enhance both performance and transparency (21). Beyond disease classification, research showed that extended CNN applications to the identification of 12 bean cultivars using hierarchical classification across species and leaf sides. With classification accuracy ranging from 86.87 % to 95.86 %, the study revealed the role of leaf orientation and inter-class variability in cultivar recognition. Integrating deep learning with mobile technologies, a MobileNetV2-based model was introduced for real-time detection of six mung bean diseases and four pest types. The model, achieving 93.65 % accuracy, was integrated into a mobile app for in-field diagnostics, showcasing how AI tools can be translated into farmer-friendly applications (23). Mobile-optimised CNN architectures were further investigated in which a comparison of MobileNetV1 and V2 for classifying angular leaf spot and bean rust has been depicted. MobileNetV2 demonstrated a test accuracy of 92.97 %, affirming its suitability for real-time classification on constrained hardware. Image processing pipelines were also explored, which indicated a statistical feature-based classifier using texture and colour descriptors extracted through segmentation and GLCM analysis. The model attained 96 % accuracy, serving as a low-compute alternative to deep networks.

SVM-based classification technique was used to identify rust disease in pea plants using microscopic leaf images (26). Their workflow incorporated Gaussian filtering, log transformation, Otsu's method and 2D wavelet transform for feature extraction. The model achieved 89.6 % accuracy on a dataset of 500 samples, effectively separating diseased from healthy leaves. This research highlights the effectiveness of integrating microscopic imaging with machine learning algorithms for precise and early detection of plant diseases in agricultural systems. A reproducible, low-cost spray inoculation method was introduced for uniformly distributing *Erysiphe pisi* spores on pea leaves, validated by a novel uniformity index (27). They complement this with semi-automated, open-source ImageJ analysis to quantify powdery mildew severity. Reproducibility is confirmed via RT-qPCR (Reverse Transcription Quantitative Polymerase Chain Reaction) and image metrics. Their workflow advances precision phenotyping by enabling consistent pathogen application and scalable disease severity assessment in plant pathology research.

The study proposed an IoT-enabled smart agriculture framework for classifying black gram leaf diseases (28). The system utilizes advanced preprocessing, feature extraction through modified LGTP, colour and shape descriptors and classification via a hybrid LSTM–BiGRU model. Achieving an accuracy of 98.69 %, it outperformed traditional models. The model also provides treatment recommendations, making it suitable for real-time precision farming and decision support in crop disease management. Similarly, the study proposed a dual-phase deep learning framework to identify diseases in black gram plant leaves (29). Initially, DeepLabv3+ with a ResNet-18 backbone performs semantic segmentation to extract leaf regions from natural field backgrounds. Subsequently, EfficientNet-B0 is employed for disease classification, attaining a high accuracy of 99.72 %. The model surpassed prior approaches in terms of precision, recall and F1-score. A complementary mobile application was also developed for on-field disease recognition and fertilizer recommendation. The introduction an ensemble-based deep learning model for detecting leaf diseases in black gram crops (30). Their method integrates several CNN architectures with optimized training configurations to enhance classification performance. The system accurately distinguishes key diseases like Yellow Mosaic and Anthracnose, achieving a high accuracy of 99.21 %. The research underscores the effectiveness of ensemble techniques in providing robust and precise disease detection, making it suitable for real-world agricultural scenarios and smart farming systems.

Materials and Methods

This section details the datasets employed, the preprocessing techniques implemented, the hyperparameters adjusted for network fine-tuning, the essential elements of the proposed model, the feature fusion strategy and the performance metrics used to evaluate the model's effectiveness.

Dataset

This study utilizes publicly available leaf image datasets of legume crops, including the Beans Leaf Dataset comprising 1,295 images categorized into three classes:Healthy, Angular Leaf Spot and Bean Rust; the Pea Leaf Dataset containing 1432 images classified into downy mildew, leaf miner, powdery mildew and healthy; and the black gram leaf dataset encompassing five categories: Anthracnose, Leaf Crinkle, Powdery Mildew and Healthy (31-33). Each dataset was independently partitioned into training, validation and testing sets in an 80:10:10 ratio. All images were resized to 224 × 224 pixels to match the model input requirements. To improve model generalization and mitigate overfitting, various data augmentation techniques including random rotations, vertical flips, RandAugment, normalization, MixUp and CutMix, were applied exclusively to the training subsets. The preprocessing and augmentation strategies are summarized in Fig. 1 & 2.

Performance metrics

To evaluate the classification effectiveness of the proposed model, a suite of widely accepted performance metrics was utilized, including accuracy, precision, recall and F1-score as shown in Equations 1-4. These metrics offer a detailed insight

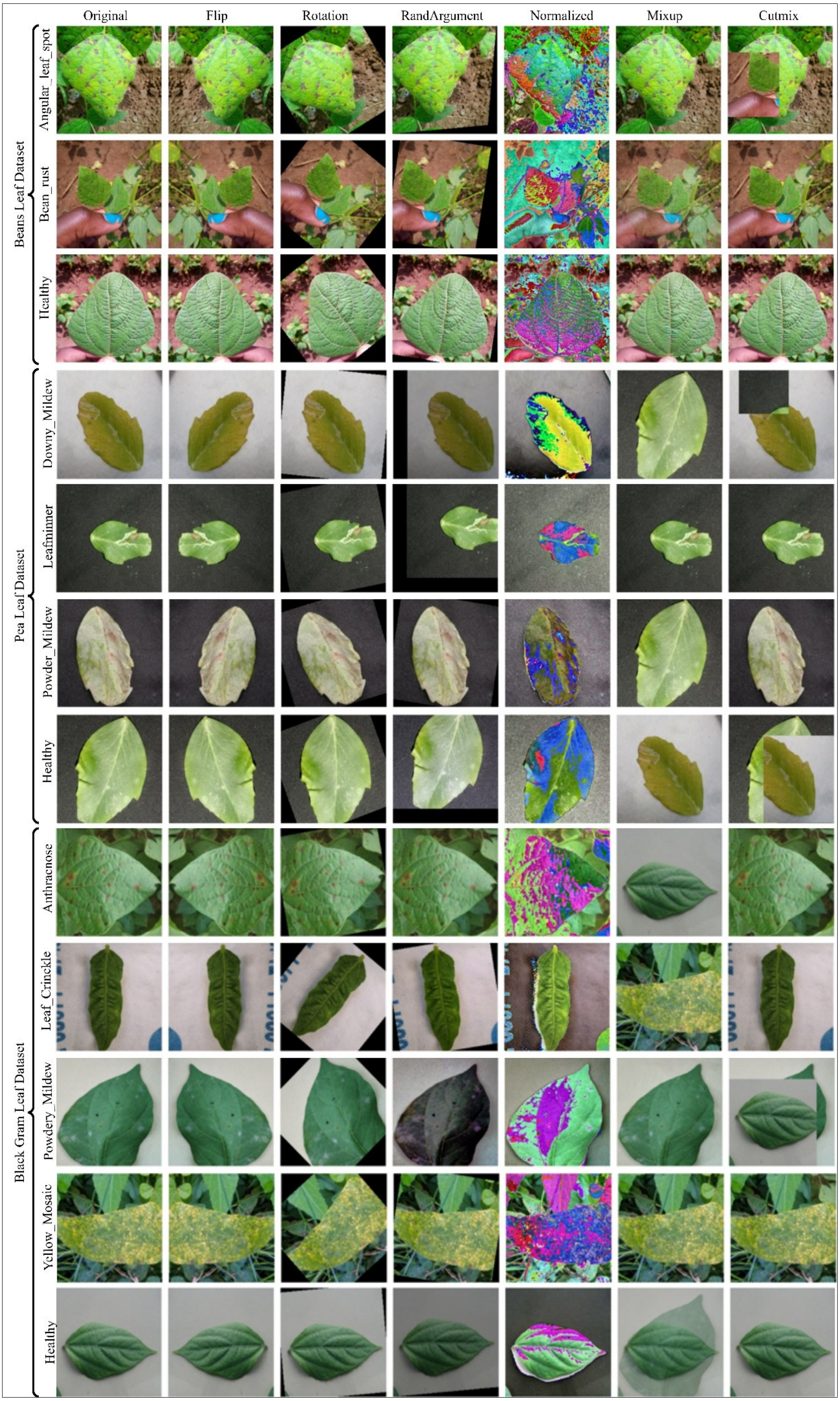


Fig. 1. Legume crop leaf original and augmented images from different classes.

