International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 206-212
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0028

Automated Medicinal Plant Identification Using Deep Convolutional Neural Networks

Shobha Chandra K[1] Shruthi N M[2], Vishwajith G Bhat[3], Fathima Zahara[4], Veditha B S[5]
[1]Assistant Professor Department of CSE, Malnad College of Engineering, Hassan, India.
[2,3,4,5]Department of CSE, Malnad College of Engineering, Hassan, India.
Emails: *ksc@mcehassan.ac.in[1], shruthiimahesh07@gmail.com[2], vishwajithgbhat@gmail.com[3] zaharafathima935@gmail.com[4], bsveditha@gmail.com[5]*

Abstract
The accurate determining different species of medicinal plants presents a significant challenge in ethnobotany and healthcare applications due to morphological similarities and environmental variations. This research work develops an automated system leveraging CNN architectures for medicinal plant recognition using leaf imagery. The system integrates preprocessing, feature extraction, and multi-class classification to enable real-time plant detection. Image preprocessing techniques such as resizing, normalization, and augmentation enhance model robustness against lighting and background variations. The CNN automatically extracts discriminative morphological features, including leaf venation, margin patterns, and textural properties, through hierarchical learning. Transfer learning with progressive fine-tuning strategies is employed to improve feature generalization and classification accuracy. Experimental evaluation demonstrates that the model achieves an accuracy level of approximately 93%, effectively distinguishing visually similar plant species. The trained model is deployed in a Flask-based interactive web interface through which users can upload leaf images for real-time identification. Along with the prediction, the system displays the plant's scientific name, medicinal properties, and therapeutic benefits, providing an accessible and intelligent platform for automated medicinal plant recognition.

Keywords: *Medicinal plant identification, Convolutional Neural Networks, Deep learning, Image classification, Transfer learning, Botanical recognition, Flask web application*

1. Introduction

1.1 Background and Motivation

Medicinal plants have a significant impact on traditional and modern healthcare systems due to their therapeutic compounds and bioactive properties. The World Health Organization (WHO) reports that nearly 80% of developing nations depend on herbal medicine as a primary source of healthcare, recognizing over 21,000 plant species with documented medicinal value. However, accurate identification of medicinal plants remains a major challenge in ethnobotany and pharmacognosy. **Several factors contribute to this difficulty:**

- **Morphological similarity -** many plant species exhibit overlapping visual features such as leaf shape, texture, and color.
- **Phenotypic plasticity -** environmental conditions like lighting, soil, and humidity alter the appearance of leaves.
- **Expert dependency -** conventional identification methods require specialized botanical expertise.
- **Seasonal variations -** changes in leaf structure across seasons complicate visual recognition.,

With recent advancements in computer vision and deep learning, automated image-based classification has have become a potential method to overcome these drawbacks. Convolutional Neural Networks (CNNs) demonstrate high capability in extracting meaningful spatial features from visual inputs,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 206-212
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0028

enabling accurate plant species identification based on leaf morphology. This research utilizes CNN-based deep learning to streamline the procedure of medicinal plant detection and information retrieval, thereby enhancing accessibility and promoting sustainable healthcare practices.

1.2 Problem Formulation

Let

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

represent a dataset of N leaf images, where $x_i \in \mathbb{R}^{H\times W\times C}$ denotes an image with height H, width W, and C colour channels, and $y_i \in \{1,2,\dots,K\}$ represents the corresponding plant species label from K possible classes. The objective is to learn a classification function

$$f: \mathbb{R}^{H\times W\times C} \rightarrow \{1,2,\dots,K\}$$

that accurately predicts the species label y_i for a given input image x_i, while minimizing prediction uncertainty and maximizing classification accuracy. Formally, this can be expressed as:

$$\hat{y} = \arg\max_{k\in\{1,\dots,K\}} P(y = k \mid x; \theta)$$

where θ denotes the optimized parameters of the CNN model.

1.3 Contributions

This research makes the following key contributions:

- **Automated Identification Framework:** A CNN-based system capable of detecting and classifying medicinal plants directly from leaf images.
- **Comprehensive Preprocessing Pipeline:** Implementation of multi-scale augmentation strategies such as rotation, zoom, brightness, and flipping to improve generalization under varying environmental conditions.
- **Optimized Training Process:** Progressive fine-tuning and transfer learning optimization to enhance model convergence and feature representation.
- **Confidence-Based Prediction:** Integration of probabilistic output metrics to evaluate the reliability of predictions.
- **Deployment-Ready Web Application**: A Flask-based user interface that enables real-time plant detection and displays scientific name, medicinal properties, and therapeutic benefits.

