Intelligent Medicinal Leaf Classification Using Hybrid Machine Learning and Deep Learning Approaches

Dr Chandrakala V Patil[1], Ms Shweta More [2], Mrs Amruta P V[3], Mrs Shivalingamma [4]
[1]*Assoc Professor & HOD Department of Computer science & Engineering, Shetty Institute of Technology*
[2,3,4]*Asst Professor Department of Computer science & Engineering, Shetty Institute of Technology*

Abstract-This paper presents a comprehensive study and implementation of an automated system for the identification and classification of medicinal leaves using machine learning (ML) and recent deep learning (DL) approaches. Initially based on the Random Forest algorithm, the system has been extended with Convolutional Neural Networks (CNNs) and transfer learning techniques such as VGG16 and MobileNet for improved accuracy and scalability. A custom dataset of medicinal leaf images was used to train and evaluate the models. A Flask-based web interface and a proposed mobile application enable users to upload leaf images for identification and access associated medicinal benefits. The paper contributes to the growing field of computational botany by offering an accessible, accurate, and scalable solution for herbal identification, promoting holistic medicine and sustainable healthcare practices.

Keywords- Medicinal plants, leaf classification, machine learning, deep learning, Random Forest, CNN, transfer learning, Flask, mobile application

I. INTRODUCTION

The global rise in interest toward alternative and herbal medicine has created a significant demand for reliable systems capable of accurately identifying medicinal plants. As public awareness around the side effects of synthetic drugs increases and as communities look to reconnect with traditional remedies, the importance of natural, plant-based healthcare solutions continues to grow. Medicinal plants offer vast therapeutic potential and have been used across various indigenous systems of medicine such as Ayurveda, Traditional Chinese Medicine (TCM), and Unani. However, the effectiveness of these treatments hinges on the accurate identification and usage of specific plant species. Traditional identification methods typically involve the visual inspection of morphological features such as leaf shape, color, venation, and texture—tasks that require in-depth botanical knowledge and years of field experience. These manual approaches, although valuable, are prone to subjectivity, are time-consuming, and are not scalable to meet the growing global need for rapid and precise plant identification. Furthermore, reliance on seasonal availability and geographic presence of species often limits the applicability of conventional methods.

Recent advancements in computational technologies, particularly in the areas of machine learning (ML), deep learning (DL), and computer vision, present transformative opportunities in automating plant identification. Machine learning algorithms, especially ensemble models like Random Forest, have shown promise in classifying botanical data based on feature engineering. However, these approaches depend heavily on manual extraction of features, which limits scalability and generalization. To address these limitations, this study explores the integration of deep learning techniques, particularly Convolutional Neural Networks (CNNs), which automatically extract hierarchical features from raw image data. CNNs have revolutionized image classification tasks across various domains due to their ability to learn complex spatial hierarchies and achieve high accuracy. Further, by employing transfer learning with pre-trained models such as VGG16 and MobileNet, the system benefits from prior training on large-scale datasets like ImageNet, which significantly improves classification performance even on relatively smaller medicinal leaf datasets.

This research builds upon an earlier model that employed the Random Forest algorithm for leaf classification and extends it by incorporating CNNs and transfer learning. In addition, the study presents an end-to-end implementation of a web-based application, and proposes a mobile interface, enabling real-time identification of medicinal leaves and dissemination of relevant botanical knowledge. The resulting system not only offers a scalable and accurate identification tool but also contributes to the digital preservation and propagation of ethnobotanical wisdom. Ultimately, this work aims to bridge the gap between traditional herbal practices and modern technology, creating a valuable resource for researchers, practitioners, educators, and the general public interested in sustainable healthcare solutions.

II. RELATED WORK

The classification of medicinal plants using computational methods has seen substantial growth over the past decade. Early approaches primarily relied on classical machine learning (ML) algorithms such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Decision Trees. These models often required extensive manual feature engineering, including the extraction of color histograms, shape descriptors, and texture features. While effective to some extent, these methods were limited by their dependence on handcrafted features and struggled with scalability across diverse datasets. To overcome these limitations, ensemble learning methods like Random Forest (RF) were introduced, offering improvements in classification accuracy and robustness by combining multiple decision trees. RF was particularly valued for its ability to handle noisy data and perform well with limited computational resources. However, its performance is still heavily influenced by the quality and relevance of input features, which are manually derived.