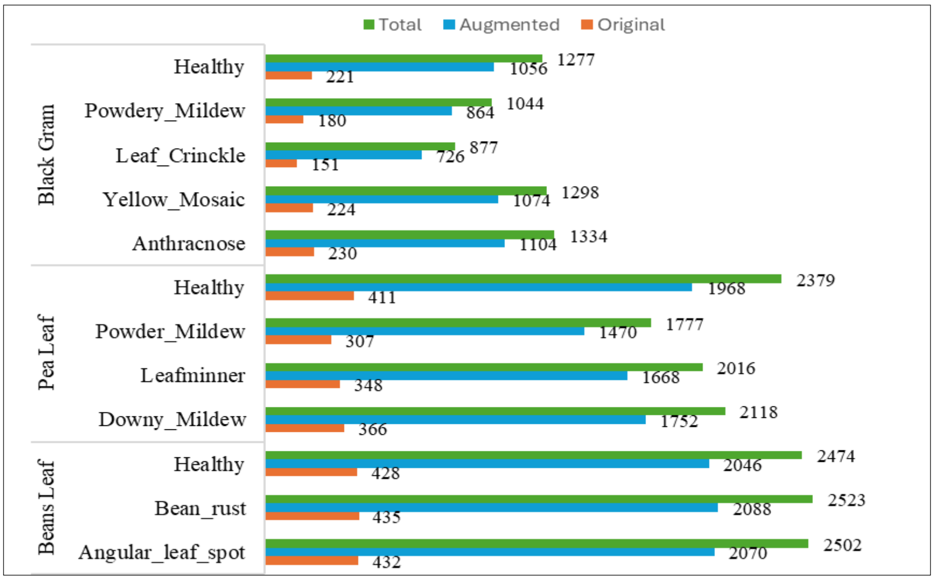


Fig. 2. Legume crop leaf original and augmented images from different classes.

into the model's behaviour across different aspects of prediction. Accuracy provides an overall assessment by calculating the proportion of correctly classified samples relative to the total number of predictions. Precision reflects the model's ability to return relevant positive results by computing the ratio of correctly predicted positives to all instances predicted as positive. This is especially important in domains where minimizing false alarms is critical. Recall, or sensitivity, measures the model's capacity to identify actual positive cases correctly, which is essential when missing positive samples carries higher consequences. The F1-score, as the harmonic mean of precision and recall, offers a balanced evaluation that is particularly useful when class distributions are skewed.

$$\text{Accuracy} = \frac{(T+) + (T-)}{(T+) + (T-) + (F+) + (F-)} \qquad \text{(Eqn. 1)}$$

$$\text{Precision} = \frac{(T+)}{(T+) + (F+)} \qquad \text{(Eqn. 2)}$$

$$\text{Recall} = \frac{(T+)}{(T+) + (F-)} \qquad \text{(Eqn. 3)}$$

$$\text{F1 - score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad \text{(Eqn. 4)}$$

Where, T+, T-, F+ and F- are true positive, true negative, false positive and false negative, respectively.

Implementation setup

The experiments were conducted on Google Colab, a cloud-based platform provided by Google. The environment was configured with an NVIDIA Tesla T4 GPU, which offers 15GB of VRAM and supports efficient computation for deep learning tasks. This GPU is typically available by default in Colab's runtime settings. The setup ensured compatibility with standard deep learning libraries using CUDA and cuDNN pre-installed in the Colab environment.

Methodology

This study presents a unified classification framework to evaluate the performance of the proposed AE-SwinFPF model, as shown in Fig. 3 and three standard CNN Baselines, VGG19, MobileNetV2 and EfficientNetB0 for legume leaf disease identification. All models were trained using transfer learning, with ImageNet-pretrained weights fine-tuned on three publicly

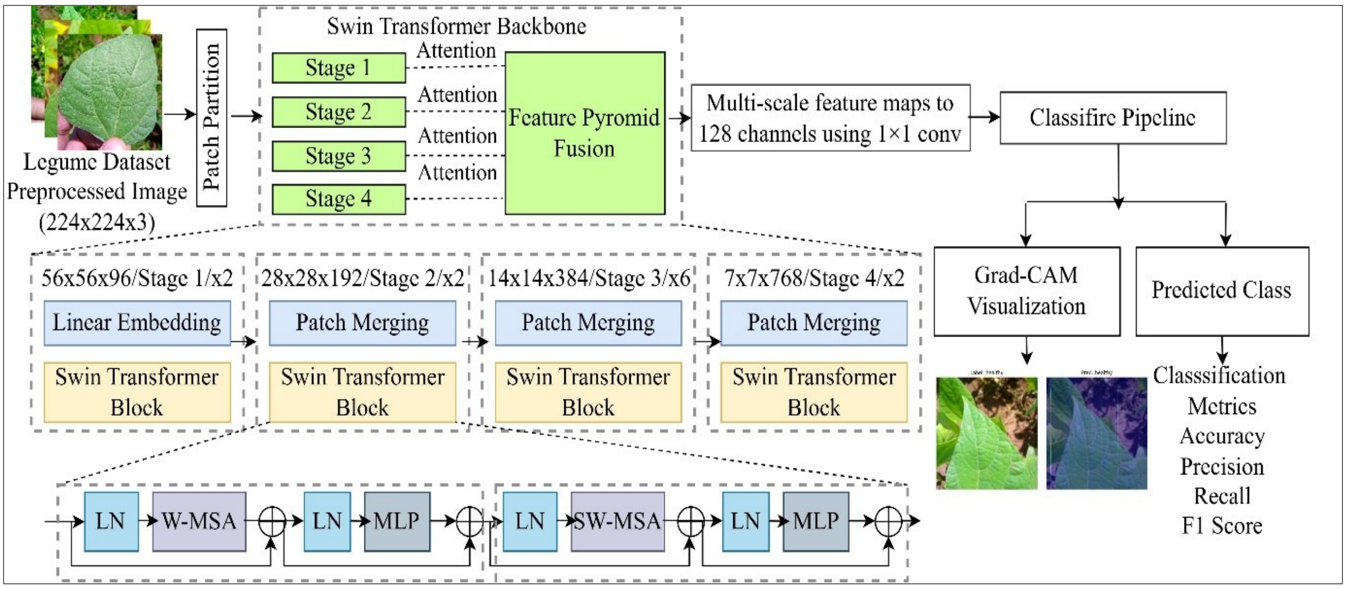


Fig. 3. Proposed AE-SwinFPF model and Grad-CAM for interpretability.

available datasets: Beans Leaf, Peas Leaf and Black Gram Leaf dataset. The AE-SwinFPF architecture leverages an attention-enhanced Swin Transformer integrated with a Feature Pyramid Fusion module to capture hierarchical, multi-scale features critical for the accurate localization of disease symptoms. A consistent preprocessing and augmentation pipeline was applied across models and evaluation was based on standard metrics including accuracy, precision, recall and F1-score. To support interpretability, Grad-CAM visualizations were employed in AE-SwinFPF, highlighting the model's attention focus during prediction.

Backbone architecture: Swin transformer

The model employs a pretrained Swin Transformer as the base feature extractor, which processes the input image hierarchically using non-overlapping shifted windows. The image is initially partitioned into fixed-size patches and passed through four successive Swin Transformer stages, where each stage outputs a feature map, with each output denoted as shown in Equation 5, where H, W and C are height, width and channel of the image, respectively. These features represent hierarchical representations at varying resolutions and depths. Unlike traditional convolutional models, the Swin Transformer preserves long-range dependencies while ensuring local attention, making it highly suitable for subtle pattern discrimination in disease-affected plant regions.

$$F_i \in R^{H_i \times W_i \times C_i}, \text{where } i \in \{1,2,3,4\} \qquad \text{(Eqn. 5)}$$

Feature projection and alignment

Due to the disparity in spatial resolution and channel depth across Swin stages, each output feature map Fi is passed through a 1 × 1 convolutional layer to project its channel dimension to a unified size C. Subsequently, bilinear interpolation is applied to resize each feature map to a common spatial dimension H × W, empirically fixed at 7 × 7. This operation ensures consistency for subsequent fusion and transformed features are defined as shown in Equation 6, where Conv 1 × 1 is the channel projector and Upsample denotes bilinear interpolation.

$$F_i' = \text{UPsample}(\text{Conv } 1 \times 1(F_i), \text{size} = 7 \times 7) \qquad \text{(Eqn. 6)}$$

Feature Pyramid Fusion

To integrate information across different scales, the projected features Fi' from all Swin stages are concatenated along the channel dimension. The resulting tensor is then compressed via another 1 × 1 convolution to generate a semantically rich, spatially consistent representation as shown in Equation 7.

$$F_{fused} = Conv\ 1 \times 1\left(Concat(F_1', F_2', F_3', F_4')\right) \qquad \text{(Eqn. 7)}$$

Classification head

The final fused feature map is flattened and passed through a fully connected layer to produce class logits. The predicted class probabilities are obtained via softmax activation as shown in (Equation 8).

$$\hat{y} = Softmax\left(FC\left(Flattern(F_{fused})\right)\right) \qquad \text{(Eqn. 8)}$$

Hyperparameters

In this study, the proposed model, along with several baseline CNN architectures, was trained using a consistent set of well-tuned hyperparameters to ensure optimal performance as shown in Table 1. A learning rate of 1e-4 was employed to

Table 1. Training hyperparameter settings across all evaluated models

Hyperparameter	Value
Batch Size	32
Epochs	50
Learning Rate	1e-4
Optimizer	Adam
Loss Function	CrossEntropyLoss

maintain stable convergence during training, while a batch size of 32 offered a balance between computational efficiency and generalization. The Adam optimizer was utilized for its adaptive learning capabilities. All models were trained for 50 epochs, with the output layer configured according to the number of disease classes. This configuration contributed to robust convergence and high classification accuracy across legume leaf datasets.

