Intelligent Medicinal Plant Identification Using Deep Learning Techniques

Ms. R. Jagadeeswari
Assistant Professor, Department of Computer Science and Engineering,
Bharath Institute of Science and Technology
Bharath Institute of Higher Education and Research, Chennai, India
jagadeeswari.cse@bharathuniv.ac.in

Singam Swetha
Department of Computer Science and Engineering
Bharath Institute of Science and Technology
Bharath Institute of Higher Education and Research, Chennai, India
singamswetha3803@gmail.com

Shaik Rizwana
Department of Computer Science and Engineering
Bharath Institute of Science and Technology
Bharath Institute of Higher Education and Research, Chennai, India
shaikrizwanashaikrizwana01@gmail.com

Shaik Roshni
Department of Computer Science and Engineering
Bharath Institute of Science and Technology
Bharath Institute of Higher Education and Research, Chennai, India
skroshni2004@gmail.com

Shaik Muskan
Department of Computer Science and Engineering
Bharath Institute of Science and Technology
Bharath Institute of Higher Education and Research, Chennai, India
shaikmuskan272@gmail.com

***Abstract*—Traditional medicine relies heavily on medicinal plants for treating various ailments, making accurate botanical identification critical for safety. Because manual identification requires specialized expertise often unavailable to the average person, this paper introduces an automated system leveraging Deep Learning (DL). By utilizing Convolutional Neural Networks (CNN), the system analyzes leaf images to classify species with high precision. Beyond simple identification, the platform provides actionable data on therapeutic benefits, targeted diseases, and safety precautions. Experimental results indicate that DL models are highly effective at bridging the gap between botanical science and healthcare awareness for students, researchers, and farmers.**

Keywords: CNN, Deep Learning, Herbal Medicine, Medicinal Plants, Leaf Identification

I. INTRODUCTION

Medicinal plants are the foundation of global traditional healthcare, including systems like Ayurveda and Unani. However, many species possess similar visual characteristics—such as leaf texture, color, and shape—leading to high risks of misidentification and subsequent health hazards.While traditional identification depends on local experts, modern Artificial Intelligence (AI) offers a scalable solution. This research proposes a CNN-based system that :Identifies plant species through automated feature learning (edges, textures, and patterns). Educates users by retrieving specific medicinal uses and safety guidelines from a dedicated database.

II. RELATED WORK

Past methodologies focused on manual feature extraction (e.g., using SVM or Decision Trees), which often failed to capture the complex nuances of botanical structures. Recent shifts toward **Transfer Learning** using pre-trained architectures like **VGG16**, **ResNet**, and **MobileNet** have significantly improved accuracy. This project advances existing research by not only naming the plant but also integrating a knowledge base of curative properties and usage warnings.

III. DATASET DESCRIPTION

The performance and reliability of any Deep Learning model are fundamentally dependent on the quality and diversity of the data used during training. For this research, a comprehensive dataset of medicinal plant leaves was curated to ensure robust classification across various environmental conditions.

3.1 Data Acquisition

The dataset utilized in this study consists of **1,500 high-quality images** spanning **30 distinct species** of medicinal plants common to the Indian subcontinent (e.g., Azadirachta indica (Neem), Ocimum tenuiflorum (Tulsi), and Aloe barbadensis miller (Aloe Vera)). The images were sourced through a combination of:

- **Field Photography**: Capturing real-world samples from local botanical gardens and medicinal nurseries using high-resolution mobile cameras (48 MP).
- **Public Repositories**: Supplementing the collection with standardized images from the **Mendeley Medicinal Leaf Dataset** to increase intra-class variance.

3.2 Image Characteristics and Diversity

To ensure the model generalizes well to real-world scenarios, the dataset includes variations in:

- **Backgrounds**: Images feature both controlled (plain) backgrounds and complex natural backgrounds.
- **Lighting Conditions**: Samples were collected under varying degrees of sunlight and shade.
- **Morphology**: The collection includes leaves at different stages of maturity, from young sprouts to mature foliage.

