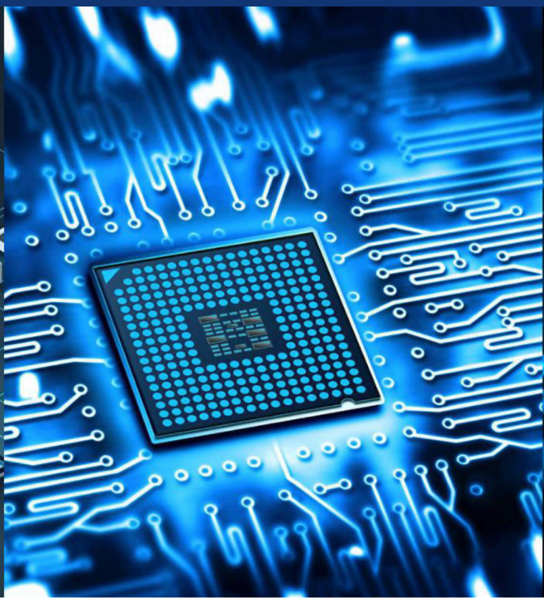
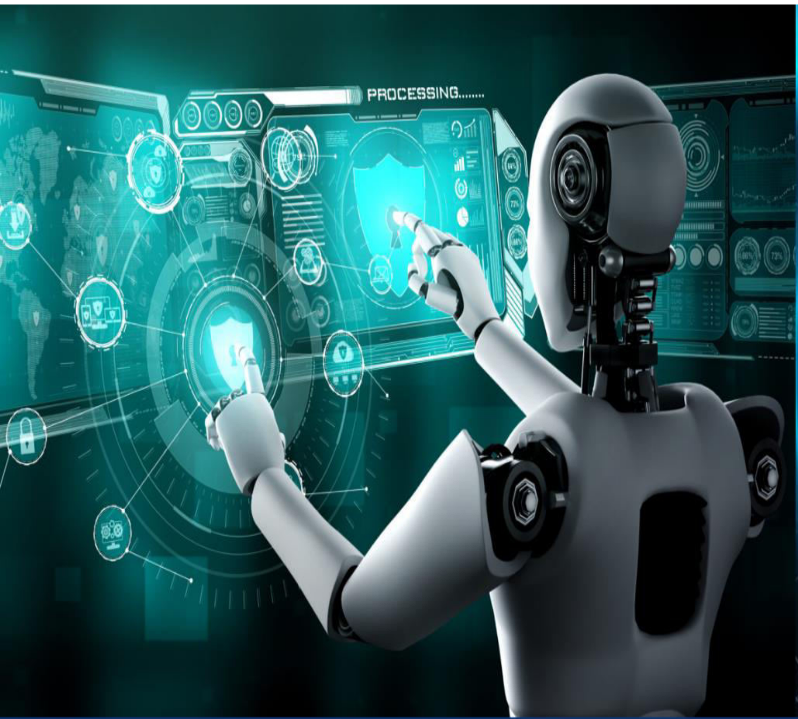
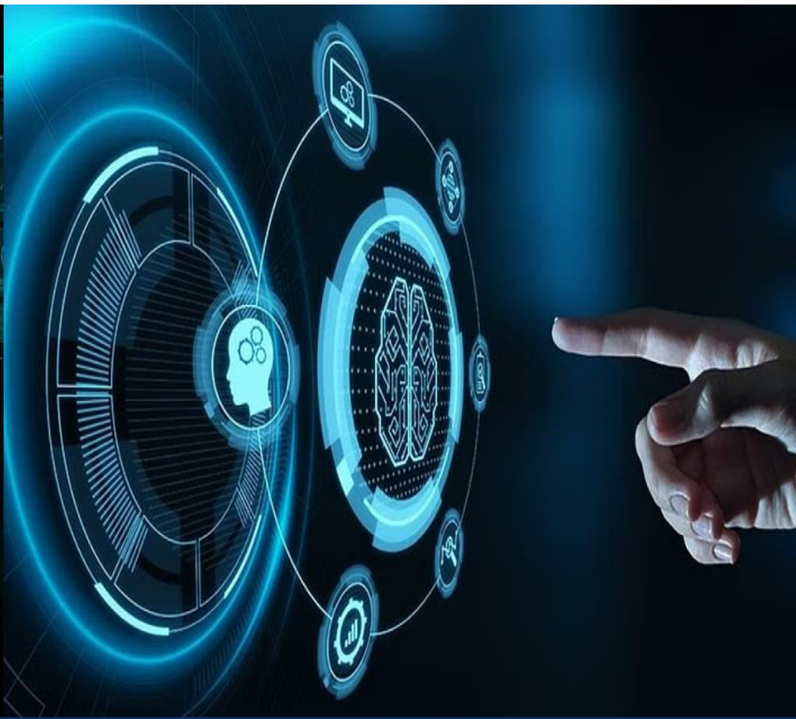
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771 **Volume 13, Issue 5, May 2025**

