Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AMnet: An Explainable Graph Based Inception Framework for Aegle marmelos Leaf Disease Classification

Munira Ferdous
Jahangirnagar University
Mahruf Zaman Utso
Jahangirnagar University
Raiyan Gani
East West University
Bashirul Islam
East West University
Shamim Ripon
dshr@ewubd.edu

East West University

Article

Keywords: Aegle marmelos leaf disease classification, AMnet, InceptionV3, fixed graph convolutional network, deep learning, plant disease diagnosis, explainable AI, precision agriculture

Posted Date: June 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9818594/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on August 11th, 2026.
See the published version at https://doi.org/10.1038/s41598-026-66692-x.

AMnet: An Explainable Graph Based Inception Framework for Aegle marmelos Leaf Disease Classification

Munira Ferdous[1], Mahruf Zaman Utso[1], Raiyan Gani[2], Bashirul Islam[2], Shamim Ripon[2]
[1]*Department of Computer Science & Engineering, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh*
[2]*Department of Computer Science and Engineering, East West University, Aftabnagar, Dhaka, 1212, Bangladesh*

*Corresponding author(s). E-mail(s): dshr@ewubd.edu
Contributing authors: muniraferdous05@gmail.com, mahrufzaman044@gmail.com, raiyan.gani02@gmail.com, basirulislamhira70@gmail.com

ABSTRACT
Aegle marmelos (bael) is a medicinally important tropical crop that remains severely underrepresented in computational plant pathology research. This study proposes AMnet, a deep learning framework integrating an InceptionV3 backbone with a fixed graph convolutional network to capture both within-region disease texture and between-region spatial propagation patterns for automated four-class classification of Aegle marmelos leaf diseases: Cercospora leaf, healthy leaf, leaf curl, and leaf spot. The graph module constructs a 25-node spatial graph directly from CNN feature maps, enabling end-to-end training without external preprocessing. Compared against MobileNetV2, InceptionV3, VGG19, and DenseNet201, AMnet achieved the best overall performance with 98.83% accuracy, 98.84% F1-score, 99.95% PR-AUC, and 98.45% MCC, alongside the fastest inference time of 1.83 ms. Robustness was confirmed through bootstrap confidence interval estimation and four-fold cross-validation. PCA-based clustering analysis with silhouette scoring and Davies–Bouldin indexing demonstrated clear class separability in the learned embeddings. Grad-CAM and LIME visualizations confirmed that predictions were grounded in biologically meaningful leaf regions rather than background artifacts. A Gradio-based prototype further demonstrated practical deployment potential. Although broader field validation remains necessary, AMnet provides an accurate, interpretable, and reproducible framework for diagnosing Aegle marmelos leaf disease.
Keywords: Aegle marmelos leaf disease classification; AMnet, InceptionV3, fixed graph convolutional network, deep learning; plant disease diagnosis, explainable AI, precision agriculture

1 Introduction

Plant diseases remain a major challenge in agricultural production because they reduce crop quality, affect yield stability, and increase management costs. Early identification of leaf diseases is therefore important for timely intervention and sustainable crop protection. In many agricultural settings, disease diagnosis still depends largely on visual inspection by farmers or plant specialists. Although expert-based diagnosis is valuable, it can be time-consuming, subjective, and difficult to apply consistently across large cultivation areas.

Aegle marmelos, commonly known as bael, is an important medicinal and fruit-bearing plant with nutritional, therapeutic, and economic value. Its leaves may be affected by several disease

conditions that cause visible changes in color, texture, venation, shape, and surface morphology. In this study, four leaf categories are considered: Cercospora leaf, healthy leaf, leaf curl, and leaf spot. These categories contain disease symptoms such as chlorosis, discoloration, deformation, and irregular lesion patterns. However, visual similarity among disease classes can make manual and automated recognition difficult, particularly when symptoms appear at early or intermediate stages.

Deep learning has shown strong potential for image-based plant disease detection by learning discriminative visual features directly from leaf images [1,2]. Convolutional neural networks can capture disease-related patterns such as lesion shape, color variation, texture irregularity, and structural deformation. Several CNN architectures, including MobileNetV2, InceptionV3, VGG19, and DenseNet201, have been widely used as baseline models for agricultural image classification. However, conventional CNN models may not always capture spatial relationships among local disease regions effectively. In addition, high accuracy alone is not sufficient for reliable agricultural deployment, because a model may still depend on background artifacts or dataset-specific visual cues rather than biologically meaningful symptoms [3].

Recent research has therefore moved toward more robust and interpretable plant disease recognition systems, including graph-based learning, attention-based feature enhancement, semi-supervised learning, and explainable AI. Graph convolutional networks provide a useful mechanism for modeling relationships among feature regions, while explainability tools such as Grad-CAM and LIME help visualize whether model predictions are based on meaningful image regions [4–6]. Despite these advances, limited attention has been given to Aegle marmelos leaf disease classification using a framework that combines strong supervised performance, feature-space validation, statistical reliability analysis, explainability, and prototype deployment.

To address this gap, this study proposes AMnet, a deep learning framework for automated four-class classification of Aegle marmelos leaf disease. The proposed model is evaluated against MobileNetV2, InceptionV3, VGG19, and DenseNet201. The main contributions of this study are as follows:

- *Graph-augmented spatial disease reasoning:* AMnet integrates a fixed spatial graph convolutional module over a 5×5 regional grid derived from InceptionV3 feature maps. This design captures both within-region disease texture, through the CNN backbone, and between-region spatial propagation patterns, through graph-based message passing. Neither a standalone CNN nor a generic GCN achieves both simultaneously, and to our knowledge this architectural combination has not been previously applied to Aegle marmelos leaf disease classification.
- *First deep learning study on Aegle marmelos using graph-based reasoning*: Aegle marmelos (bael) is a medicinal and economically important plant widely cultivated across South and Southeast Asia, yet it has received almost no attention in the computational plant pathology literature. This study represents, to our knowledge, the first application of graph convolutional reasoning to automated disease classification in this species, addressing a clear gap in precision agriculture research for underrepresented tropical crops.
- *Comprehensive and reproducible evaluation pipeline*: Beyond predictive metrics, the proposed framework is validated through confidence interval estimation, four-fold cross-validation, PR-AUC and MCC reporting, PCA-based feature clustering with silhouette and Davies–Bouldin indices, centroid distance analysis, and explainability through both Grad-CAM and LIME. Most comparable plant disease studies report only accuracy and F1-score. This broader evaluation pipeline is itself a methodological contribution toward more trustworthy and reproducible agricultural AI.

- *Statistical and cross-validation analysis*: Confidence interval estimation and four-fold cross-validation are conducted to assess the reliability and generalizability of model performance across different data partitions.
- *Explainable and deployable diagnosis*: Grad-CAM and LIME are applied to interpret model decisions with respect to biologically meaningful leaf regions, and a Gradio-based prototype is developed to demonstrate practical deployment potential for real-time disease diagnosis support.

It is important to distinguish the novelty of AMnet from prior work that applies CNNs or GCNs independently. Conventional CNN-based plant disease models, including the baseline architectures evaluated here, extract disease features from individual image regions but do not explicitly model spatial dependencies between neighboring leaf areas. Graph convolutional approaches applied elsewhere in agricultural imaging typically construct graphs from superpixels or external structural priors, requiring complex preprocessing pipelines. AMnet avoids this by deriving graph nodes directly from the CNN feature map through adaptive average pooling, producing a lightweight and end-to-end trainable architecture that requires no external graph construction. Furthermore, unlike studies that report accuracy on a single dataset split without statistical validation, this work provides bootstrap confidence intervals, cross-validation stability analysis, and unsupervised feature-space evaluation to confirm that reported performance reflects genuine model capability rather than favorable data partitioning. Together, these characteristics position AMnet as a methodologically complete and domain-specific framework for Aegle marmelos leaf disease diagnosis that advances both the applied and methodological aspects of precision agriculture research.

The remainder of this paper is organized as follows. **Section 2** reviews recent studies on attention-based deep learning, graph convolutional and attention-based learning, semi-supervised learning, and plant leaf disease detection. **Section 3** describes the materials and methods, including the Aegle marmelos dataset, preprocessing steps, baseline model architectures, proposed AMnet framework, unsupervised feature analysis, and explainable AI techniques. **Section 4** presents the experimental results and analysis, covering supervised model performance, computational efficiency, loss behavior, confusion matrices, ROC and precision–recall curves, XAI outcomes, PCA-based clustering analysis, statistical validation, ablation study, cross-validation, prototype application, and comparison with related studies. **Section 5** discusses the key findings, practical implications, limitations, and future research directions. Finally, **Section 6** concludes the paper by summarizing the major contributions and significance of the proposed AMnet framework.

2 Related Works

2.1 Attention-Based Deep Learning Approaches

Attention-based deep learning methods improve classification by emphasizing discriminative features and reducing irrelevant information. Kanumuri et al. [7] proposed a ResNet152V2-CBAM model for pneumonia classification, achieving 93.10% accuracy, while Gao et al. [8] introduced X-Profiler, combining a CNN with Transformer-based attention for biomedical image analysis. Miao et al. [9] developed a CNN–BiGRU–Transformer model for human activity recognition, reporting accuracies of 98.41%, 95.62%, and 96.74% on three benchmark datasets.

In medical imaging, N and U [10] proposed SkinProNet for skin disease classification and segmentation, achieving 94.04% accuracy. Gupta, Kumar, and Tripathi [11] introduced ESADL for diabetic retinopathy grading, while Yousafzai et al. [12] developed C-Swin for lung cancer classification with 96.26% accuracy. Ayesha and Khalidi [13] applied ViT and self-attention with CNNs for blood cell classification, achieving 98% validation and testing accuracy. Kuang [14] further demonstrated the effectiveness of tri-domain attention in music genre classification,

achieving 96.96% accuracy. Overall, these studies confirm the value of attention mechanisms, although computational complexity, benchmark dependency, and limited real-world validation remain key limitations.

2.2 Graph Convolutional and Attention-Based Learning Approaches

Graph-based deep learning approaches are effective for modeling relational, contextual, and spatial–temporal dependencies in structured data. Xu and Zhang [15] proposed GCNAA for digital ink Chinese character recognition, representing sampling points as graph vertices and achieving 98.7% accuracy. Yan et al. [16] introduced RDD-GCN for sarcasm detection using reinforced dependency graphs and denoising mechanisms, reporting F1-scores of 84.97%, 71.11%, and 83.83% across multiple datasets.

Graph convolutional models have also been widely applied in traffic prediction. Yang et al. [17] developed AT-GCN by combining GCNs, attention mechanisms, and GRUs for spatial–temporal traffic modeling. Su, Huang, and Li [18] proposed an adaptive graph convolutional framework with spatial and frequency-domain attention, while Luo, Zhu, Zhang, and Li [19] introduced a dynamic GCN with attention fusion that outperformed 18 baseline methods on four traffic datasets. Yang et al. [20] further proposed TARGCN by integrating node-embedded GCNs, GRUs, and temporal attention for long-term traffic forecasting.

Beyond transportation, Chang, Li, and Duan [21] proposed GANM for fake news detection using graph-structured propagation networks and memory-enhanced global attention. Overall, these studies show that graph convolutional and attention-based models improve relational feature learning and contextual representation. However, their practical use is often limited by computational complexity, dependence on graph quality, scalability issues, and reduced robustness in noisy real-world environments.

2.3 Semi-Supervised Learning Approaches

Semi-supervised learning improves model performance by using both labeled and unlabeled data, reducing dependency on costly manual annotation. Thekumparampil et al. [22] proposed AGNN for graph-structured semi-supervised learning, replacing conventional propagation layers with graph attention mechanisms and outperforming several benchmark methods. Ilsever and Baz [23] introduced a ResNet-50 Mean-Teacher framework for plant disease recognition, achieving up to 5.52% higher accuracy than fully supervised learning under limited labeled data.

In medical imaging, Qiu, Meng, and Li [24] developed a Strong-Teacher Consistency Network for few-shot cardiac MRI segmentation, while Huang et al. [25] proposed SEMI-PLC to improve pseudo-label correction using CutMix augmentation and perturbation-based consistency. Rahmati, Shirani, and Keshavarz-Motamed [26] further applied Vision Transformers and masked autoencoder-based reconstruction for pseudo-label refinement in low-label medical image analysis.

