Accurate and Efficient Indian Medicinal Leaf Classification Using Lightweight Deep Learning Architectures: A Comparative Study of the Impact of Fine-Tuning

Ashika K A[1*], S Veni[2]
[1,2]Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India.
E-mail ID's :- ka.ashika21@gmail.com , venics@kahedu.edu.in

Abstract
Accurate identification of medicinal plant leaves is crucial for their effective and safe utilization in traditional and modern medicine. Despite transfer learning's widespread application in medicinal leaf classification, the impact of fine-tuning different lightweight convolutional neural network (CNN) architectures with a unique fine-tuning strategy (without the use of specific hyper-parameter) is still not well understood. This study evaluated three lightweight CNN architectures: MobileNetV2, MobileNetV4 Conv Small and ShuffleNetV2. Each model was trained using a dataset of medicinal leaves from India comprising 5,627 images spanning 56 species. Initial training involved transfer learning utilizing ImageNet pre-trained weights, followed by fine-tuning within the identical experimental framework. Evaluation metrics included Accuracy, Macro Precision, Macro Recall, Macro F1-score, Cohen's Kappa and Macro AUC. Fine-tuning consistently boosted the performance of all three architectures while maintaining resource efficiency. The best-performing model was MobileNetV4 Conv Small, improving in accuracy from 92.85% (transfer learning) to 96.42% and in macro F1-score from 92.86% (transfer learning) to 96.23% (fine tuning). While ShuffleNetV2 had the lowest baseline performance during transfer learning, fine-tuning resulted in the largest improvement, matching the classification performance of MobileNetV4 Conv Small with fewer parameters and lower GPU memory consumption. Further assessment of model size, inference speed, throughput and memory consumption indicated the suitability of lightweight CNNs for resource-limited applications. This study illustrates that architecture-specific fine-tuning improves the performance of lightweight CNNs for medicinal leaf classification. The confusion matrix and Grad-CAM visualizations confirmed that the models effectively generated reliable predictions by focusing on critical leaf regions.

Keywords: Medicinal leaf classification, Transfer learning, Fine-tuning, MobileNetV2, MobileNetV4 Conv Small, ShuffleNetV2, Lightweight convolutional neural networks, Grad-CAM

How to cite this article: Ashika KA, Veni S. Accurate and Efficient Indian Medicinal Leaf Classification Using Lightweight Deep Learning Architectures: A Comparative Study of the Impact of Fine-Tuning. Int J Drug Deliv Technol. 2026;16(67s):807-820. DOI: 10.25258/ijddt.16.67s.66

1. Introduction

Due to their medicinal properties and diverse bioactive components, medicinal plants have traditionally served as crucial elements of healthcare and still hold significant importance in modern medicine [1,2]. Species identification is critical for their safe and effective utilization as India possesses rich variety of therapeutic species used in Ayurveda and Siddha systems of medicine. Manual identification relying on experts to visually examine distinguishing characteristics can be slow and challenging when different species have similar features. Thus, automated image identification has gained considerable interest for boosting both accuracy and efficiency [3]. Recent deep learning progressions have dramatically advanced plant image identification. Convolutional neural networks (CNNs), which are combined with transfer learning, have achieved an outstanding classification performance by fine-tuning pretrained models for specific classification issues [4, 11]. Notably, lightweight deep learning designs have attracted increasing attention. These layouts offer similar levels of predictive performance but run with less computational resources allowing their use in a mobile and edge-based environment [12, 13, 14]. Explainable artificial intelligence (XAI) approaches i.e., Grad-CAM, SHAP or LIME also further improved the interpretability of deep learning models by illustrating image regions influencing predictions of the given input [13,14,15,16]. Regardless of outstanding progress achieved in deep learning based medicinal plant identification problems, research gaps still exist. Several studies only concentrate on the maximization of the classification accuracy, giving limited consideration to how different lightweight models respond toward fine-tuning. Likewise, the relatively small number of medicinal plant datasets comprising only little classes restrains the assessment of model scalability and generalization capability. Comparable evaluations using a plethora of lightweight transfer learning architectures in medical leaf identification considering uniform experimental conditions are also limited. Similarly, the recently introduced MobileNetV4 architecture needs to be largely explored, thus its possibility of utilization in classifying medicinal species remains unexplored.

Explainable artificial intelligence techniques have furthermore gained in importance in computer vision but still few studies connect lightweight transfer learning models with methods for explainability in medicinal leaf classification.

Motivated by the mentioned problems, this article explores the effects of architecture-specific fine-tuning on three lightweight convolutional neural networks: MobileNetV2, MobileNetV4 Conv Small and ShuffleNetV2. To this aim the experiments used 5,627 images of 56 Indian leaves. Each layout was individually subjected to an evaluation of both transfer learning and subsequent refining through fine-tuning under uniform experimental settings. Aside from classification performance, the models were compared in terms of their computational efficiencies: parameter count, model size, inference time, throughput and GPU memory consumption. Furthermore, Grad-CAM outputs were utilized to explore regions influencing model predictions and to improve interpretability for the developed framework.

The major contributions of the current article can be concluded as follows:

1. A comparative study on MobileNetV2, MobileNetV4 Conv Small and ShuffleNetV2 in Indian medicinal leaf identification
2. A study to determine the effects of architecture-specific fine-tuning on three lightweight transfer learning models by subjecting them to the same experimental setup.
3. A broad evaluation of a 56-class medicinal leaf dataset which comprises 5,627 images alongside an analysis of predictability performance and computational efficiency.
4. An explainability analysis with Grad-CAM that provides qualitative insight into the prediction behavior of the explored lightweight architectures.

2. Related Work

2.1 Medicinal Leaf Classification Using Deep Learning

Deep learning has become effective method for medicinal leaves classification as it discovers discriminate visual features directly from leaf images automatically, reducing the dependency on manually designed feature extraction methods. A new systematic assessment by **Mulugeta et al [8]** describes the fast growing adoption of deep learning in medicinal plants recognition but also acknowledges problem of diversity on datasets, range of species cover and evaluation methods. **Dileep and Pournami** developed the AyurLeaf dataset, 40 medicinal plant species collected from Kerala and showed that a CNN-based classification scheme was effective [17]. **Roopashree and Anitha** introduced DeepHerb dataset which involved 40 species of medicinal Indian herb species on which transfer learning approach has been applied using existing deep learning models [18]. **Reddy et al.** have investigated into comparison of transfer-learning models VGG16, Xception and EfficientNetV2 on their medicinal leaves classifying task [19], whereas **Karwa et al.** have evaluated contemporary deep-learning models such MobileNetV3-Large, EfficientNetB3, YOLOv11n and Vision Transformer to carry out medicinal leaves classification [20]. Furthermore, **Gautam et al.** presented a lightweight fusion of attention & lightweight convolutional neural network [12], whereas **Kavitha et al.** and **Praveena et al [21]** showcased the utility of using deep learning within the realm of real-time medicinal leaves identification as well as in aiding to choose medicinal usage of leaves effectively [22].

