Article

SAH-Net: Species-Aware Hierarchical Learning for Medicinal Plant Disease Recognition

Wang Hao Lin [1,*] and Mir Muhammad Nizamani [2,*]

[1] School of Geography, Nanjing Normal University, Jiangsu Province 210023, China
[2] Research Group for Lowland Plain and Human Activity, Institute of Marine Sciences, Shantou University, Shantou 513063, China
* Correspondence: 31203043@nnu.edu.cn (W.H.L.); mirmohammadnizamani@outlook.com (M.M.N.)

Abstract

Accurate recognition of medicinal plant diseases is important for maintaining plant health and product quality. Although existing deep learning approaches have achieved high classification accuracy, most formulate disease recognition as a flat multiclass problem without explicitly supervising species-related and disease-related representations. This study proposes a Species-Aware Hierarchical Network (SAH-Net) to investigate the integration of these representations within a unified classification framework. SAH-Net employs an ImageNet-pretrained EfficientNet-B0 backbone with two query-based attention branches for species-aware and disease-aware feature extraction, followed by adaptive gated hierarchical fusion. The model jointly performs three-class species recognition, seven-class disease/condition recognition, and 12-class species–condition classification. Experiments were conducted on 2,547 images of *Kalanchoe pinnata*, *Azadirachta indica*, and *Ocimum tenuiflorum*, using a stratified 70:15:15 training, validation, and test split. On the 383-image test subset, SAH-Net achieved 100% accuracy and macro F1-score across all three tasks, with an AUC of 1.000 for each final classification category. Explainability visualizations provided qualitative insights into the model's spatial attention patterns. SAH-Net required 6.019 million parameters and 0.893 GFLOPs, achieving 138.29 images/s using MPS acceleration on an Apple M5 Pro MacBook Pro. Although the results demonstrate the feasibility of hierarchical learning, external validation remains necessary to establish generalization beyond the evaluated acquisition conditions.

Keywords: Medicinal plants; plant disease recognition; hierarchical learning; attention mechanism; deep learning; explainable artificial intelligence; SAH-Net.

Citation: Lin, W.H.; Nizamani, M.M. SAH-Net: Species-Aware Hierarchical Learning for Medicinal Plant Disease Recognition. *Impact in Agriculture* **2026**, 2, 6. https://doi.org/10.65500/agriculture-2026-006

Received: 13 May 2026 | Revised: 7 July 2026 | Accepted: 12 July 2026 | Published: 21 July 2026

Copyright: © 2026 by the authors. Licensee Impaxon, Malaysia. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Medicinal plants represent an important natural source of bioactive compounds and continue to contribute to traditional medicine, pharmaceutical development, and healthcare applications [1,2]. Their therapeutic importance is closely associated with phytochemical constituents exhibiting antimicrobial, antioxidant, anti-inflammatory, and other biological activities[2,3]. However, preserving the medicinal value of these plants requires maintaining both plant health and the quality of harvested material. Foliar diseases can adversely affect plant growth, productivity, and quality, making timely disease management an important component of sustainable

medicinal plant cultivation [4]. This consideration is particularly relevant for *Kalanchoe pinnata*, *Azadirachta indica* (Neem), and *Ocimum tenuiflorum* (Tulsi), which have been associated with different therapeutic applications [5–7]. Consequently, accurate recognition of both plant species and their corresponding health conditions is important for medicinal plant management.

Leaves provide a convenient visual source for automated plant analysis because they are commonly available and contain useful information related to shape, venation, texture, and color [8]. Nevertheless, manual identification can become difficult when different medicinal species exhibit similar morphological characteristics or when disease symptoms share comparable patterns of discoloration, lesions, and texture variation. Mukherjee et al. [9], for example, demonstrated the feasibility of simultaneously recognizing medicinal leaf species and maturity levels using computer vision, obtaining approximately 99% classification accuracy for Neem, Tulsi, and Kalmegh. Earlier feature-based approaches relied heavily on manually extracted characteristics [10,11], whereas the emergence of deep learning has enabled discriminative representations to be learned directly from leaf images with considerably less dependence on handcrafted feature engineering [8].

Transfer learning and convolutional neural networks (CNNs) have subsequently become dominant approaches for medicinal leaf recognition. Sachar and Kumar [12] combined MobileNetV2, InceptionV3, and ResNet50 through ensemble learning and achieved 99.66% test accuracy, while their three- and five-fold validation experiments produced approximately 99.94% average accuracy. Tiwari et al. [13] investigated multiclass leaf classification involving 22 categories and obtained 97.69% average test accuracy together with 98.68% cross-validation accuracy. Machine learning approaches have also been investigated specifically for medicinal leaf disease diagnosis; Meenakshi [14] combined adaptive enhancement, segmentation, handcrafted features, and modified logistic regression for disease identification. Meanwhile, modified CNN architectures such as AousethNet achieved 99.89% accuracy for medicinal leaf identification [15]. These findings collectively demonstrate the considerable improvement obtained when learned visual representations replace or complement conventional feature engineering.

More recent studies have concentrated on transfer learning, attention mechanisms, multi-scale feature extraction, and joint learning. Kayaalp [16] evaluated seven pretrained deep learning architectures for medicinal plant type and disease classification and reported a highest test accuracy of 98.69% using DenseNet121. Spatial and channel-related information has also received increasing attention. Azadnia et al. [17] utilized visual characteristics and attention-oriented deep representations for medicinal and poisonous plant classification, whereas Dey et al. [18] compared multiple CNN architectures on 5,878 images from 30 medicinal plants. Lightweight recognition has similarly been explored using MobileNet [19], while transfer-learning frameworks based on InceptionV3, ResNet, and VGG architectures have demonstrated their effectiveness in plant disease recognition [20]. For early-stage disease diagnosis, Kini et al. [21] investigated pretrained convolutional networks, and modified transfer-learning strategies have also been developed for potato [22] and tea leaf diseases [23].

Attention and multi-task learning have further changed the manner in which medicinal leaf features are represented. Sharma and Vardhan [24] proposed AELGNet by integrating local and global representations with attention, obtaining 99.71% accuracy, 99.80% precision, 99.75% recall, and 99.77% F1-score. Their MTJNet framework [25] subsequently demonstrated the potential of multi-task learning, reporting 99.71% accuracy and 99.58% F1-score. Hybrid CNN designs incorporating Squeeze-and-Excitation mechanisms have also been explored for medicinal plant identification [26]. Other investigations have combined CNN feature extraction with optimized conventional classifiers [27], while optimized deep architectures have been developed for disease recognition in other crops [28]. These developments indicate a gradual transition from conventional flat CNN classification toward architectures capable of emphasizing complementary visual information.

Despite these advances, plant disease classification remains affected by differences in acquisition conditions, disease appearance, background characteristics, and inter-class similarity [29,30]. Recent reviews consequently emphasize the importance of robust visual representation learning for practical disease recognition [31]. Vision Transformers (ViTs) provide an additional mechanism for capturing long-range spatial relationships [32], and hybrid CNN–Transformer architectures have increasingly been investigated to combine convolutional locality with global contextual modeling [33]. At the same time, explainable artificial intelligence has become important for examining whether high predictive performance is supported by meaningful image evidence [34]. LIME [35], Grad-CAM

[36], and feature-space visualization approaches such as t-SNE [37] have therefore been increasingly adopted to investigate model decisions rather than relying exclusively on classification metrics.

For medicinal leaf disease classification specifically, Leelavathi and Kalamani [38] introduced a framework involving multi-scale feature extraction, optimized feature selection, and an attention-based dilated DenseNet. More recently, Ahmmed et al. [39] proposed LSeTNet for the same three medicinal species considered in the present study. Their CNN–Transformer model achieved 99.72% accuracy, a macro F1-score of 1.00, and an AUC of 1.00 across 12 classes while requiring 9.38 million parameters and 2.50 GFLOPs. Five-fold cross-validation resulted in $99.74 \pm 0.14\%$ accuracy, while external validation produced 99.42% accuracy. The corresponding medicinal leaf disease dataset provides 2,547 original images representing healthy and diseased conditions of Kalanchoe, Neem, and Tulsi [40].