Transformer-based and pseudo-label-driven strategies have also shown promising results. Faysal et al. [27] proposed SSMAE, achieving up to 9.24% improvement on CIFAR-10 using only 10% labeled data. Pham, Nguyen, and Nguyen [28] introduced an ambiguity-aware semi-supervised framework for coffee and banana leaf disease classification, reporting 99.46% and 100.0% precision using 50% labeled data. Islam et al. [29] compared supervised, self-supervised, and semi-supervised learning for lung cancer classification, where SimCLR with ResNet18 achieved 82.3% accuracy and 0.81 AUC. Overall, semi-supervised methods reduce annotation dependency,

but pseudo-label noise, computational cost, dataset diversity, and scalability remain major limitations.

2.4 Plant Leaf Disease Detection Approaches

Plant leaf disease detection has become an important area in precision agriculture because of its role in improving crop productivity and reducing economic losses. Surekha and Rao [30] proposed AdjLeafGNN, combining CNN, ASPP, CSAM, and GNN modules for multi-class leaf disease classification and adjacent-leaf disease spread prediction on the PlantVillage dataset, achieving 98.88% accuracy and 98.71% F1-score. Sundhar et al. [31] introduced a hybrid GAT-GCN framework using superpixel segmentation and edge augmentation, showing strong performance on apple, potato, and sugarcane leaf disease datasets.

Hybrid CNN, Transformer, and attention-based models have also shown strong results in plant disease classification. Aboelenin et al. [32] combined CNN architectures with Vision Transformers for Apple and Corn leaf disease recognition, achieving 99.24% and 98% accuracy, respectively. Singh, Sharma, and Dey [33] proposed CBAM-VGG16 with Grad-CAM, Grad-CAM++, and LRP-based interpretability, reporting up to 98.87% accuracy. D, T. L. N., and Prusty [34] developed ALNet, a lightweight attention-based CNN for grapevine disease classification, achieving up to 100% accuracy with only 0.17 million parameters.

Semi-supervised approaches have also reduced annotation dependency in plant disease detection. Sanyal et al. [35] proposed Pseudo CNN using soft labeling and pseudo-labeling, achieving 96.4% accuracy. Overall, existing studies confirm the effectiveness of graph learning, attention mechanisms, Transformers, and semi-supervised learning for plant disease classification. However, many methods remain limited by controlled benchmark datasets, computational complexity, limited field validation, and weak generalization across diverse crop species and environmental conditions. **Table 1** shows the overall study summaries.

Table 1. Comparative Analysis of Existing Deep Learning, Graph-Based, and Semi-Supervised Approaches

Ref	Dataset	Method	Model	Accuracy	Limitation
[10]	Dermatological Skin Datasets	Skin disease classification and segmentation	EfficientNetV2Small + U2-Net++ + ACRNN	94.04%	Limited dataset diversity and computational overhead
[12]	IQ-OTH/NCCD CT Dataset	CNN-Transformer lung cancer classification	C-Swin	96.26%	Small dataset and overfitting risk
[14]	GTZAN	Music genre classification	CNN–Transformer + Tri-domain Attention	96.96%	Single benchmark dataset dependency
[15]	CASIA, SCUT-COUCH2009, HIT-OR3C	Graph-based handwritten character recognition	GCNAA	98.7%	Sensitivity to graph construction quality

[16]	Sarcasm Detection Datasets	Dependency graph-based sarcasm detection	RDD-GCN	84.97%, 71.11%, 83.83%	Difficulty capturing implicit sarcasm
[17]	PeMS	Traffic flow prediction	AT-GCN	Improved prediction performance	Limited real-time validation
[18]	Real-world Traffic Datasets	Adaptive graph traffic forecasting	Adaptive GCN + Attention	Improved MAE/RMSE	Long-term prediction degradation
[19]	Four Traffic Datasets	Dynamic graph traffic prediction	Dynamic GCN + Attention Fusion	Outperformed 18 baselines	High computational complexity
[21]	Social Media Propagation Datasets	Fake news detection	GANM	Promising detection performance	Sensitivity to noisy interactions
[22]	Citation Network Datasets	Semi-supervised graph learning	AGNN	Outperformed competing methods	Limited scalability analysis
[23]	CIFAR-10, CIFAR-100	Semi-supervised masked autoencoding	SSMAE	+9.24% improvement	Limited real-world validation
[26]	Medical Imaging Datasets	Pseudo-label refinement	ViT + MAE + Pseudo-labeling	Improved low-label performance	Sensitivity to pseudo-label noise
[27]	Three Medical Segmentation Datasets	Semi-supervised segmentation	SEMI-PLC	Improved Dice and 95HD	Increased training complexity
[31]	Apple, Potato, Sugarcane Datasets	Hybrid graph-based leaf classification	GAT-GCN	F1: 0.9818, 0.9743, 0.8799	Lower performance on sugarcane dataset
[33]	Five Plant Disease Datasets	Interpretable attention-enhanced classification	CBAM-VGG16	98.87%	High computational requirements

3 Materials and Methods

Fig. 1 illustrates the overall experimental workflow of the proposed AMnet framework for Aegle marmelos leaf disease classification. The pipeline begins with dataset acquisition, preprocessing, normalization, and augmentation to improve data diversity and model generalization. Multiple supervised baseline models, including MobileNetV2, InceptionV3, VGG19, and DenseNet201, are trained and compared with the

proposed AMnet architecture. AMnet integrates an InceptionV3 backbone with a Fixed-GCN module to capture both deep visual features and spatial relationships among leaf regions. The learned representations are further evaluated using classification metrics, explainable AI techniques, PCA-based clustering analysis, and statistical validation methods. Finally, a Gradio-based prototype application is developed to demonstrate the practical deployment capability of the proposed system for real-time and explainable leaf disease diagnosis.

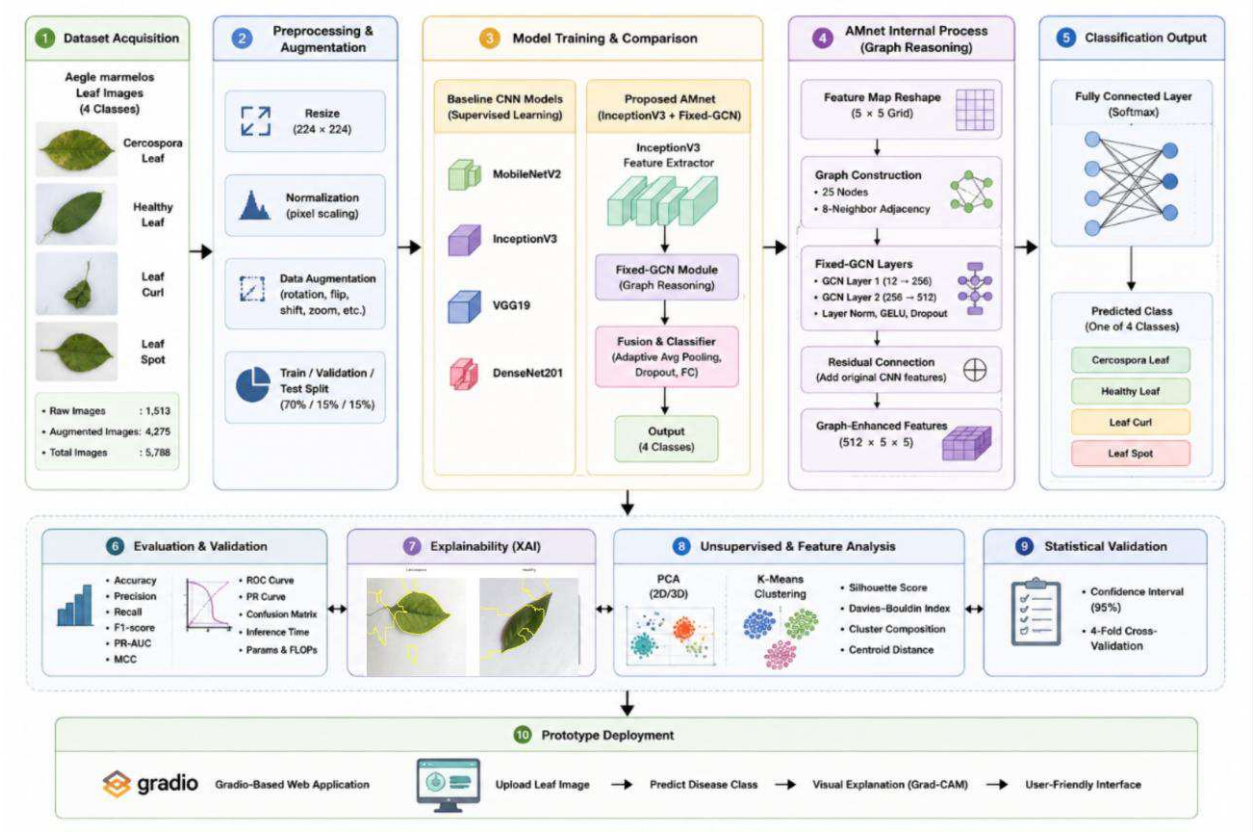


Fig 1. Overall Working Procedure

3.1 Dataset Description

This study uses a publicly available Aegle marmelos leaf image dataset for automated disease classification in the plant pathology domain [36]. The dataset contains four classes: Cercospora, Healthy Leaf, Leaf Curl, and Leaf Spot. Diseased samples show visible symptoms such as discoloration, chlorosis, deformation, and texture variation, while healthy leaves exhibit normal structure and uniform green pigmentation.

The raw images are provided in 700 × 700 JPG format, and the augmented images are resized to 256 × 256 JPG. Data augmentation was applied to increase image diversity and reduce overfitting by introducing variations in orientation, scale, and spatial transformation. To ensure experimental integrity, the train/validation/test split was performed on the raw images prior to any augmentation. Data augmentation was subsequently applied exclusively to the training partition. No augmented derivatives of validation or test images were introduced into the training set at any stage. This protocol ensures that the reported accuracy, precision, recall, F1-score, and PR-AUC values reflect genuine generalization to unseen samples rather than inflated performance arising from augmented duplicates of test images appearing during training. **Fig. 2** shows representative healthy and diseased leaf samples, highlighting the visual differences used for model training and classification.



Cercospora Leaf



Healthy Leaf



Leaf Curl



Leaf Spot

Fig 2. Dataset Sample Images

Table 2 summarizes the class-wise distribution of raw and augmented Aegle marmelos leaf images used in this study. The dataset contains 1,513 raw images, which were expanded to 4,275 augmented images through data augmentation. The four classes are relatively balanced, with raw image counts ranging from 370 images for Leaf Spot to 386 images for Cercospora. A similar balance is maintained after augmentation, where each class contains more than 1,000 images. This balanced distribution helps reduce class bias during model training and supports fair evaluation of classification performance across healthy and diseased leaf categories.

Table 2. Distribution of the Aegle Marmelos leaf disease dataset

Class	Raw Images	Augmented Images
Leaf Spot	370	1046
Leaf Curl	380	1067
Cercospora	386	1110
Healthy	377	1052

Total	**1,513**	**4,275**

3.2 Baseline Models Architecture

This section describes the architectural characteristics of the baseline deep learning models used for comparative evaluation, including MobileNetV2, InceptionV3, VGG19, and DenseNet201.

3.2.1 MobileNetV2 Architecture

MobileNetV2 was used as a lightweight convolutional baseline model for leaf disease classification. The architecture is based on inverted residual blocks, depthwise separable convolutions, and linear bottleneck layers, which reduce computational cost while preserving useful spatial features. The pretrained feature extraction layers were used to extract leaf texture, shape, and color-related patterns. The original classification head was replaced with a task-specific fully connected layer followed by softmax activation for four-class classification.

3.2.2 InceptionV3 Architecture

InceptionV3 was employed as a multi-scale feature extraction model to capture disease patterns at different receptive-field sizes. The architecture uses inception modules with factorized convolutions, allowing the network to learn both local and global visual features efficiently. In this study, the pretrained InceptionV3 backbone was used to extract discriminative leaf representations from the input images. Its final classification layer was modified to produce four output classes corresponding to the target leaf categories.

3.2.3 VGG19 Architecture

VGG19 was selected as a deep convolutional baseline with a simple and uniform architecture. It consists of sequential convolutional layers using small 3 × 3 kernels, followed by max-pooling layers to progressively reduce spatial dimensions while increasing feature depth. This structure enables the model to learn hierarchical visual patterns from low-level edges to high-level disease-specific textures. The original classifier was replaced with a four-class output layer for *Aegle marmelos* leaf disease classification.