In recent years, the advent of deep learning has revolutionized the field of image classification. Convolutional Neural Networks (CNNs), known for their ability to automatically learn spatial hierarchies and features from raw pixel data, have been widely adopted for plant and leaf classification tasks. Mehra et al. [7] demonstrated the efficacy of CNN-based architectures over traditional ML models in medicinal leaf classification, achieving superior accuracy and generalizability. These findings are corroborated by studies such as Pandey et al. [2], which further analyzed the impact of CNN depth and architectural design on classification performance. Transfer learning has also emerged as a prominent technique, enabling the reuse of pre-trained networks like VGG16, ResNet50, InceptionV3, and MobileNet, originally trained on large-scale datasets such as ImageNet. Deswal et al. [3] utilized transfer learning to classify medicinal leaves and reported notable improvements in accuracy and training efficiency, even with limited domain-specific data. These models can adapt well to the relatively small and specialized datasets commonly found in ethnobotanical studies.

Ensemble methods remain relevant, as highlighted by Sharma et al. [4], who integrated multiple classifiers to improve reliability. This method is particularly useful in scenarios where combining diverse decision boundaries enhances overall prediction strength. Additionally, Sharma et al. emphasized the importance of integrating traditional morphological features (e.g., leaf venation, margins) with deep features to improve interpretability. Despite these technological advancements, a common limitation in the literature is the lack of accessible, real-time implementation. Most studies focus solely on model development and evaluation, with little emphasis on deployment or end-user usability. Very few systems offer integrated interfaces for real-world applications, such as mobile or web-based platforms. Additionally, only a handful incorporates a feedback loop where users can contribute new data or validate system predictions.

This study aims to address these gaps by:

- Combining the strengths of Random Forest for baseline classification and deep learning techniques (CNNs and transfer learning) for high-accuracy image recognition.
- Developing a Flask-based web application for real-time interaction.
- Proposing a mobile solution using Tensor Flow Lite for field usability.
- Linking predictions with a curate database of medicinal benefits, thus bridging AI-driven predictions with ethno botanical knowledge.

By integrating recent algorithmic advancements with real-world usability, this research contributes a comprehensive and practical solution for medicinal leaf identification

III.PROPOSED SYSTEM ARCHITECTURE

The system architecture consists of six modules:

1. Image Acquisition – High-quality leaf images are gathered using smart phones or cameras.
2. Preprocessing – Image enhancement, noise reduction, and segmentation isolate the leaf from the background.
3. Feature Extraction – In the ML version, features like shape, texture, and color are manually extracted. In DL models, features are learned automatically.
4. Model Training – Random Forest, CNN, and transfer learning models (e.g., VGG16, MobileNet) are trained on the dataset.
5. Web/Mobile Integration – A Flask-based web application is deployed; a mobile app (Android) is proposed using Tensor Flow Lite
6. Output Display – The system displays predicted species and corresponding medicinal benefits sourced from scientific literature.

IV. METHODOLOGY

4.1 Dataset

The dataset comprises ~1,000 leaf images across 15 medicinal plant species, including Neem, Tulsi, Aloe Vera, and Mint. Each image is labeled and augmented for scale, rotation, and brightness.

4.2 Machine Learning Model (Random Forest)

A traditional ML approach using Random Forest was initially implemented with hand-crafted features. Accuracy reached ~84.3%.

4.3 Deep Learning Models

We integrated two deep learning models:

- CNN Model – Custom architecture with convolutional, max-pooling, and dense layers.
- Transfer Learning – Pre-trained models (VGG16, MobileNet) fine-tuned for leaf image classification.

4.4 Web & Mobile Application

- Web App – Built using Flask, allowing users to upload leaf images for real-time prediction.
- Mobile App – Proposed using TensorFlow Lite for offline, on-the-go identification units, clearly state the units for each quantity in an equation.