Although existing deep learning approaches have achieved high classification accuracy in medicinal plant disease recognition, most studies consider plant species and disease conditions as combined categories within a flat multiclass classification framework. These approaches have demonstrated their effectiveness in identifying different medicinal plant diseases. However, the explicit learning of species-related and disease-related characteristics through separate supervisory objectives remains comparatively less explored. Previous studies have demonstrated the potential of multi-task learning for extracting complementary plant characteristics [25]. Nevertheless, the integration of species-aware and disease-aware representations through a hierarchical learning mechanism requires further investigation. In this regard, hierarchical learning offers an alternative approach for incorporating species and disease information at different levels of the classification process, without assuming that conventional flat classification models are insufficient for achieving high predictive performance.

Considering these aspects, this study proposes a Species-Aware Hierarchical Network (SAH-Net) for medicinal plant disease recognition. The proposed framework employs an ImageNet-pretrained EfficientNet-B0 backbone to extract shared visual features, followed by two query-based attention branches for learning species-aware and disease-aware representations. These representations are subsequently combined through an adaptive gated hierarchical fusion mechanism to perform the final species–condition classification. Unlike conventional flat classification approaches, SAH-Net incorporates three interconnected learning objectives comprising species recognition, disease/condition recognition, and final species–condition diagnosis. Through this hierarchical learning strategy, the present study investigates the feasibility of explicitly incorporating species-related and disease-related information into a unified classification framework.

The major contributions of this study are summarized as follows:

- A species-aware hierarchical learning framework, termed SAH-Net, is proposed to jointly model three-class species recognition, seven-class disease/condition recognition, and 12-class species–condition diagnosis rather than treating medicinal leaf diagnosis exclusively as a flat multiclass classification problem.
- Two query-based attention branches are introduced to explicitly learn species-related and disease-related representations from a shared feature space, followed by an adaptive gated hierarchical fusion mechanism that integrates their complementary information for final classification.
- The hierarchical decision process is examined through species attention, disease attention, and Grad-CAM, providing visual evidence at different levels of the proposed architecture rather than explaining only the final prediction.
- The performance of SAH-Net is examined from different aspects, including classification, ROC-AUC, model explanation, and computational cost. On the 383 test images, the model obtained 100% accuracy using 6.019 million parameters and 0.893 GFLOPs.

2. Materials and Methods

2.1. Dataset Description

A total of 2,547 images covering three medicinal plants were considered in this work: *Kalanchoe pinnata* (PatharKuchi), *Azadirachta indica* (Neem), and *Ocimum tenuiflorum* (Tulsi). The leaf samples originated from fields and household medicinal gardens located in Senbag Upazila, Noakhali, Bangladesh, and image acquisition was carried out in August 2025. A 50 MP Honor 200 smartphone was used to photograph the leaves under daylight conditions. The acquired images had a resolution of 4000 × 3000 pixels. Instead of maintaining a fixed camera position, photographs were taken at varying distances and viewing angles. Therefore, differences in leaf orientation, illumination, and surrounding background are present among the images. The disease identification and labeling

process was conducted under the supervision of the Upazila Agriculture Officer, Senbag, and the assigned labels were verified on 19 October 2025. The complete dataset is publicly available through Mendeley Data [40].

The dataset contains 12 classes representing healthy leaves and different disease or symptom conditions associated with the three medicinal plant species. For *Kalanchoe pinnata*, the dataset contains Healthy, Web Blight, Yellow, and Yellow Blight classes, comprising a total of 839 images. Similarly, *Azadirachta indica* is represented by Healthy, Leaf Spot, Web Blight, and Yellow classes with 885 images, whereas *Ocimum tenuiflorum* contains Healthy, Downy Mildew, Web Blight, and Yellow Spot classes with a total of 823 images. Thus, the dataset provides both species-level and disease-level information, which is particularly suitable for the hierarchical learning strategy considered in this study. Unlike conventional flat classification, where each species–condition combination is treated as an independent class, the available labels allow the images to be represented simultaneously according to plant species, disease condition, and their final species–condition category. The detailed class-wise distribution of the dataset is presented in Table 1 and samples of the dataset are shown in Figure 1.

2.2. Dataset Preparation and Experimental Split

All images were resized to 224 × 224 pixels before being provided to the model, while the original images were retained without modification. A stratified 70:15:15 split was applied to the dataset. Accordingly, 1,782 images were used for training, 382 for validation, and 383 for testing. The class proportions were maintained during the splitting process, and these subsets were kept unchanged for all experiments. The training, validation, and test subsets were mutually exclusive, ensuring that each image was assigned to only one subset. Duplicate-image verification across the three subsets confirmed that no identical images were shared between training, validation, and testing.

During training, random horizontal flipping ($p = 0.5$), rotation of up to ±15°, and color jitter were applied. The brightness and contrast factors were set to 0.15, while 0.10 was used for saturation. These augmentation operations were not applied to the validation and test images. Finally, the images were converted into tensors and normalized according to the ImageNet mean and standard deviation. A random seed of 42 was maintained throughout the experiments.

Table 1. Class-wise distribution of the medicinal plant leaf dataset.

Plant species	Class	Number of images
Kalanchoe pinnata (PatharKuchi)	Healthy	209
	Web Blight	213
	Yellow	204
	Yellow Blight	213
Subtotal		839
Azadirachta indica (Neem)	Healthy	244
	Neem Leaf Spot	211
	Web Blight	205
	Yellow	225
Subtotal		885
Ocimum tenuiflorum (Tulsi)	Healthy	213
	Downy Mildew	200
	Web Blight	205
	Yellow Spot	205
Subtotal		823
Total	12 classes	2,547



Figure 1. Representative samples of the medicinal plant leaf dataset.

2.3. Proposed SAH-Net Architecture

This study proposes a Species-Aware Hierarchical Network (SAH-Net) for medicinal plant disease recognition by explicitly incorporating the relationship between plant species and disease characteristics into the learning process. Unlike conventional flat classification, where each species–condition combination is learned directly as an independent class, SAH-Net decomposes the learning process into species-aware and disease-aware representations and subsequently integrates them for final classification. The overall architecture consists of four major components: a shared visual feature encoder, two parallel attention branches, an adaptive gated hierarchical fusion module, and a final species–condition classifier.

Given an input leaf image I, the shared encoder first generates a common feature representation F. Two independent attention branches then transform F into a species-aware embedding z_s and a disease-aware embedding z_d. These representations are additionally supervised through separate species and disease classification heads. Subsequently, an adaptive gating mechanism determines the relative contribution of the two representations and generates a fused hierarchical representation z_f, which is used to predict the final species–condition class. Therefore, the overall processing flow can be summarized as

$$I \rightarrow F \rightarrow \{z_s, z_d\} \rightarrow z_f \rightarrow \hat{y},$$

where $\hat{y}$ denotes the final prediction among the 12 species–condition categories. The individual components of SAH-Net are described in the following subsections, while the complete architecture is illustrated in Figure 2.

