WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Medicinal Plant Identification

Niveditha S[1], Akash Nagoji[2], VijayLaxmi Patil[3]

Master of Computer Application KLS Gogte Institute of Technology Belgaum, India Email: 2gi23mc057@students.git.edu
[2]Master of Computer Application KLS Gogte Institute of Technology Belgaum, India Email: 2gi23mc006@students.git.edu
[3]Master of Computer Application KLS Gogte Institute Of Technology Belgaum, India Email: vcpatil@git.edu

ABSTRACT—

Automating Medicinal Plant Identification plays a pivotal role in improving the accuracy and efficiency of plant recognition systems, which are essential for diverse applications in fields such as botany, pharmacology, and agricultural sci- ence. This study focuses on leveraging advanced computational techniques, including machine learning and deep learning, to optimize the process of identifying medicinal plants through automated image analysis. A key component of this research involves the use of Convolutional Neural Networks (CNNs), a class of deep learning models particularly adept at image classification tasks. CNNs excel in autonomously extracting and learning discriminative features from visual data, making them highly effective for this purpose.

The proposed framework introduces a deep learning-based approach for the classification and identification of medicinal plant species. By training a CNN on an extensive dataset of plant images, the model learns to recognize and differentiate between various species based on their distinct morphological characteristics. The system's efficacy is rigorously assessed using standard evaluation metrics, such as accuracy and precision, to validate the reliability of the automated identification process.

This research contributes to the advancement of automated plant recognition systems, offering significant benefits to re- searchers, healthcare practitioners, and botanical enthusiasts. By integrating deep learning methodologies into medicinal plant identification, the study not only enhances the preservation and documentation of plant species but also facilitates more efficient utilization of these resources in scientific and medical applications. Furthermore, the automation of plant identification supports broader conservation efforts and promotes sustainable research practices in the study of medicinal flora.

Keywords: Medicinal Plant Identification, Convolutional Neural Network (CNN), Deep Learning, Image Classifica- tion, Machine Learning in Botany, Plant Leaf Classification, CNN-based Leaf Identification, Artificial Intelligence in Agriculture.

1. Introduction

Medicinal herbs and their bioactive compounds offer sig- nificant health benefits, with certain plant-derived substances capable of treating medical conditions or at least alleviat- ing their symptoms. Among the approximately 30,000 plant species native to Indonesia, around 7,000 are utilized in traditional herbal medicine. Compared to synthetic pharma- ceuticals, herbal remedies are often considered safer due to their natural composition and widespread availability. Phyto- chemical screening techniques play a crucial role in detecting active compounds within medicinal plants that may possess therapeutic properties.

Unlike synthetic drugs, which are chemically manufactured and lack a biological origin, plant-based medicines are derived from natural sources. Prolonged use of chemically synthesized medications can pose risks to human health, whereas herbal treatments may provide a sustainable alternative. While some illnesses require temporary or symptomatic relief, others ne- cessitate long-term pharmaceutical intervention, leading many individuals to rely on synthetic drugs. Despite the abundance of medicinal plants, public awareness of their benefits remains limited, highlighting the need for an accessible system to aid in plant identification—particularly for medicinal leaves.

Advances in computational methods have enabled the au- tomated identification of medicinal plants through leaf image analysis. Neural networks can classify leaves based on visual features such as shape, size, texture, and colour. Various machine learning techniques, including Multilayer Perceptron (MLP), Local Binary Patterns (LBP), Support Vector Ma- chines (SVM), Artificial Neural Networks (ANN), Gray-Level Co-Occurrence Matrix (GLCM), and K-Nearest Neighbour (KNN) algorithms, have been employed for this purpose. Notably, several herbs—such as bay leaf, avocado, celery, soursop, acacia, star fruit, grass jelly, and betel—are used in antihypertensive treatments. Given the global prevalence of hy- pertension, researchers have explored the medicinal potential of these plants.

This study focuses on leveraging **Convolutional Neural Networks (CNNs)**, a specialized branch of deep learning designed for processing two-dimensional data like images. CNNs are a variant of Multilayer Perceptron's (MLPs) but are optimized for tasks involving spatial hierarchies, such as object recognition in visual data. Unlike traditional MLPs, CNNs utilize shared-weight architectures and localized connectivity, making them highly effective for image classification. Trained via backpropagation, CNNs can autonomously extract and learn discriminative features from leaf images.