Collectively, these studies are demonstration of deep learning which has enhanced the medicinal leaves classification. However, dataset; diversity of the species; imaging situation; evaluation methods still hinder the direct comparison of findings from studies [8].

2.2 Transfer Learning and Fine-Tuning Approaches

Transfer learning has the most adopted strategy for image classification that include restricted label dataset. In the recent past, the researchers have found that fine-tuning pretrained model result into better generalization into the tasks, increasing their classification performance [23, 25].Among these research works was **Ayumi et al. 's**, which directly addressed comparison among models built on transfer learning with and without prior fine-tuning, thus showing that fine-tuning was outperforming those involving only feature-extraction methods with MobileNetV2 reaching at the maximum performance [26]. Similarly, **Kiflie et al. 's** work on reviewing some CNN models built via transfer learning to study the question related to plant identification reported an essential accuracy increase achieved by means of retraining regardless how different models were [27]. In addition, **Hajam et al [28]** in their paper, developed an ensemble framework by combining transfer learning and fine-tuning. It showed improvements on robustness including a more precise classification when juxtaposed to independent pretrained models.In summary, what this research indicated that with the use of transfer learning further enhancement is possible which fine-tune and tune deep learning is allowed to adopt the specific image characters of to give enhanced outcome.

2.3 Lightweight Deep Learning Architectures

Lightweight convolutional neural networks (CNNs) have gained considerable attention in plant image classification owing to their ability to achieve a balance between classification accuracy and computational efficiency, thereby allowing for deployment on resource-constrained devices. Recently, lightweight architectures such as MobileNet, ShuffleNet, EfficientNet, SqueezeNet and GhostNet have been investigated for medicinal plant and leaf classification [9, 10, 23, 29-31]. Several studies aimed to propose lightweight models for medicinal plant classification. Authors in [32] developed a lightweight CNN model named as Ayur-PlantNet that achieve impressive accuracy with reduced computational complexity. Authors in [33] proposed AELGNet model for medicinal leaf classification by incorporating attention mechanisms, achieving higher accuracy along with the compact architecture using local and global features. Similarly, authors in [12] proposed a lightweight feature-fusion framework for medicinal leaf classification. Comparative studies have also demonstrated the effectiveness of lightweight CNNs. The authors in [31] employed MobileNetV3 for transfer learning to build a computationally efficient medicinal plant recognition system. The authors in [10] conducted a study to compare the performance of MobileNetV2, ShuffleNet, EfficientNetB0 and SqueezeNet models, revealing that ShuffleNet outperforms the MobileNetV2, EfficientNetB0 and SqueezeNet models when dealing with varied datasets. There is a need for the lightweight models to be evaluated by computation efficiency such as memory consumption and Flops in addition to accuracy is stated in [23].

Despite these advances, few studies conducted comprehensive comparative evaluations involving several state-of-the-art CNN architectures for medicinal leaf classification. The studies compare a few models using different environmental conditions or propose one lightweight model. Therefore, the relative performance of modern lightweight CNNs, in particular, remain poorly understood for medicinal plant image classification therefore there lies a need for comparative evaluations as is performed in this study.

2.4 Explainable Artificial Intelligence for Medicinal Leaf Classification

In deep learning, Explainable Artificial Intelligence (XAI) is vital for understanding model predictions and for raising transparency to increase the confidence of researchers in the classification. Through the highlighting of specific positions and aspects related to the classification tasks, in an effort to explain the behavior of developed models. In [34], the author integrated Grad-CAM and SHAP into a web-based medicinal plant identification system, leveraging these methods for visualizing the decision-making process of the CNN- and transformer-based models. **Ahmmed et al [35]** employed Grad-CAM, LIME and t-SNE techniques within their lightweight hybrid CNN-Transformer architecture to provide visual and feature-level explanations for disease diagnosis and medicinal leaf classification tasks. In a similar vein, **Kiflie** fused knowledge distillation with a lightweight MobileNetV2 model and leveraged LIME to enhance the interpretability of student model while maintaining high classification accuracy [13]. K**handaker et al [36]** utilized LIME to explain model predictions for medicinal plant recognition. Recently, **N. S. et al [37]** proposed a multimodal CNN-(Transformer)-based framework for the classification of Indian medicinal leaves. They also employed XAI methods for interpreting the results of single and multimodal classifiers using Grad-CAM and LIME respectively.

While XAI is gaining importance in herbal plant research, its application generally lags when compared to the vast adoption of deep learning techniques. Hence, there is a clear imperative for more research aiming at enhancing the interpretability of deep learning models, thereby fostering improved transparency and reliability of automated medicinal leaf classification systems.

Table 1 briefly summarizes recent studies for medicinal leaf classification, including reference, dataset or region, number of classes, models used, lightweight model adoption, transfer learning or fine-tuning, Explainable AI (XAI) and classification accuracy.

Table 1. Summary of Representative Studies on Medicinal Leaf Classification

Ref.	**Dataset / Region**	**Classes**	**Models Used**	**Lightweight Models**	**Transfer Learning**	**Fine-Tuning**	**XAI**	**Accuracy(%)**
Roopashree & Anitha 2021 [18]	Deep Herb dataset (South India, India)	40	VGG16, VGG19, InceptionV3, Xception + ANN/ SVM	✕	☑	☑	✕	97.50% (Xception + ANN)
Redd	Indi	30	VGG1	✕	☑	✕	✕	99.99%

y et al. 2022 [19]	an Medicinal Herbs Dataset (India)		6, Xception, EfficientNetV2					(Xception)
Ayumi et al. 2022 [26]	Medicinal plant leaf dataset	30	VGG16, VGG19, MobileNetV2	mixed	✓	✓	✗	81.82% (MobileNetV2)
Hajam et al. 2023 [28]	Mendeley Medicinal Leaf Dataset	30	VGG16, VGG19, DenseNet201, Ensemble CNN	✗	✓	✓	✗	99.12% (VGG19+DenseNet201)
Kiflie et al. 2024 [27]	Ethiopian indigenous medicinal	35	VGG16, VGG19, InceptionV3, Xception	✗	✓	✓	✗	94.00% (VGG 19)
	plant dataset							
Pushpa & Rani 2023 [32]	Indian Ayurvedic plant dataset	40	Ayur-PlantNet, ResNet34, ResNet50, VGG16, MobileNetV3-Large, EfficientNet-B4, DenseNet121	mixed	✗	✗	✗	92.27% (Ayur-PlantNet)
Gautam et al. 2025 [12]	Indian medicinal leaf dataset	80	FF-NCA-CNN	✓	✗	✗	✗	98.90%
Martono & Triwijoyo 2025	Hugging Face medicinal	30	ResNet18, MobileNetV2, EfficientNetB	mixed	✓	✗	✗	99.50% (ShuffleNet)

[10]	plant dataset		0, SqueezeNet, ShuffleNet, AlexNet					
Nahiduzzaman et al. 2023 [14]	Bangladesh mulberry leaf dataset	3	PDS-CNN + transfer learning models	☑	☑	✕	☑ (SHAP)	95.05%
Kiflie 2026 [13]	Ethiopian medicinal plant dataset	35	MobileNetV2 (knowledge distillation)	☑	☑	☑	☑ (LIME)	99.46%

2.5. Research Gap and Motivation

While significant success has been achieved in medicine plant as well as medicine leaf area classification it suffers from the few of following shortcomings are as described below:

1. The data sets used by different researchers vary in content, size and experimental procedure. Thus the results obtained by different researchers might not be comparable directly.
2. Use of Transfer learning technique and fine tuning has shown greater success rates in classification problems but comparison between the different lightweight architecture in same experiment conditions have not been carried out for medicine leaf whose number of classes is moderately large.
3. Different version architectures based on MobileNet have found greater acceptance while the use of ShuffleNet architecture has been largely ignored. Also the latest introduced MobileNetV4 needs to be largely explored for medicine leaf classification.
4. Explainable techniques namely Grad-CAM, SHAP and LIME are being used in image analysis for plants. But their uses for medicinal leaf models are limited.