Results

The effectiveness of the proposed model AE-SwinFPF was thoroughly assessed using three legume crop leaf image datasets beans, peas and black gram. Each dataset comprised images representing both healthy and diseased leaf categories. To enhance model generalization and mitigate overfitting, advanced data augmentation techniques, including random rotations, vertical flips, RandAugment, normalization, MixUp and CutMix, were applied. The model's performance was comprehensively measured using standard evaluation metrics such as accuracy, precision, recall and F1-score, providing an in-depth analysis of its classification capabilities across diverse disease classes.

Training progress and convergence behaviour

The training process of the AE-SwinFPF model was carried out using consistent hyperparameter configurations across datasets, which included a learning rate of 1×10^{-4}, batch size of 32 and 50 epochs. Optimization was performed using the Adam optimizer in conjunction with the cross-entropy loss function. The resulting training and validation accuracy and loss curves, as shown in Fig. 4-7, exhibit consistent and stable convergence across all three legume crop leaf datasets. Compared to traditional CNN-based models like VGG19, MobileNetV2 and EfficientNetB0, the AE-SwinFPF model demonstrated improved convergence speed and stability. This highlights the strength of its hierarchical attention mechanism and multi-scale feature learning.

Comparative evaluation with baseline models

A comprehensive performance comparison of the proposed AE-SwinFPF model with conventional CNN architectures is provided in Table 2, which reports accuracy, precision, recall and F1-score

Table 2. Performance comparison of models on legume leaf datasets using Precision, Recall, F1-score and Accuracy

Dataset	Model	Precision	Recall	F1-score	Accuracy
beans leaf	VGG19	0.9632	0.9623	0.9624	0.9624
	MobileNetV2	0.9704	0.9700	0.9700	0.9699
	EfficientNetB0	0.9778	0.9774	0.9775	0.9774
	AE-SwinFPF	0.9858	0.9848	0.9850	0.9850
pea leaf	VGG19	0.9573	0.9579	0.9554	0.9574
	MobileNetV2	0.9654	0.9645	0.9645	0.9645
	EfficientNetB0	0.9680	0.9688	0.9674	0.9687
	AE-SwinFPF	0.9705	0.9714	0.9708	0.9716
black gram leaf	VGG19	0.9830	0.9776	0.9798	0.9800
	MobileNetV2	0.9913	0.9889	0.9898	0.9900
	EfficientNetB0	0.9808	0.9780	0.9788	0.9800
	AE-SwinFPF	0.9999	0.9999	0.9999	0.9999

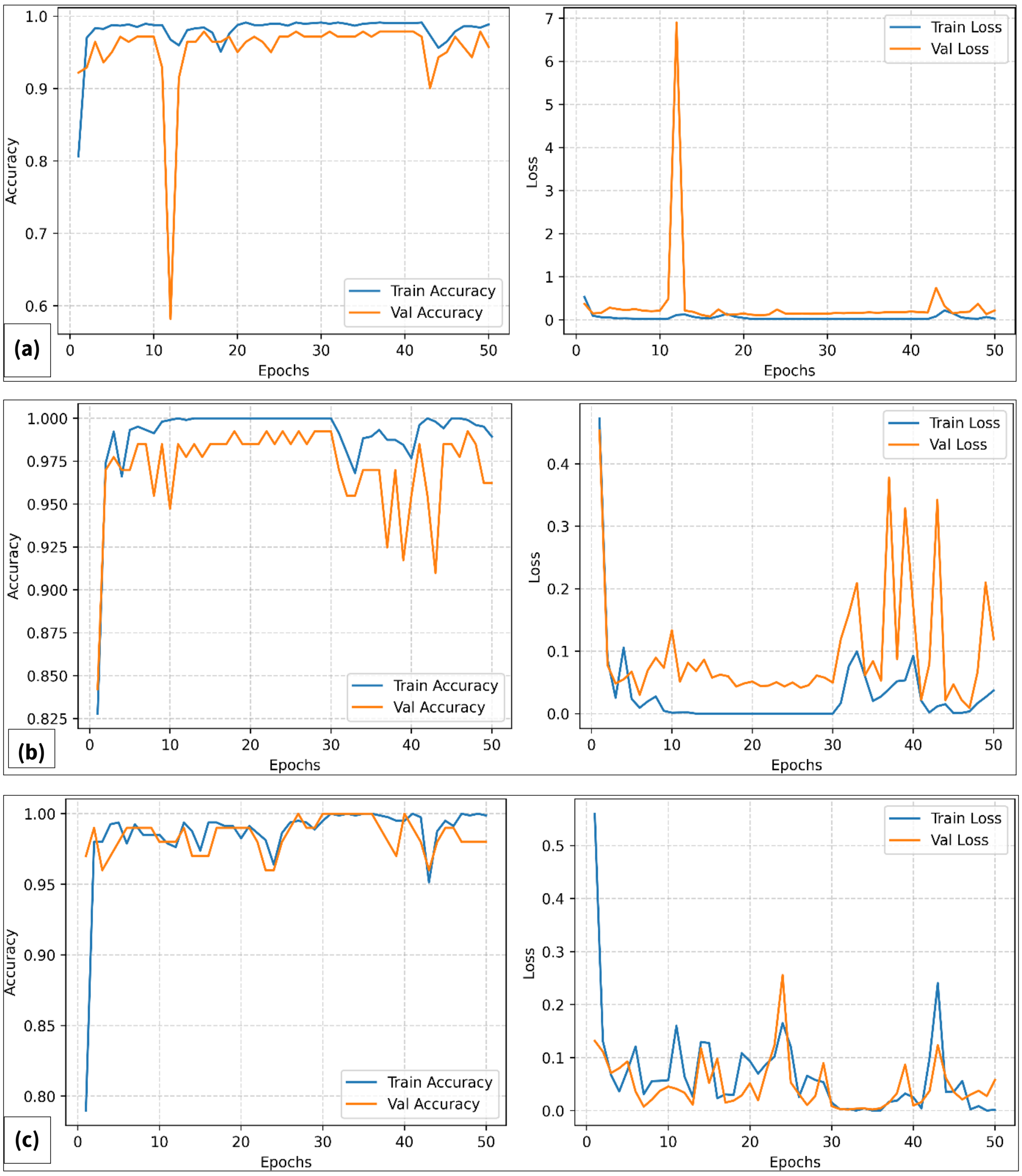
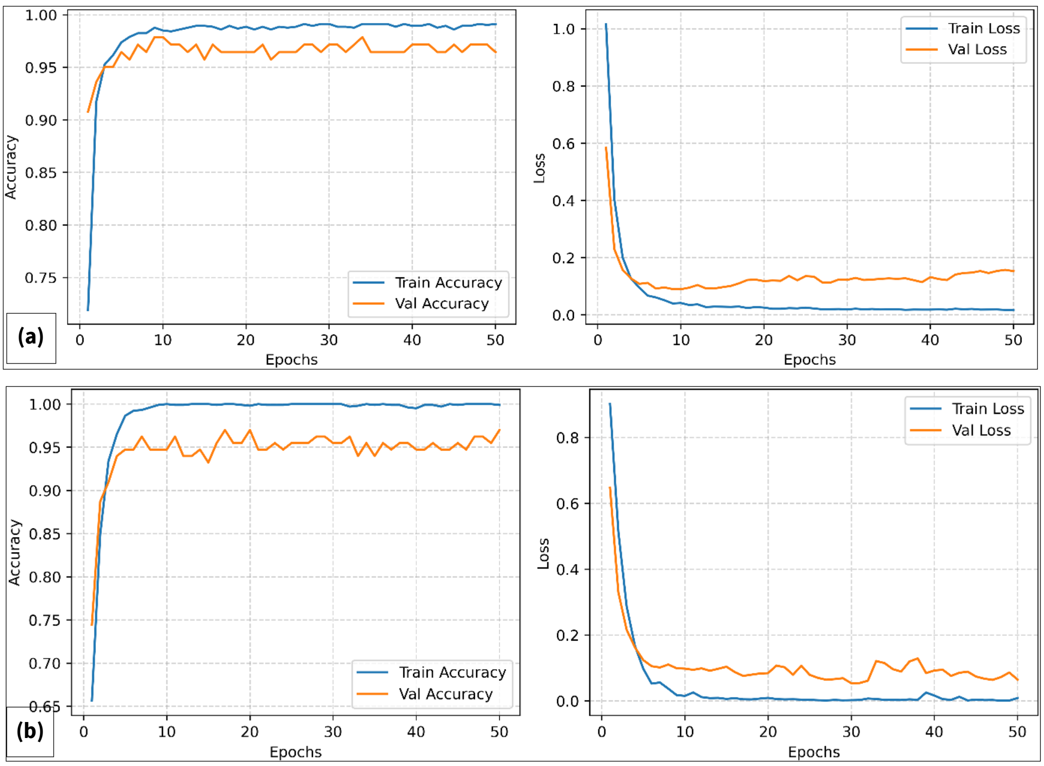


Fig. 4. Evaluation of the VGG19 CNN model using various legume leaf image datasets with accuracy and loss metrics for (a) Peas leaf, (b) Bens leaf and (c) Black gram leaf.