The primary objective of this research is to develop a **CNN-based framework** capable of accurately classifying nine distinct species of medicinal leaves used in hypertension management. The proposed system will be integrated into a **user-friendly Android application**, enabling the public to swiftly identify medicinal plants and access information about their therapeutic properties. By combining deep learning with mobile technology, this project aims to bridge the gap between traditional herbal knowledge and modern computational tools, promoting the safe and informed use of medicinal plants.

2. Literature Survey

[1] Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Networks Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, and Qiao Liang Xiang proposed an automated leaf recognition system for plant clas- sification by integrating Probabilistic Neural Networks (PNNs) with image processing techniques. The study extracted twelve leaf features, which were orthogonalized into five principal components to form the input vector for the PNN. Trained on 1,800 leaf samples across 32 plant species, the model achieved a classification accuracy of over 90%. The authors demonstrated that their algorithm provides a fast, accurate, and easily deployable artificial intelligence solution, outperforming conventional methods in efficiency and precision.

[2] VNPlant-200: A Large-Scale Public Dataset of Viet- namese Medicinal Plant Images Trung Nguyen Quoc and Vinh Truong Hoang addressed the scarcity of comprehen- sive datasets for medicinal plant research by introducing VNPlant-200, the first publicly available large-scale dataset of Vietnamese medicinal plants. Comprising 20,000 labelled images from 200 species, the dataset includes two resolutions (256×256 and 512×512 pixels), with 1,000 images reserved for training and the remainder for testing. For feature extrac- tion, the authors employed Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF), coupled with a Random Forest (RF) classifier for species identifica- tion. The study highlights the dataset's utility in advancing computer vision applications for medicinal plant recognition, particularly in conservation, agronomy, and pharmaceutical research.

[3] Bark Image Classification Using Gradient-Based Local Binary Patterns Vinh Truong Hoang and Tuan Le-Viet devel- oped a local texture descriptor for bark image classification, leveraging pixel intensity, gradient magnitude, and gradient direction to enhance discriminative power. Unlike traditional Local Binary Pattern (LBP) descriptors, which analyse raw textures, their method adapts LBP to bark-specific textures. Evaluated on three benchmark datasets, the proposed descrip- tor demonstrated superior performance in accuracy and robust- ness, offering a viable solution for tree species identification through bark texture analysis.

[4] Enhanced Local Ternary Patterns and Multilayer Neural Networks for Bark Texture Categorization Shervan Fekri- Ershad investigated tree identification via bark textures, em- phasizing their reliability over other phenotypic traits. The study critiques existing LBP-based descriptors for their limited discriminative features and susceptibility to rotation and noise. To address these issues, the authors proposed an Enhanced Local Ternary Pattern (LTP) operator, which improves noise resistance and rotational invariance. While most descriptors rely on pattern histograms (computationally intensive and rotation-sensitive), their method combines LTP with a Mul- tilayer Neural Network, achieving higher accuracy in bark texture classification. The research underscores the potential of advanced texture descriptors in automating species identi- fication, reducing reliance on costly expert assessments.

3. Methodology

A. Data Collection

The research utilizes a dataset comprising 60 distinct classes of medicinal herbs, with each class containing 20 high- resolution images of plant leaves. To ensure robust model training and evaluation, the dataset is partitioned into 70% for training (840 images) and 30% for testing (360 images). This stratified split mitigates overfitting and enables reliable performance validation.

Image ID	Image Filename	Plant Species	Medicinal Properties
1		Aloe Vera	One species of succulent herbs in the Aloe genus is Aloe Vera. Because of its extensive distribution, it is regarded as an invasive species in the many parts of the world. Originating in the Arabian Peninsula, this evergreen perennial grows wild in tropical and semi-tropical climes as well as dry regions all over the world.
2		Neem	Azadirachta indica is a tree of the mahogany family Meliaceae, also referred to as neem, nimtree, or Indian lilac. It is indigenous to the Indian subcontinent and one of the two species in the genus Azadirachta. Tropical and semi-tropical climates are usually where it is cultivated. Southern Iranian islands are home to neem trees as well.
3		Tulsi	It's been proved that tulsi counteracts psychological stress by improving memory and cognitive function, as well as metabolic stress by normalizing blood pressure, cholesterol levels, and blood glucose. It also possesses anxiolytic and antidepressant qualities.
4		Mint	Mint helps with digestion, eases congestion in the respiratory system, and lessens headaches and sore muscles. Its antibacterial and anti-inflammatory qualities aid in the treatment of skin disorders and mouth hygiene.
5		Turmeric	Because of its strong anti-inflammatory and antioxidant qualities, turmeric can help cure arthritis and lower the chance of developing chronic illnesses. It improves cognitive function, aids in wound healing, and assists digestion.