2. Related Work

Janani and Gopal [1] pioneered early efforts in application of Artificial Neural Networks (ANNs) for medicinal plant identification based on extracted image features. Their approach extracted morphological features including leaf area, perimeter, major and minor axis lengths, and textural properties. The system employed multi-layer perceptron networks for classification, achieving reasonable accuracy for a limited set of medicinal species. However, the reliance on hand-crafted feature engineering restricted scalability and generalization capabilities across diverse plant varieties. Simandla et al. [2] advanced traditional machine learning by implementing an auto-detection system using ensemble methods. Their framework incorporated Support Vector Machines (SVM) and Random Forest classifiers operating on extracted shape descriptors, color histograms, and texture features derived from Grey Level Co-occurrence Matrix (GLCM). The study emphasized feature selection optimization through Principal Component Analysis (PCA) to reduce dimensionality. While demonstrating improved computational efficiency, the approach still necessitated utilization of expert insights for extracting meaningful features and showed performance degradation under varying illumination conditions. Krishna et al. [3] introduced a novel hybrid methodology incorporating image processing strategies alongside machine learning classifiers for medicinal plant identification. Their pipeline incorporated edge detection algorithms (Canny, Sobel), morphological operations, and contour analysis to extract leaf boundary characteristics. Feature vectors comprising shape coefficients, vein patterns, and texture descriptors were fed into k-Nearest Neighbors (k-NN) and Decision Tree classifiers. The research highlighted

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 206-212
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0028

computational efficiency with processing times under 5 seconds per image. However, the system exhibited sensitivity to background complexity and leaf orientation variations, limiting real-world applicability. The transition toward deep learning architectures began with Quoc and Hoang [4], who deployed Convolutional Neural Networks for medicinal plant recognition in natural settings. Their approach utilized an optimized CNN structure incorporating five convolutional layers, incorporating batch normalization and dropout regularization to mitigate overfitting. The study specifically addressed challenges posed by uncontrolled environmental conditions including variable lighting, partial occlusion, and background clutter. By training on 15,000 images across 30 medicinal species, they achieved classification accuracy exceeding traditional methods by 18-22%. The work demonstrated CNN's capability to autonomously capture hierarchical patterns and representations without manual feature engineering. Rani et al. [5] conducted comprehensive comparative analysis of state-of-the-art deep learning architectures for medicinal leaf identification. Their investigation evaluated VGG16, ResNet50, InceptionV3, and MobileNetV2 using transfer learning strategies. The research employed data augmentation techniques including geometric transformations, color jittering, and mixup regularization to enhance model robustness. Performance evaluation revealed ResNet50 achieving superior accuracy (94.7%) attributed to its residual connections supporting deeper architecture training without gradient vanishing. The study established benchmarks for architecture selection in botanical classification tasks and demonstrated the efficacy of fine-tuning pre-trained ImageNet weights on domain-specific datasets. Patil et al. [6] developed FloraMediVision, a comprehensive computer vision system integrating deep learning with user-centric web interfaces. Their architecture employed EfficientNetB0 for feature extraction, optimized for mobile deployment through quantization techniques. The system incorporated a three-stage classification pipeline: leaf segmentation using U-Net architecture, feature extraction via EfficientNet, and species prediction with confidence scoring. Additionally, they integrated a knowledge base providing medicinal properties, phytochemical compositions, and therapeutic applications for identified species. The deployment framework achieved inference times below 2 seconds on standard smartphones, demonstrating practical viability for field applications. While existing research demonstrates significant progress, several limitations persist:

- **Limited Dataset Diversity:** Many studies operate on controlled datasets with uniform backgrounds and lighting conditions, reducing real-world applicability.
- **Architecture Optimization:** Insufficient exploration of progressive fine-tuning strategies that balance between preserving pre-trained knowledge and adapting to domain-specific features.
- **Scalability Concerns:** Most systems focus on fixed species sets without addressing incremental learning for expanding plant databases [7].
- **Deployment Accessibility:** Limited emphasis on creating user-friendly interfaces accessible to non-technical stakeholders including traditional medicine practitioners and healthcare workers.
- **Confidence Quantification:** Inadequate attention to prediction uncertainty estimation, important for medical uses where misidentification carries significant consequences.