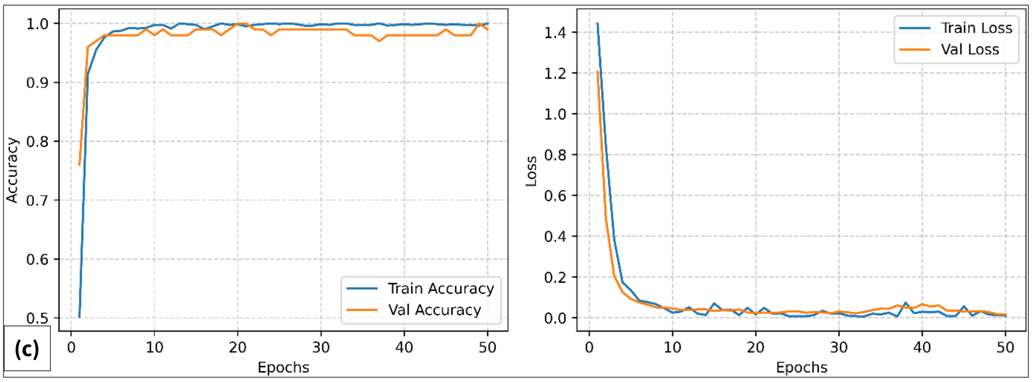


Fig. 5. Evaluation of MobileNetV2 CNN model using various legume leaf image datasets with accuracy and loss metrics for (a) Peas leaf, (b) Bens leaf and (c) Black gram leaf.

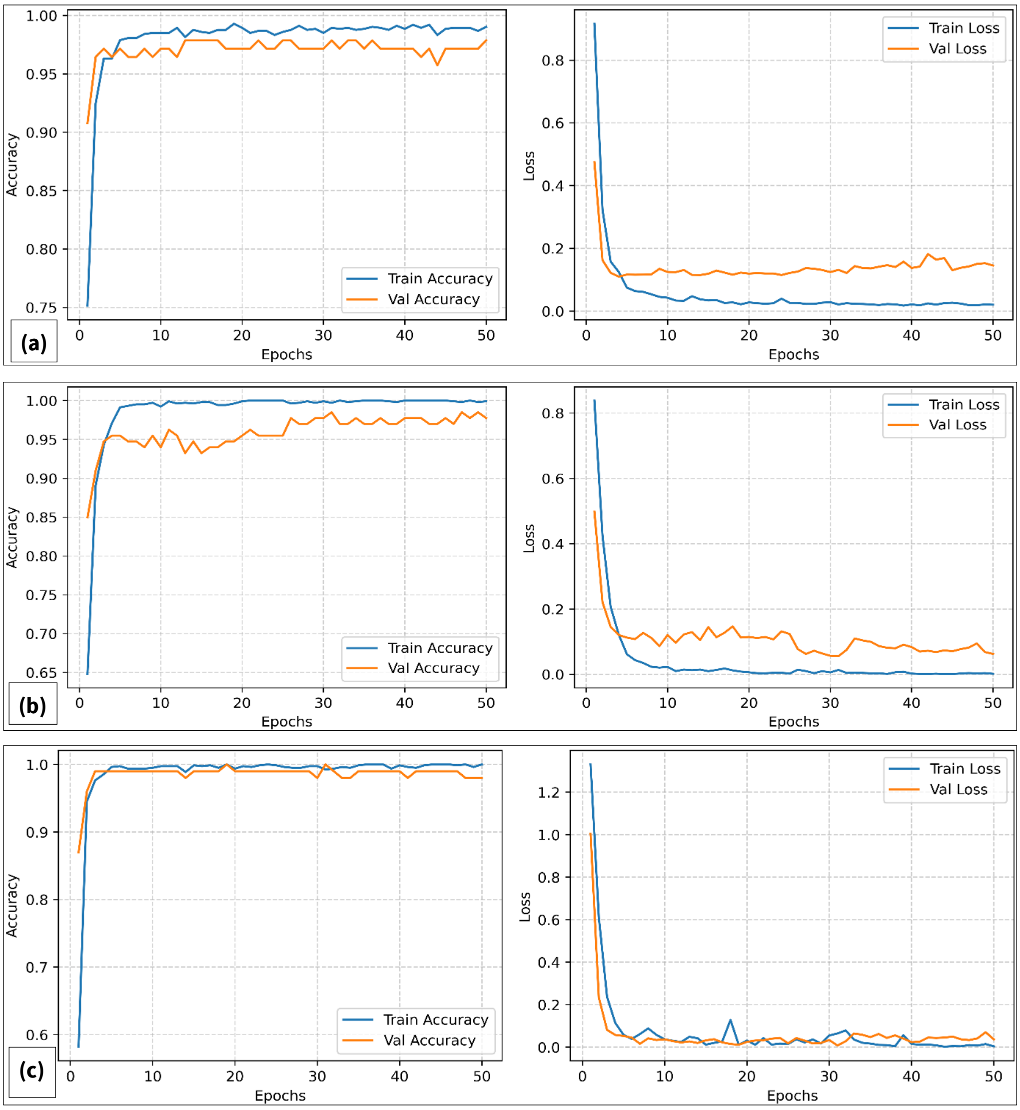
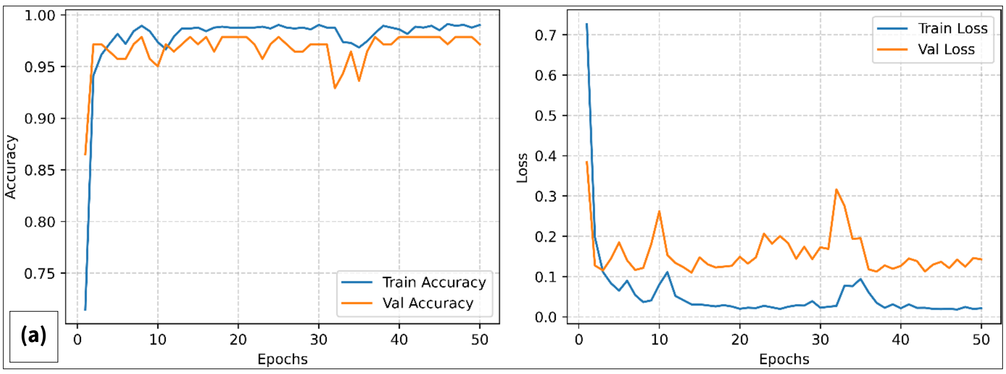


Fig. 6. Evaluation of EfficientNetB0 CNN model using various legume leaf image datasets with accuracy and loss metrics for (a) Peas Leaf, (b) Bens Leaf and (c) Black Gram Leaf.



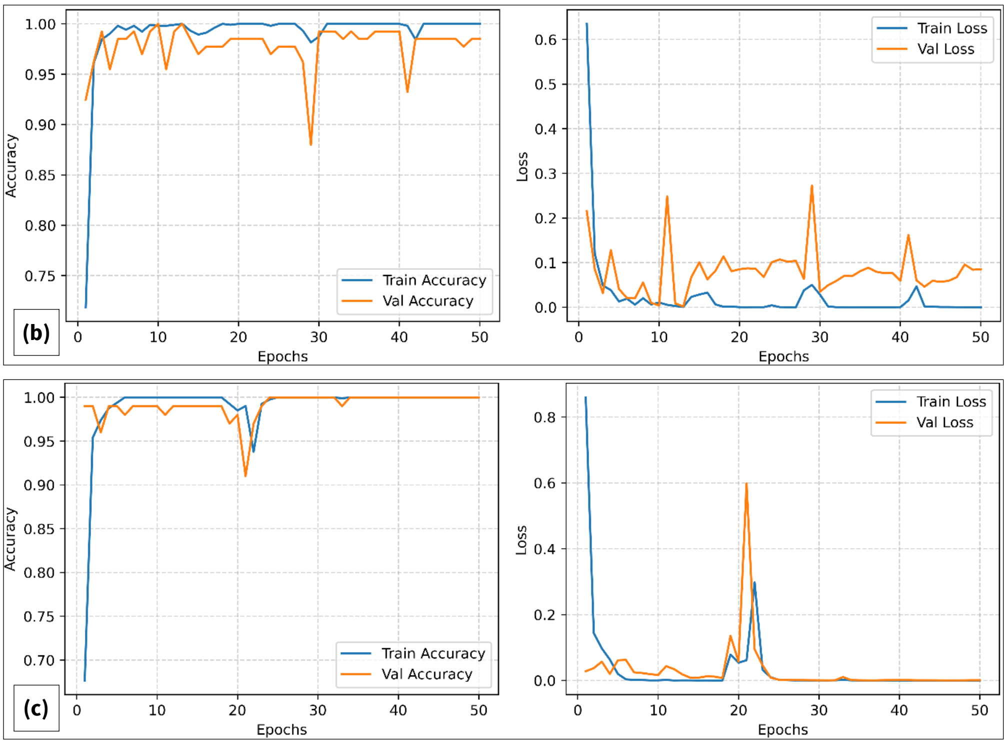


Fig. 7. Evaluation of the proposed model using various legume leaf image datasets with accuracy and loss metrics for (a) Peas Leaf, (b) Bens Leaf and (c) Black Gram Leaf.

across three legume crop leaf datasets. The AE-SwinFPF model achieved superior classification results, attaining 98.50 % accuracy on the Beans Leaf dataset, compared to 97.74 %, 96.99 % and 96.24 % obtained by EfficientNetB0, MobileNetV2 and VGG19, respectively. Similar gains were recorded for the Pea Leaf dataset, where the proposed model reached 97.16 % accuracy, surpassing the best baseline of 96.87 %. On the Black Gram Leaf dataset, AE-SwinFPF nearly achieved perfect classification with 99.99 % accuracy, outperforming the most competitive baseline (EfficientNetB0) at 99.00 %. The detailed class-wise performance, as presented in Table 3, further underscores the model's robustness, with consistently high precision, recall and F1-scores across all categories. These consistent improvements highlight the effectiveness of combining hierarchical attention with feature pyramid fusion, enabling the model to deliver enhanced feature representation and strong generalization capabilities for diverse leaf morphologies and disease conditions in real-world agricultural applications.