Therefore, to bridge the mentioned gap this papers main motive is to study the effect of fine tuning for three lightweight architecture MobileNetV2, MobileNetV4 Conv small and ShuffleNetV2 with 56 class Indian medicine leaf data set and compare them across the parameters using same experimental setup for classification accuracy and time and to evaluate their computational time complexity. We also use Grad CAM to explain the results of the model to shed light on their prediction performance.

3. Materials and Methods

3.1 Data acquisition and pre-processing

The experiment is conducted by training the neural network models listed later on a set of 56 Indian medicine leaf species with a total of 5627 labelled images as described in [38]. This image dataset, similar to most real-word plant image collections, is characterized by image inequality across leaf classes. Stratified sampling method was considered for constructing three data sets: training data, validation data and test data in the ratio of 70%, 15% and 15% consequently. Thus, we reserved 3949 images, 839images and 839 images for training, validation and testing respectively.

All the input images used for training with fine tuning were scaled along x-axis and y-axis to (224 x 224) pixels to fit the input format of transfer learning models during runtime; subsequently, we changed the colour format into RGB to use them effectively for transfer learning, after converting into numpy array (tensor); thus, we converted all the pixel values to (0,1). The pixel values were further scaled using the mean and standard deviation of images available on ImageNet as our baseline was ImageNet.

Some sample images of the selected categories of Indian medicinal plants can be viewed in Figure 1, while the class distribution is shown in Figure 2. From Figure 2 we can infer that the data set is slightly imbalanced.

Fig. 1 Representative leaf sample images from the dataset

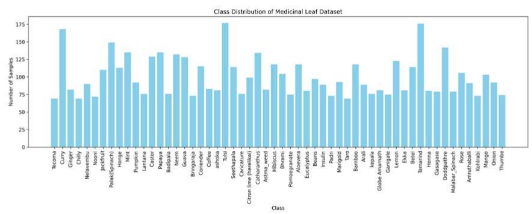

Fig 2: Class Distribution of Medicinal Leaves

3.2 Lightweight Deep Learning Architectures

We investigated the effect of fine tunning for three lightweight Convolutional Neural Network (CNN) namely, **MobileNetV2**, **MobileNetV4 Conv Small** and **ShuffleNetV2**. These model were selected because they provide a balance between classification accuracy while having less computational cost and thus can be integrated into resource-constrained devices at the edge of the network.

3.2.1 MobileNetV2

It was proposed by Sandler et al [39]. It is a lightweight CNN for mobile and embedded vision. It builds on MobileNetV1 with inverted residual blocks and linear bottlenecks for improved feature representation at low computational cost. Each block has a pointwise expansion layer, a depthwise convolution for spatial features and a pointwise projection layer to compress maps. Residual connections are used when dimensions match. MobileNetV2 uses depthwise separable convolutions plus inverted residuals and linear bottlenecks to deliver competitive classification with few parameters and low complexity, suitable for resource-efficient image tasks.

3.2.2 MobileNetV4 Conv Small

Recent version MobileNetV4 is proposed by Qin et al [40]. Previous generation MobileNet models focus on computational complexity, while the V4 was optimized from classification accuracy, inference latency and resource used perspectives with hardware-aware method. The novelty of the models is a redesigned building block called Universal Inverted Bottleneck (UIB) with further extended bottleneck, where the depthwise convolution with adjustable kernel size can be enabled for more flexibility. Hardwareaware Neural Architecture Search (NAS) is employed to efficiently search for well-behaved blocks. In this study, one of their models called MobileNetV4 Conv Small is chosen, has around 3.8 million parameters and needs approximately 0.2 billions multiply adds for forward and backward pass, which is a great fit for the medical leaf classification task on resource-constrained devices. Another motivation was that V4 is proposed recently and has not been implemented often so this study tested the potential of this architecture.

3.2.3 ShuffleNetV2

It was roposed by Ma et al [41]. It enhances practical inference speed by considering hardware efficiency alongside computational complexity. It adheres to design guidelines minimizing memory access cost through improved parallelism. The core ShuffleNetV2 element uses a channel split: one branch for light convolutions, the other an identity path. Outputs merge and channel shuffle ensures efficient information exchange between channel groups. ShuffleNetV2 reduces previous excessive group convolutions and multi-branch complexity, making inference faster with competitive classification. All of them make it beneficial for efficient medicinal leaf classification.

3.3 Methodology Workflow

The overall workflow of this study is in Figure 3. Dataset preparation and splitting in a stratified fashion for the training, validation and test were performed for all classes with a fixed random seed used to maintain class distribution across splits. Train, validation and test images are then pre-processed, including resizing and normalization. After data preparation, three lightweight CNNs (MobileNetV2, MobileNetV4 Conv Small and ShuffleNetV2) are used for classification. Stage one involves initial transfer learning: keeping pretrained feature extraction layers and only training a newly added classification head for the Medicinal leaf task. Baseline transfer learning model performance is evaluated using validation and test datasets.

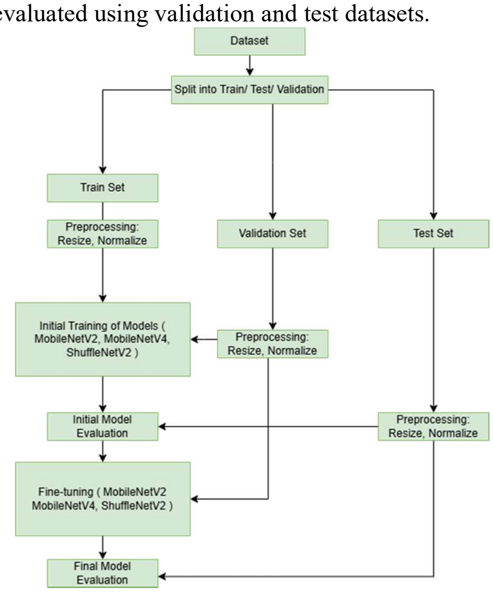

Fig 3. Overview of the methodology adopted for medicinal leaf classification in this study.