This work addresses the identified gaps by implementing a robust deep learning framework with progressive fine-tuning, comprehensive data augmentation, confidence-based prediction mechanisms, and accessible web deployment specifically tailored for medicinal plant identification in diverse real-world scenarios.

3. Proposed Methodology

The proposed work aims to build an end-to-end an automated framework for identifying medicinal plants from leaf photographs using a Convolutional Neural Network (CNN). The system includes several

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 206-212
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0028

major stages: dataset preparation, image preprocessing, model design and training, and finally, web deployment for real-time use. Each module is specifically designed to ensure reliability, adaptability, and ease of access for users outside the technical domain.

3.1 Dataset Description

The experiment uses a publicly available collection of Indian leaf images of medicinal plants covering multiple species captured under different lighting and background conditions. The dataset is divided into three parts-training, validation, and testing in the ratio 80: 10: 10. Each image is treated as a three-channel RGB input to the CNN. The heterogeneity of the dataset in terms of angles, lighting, and texture differences helps the model learn more general and transferable representations, allowing it to perform well in unseen scenarios [8].

3.2 Image Preprocessing

Image preprocessing ensures that all inputs share a uniform format and improves the stability of model training. The following steps are applied sequentially:

- **Resizing** – All images are scaled to 224 × 224 pixels so that the input dimension remains constant.
- **Normalization** – Pixel intensities are divided by 255 to bring values into the [0, 1] range.
- **Noise Filtering** – Minor background artifacts are reduced using a light smoothing filter.
- **Augmentation** – Random transformations such as rotation, brightness shifts, zooming, and horizontal flipping expand the dataset virtually, improving the model's ability to generalize to real-world variations.

This pipeline minimizes the effect of camera quality, lighting, and orientation on prediction performance.

3.3 CNN-Based Feature Extraction and Classification

Convolutional Neural Networks are designed to automatically discover useful spatial features from images without manual engineering. The architecture starts with convolutional layers designed to detect basic visual characteristics, including edges and textures. As depth increases, these layers learn complex structures like venation patterns or margin details that are unique to individual plant species [9]. Once the convolution and pooling steps are performed, the resulting feature representations are flattened and processed through dense layers, concluding with a Softmax activation to determine class probabilities. The learning objective is to minimize the categorical cross-entropy loss:

$$L = -\sum_{k=1}^{K} y_k \log(\hat{y}_k)$$

where y_k represents the true class indicator and $\hat{y}_k$ is the model's predicted probability for class k. Optimization is executed with the help of the Adam algorithm, and learning-rate scheduling is applied to maintain stable convergence throughout training.

3.4 Model Training and Performance Evaluation

Training is carried out in three stages.

- **Base Training:** Initial layers are frozen, allowing only the classifier head to learn from the dataset.
- **Fine Tuning:** Selected convolutional blocks are unfrozen for domain-specific adaptation.
- **Final Optimization:** Hyperparameters such as batch size, learning rate, and dropout rate are tuned based on validation performance [10].

Model effectiveness is measured using standard metrics—accuracy, precision, recall, and F1-score—defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP, TN, FP, and FN indicate true positives, true negatives, false positives, and false negatives, respectively.

3.5 Web Application Deployment

To make the trained model accessible, it is integrated into a lightweight web interface built using Flask. The application allows users to upload or capture a leaf image, after which the backend performs

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 206-212
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0028

prediction and returns:

- The common and scientific name of the detected plant
- A brief description of its medicinal benefits and applications
- The confidence score indicating prediction reliability [11]

This deployment enables real-time identification directly from a browser or mobile device, serving as a connection between contemporary machine learning models and everyday healthcare usability.The complete workflow of the proposed system, covering data preparation, model training, deployment, and real-time prediction, is illustrated in Figure 1.