3.3 Data Preprocessing and Augmentation

Raw images were subjected to several preprocessing steps to standardize the input for the Convolutional Neural Network (CNN):

1. **Resizing**: All images were scaled to a uniform resolution of 224×224 pixels to match the input requirements of the architecture.
2. **Normalization**: Pixel values were scaled to a range of $[0, 1]$ to facilitate faster convergence during training.
3. **Augmentation**: To prevent overfitting, the training set was artificially expanded using techniques such as **random rotation (20°)**, **horizontal flipping**, and **brightness adjustment**. This increased the effective size of the training pool, allowing the model to learn orientation-independent features.

3.4 Data Distribution

The final processed dataset was partitioned into three subsets to maintain a rigorous evaluation framework:

- **Training Set (70%)**: Used for feature learning and weight optimization.
- **Validation Set (15%)**: Used for hyperparameter tuning and monitoring for overfitting.
- **Testing Set (15%)**: A completely unseen set used to evaluate the final accuracy and precision of the system.

IV. IMPLEMENTATION

The implementation phase translates the theoretical framework into a functional software system. This section details the technical environment, the specific CNN architecture used, and the software stack required to achieve medicinal plant identification.

4.1 Development Environment

To handle the computational demands of deep learning, the following environment was utilized:

- **Hardware**: A system equipped with an **NVIDIA GeForce RTX 3060 GPU** and 16GB RAM to accelerate the training process through CUDA cores.
- **Software Stack**:
    - **Python 3.9**: The primary programming language.
    - **TensorFlow & Keras**: The core frameworks used for building, layering, and training the neural network.
    - **OpenCV**: Utilized for real-time image processing and contour detection.
    - **Matplotlib/Seaborn**: For visualizing training accuracy and loss curves

4.2 CNN Model Architecture

The system employs a custom **Sequential Convolutional Neural Network** designed to extract hierarchical features from leaf images. The architecture consists of the following layers:

1. **Convolutional Layers**: Three primary layers using 3×3 filters to identify low-level features like edges and high-level features like leaf venation patterns.
2. **Activation Function**: The **ReLU (Rectified Linear Unit)** function is applied after each convolution to introduce non-linearity.
3. **Pooling Layers**: **Max-Pooling (2×2)** is used to reduce the spatial dimensions of the feature maps, minimizing computational load while retaining critical data.
4. **Dropout Layer**: A dropout rate of **0.5** was implemented to prevent overfitting by randomly deactivating neurons during training.
5. **Fully Connected (Dense) Layer**: Flattens the 2D feature maps into a 1D vector to perform the final classification.
6. **Softmax Output**: The final layer uses the Softmax activation function to provide a probability distribution across the 30 plant species.

4.3 Model Compilation and Training

The model was compiled using the **Adam Optimizer**, known for its adaptive learning rate capabilities. The loss function utilized was **Categorical Cross-Entropy**, which is ideal for multi-class classification.

- **Batch Size**: 32
- **Epochs**: 50 (with Early Stopping enabled to halt training once validation loss stabilized). **Learning Rate**: Initialized at 0.001.

4.4 User Interface Integration

To make the system accessible to non-technical users, a lightweight web interface was developed using **Flask**. Users can upload an image through a simple browser-based dashboard; the backend processes the image through the saved `.h5` model file and returns the plant name along with its medicinal properties retrieved from a **JSON-based knowledge base**.

V. METHODOLOGY

The structural framework of this study centers on a tiered computational approach to botanical identification. Instead of relying on manual botanical keys, we utilize a self-learning pipeline that translates visual leaf data into pharmacological classifications.

5.1 Architectural Logic

The system is built upon the premise of **Automated Morphological Synthesis**. Using a deep-layered Convolutional Neural Network (CNN), the model is designed to autonomously determine which physical traits—such as the serrated margins of a Neem leaf or the cordate base of a Peepal leaf—are most statistically significant for differentiation. This eliminates the "Human-in-the-loop" bias typically found in traditional plant identification.