In stage two, fine-tuning is conducted by unfreezing selected deeper layers of each architecture, while keeping the parameters of earlier layers frozen. This enables the models to learn more specific, high-level feature representations of medicinal leaf images while retaining the broad visual knowledge acquired from large-scale image datasets. Various Performance metrics were used to determine the effect of fine-tuning on classification. The best fine-tuned models were further analysed through their confusion matrices and with Grad-CAM to establish their strengths and weaknesses on individual class levels.

3.4 Experimental Setup

3.4.1 Transfer Learning Strategy

The ImageNet-1K dataset provided rich visual representation learning opportunities to transfer learning models. The convolutional backbone architectures, MobileNetV2, MobileNetV4 Conv Small and ShuffleNetV2 served as base models during transfer learning for medicinal leaf classification. The original classification head was removed and replaced with a new classification head with a newly added fully connected layer corresponding to the medicinal leaf dataset (featuring 56 classes) to each network. At this point, all feature extracting layers which were present up till this stage remained unfrozen, with only the newly installed classification head being made trainable on the medicinal leaf dataset. The classification heads were to undergo parameter updates exclusively, such that the pretrained parameters of the backbone could remain preserved as they were originally obtained from ImageNet. Validation Macro F1-score was selected as the performance metric for learning-rate scheduling, checkpoint monitoring and model selection. During model selection, the checkpoint having the highest validation Macro F1-score was selected; it would undergo fine-tuning as an unfreezable network and will be retrained with appropriate configuration for weights.

4.4.2 Fine-Tuning Strategy

Deep transfer learning models which were previously trained on the medicinal leaf dataset as a newly trained classification head along with original ImageNet classification backbone; they were further adjusted to a higher and more specific level for medicinal leaf classification specifically by unfreezing selected deeper layers of each architecture beyond the simple transfer learning which did not even involve retraining the backbone for medicinal leaf classification and training them as fine-tuned neural networks with the preserved ImageNet and pretrained medicinal leaf data. This approach permits networks to safeguard knowledge gathered from the prior learning stage and further finetune the upper feature layers of the network according to the feature sets within pictures of medicinal plants.

Once again the stratified division was applied, 70:15:15 training, validation and testing split, where it would be possible, throughout all the experimentation, thus to permit a more fair and reliable comparison between deep transfer learning and deep fine-tuning. The unfrozen layers are only limited to a certain number of layers but exclude the early layers whilst retraining with the classification layer in all subsequent passes while updating fine-tuned training. Moreover, batch normalization statistics have been frozen so as to make for more steady training while at the same time mitigating the potential risks of model overfitting.

The performance indicator Macro validation F1 score determines the learning rates of the models, when model checkpoints are set and whether the models pass beyond certain checks of performance or not. Again, at checkpoints which have obtained the maximum Macro Validation F1-score was once more chosen. The architectures of each of the models along with the specifics of the training process about each architecture are provided in tabular form (see table 2) The number of layers which were unfrozen for each model along with the range of learning rates is displayed too.

Table 2. Fine-Tuning Configuration of the Evaluated Lightweight Models

Model	Unfrozen Layers	Learning Rate	Weight Decay
MobileNet V2	Final four feature blocks + classifier	1×10^{-4}	1×10^{-5}
MobileNet V4 Conv Small	Deeper feature blocks + classifier	1×10^{-5} (backbone), 2×10^{-4} (classifier)	1×10^{-4}
ShuffleNet V2	Stage3, Stage4, Conv5 + classifier	8×10^{-6}(Stage3), 2×10^{-5}(Stage 4), 2×10^{-5}(Stage 5), 2×10^{-4}(classifier)	1×10^{-3}

4.4.3 Hyperparameter Configuration

To acquire an impartial comparison between these selected light architectures without having any side effect from factors other than the architecture itself, a common set of hyperparameters was used as widely as possible. Parameters such as image input size, batch size, optimizer to be used and its configurations, no. Of Epochs and checkpoint selection method were the same for every models in the experiments. There are distinct optimizer settings which were utilized in this study during the stage deep transfer learning and deep fine-tuning for models in respectively order in terms of differences between the goals of two tasks and how exactly that affects the learning process during the training. Table 3 presented in its entirety the list of hyperparameters for both stages.

Table 3. Experimental Hyperparameter Configuration

Parameter	Transfer Learning	Fine-Tuning
Input image size	224 × 224	224 × 224
Batch size	64	64
Optimizer	AdamW	AdamW
Maximum epochs	100	100
Loss function	Cross-Entropy Loss	Cross-Entropy Loss
Learning-rate scheduler	ReduceLROnPlateau	ReduceLROnPlateau
Model selection criterion	Validation Macro F1-score	Validation Macro F1-score
Early stopping patience	6	10

3.5 Evaluation and Explainability Methods

3.5.1 Evaluation Metrics

In this stage, both transfer learning and fine-tuning models were evaluated according to different assessment metrics, such as Accuracy, Precision, Recall, Macro F1-score, Cohen 's Kappa and Macro AUC. These metrics collectively offer a thorough understanding into the performance of the models concerning the task of medicinal leaf classification with multiple classes [42].

The metric of Accuracy measured how well models can label classes accurately. Precision helped in assessing the credibility of predicted class labels and Recall in assessing the capacity of models in discovering all positive cases in each label. The F1-score which considered the effect of both false positives and false negatives was also computed for models and was a harmonic mean of Precision and Recall. Specifically, Macro F1 score was used as it takes a considerate average of all the class-wise F1 scores due to the nature of present dataset. This results in the fact that all the leaf types were given equal consideration irrespective of the number of samples in them. Cohen's Kappa was also determined for every model to find the agreement between actual and predicted class labels by ignoring agreements which could most probably happen by chance [43]. Moreover, to find out the discriminative power from each model under test during the present analysis Macro AUC has also been calculated additionally for all the tested models.

In the process of fine-tuning or transfer learning processes (mentioned in previous stages), the learning rate scheduling, model checkpointing and model selection was primary ruled by the validation Macro F1 score. Furthermore, the analysis of performance per class was achieved by analysis of confusion matrices.

3.5.2 Explainability Analysis Using Grad-CAM

The proposed leaf-classification framework based on various lightweight models were enhanced on its overall interpretation by Grad-CAM. A novel Gradient-weighted Class Activation Mapping (Grad-CAM) technique capable of generating visual explanations was invented by Selvaraju et. al [44]. It helped in identifying the key regions of an Image that were found to be involved at the time of predictions hence allowing a deeper understanding into reasoning and working of the applied lightweight architectures (CNN's).

4. Results and Discussion

4.1 Transfer Learning Performance

The initial transfer learning results of the three networks during step 1 of transfer learning are tabulated in Table 4. The networks have obtained excellent image retrieval.

Table 4. Initial Transfer Learning Performance of the Evaluated Lightweight Architectures.