3.2.4 DenseNet201 Architecture

DenseNet201 was used as a strong supervised baseline because of its dense connectivity between convolutional layers. In this architecture, each layer receives feature maps from all preceding layers, which encourages feature reuse and improves gradient flow during training. This design helps the model capture fine-grained disease symptoms such as spots, curling, discoloration, and texture variation. The pretrained DenseNet201 classifier was adapted by replacing the final layer with a four-neuron softmax classifier for the target classes.

3.3 AMnet Model Architecture

AMnet adopts an InceptionV3-based fixed graph convolutional architecture for Aegle marmelos leaf disease classification (shown in **Fig. 3**). The input image is resized to 224 × 224 × 3 and passed through the pretrained InceptionV3 backbone up to the Mixed_7c block to extract high-level convolutional feature maps. A 1 × 1 convolutional projection then reduces the channel dimension from 2048 to 512, followed by batch normalization and GELU activation. The reduced feature map is adaptively pooled into a 5 × 5 spatial grid, producing 25 graph nodes, where each node represents a local CNN-derived leaf feature region.

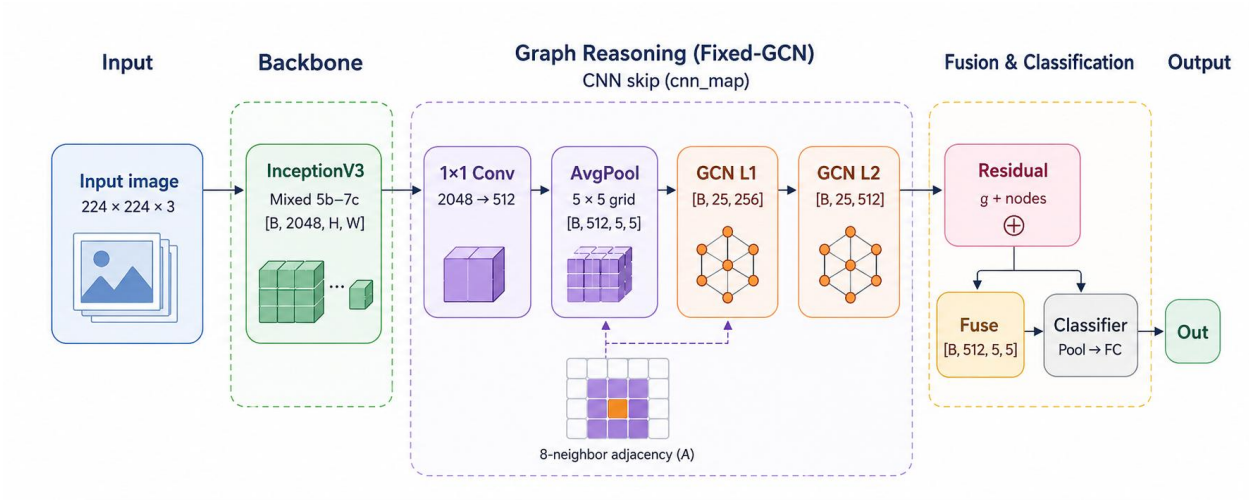


Fig 3. AMnet Architecture

To model spatial relationships among neighboring leaf regions, AMnet constructs a fixed 8-neighbor adjacency matrix with self-connections. This adjacency matrix connects each node to its surrounding spatial neighbors and applies symmetric degree normalization for stable graph propagation. The graph node features are then processed through two fixed-GCN layers, where the first layer maps node features from 512 to 256 dimensions, and the second layer restores them from 256 to 512 dimensions. Each GCN layer applies adjacency-based message passing, linear transformation, layer normalization, GELU activation, and dropout regularization. A residual connection adds the original CNN node features to the GCN-enhanced features, preserving both local convolutional information and graph-refined spatial context.

After graph reasoning, the node features are reshaped back into a 512 × 5 × 5 spatial feature map. This graph-enhanced map is concatenated with the original CNN feature map and passed through a fusion block consisting of 1 × 1 convolution, batch normalization, and GELU activation. Finally, adaptive average pooling, dropout, and a fully connected layer generate the class logits for the four leaf categories. In the final optimized AMnet variant, a higher dropout setting is applied to improve generalization and reduce overfitting during classification.

Algorithm 1 summarizes the complete computational workflow of AMnet, beginning with input image preprocessing and InceptionV3-based feature extraction, followed by channel reduction, fixed 8-neighbor graph construction, two-layer GCN reasoning, CNN-GCN feature fusion, and final classification. The algorithm shows how the model transforms a leaf image into spatial graph nodes and applies graph-based message passing to capture local relationships among neighboring leaf regions. This design allows AMnet to combine the strong visual representation capability of InceptionV3 with structured spatial reasoning from the Fixed-GCN module. As a result, the model can learn both disease-specific visual patterns and region-level dependencies, supporting more reliable *Aegle marmelos* leaf disease classification.

Algorithm 1 AMnet: InceptionV3 with Fixed-GCN Reasoning
Input: Leaf image **Output:** Predicted class label $\hat{y}$ and class probabilities

Input: Leaf image $I \in R^{224\times224\times3}$
Output: Predicted class label $\hat{y}$.

1. Receive the input leaf image I.
2. Resize and normalize the image for InceptionV3 input.
3. Extract deep convolutional feature maps using the pretrained InceptionV3 backbone up to the Mixed_7c block.
4. Apply a 1 × 1 convolution to reduce the feature dimension from 2048 to 512.
5. Apply adaptive average pooling to generate a 5×5 spatial feature grid.
6. Convert the 5 × 5 feature grid into 25 graph nodes.
7. Construct a fixed 8-neighbor adjacency matrix with self-loops for the graph nodes.
8. Normalize the adjacency matrix using symmetric degree normalization.
9. Pass the node features through the first Fixed-GCN layer to obtain 256-dimensional graph features.
10. Pass the output through the second Fixed-GCN layer to restore 512-dimensional graph features.
11. Add the original CNN node features to the GCN output using a residual connection.
12. Reshape the graph-enhanced node features back into a spatial feature map.
13. Concatenate the original CNN feature map and the graph-enhanced feature map.
14. Apply the fusion block using 1×1 convolution, batch normalization, and GELU activation.
15. Apply adaptive average pooling and dropout.
16. Pass the fused feature representation through the fully connected classifier.
17. Generate the final predicted class label $\hat{y}$

Return: Predicted class label $\hat{y}$ and class probabilities.

3.4 Unsupervised Learning-Based Feature Analysis

This study employs unsupervised learning to examine the separability of AMnet feature embeddings without relying on class labels during clustering. The analysis evaluates whether the learned deep features naturally form meaningful groups corresponding to the four leaf classes.

The elbow method identifies the suitable number of clusters by measuring the change in inertia across different K values. The silhouette score assesses cluster compactness and separation, while the Davies–Bouldin index provides an additional measure of clustering quality, where lower values indicate better separation. After determining the appropriate cluster number, K-means clustering groups the AMnet feature embeddings into distinct clusters. Finally, PCA reduces the high-dimensional embeddings into a two-dimensional space to visualize class-wise distribution, overlap, and separation.

3.5 Explainable AI (XAI)

This study incorporates explainable AI to interpret the prediction behavior of AMnet and assess whether the model focuses on disease-relevant leaf regions. Grad-CAM generates class-discriminative heatmaps by highlighting important image regions that contribute strongly to the final prediction, such as lesions, discoloration, vein deformation, and texture changes.

LIME provides a complementary local explanation by perturbing image segments and identifying the regions that most influence the model's decision. Together, Grad-CAM and LIME enhance

model transparency by showing whether AMnet bases its predictions on meaningful leaf symptoms rather than background or irrelevant visual patterns.

4 Results and Analysis

4.1 Performance Analysis of Supervised Models

As shown in **Table 3**, AMnet achieved the best overall supervised classification performance, with the highest accuracy (98.83%), precision (98.86%), recall (98.85%), F1-score (98.84%), PR-AUC (99.95%), and MCC (98.45%). VGG19 was the second-best model, showing very competitive performance with 98.13% accuracy and 97.51% MCC. DenseNet201 and MobileNetV2 also produced acceptable results, whereas InceptionV3 showed the weakest performance with only 68.22% accuracy and 59.14% MCC. Overall, AMnet demonstrated the most reliable and balanced classification ability among the evaluated supervised models.

Table 3. Performance comparison of the models

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	PR AUC (%)	MCC (%)
MobileNetV2	92.06	92.05	92.13	92.01	97.19	89.46
InceptionV3	68.22	75.46	67.56	68.88	80.96	59.14
VGG19	98.13	98.14	98.17	98.15	99.84	97.51
DenseNet201	96.26	96.35	96.21	96.23	99.15	95.04
AMnet	**98.83**	**98.86**	**98.85**	**98.84**	**99.95**	**98.45**

4.2 Computation Analysis of Supervised Models

All inference time measurements were conducted on a GPU-accelerated computing environment using PyTorch, with a fixed batch size of one image to reflect single-sample deployment conditions. Each measurement was averaged over 100 forward passes following 10 warmup iterations to eliminate initialization overhead and ensure stable timing estimates. FLOPs were computed using a single input of size 224 × 224 × 3. Exact hardware specifications are available from the corresponding author upon request. **Table 4** illustrates that AMnet achieved the fastest inference time (1.83 ms) and highest processing speed (547.89 FPS), indicating strong real-time deployment potential despite having the largest model size (90.13 MB) and parameter count (23.63M). MobileNetV2 was the lightest model, with only 2.27M parameters and 0.613 GFLOPs, but its predictive performance was lower than that of AMnet. VGG19 required the highest computational cost (39.038 GFLOPs) and showed the slowest inference speed (8.99 ms), making it less suitable for efficient deployment. Overall, AMnet provided the best trade-off between classification performance and inference efficiency among the evaluated models.

Table 4. Parameter summary of the models

Model Name	Inference Time (ms)	Frames/sec (approx.)	FLOPs (G)	No. of Params	Size (MB)

MobileNetV2	4.39	227.74	0.613	2, 273, 358	9.08
InceptionV3	5.06	197.55	5.693	21, 827, 374	84.05
VGG19	8.99	111.19	39.038	20, 030, 542	76.48
DenseNet201	7.64	130.84	8.632	18, 345, 038	71.72
AMnet	**1.83**	**547.89**	**5.764**	**23,626,212**	**90.13**

4.3 Loss Curve Analysis of Supervised Models

Fig. 4 illustrates that VGG19 and DenseNet201 achieved smoother convergence than MobileNetV2 and InceptionV3, with steadily decreasing training and validation losses. MobileNetV2 showed gradual improvement but retained a noticeable gap between training and validation loss, while InceptionV3 exhibited unstable validation behavior, indicating weaker convergence. **Fig. 5** shows that AMnet converged rapidly within a few epochs and maintained a low validation loss after convergence. Overall, the loss-curve analysis supports the superior learning stability of AMnet compared with the baseline models.

