International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 6, Nov-Dec 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Medicinal Plants Recognition Using Machine Learning and Image Processing

[1]Natasha Bora, [2]Dr. Bobby Sharma
[1](M.Tech CSE Student, The Assam Don Bosco University, Guwahati, Assam, India.)
[2](Professor (Guide), The Assam Don Bosco University, Guwahati, Assam, India.)

--************************----------------------------------

Abstract:

The identification of medicinal plants is crucial for their conservation and utilization in healthcare and pharmaceutical applications. Traditional methods of plant recognition are often time-consuming and prone to errors, highlighting the need for an automated approach. This project addresses the problem by leveraging machine learning and image processing techniques for accurate medicinal plant recognition. A deep learning-based model, specifically a Convolutional Neural Network (CNN), is trained on plant images to classify species based on their visual features. The methodology involves image preprocessing, augmentation, and model optimization to improve classification accuracy. Experimental results demonstrate the effectiveness of deep learning in medicinal plant recognition, contributing to biodiversity conservation and the advancement of herbal medicine research.

Keywords — **Medicinal plant recognition, Machine learning, Deep learning, Convolutional Neural Networks (CNN), Feature extraction, Model optimization, Pattern recognition.**

--************************----------------------------------

I. INTRODUCTION

Medicinal plants have played an essential role in traditional healing practices and continue to contribute greatly to modern healthcare and pharmaceutical research. Accurate identification of these plants is crucial for safe usage, conservation, and scientific advancement, but manual recognition is time-consuming, expertise-dependent, and prone to human errors. With the growth of artificial intelligence and computer vision, automated plant recognition offers a faster and more reliable solution. This project focuses on developing a machine learning-based system for classifying medicinal plants using image processing and deep learning techniques, particularly Convolutional Neural Networks (CNNs) for feature extraction and species identification. The system aims to support researchers, conservationists, and healthcare professionals by providing an efficient tool for accurate plant recognition and contributing to sustainable and AI-driven botanical research.

II. LITERATURE REVIEW

The reviewed studies collectively explore medicinal and general plant recognition using conventional image processing, deep learning, and transfer learning techniques across various datasets and applications. Pearline et al. [1] compared traditional handcrafted features (Hu Moments, LBP, Haralick, color statistics) with CNNs (VGG16, VGG19, Inception-v3, Inception-ResNet-v2) on Folio, Swedish Leaf, Flavia, and Leaf12 datasets, showing CNNs outperforming traditional methods with up to 99.42% Rank-5 accuracy, though computational cost and limited real-time deployment were noted. Ayumi et al. [2] applied transfer learning with VGG16, VGG19, and MobileNetV2 on the Medicinal Leaf Dataset (30 species, 1500 training, 43 testing), finding MobileNetV2 with fine-tuning achieved superior performance (training 98.41%, validation 96.02%, test 81.82%, F1 0.76), highlighting the importance of fine-tuning, model selection, and preprocessing. Ghosh et al. [3] combined PCA with VGG16 for hybrid transfer learning, achieving 95.25% accuracy and 0.948 F1-score on 30 species, demonstrating improved feature extraction but noting resource intensity and limited datasets. Tiwari et al. [4] evaluated CNNs (5/7 layers, AlexNet, ResNet) on 30-species leaves, with CNN 5 Layer achieving 96.82% accuracy, emphasizing the need for dataset expansion and hybrid approaches. Kavitha et al. [5] implemented a MobileNet-based mobile app for six species (3,000 images), achieving 98.33% test accuracy, demonstrating lightweight cloud-based real-time recognition but limited species coverage. Thenral and Mahalakagan [6] proposed AyurLeaf, a CNN-based system for 20 Ayurvedic plants, attaining 95% accuracy with preprocessing including resizing, cropping, background removal, and segmentation, highlighting scalability yet limited dataset size. Study [7] employed DenseNet on 1,835 images (30 species) for mobile deployment, achieving 99% accuracy, outperforming prior methods, with limitations in dataset size and diversity. Study [8] integrated CBAM attention mechanisms with DenseNet-121, Vision Transformer, and Swin Transformer on DIMPSAR (5945 images, 40 species), achieving 99.66% accuracy with attention-enhanced feature extraction but high computational cost. Study [9] applied ensemble methods (VGG16, VGG19, ResNet50, InceptionV3, EfficientNetB2) with Adam optimizer on MepcoTropicLeafV1 and Spinach datasets, where ensemble ResNet101V2 reached 97.37% test accuracy, highlighting robustness but narrow species coverage. Paper [10] evaluated CNNs (ResNet34, DenseNet121, VGG11, ConvNeXt, Swin Transformer) on IndoHerb and Vietnam Medicinal Plant datasets (10,000–20,000 images, 100–200 classes), with ConvNeXt achieving 92–92.78% accuracy, emphasizing transfer learning and dataset creation, though computational needs remained high. Kumar et al. [11] implemented MobileNetV2 with transfer learning in a Streamlit web app, achieving 97% accuracy with real-time chatbot support, highlighting mobile integration but unspecified dataset size. Study [12] introduced PSR-LeafNet, combining subnetworks (P-Net, S-Net, R-Net) with SVM for leaf classification, achieving 95.88–98.10% accuracy on Flavia, Malayakew, and Indian Medicinal Plant datasets, demonstrating robustness yet limited to leaf-based identification. Bouakkaz et al. [13] developed a hybrid CNN with residual and inverted residual blocks optimized via Binary Chimp Optimization on a 30-class Kaggle dataset, achieving 99.9% accuracy with Grad-CAM interpretability, though computationally intensive and not real-time. Thapa et al. [14] created a mobile app using DenseNet121, ResNet50, VGG16, InceptionV3, EfficientNetV2, and ViT on 12,000 images (50 species), achieving 95.9% accuracy for real-time Nepalese herb recognition, highlighting mobile deployment capabilities yet noting computational load and decreased performance for unseen species. Collectively, these studies underscore the efficacy of CNNs, transfer learning, attention mechanisms, hybrid models, and ensembles for high-accuracy plant recognition while identifying common limitations such as dataset size, computational cost, real-time deployment challenges, and the need for mobile-friendly, scalable solutions.