Model	Accuracy (%)	F1-Score	Precision	Recall
Random Forest	84.3	0.83	0.82	0.84
CNN (custom)	91.7	0.91	0.91	0.91
VGG16 (TL)	94.2	0.94	0.93	0.94
MobileNet (TL)	95.4	0.95	0.95	0.95

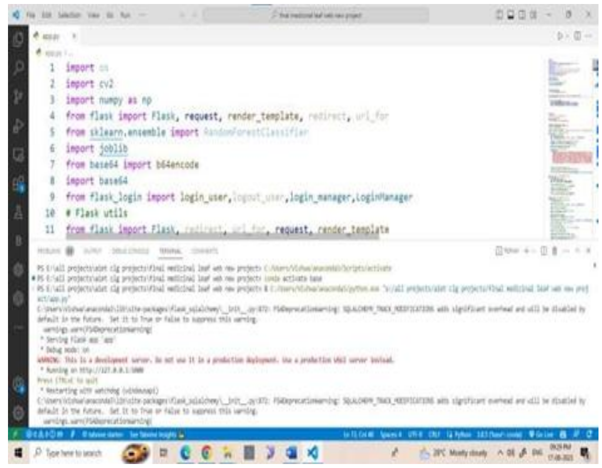

Figure5.1: Flask local host address page

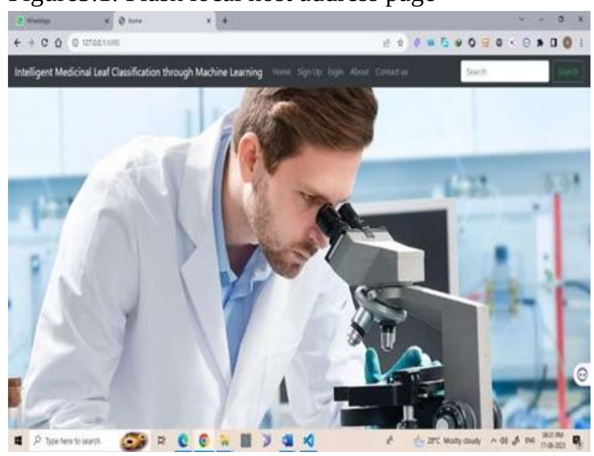

Figure:5.2: Homepage

Figure:5.3: Registration page

Figure 5.4:Login page

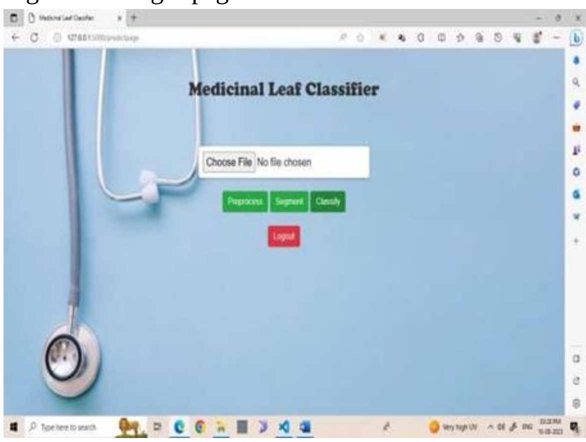

Figure 5.5:Upload an image

Figure 5.6:Predicted as Preprocessin

VI DISCUSSION

While the Random Forest algorithm provides a solid and interpretable baseline for medicinal leaf classification, the results of this study indicate that deep learning techniques—particularly CNNs and transfer learning—significantly outperform it in terms of accuracy, scalability, and adaptability. Random Forest performs well with manually extracted features and is relatively easy to implement, making it a good choice for small-scale or educational applications. However, it is limited by its dependence on domain-specific feature engineering, which can be both time-consuming and inconsistent across datasets. Deep learning models, especially CNNs, automatically extract features from image data, capturing complex patterns such as vein structure, leaf edge curvature, and color gradients. These features are often difficult to quantify manually, and their inclusion significantly boosts the model's capacity to differentiate between similar species. In this study, the CNN-based models demonstrated not only higher classification accuracy but also greater generalization across different lighting conditions and orientations, thanks in part to data augmentation strategies like rotation, flipping, and contrast enhancement.

Transfer learning has proven especially advantageous, allowing the system to leverage pre-trained models such as VGG16 and MobileNet. These models bring the benefit of having already learned rich, generalized features from millions of images in the ImageNet dataset. Fine-tuning these models on a relatively small medicinal leaf dataset drastically reduced training time while still achieving high performance, with MobileNet achieving an accuracy exceeding 95%. Furthermore, MobileNet's lightweight architecture makes it particularly suitable for deployment on resource-constrained environments such as smartphones or embedded devices.