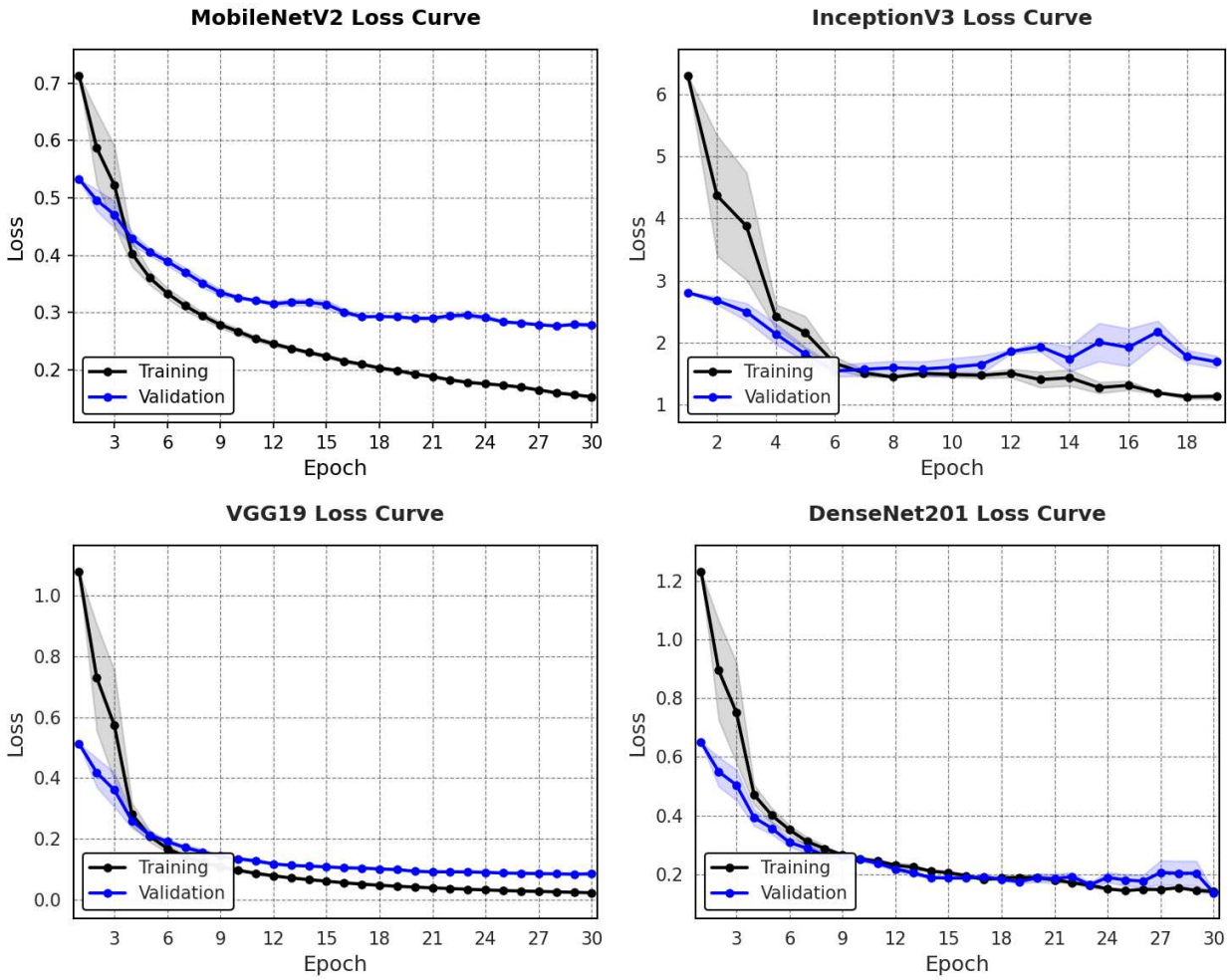


Fig 4. Baseline Models Loss Curves

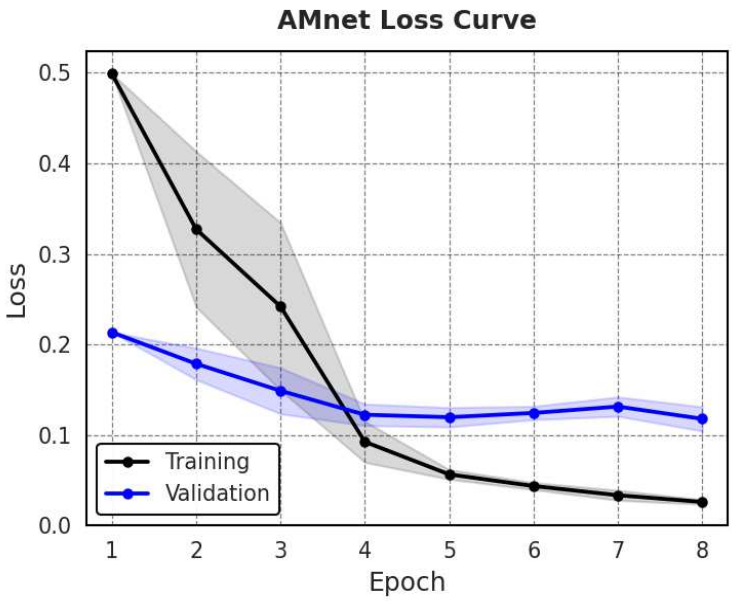


Fig 5. AMnet Loss Curves

4.4 Confusion Matrix Analysis of Supervised Models

Fig. 6 presents the baseline-model confusion matrices, where VGG19 and DenseNet201 show strong diagonal dominance, indicating reliable class-wise classification, while MobileNetV2 produces acceptable but slightly weaker predictions with some confusion among healthy leaf and leaf curl samples. InceptionV3 shows the highest misclassification rate, particularly for leaf curl and leaf spot, suggesting limited ability to separate visually similar disease patterns. **Fig. 7** further demonstrates that AMnet provides more consistent class-level recognition, with very high correct classification rates for healthy leaf, leaf curl, and leaf spot, and only minor errors mainly involving Cercospora leaf and leaf spot. Overall, the sequential comparison confirms that AMnet improves class-wise discrimination over the baseline models and provides the most stable confusion-matrix performance.

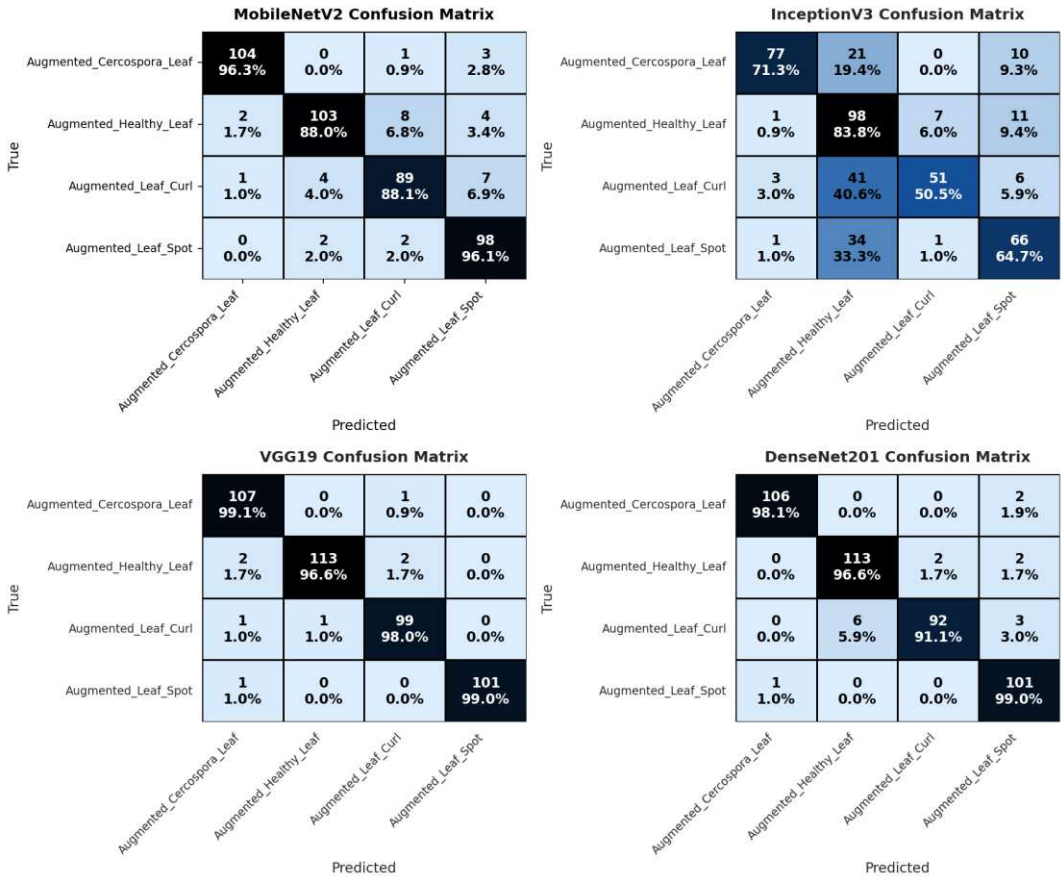


Fig 6. Baseline Models Confusion Matrix

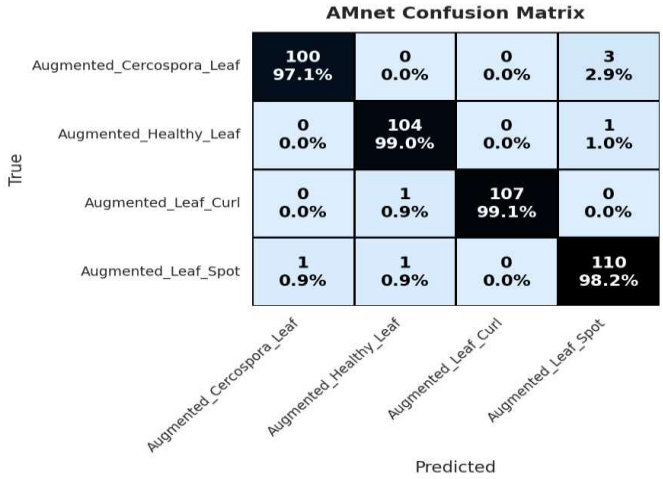


Fig 7. AMnet Confusion Matrix

4.5 ROC and PR Curve Analysis of Supervised Models

Fig. 8 demonstrates that VGG19 and DenseNet201 achieved near-perfect ROC behavior across all classes, while MobileNetV2 also maintained strong separability; in contrast, InceptionV3 showed weaker ROC performance, especially for healthy leaf and leaf spot classes. **Fig. 9** further supports this pattern, where VGG19 and DenseNet201 produced consistently high precision across recall levels, whereas InceptionV3 showed a sharper precision decline, indicating less reliable positive-class prediction. **Fig. 10** confirms the superior discriminative ability of AMnet,

with ROC-AUC values reaching 1.00 for all classes and AP values ranging from 0.997 to 1.000. Overall, the ROC and precision–recall analyses show that AMnet provides the most stable and class-consistent classification performance compared with the baseline models.