Model	Macro F1 (%)	Accuracy (%)	Kappa	Macro Precision (%)	Macro Recall (%)	Macro AUC
MobileNetV2	90.31	89.87	0.8967	91.01	90.22	0.9986
MobileNetV4	92.86	92.85	0.9271	93.58	92.63	0.9990
ShuffleNetV2	87.71	87.60	0.8736	88.76	87.84	0.9966

MobileNetV4 Conv Small gave the best baseline performance of the different architectures, achieving the highest Macro F1-score (92.86%), Accuracy (92.85%), Cohen' s Kappa (0.9271), Macro Precision (93.58%) and Macro ROC-AUC (0.9990). The excellent performance of MobileNetV4 Conv

Small indicates that the pretrained feature representations were successfully able to recognise subtle differences between species without the deep layers needing to adapt in the slightest. MobileNetV2 also performed well and achieved a Macro F1 score of 90.31% and Accuracy of 89.87%. It gave a consistently high Macro Precision (91.01%) and Macro Recall (90.22%) and was able to classify the dataset well purely using the newly trained head. Even though ShuffleNetV2 gave the lowest scores across these models, the performance of the model as measured with the Macro F1 score of 87.71% and Accuracy of 87.60% were comparable to the other architectures. This showed considering the small relatively less number of parameters that it could still develop a good level of representation of the unique features during transfer learning. An observation common across all these architectures was their performance relative to the Macro ROC-AUC, that were high in values larger than 0.99 which reflected the fact that each model was sufficiently separate the various species of medicinal leaves even in cases when its accuracy might not be as high. All transfer learning models provided good basis for later on evaluating performance further during architecture specific finene-tuning.

4.2 Fine-Tuning Performance

The performance of each of the three lightweight architectures is summarised in Table 5 following architecture-specific fine-tuning. By tuning these selected deeper layers, the networks became well suited to the visual context of the dataset and each achieved a high classification performance under all evaluation metrics.

Table 5. Fine Tuning Performance of the Evaluated Lightweight Architectures.

Model	**Macro F1 (%)**	**Accuracy (%)**	**Kappa**	**Macro Precision (%)**	**Macro Recall (%)**	**Macro AUC**
MobileNetV2	94.47	94.28	0.9417	94.86	94.41	0.9997
MobileNetV4 Conv Small	**96.23**	**96.42**	**0.9635**	96.64	96.25	0.9996
ShuffleNetV2	96.16	96.19	0.9611	96.40	96.23	0.9997

Among all the architectures measured MobileNetV4 Conv Small gave the best result out of all of them when fine-tuned. The architecture achieved the highest Macro F1 score of 96.23% as well as 96.42% accuracy, also providing the highest Cohen's Kappa score of 0.9635 and Macro Precision of 96.64%. This reflected that during fine-tuning features had updated to be best suited to discern subtle differences in the morphology among the species of medicinal leaves. ShuffleNetV2 performed surprisingly comparable to MobileNetV4 Conv Small when considering the significant differences arising from having a smaller architecture, having only half of the parameters of the former, while achieving values of approximately 96.16% in Macro F1 score, 96.19% in Accuracy and 96.40% in Macro Precision. This reflected how adequately lightweight models having small numbers of parameters have great potential for recognition if they are adequately adapted to suit the architecture specifically for the required task. MobileNetV2 however maintained its ability of classifying well on the dataset, giving a Macro Fl score of 94.47% and an Accuracy score of 94.28%. Its Macro Precision (94.86%), Macro Recall (94.41%) and Cohen's Kappa (0.9417) values also displayed the ability of the model predict reliably. Another remarkable result was given by the other two fine-tuned models which too had ROC-AUC value approximately 1.0 which reiterated the conclusion made earlier about the excellent ability to distinguish different species of medicinal leaves for all models after fine tuning has made them more discriminative.

4.3 Comparative Analysis of Transfer Learning and Fine-Tuning

Figures 4 and 5 indicate the change in performance exhibited by the three lightweight architectures after finetuning. While all models gained higher performance when the layers for feature extraction were modified, the increase was variable between the architectures which suggested that how effective fine-tuning can be depends heavily on the architecture.

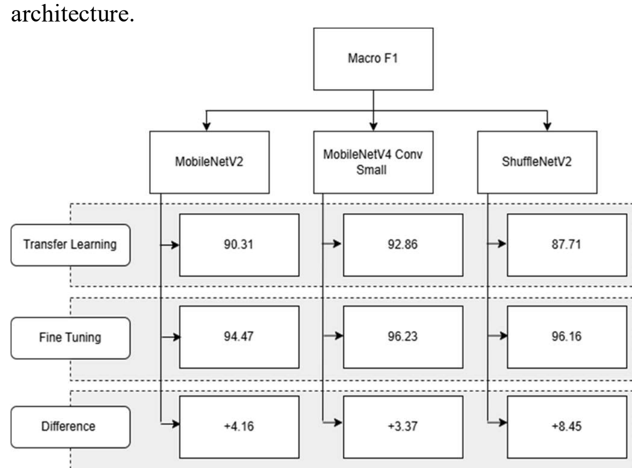

Figure 4. Effect of Fine tuning on Test Macro-F1 of the evaluated models

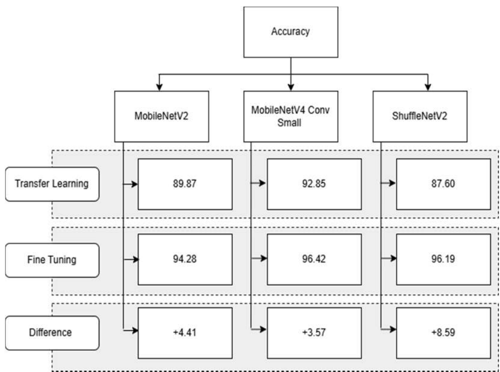

Figure 5. Effect of Fine tuning on Test accuracy of the evaluated models

Of all the studied models ShuffleNetV2 has showed the greatest rise in performance. It has gained 8.59 point in Accuracy and 8.45 points in Macro F1 score. Such a dramatic rise suggests that pretrained representations of transfer learning are not sufficient for covering all of the discriminant features of the medicinal leaves dataset. Adapting deeper layers of the model to the target domain would largely improve the model efficiency at distinguishing similar appearing species.

MobileNetV2 has reacted well to fine-tuning too. Model has improved its performance by 4.41 point in Accuracy scoring and by 4.16 point in Macro F1 score. Rise is not as drastic and dramatic as in ShuffleNetV2, but still shows that refining the higher level feature representations of the model can lead to more reliable classification results among 56 categories of medicinal leaves.

MobileNetV4 Conv Small showed the slightest rise among studied models - 3.57 point rise in Accuracy and 3.37 point rise in Macro F1 score. This, however, is due to the initially good performance of the model even with just transfer learning done. If the pretrained model has already shown the topmost performance, there is just not much left to improve on. Even with slight improvements given to it by fine-tuning, MobileNetV4 Conv Small still has performed the strongest among all studied model architectures!

In summary of the results shown in Table 4 and Table 5, it is clear there are two major take-away lessons about lightweight transfer learning. Firstly, every model, regardless of being finetuned, has shown improved classification results. Secondly, the level of improvement was dependable on the initial pretrained model performance level with medicinal leaves recognition task. Models that performed worse, gained more from the fine-tuning that was done to them. Models that came to perform great with just transfer learning done to them and came to have small improvements after fine-tuning.