Benchmarking against prior studies

As shown in Table 4, the proposed AE-SwinFPF model outperforms prior approaches on all three legume leaf datasets. On the Beans Leaf dataset, earlier models, research indicates that accuracies of 92.00 % and 92.87 %, respectively, while AE-SwinFPF achieved 98.50 %, demonstrating a significant performance gain (24, 34). Similar improvements were observed for the Pea Leaf dataset, with the proposed method attaining 97.16 %, outperforming both traditional and hybrid models such as Federated CNNs and watershed-based techniques. The most significant enhancement was seen on the Black Gram Leaf dataset, where AE-SwinFPF reached 99.99 %, outperforming existing models (29). These results confirm the proposed model as a new benchmark for legume leaf disease classification, offering both high accuracy and cross-dataset generalization through its attention-driven feature fusion framework.

Confusion matrix evaluation

The confusion matrices illustrated in Fig. 8-10 provide detailed insights into the model's performance across individual classes. For the Beans Leaf dataset, the proposed AE-SwinFPF model exhibited highly accurate classification results, correctly identifying 43 out of 44 instances in both the angular leaf spot and healthy categories and achieving a perfect score of 45 out of 45 for Bean Rust. These outcomes underscore the model's strong ability to distinguish between closely related disease symptoms. Similar performance trends were observed for the pea and black gram datasets, where the model maintained high class-wise accuracy with minimal misclassification, demonstrating its robust sensitivity and specificity across diverse crop conditions.

Visual explainability using Grad-CAM

To improve transparency and foster user trust, Grad-CAM-based visual explanations were incorporated into the evaluation pipeline. As illustrated in Fig. 11, the AE-SwinFPF model consistently directed its attention toward disease-specific regions on the leaf surface, accurately highlighting symptomatic areas. These visual cues provide meaningful interpretability by revealing the rationale for the model's classification decisions. Such interpretability is particularly critical for real-world deployment in precision agriculture, as it enables agricultural experts to validate and comprehend the model's focus and decision pathways, thereby supporting its practical utility in field-based plant disease diagnosis.

Discussion

The enhanced classification performance of the proposed AE-SwinFPF model can be attributed to its integrated hierarchical attention mechanism and multi-scale feature fusion strategy. In contrast to conventional CNN-based architectures that often exhibit limitations in capturing subtle disease traits or spatially variant features, AE-SwinFPF excels in modelling both global contextual dependencies and localized disease patterns. This

Table 3. Class-wise precision, recall, F1-score and support on legume leaf datasets

Dataset	Model	Class	Precision	Recall	F1-score	Support
Beans Leaf	VGG19	Angular_leaf_spot	0.98	0.98	0.98	44
		Bean_rust	0.94	0.98	0.96	45
		Healthy	0.98	0.93	0.95	44
	MobileNetV2	Angular_leaf_spot	0.96	0.97	0.98	44
		Bean_rust	1.00	0.96	0.98	45
		Healthy	0.96	0.98	0.97	44
	EfficientNetB0	Angular_leaf_spot	0.98	0.95	0.97	44
		Bean_rust	0.96	0.98	0.97	45
		Healthy	1.00	1.00	1.00	44
	AE-SwinFPF	Angular_leaf_spot	1.00	0.95	0.98	44
		Bean_rust	0.96	1.00	0.98	45
		Healthy	1.00	1.00	1.00	44
Pea Leaf	VGG19	Downy_mildew	1.00	0.86	0.93	36
		Leafminner	0.97	0.97	0.97	34
		Powder_mildew	0.88	1.00	0.94	30
		Healthy	0.98	1.00	0.99	41
	MobileNetV2	Downy_mildew	0.97	0.93	0.94	36
		Leafminner	0.97	0.94	0.96	34
		Powder_mildew	0.97	1.00	0.98	30
		Healthy	0.95	1.00	0.98	41
	EfficientNetB0	Downy_mildew	0.97	0.94	0.96	36
		Leafminner	0.97	0.97	0.97	34
		Powder_mildew	0.97	0.97	0.98	30
		Healthy	1.00	1.00	1.00	41
	AE-SwinFPF	Downy_mildew	0.94	0.94	0.94	36
		Leafminner	0.97	0.94	0.96	34
		Powder_mildew	0.97	1.00	0.98	30
		Healthy	1.00	1.00	1.00	41
Black Gram Leaf	VGG19	Anthracnose	0.96	1.00	0.98	23
		Leaf_crinckle	1.00	0.93	0.97	15
		Powdery_mildew	1.00	1.00	1.00	18
		Yellow_mosaic	1.00	0.95	0.98	22
		Healthy	0.96	1.00	0.98	22
	MobileNetV2	Anthracnose	1.00	1.00	1.00	23
		Leaf_crinckle	1.00	1.00	1.00	15
		Powdery_mildew	1.00	0.94	0.97	18
		Yellow_mosaic	0.96	1.00	0.98	22
		Healthy	1.00	1.00	1.00	22
	EfficientNetB0	Anthracnose	1.00	0.96	0.98	23
		Leaf_crinckle	1.00	0.93	0.97	15
		Powdery_mildew	0.95	1.00	0.97	18
		Yellow_mosaic	1.00	1.00	1.00	22
		Healthy	0.96	1.00	0.98	22
	AE-SwinFPF	Anthracnose	0.99	0.99	0.99	23
		Leaf_crinckle	0.99	0.99	0.99	15
		Powdery_mildew	0.99	0.99	0.99	18
		Yellow_mosaic	0.99	0.99	0.99	22
		Healthy	0.99	0.99	0.99	22

Table 4. Benchmarking the proposed model against existing methods on legume data

Reference	Dataset Used	Approach adopted	Accuracy (%)
(24)	Beans leaf	MobileNet	92.00
(34)		Federated CNNs	92.87
(35)		Hybrid model	96.26
(18)		Combined ViT	97.66
Proposed Model		AE-SwinFPF	98.50
(26)	Pea leaf	Support vector machine	89.60
(27)		Watershed Segmentation	95.00
Proposed Model		AE-SwinFPF	97.16
(36)	Black gram leaf	Inception-V3+ ITL-CHB	98.00
(37)		CNN50+SVM	98.69
(29)		DeepLabv3+MobileNetV2	99.54
(38)		Efficient AttentionNET	99.50
Proposed Model		AE-SwinFPF	99.99