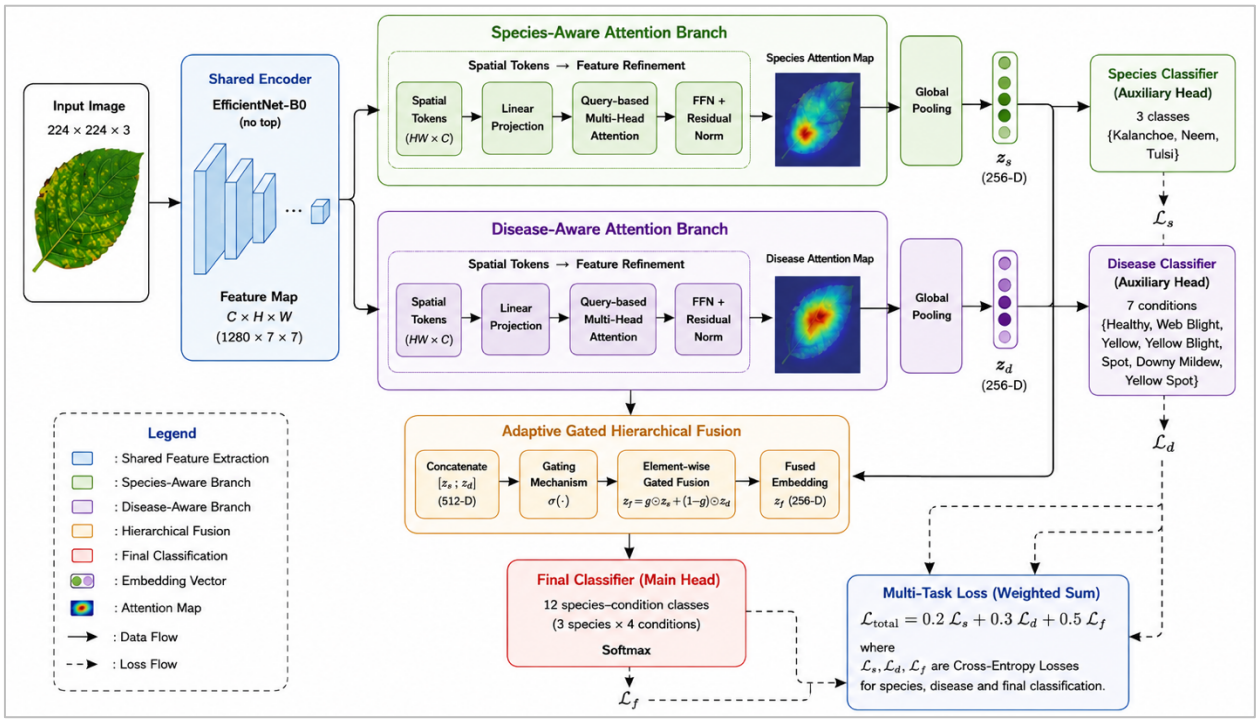


Figure 2. Architecture of proposed model: SAH-Net.

2.3.1. Shared Visual Feature Encoder

Given an input leaf image $I \in \mathbb{R}^{3\times224\times224}$, SAH-Net first extracts a common visual representation using an ImageNet-pretrained EfficientNet-B0 backbone. The original classification layer of EfficientNet-B0 is removed, allowing its convolutional layers to operate as a shared feature encoder for the hierarchical learning tasks. The feature extraction process can be expressed as

$$F = E(I),$$

where $E(\cdot)$ represents the shared EfficientNet-B0 encoder and $F \in \mathbb{R}^{1280\times H\times W}$ denotes the resulting high-level feature map. Importantly, the spatial dimensions of F are retained at this stage instead of directly applying global pooling, enabling the subsequent attention branches to operate over spatial feature tokens.

The extracted feature map F is shared between two independent branches designed to learn species-aware and disease-aware representations. This shared encoding strategy provides both branches with the same underlying visual information while allowing each branch to specialize according to its respective supervision. Consequently, species-related morphological characteristics and disease-related visual patterns can be learned separately before being adaptively combined for the final species–condition classification.

2.3.2. Species-Aware and Disease-Aware Attention Branches

The shared feature map obtained from the EfficientNet-B0 encoder is forwarded to two independent attention branches to learn species-related and disease-related information separately. Although both branches receive the same feature map F, their parameters are independently optimized through their corresponding supervision. The species-aware branch focuses on morphological characteristics associated with the three plant species, whereas the disease-aware branch learns visual patterns related to the seven disease or health conditions considered in this study.

For attention-based representation learning, the spatial feature map is first converted into a sequence of $N = H \times W$ feature tokens and projected from 1280 channels to a 256-dimensional embedding space. Let the projected token sequence be represented by

$$X = P\big(\text{Flatten}(F)\big),\ X \in \mathbb{R}^{N\times d},$$

where $P(\cdot)$ denotes the linear projection and $d = 256$ is the embedding dimension. Each branch contains a learnable query vector $q \in \mathbb{R}^{1\times d}$, which attends to the spatial tokens through an eight-head attention mechanism. The attention operation can be expressed as

$$a = \text{MHA}(q, X, X),$$

where MHA denotes multi-head attention, with the learnable query serving as the query and the projected spatial tokens acting as the keys and values. This operation enables each branch to aggregate spatial information

according to the representation required by its respective task.

The attended representation is subsequently refined using residual normalization followed by a feed-forward network consisting of two linear transformations with GELU activation and dropout. This process produces the species-aware embedding $z_s$and disease-aware embedding z_d, respectively. The two embeddings are therefore derived from a common visual feature space but learn different task-specific characteristics before hierarchical fusion.

2.3.3. Species and Disease Auxiliary Classification

The species-aware and disease-aware embeddings are independently supervised through two auxiliary classification heads. The species branch predicts three plant species, whereas the disease branch predicts seven disease or health conditions. The seven disease/condition categories comprise Healthy, Web Blight, Yellow, Yellow Blight, Leaf Spot, Downy Mildew, and Yellow Spot. These categories are derived from the original 12 species–condition classes by grouping identical condition labels across different plant species. Specifically, Healthy and Web Blight are shared across all three species, while Yellow occurs in both *Kalanchoe pinnata* and *Azadirachta indica*. Yellow Blight is associated with *Kalanchoe pinnata*, Leaf Spot with *Azadirachta indica*, and Downy Mildew and Yellow Spot with *Ocimum tenuiflorum*. Accordingly, each of the 12 species–condition classes is mapped to one species label and one disease/condition label, enabling separate supervision of the two hierarchical classification tasks.

Given the learned embeddings (z_s) and (z_d), the corresponding classification outputs are obtained as

$$[\hat{y}_s = C_s(z_s), \qquad \hat{y}_d = C_d(z_d),]$$

where $\left(C_s(\cdot)\right)$ and $\left(C_d(\cdot)\right)$ denote the species and disease classifiers, respectively.

These auxiliary classification outputs provide direct supervision to the two attention branches, encouraging them to preserve information relevant to their assigned hierarchical tasks rather than relying solely on the final 12-class objective. The resulting task-specific embeddings are subsequently forwarded to the adaptive gated fusion module for joint representation learning.

2.3.4. Adaptive Gated Hierarchical Fusion

Following task-specific representation learning, the species-aware and disease-aware embeddings are integrated through an adaptive gated hierarchical fusion module. Instead of directly concatenating the two representations for classification, SAH-Net learns a gating vector that controls their relative contribution. The two embeddings are first concatenated and transformed to obtain the gate:

$$g = \sigma\big(G([z_s; z_d])\big),$$

where $G(\cdot)$ represents the gating network, $[\cdot;\cdot]$denotes concatenation, and $\sigma$is the sigmoid activation function. The fused representation is subsequently computed as

$$z_f = g \odot z_s + (1 - g) \odot z_d,$$

where $\odot$ represents element-wise multiplication.

The gating mechanism therefore allows the network to adaptively balance species- and disease-related information at the feature level rather than assigning a fixed contribution to either representation. The resulting fused feature is further refined through a nonlinear transformation before being forwarded to the final classification stage.

2.3.5. Final Species–Condition Classification

The fused hierarchical representation z_f is used to perform the final classification across the 12 species–condition categories. Before classification, the fused representation is passed through a fully connected layer that reduces its dimensionality from 256 to 128, followed by GELU activation and dropout. The resulting feature vector is then mapped to the final class logits as

$$\hat{y} = C_f\big(z_f\big),$$

where $C_f(\cdot)$ denotes the final classification head and $\hat{y} \in \mathbb{R}^{12}$represents the logits corresponding to the 12 species–condition classes.

In this manner, the final prediction is derived from the adaptively integrated species- and disease-aware information rather than directly from the shared backbone features. This completes the hierarchical processing pipeline of SAH-Net from shared visual feature extraction to species–condition recognition.