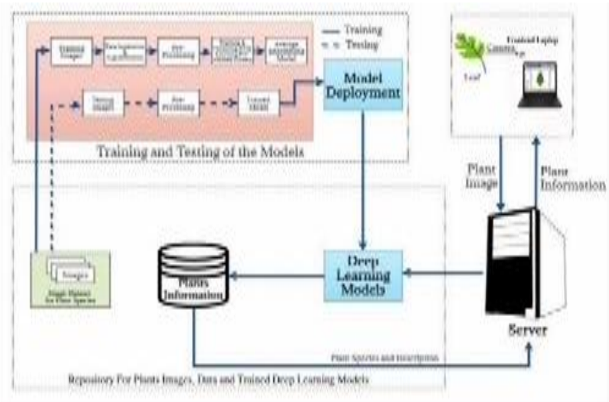

Figure 1 Overall System Architecture for Medicinal Plant Detection and Information Retrieval

4. Expreimental Setup & Implementation

The designed framework was executed using the Python 3.8, TensorFlow, and Keras libraries on a workstation equipped with an Intel i5 processor, 8 GB RAM, and NVIDIA GPU acceleration. Data preprocessing and augmentation were performed using OpenCV and Pillow, while Matplotlib and NumPy were used for visualization and data handling. The available data were distributed into training (80%), validation (10%), and testing (10%) subsets. Each input image was scaled to 224 × 224 pixels and normalized between 0 and 1. Training was conducted over 20 epochs using the Adam optimization algorithm with a categorical cross-entropy objective and an initial learning rate of 0.001. To ensure consistent convergence, the learning rate was dynamically adjusted based on validation results. During training, early stopping and model checkpoint techniques were applied to prevent overfitting and retain the best-performing model. The evaluation metrics included accuracy, recall, precision, and F1-score, which were computed on the unseen test data. After successful training, the optimized model was deployed using a Flask-based web application. The interface allows users to upload or capture leaf images for real-time identification. Upon prediction, the system displays the plant's common and scientific names, medicinal properties, and the associated confidence score. This setup enables practical deployment for both research and healthcare applications [12].

5. Results and Analysis

The experimental evaluation of the suggested framework was carried out to assess its performance in accurately identifying medicinal plant species from leaf images. The CNN model underwent training and validation using the prepared dataset, which incorporated multiple species with diverse color tones, shapes, and vein structures. Through iterative training and fine-tuning, the network demonstrated consistent improvement in both training and validation accuracy, converging to a stable performance level without signs of overfitting. The model, after training, obtained a strong capability to distinguish between morphologically similar species, confirming the effectiveness of deep convolutional representations in capturing complex botanical features [13].

6. Web Application Deployment

To ensure that the developed deep learning model could be utilized effectively by non-technical users, a web-based application was designed and deployed. The system follows a client–server architecture where the front-end interface communicates with a Flask-based back-end server for real-time plant recognition. The application allows users to upload or capture images of medicinal plant leaves using any device with a browser. Once an image is submitted, it undergoes preprocessing steps such as resizing and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 01 January 2026
Page No: 206-212
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0028

normalization before being passed to the trained CNN model for inference. The backend then returns the predicted plant species along with the associated confidence level and relevant medicinal information, which is displayed through an intuitive and responsive web interface. The front-end of the application was developed using standard web technologies including HTML5, CSS3, and JavaScript, ensuring compatibility across different platforms and screen sizes. The Flask framework, implemented in Python, manages communication between the user interface and the deep learning model. The trained CNN model was exported in TensorFlow's SavedModel format, allowing efficient loading and inference during runtime. This setup ensures that the prediction pipeline remains lightweight and responsive while maintaining high accuracy. Experimental evaluation of the deployed model showed an average latency of only a few seconds per image, making the system suitable for real-time use. The web application effectively connects the gap between advanced deep learning models and end users by providing an accessible tool for plant recognition and knowledge dissemination. Through this integration, the framework contributes toward promoting sustainable healthcare practices and supporting research in medicinal botany [14 - 15].