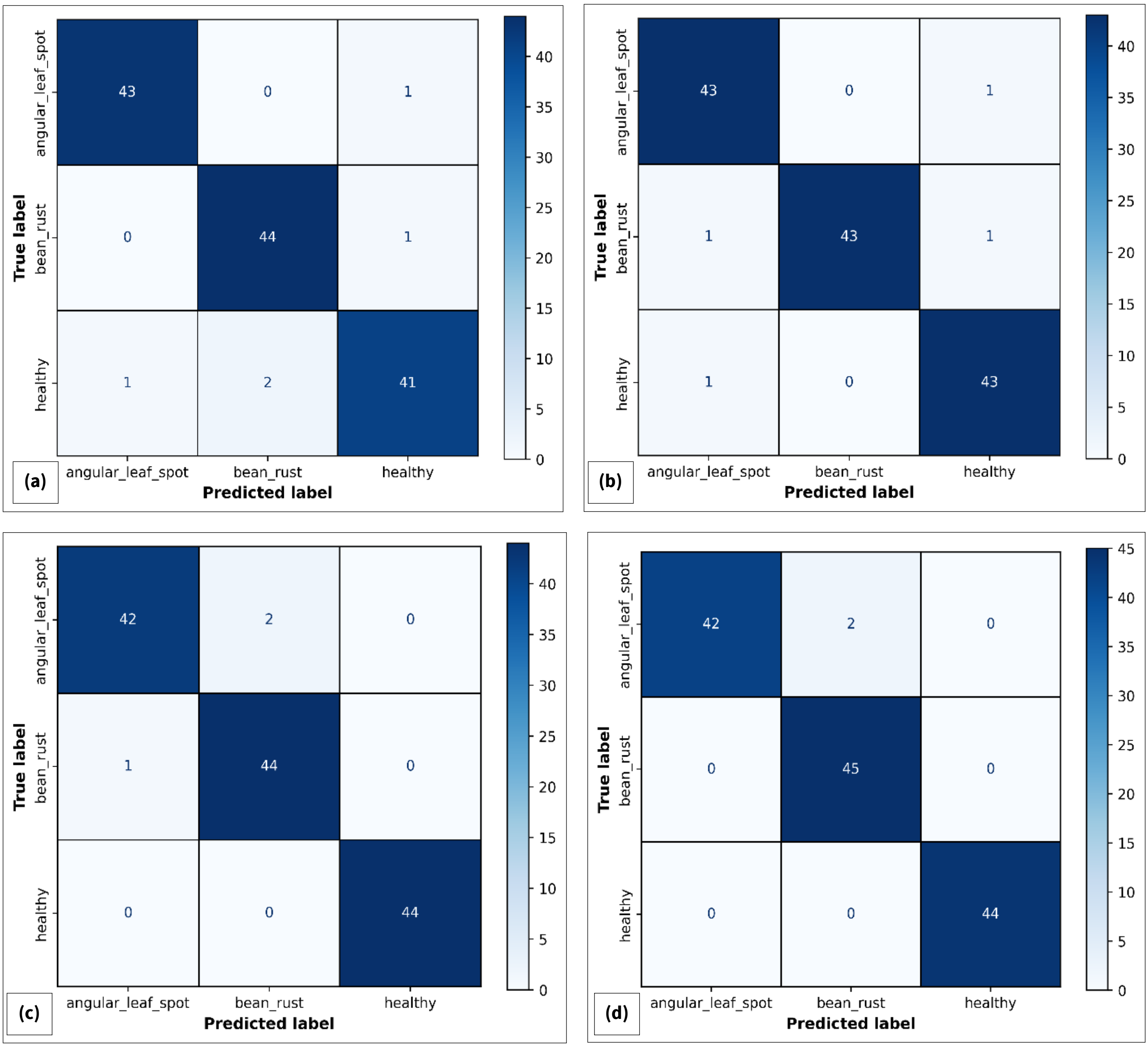
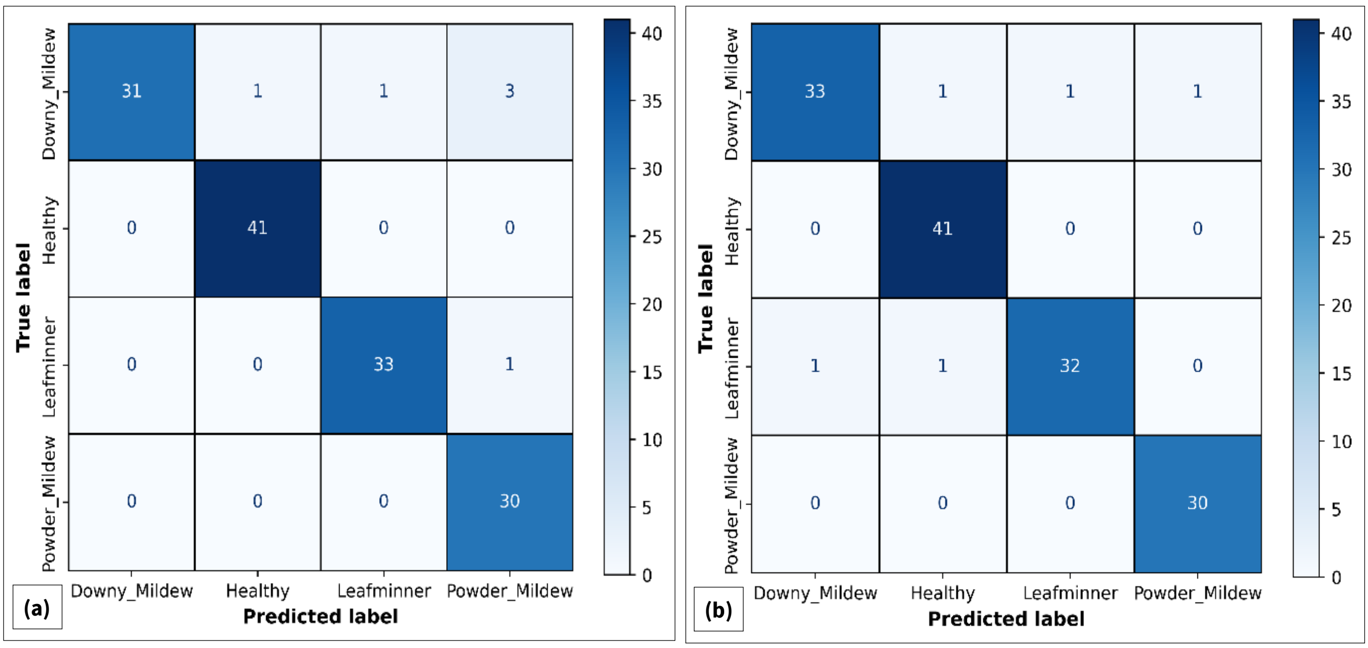


Fig. 8. Confusion matrices for beans leaf classification using (a) VGG19, (b) MobileNetV2, (c) EfficientNetB0 and (d) Proposed Model.



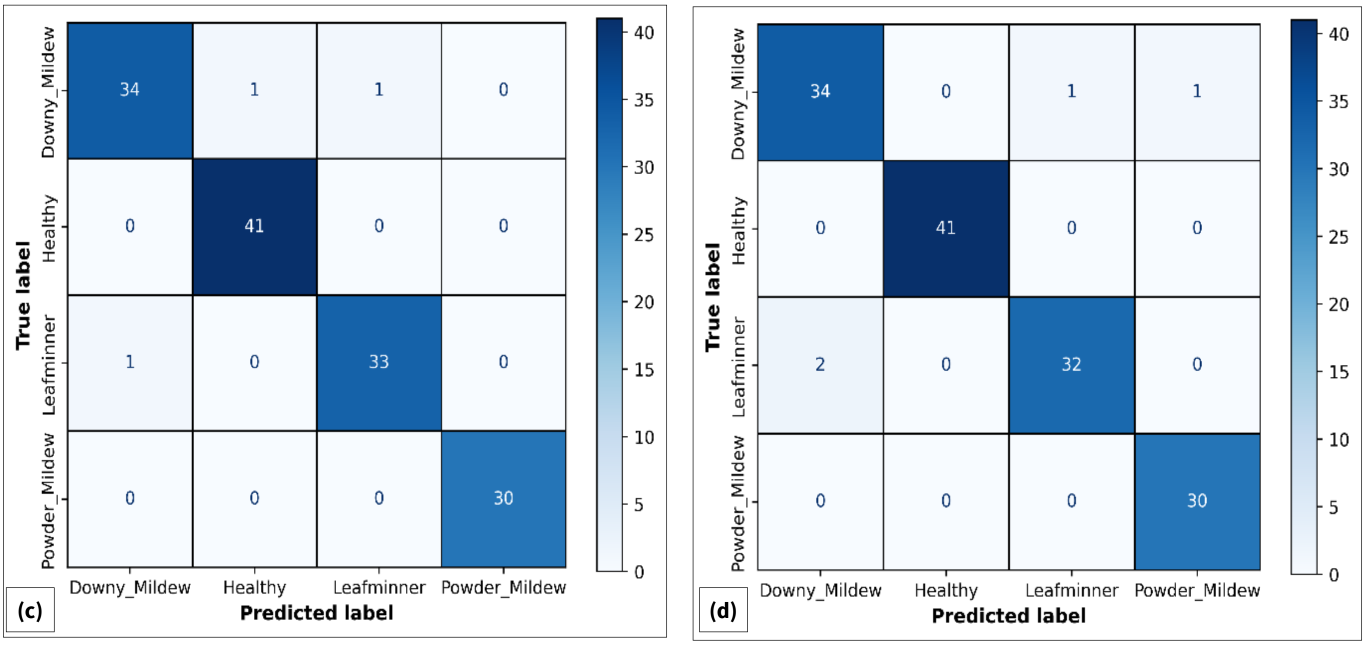


Fig 9. Confusion matrices for Pea leaf classification using (a) VGG19, (b) MobileNetV2, (c) EfficientNetB0 and (d) Proposed Model.