III. METHODOLOGY

A. *Dataset Description*

The dataset used in this research consists of the Indian Medicinal Plant Image Dataset combined with additional self-captured images to enhance diversity and real-world applicability. A total of 80 medicinal plant species were included in this study, with an overall collection of 20,000 images. The images capture variations in plant appearance under different environmental conditions such as changes in lighting, background, leaf orientation, and growth stages, which helps improve the robustness and generalization capability of the developed recognition system. All images were resized to a uniform resolution of 224 × 224 pixels to match the input requirements of the deep learning models. The dataset was systematically divided into training, validation, and testing subsets, where 14,080 images were used for training, while 2,960 images each were allocated for validation and testing purposes. This balanced dataset distribution ensures effective learning, accurate performance evaluation, and reliable generalization of the proposed deep learning-based medicinal plant recognition system.

B. Image Preprocessing

To improve the quality of input images and enhance model performance, several image preprocessing techniques were applied. All images were resized to 224 × 224 pixels and normalized by scaling pixel values between 0 and 1. Noise removal and image segmentation techniques were used to eliminate irrelevant background information and enhance the visibility of leaf features. Additionally, data augmentation techniques such as rotation, horizontal and vertical flipping, zooming, and brightness variation were employed to increase dataset diversity and reduce overfitting during training

C. Model Architecture

In this research, two convolutional neural network architectures based on transfer learning were implemented, namely MobileNetV2 and EfficientNetB0. Both models were initialized using pretrained ImageNet weights to leverage previously learned feature representations. The original fully connected layers of both networks were modified and customized for 80-class medicinal plant classification. Fine-tuning was performed by unfreezing selected convolutional layers and training them with a low learning rate to further improve classification accuracy while avoiding overfitting.

D. Training and Optimization

The training process was carried out using the Adam optimizer due to its fast convergence and adaptive learning capabilities. Different loss functions were applied for each model, where categorical cross-entropy was used for MobileNetV2 and sparse categorical cross-entropy was used for EfficientNetB0. To enhance training efficiency and prevent overfitting, several callback functions including EarlyStopping, ReduceLROnPlateau, and ModelCheckpoint were implemented. These techniques helped monitor validation performance, adjust learning rates dynamically, and save the best-performing model during training.