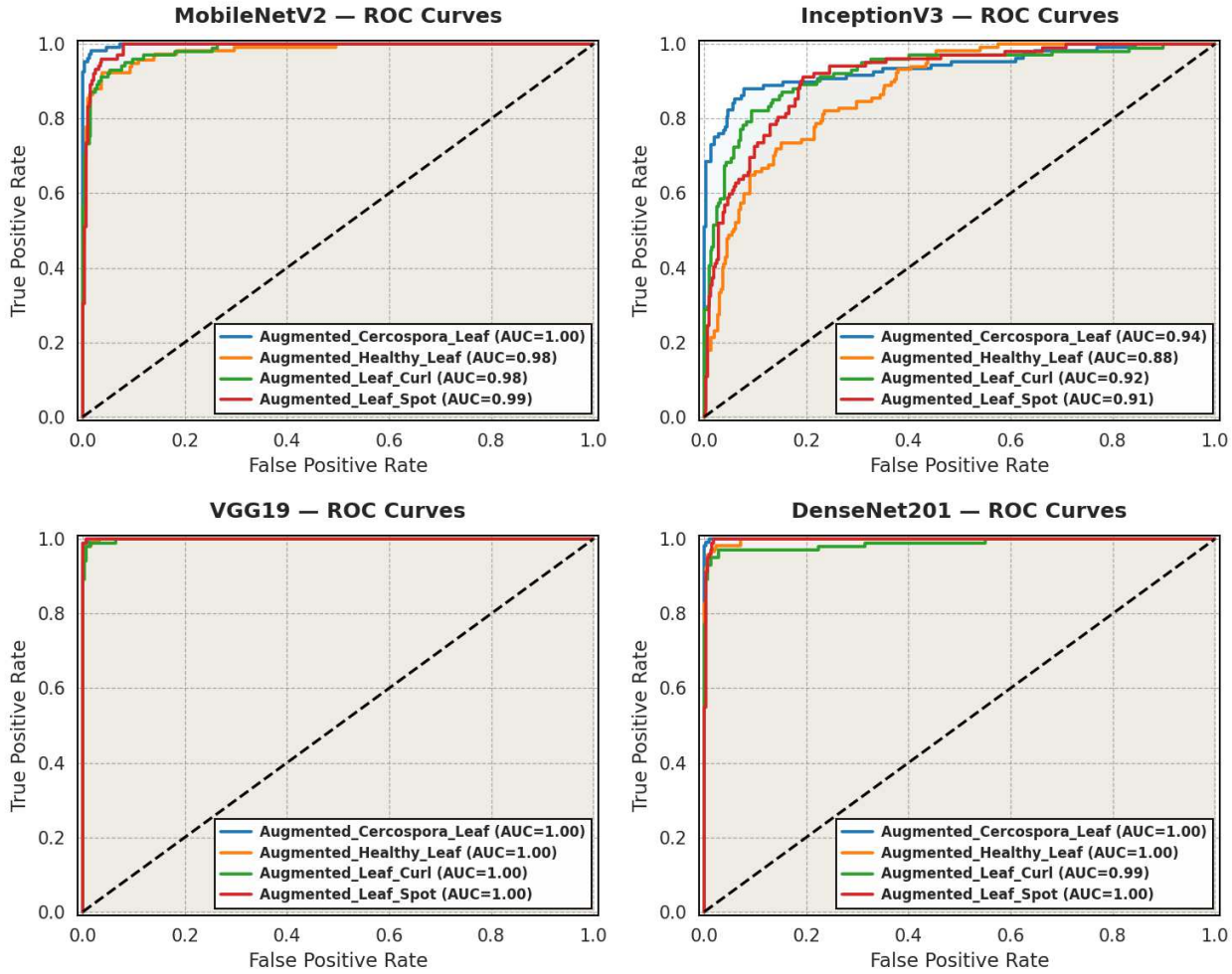


Fig 8. Baseline Models ROC Curves

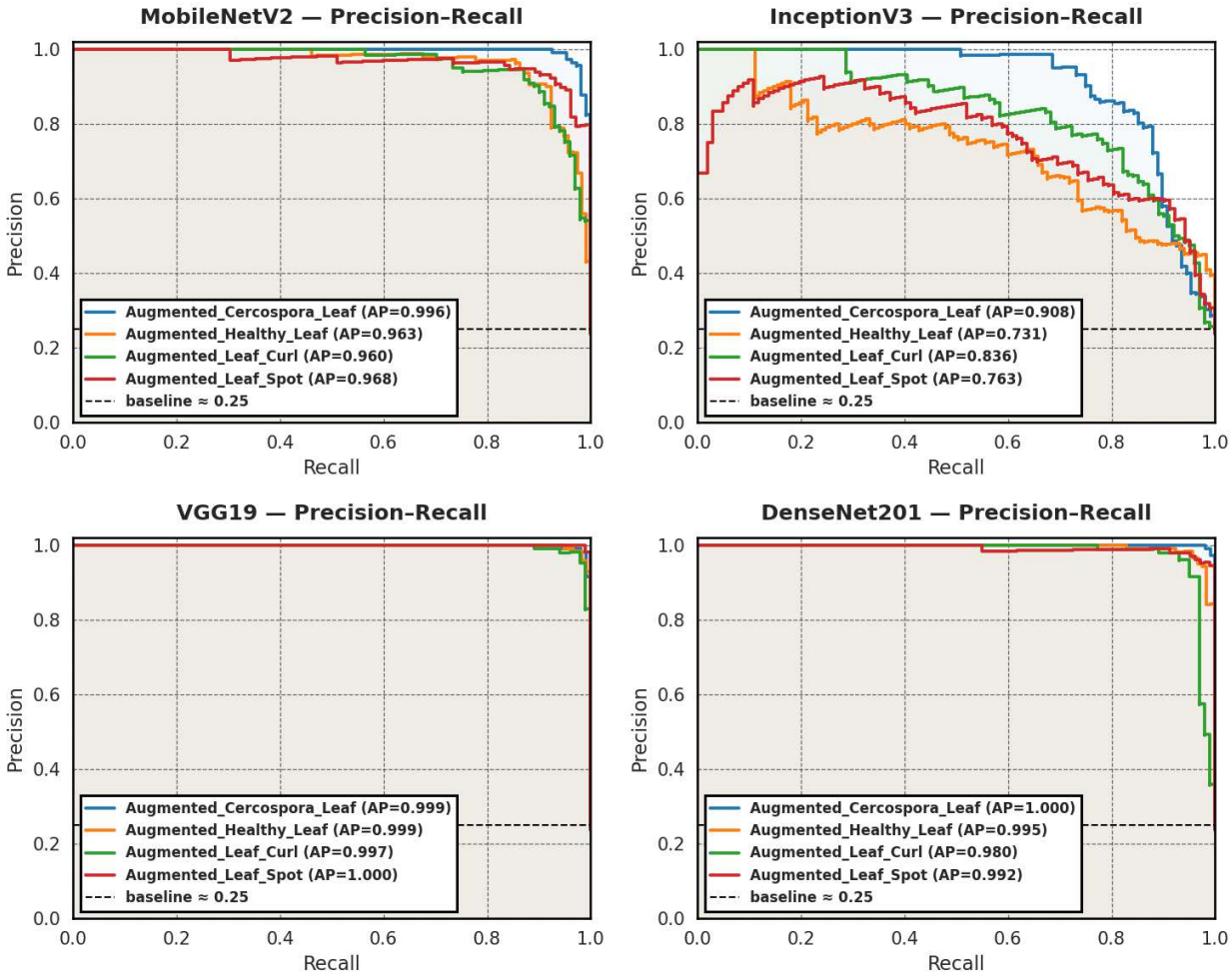


Fig 9. Baseline Models Precision Recall Curves

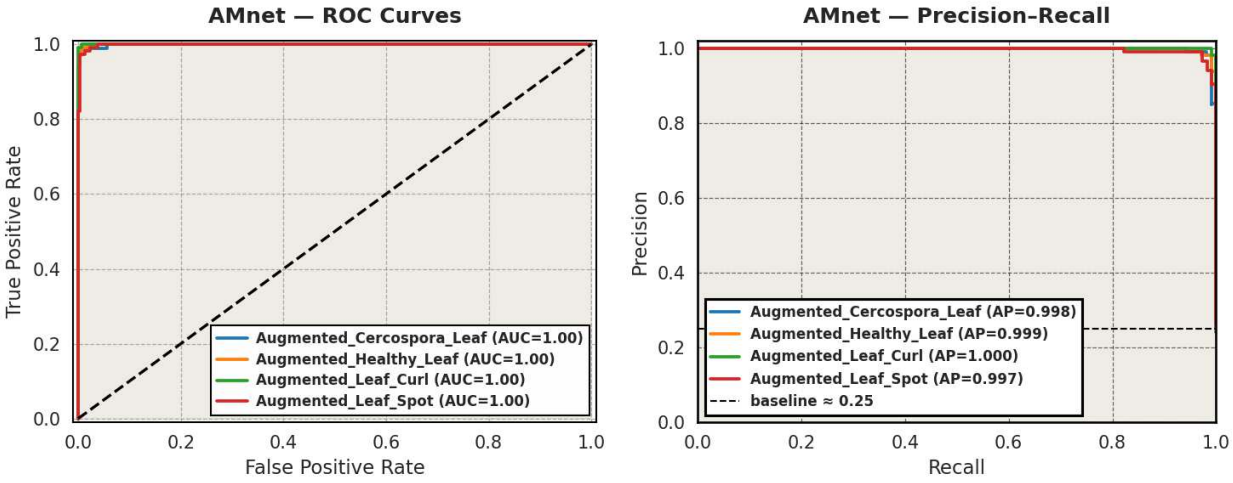


Fig 10. AMnet ROC and Precision Recall Curves

4.6 Explainable AI (XAI) Outcomes of AMnet

Fig. 11 presents the XAI outcome of AMnet using heatmap-based and boundary-based visual explanations for the four leaf classes. The highlighted regions are mainly concentrated on the leaf surface, venation structure, and visually affected areas rather than irrelevant background regions.

For Cercospora, leaf curl, and leaf spot samples, the model focuses on disease-related texture and shape variations, while the healthy sample shows attention on the intact leaf region. This indicates that AMnet makes predictions based on biologically meaningful visual cues, supporting the interpretability and reliability of the proposed model.

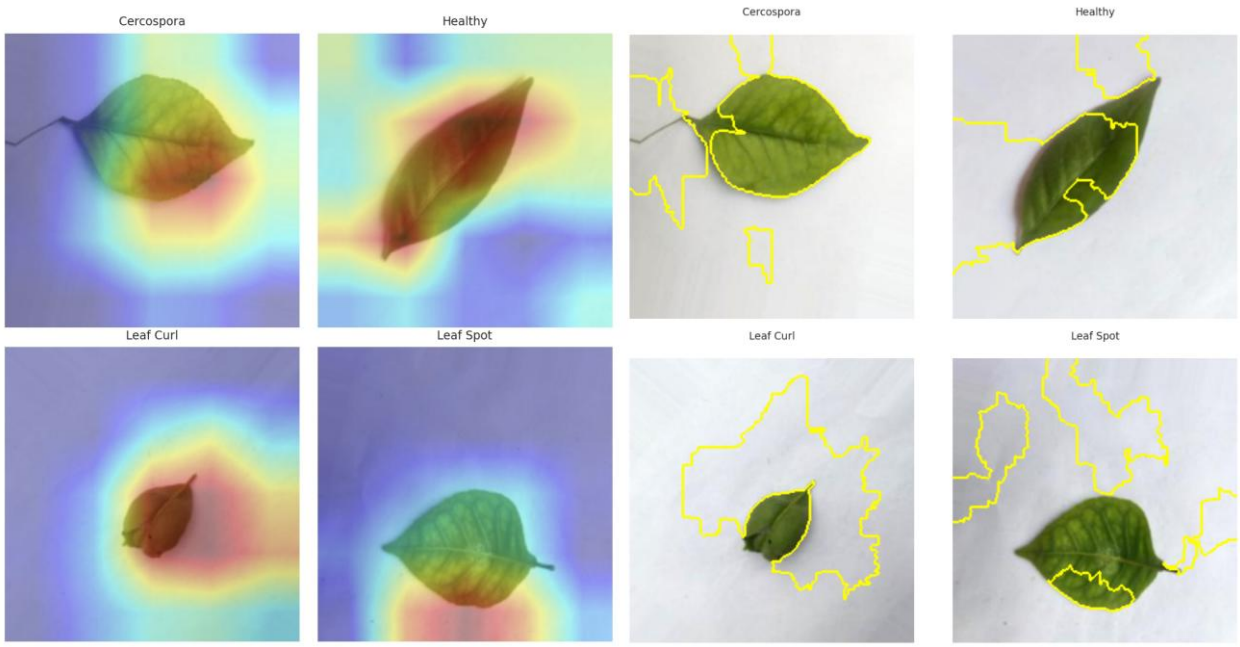


Fig 11. AMnet XAI Outcome

4.7 Principle Component Analysis of AMnet

Fig. 12 illustrates the elbow-method analysis for selecting the optimal number of clusters using K-means inertia. The inertia decreases sharply from K = 2 to K = 4, indicating a substantial improvement in within-cluster compactness as the number of clusters increases. After K = 4, the reduction in inertia becomes gradual, suggesting limited additional benefit from adding more clusters. Therefore, K = 4 can be considered the most appropriate cluster number, which is also consistent with the four target leaf classes used in this study.

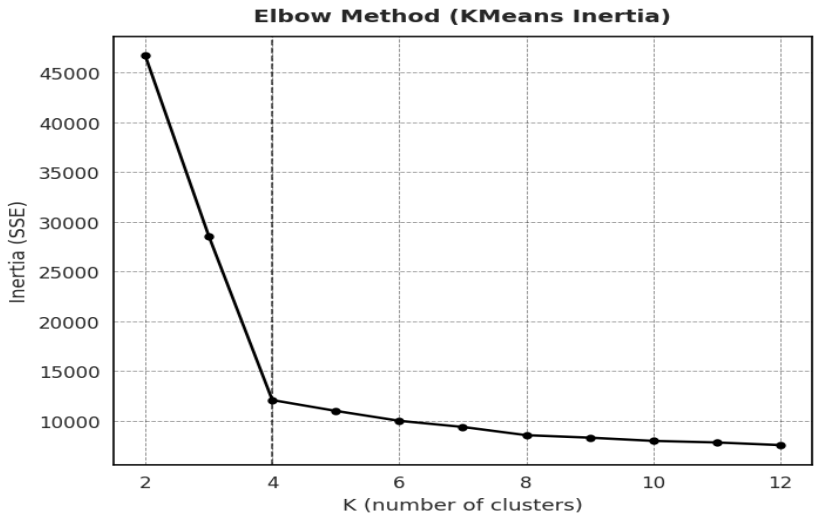


Fig 12. Elbow Method Plot

The clustering evaluation in **Fig. 13** supports K = 4 as the optimal cluster setting, with the highest Silhouette score (0.597) and the lowest Davies–Bouldin index (0.568). This indicates that the PCA-reduced feature space forms the most compact and well-separated clusters when four clusters are used. The result is also consistent with the dataset structure, since the classification task contains four leaf classes. Therefore, the clustering analysis confirms that the learned feature representation preserves meaningful class-level separability.