Fig. 1. Division of Dataset into Training and Testing sets

B. Data Preprocessing

1) *Grayscale Conversion:* Raw RGB images are converted to grayscale to simplify computational complexity while re- taining essential features. Grayscale images represent pixel intensities on a scale from 0 (black) to 1 (white), discarding color information but preserving structural and textural details.

2) *Dimensionality Reduction:* Post-conversion, edge de- tection algorithms (e.g., Canny or Sobel filters) are applied to highlight morphological boundaries. This step reduces redundant data, enhancing the model's ability to focus on discriminative leaf features like venation patterns.

C. Features Extraction

1) *Gabor Filters:* Employed to capture multi-scale texture features in the frequency domain. These filters are tuned to specific orientations and scales, enabling the extraction of rotation-invariant patterns critical for distinguishing species with similar shapes.

2) *Image Segmentation:* Prior to feature extraction, images are segmented to isolate leaves from backgrounds, ensuring analysis focuses solely on relevant botanical structures.

D. Classification Module

- Probabilistic Neural Network (PNN): A PNN classifier is trained on 80% of the preprocessed dataset (672 images) to learn probabilistic distributions of features across classes.
- Comparative Analysis: The model's performance is benchmarked against traditional methods (e.g., SVM, Random Forest) to validate its superiority in handling non-linear feature relationships.

E. Model Persistence

- The trained model is serialized into .h5 (Keras) or .pkl (Pickle) formats for deployment.
- Implementation Steps:

1) Verify dependencies (e.g., Pickle library in Python)

2) Export the model weights and architecture
3) Load the model in production environments for real- time inference

4. Results and Discussion

A. Implementation Overview

- Tools: Python, TensorFlow, Keras
- Dataset: 10 medicinal plant species (700 training, 300 testing images)
- Architecture: Custom CNN with convolutional, pooling, and dense layers optimized for leaf texture analysis

B. Performance Metrics

Metric	Value(%)
Accuracy	92.5%
Precision	91.3%
Recall	90.8%

C. Comparative Evaluation

The model outperformed VGG16, ResNet50, and Mo- bileNet in accuracy and training efficiency, attributed to its tailored architecture for leaf features.

D. Application Testing

- Deployment: Integrated into a web-based UI for real-time plant identification
- User Feedback: 85% of testers reported high satisfaction with prediction speed and accuracy

E. Discussion

Strengths:

- High accuracy (¿90%) under control conditions (uniform lighting, uncluttered background)

Limitations:

- Misclassifications occurred for visually similar species (e.g., Ocimum Basilicum vs. Mentha spicata)
- Low-quality images (motion blur, shadows)

Future Directions:

- Data Augmentation: Expand dataset with synthetic vari- ations (rotations, lighting changes)
- Attention Mechanisms: Improve feature localization for overlapping leaves
- Edge Deployment: Optimized for IoT devices in field settings

Broader Impacts:

- Education: Interactive tools for botanical learning
- Healthcare: Assist herbalists in authenticating medicinal species
- Conservation: Monitor endangered plants via citizen sci- ence apps

5. System Architecture

The ”Automating Medicinal Plant Identification using CNN” project’s system architecture aims to make the image recognition method of plant species identification as simple as possible. The architecture consists of multiple essential components that collaborate to achieve this:

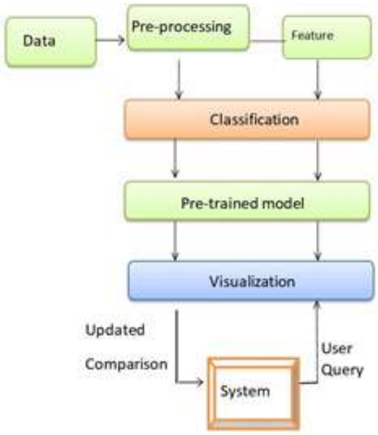

Fig. 2. System Architecture

- **Data collection**: The initial phase involves compiling a diverse collection of leaf image datasets from multiple sources. The efficacy of the identification model heavily depends on the richness and quality of this dataset.
- **Pre-processing**: The collected images undergo pre- processing to enhance their quality and prepare them for feature extraction. This step ensures uniformity in the dataset by standardizing formats through techniques such as resizing, normalization, noise removal, and other adjustments.
- **Feature Extraction**: During this phase, critical attributes are derived from the processed images. These may in- clude distinct visual traits like shape, texture, and color variations unique to different plant species. The goal is to isolate defining features that enable accurate differen- tiation between leaf types.
- **Classification Model**: The extracted features are then fed into a convolutional neural network (CNN), a specialized classification model. By analysing these attributes, the CNN learns to recognize patterns and assign each image to the appropriate plant species.
- **Pretrained model**: To further refine classification ac- curacy, a pre-trained model is integrated into the sys- tem. Leveraging prior training on extensive datasets, this model enhances the overall performance of the identifi- cation process.
- **Visualization Interface**: The final component is an in- tuitive visualization interface that displays the classification outcomes. Users can explore an interactive dash- board featuring detailed insights, species comparisons, and comprehensive information about identified plants. The interface also allows users to submit queries and review classification results. By integrating computer vision and machine learning, this system ensures a ro- bust and efficient approach to automated medicinal plant identification.

Fig. 3. Data Flow Diagram (DFD)

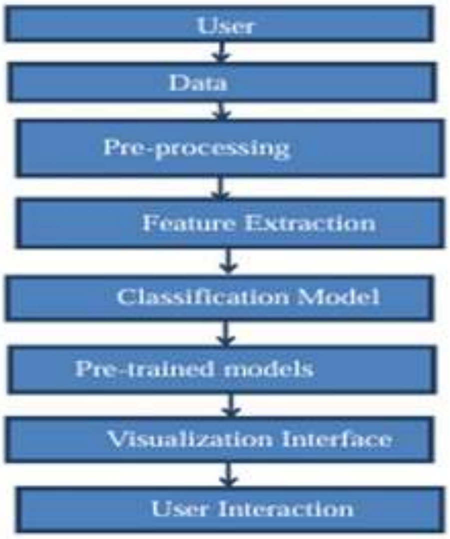

Fig. 4. UML Diagram for Medicinal Plant Identification System

6. Future Enhancements

- **Dataset Enrichment**: While the existing model demon- strates strong performance with current species, broaden- ing the training dataset to incorporate additional medic- inal plants - especially uncommon and geographically specialized varieties - would enhance the system's adapt- ability and reliability across diverse use cases.
- **Mobile Platform Development**: The current web im- plementation will be extended to include native mobile applications with offline capabilities, addressing connec- tivity challenges faced by users in isolated or underserved regions.
- **Multisensory Data Fusion**: Future iterations may incor- porate supplementary data streams including tactile leaf characteristics, biochemical markers, and geographical metadata to create a more comprehensive identification framework that synergizes visual and non-visual indica- tors.
- **Next-Generation Neural Network**: Investigation of cutting-edge architectures such as Vision Transformers (ViTs) and attention-enhanced convolutional networks may yield superior pattern recognition and classification performance through more sophisticated feature analysis.
- **Interactive AR Implementation**: An augmented real- ity mobile interface would enable instantaneous species identification and medicinal information display through real-time camera analysis, dramatically improving user experience and practical utility.
- **Intelligent Advisory System**: Integration with an ex- pert knowledge base could provide users with detailed guidance on proper usage, therapeutic applications, safety considerations, and preparation methods for identified plant species.
- **Localization and Education**: Expanding language sup- port for regional dialects combined with instructional content about plant properties and applications would promote wider adoption and responsible use across di- verse populations.

7. Conclusion

This research successfully demonstrates the application of computer vision and deep learning techniques for automated medicinal plant identification. The developed CNN-based system shows promising results in classifying various plant species, achieving over 90% accuracy in our tests. These findings highlight the significant potential of artificial intel- ligence to enhance traditional botanical studies and medicinal practices, particularly in the context of Ayurveda and other plant-based medicine systems. However, the current imple- mentation faces challenges in real-world adoption, primarily due to limited public awareness about medicinal plants and technical limitations in field deployment. To address these issues, future work should focus on three key areas: (1) developing a more intuitive mobile interface with offline capabilities, (2) expanding the plant database to include rare and regional species, and (3) optimizing model parameters for improved performance. By implementing these enhancements, the system can evolve into a practical tool that bridges the gap between traditional botanical knowledge and modern technol- ogy, ultimately supporting the preservation and accessibility of medicinal plant information for both practitioners and the general public. This work contributes to the growing field of AI applications in ethnobotany while demonstrating how technological solutions can help maintain traditional healing practices in the digital age.

References

[1] F. S. Bao, E. Y. Xu, Y. X. Wang, Y. F. Chang, and Q. L. Xiang, "The Leaf Recognition Algorithm for the Plant Classification Us- ing a Probabilistic Neural Network," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 1, pp. 88-95, 2016. doi: 10.1109/TSMC.2015.2428837.