Unlike conventional EfficientNet-B0-based classification, which directly maps extracted features to the final class labels, SAH-Net incorporates explicit species-level and disease-level supervision through two independently parameterized attention branches with task-specific supervision. These branches are designed to capture complementary representations, which are subsequently integrated through adaptive gated hierarchical fusion for the final species-condition diagnosis. The principal architectural contribution therefore lies in the explicit hierarchical formulation of the classification problem rather than relying solely on conventional flat classification. Nevertheless, the present study does not include controlled baseline or ablation experiments to quantify the individual contributions of the attention branches and fusion mechanism. Accordingly,

their independent effects on classification performance cannot be established from the current results and require further investigation.

2.4. Hierarchical Multi-Task Learning Objective

SAH-Net is trained using a hierarchical multi-task learning objective that jointly supervises species recognition, disease recognition, and final species–condition classification. Cross-entropy loss is independently computed for the three prediction tasks and can be expressed as

$$[\mathcal{L}s = \mathrm{CE}(\hat{y}s, y_s), \quad \mathcal{L}d = \mathrm{CE}(\hat{y}d, y_d), \quad \mathcal{L}f = \mathrm{CE}(\hat{y}, yf),]$$

where (y_s), (y_d), and (y_f) denote the ground-truth species, disease, and final species–condition labels, respectively. The three losses are jointly optimized using the weighted objective

$$[\mathcal{L}_{\text{total}} = \lambda_s \mathcal{L}_s + \lambda_d \mathcal{L}_d + \lambda_f \mathcal{L}_f,]$$

where (λ_s), (λ_d), and (λ_f) control the contribution of the individual tasks. In the present experiments, these weights were heuristically assigned as 0.2, 0.3, and 0.5, respectively, based on the relative importance of the three classification objectives. The highest weight was assigned to the final 12-class species–condition classification task, which represents the primary diagnostic objective. A relatively higher weight was assigned to disease/condition recognition than to species recognition to emphasize disease-related characteristics while maintaining explicit species-level supervision. These weights reflect the intended task priorities rather than an experimentally optimized weighting configuration.

The joint objective enables the shared encoder and task-specific attention branches to be optimized within a single end-to-end learning framework. Consequently, the auxiliary tasks guide the network toward learning the intended hierarchical representations, while the final classification loss optimizes their combined representation for species–condition recognition.

2.5. Experimental Setup

All experiments were conducted using Python 3.13.13 and PyTorch 2.13.0, together with torchvision 0.28.0 and timm 1.0.28. The experiments were performed on a MacBook Pro equipped with an Apple M5 Pro chip with 18 CPU cores and 64 GB of unified memory, while model training was accelerated through the Apple Metal Performance Shaders (MPS) backend. A fixed random seed of 42 was maintained throughout the experimental pipeline to improve the reproducibility of the experiments.

SAH-Net was initialized using an ImageNet-pretrained EfficientNet-B0 backbone and optimized using AdamW with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-4}. A batch size of 16 was used, and the model was trained for a maximum of 50 epochs. The learning rate was adaptively reduced by a factor of 0.5 when the validation loss did not improve for two consecutive epochs.

Model selection was based on the total validation loss, with the checkpoint corresponding to the lowest validation loss retained as the best-performing model. Early stopping with a patience of seven epochs was additionally employed to terminate the training process when no further improvement in the validation loss was observed. Following completion of the training process, the best checkpoint was restored and evaluated on the held-out test subset to obtain the final performance results.

The classification performance was evaluated using accuracy, precision, recall, and F1-score. In addition to the overall classification accuracy, macro-averaged precision, recall, and F1-score were computed to assign equal importance to each disease class irrespective of the number of samples available in individual classes. A confusion matrix was further generated to provide a class-wise assessment of the predictions and identify possible misclassification patterns among visually similar disease categories. These evaluation measures were calculated exclusively on the held-out test subset using the best model checkpoint selected based on the validation loss.

3. Results and Discussion

3.1. Training and Validation Performance

Figure 3 presents the learning behaviour of SAH-Net during training. The initial training loss was 1.23, which decreased considerably during the first few epochs. The validation loss followed the same trend, decreasing from 0.53 to less than 0.1 during the early stage of training. Subsequently, only small variations were observed, and the loss values remained close to zero for most of the remaining epochs. This shows that most of the learning occurred during the early epochs, after which the model reached a stable state.

For the final 12-class task, the training accuracy increased from approximately 54% in the first epoch to above 94% in the second epoch. During the same two epochs, the validation accuracy improved from approximately 92% to above 98%. The rapid increase in accuracy is consistent with the reduction observed in the loss values. After the initial epochs, both training and validation accuracies remained high with relatively small variations until the training process was completed.

Thereafter, both curves rapidly approached 100% and remained almost stable for the remaining epochs. Although minor variations can be observed in both curves, no persistent reduction in the validation accuracy was recorded as the training progressed. These results are consistent with the loss behaviour presented in Figure 3(a) and demonstrate that the proposed model was able to learn the discriminative characteristics of the medicinal leaf images within a relatively small number of epochs.

Another important observation from Figure 3 is the small difference between the training and validation curves after convergence. The validation loss did not show a progressive increase when the training loss approached zero, while the validation accuracy remained comparable to the training accuracy. This indicates that the optimization of SAH-Net remained stable during the later stages of training. The obtained convergence can be attributed to the joint optimization of the species, disease/condition, and final species-condition classification objectives, through which the shared representation is simultaneously guided by different levels of supervisory information. The generalization capability of the trained model is further examined on the independent test set in the subsequent section.

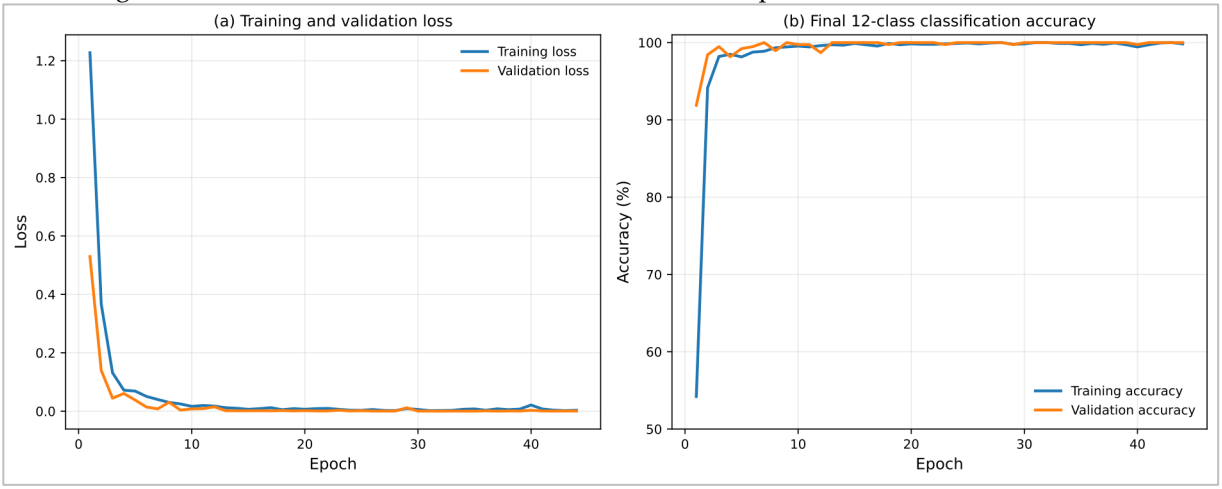


Figure 3. Training and validation performance of the proposed SAH-Net: (a) training and validation loss and (b) training and validation accuracy for the final 12-class species-condition classification task.

3.2. Hierarchical Classification Performance

The classification performance of the proposed SAH-Net was further evaluated at the three hierarchical levels, namely species recognition, disease/condition recognition, and final species-condition diagnosis. As presented in Table 2, the model achieved an accuracy of 1.000 for all three classification tasks on the independent test set comprising 383 images. Similarly, the macro precision, macro recall, macro F1-score, and weighted F1-score reached 1.000 at each hierarchical level. These results indicate that the learned representations were sufficiently discriminative not only for distinguishing the three medicinal plant species but also for identifying the seven disease/condition categories and their corresponding 12 species-condition combinations.