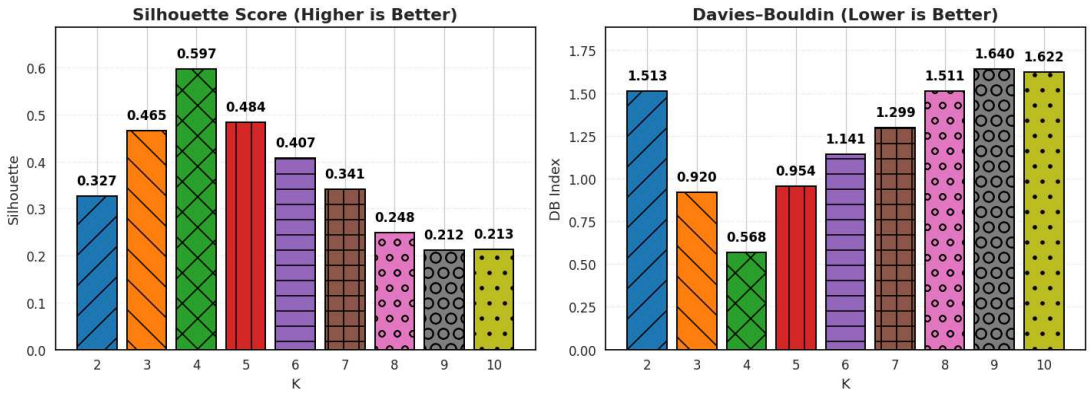


Fig 13. Clustering Evaluation Metrics for PCA-Reduced Feature Space

The PCA projection in **Fig. 14** shows that AMnet learned a clearly structured feature space, with the four leaf classes forming mostly distinct clusters. Healthy leaf, leaf curl, and leaf spot samples appear well separated, indicating strong class-specific feature representation. Cercospora leaf samples are positioned between neighboring clusters with slight overlap, suggesting some visual similarity with other disease patterns. Overall, the PCA visualization supports the quantitative clustering results and confirms that AMnet preserves meaningful class-level separability in the learned embeddings.

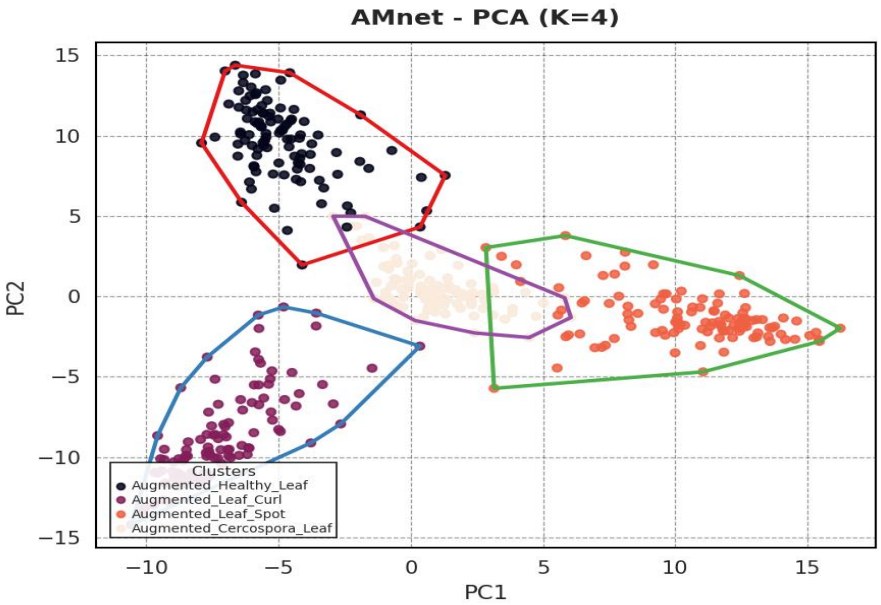


Fig 14. PCA Visualization of Learned Feature Embeddings from AMnet

Fig. 15 presents the cluster composition of AMnet feature embeddings, showing that each cluster is dominated by a single leaf class. Cluster 0 mainly represents healthy leaf samples (98.1%), Cluster 1 perfectly corresponds to leaf curl (100.0%), Cluster 2 is dominated by leaf spot (97.3%), and Cluster 3 mainly contains Cercospora leaf samples (99.0%). The very low cross-class contamination indicates strong intra-class compactness and inter-class separation in the learned feature space. This result further supports the PCA and clustering analyses, confirming that AMnet learns discriminative and class-consistent representations.

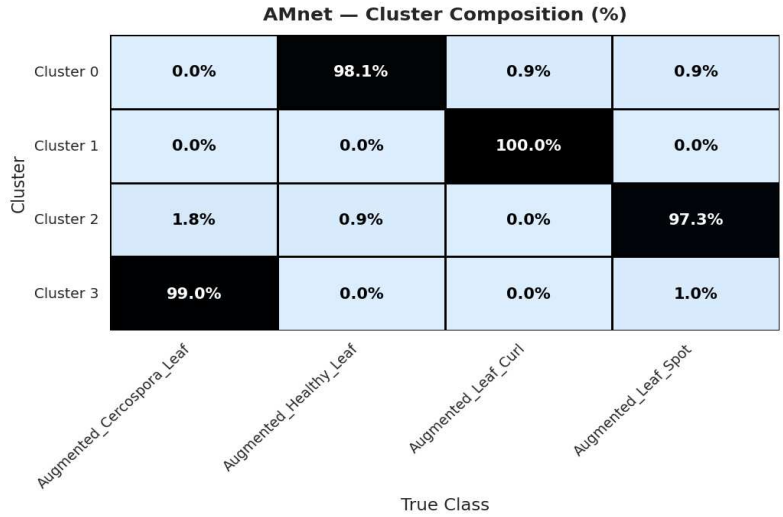


Fig 15. Cluster Composition (%)

Fig. 16 shows the class centroid distance matrix of AMnet embeddings, where all off-diagonal distances are high, ranging from 17.42 to 18.93. The largest separation is observed between leaf curl and leaf spot (18.93), while the smallest distance occurs between Cercospora leaf and leaf spot (17.42), suggesting relatively closer feature similarity between these two disease classes. The consistently large centroid distances indicate strong inter-class separation in the learned feature space. This further supports the clustering and PCA results, confirming that AMnet generates discriminative class-level representations.

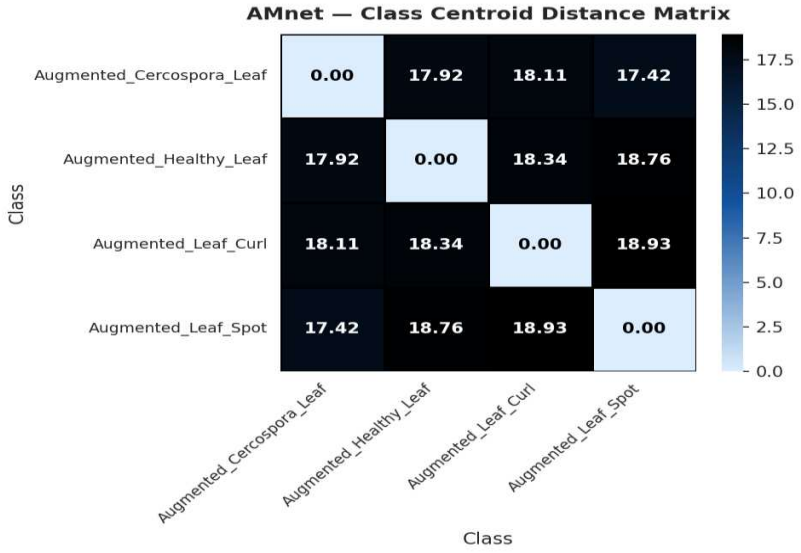


Fig 16. Class Centroid Distance Matrix

4.8 Statistical Analysis of AMnet

Table 5 summarizes the overall performance of AMnet with 95% confidence intervals, providing an estimate of the reliability of the reported results. AMnet achieved an accuracy of 0.9836, with a Wilson confidence interval of 0.9666–0.9921 and a bootstrap confidence interval of 0.9720–0.9953, indicating consistently high predictive performance. The macro F1-score of 0.9837 and MCC of 0.9781 also showed narrow bootstrap confidence intervals, suggesting stable class-balanced performance and strong agreement between predicted and true labels. Overall, the statistical analysis confirms that AMnet maintains robust classification performance with limited uncertainty across the evaluated test samples.

Table 5. Overall Performance Metrics with 95 % Confidence Intervals

Metric	Value	95 % CI (Wilson / Bootstrap)
Accuracy	0.9836	(Wilson CI 0.9666–0.9921; Bootstrap CI 0.9720–0.9953)
Macro F1	0.9837	(Bootstrap CI 0.9699–0.9953)
MCC	0.9781	(Bootstrap CI 0.9599–0.9908)

The per-class bootstrap evaluation in **Table 6** shows that AMnet achieved consistently high precision, recall, and F1-score across all four classes. Leaf curl produced the strongest class-wise performance, with perfect precision (1.0000) and the highest F1-score (0.9955), indicating highly reliable identification of this class. Cercospora leaf, healthy leaf, and leaf spot also achieved strong F1-scores of 0.9804, 0.9860, and 0.9737, respectively, with relatively narrow confidence intervals. Overall, the per-class confidence intervals confirm that AMnet provides stable and balanced recognition performance across both healthy and diseased leaf categories.

Table 6. Per-Class Evaluation Metrics with 95 % Bootstrap Confidence Intervals

Class	Precision [95 % CI]	Recall [95 % CI]	F1 Score [95 % CI]
Cercospora Leaf	0.9904 [0.9703,1.0000]	0.9707 [0.9320,1.0000]	0.9804 [0.9600,0.9952]
Healthy Leaf	0.9813 [0.9537,1.0000]	0.9908 [0.9714,1.0000]	0.9860 [0.9671,1.0000]
Leaf Curl	1.0000 [1.0000,1.0000]	0.9910 [0.9722,1.0000]	0.9955 [0.9859,1.0000]
Leaf Spot	0.9657 [0.9316,0.9912]	0.9821 [0.9554,1.0000]	0.9737 [0.9502,0.9912]

4.9 Ablation Study Analysis of AMnet

Table 7 outlines the ablation variants designed to examine the contribution of incremental architectural changes in the InceptionV3-GCN framework. The base variant, A1, introduces fixed spatial graph reasoning with a 5 × 5 grid, 512-dimensional node features, and a 256-dimensional GCN hidden layer to evaluate whether graph-based feature propagation improves CNN representation. Variants A2 and A4 modify only the dropout rate to assess the effect of regularization strength on underfitting and generalization, while A3 reduces the GCN hidden

dimension to test whether a lighter graph module can improve efficiency and reduce overfitting. Overall, these controlled variants allow a focused analysis of how dropout and GCN capacity influence model performance.

Table 7. Description of model variants used in the ablation study, highlighting incremental architectural modifications

Model ID	Variant Name	Key Modification	Expected Outcome
A1	InceptionV3-FixedGCN	node_dim=512, gcn_hidden=256, dropout=0.3, grid_size=5	Tests whether spatial graph reasoning improves over the plain CNN baseline.
A2	InceptionV3-FixedGCN-LowDropout	Only dropout is reduced from 0.3 to 0.2	Tests whether lower dropout improves accuracy by reducing underfitting.
A3	InceptionV3-FixedGCN-LightGCN	GCN hidden dimension is reduced from 256 to 128	Tests whether a smaller GCN can reduce overfitting and computation.
A4	InceptionV3-FixedGCN-HighDropout	Only dropout is increased from 0.3 to 0.4	Tests whether stronger dropout improves generalization.

Table 8 presents the ablation performance of the four InceptionV3-GCN variants, where A4 achieved the best results across all metrics, including accuracy (98.83%), F1-score (98.84%), PR-AUC (99.95%), and MCC (98.45%). Compared with A1, increasing dropout from 0.3 to 0.4 substantially improved classification performance, suggesting that stronger regularization enhanced generalization. A2 also performed strongly with 98.36% accuracy, indicating that dropout adjustment had a greater effect than reducing the GCN hidden dimension. Overall, the ablation results confirm that the high-dropout GCN variant A4 provides the most effective configuration for AMnet.