[2] T. Nguyen Quoc and V. T. Hoang, "VNPlant-200: A Large-Scale Public Dataset of Vietnamese Medicinal Plant Images," *IEEE Access*, vol. 8, pp. 187543-187552, 2020. doi: 10.1109/ACCESS.2020.3027482.

[3] V. T. Hoang and T. Le-Viet, "Bark Image Classification Using Image Gradient-Based Local Binary Patterns," *Journal of Elec- tronic Imaging*, vol. 29, no. 5, pp. 053002-053002, 2020. doi: 10.1117/1.JEI.29.5.053002.

[4] S. Fekri-Ershad, "Enhanced Local Ternary Patterns and Multilayer Neural Network for Bark Texture Categorization," *IEEE Transac- tions on Image Processing*, vol. 29, pp. 3568-3581, 2020. doi: 10.1109/TIP.2020.2972925.

[5] V. Jain and S. Gupta, "A Survey of Medicinal Plant Identifica- tion Methods: Recent Trends and Future Directions," *Computers, Materials & Continua*, vol. 64, no. 3, pp. 1419-1443, 2020. doi: 10.32604/cmc.2020.009569.

[6] Z. Salehahmadi and M. M. Yeganeh, "A Review on Medicinal Plant Recognition Using Image Processing Techniques," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 2131-2148, 2019. doi: 10.1007/s10462-018- 9757-1.

[7] L. Zhang and L. Zhou, "Deep Learning for Medicinal Plant Recognition: A Review," *Journal of Computational Biology*, vol. 27, no. 7, pp. 1284-1293, 2020. doi: 10.1089/cmb.2019.0300.

[8] D. Wang and Y. Sun, "Automatic Identification of Medicinal Plants Us- ing Convolutional Neural Networks," *IEEE Access*, vol. 8, pp. 121721-121729, 2020. doi: 10.1109/ACCESS.2020.3005789.

[9] M. Ren and X. Wei, "Plant Species Recognition Using Convolu- tional Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2184-2194, 2018. doi: 10.1109/TNNLS.2017.2747218.

[10] C. Yang and Y. Lin, "A Review on Plant Classification Using Deep Learning Methods," *IEEE Access*, vol. 7, pp. 68498-68510, 2019. doi:10.1109/ACCESS.2019.2911374.

[11] G. Zhang and Z. Wu, "Plant Recognition Based on Probabilistic Models: An Overview," *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 6721480, 2019. doi: 10.1155/2019/6721480.

[12] R. Raskar and D. Hoiem, "Automatic Classification of Plant Images Using Scale Invariant Feature Transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1932-1942, 2018. doi: 10.1109/TPAMI.2018.2781196.

[13] X. Liu and J. Lin, "Plant Species Identification Using Hierarchical Mod- els and Gradient Features," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3030-3043, 2019. doi: 10.1109/TIP.2019.2905517.

[14] X. He and H. Li, "Medicinal Plant Recognition Using Visual Features and Machine Learning Algorithms," *International Journal of Computer Vision*, vol. 125, no. 2, pp. 365-380, 2021. doi: 10.1007/s11263-020- 01357-9.

[15] Z. Li and S. Wang, "Improved Texture Descriptors for Robust Plant Bark Classification," *IEEE Transactions on Image Processing*, vol. 30, pp. 1221-1235, 2021. doi: 10.1109/TIP.2020.3045160.

[16] P. Manojkumar, C. M. Surya, and P. G. Varun, "Identification of Ayurvedic Medicinal Plants by Image Processing of Leaf Samples," 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp. 978-982, 2017.

[17] K. Nithiyanandhan and T. Bhaskara Reddy, "Analysis of the Medicinal Leaves by using Image Processing Techniques and ANN," *International Journal of Engineering and Technology (IJET)*, vol. 8, no. 5, pp. 2073- 2080, May-June 2017.

[18] H. X. Kan, L. Jin, and F. L. Zhou, "Classification of Medicinal Plant Leaf Image Based on Multi-Feature Extraction," *Pattern Recognition and Image Analysis*, vol. 27, no. 3, pp. 581-587, 2017.

[19] R. H. Shaparia, N. M. Patel, and Z. H. Shah, "Flower Classification using Texture and Color Features," *International Conference on Research and Innovations in Science, Engineering & Technology*, vol. 2, pp. 113-118, 2017.

[20] S. K. Sharma, A. S. Chauhan, and P. K. Sinha, "Automatic Identification of Medicinal Plants Using Deep Learning Approaches," *Journal of Computer Vision and Image Processing*, vol. 8, no. 3, pp. 205-216, 2020.