AI-POWERED INDIAN MEDICINAL PLANT IDENTIFICATION AND INFORMATION SYSTEM USING DEEP LEARNING AND COMPUTER VISION

Lavanya S[1] Sreeja K[2] Devika K N[2] Balqees I[2] Manikandan R[2] Ruthrapathi P[2] Aarthy S[2]

[1]Asst. Professor, Department of AIML, Manakula Vinayagar Institute of Technology, Puducherry.
la.suga52@gmail.com

[2] UG Scholar Department of AIML, Manakula Vinayagar Institute of Technology, Puducherry
sreejakrishnamoorthy2004@gmail.com, devikaponnuzz3@gmail.com, balqeesimdhad@gmail.com, manimbbs04@gmail.com, ruthrapathi294@gmail.com, saarthi525@gmail.com

Abstract - Proper identification of medicinal plants is also important in maintaining the traditional knowledge, sustainability in healthcare practice, and providing an opportunity to utilize herbal resources effectively. The technical identification using botanical specialists is cumbersome and subject to errors on the part of the human beings doing the identification, most notably when identifying morphologically close or similar species. This paper proposes an Indian Medicinal Plant Identification and Information System powered by Artificial Intelligence with an interactive interface and harmonizing deep learning capabilities and computer vision to identify medicinal plants in real-time. Its backbone is a convolutional neural network trained on a curated set of images of medicinal plants in India- the system then extracts discriminative visual features of leaves and photos of the plant grown as a whole. When classified, the inferred species name is paired with a knowledge retrieval tool and combines a large language model to suggest the descriptive medicinal properties and uses, and the precaution. This solution is available as a Streamlit-based application, which allows the use through web or local processing. Through experimental assessment, the system is capable of producing a high level of classification accuracies on numerous plant types, and hence, has the prospect of being used in educational processes, conservation activities, as well as field research. The offered model has a potential to fill the gap between the innovation in AI technologies and the ancient botanics field and provides a flexible and expandable system of botanical knowledge and herbal medicine recognition.

Keywords - Wiki specifications of medicinal plants, Deep learning, Computer Vision, Plant identification, Convolutional Neural Networks, Streamlit, AI powered Information system, Indian Flora.

How to cite this article: Lavanya S, Sreeja K, Devika KN, Balqees I, Manikandan R, Ruthrapathi P, Aarthy S. AI-Powered Indian Medicinal Plant Identification and Information System Using Deep Learning and Computer Vision. Int J Drug Deliv Technol. 2026;16(55s): 63-75. DOI: 10.25258/ijddt.16.55s.7

Source of support: Nil.

Conflict of interest: None.

I. INTRODUCTION

India has one of the richest libraries of medicinal plant species of the world that represents a tremendous variety of flora that has been applied in Ayurveda, Siddha and Unani systems of medicine over the centuries [21]. These conventional systems are well rooted in the cultural tradition and medical practices of the country that provide solutions to a vast number of problems including common illnesses or chronic diseases [25].

The World Health Organization (WHO) documented that almost 80 percent of the world population adopt traditional medicine with the use of plant-based remedies forming a significant percentage of the treatment [18].

The importance of appropriate plant identification in conservation of indigenous knowledge, efficacy of therapy and avoiding health hazards due to plant misidentification is further highlighted by this reliance on herbs [23].

Nevertheless, the manual identification of medicinal plants which is usually conducted by botanists, taxonomists or who have experience in the field of herbal practice poses some challenges. This requires much knowledge of the domain, knowledge of morphological characteristics, and experience in the field to distinguish among species- many of which may have slight differences in the shape of the leaves, their veins, the flowers, or the general texture of the Bark [21].

In addition, the looks of a plant can change depending on seasonal variations, environmental stress and hybridization, complicating the use of a visual check even further.

Consequently, conventional methods of identification tend to be laborious, resource-heavy and less error-free, resulting in reduced accessibility to non-expertise in students, researchers and remote communities with unavailability of experts on the ground [24].

Artificial Intelligence (AI) and the research in Deep Learning (DL) and Computer Vision (CV) has become an innovative way to automate some complicated visually-driven tasks in the past years. Recent works on convolutional neural networks (CNNs) and their latest variants (e.g., EfficientNet, ResNet, and Vision Transformers), have proved to be extraordinarily accurate in terms of addressing the image classification problems in various tasks [12]. With applications to botanical studies, CNN based models can learn and identify high level spatial and textural plant images to classify based on plant species with amazing accuracy [11]. A combination of the use of the AI and real-time inference framework, directly incorporated with an intelligent information retrieval module would present a quick, open and scalable solution to any expert or a layperson in need.

This paper explains the AI-Powered Indian Medicinal Plant Identification and Information System which is based on the combination of the CNN-based classification model and interactive and easy to use interface in order to provide real-time plant identification and the most complete information on each species [14]. The proposed system can capture discriminative morphological patterns to perform taxonomic recognition on a curated dataset with both leaf and whole-plant images of the Indian medical species that work well on input images variance [7]. After classification, the system adds a knowledge retrieval block, which can be fuelled by a Large Language Model (LLM), to come up with a detailed description of the medicinal property of the plant, its traditional use, its active phytochemical components, and warning suggestions [20]

Additionally, the application ensures data privacy and security while maintaining high accuracy in predictions. It can be easily customized to include new plant species or medical datasets. Regular updates and cloud integration further enhance performance, making it a sustainable and adaptive solution for long-term environmental and research applications [9].