Table 8. Performance comparison of model variants in the ablation study across multiple evaluation metrics

Model ID	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	PR AUC (%)	MCC (%)
A1	95.09	95.64	95.21	95.23	99.69	93.60
A2	98.36	98.40	98.36	98.37	99.83	97.82
A3	95.79	96.04	95.94	95.79	99.72	94.53
A4	**98.83**	**98.86**	**98.85**	**98.84**	**99.95**	**98.45**

4.10 Model Performance on Four Fold Cross-Validation

Table 9 reports the four-fold cross-validation results, demonstrating that the proposed model maintained consistently high performance across all evaluation metrics. The mean accuracy, precision, recall, and F1-score were highly comparable, ranging from 97.35% to 97.36%, which indicates balanced predictive performance across the folds. The very low standard deviation values, between 0.0023 and 0.0025, suggest minimal performance variation and strong training stability. These results confirm that the model generalizes reliably across different data partitions rather than depending on a single train–test split.

Table 9. Performance comparison of the models on four fold cross-validation

Model Name	Mean Results	Standard Deviation
Accuracy	97.35	0.0024
Precision	97.36	0.0023
Recall	97.36	0.0025
F1 Score	97.35	0.0024

4.11 Model Performance on Additional Dataset

Table 10 shows the performance comparison of the evaluated models on the additional dataset. Among all models, AMnet achieved the best overall performance with the highest accuracy (81.25%), F1 score (80.11%), PR AUC (87.51%), and MCC (76.55%), indicating superior classification capability and robustness. VGG19 and DenseNet201 demonstrated competitive performance, whereas MobileNetV2 produced the lowest results across most evaluation metrics. Overall, the findings confirm the better generalization ability and effectiveness of AMnet compared with the baseline models.

Table 10. Performance comparison of the models on additional dataset

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	PR AUC (%)	MCC (%)
MobileNetV2	68.50	70.46	68.38	67.43	76.09	61.41
VGG19	77.50	77.97	77.38	77.04	86.10	72.18
DenseNet201	77.25	81.41	76.88	75.71	85.17	73.00
AMnet	**81.25**	**80.37**	**80.30**	**80.11**	**87.51**	**76.55**

4.12 Prototype Application

The prototype Gradio application in **Fig. 17** demonstrates the practical deployment of the proposed AMnet model for Aegle marmelos leaf disease classification. The interface allows users to upload a leaf image and automatically generates the predicted disease class, class-wise probability distribution, and Grad-CAM-based visual explanation. In the displayed example, the

system predicts the input image as Cercospora leaf with high confidence, while the Grad-CAM output highlights the image regions that contributed to the decision. This prototype confirms the feasibility of using the proposed model as an interactive and explainable decision-support tool for leaf disease diagnosis.

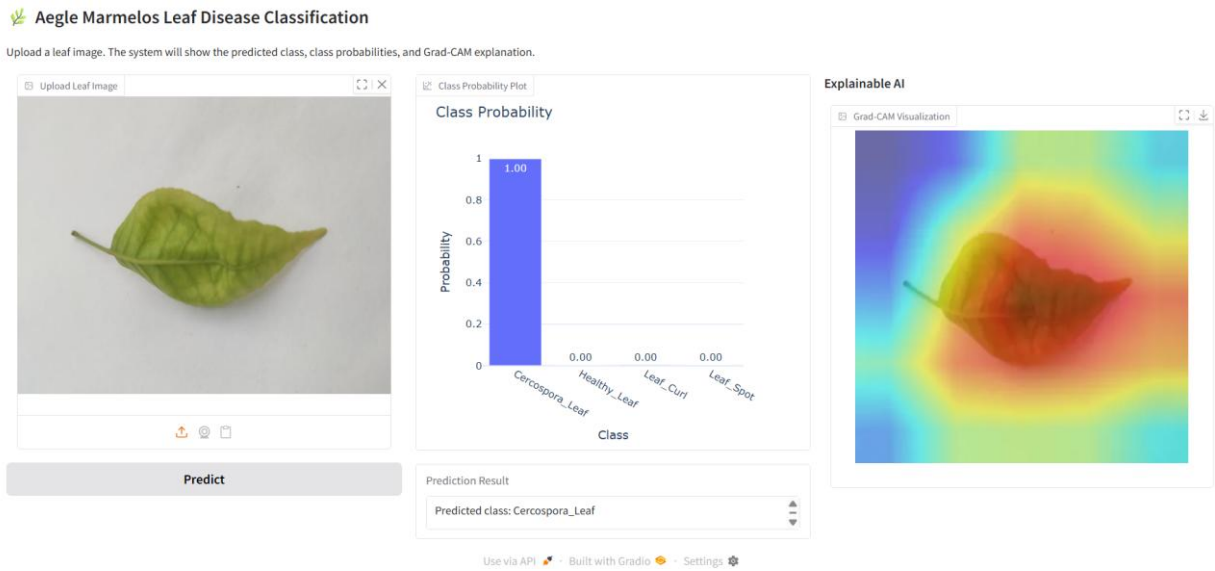


Fig. 17 Prototype Gradio Application

4.13 Study Comparison

Table 11 positions the proposed AMnet model against recent graph-based, attention-driven, and semi-supervised plant disease classification approaches. Compared with prior studies, AMnet achieves highly competitive performance, with 98.83% accuracy, 98.84% F1-score, and 98.45% MCC, while also incorporating explainability, clustering-based feature validation, and prototype deployment. Although some existing methods report slightly higher accuracy on specific datasets, many are limited by narrow crop coverage, pseudo-label uncertainty, or restricted field validation. Overall, the comparison shows that AMnet provides a strong balance between predictive performance, interpretability, and practical applicability for Aegle marmelos leaf disease diagnosis.

Table 11. Comparative positioning of the proposed AMnet against graph-based, attention-driven, and semi-supervised plant disease classification studies

Ref	Dataset	Task	Methodology	Model	Performance
[19]	CIFAR-10, CIFAR-100	Image classification	Semi-supervised masked autoencoding	SSMAE	+9.24% improvement
[22]	Coffee and banana leaf datasets	Leaf disease classification	Ambiguity-aware semi-supervised learning	Pseudo-label Rejection Framework	99.46%, 100%

[24]	PlantVillage	Adjacent leaf disease prediction	Graph-based disease propagation modeling	AdjLeafGNN	Accuracy: 98.88%; F1: 98.71%
[25]	Apple, potato, and sugarcane datasets	Leaf disease classification	Hybrid graph convolution and attention learning	GAT-GCN	F1: 0.9818, 0.9743, 0.8799
[27]	Five plant disease datasets	Plant disease classification	Attention-enhanced interpretable deep learning	CBAM-VGG16	98.87%
[29]	Plant disease datasets	Semi-supervised leaf disease detection	Pseudo-label-guided semi-supervised learning	Pseudo CNN	96.4%
This Study	**Aegle marmelos leaf image dataset**	**Four-class leaf disease classification**	**Supervised CNN-GCN feature learning with regularized graph-based representation, XAI analysis, clustering validation, and prototype deployment**	**AMnet**	**Accuracy: 98.83%; F1-score: 98.84%; MCC: 98.45%**

5 Discussion

5.1 Summary of major findings

The present study demonstrates that the proposed AMnet model provides strong performance for four-class Aegle marmelos leaf disease classification. Across the supervised model comparison, AMnet achieved the highest overall performance, with 98.83% accuracy, 98.86% precision, 98.85% recall, 98.84% F1-score, 99.95% PR-AUC, and 98.45% MCC. The improvement over strong baseline models such as VGG19 and DenseNet201 indicates that the proposed architecture was able to learn more discriminative representations for separating Cercospora leaf, healthy leaf, leaf curl, and leaf spot classes. In addition to classification performance, AMnet also showed the fastest inference time of 1.83 ms, suggesting practical suitability for real-time or near-real-time disease recognition.

5.2 Comparison with supervised baseline models

The baseline analysis revealed clear performance differences among the evaluated CNN architectures. VGG19 and DenseNet201 performed strongly, whereas MobileNetV2 showed moderate performance and InceptionV3 produced the weakest results, achieving only 68.22% accuracy. The confusion-matrix analysis further confirmed this trend, with InceptionV3 showing considerable misclassification among visually similar classes, particularly leaf curl and leaf spot.

The weak standalone performance of InceptionV3 warrants explanation, since AMnet uses it as a backbone. Three factors contribute. InceptionV3 was designed for 299×299 inputs, and resizing to 224×224 degrades its multi-scale receptive field behavior. The relatively small dataset of 1,513 raw images increases fine-tuning instability in deeper layers, consistent with the oscillating validation loss in Figure 4. Additionally, standalone InceptionV3 was evaluated without the dropout regularization confirmed as essential by the ablation study. AMnet addresses all three weaknesses structurally: the fixed-GCN module aggregates information across neighboring feature regions to smooth locally noisy activations, the residual connection preserves the original CNN features alongside graph-refined representations, and the higher dropout rate improves generalization. Ablation variant A1, which adds only the GCN module without dropout adjustment, already recovers performance to 95.09%, confirming that the graph module is the primary compensating mechanism. AMnet therefore represents a purposeful architectural response to the known limitations of standalone InceptionV3 in small-dataset fine-grained classification, rather than an incidental improvement.

5.3 Computational efficiency and deployment relevance

An important outcome of this study is that AMnet achieved the best predictive performance while also maintaining the fastest inference speed. Although AMnet had the largest parameter count and model size compared with the baseline models, its inference time was substantially lower, reaching approximately 547.89 frames per second. This indicates that the architecture is computationally efficient during prediction, even though it requires more storage than lightweight models such as MobileNetV2. Therefore, AMnet may be suitable for web-based or workstation-based diagnostic applications. However, deployment on highly resource-constrained mobile or embedded devices may still require model compression, pruning, quantization, or lightweight architectural redesign.

5.4 Effect of architectural modification in the ablation study

The ablation study provides evidence that architectural and regularization choices had a direct effect on model performance. Among the tested variants, the high-dropout configuration, A4, produced the best results, achieving 98.83% accuracy and 98.84% F1-score. This suggests that stronger dropout regularization helped improve generalization in the proposed CNN-GCN setting. In contrast, reducing the GCN hidden dimension improved efficiency but did not produce the highest classification performance. These findings indicate that AMnet benefits from a sufficiently expressive graph-learning component combined with appropriate regularization.

5.5 Reliability of predictions and statistical confidence

The statistical analysis supports the robustness of AMnet's performance. The overall accuracy of 0.9836 was supported by a Wilson confidence interval of 0.9666–0.9921 and a bootstrap confidence interval of 0.9720–0.9953. Similarly, the macro F1-score and MCC showed narrow bootstrap confidence intervals, indicating limited uncertainty in the reported results. Per-class evaluation also showed strong performance across all categories, with F1-scores ranging from 0.9737 for leaf spot to 0.9955 for leaf curl. These results suggest that AMnet did not perform well only on one dominant class, but maintained balanced recognition across both healthy and diseased leaf samples.

5.6 Feature-space separability and explainability

The clustering and PCA analyses provide additional evidence that AMnet learned meaningful class-level representations. The elbow method, silhouette score, and Davies–Bouldin index all supported K = 4, which is consistent with the four target classes. The PCA visualization showed largely separated feature clusters, while the cluster composition analysis demonstrated that each

cluster was dominated by a single class. The class centroid distance matrix further confirmed strong inter-class separation, with high centroid distances among all class pairs. Together, these analyses indicate that AMnet does not only produce accurate predictions, but also organizes the learned feature space in a biologically and semantically meaningful manner.

The XAI results further strengthen the interpretability of the proposed model. The Grad-CAM and boundary-based visual explanations showed that AMnet generally focused on leaf regions, venation structures, and disease-related visual patterns rather than irrelevant background areas. This is important for plant disease diagnosis because high accuracy alone is not sufficient for practical use; the model should also provide evidence that its decisions are based on meaningful visual cues. The prototype Gradio application demonstrates this practical value by combining image upload, class prediction, probability visualization, and explainable output in a single interface.