IV. EVALUATION METRICS

The performance of the proposed medicinal plant recognition system was evaluated using standard classification metrics including accuracy, precision, recall, F1-score, and confusion matrix. Accuracy was used to measure the overall correctness of the model by calculating the proportion of properly classified images. Precision was employed to determine the reliability of the model in predicting correct positive instances, while recall measured the model's ability to correctly identify all relevant plant classes. The F1-score, which represents the harmonic mean of precision and recall, was used to provide a balanced measure of classification performance. In addition, the confusion matrix was analyzed to visualize the classification results and to understand misclassification patterns across different medicinal plant categories.

V. EXPERIMENTAL RESULTS

The performance of the proposed medicinal plant recognition system was evaluated using two deep learning models, MobileNetV2 and EfficientNetB0, both with and without fine-tuning. Without fine-tuning, MobileNetV2 achieved a classification accuracy of 68.75%, which significantly improved to 90.98% after fine-tuning. Similarly, EfficientNetB0 achieved an initial accuracy of 92.80% without fine-tuning, which further increased to 95.14% after fine-tuning. Among the two models, EfficientNetB0 achieved the highest classification accuracy after fine-tuning, demonstrating superior learning capability and generalization performance. The training and validation accuracy and loss curves indicated stable learning behavior with minimal overfitting, confirming the effectiveness of the applied preprocessing, transfer learning, and optimization techniques.

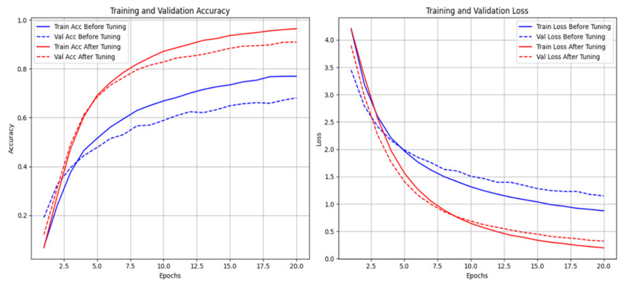

Fig 1: Training, validation accuracy and loss curves of MobileNetV2

The training and validation accuracy/loss graphs for the MobileNetV2 model clearly illustrate the model's learning progression. Initially, both training and validation accuracy gradually increased, while the loss consistently decreased, indicating effective learning. Without fine-tuning, the model achieved a moderate validation accuracy of 68.75%, but after applying fine-tuning, the validation accuracy significantly improved to 90.98%. This demonstrates that unfreezing and retraining the top layers of the base model allowed it to better adapt to the domain-specific features of medicinal plant images. The training curves also showed no significant signs of overfitting, suggesting good generalization performance.

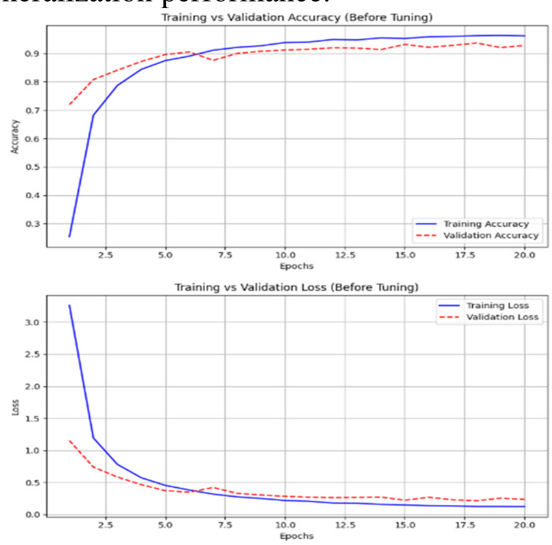

Fig 2: Training, validation accuracy and loss curves of EfficientNetB0 (Before Tuning)