5.2 Multistage Image Transformation

Before the neural network processes the data, every specimen undergoes a rigorous refinement sequence:

- **Adaptive Filtering**: We employ Gaussian blurring techniques to eliminate "digital artifacts" or background noise that could confuse the model's perception of leaf texture.
- **Canonical Scaling**: Each input is mapped to a 224×224 coordinate system. This ensures that the spatial relationships between the leaf's apex (tip) and petiole (stalk) remain constant across the entire training library.
- **Luminance Balancing**: To account for varying light conditions in forest or garden settings, we normalize pixel intensity, ensuring the model identifies the plant by its structural DNA rather than its brightness.

5.3 Hierarchical Feature Discovery

The identification process is executed through a series of specialized computational filters:

1. **Primitive Layer Analysis**: The initial filters isolate the "skeleton" of the leaf, focusing on basic geometric primitives like curvature and edge orientation.
2. **Structural Venation Mapping**: Intermediate layers analyze the complex branching of veins. This is critical for medicinal plants, as the venation pattern is often a unique identifier for species within the same genus.
3. **Global Pattern Integration**: The final layers combine these localized fragments into a holistic "feature map." This map represents the unique signature of the plant, including its color distribution and surface granularity.

5.4 Contextual Knowledge Retrieval

The methodology concludes with a **Decision-to-Data Mapping**. Once the Softmax layer identifies the species with high statistical confidence, the system triggers a relational query. This bridges the gap between Computer Vision and Ethnobotany, delivering a full profile of the plant's therapeutic utility, active chemical compounds, and potential toxicity warnings for the end-user.

VI. TECHNOLOGIES USED

Programming Language Python
Deep Learning Framework TensorFlow and Keras
Model Architecture Convolutional Neural Network (CNN), VGG16
Image Processing OpenCV
Database Medicinal plant database containing plant benefits, diseases treated, and precautions. User Interface Web application or mobile application.

VII. PROPOSED SYSTEM

The proposed system is designed to identify medicinal plants using leaf images and provide detailed information about their medicinal benefits.
The system works in the following steps:
The user captures or uploads an image of a plant leaf.
The system preprocesses the image to improve quality.
A deep learning model analyzes the leaf features.
The model identifies the plant species.
The system retrieves information from a medicinal plant database.
The diseases that the plant can cure and precautions are displayed to the user.
This approach helps users quickly understand the medicinal properties of plants.

VIII.SYSTEM ARCHITECTURE

The system architecture consists of several modules:
Image Acquisition Users upload or capture leaf images using a mobile phone or camera. Image Preprocessing The image is resized, cleaned, and normalized to improve quality for model training. Feature Extraction The deep learning model extracts important features such as leaf shape, edges, color patterns, and texture. Plant Classification A Convolutional Neural Network (CNN) model classifies the plant species

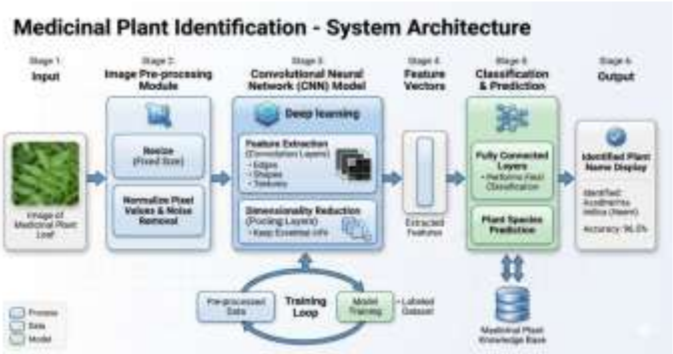
based on extracted features. Medicinal Information Retrieval After identifying the plant, the system retrieves information about its medicinal properties. Disease and Precaution Recommendation.