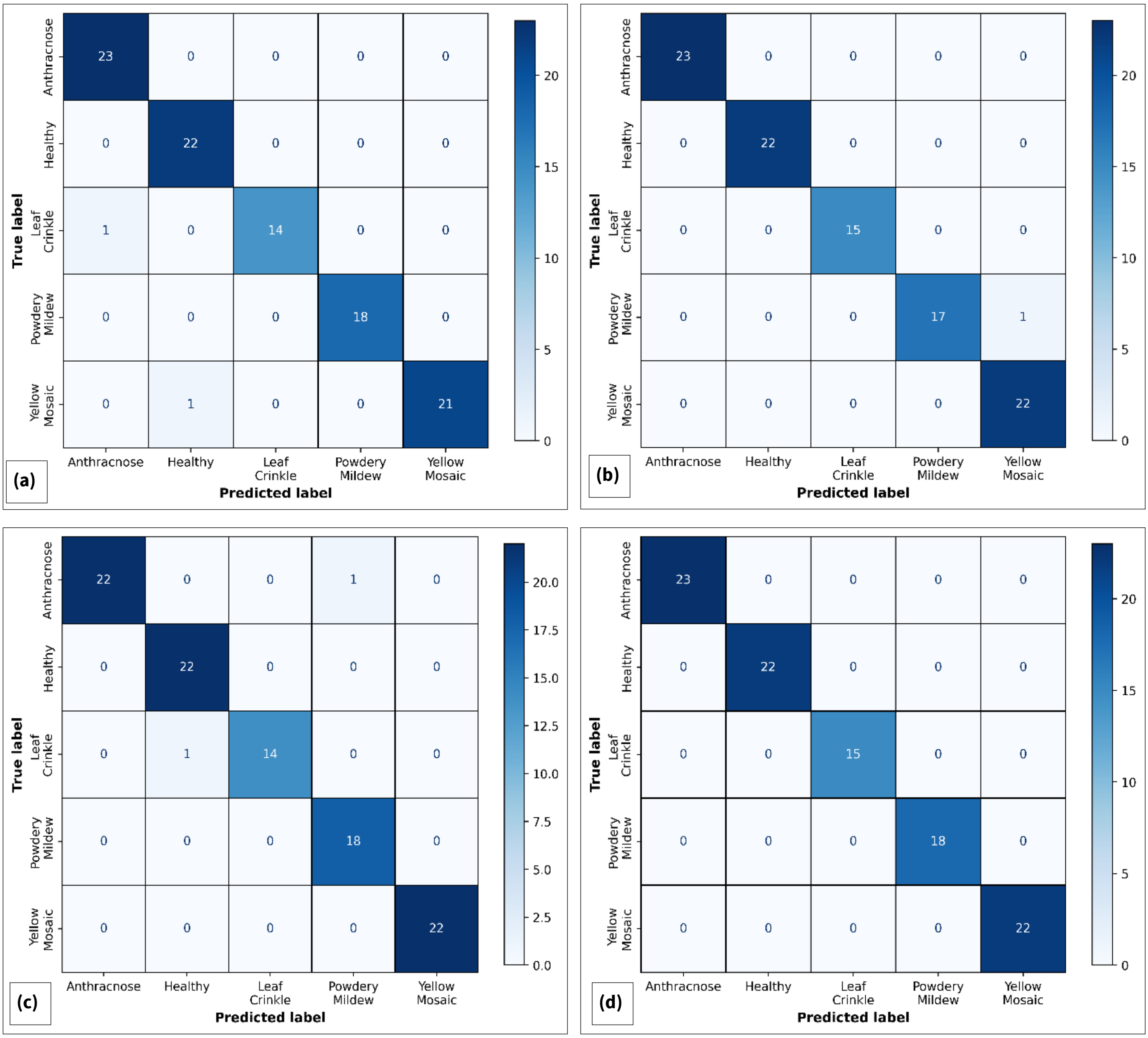


Fig. 10. Confusion matrices for Black Gram Leaf classification using (a) VGG19, (b) MobileNetV2, (c) EfficientNetB0 and (d) Proposed Model.

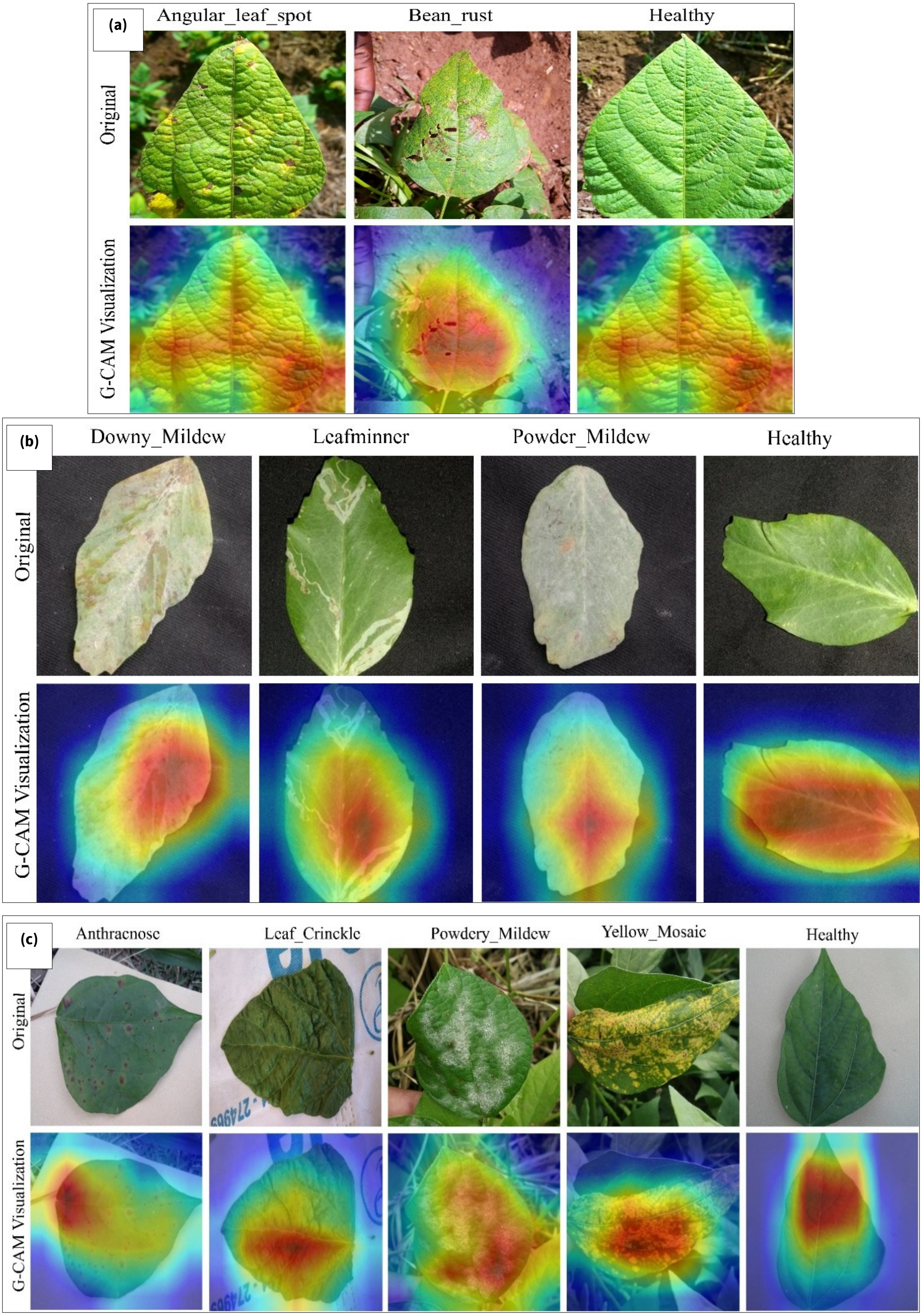


Fig. 11. Grad-CAM attention maps overlaid on original leaf images to demonstrate model interpretability and focus during classification for (a) Beans Leaf, (b) Pea Leaf and (c) Black Gram Leaf.

capability significantly improves its generalization, particularly on datasets characterized by high intra-class similarity and visual ambiguity between classes. Furthermore, the incorporation of Grad-CAM-based visual interpretability not only adds transparency to the decision-making process but also facilitates expert validation in agricultural diagnostics. Coupled with its lightweight and modular design, the model demonstrates strong potential for real-world deployment in precision agriculture settings, where accuracy, explainability and computational efficiency are critical for continuous disease monitoring and early intervention.

Conclusion

This study presents AE-SwinFPF, an interpretable and high-performing deep learning framework for legume leaf disease classification, leveraging an attention-enhanced Swin Transformer combined with Feature Pyramid Fusion. The proposed model effectively captures both spatial and semantic hierarchies, enabling it to distinguish complex disease patterns with high precision. Unlike traditional CNN-based methods, which often struggle with limited receptive fields and loss of fine-grained details, the integration of hierarchical attention mechanisms ensures robust feature representation across multiple scales. Empirical evaluation across three legume datasets, beans, peas and black gram- demonstrated superior classification accuracies of 98.50 %, 97.16 % and 99.99 %, respectively, outperforming conventional CNN models (VGG19, MobileNetV2, EfficientNetB1). Additionally, the inclusion of Grad-CAM-based visualization enhances model interpretability, supporting greater transparency and trust in decision-making processes. Looking ahead, this architecture offers significant potential for extension to multi-label disease classification tasks, particularly relevant for detecting co-occurring infections in real-world environments. Further research may explore deployment optimizations, including model compression techniques such as pruning, quantization and lightweight transformer adaptations for mobile or edge devices. Field-scale validation using diverse and uncurated image datasets will be critical to assess generalizability and operational scalability. Overall, this work lays a strong foundation for the development of explainable, scalable and precision-focused disease detection systems in smart agriculture.

Acknowledgements

This research is supported by the Department of Science and Technology (DST) under project reference number T-319/2018-19. We gratefully acknowledge their generous support and funding for this study.