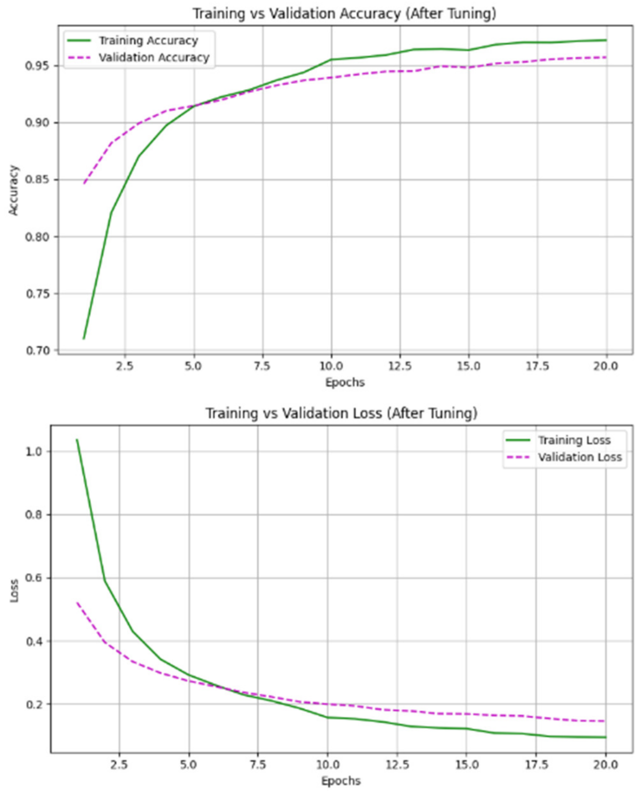

Fig 3: Training, validation accuracy and loss curves of EfficientNetB0 (After Tuning)

The training and validation accuracy/loss graphs for the EfficientNetB0 model demonstrate strong and stable learning performance. From the beginning, the model achieved high accuracy and low loss on both training and validation sets, even without fine-tuning. Initially, the validation accuracy reached 92.80%, indicating that the pretrained EfficientNetB0 was already well-suited for feature extraction in medicinal plant images. After applying fine-tuning, the validation accuracy further improved to 95.14%, showing that the model could effectively adapt and refine its parameters for the specific classification task. The smooth curves and minimal gap between training and validation accuracy indicate excellent generalization and minimal overfitting.

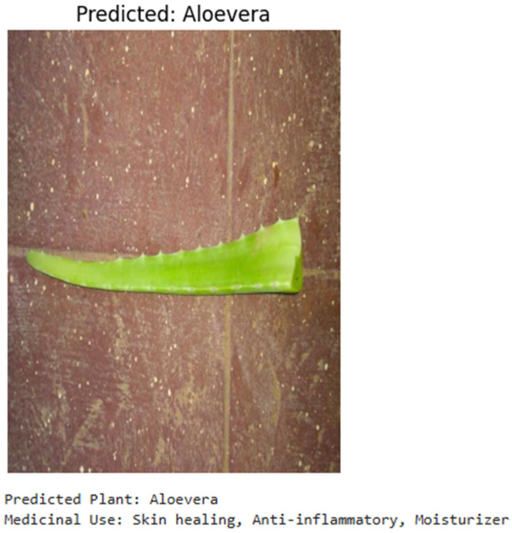

Fig 4: Prediction and Display of Test Image with Medicinal Use

VI. DISCUSSION

The experimental results demonstrate that MobileNetV2 showed a significant improvement in classification performance after fine-tuning, indicating its suitability for resource-constrained environments such as mobile and embedded applications. Its lightweight architecture and reduced computational cost make it an efficient choice for real-time deployment. However, EfficientNetB0 outperformed MobileNetV2 in terms of classification accuracy due to its compound scaling approach, which optimally balances network depth, width, and input resolution. The consistently high accuracy achieved by EfficientNetB0 highlights the robustness and effectiveness of CNN-based transfer learning for medicinal plant recognition. These findings confirm that deep learning models, when properly fine-tuned, can reliably identify medicinal plant species with high precision and generalization capability.

VII. CONCLUSION

This research presents an efficient and accurate medicinal plant recognition system using deep learning and image processing. The system successfully classifies 80 medicinal plant species with a maximum accuracy of 95.14%. The integration of transfer learning and fine-tuning significantly improves classification performance. The proposed system can assist botanists, healthcare professionals, researchers, and farmers in real-time identification of medicinal plants. Future work will focus on mobile deployment, larger datasets, and explainable AI.