5.7 Comparison with related studies

Compared with existing graph-based, attention-driven, and semi-supervised plant disease classification studies, AMnet achieved competitive performance while offering additional interpretability and deployment-oriented analysis. Some prior works reported very high accuracy on specific datasets, but many were limited by narrow crop coverage, dependence on pseudo-labeling, or insufficient real-world validation. The proposed method contributes by combining strong supervised classification performance with XAI visualization, feature-space clustering analysis, statistical confidence estimation, and a working prototype interface. This broader evaluation makes the study more complete than accuracy-only comparisons.

5.8 Limitations and threats to validity

Despite the promising results, several limitations should be acknowledged. First, the study was conducted on a single crop species, *Aegle marmelos*, and only four leaf categories were considered. Therefore, the current findings should not be interpreted as direct evidence that AMnet will generalize equally well to other crops, disease types, or field environments without further validation. Second, the images seem to have been collected under relatively controlled or semi-controlled imaging conditions. Real field deployment may involve complex backgrounds, overlapping leaves, shadows, motion blur, variable illumination, occlusion, different disease stages, and mixed infections. These factors can reduce model performance compared with controlled experimental results. Third, although AMnet achieved fast inference, it also had the largest number of parameters and model size among the compared models. This may limit its direct use on low-memory mobile devices or embedded agricultural sensors unless additional optimization is performed. The current prototype demonstrates feasibility in a web-based environment, but it does not yet confirm performance under offline mobile deployment conditions. Fourth, the XAI analysis provides useful visual evidence, but Grad-CAM-style explanations are qualitative and should not be treated as definitive biological validation. Expert plant pathologist assessment would be needed to confirm whether the highlighted regions consistently correspond to true disease symptoms. In addition, the model may still rely on subtle dataset-specific visual cues that are not obvious from the heatmaps.

Finally, the study used four-fold cross-validation and bootstrap confidence intervals, which improve reliability, but larger external validation datasets from different locations, seasons, cameras, and disease severities are still needed. Such validation would provide stronger evidence of generalizability and clinical-agricultural usefulness.

5.9 Future directions

Future work should focus on validating AMnet using larger, multi-location field datasets collected under diverse environmental conditions. The model should also be tested on additional *Aegle marmelos* disease categories and visually similar leaf disorders to assess its diagnostic boundary. From a technical perspective, model compression, pruning, quantization, and mobile deployment should be explored to reduce memory requirements. Expert-verified XAI evaluation can further strengthen interpretability, while integration with disease severity estimation and management recommendations may transform the system from a classification tool into a more complete plant-health decision-support platform.

6 Conclusion

This study proposed AMnet for four-class *Aegle marmelos* leaf disease classification and demonstrated its effectiveness through supervised evaluation, ablation analysis, statistical validation, explainability, clustering analysis, and prototype deployment. AMnet achieved the best overall performance among the evaluated models, with 98.83% accuracy, 98.84% F1-score, 99.95% PR-AUC, and 98.45% MCC, while also providing the fastest inference time. The confusion matrix, ROC, and precision–recall analyses confirmed its strong class-wise discrimination, and the ablation study showed that the high-dropout GCN configuration provided the most effective model variant. Statistical confidence intervals further supported the robustness of the results, while PCA, clustering metrics, cluster composition, and centroid-distance analysis demonstrated meaningful feature-space separability. Grad-CAM-based explanations and the Gradio prototype also showed the practical potential of AMnet as an explainable decision-support tool. However, the study remains limited to a single crop dataset and controlled imaging conditions; therefore, external validation using larger field-based datasets is required. Overall, AMnet offers a promising foundation for accurate, interpretable, and deployable leaf disease diagnosis.

Data availability

The dataset analysed of this study is publicly available in the Mendeley Data repository at (https://data.mendeley.com/datasets/54r883j5zr/1).

Declarations

Competing interests

The authors declare no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Generative AI and AI-Assisted Technologies

During the preparation of this manuscript, the authors utilized AI-assisted tools to enhance language clarity and readability. All generated or refined content was carefully reviewed and edited by the authors, who take full responsibility for the accuracy, integrity, and originality of the work.

References

[1] Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419. https://doi.org/10.3389/fpls.2016.01419.

[2] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311–318. https://doi.org/10.1016/j.compag.2018.01.009.
[3] Barbedo, J. G. A. (2018). Factors influencing the use of deep learning for plant disease recognition. Biosystems Engineering, 172, 84–91. https://doi.org/10.1016/j.biosystemseng.2018.05.013.
[4] Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations.
[5] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision, 618–626. https://doi.org/10.1109/ICCV.2017.74.
[6] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations, 97–101. https://doi.org/10.18653/v1/N16-3020.
[7] Deepika Lakshmi Kanumuri, Madhuri Kamma, Somitha Anna, and Pavan Kumar Kolluru (2025). Attention-based deep learning model for robust pneumonia classification and categorization using image processing. In Proceedings of the ICITSM Part I. EAI. https://doi.org/10.4108/eai.28-4-2025.2357765
[8] Gao, X., Zhang, F., Guo, X. et al. Attention-based deep learning for accurate cell image analysis. Sci Rep 15, 1265 (2025). https://doi.org/10.1038/s41598-025-85608-9
[9] Miao, M., Yan, W., Gao, X., Yang, L., Zhou, J., & Zhang, W. (2025). Attention-Based CNN-BiGRU-Transformer Model for Human Activity Recognition. Applied Sciences, 15(23), 12592. https://doi.org/10.3390/app152312592
[10] N, A., & S, U. (2025). SkinProNet: An attention-based deep learning system for skin disease classification and segmentation. Softw. Impacts, 26, 100798.
[11] Gupta, R., Kumar, V., & Tripathi, V. (2025). ESAM: Efficient Spatial Attention based Deep Learning Approach for Fundus Images Classification. Biomedical & Pharmacology Journal.
[12] Yousafzai, S.N., Nasir, I.M., Mansour, S. et al. A hybrid deep learning approach integrating CNN and transformer for lung cancer classification using CT scans. Sci Rep (2026). https://doi.org/10.1038/s41598-026-41161-7
[13] Ayesha, N., & Khalidi, H. (2026). A Vision Transformer with a Self-Attention Mechanism for High-Accuracy Blood Cell Classification. Engineering, Technology & Applied Science Research, 16(1), 32557–32563. https://doi.org/10.48084/etasr.16600
[14] Fuqiang Kuang. 2026. A CNN-Transformer Hybrid Model Based on Tri-domain Attention for Music Genre Classification. In Proceedings of the 2025 5th International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR '25). Association for Computing Machinery, New York, NY, USA, 378–385. https://doi.org/10.1145/3800227.3800285
[15] Xu, H., & Zhang, X. (2025). Graph Convolutional Network with Agent Attention for Recognizing Digital Ink Chinese Characters Written by International Students. Information, 16(9), 729. https://doi.org/10.3390/info16090729
[16] Yan, P., An, T., Yu, J. et al. Graph convolutional network with reinforced dependency graph and denoising mechanism for sarcasm detection. Sci Rep 15, 35094 (2025). https://doi.org/10.1038/s41598-025-18849-3
[17] Yang, G., Li, Y., Zhou, W., Wu, Y., Wu, W., Gu, X. (2022). Attention Mechanism Based on Temporal Graph Convolutional Neural Network for Traffic Flow Prediction. In: Jia, Y., Zhang, W., Fu, Y., Yu, Z., Zheng, S. (eds) Proceedings of 2021 Chinese Intelligent Systems

Conference. Lecture Notes in Electrical Engineering, vol 804. Springer, Singapore. https://doi.org/10.1007/978-981-16-6324-6_44
[18] Su, L., Huang, J., & Li, F. (2026). Adaptive graph convolution neural networks with spatial and frequency domain attention mechanism for traffic flow forecasting. Applied Soft Computing, 197, 115156. https://doi.org/10.1016/j.asoc.2026.115156
[19] Luo, X., Zhu, C., Zhang, D., & Li, Q. (2023). Dynamic Graph Convolutional Network with Attention Fusion for Traffic Flow Prediction. ArXiv.org. https://arxiv.org/abs/2302.12598
[20] Yang, H., Jiang, C., Song, Y. et al. TARGCN: temporal attention recurrent graph convolutional neural network for traffic prediction. Complex Intell. Syst. 10, 8179–8196 (2024). https://doi.org/10.1007/s40747-024-01601-1
[21] Chang, Q., Li, X., & Duan, Z. (2024). Graph global attention network with memory: A deep learning approach for fake news detection. Neural Networks, 172, Article 106115. https://doi.org/10.1016/j.neunet.2024.106115
[22] Thekumparampil, Kiran K, Wang, C., Oh, S., & Li, L.-J. (2018). Attention-based Graph Neural Network for Semi-supervised Learning. ArXiv.org. https://arxiv.org/abs/1803.03735
[23] Ilsever, M., & Baz, I. (2024). Consistency Regularization Based Semi-supervised Plant Disease Recognition. Smart Agricultural Technology.
[24] Qiu, Y., Meng, J., & Li, B. (2025). Semi-supervised Strong-Teacher Consistency Learning for few-shot cardiac MRI image segmentation. Computer methods and programs in biomedicine, 261, 108613. https://doi.org/10.1016/j.cmpb.2025.108613
[25] Huang, S., Wang, S., Qiao, S., Liu, Y., & Liu, T. (2025). SEMI-PLC: A framework for semi-supervised medical images segmentation with pseudo label correction. Computer methods and programs in biomedicine, 271, 109027. https://doi.org/10.1016/j.cmpb.2025.109027
[26] Rahmati, B., Shirani, S. & Keshavarz-Motamed, Z. A hybrid approach for enhancing pseudo-labeling in medical images through pseudo-label refinement. Sci Rep 15, 35161 (2025). https://doi.org/10.1038/s41598-025-19121-4
[27] Faysal, A., Rostami, M., Roshan, R. G., Muralidhar, N., & Wang, H. (2026). Semi-Supervised Masked Autoencoders: Unlocking Vision Transformer Potential with Limited Data. ArXiv.org. https://arxiv.org/abs/2601.20072
[28] Pham, TC., Nguyen, TN. & Nguyen, VD. Ambiguity-aware semi-supervised learning for leaf disease classification. Sci Rep 15, 14070 (2025). https://doi.org/10.1038/s41598-025-95849-3
[29] B. Islam et al., "Annotation-Efficient Lung Cancer Classification: A Comparative Study of Supervised, Self-Supervised, and Semi-Supervised Learning," 2025 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), Dhaka, Bangladesh, 2025, pp. 835-840, doi: 10.1109/BECITHCON69222.2025.11504014.
[30] Surekha, B., Subha Mastan Rao, T. AdjLeafGNN: a hybrid deep learning and graph neural network framework for probabilistic modeling of adjacent leaf disease spread in precision agriculture. Sci Rep (2026). https://doi.org/10.1038/s41598-026-45061-8
[31] Sundhar S, Sharma R, Maheshwari P, Kumar SR and Kumar TS (2025) Enhancing leaf disease classification using GAT-GCN hybrid model. Front. Plant Sci. 16:1569821. doi: 10.3389/fpls.2025.1569821
[32] Aboelenin, S., Elbasheer, F.A., Eltoukhy, M.M. et al. A hybrid Framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer. Complex Intell. Syst. 11, 142 (2025). https://doi.org/10.1007/s40747-024-01764-x
[33] Singh, B., Sharma, R. P., & Dey, S. (2025). Interpretable Plant Leaf Disease Detection Using Attention-Enhanced CNN. ArXiv.org. https://arxiv.org/abs/2512.17864

[34] D, A., T, L. N., & Prusty, M. R. (2026). An attention-augmented lightweight convolutional framework for fine-grained plant leaf disease classification. Frontiers in plant science, 17, 1762956. https://doi.org/10.3389/fpls.2026.1762956

[35] Sanyal, M., Dasgupta, S., Chaki, R. et al. Pseudo CNN: An Updated CNN Based Semi Supervised Learning Model for Leaf Disease Detection. Innovations Syst Softw Eng 21, 1511–1520 (2025). https://doi.org/10.1007/s11334-025-00617-1

[36] Huq, Rezwan; Ahmed, Jubaer; Nessa, Maherun; Islam, Tasmia; Gani, Raiyan (2024), "Smartphone Image Dataset for Aegle Marmelos, Hog Plum and Lemon Plant Leaf ", Mendeley Data, V1, doi: 10.17632/54r883j5zr.1