4.4 Computational Efficiency Analysis

Beyond classification results, the practical applications of deep learning models also consider their resource requirements. Following results in Table 6 show the parameter count, model size, inference time, throughput and GPU memory consumption for each lightweight transfer learning model, assisting decision-makers with additional deployment criteria.

Table 6. Computational Efficiency Comparison of Fine-Tuned Models

Model	Parameters (M)	Model Size (MB)	Inference Time (ms/image)	Throughput (images/s)	GPU Memory (MB)
MobileNetV2	2.30	9.00	28.01	35.71	438.60
MobileNetV4 Conv Small	2.56	9.99	25.16	39.75	249.49
ShuffleNetV2	1.31	5.19	26.10	38.32	188.48

ShuffleNetV2 was the most compact model by far - 1.31 million trainable parameters, by replacing only model' s original ImageNet classification layer with the 56 classification layer. It also produced the smallest model size - of 5.19 Megabits and consumed the least GPU memory while inference - of 188.48 Megabits. Despite its small size, ShuffleNetV2 has provided inference speed similar to other models while showing strong classification rates!

MobileNetV4 Conv Small is slightly larger by parameters and size, but it has won on fastest inference period and throughput. It provides inference speed of 25.16 millisecond per image and throughput of 39.75 image per second. Overall, it means that MobileNetV4 Conv Small manages to find a great balance of predictive accuracy and computation that needs to be invested in the learning for all of its features.

MobileNetV2 consumed the highest GPU memory of 438.60 megabites, yet still performed similarly in inference by time of its MobileNet family models. It had the lowest throughput of all three architectures too. But it was not too slow either; and since MobileNetV2 still showed accurate results, it is a possible selection for applications when there are sufficient computational resources to use.

In summary - three architectures use their parameters differently to obtain different types of advantages. ShuffleNeV2 would be used if there were limited resources for training/prediction. It is suitable, since it has the smallest size. Yet a good balance between the computational power used and resources should favour MobileNetV4 Conv Small, to show strong prediction efficiency through efficient model inference.

4.5 Error Analysis and Explainability

4.5.1 Confusion Matrix Analysis

While the evaluation metrics reported in the previous sections give an idea about the overall performance of the model, a fine idea about the confusion in recognition of individual classes are given by confusion matrices. Confusions matrix of the fine-tuned MobilenetV4 Conv Small model on all testing samples (refer to Figure 6) is shown below.

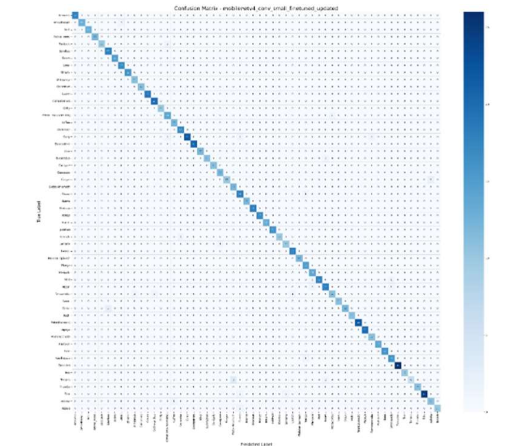

Figure 6. Confusion Matrix of the Fine-Tuned MobileNetV4 Conv Small Model

Almost all predictions are clustered in the main diagonal which means the majority of the testing images are assigned to actual classes. The instances which fall in categories different from their actual are few in counts. They can be found spread between many classes suggesting random confusion in case of 56 medicinal plants. This is consistent with its high Accuracy, 96.42% and high Marco F1 score 96.23%, as previously depicted.

Most of the leaves are recogized with little to no confusions in it, whereas instances of a plant being misrecognized are in most cases equal to once or twice in count than being a confusion of a particular pair of plants. Among the medicinal plant leaves taken up, Tecoma proves to be the most challenging (only 6 of 10 test image were recogizized). Recall rates are lower only for Nelavembu, Ginger and Curry, whereas Guava being misrecgonized in few cases. Therefore, the scattering of error depicts the model doesn not face a systematic confusion but only when few cases of medicinal plant leaves appear visually similar to each other.

Collectively, the confusion matrix depicts representations are learned by the fine-tuned MobilenetV4 Conv Small which distinguishes clearly between medicinal plant leaves with remaining minor confusions associated with only few medicinal plants. Although the previous sections give an exact idea of recognition, it can be visualized more clearly here only certain leaves suffer from misrecognition in particular classes. Further, analysis of the activation patterns shows the reasoning behind the network's prediction.

4.5.2 Grad-CAM Visualizations

To show how the network makes predictions, two correctly and one incorrectly predicted leaf class examples are shown in Figure 7.

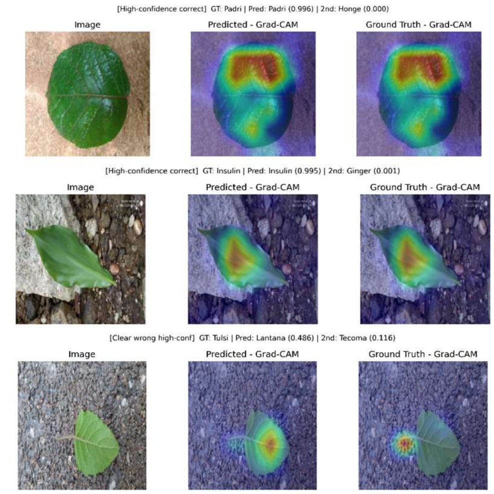

Figure 7. Grad-CAM visualizations generated by the fine-tuned MobileNetV4 Conv Small model

The leaf varieties include Padri and Insulin (accurately classified) and Tulsi (incorrectly identified as Lantana). In the successful cases, only the leaf region in the image contains activation maps, showing nearly no attention paid to the surrounding background. The highlighted regions mostly encompass clear structural characteristics evident in the images, like leaf shape, venation and texture, indicating a strong influence on the network's prediction. The close correlation between activation patterns of the predicted and ground classes suggests that the network is capable of learning stable visual representations.

Additional information on decision making by the network is provided by sample analysis of Tulsi. Even with an incorrect prediction, the activation map is seen clustered to the leaf not having its attention fully being diverted over background as observed above with some class being randomly recognized or misrecognized suggesting that error of prediction was not created due to improper localization, but with commonality in leaf morphology of both Tulsi and predicted class. Also, the difference between activation maps of predicted and ground with each passing layer suggests how change in feature representations leads in changing the decision while predicting leaf categories.

Taken together, the Grad CAM visualisations shows that the model relied on bio-relevant leaf features to generate predictions. Even for the misclassified example the network paid attention towards the important areas of the image allowing for interpretation of the proposed classification framework.