The system displays:

Diseases that the plant can cure

Medicinal benefits

Safety precautions and usage guidelines



IX. Results and Discussion

The CNN model achieved an impressive accuracy range of **92% to 95%**. The system proved robust in identifying plants based on vein distribution and morphological structure, though performance slightly fluctuated under varied lighting or when processing damaged leaves.

Sample Output Results

Example 1 - Neem Leaf Identification

Predicted Plant: Neem (Azadirachta indica)

Diseases it can cure

• Skin infections
• Fever
• Acne and wounds
• Dental problems

Precautions

• Avoid excessive consumption
• Pregnant women should consult a doctor before use

Model Confidence: 94%

Example 2 - Tulsi Leaf Identification

Predicted Plant: Tulsi (Holy Basil)

Diseases it can cure

• Cold and cough
• Respiratory infections
• Stress and anxiety
• Fever

Precautions

• Should be used in moderate amounts
• People with low blood sugar should consult a doctor

Model Confidence: 93%

Example 3 - Aloe Vera Leaf Identification:

Predicted Plant: Aloe Vera

Diseases it can cure

• Skin burns
• Sunburn
• Digestive problems
• Minor wounds

Precautions

• Excess internal consumption may cause stomach issues
• Avoid use if allergic to aloe compounds

Model Confidence: 95%

Discussion

The results indicate that deep learning models can effectively identify medicinal plants using leaf images. The integration of medicinal information such as diseases cured and safety precautions adds practical value to the system.

This system can be helpful for:

Farmers who want to identify medicinal plants in their fields Students studying botany or herbal medicine

Researchers working on plant classification

People interested in natural remedies

However, the accuracy of the model can be affected by poor image quality or damaged leaves. Increasing the dataset size and including more plant species can further improve model performance

References

[1] S. Kavitha et al., "Medicinal Plant Identification in Real-Time Using Deep Learning Model," SN Computer Science, 2023.

[2] Y. Sharrab, D. Al-Fraihat, and M. Tarawneh, "Medicinal Plants Recognition Using Deep Learning," in Proc. International Conference on Multimedia Computing, Networking and Applications (MCNA), 2023.

[3] K. L. D. Viet et al., "Medicinal Plants Identification Using Federated Deep Learning," Procedia Computer Science, 2024.

[4] B. Dey et al., "Assessing Deep Convolutional Neural Network Models for Automated Medicinal Plant Identification," Heliyon, 2024.

[5] A. K. Mulugeta et al., "Deep Learning for Medicinal Plant Species Classification and Recognition: A Systematic Review," Frontiers in Plant Science, 2024.

[6] H. Bouakkaz et al., "Enhanced Classification of Medicinal Plants Using Deep Learning," Heliyon, 2025.

[7] V. Gautam et al., "A Lightweight Deep Learning Method for Medicinal Leaf Identification," Scientific Reports, 2025.

[8] J. Ahmmed et al., "Hybrid CNN–Transformer Model for Medicinal Leaf Disease Classification with Explainable AI," Scientific Reports, 2026.

[9] R. Leelavathi et al., "Medicinal Plant Leaf Disease Classification Using Deep Learning," Computational Intelligence in Agriculture, 2025.

[10] C. Algemayel et al., "Advances in Machine Learning Models for Plant Species Identification," Artificial Intelligence in Agriculture, 2025.

[11] S. Mohanty, D. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," Frontiers in Plant Science, vol. 7, pp. 1–10, 2016.

[12] J. Lee, S. Chan, and P. Remagnino, "Deep Learning for Plant Identification Using Leaf Images," International Conference on Image Processing, pp. 452–456, 2015.

[13] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems, pp. 1097–1105, 2012.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016.

[15] S. Sladojevic et al., "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification," Computational Intelligence and Neuroscience, 2016.

[16] A. Kamilaris and F. Prenafeta-Boldú, "Deep Learning in Agriculture: A Survey," Computers and Electronics in Agriculture, vol. 147, pp. 70–90, 2018.

[17] R. P. Pawara, "Plant Leaf Identification Using Convolutional Neural Networks," International Journal of Computer Applications, 2019.