The class-wise performance of the final diagnostic task is further illustrated through the confusion matrix in Figure 4. All 383 test images were correctly assigned to their respective classes, with no misclassification observed among any of the 12 species-condition categories. The diagonal entries ranged from 30 correctly classified samples for Kalanchoe Yellow, Neem Web Blight, and Tulsi Downy Mildew to 37 samples for Neem Healthy, reflecting the variation in class-wise test support. More importantly, all off-diagonal entries remained zero, demonstrating that the performance was consistently maintained across the complete set of classes rather than being dominated by classes containing a relatively larger number of samples.

The perfect classification performance across the three hierarchical tasks may be partly attributed to the visually distinguishable characteristics of the medicinal plant species and their associated disease conditions. Inspection of the dataset samples suggests noticeable differences in leaf morphology, discoloration, lesion patterns, and disease symptoms, which may have facilitated discrimination among the evaluated classes. Furthermore, the training, validation, and test images originated from the same dataset and acquisition context, potentially limiting the variability encountered during evaluation. Although the species-aware and disease-aware branches, together with gated hierarchical fusion, supported accurate

classification of all 383 test images, their individual contributions cannot be established from these results alone. Therefore, the achieved performance should be interpreted within the evaluated test partition rather than as evidence of equivalent performance under different acquisition conditions or real-world field environments. External validation using more heterogeneous datasets would be necessary to establish the generalization capability of SAH-Net.

Table 2. Hierarchical classification performance of SAH-Net

Task	Classes	Test samples	Accuracy	Macro Precision	Macro Recall	Macro F1	Weighted F1
Species recognition	3	383	1.0	1.0	1.0	1.0	1.0
Disease/condition recognition	7	383	1.0	1.0	1.0	1.0	1.0
Final species-condition diagnosis	12	383	1.0	1.0	1.0	1.0	1.0

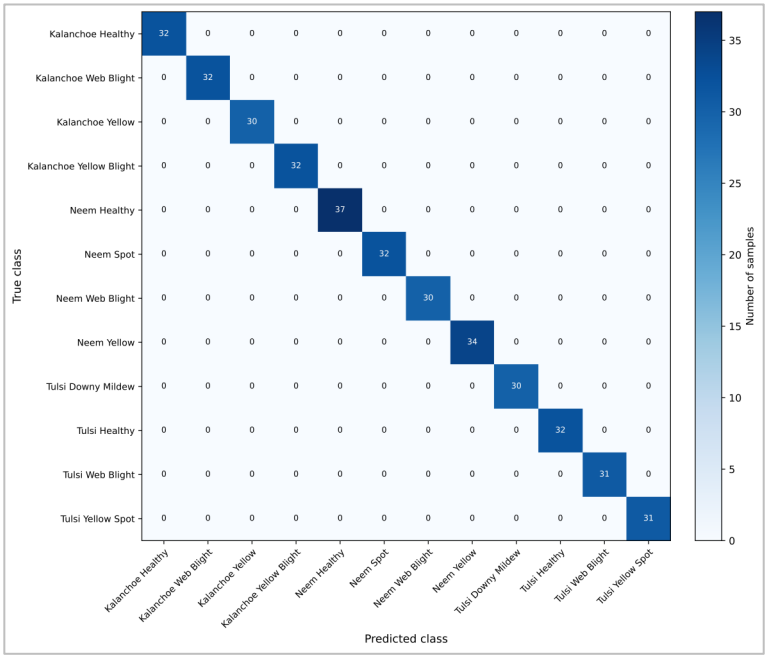


Figure 4. Confusion matrix for the final 12-class diagnosis.

3.3 ROC and AUC Analysis

The discriminative capability of the proposed SAH-Net was further examined using receiver operating characteristic (ROC) analysis for the final species-condition classification task. Figure 5 presents the one-versus-rest ROC curves for all 12 classes together with the macro- and micro-average curves. The proposed model achieved an AUC of 1.000 for each individual class, while both the macro-average and micro-average AUC also reached 1.000. The ROC curves are concentrated at the upper-left boundary of the plot and remain substantially above the diagonal random-classifier reference, indicating complete separation between the positive and negative samples within the evaluated test set.

The consistency of the class-wise AUC values is particularly important considering that the final classification task combines three medicinal plant species with different disease and health conditions. No reduction in discriminative performance was observed for classes sharing similar disease terminology across species, such as Web Blight, or for visually related leaf conditions. Furthermore, the agreement between the macro- and micro-average AUC values indicates that the observed performance was maintained across classes despite the differences in their test sample frequencies. These findings are consistent with the confusion-matrix results presented in Figure 4 and further demonstrate the strong discriminative capability of the learned species-condition representations on the present test dataset.

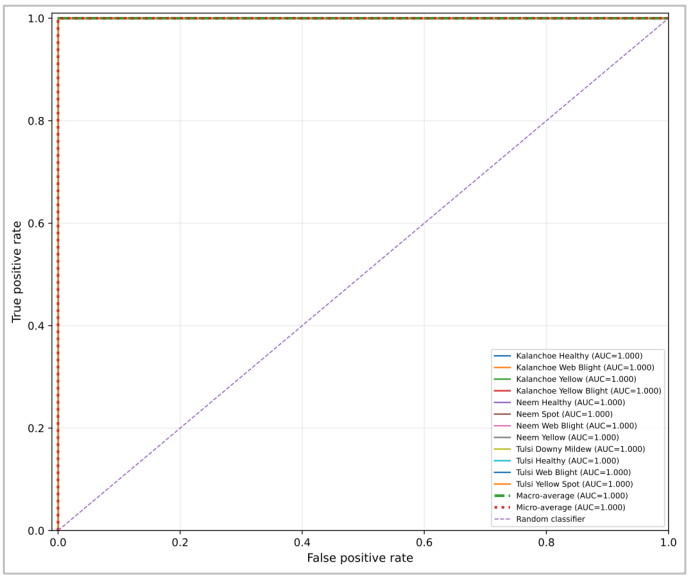


Figure 5. Class-wise ROC curves of SAH-Net for the 12 species-condition categories.

3.4 Explainability Analysis

To further examine the decision-making behavior of the proposed SAH-Net, the learned attention patterns were visualized at different stages of the hierarchical classification process. Figure 6 presents three illustrative diseased leaf samples, namely Kalanchoe Yellow Blight, Neem Spot, and Tulsi Downy Mildew, together with their corresponding species-attention, disease-attention, and Grad-CAM visualizations. These examples were selected to include one diseased sample from each medicinal plant species and to examine the attention patterns associated with different disease manifestations. The selected samples provide a qualitative illustration of the model's attention behavior rather than a comprehensive assessment across all 12 species–condition categories.

The species-attention maps presented in Figure 6 show that the highlighted regions vary across the three examined leaf images. In the Kalanchoe Yellow Blight sample, higher attention is concentrated around the discolored and damaged portion of the leaf. For Neem Spot, the attention is primarily localized within the central region of the leaf, whereas the Tulsi Downy Mildew sample shows a comparatively broader attention distribution across the leaf surface. Similar spatial patterns can be observed in the disease-attention maps, where the highlighted regions largely overlap with those identified by the species-attention branch. For all three illustrated samples, both auxiliary branches produced the correct predictions with reported confidence values of 1.000. These observations suggest that the two branches assign greater attention to selected leaf regions, although the similarity between their attention maps does not independently establish that they learn distinct visual characteristics.

The Grad-CAM visualizations further illustrate the spatial activation patterns associated with the final species–condition predictions. For Kalanchoe Yellow Blight, the major activation region overlaps with the visibly damaged portion of the leaf, while the Neem Spot sample shows stronger activation around the central leaf region containing visible symptoms. In the Tulsi Downy Mildew sample, the activation is more prominent toward the lower portion of the leaf and extends beyond the regions highlighted by the auxiliary attention branches. These differences illustrate that the spatial emphasis associated with the final classification output is not necessarily identical to that of the species-aware and disease-aware attention branches.