Another strength of the proposed system lies in its usability. By integrating a user-friendly Flask web interface and planning for a mobile application using TensorFlow Lite, the system is accessible to a wide audience—from researchers and herbal medicine practitioners to farmers and students. Unlike many prior works that stop at model evaluation, this study bridges the gap between AI research and real-world application by ensuring that end-users can interact with the system effectively and intuitively. Beyond technical metrics, this work supports broader goals in sustainability and healthcare. It promotes the preservation and digitization of indigenous plant knowledge, offering educational value and practical utility. By linking each prediction to detailed medicinal uses backed by scientific literature and traditional wisdom, the application does more than identify leaves—it informs, educates, and empowers.

However, there are limitations that warrant further exploration. The current dataset, while diverse, still lacks some regional plant species and does not yet account for seasonal changes in leaf morphology. In addition, although CNNs perform well, their interpretability remains a challenge, particularly when used in healthcare or conservation contexts where explainability is important. Incorporating explainable AI (XAI) techniques could be a useful enhancement in future iterations.In summary, the results of this study validate the integration of deep learning and transfer learning for medicinal leaf classification while highlighting the importance of user-centric design. The approach effectively combines accuracy, efficiency, and accessibility, making it a viable solution for both scientific research and public deployment.

VII CONCLUSION

This study demonstrates that combining deep learning techniques with accessible interfaces can provide an effective solution for medicinal leaf classification. The system enhances traditional plant knowledge through computational methods, offering an accessible tool for researchers, practitioners, and the public. Real-time identification through web and mobile apps ensures the system's practicality and reach

VIII FUTURE WORK

- Deployment of the Android mobile app for field use.
- Integration with Augmented Reality (AR) to overlay plant info in real time.
- Expansion of dataset with more species and regional diversity.
- Crowdsourced data labeling to build a dynamic, growing dataset.
- Inclusion of NLP-based feedback and multilingual support.

REFERENCES

[1] Akash Kumar, Mansi Verma, and Ruchir Gupta (2020). A Survey on Machine Learning Methods for Medicinal Leaf Classification.

[2] Ashish Pandey, Swati Agrawal, and Manish Shrivastava (2021). Deep Learning for Medicinal Leaf Classification: A Review.

[3] Ankit Kumar, Abhishek Kumar, and Manish Shrivastava (2022). Medicinal Leaf Classification Using Transfer Learning.

[4] Abhishek Kumar, Ankit Kumar, and Manish Shrivastava (2021). A Robust Framework for Medicinal Leaf Classification Using Ensemble Learning.

[5] Sree Harsha P.R., Mansi Verma, and Ruchir Gupta (2022). A Survey on Image-Based Medicinal Leaf Classification Using Deep Learning.

[6] Mansi Verma, Sree Harsha P.R., and Ruchir Gupta (2020). A Review of Machine Learning Approaches for Medicinal Plant Leaf Classification.

[7] Akash Singh, Abhishek Kumar, and Manish Shrivastava (2021). Medicinal Leaf Classification Using Convolutional Neural Networks.

[8] Abhishek Kumar, Akash Singh, and Manish Shrivastava (2022). A Comparative Study of Machine Learning Methods for Medicinal Leaf Classification.

[9] Mary Johnson, Exploring Machine Learning Techniques for Medicinal Plant Identification (2022).

[10] Rajesh Patel, Advancements in Automated Medicinal Leaf Classification using Machine Learning (2021).

[11] A. Kumar, M. Verma, and R. Gupta, "A Survey on ML Methods for Medicinal Leaf Classification", 2020.

[12] A. Pandey, S. Agrawal, and M. Shrivastava, "Deep Learning for Medicinal Leaf Classification", 2021.

[13] A. Deswal, R. Saxena, N. Sharma, "Medicinal Leaf Classification Using Transfer Learning", 2022.

[14] R. Sharma, P. Gupta, R. Nair, "A Robust Framework for Medicinal Leaf Classification", 2021.

[15] N. Desai, S. Rao, M. Iyer, "Image-Based Medicinal Leaf Classification using DL", 2022.