Authors' contributions

VS contributed to the conceptualization, methodology and study design. AC was involved in the study design, coordination and visualization. APS managed project administration, supervision and funding acquisition. All authors reviewed and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Grammarly to assist with grammar and language enhancement. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

1. Pierre JF. Legumes: cornerstones of global food security and sustainable agriculture. In: Pierre JF, editor, Legume crops - cultivation, uses and benefits. London: IntechOpen; 2024. https://doi.org/10.5772/intechopen.1006180
2. Stagnari F, Maggio A, Galieni A, Pisante M. Multiple benefits of legumes for agriculture sustainability: an overview. Chem Biol Technol Agric. 2017;4(1):2. https://doi.org/10.1186/s40538-016-0085-1
3. Karavidas I, Ntatsi G, Vougeleka V, Karkanis A, Ntanasi T, Saitanis C, et al. Agronomic practices to increase the yield and quality of common bean (*Phaseolus vulgaris* L.): a systematic review. Agronomy. 2022;12(2):271. https://doi.org/10.3390/agronomy12020271
4. Vemulakonda AL, Saad AA, Kumar SS, Wani OA, Singh L, Babu S, et al. Revealing the nexus between fertilizer composition and the performance of common bean (*Phaseolus vulgaris* L.) genotypes in the himalayan heartland of india. Sustainability. 2024;16(14):6234. https://doi.org/10.3390/su16146234
5. Girmaw DW, Muluneh TW. Deep convolutional neural network model for classifying common bean leaf diseases. Discov Artif Intell. 2024;4(1):96. https://doi.org/10.1007/s44163-024-00212-6
6. Magrini MB, Anton M, Chardigny JM, Duc G, Duru M, Jeuffroy MH, et al. Pulses for sustainability: breaking agriculture and food sectors out of lock-in. Front Sustain Food Syst. 2018;2:64. https://doi.org/10.3389/fsufs.2018.00064
7. Nemecek T, von Richthofen JS, Dubois G, Casta P, Charles R, Pahl H. Environmental impacts of introducing grain legumes into European crop rotations. Eur J Agron. 2008;28(3):380–93. https://doi.org/10.1016/j.eja.2007.11.004
8. Gebre GG, Amekawa Y, Fikadu AA, Rahut DB. Farmers' use of climate change adaptation strategies and their impacts on food security in Kenya. Clim Risk Manag. 2023;40:100495. https://doi.org/10.1016/j.crm.2023.100495
9. Hwang SF, Wang H, Gossen BD, Chang KF, Turnbull GD, Howard RJ. Impact of foliar diseases on photosynthesis, protein content and seed yield of alfalfa and efficacy of fungicide application. Eur J Plant Pathol. 2006;115(4):389–99. https://doi.org/10.1007/s10658-006-9027-2
10. Kumari T, Kannan MKJ, Nithya V. A survey on plant leaf disease detection. Int J Comput Appl. 2022;184(17):23–30. https://doi.org/10.5120/ijca2022922170
11. Osuna-Caballero S, Rispail N, Barilli E, Rubiales D. Management and breeding for rust resistance in legumes. J Plant Pathol. 2024;106(1):1–12. https://doi.org/10.1007/s42161-024-01679-z
12. Pei G, Qian X, Zhou B, Liu Z, Wu W. Research on agricultural disease recognition methods based on very large kernel convolutional network-RepLKNet. Sci Rep. 2025;15(1):16843. https://doi.org/10.1038/s41598-025-01553-7
13. Oerke EC. Crop losses to pests. J Agric Sci. 2006;144(1):31–43. https://doi.org/10.1017/S0021859605005708

14. Garrett KA, Dendy SP, Frank EE, Rouse MN, Travers SE. Climate change effects on plant disease: genomes to ecosystems. Annu Rev Phytopathol. 2006;44(1):489–509. https://doi.org/10.1146/annurev.phyto.44.070505.143420
15. Ferentinos KP. Deep learning models for plant disease detection and diagnosis. Comput Electron Agric. 2018;145:311–8. https://doi.org/10.1016/j.compag.2018.01.009
16. Pacal I, Işık G. Utilizing convolutional neural networks and vision transformers for precise corn leaf disease identification. Neural Comput Appl. 2025;37(4):2479–96. https://doi.org/10.1007/s00521-024-10769-Z
17. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin transformer: hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE. 2021. p. 9992–10002. https://doi.org/10.1109/ICCV48922.2021.00986
18. Karthik R, Aswin R, Geetha KS, Suganthi K. An explainable deep learning network with transformer and custom CNN for bean leaf disease classification. IEEE Access. 2025;13:38562–73. https://doi.org/10.1109/access.2025.3546017
19. Sharma V, Tripathi AK, Mittal H, Nkenyereye L. SoyaTrans: A novel transformer model for fine-grained visual classification of soybean leaf disease diagnosis. Expert Syst Appl. 2025;260:125385. https://doi.org/10.1016/j.eswa.2024.125385
20. Katumba A, Okello WS, Murindanyi S, Nakatumba-Nabende J, Bomera M, Mugalu BW, et al. Leveraging edge computing and deep learning for the real-time identification of bean plant pathologies. Smart Agric Technol. 2024;9:100627. https://doi.org/10.1016/j.atech.2024.100627
21. Noviandy TR, Maulana A, Khowarizmi F, Muchtar K. Effect of CLAHE-based enhancement on bean leaf disease classification through explainable AI. In: 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE). IEEE. 2023. p. 544–47. https://doi.org/10.1109/gcce59613.2023.10315394
22. Tavakoli H, Alirezazadeh P, Hedayatipour A, Banijamali Nasib AH, Landwehr N. Leaf image-based classification of some common bean cultivars using discriminative convolutional neural networks. Comput Electron Agric 2021;181:105935. https://doi.org/10.1016/j.compag.2020.105935
23. Mallick MT, Biswas S, Das AK, Saha HN, Chakrabarti A, Deb N. Deep learning based automated disease detection and pest classification in Indian mung bean. Multimed Tools Appl. 2023;82(8):12017–41. https://doi.org/10.1007/s11042-022-13673-7
24. Elfatimi E, Eryigit R, Elfatimi L. Beans leaf diseases classification using MobileNet models. IEEE Access. 2022;10:9471–82. https://doi.org/10.1109/access.2022.3142817
25. Abed S, Esmaeel AA. A novel approach to classify and detect bean diseases based on image processing. In: 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). IEEE. 2018. p. 297–302. https://doi.org/10.1109/iscaie.2018.8405488
26. Singh K, Kumar S, Kaur P. Support vector machine classifier-based detection of fungal rust disease in pea plant (*Pisum sativum*). Int J Inf Technol. 2019;11(3):485–92. https://doi.org/10.1007/s41870-018-0134-Z
27. Ray P, Chandran D. Spray inoculation and image analysis-based quantification of powdery mildew disease severity on pea leaves. MethodsX. 2024;13:102980. https://doi.org/10.1016/j.mex.2024.102980
28. Hajare N, Rajawat AS. IoT based smart agri system: deep classifiers for black gram disease classification with modified feature set. Int J Syst Assur Eng Manag. 2024;15(7):3368–84. https://doi.org/10.1007/s13198-024-02347-2
29. Talasila S, Rawal K, Sethi G. Black gram disease classification using a novel deep convolutional neural network. Multimed Tools Appl. 2023;82(28):44309–33. https://doi.org/10.1007/s11042-023-15220-4
30. Hajare N, Rajawat AS. Black gram disease classification via deep ensemble model with optimal training. Int J Image Graph. 2023;23(8):2550033. https://doi.org/10.1142/S0219467825500330
31. Makerere AI Lab. ibean: Bean disease dataset. 2020. https://github.com/AI-Lab-Makerere/ibean/
32. Zunorain. Pea plant dataset. 2022. https://www.kaggle.com/datasets/zunorain/pea-plant-dataset
33. Andy Ting Zhi Wei. Black gram plant leaf disease. 2023. https://www.kaggle.com/datasets/andytingzhiwei/black-gram-plant-leaf-disease
34. Mehta S, Kukreja V, Gupta A. Collaborative intelligence in AgriTech: federated learning CNN for bean leaf disease classification. In: 2023 World Conference on Communication & Computing (WCONF). IEEE. 2023. p. 1–6. https://doi.org/10.1109/wconf58270.2023.10235221
35. Wang Y, Wang Q, Su Y, Jing B, Feng M. Detection of kidney bean leaf spot disease based on a hybrid deep learning model. Sci Rep. 2025;15(1):11185. https://doi.org/10.1038/s41598-025-93742-7
36. Pravin SS, Surendiran A, Sriraman P, Mariyappan I. Early detection and classification of black gram plant leaf diseases using ITL-CHB method. IETE J Res. 2025;71(2):465–78. https://doi.org/10.1080/03772063.2024.2425775
37. Prasanth K, Kabilamani P, Sangar G, Kaliraj V, Rajasekar V. Enhanced disease recognition and classification in black gram plant leaves using deep learning. In: Geetha R, Dao NN, Khalid S, editors. Advances in artificial intelligence and machine learning in big data processing. AAIMB 2023. Communications in Computer and Information Science, vol 2202. Cham: Springer. 2025. p. 213–24. https://doi.org/10.1007/978-3-031-73065-8_17
38. Sangar G, Rajasekar V. An optimized attention-based deep learning model for black gram leaf disease classification. J Theor Appl Inf Technol. 2025;28(4). https://doi.org/10.5281/zenodo.15118315

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.