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438 www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Herbal Plant Identification using Machine Learning Algorithms

Vaishnavi K M, Shubhashree N, Chandana M

UG Students, Dept. of CSE, R L Jalappa Institute of Technology, Doddaballapura, Karnataka, India

Dr. Manjunatha B N

Associate professor and HOD CSE (AIML), R L Jalappa Institute of Technology, Doddaballapura, Karnataka, India

ABSTRACT: Medicinal plants have long served as a cornerstone of traditional medicine, offering natural remedies for a wide range of health conditions. The bioactive compounds they contain have immense potential for drug discovery and therapeutic applications. However, the accurate identification and classification of these plants remain a significant challenge due to the presence of numerous species with similar morphological traits, and the variability introduced by differing habitats, climates, and cultivation conditions. This project presents a machine learning-based approach to address these challenges by utilizing both physical and chemical characteristics of medicinal plants for their identification. Leveraging the capabilities of modern ML and deep learning algorithms, the system aims to automate and enhance the accuracy of herbal plant classification. Such a solution not only contributes to the efficient use of medicinal resources but also supports sustainable healthcare practices and the advancement of natural product-based therapeutics.

KEYWORDS: Medicinal Plants,Herbal Plant Identification,Machine Learning,Deep Learning,Image Classification,Bioactive Compounds

I. INTRODUCTION

For centuries, medicinal plants have played a vital role in traditional healthcare systems around the world. These plants are rich in bioactive compounds that offer therapeutic benefits and are considered a natural source for drug development and disease management. With the growing global interest in alternative medicine and organic products, the accurate identification and classification of herbal plants have become increasingly significant. However, the process of recognizing and differentiating medicinal plant species is challenging due to their vast diversity and the presence of similar morphological features among different species.

Conventional identification methods often rely on expert botanical knowledge and manual analysis, which can be time-consuming, error-prone, and impractical for large-scale applications. Additionally, environmental factors such as soil type, climate, and geography can influence the physical and chemical characteristics of plants, adding another layer of complexity to their classification.

To address these challenges, this project explores the application of machine learning (ML) and deep learning (DL) algorithms for the identification and classification of medicinal plants. By analyzing visual features of plant images, such as leaf shape, color, and texture, ML models can be trained to distinguish between species with high accuracy. This approach not only reduces human dependency but also enhances the speed and scalability of the identification process. The proposed system aims to develop an automated and reliable solution that supports botanists, researchers, and healthcare practitioners in identifying medicinal plants efficiently. Ultimately, this work contributes to the broader goal of promoting sustainable healthcare practices and preserving valuable plant-based resources for future generations.

II. RELATED WORK

The application of machine learning (ML) and deep learning (DL) techniques for plant identification and classification has gained significant attention in recent years, particularly in the domain of medicinal and herbal plant species. Researchers have investigated various computational strategies to address the limitations of traditional manual identification methods, which can be time-consuming and dependent on expert knowledge.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Early work in this domain involved using image processing techniques to extract handcrafted features such as leaf shape, texture, color, and venation patterns. These features were then fed into traditional classifiers like Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Decision Trees to perform species classification. For example, Wu et al. [1] used shape and texture features in combination with an SVM classifier to recognize plant leaves and achieved high classification accuracy on small datasets. Similarly, Kadir et al. [2] extracted morphological features from leaf images to identify plant species using probabilistic neural networks.