Overall, the explainability results provide a qualitative illustration of the spatial regions highlighted by SAH-Net for the three selected samples. Although several highlighted regions correspond to visible disease-related characteristics, these observations should not be interpreted as evidence of consistent attention behavior or reliable interpretability across the complete dataset. Moreover, visual correspondence between attention maps and disease symptoms does not establish a causal relationship between the highlighted regions and the predicted classes. A more comprehensive evaluation involving additional samples from all species–condition categories and quantitative assessment of explanation consistency would be necessary to establish the reliability and generalizability of the observed attention patterns.

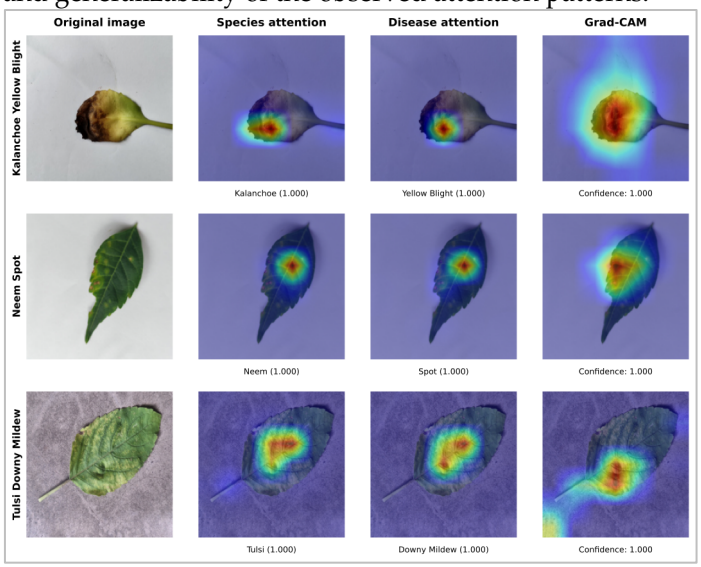


Figure 6. Explainability analysis of SAH-Net using species attention, disease attention, and Grad-CAM.

3.5. Computational Efficiency

In addition to classification performance, the computational characteristics of SAH-Net were evaluated to examine the resource requirements of the proposed architecture. As presented in Table 3, SAH-Net contains 6.019 million parameters and requires 0.447 GMACs, corresponding to approximately 0.893 GFLOPs for a single forward pass. The trained model occupies only 23.25 MB of storage, indicating that the incorporation of the species-aware and disease-aware attention branches together with the gated hierarchical fusion mechanism does not result in an excessively large model.

The inference performance of SAH-Net was further evaluated using the CPU and Apple Metal Performance Shaders (MPS) on a MacBook Pro equipped with an Apple M5 Pro processor. The model achieved an average inference latency of 194.00 ms per image on the CPU, corresponding to a throughput of 5.15 images/s. When evaluated using MPS acceleration, the inference latency decreased to 7.231 ms per image, with a corresponding throughput of 138.29 images/s. These results indicate that SAH-Net can achieve comparatively faster inference using hardware acceleration within the evaluated computing environment. However, the reported latency and throughput values are specific to the MacBook Pro/Apple M5 Pro hardware and software configuration used in this study. Therefore, these measurements should not be interpreted as device-independent deployment performance, as inference efficiency may vary across different processors, hardware accelerators, and deployment environments. Further evaluation on alternative computing platforms, including resource-constrained devices, would be necessary to establish the practical deployment efficiency of the proposed model.

Table 3. Computational profile of the proposed SAH-Net.

Metric	SAH-Net
Parameters	6.019 M
MACs	0.447 GMACs
FLOPs	0.893 GFLOPs
Checkpoint size	23.25 MB
CPU latency	194.00 ms/image
CPU throughput	5.15 images/s
MPS latency	7.231 ms/image
MPS throughput	138.29 FPS

3.6. Comparison with Existing Studies

To further position the performance of SAH-Net, the obtained results were compared with representative studies on medicinal plant leaf identification and disease classification, as summarized in Table 4. Tiwari et al.[13] developed a deep neural network for the classification of 22 medicinal plant categories and reported an average test accuracy of 97.69% on unseen images, together with a five-fold cross-validation accuracy of 98.68%. Similarly, Kayaalp [16] employed several pretrained deep learning architectures for medicinal plant type and disease classification and obtained a test accuracy of 98.69% using DenseNet121 across ten categories. These studies demonstrate the effectiveness of conventional deep CNN architectures; however, the plant species and disease information are generally handled within a flat classification framework.

More recently, Leelavathi and Kalamani [38] introduced an optimization-based A-DADenseNet framework for medicinal plant leaf disease classification using preprocessing, feature selection, and attention-based classification. Their proposed framework reported an overall accuracy of 89.14%. Reddy et al. [41] proposed BiFPNet for medicinal leaf identification by integrating EfficientNet-B0 with bidirectional feature-pyramid fusion. The model was evaluated on 12,813 images from 98 medicinal plant species and achieved an accuracy of 97%, while Grad-CAM was incorporated to improve the interpretability of the predictions. Although this study considers a substantially larger number of species, its primary objective is species identification rather than joint species and disease recognition.

Among the existing studies, LSeTNet provides the closest comparison with the present work because it considers the same three medicinal plants, namely Kalanchoe, Neem, and Tulsi, across 12 disease-related classes. Ahmmed et al.[39] reported an accuracy of 99.72%, macro F1-score of 1.00, and AUC of 1.00 using a lightweight CNN–Transformer architecture with Squeeze-and-Excitation modules. The model contains 9.38 million parameters and requires 2.50 GFLOPs. Furthermore, five-fold cross-validation resulted in an average accuracy of $99.74 \pm 0.14\%$, and additional external validation was performed on the BD-MediLeaves dataset.

In comparison, the proposed SAH-Net achieved 100% accuracy, macro precision, macro recall, and macro F1-score on the independent 383-image test subset while requiring 6.019 million parameters and 0.893 GFLOPs. Unlike the existing approaches, SAH-Net explicitly decomposes the learning process into species-aware and

disease-aware representations and subsequently integrates them through adaptive gated hierarchical fusion. The proposed framework therefore differs not only in its classification performance but also in the manner through which species morphology and disease-related visual information are incorporated into the final diagnosis. Furthermore, the use of species attention, disease attention, and Grad-CAM provides multiple levels of visual interpretation corresponding to the hierarchical structure of the network.

Nevertheless, the numerical values in Table 4 should not be interpreted as a direct ranking of the models because the studies employ different datasets, class definitions, data partitions, and evaluation protocols. This consideration is particularly important when comparing SAH-Net with LSeTNet. Although SAH-Net achieved a slightly higher accuracy on the present test partition with lower computational complexity, LSeTNet additionally employed five-fold cross-validation and external validation, providing stronger evidence regarding performance stability across different data partitions and datasets. Therefore, the principal contribution of SAH-Net should be viewed in terms of its species-aware hierarchical learning strategy, compact computational profile, and architecture-specific interpretability, rather than solely the difference in classification accuracy.

Table 4. Comparison of SAH-Net with existing medicinal plant leaf classification studies.

Study	Task / classes	Method	Accuracy (%)	XAI	Main characteristic
Tiwari et al. (2022) [13]	Medicinal plant leaf classification, 22 categories	Deep neural network	97.69	No	Five-fold CV and independent testing
Kayaalp (2024) [16]	Plant type and disease classification, 10 categories	DenseNet121	98.69	No	Pretrained CNN-based classification
Leelavathi & Kalamani (2025) [38]	Medicinal leaf disease classification	HZKHO–A-DADenseNet	89.14	Attention mechanism	Optimized feature selection and classification
Ahmmed et al. (2026) [39]	Kalanchoe, Neem and Tulsi, 12 classes	LSeTNet	99.72	Grad-CAM, LIME, t-SNE	Lightweight CNN–Transformer; external validation
Reddy et al. (2026) [41]	Medicinal species identification, 98 classes	BiFPNet	97.00	Grad-CAM	Large-scale medicinal species identification
Proposed SAH-Net	Kalanchoe, Neem and Tulsi, 12 species-condition classes	Species-aware hierarchical learning	100.00	Species attention, disease attention, Grad-CAM	Explicit species/disease learning with gated hierarchical fusion

4. Conclusion

This study proposed SAH-Net, a species-aware hierarchical deep learning framework for medicinal plant disease recognition. Unlike conventional flat classification approaches, the proposed model explicitly learns species-related and disease-related visual representations through two independent query-based attention branches and subsequently integrates them using an adaptive gated hierarchical fusion mechanism. The network was evaluated on 2,547 medicinal leaf images representing Kalanchoe pinnata, Azadirachta indica, and Ocimum tenuiflorum across 12 species-condition classes. On the independent test subset of 383 images, SAH-Net achieved an accuracy, macro precision, macro recall, macro F1-score, and weighted F1-score of 1.000 for species recognition, disease/condition recognition, and final species-condition diagnosis. The class-wise ROC analysis further produced an AUC of 1.000 for all 12 classes, with both macro- and micro-average AUCs reaching 1.000.