5. Discussion

Based on the results, fine-tuning of the pre-trained models increased the performance of all three lightweight CNN models for medicinal leaf classification. Even though MobileNetV4 Conv Small performed best with transfer learning alone, each model benefited from fine-tuning its features for this task as it learnt fine-grained differences in leaf species; MobileNetV4 Conv Small had better fine-tuning but improvement in ShuffleNetV2 was greater in between stages. Pre-trained ImageNet features provide a good starting-point for image classification task, further fine-tuning improved results. MobileNetV4 Conv Small gave the highest overall validation accuracy and macro F1 scores after fine-tuning. Therefore, MobileNetV4 Conv Small had better task specific features as its features were extracted well. ShuffleNetV2 performed worse than other two networks in the first stage but showed huge improvements over the second training stage, although with higher number of epochs. With fine-tuning, MobileNetV2 performs near to the accuracy of MobileNetV4 Conv Small and its updated features with the changes in parameters (e.g. learning rate) gives good results with fine-tuning. Therefore, fine-tuning helps all three neural networks alike, although some show rapid improvement whereas others take more time or training.

As indicated by the computations, using lightweight CNN models can be effective in this scenario. Using the highest accuracy with minimum inference times gives MobileNetV4 Conv Small better efficiency over the other networks. With lowest number of parameters using the lowest model size consuming minimal GPU memory, ShuffleNetV2 has an advantage over the two MobileNet models if resource scarcity is a limiting constraint. These all are valid findings thus depending on the purpose one can use either of them for maximum accuracy of minimum price to pay and still get good results. The qualitative findings include confusion matrix and Grad-CAM visualizations. As seen in the confusion matrix most of the predictions with only a small number of incorrect predictions lie on the prediction path. Grad-CAM shows that model makes predictions on relevant parts of image leaving out the background details. Hence transfer-learning along with fine-tuning is an effective way of improvement of the lightweight CNNs performance retaining their properties of light weight and faster execution times (speed). Along with MobileNetV4 being accurate and efficient, ShuffleNetV2 retains its lightweight nature and can be effective especially when computational resources (memory is a constraint) becomes an issue.

References

[1]. Davis CC, Choisy P. Medicinal Plants Meet Modern Biodiversity Science. Current Biology. 2024;34(4):R158–R173. doi:10.1016/j.cub.2023.12.038.

[2]. Wang X, Anwar T, Qureshi H, El-Beltagi HS, Sehar Z, Solieva D, Azizov B, Rebouh NY, Abbasov MA, Yakubov F, Alomran MM. Plant-Based Traditional Remedies and Their Role in Public Health: Ethnomedicinal Perspectives for a Growing Population. Journal of Health, Population and Nutrition. 2025;44(1):300. doi:10.1186/s41043-025-01036-5.

[3]. Lankasena N, Nugara RN, Wisumperuma D, Seneviratne B, Chandranimal D, Perera K. Misidentifications in Ayurvedic Medicinal Plants: Convolutional Neural Network (CNN) to Overcome Identification Confusions. Computers in Biology and Medicine. 2024;183:109349. doi:10.1016/j.compbiomed.2024.109349.

[4]. Malik OA, Ismail N, Hussein BR, Yahya U. Automated Real-Time Identification of Medicinal Plant Species in Natural Environment Using Deep Learning Models—A Case Study from Borneo Region. Plants. 2022;11(15):1952. doi:10.3390/plants11151952.

[5]. Azadnia R, Al-Amidi MM, Mohammadi H, Cifci MA, Daryab A, Cavallo E. An AI Based Approach for Medicinal Plant Identification Using Deep CNN Based on Global Average Pooling. Agronomy. 2022;12(11):2723. doi:10.3390/agronomy12112723.

[6]. 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(1):e23655. doi:10.1016/j.heliyon.2023.e23655.

[7]. Bouakkaz H, Bouakkaz M, Kerrache CA, Dhelim S. Enhanced Classification of Medicinal Plants Using Deep Learning and Optimized CNN Architectures. Heliyon. 2025;11(3):e42385. doi:10.1016/j.heliyon.2025.e42385.

[8]. Mulugeta AK, Sharma DP, Mesfin AH. Deep Learning for Medicinal Plant Species Classification and Recognition: A Systematic Review. Frontiers in Plant Science. 2024;14:1286088. doi:10.3389/fpls.2023.1286088.

[9]. Ibrahim NM, Gabr DG, Rahman A, Musleh D, AlKhulaifi D, AlKharraa M. Transfer Learning Approach to Seed Taxonomy: A Wild Plant Case Study. Big Data and Cognitive Computing. 2023;7(3):128. doi:10.3390/bdcc7030128.

[10]. Martono GH, Triwijoyo BK. Performance Evaluation of CNN Models for Image-Based Medicinal Plant Classification. In: 2025 9th International Conference on Electrical,

Electronics and Information Engineering (ICEEIE). IEEE; 2025. p. 1–6. doi:10.1109/ICEEIE66203.2025.11252300.

[11]. Hou Q, Yang W, Liu G. Chinese Herbal Medicine Recognition Network Based on Knowledge Distillation and Cross-Attention. Scientific Reports. 2025;15:1687. doi:10.1038/s41598-025-85697-6.

[12]. Gautam V, Kaur G, Ghantasala GSP, et al. A Lightweight Deep Learning Method for Medicinal Leaf Image Classification Using Feature Fusion. Scientific Reports. 2025;15:34417. doi:10.1038/s41598-025-17436-w.

[13]. Kiflie MA. Leveraging Knowledge Distillation for Lightweight and Interpretable Deep Learning in Ethiopian Medicinal Plant Classification. Intelligent Systems with Applications. 2026;30:200638. doi:10.1016/j.iswa.2026.200638.

[14]. Nahiduzzaman M, Chowdhury MEH, Salam A, Nahid E, Ahmed F, Al-Emadi N, Ayari MA, Khandakar A, Haider J. Explainable Deep Learning Model for Automatic Mulberry Leaf Disease Classification. Frontiers in Plant Science. 2023;14:1175515. doi:10.3389/fpls.2023.1175515.

[15]. Kondaveeti HK, Simhadri CG. Evaluation of Deep Learning Models Using Explainable AI with Qualitative and Quantitative Analysis for Rice Leaf Disease Detection. Scientific Reports. 2025;15:31850. doi:10.1038/s41598-025-14306-3.

[16]. Alhammad SM, Khafaga DS, El-Hady WM, Samy FM, Hosny KM. Deep Learning and Explainable AI for Classification of Potato Leaf Diseases. Frontiers in Artificial Intelligence. 2025;7:1449329. doi:10.3389/frai.2024.1449329.

[17]. Dileep MR, Pournami PN. AyurLeaf: A Deep Learning Approach for Classification of Medicinal Plants. In: 2019 IEEE Region 10 Conference (TENCON). Kochi, India: IEEE; 2019. p. 321–325. doi:10.1109/TENCON.2019.8929394.

[18]. Roopashree S, Anitha J. DeepHerb: A Vision Based System for Medicinal Plants Using Xception Features. IEEE Access. 2021;9:135927–135941. doi:10.1109/ACCESS.2021.3116207.