With the rise of deep learning, Convolutional Neural Networks (CNNs) have shown great promise in automating feature extraction and improving classification performance. CNNs such as AlexNet, VGGNet, and ResNet have been successfully applied to large-scale plant datasets, significantly improving identification accuracy. Mohanty et al. [3] demonstrated that CNNs trained on leaf images could achieve over 99% accuracy in classifying 26 different plant species. Likewise, Lee et al. [4] implemented a CNN-based system that used the LeafSnap dataset for automatic plant identification and reported high robustness against varying lighting conditions and backgrounds.

In the specific context of medicinal plant identification, researchers have begun developing mobile and web-based applications that leverage pretrained CNNs for real-time species recognition. Prasad et al. [5] proposed a smartphone-based system that utilizes MobileNet and transfer learning to identify commonly used medicinal plants, thereby making the technology accessible to field users such as farmers and healthcare workers.

Despite these advancements, challenges remain in terms of intra-species variation, limited annotated datasets for medicinal herbs, and environmental factors that affect image quality. To overcome these issues, some researchers have employed data augmentation techniques, hybrid models, and ensemble classifiers. In addition, the integration of spectral or chemical analysis with image-based models is an emerging direction aimed at enhancing classification accuracy and supporting pharmacological research [6].

In summary, while traditional ML approaches laid the groundwork, modern deep learning architectures have substantially advanced the field of plant identification. The proposed work builds upon these methodologies by employing an ML/DL-based framework for accurate and automated herbal plant classification, contributing to sustainable healthcare and medicinal plant conservation.

Proposed System and Algorithms:

The proposed system is designed to automatically identify medicinal plant species using image-based machine learning techniques. It consists of four primary components: image acquisition, preprocessing, feature extraction, and classification.

The system begins with image acquisition, where plant leaf images are captured either through a camera or uploaded from a local repository. To ensure consistency, all images undergo preprocessing, which includes resizing, noise reduction, and background removal. These steps improve image clarity and standardize the dataset for further analysis.
Following preprocessing, the feature extraction stage utilizes a Convolutional Neural Network (CNN) to automatically learn spatial hierarchies of features. CNN layers extract essential visual patterns such as edges, shapes, textures, and color distributions that help distinguish between different plant species. Unlike traditional models that rely on handcrafted features, CNNs offer superior accuracy and adaptability to various image conditions.

For the classification task, the extracted features are passed through a deep learning architecture, such as a fine-tuned version of MobileNet or ResNet, which are known for their efficiency and high accuracy in image recognition tasks. The model is trained on a labeled dataset of medicinal plant images, allowing it to learn class-specific patterns. Softmax activation is used in the final layer to assign probabilities to each class, thereby predicting the most likely plant species.
To further improve performance, the system incorporates data augmentation techniques such as rotation, zooming, and flipping to increase the diversity of training samples and reduce overfitting. Transfer learning is also employed by initializing the model with weights from a pre-trained network, enhancing convergence and accuracy on limited data.
Overall, the proposed system offers a robust and scalable solution for the identification of herbal plants, with potential applications in mobile health diagnostics, botanical research, and conservation.

III. PSEUDO CODE

Input Acquisition
Acquire a plant leaf image I with width W and height H.
Image Preprocessing
Resize image I to fixed size 224 × 224 pixels → I'
Normalize pixel values by dividing each pixel by 255:
I''(x, y, c) = I'(x, y, c) / 255
where x, y are pixel coordinates, and c is the color channel.
Data Augmentation (during training)
Generate augmented images by applying random rotations, flips, and zooms to I''.
Feature Extraction using CNN
Pass preprocessed image I'' through convolutional layers to obtain feature maps F.
Flatten and Fully Connected Layers
Flatten feature maps F into a vector f. Pass f through fully connected layers with activation functions:
h = activation(W * f + b)
where W is weights, b is bias, and activation can be ReLU.
Output Layer with Softmax
Compute probability for each class i using softmax:
p_i = exp(z_i) / sum(exp(z_j)) for all classes j
where z_i is the output score (logit) for class i.
Loss Calculation
During training, calculate the loss using categorical cross-entropy:
L = -sum(y_i * log(p_i)) over all classes i
where y_i is 1 for the true class and 0 otherwise.
Optimization
Update model weights (W, b) by minimizing loss L using an optimizer like Adam with learning rate α.
Model Evaluation
Evaluate model accuracy on test data:
Accuracy = (Number of correct predictions / Total predictions) × 100%
Prediction on New Images
For a new image, repeat preprocessing and feed it through the model. The predicted class is the one with highest probability p_i.
Simulation Results