In addition to predictive performance, the proposed model maintained a relatively compact computational profile with 6.019 million parameters, 0.447 GMACs, and 0.893 GFLOPs, while the trained checkpoint required only 23.25 MB of storage. The model achieved an inference latency of 7.231 ms/image on MPS, corresponding to 138.29 FPS, whereas CPU inference required 194.00 ms/image. The explainability analysis further showed that the species-attention and disease-attention branches concentrated on discriminative leaf regions, while Grad-CAM provided complementary information regarding the areas contributing to the final species-condition prediction. These findings demonstrate that hierarchical supervision can be effectively incorporated into medicinal leaf disease recognition without introducing excessive computational complexity.

Although the proposed SAH-Net achieved promising results, certain limitations of the present study need to be considered. First, the experimental evaluation was limited to three medicinal plant species and 12 species–condition classes. Therefore, the effectiveness of the proposed model for identifying other medicinal plants and disease conditions remains to be investigated. Another important limitation is that all images used for training, validation, and testing were obtained from the same dataset and acquisition context. Although the dataset was divided into mutually exclusive subsets, the images may still share similar characteristics in terms of illumination, background, camera settings, and environmental conditions. Consequently, the 100% classification accuracy achieved on the 383 test images does not necessarily indicate that the proposed model will maintain similar performance when evaluated on images collected under different field conditions. Factors such as variations in lighting, disease severity, image quality, and background complexity may influence the recognition performance of SAH-Net in real-world environments. Furthermore, the performance of SAH-Net under natural canopy conditions, including overlapping leaves, partial occlusions, and complex vegetation backgrounds, has not been independently evaluated in the present study.

Therefore, further evaluation using independently collected datasets from different geographical locations and acquisition conditions is required to establish the generalization capability of the proposed model. In future work, we intend to extend the present study by incorporating additional medicinal plant species and disease categories, evaluating the model using more diverse field images, and investigating its implementation on mobile and edge devices for practical medicinal plant disease recognition.

Author Contributions: W.H.L. and M.M.N. contributed equally to the conceptualization, methodology, software development, validation, formal analysis, investigation, data curation, visualization, writing—original draft preparation, and writing—review and editing. Both authors have read and approved the final version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This study used a publicly available Medicinal Plant Leaf Disease Dataset developed by Ahmmed et al. [40], available through Mendeley Data at https://doi.org/10.17632/ncg7kk3gwx.1.

Acknowledgments: None.

Conflicts of Interest: The authors declare no conflicts of interest.

AI Use Declaration: Generative AI was used solely for language editing and proofreading. The authors reviewed and approved all content and take full responsibility for the manuscript.

References:

1. Manisha, M.; Babu, R.; Begam, A.M.; Shakya Chahal, K.; Ashok Harale, A. Medicinal Plants and Traditional Uses and Modern Applications. *Journal of Neonatal Surgery* **2025**, *14*, 162–175 https://doi.org/10.52783/JNS.V14.2210.
2. Kussmann, M.; Abe Cunha, D.H.; Berciano, S. Bioactive Compounds for Human and Planetary Health. *Frontiers in Nutrition* **2023**, *10*, 1193848 https://doi.org/10.3389/FNUT.2023.1193848.
3. Sasidharan, S.P.; Viji, Z.; Sankaran, R.P.; Muraleedharan, S.; K, J.; R, P.; S, B.; S, S.; S, S.; P, A.; et al. Exploring Folklore Herbal Knowledge: A Compilation from Palakkad District, Kerala, India: Exploring Folklore Herbal Knowledge from Palakkad District. *Ethnobotany Research and Applications* **2024**, *27*, 1–28 https://doi.org/10.32859/era.27.23.1-28.
4. Tyagi, A.; Lama Tamang, T.; Kashtoh, H.; Mir, R.A.; Mir, Z.A.; Manzoor, S.; Manzar, N.; Gani, G.; Vishwakarma, S.K.; Almalki, M.A.; et al. A Review on Biocontrol Agents as Sustainable Approach for Crop Disease Management: Applications, Production, and Future Perspectives. *Horticulturae* **2024**, *10*, 805 https://doi.org/10.3390/HORTICULTURAE10080805.
5. Bhattarai, K.; Bhattarai, R.; Pandey, R.D.; Paudel, B.; Bhattarai, H.D. A Comprehensive Review of the Phytochemical Constituents and Bioactivities of Ocimum Tenuiflorum. *The Scientific World Journal* **2024**, *2024*, 8895039 https://doi.org/10.1155/2024/8895039.
6. Mehnaz, S.; Shamsi, Y.; Akhtar, M.W.; Mohanty, S.; Ahmad, S. Therapeutic Applications of Neem (Azadirachta Indica): A Narrative Review. *Advances in Mind-body Medicine* **2024**, *38*, 14–18.
7. Hernández-Caballero, M.E.; Sierra-Ramírez, J.A.; Villalobos-Valencia, R.; Seseña-Méndez, E. Potential of Kalanchoe Pinnata as a Cancer Treatment Adjuvant and an Epigenetic Regulator. *Molecules* **2022**, *27*, 6425 https://doi.org/10.3390/MOLECULES27196425.
8. Pushpanathan, K.; Hanafi, M.; Mashohor, S.; Fazlil Ilahi, W.F. Machine Learning in Medicinal Plants Recognition: A Review. *Artificial Intelligence Review* **2020**, *54*, 305–327 https://doi.org/10.1007/S10462-020-09847-0.
9. Mukherjee, G.; Tudu, B.; Chatterjee, A. A Convolutional Neural Network-Driven Computer Vision System toward Identification of Species and Maturity Stage of Medicinal Leaves: Case Studies with Neem, Tulsi and Kalmegh Leaves. *Soft Computing* **2021**, *25*, 14119–14138 https://doi.org/10.1007/s00500-021-06139-9.
10. Naeem, S.; Ali, A.; Chesneau, C.; Tahir, M.H.; Jamal, F.; Sherwani, R.A.K.; Hassan, M.U. The Classification of Medicinal Plant Leaves Based on Multispectral and Texture Feature Using Machine Learning Approach. *Agronomy* **2021**, *11*, 263 https://doi.org/10.3390/AGRONOMY11020263.