[19]. Reddy KUSD, Shaik A, Balasundaram A, Nithin MS, Kakarla LSR, Noor Mahammad S. Classification of Indian Medicinal Leaves Using Transfer Learning Based Convolutional Neural Networks. In: 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC). Trichy, India: IEEE; 2022. p. 1050–1058. doi:10.1109/ICOSEC54921.2022.9952074.

[20]. Karwa D, Mittal U, Wadhawan A. Comparative Study of Deep Learning Models for Image-Based Classification of Medicinal Plants. In: 2025 12th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO). Noida NCR, India: IEEE; 2025. p. 1–6. doi:10.1109/ICRITO66076.2025.11241695.

[21]. Kavitha S, Kumar TS, Naresh E, et al. Medicinal Plant Identification in Real-Time Using Deep Learning Model. SN Computer Science. 2024;5:73. doi:10.1007/s42979-023-02398-5.

[22]. Praveena S, Pavithra SM, Kumar ADV, et al. CNN-Based Indian Medicinal Leaf Type Identification and Medical Use Recommendation. Neural Computing and Applications. 2024;36:5399–5412. doi:10.1007/s00521-023-09352-9.

[23]. Ali H, Shifa N, Benlamri R, et al. A Fine Tuned EfficientNet-B0 Convolutional Neural Network for Accurate and Efficient Classification of Apple Leaf Diseases. Scientific Reports. 2025;15:25732. doi:10.1038/s41598-025-04479-2.

[24]. Khalid F, Romle AA. Herbal Plant Image Classification Using Transfer Learning and Fine-Tuning Deep Learning Model. Journal of Advanced Research in Applied Sciences and Engineering Technology. 2024;35(1):16–25. doi:10.37934/araset.34.3.1625.

[25]. Shafik W, Tufail A, De Silva Liyanage C, et al. Using Transfer Learning-Based Plant Disease Classification and Detection for Sustainable Agriculture. BMC Plant Biology. 2024;24:136. doi:10.1186/s12870-024-04825-y.

[26]. Ayumi V, Ermatita, Abdiansah, Noprisson H, Jumaryadi Y, Purba M, Utami M, Putra ED. Transfer Learning for Medicinal Plant Leaves Recognition: A Comparison with and Without a Fine-Tuning Strategy. International Journal of Advanced Computer Science and Applications. 2022;13(9). doi:10.14569/IJACSA.2022.0130916.

[27]. Kiflie MA, Sharma DP, Haile MA. Deep Learning for Ethiopian Indigenous Medicinal Plant Species Identification and Classification. Journal of Ayurveda and Integrative Medicine. 2024;15(6):100987. doi:10.1016/j.jaim.2024.100987.

[28]. Hajam MA, Arif T, Khanday AMUD, Neshat M. An Effective Ensemble Convolutional Learning Model with Fine-Tuning for Medicinal Plant Leaf Identification. Information. 2023;14(11):618. doi:10.3390/info14110618.

[29]. Ahmed S, Hasan MB, Ahmed T, Sony R, Kabir MH. Less Is More: Lighter and Faster Deep Neural Architecture for Tomato Leaf

Disease Classification. IEEE Access. 2022;10:69368–69384. doi:10.1109/ACCESS.2022.3187203.

[30]. Chowdhuri S, Banerjee S, Bhargav A, Chaudhuri SS. Therapeutic Foliage: A Balanced Dataset Designed for Automated Identification of Leaves of Common Medicinal Plants. In: 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE; 2023. p. 362–367. doi:10.1109/ICIRCA57980.2023.10220905.

[31]. Ashika KA, Veni S. Optimizing Data Augmentation, Class Weighting, and Fine-Tuning in MobileNetV3 for Indian Medicinal Leaf Classification. In: Kumar S, Shorif Uddin M, Rawat TK, editors. Sixth Congress on Intelligent Systems (CIS 2025). Lecture Notes in Networks and Systems. Vol. 1837. Singapore: Springer; 2026. doi:10.1007/978-3-032-18282-1_14.

[32]. Pushpa BR, Rani N. Ayur-PlantNet: An Unbiased Lightweight Deep Convolutional Neural Network for Indian Ayurvedic Plant Species Classification. Journal of Applied Research on Medicinal and Aromatic Plants. 2023;35:100459. doi:10.1016/j.jarmap.2023.100459.

[33]. Sharma S, Vardhan M. AELGNet: Attention-Based Enhanced Local and Global Features Network for Medicinal Leaf and Plant Classification. Computers in Biology and Medicine. 2024;184:109447. doi:10.1016/j.compbiomed.2024.109447.

[34]. Kikome C, Okumu G, Wagisha E, Jjingo D, Kizito J, Marvin G. Transparent Machine Vision Techniques for Medicinal Plant Species Identification. In: Proceedings of the 5th International Conference on Image Processing and Capsule Networks (ICIPCN 2024). IEEE; 2024. p. 135–142. doi:10.1109/ICIPCN63822.2024.00030.

[35]. Ahmmed J, Kabir MA, Rehman AU, et al. A Lightweight Hybrid CNN and Transformer Model for Medicinal Leaf Disease Classification with Explainable AI. Scientific Reports. 2026;16:8243. doi:10.1038/s41598-026-39182-3.

[36]. Khandaker S, Rahman MM, Azad AKM, Moni MA, Hossain MA. Transparent Deep Learning for Medicinal Plant Recognition: A Hybrid CNN–ViT Approach with Explainable AI on BDHerbalPlants. Smart Agricultural Technology. 2026;14:101862. doi:10.1016/j.atech.2026.101862.

[37]. N S, H R, Reddy KSV, R PT, Kasturi N. A Multimodal Explainable AI Framework for Classification of Indian Medicinal Leaves Using Transformer and CNN-Based Architectures. In: 2025 IEEE 22nd India Council International Conference (INDICON). IEEE; 2025. p. 1–5. doi:10.1109/INDICON68490.2025.11393028.

[38]. Pushpa BR, Rani S. Indian Medicinal Leaves Image Datasets. Mendeley Data. Version 3; 2023. DOI: 10.17632/748f8jkphb.3

[39]. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018:4510-4520. DOI: 10.1109/CVPR.2018.00474.

[40]. Qin D, Leichner C, Delakis M, Fornoni M, Luo S, Yang F, Wang W, Banbury C, Ye C, Akin B, et al. MobileNetV4: Universal Models for the Mobile Ecosystem. In: European Conference on Computer Vision (ECCV). 2024:78-96. DOI: 10.1007/978-3-031-73661-2_5.

[41]. Ma N, Zhang X, Zheng HT, Sun J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In: Computer Vision – ECCV 2018. Cham: Springer; 2018. p. 122-138. DOI: 10.1007/978-3-030-01264-9_8.

[42]. Sokolova M, Lapalme G. A Systematic Analysis of Performance Measures for Classification Tasks. Information Processing & Management. 2009;45(4):427-437. DOI: 10.1016/j.ipm.2009.03.002.

[43]. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 1960;20(1):37-46. DOI: 10.1177/001316446002000104.

[44]. Selvaraju RR, 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). 2017:618-626. DOI: 10.1109/ICCV.2017.74.