Conclusion and Future Work

This work shows a complete deep learning framework for the automatic detection of medicinal plants based on leaf images. The proposed system employs a Convolutional Neural Network (CNN) architecture combined with an optimized preprocessing and augmentation pipeline to enhance classification accuracy. Through progressive fine-tuning and adaptive learning rate scheduling, the model effectively learns discriminative morphological features such as venation patterns, margins, and texture variations. Experimental results demonstrate that the approach achieves strong generalization and robust performance across multiple plant species. The trained model's deployment through a Flask-based web interface further extends the practical usability of the system. By allowing real-time image uploads and instant predictions, the application bridges the gap between deep learning research and field-level accessibility. This tool can support students, researchers, and healthcare practitioners by providing accurate plant identification along with detailed medicinal properties. Future work will focus on expanding the system's scope to include multiple plant organs such as flowers, stems, and roots to achieve more comprehensive species recognition. Further improvements may involve incorporating attention mechanisms and explainable AI methods to enhance interpretability and reliability. Additionally, integrating multilingual support and mobile deployment could increase accessibility for a wider audience, strengthening the role of artificial intelligence in sustainable healthcare and ethnobotanical research.

References

[1]. D. S. S. V. Krishna, E. V. Kumar, E. M. Goud, and S. V. Y. Dr., "A Novel Approach to Identify Medicinal Plant Based on Image Processing," IRE Journals, vol. 6, no. 11, pp. 87–91, May 2023.

[2]. O. Simandla, S. Ngwenya, and P. Sibandze, "Auto-Detection of Medicinal Plants using Machine Learning Approach," in IST-Africa 2024 Conference Proceedings, 2024. ISBN: 978-1-905824-73-1.

[3]. P. Patil, K. Kulkarni, M. Sonkar, and P. Deshmukh, "FloraMediVision: A Computer Vision-Based System for Medicinal Plant Leaf Identification," in Proc. 6th Int. Conf. Adv. Sci. Technol., 2023, pp. 372–376.

[4]. R. Janani and A. Gopal, "Identification of Selected Medicinal Plant Leaves Using Image Features and ANN," IEEE Xplore, accessed Apr. 2025.

[5]. L. Rani, A. G. Karegowda, D. G., V. S., V. H. R., and S. Bhat, "Identification of Medicinal Leaves Using State of Art Deep Learning Techniques," in Proc. 2022 IEEE ICDCECE, pp. 1–6, 2022.

[6]. T. N. Quoc and V. T. Hoang, "Using Convolutional Neural Networks for

Medicinal Plant Recognition in Natural Settings," in Proc. Int. Conf. on Information and Communication Technology Convergence (ICTC), Jeju, Korea, 2020, pp. 1–6.

[7]. K. Kavitha, R. V. S. Lalitha, P. Sharma, and S. Gupta, "Medicinal Plant Species Detection using Deep Learning," in Proc. 2022 IEEE ICEEICT, pp. 1–5, 2022.

[8]. G. Priyanga, A. Aravindh, K. I. Ashwin, V. G. Shekar, and M. Manjunathan, "Automated Identification of Medicinal Plants Using Machine Learning," IJCRT, vol. 12, no. 3, pp. 2777–2783, Mar. 2024.

[9]. N. G. Gavhale and A. P. Thakare, "Identification of Medicinal Plant Using Machine Learning Approach," International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 7, pp. 1116 1118, Jul. 2020.

[10]. A.Begue, V. Kowlessur, U. Singh, F. Mahomoodally, and S. Pudaruth, "Automatic Recognition of Medicinal Plants Using Machine Learning Techniques," IJACSA, vol. 8, no. 4, pp. 166–173, 2017.

[11]. R. Meenakshi, S. Priya, and M. Kumar, "Deep Learning based Medicinal Plant Recognition using CNN," Int. J. of Computer Applications, vol. 176, no. 35, pp. 12–16, 2020. [12] R. Sharma and A. Kumar, "Texture and Color Feature Based Plant Recognition using SVM," Procedia Computer

[12]. R. Sharma and A. Kumar, "Texture and Color Feature Based Plant Recognition using SVM," Procedia Computer Science, vol. 165, pp. 539 546, 2019.

[13]. R. Patil, T. Singh, and H. Kaur, "Medicinal Plant Classification Using Transfer Learning," Journal of Intelligent & Fuzzy Systems, vol. 41, no. 2, pp. 1237–1244, 2021.

[14]. A. Ali, V. Kumar, and R. Singh, "Automatic Identification of Indian Medicinal Plants using Deep CNNs," Computers and Electronics in Agriculture, vol. 195, p. 106841, 2022.

[15]. R. Rajeswari, K. Vani, and R. Menaka, "Hybrid Feature Extraction for Plant Recognition using Machine Learning," International Journal of Engineering and Advanced Technology, vol. 9, no. 6, pp. 1124 1130, 2020