11. Kan, H.X.; Jin, L.; Zhou, F.L. Classification of Medicinal Plant Leaf Image Based on Multi-Feature Extraction. *Pattern Recognition and Image Analysis* **2017**, *27*, 581–587 https://doi.org/10.1134/S105466181703018X.
12. Sachar, S.; Kumar, A. Deep Ensemble Learning for Automatic Medicinal Leaf Identification. *International Journal of Information Technology (Singapore)* **2022**, *14*, 3089–3097 https://doi.org/10.1007/s41870-022-01055-z.
13. Tiwari, V.; Joshi, R.C.; Dutta, M.K. Deep Neural Network for Multi-Class Classification of Medicinal Plant Leaves. *Expert Systems* **2022**, *39* https://doi.org/10.1111/exsy.13041.
14. Meenakshi Automatic Detection of Diseases in Leaves of Medicinal Plants Using Modified Logistic Regression Algorithm. *Wireless Personal Communications* **2023**, *131*, 2573–2597 https://doi.org/10.1007/s11277-023-10555-5.
15. Blesslin Elizabeth CP, Baulkani S. Novel Network for Medicinal Leaves Identification. *IETE Journal of Research* **2023**, *69*, 1772–1782 https://doi.org/10.1080/03772063.2021.2016504.
16. Kayaalp, K. Classification of Medicinal Plant Leaves for Types and Diseases with Hybrid Deep Learning Methods. *Information Technology and Control* **2024**, *53*, 19–36 https://doi.org/10.5755/j01.itc.53.1.34345.
17. Azadnia, R.; Noei-Khodabadi, F.; Moloudzadeh, A.; Jahanbakhshi, A.; Omid, M. Medicinal and Poisonous Plants Classification from Visual Characteristics of Leaves Using Computer Vision and Deep Neural Networks. *Ecological Informatics* **2024**, *82*, 102683 https://doi.org/10.1016/J.ECOINF.2024.102683.
18. Dey, B.; Ferdous, J.; Ahmed, R.; Hossain, J. Assessing Deep Convolutional Neural Network Models and Their Comparative Performance for Automated Medicinal Plant Identification from Leaf Images. *Heliyon* **2024**, *10*, e23655 https://doi.org/10.1016/j.heliyon.2023.e23655.
19. Kavitha, S.; Kumar, T.S.; Naresh, E.; Kalmani, V.H.; Bamane, K.D.; Pareek, P.K. Medicinal Plant Identification in Real-Time Using Deep Learning Model. *SN Computer Science* **2023**, *5*, 73- https://doi.org/10.1007/S42979-023-02398-5.
20. Panchal, A. V.; Patel, S.C.; Bagyalakshmi, K.; Kumar, P.; Khan, I.R.; Soni, M. Image-Based Plant Diseases Detection Using Deep Learning. *Materials Today: Proceedings* **2023**, *80*, 3500–3506 https://doi.org/10.1016/J.MATPR.2021.07.281.
21. Kini, A.S.; Prema, K. V.; Pai, S.N. Early Stage Black Pepper Leaf Disease Prediction Based on Transfer Learning Using ConvNets. *Scientific Reports* **2024**, *14*, 1404- https://doi.org/10.1038/s41598-024-51884-0.
22. Lanjewar, M.G.; Morajkar, P.; P, P. Modified Transfer Learning Frameworks to Identify Potato Leaf Diseases. *Multimedia Tools and Applications 2023 83:17* **2023**, *83*, 50401–50423 https://doi.org/10.1007/S11042-023-17610-0.
23. Panchbhai, K.G.; Lanjewar, M.G. Enhancement of Tea Leaf Diseases Identification Using Modified SOTA Models. *Neural Computing and Applications 2024 37:4* **2024**, *37*, 2435–2453 https://doi.org/10.1007/S00521-024-10758-2.
24. Sharma, S.; Vardhan, M. AELGNet: Attention-Based Enhanced Local and Global Features Network for Medicinal Leaf and Plant Classification. *Computers in Biology and Medicine* **2025**, *184* https://doi.org/10.1016/j.compbiomed.2024.109447.
25. Sharma, S.; Vardhan, M. MTJNet: Multi-Task Joint Learning Network for Advancing Medicinal Plant and Leaf Classification. *Knowledge-Based Systems* **2024**, *299*, 112147 https://doi.org/10.1016/J.KNOSYS.2024.112147.
26. Pushpa, B.R.; Jyothsna, S.; Lasya, S. HybNet: A Hybrid Deep Models for Medicinal Plant Species Identification. *MethodsX* **2025**, *14*, 103126 https://doi.org/10.1016/J.MEX.2024.103126.
27. Diwedi, H.K.; Misra, A.; Tiwari, A.K. CNN-Based Medicinal Plant Identification and Classification Using Optimized SVM. *Multimedia Tools and Applications 2023 83:11* **2023**, *83*, 33823–33853 https://doi.org/10.1007/S11042-023-16733-8.
28. Talaat, F.M.; Shams, M.Y.; Gamel, S.A.; ZainEldin, H. DeepLeaf: An Optimized Deep Learning Approach for Automated Recognition of Grapevine Leaf Diseases. *Neural Computing and Applications 2025 37:15* **2025**, *37*, 8799–8823 https://doi.org/10.1007/S00521-025-11038-3.
29. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in plant science* **2016**, *7*, 1419.
30. Arnal Barbedo, J.G. Plant Disease Identification from Individual Lesions and Spots Using Deep Learning. *Biosystems Engineering* **2019**, *180*, 96–107 https://doi.org/10.1016/J.BIOSYSTEMSENG.2019.02.002.
31. Upadhyay, A.; Chandel, N.S.; Singh, K.P.; Chakraborty, S.K.; Nandede, B.M.; Kumar, M.; Subeesh, A.; Upendar, K.; Salem, A.; Elbeltagi, A. Deep Learning and Computer Vision in Plant Disease Detection: A Comprehensive Review of Techniques, Models, and Trends in Precision Agriculture. *Artificial Intelligence Review 2025 58:3* **2025**, *58*, https://doi.org/10.1007/S10462-024-11100-X.
32. Khan, A.; Rauf, Z.; Sohail, A.; Khan, A.R.; Asif, H.; Asif, A.; Farooq, U. A Survey of the Vision Transformers and Their CNN-Transformer Based Variants. *Artificial Intelligence Review 2023 56:3* **2023**, *56*, 2917–2970 https://doi.org/10.1007/S10462-023-10595-0.
33. Shao, R.; Bi, X.J.; Chen, Z. Hybrid ViT-CNN Network for Fine-Grained Image Classification. *IEEE Signal Processing Letters* **2024**, *31*, 1109–1113 https://doi.org/10.1109/LSP.2024.3386112.
34. Chamola, V.; Hassija, V.; Sulthana, A.R.; Ghosh, D.; Dhingra, D.; Sikdar, B. A Review of Trustworthy and Explainable Artificial Intelligence (XAI). *IEEE Access* **2023**, *11*, 78994–79015 https://doi.org/10.1109/ACCESS.2023.3294569.
35. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should i Trust You?" Explaining the Predictions of Any Classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* **2016**, *13-17-August-2016*, 1135–1144 https://doi.org/10.1145/2939672.2939778.
36. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017; pp. 618–626.
37. Van Der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *Journal of Machine Learning Research* **2008**, *9*, 2579–2605.
38. Leelavathi, R.; Kalamani, M. Medicinal Plant Leaf Disease Classification Using Optimal Weighted Features with Dilated Adaptive DenseNet and Attention Mechanism. *Scientific Reports 2025 15:1* **2025**, *15*, 36852 https://doi.org/10.1038/s41598-025-20629-y.
39. Ahmmed, J.; Kabir, M.A.; Rehman, A. ur; Bermak, A. A Lightweight Hybrid CNN and Transformer Model for Medicinal Leaf Disease Classification with Explainable AI. *Scientific Reports* **2026**, *16*, 8243 https://doi.org/10.1038/s41598-026-39182-3.

40. Ahmmed, J.; Kabir, M.A.; Rehman, A. ur; Bermak, A. Medicinal Plant Leaf Disease Dataset. **2025**, *1* https://doi.org/10.17632/NCG7KK3GWX.1.
41. Padmanabha Reddy; Poojith, J.; Sidhartha, P.; Rama Devi; Kishan Babu, K. BiFPNet: Bidirectional Feature Pyramid Network for Medicinal Leaf Identification with Therapeutic Usage. *International Journal of Computers and Applications* **2026**, *48*, 127–143 https://doi.org/10.1080/1206212X.2025.2608641.

Disclaimer: All views, interpretations, and data presented in Impaxon publications are the sole responsibility of the respective authors. These do not necessarily reflect the opinions of Impaxon or its editorial team. Impaxon and its editors assume no liability for any harm or loss arising from the use of information, procedures, or materials discussed in the published content.

Publisher's Note: Impaxon remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.