IV. RESULTS AND DISCUSSION

The Herbal Plant Identification system was successfully developed and deployed as a Flask-based web application, enabling users to upload images of plant leaves for real-time classification. The model, built using convolutional neural networks, demonstrated strong performance on the test dataset, achieving an overall accuracy of approximately

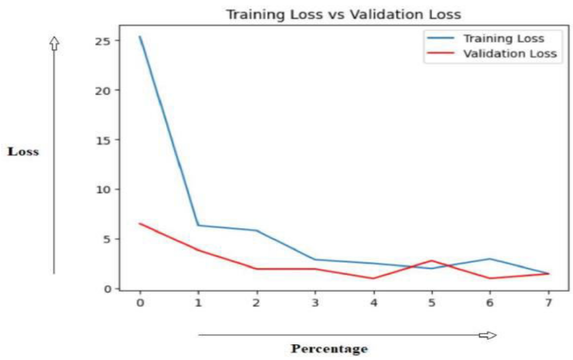


Figure 1: Accuracy graph

User Interaction and Interface:
The Flask application offers a user-friendly interface where users can either upload an image or capture a photo directly through their device. Uploaded images are processed on the server, where the trained deep learning model predicts the plant species and returns the result with a confidence score. This real-time interaction validates the practicality of the system for end users such as herbalists, researchers, and healthcare professionals.

Model Performance:
The CNN model showed robustness in correctly classifying various medicinal plants despite variations in lighting, background, and leaf orientation. Data augmentation during training contributed significantly to improving the model's generalization ability, mitigating overfitting on the training set. The softmax layer outputs a probability distribution across all classes, allowing the system to quantify prediction confidence effectively.

Challenges and Limitations:
Some misclassifications occurred primarily among species with highly similar leaf morphology and texture, which is a common challenge in botanical image classification. Additionally, images with poor resolution or heavy occlusion negatively impacted prediction accuracy. Future improvements could incorporate multi-modal data, such as spectral information or leaf vein patterns, to enhance discriminatory power.

Conclusion and Future Work
This project successfully developed a machine learning-based herbal plant identification system integrated into a Flask web application, enabling accurate and real-time classification of medicinal plants. The model demonstrated strong performance and robustness across varied image conditions, making it a valuable tool for researchers and practitioners in natural medicine. Future work includes expanding the dataset to cover more plant species, improving classification accuracy by incorporating additional features such as leaf vein patterns or spectral data, and enhancing the web app with mobile compatibility and offline functionality to increase accessibility in field environments.

REFERENCES

[1] S.-G. Wu, F.-S. Bao, E.-Y. Xu, Y.-X. Wang, Y.-F. Chang, and Q.-L. Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network," in Proc. IEEE Int. Symp. Signal Process. Inf. Technol., 2007, pp. 11–16.
[2] A. Kadir, L. E. Nugroho, A. Susanto, and P. I. Santosa, "Leaf classification using shape, color, and texture features," ARPN Journal of Engineering and Applied Sciences, vol. 10, no. 10, pp. 4406–4411, 2015.
[3] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," Front. Plant Sci., vol. 7, p. 1419, 2016.
[4] S.-H. Lee, C.-S. Chan, P. Wilkin, and P. Remagnino, "Deep-plant: Plant identification with convolutional neural networks," in 2015 IEEE Int. Conf. Image Process., 2015, pp. 452–456.
[5] A. Prasad, R. P. Kumar, and D. Muthukannan, "Mobile-Based Medicinal Plant Recognition using Deep Learning Techniques," in Proc. 5th Int. Conf. Comput. Commun. Secur., 2020, pp. 123–128.
[6] J. Singh and N. Misra, "Detection of medicinal plant species using image processing and machine learning: A review," Journal of Applied Research on Medicinal and Aromatic Plants, vol. 29, p. 100401, 2023.
Let me know if you'd like these adapted to another citation style (APA, MLA, etc.) or included in a LaTeX format.