REFERENCES

[1] S. A. Pearline, V. S. Kumar, and S. Harini, "A study on plant recognition using conventional image processing and deep learning approaches," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 3, pp. 1997–2004, Jan. 2019, doi: 10.3233/JIFS-169911.

[2] V. Ayumi, A. Dharmawan, and A. Haris, "Transfer learning for medicinal plant leaves recognition: A comparison with and without a fine-tuning strategy," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, pp. 138–142, Sept. 2022, doi: 10.14569/IJACSA.2022.0130916.

[3] S. Ghosh, A. Singh, and S. Kumar, "Identification of medicinal plant using hybrid transfer learning technique," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 3, pp. 1605–1615, Sept. 2023, doi: 10.11591/ijeecs.v31.i3.pp1605-1615.

[4] N. Tiwari, B. K. Gupta, M. Gupta, and A. Yadav, "Developing a neural network-based model for identifying medicinal plant leaves using image recognition techniques," *Journal of Advanced Zoology*, vol. 44, no. S5, Nov. 2023, doi: 10.17762/jaz.v44iS-5.1564.

[5] S. Kavitha, T. S. Kumar, E. Naresh, V. H. Kalmani, and K. D. Bamane, "Medicinal plant identification in real-time using deep learning model," *SN Computer Science*, vol. 5, art. 73, 2024, doi: 10.1007/s42979-023-02398-5.

[6] R. Thenral and M. S. Mathiazhagan, "Medicinal plant recognition for Ayurvedic treatment using deep learning approach," *Journal of Current Research in Engineering and Science*, vol. 7, no. 1, pp. 23–30, Jan. 2024.

[7] K. Vyadande, P. P. Ghadekar, A. Sawant, M. P. Shelke, B. Shirsath, B. Bhande, H. Sawai, S. Gawade, and S. Samgir, "Identification of Ayurvedic medicinal plant using deep learning," *International Journal of Intelligent Systems and*

Applications in Engineering, vol. 12, no. 15s, 2024.

[8] S. Kanakasabai and M. K. Kar, "Classification of Indian medicinal plant species using attention module with transfer learning," in *Proc. IEEE Int. Conf. Signal Process., Informatics, Commun. Energy Syst. (SPICES)*, Kottayam, India, Sept. 2024, doi: 10.1109/SPICES62143.2024.10779801.

[9] P. K. Sharma, "Medicinal plant identification by leaf structure using ensemble methods on deep learning algorithms," *International Journal of Recent Advances in Multidisciplinary Topics*, vol. 5, no. 9, pp. 88–93, Sept. 2024.

[10] M. S. I. Musyaffa'a, N. Yudistira, M. A. Rahman, A. H. Basori, A. B. F. Mansur, and J. Batoro, "IndoHerb: Indonesia medicinal plants recognition using transfer learning and deep learning," *Heliyon*, vol. 10, no. 23, art. e40606, Dec. 2024, doi: 10.1016/j.heliyon.2024.e40606.

[11] G. K. Kumar, A. Rama, D. S. C. Reddy, and R. Vinayak, "Medicinal plants identification through image processing and machine learning," *International Journal of Engineering Innovations and Management Strategies*, vol. 1, no. 1, pp. 1–11, Dec. 2024.

[12] P. K. Sekharamantry, M. S. Rao, Y. Srinivas, and A. Uriti, "PSR-LeafNet: A deep learning framework for identifying medicinal plant leaves using support vector machines," *Big Data and Cognitive Computing*, vol. 8, no. 12, art. 176, Dec. 2024, doi: 10.3390/bdcc8120176.

[13] H. Bouakkaz, M. Bouakkaz, C. A. Kerrache, and S. Dhelim, "Enhanced classification of medicinal plants using deep learning and optimized CNN architectures," *Heliyon*, vol. 11, no. 3, art. e42385, Jan. 2025, doi: 10.1016/j.heliyon.2025.e42385.

[14] P. Thapa, M. Sharma, J. Nyachhyon, and Y. R. Pandeya, "Local herb identification using transfer learning: A CNN-powered mobile application for Nepalese flora," *Vietnam Journal of Computer Science*, preprint (arXiv:2505.02147), Mar. 2025.