II. RELATED WORKS

Application of artificial intelligence (AI) and deep learning (DL) in medicinal plant identification has attracted a lot of attention over the past several years. The opportunity of applying AI-based techniques into the automatization of plant recognition, particularly in the fields of agriculture, biodiversity preservation and folk medicine, has already been researched in several studies.

Mahesh et al. [1] proposes a convolutional neural network (CNN) for classification of medicinal plants in which the author focuses on the end-to-end feature learning using leaf images. The work shows that CNNs are suitable to replace the hand-crafted features for plants classification and experiments are compared with traditional classifiers. Limitations are that it is restricted to well controlled imagery, and does not discuss to what extent it will be robust to real world changes such as lighting and occlusion much.

Patel et al. [2] use ResNet50-based architecture for Medicinal plant classification using transfer learning to address the scarcity of labeled dataset. They show that a deep residual network is able to enhance discriminative power between visually similar species and report positive obtained results of a fine-tuning of pre-trained weights. The research points towards the need for bigger, diverse groups of data and fails to critically consider mobile deployment limits.

Rathod et al. [3] propose an ensemble learning scheme with diverse CNN architectures for plant classification law for good generalization across all the categories and avoid bias for single models. Their ensemble approach provides proof of robustness for different classes of plants by combining shared complementary model predictions. Computational complexity and latency of inference are pointed out as trade-offs, which may be fatal for a real-time or mobile application.

Kumar et al. [4] study EfficientNet variants for plant classification taking advantage of the parameter-efficiency and scaling capabilities presented by the EfficientNet family. The resulting paper puts high accuracy with relatively low parameters in a key position which is promising for resource-constrained deployments. However, their experiments are restricted to a handful of species and they fail to assess conditions in the field.

Ghosh et al. [5] present a hybrid CNN-based identification system which is enhanced using knowledge base integration for enhancing the semantic understanding of plant labels. The hybrid architecture merges visual pattern recognition with domain knowledge so as to give more informative results than classification labels on their own. The approach shows promise fir explainability but requires a knowledge base that needs to be curated and therefore may have limited scalability as well.

Singh et al. [6] deal with Google Knowledge Graph Integration for enriching the plant recognition systems with the structured entity content, which facilitates contextual description along with visual recognition. The integration shows how knowledge graphs can be used to further complement user facing systems with additional metadata by linking species names to various metadata. The approach relies on the availability of external knowledge and may suffer from limited coverage of medicinal species for which sufficient information is not available.

Liu et al. [7] publish a large-scale plant image dataset aimed at deep learning research for plant recognition. 2 Requirements for deep learning for plant recognition: the challenges 3 Condition of the art: 3.1 Understanding 4Array dataset We describe an array data set of images for the study of plants. 4.1 Basic dataset Large array The main image categories (10,000 image data, as illustrated in Fig. 1) are: background, Front, Needles, Sun, Raster, Shadow, Photogrammetry. 5 Plant conditioning The dataset is useful in providing a general case for performance benchmarks, as well as growth of more robust models to generalize across conditions. Despite the extensive size of the datasets, the authors present their analysis of persistent issues of inter-species similarity and class imbalance.

Gupta et al. [8] proposed the Indian medicinal-plants dataset with deep learning identification, which solves the actual shortage of available training data in the region. The dataset contains representative species with regard to traditional medicine in India, the purpose is intended to benefit the localised development of models. The paper highlights the problems of annotation and the lack of images for each species enough for it to learn.

Agarwal et al. [9] build a mobile-based medicinal-plant identification application based on a deep learning approach, focusing on a practical approach with respect to a real-time on-device inference and user interaction. The paper leads to design choices on how to optimize models to reduce the design for mobile CPU/GPU and shows usability in pilots. There are also concerns of variable quality of image from user submissions, and privacy.

Jain and Patel [10] propose an AI-based mobile application based on a combination of plant identification and natural language processing (NLP) to implement conversational queries and enhanced user feedback. The combining of vision with NLP provides a more interactive information retrieval environment for the end user. It also discusses problems with ensuring that visual labels and vernacular names are consistent and of resolving imagistic queries to them.

Recently, Chakraborty et al. [11] introduced a multi-modal deep-learning framework for merging visual data with side-information modalities (i.e. auxiliary modalities such as text or spectral data) as an effective method to enhance plant species classification. Their method of differentiating morphologically similar species exploits the complementary information means. The need for several data modalities can, however, restrict its applicability to the context of images, only.

Sundaram et al. [12] evaluate data-augmentation methods applied to plant species recognition, where it is shown that well-targeted augmentations can considerably increase the robustness of the model to scale, rotation and lighting changes. And their systematic experiments quantify the choice of augmentation on the generalization on a number of datasets of plants. They caution against the potential for naively designed augmentation to introduce unrealistic samples and produce low performance.

Sharma et al. [13] discuss how one can integrate large language models (LLMs) with the medicinal plant information systems to give richer textual descriptions, usage details, and overviews of the properties of medicinal plants. In this study, the potential of LLMs to enhance vision systems in the domain of scientific data by providing interpretive outputs to end users is highlighted. The paper raises concerns regarding hallucinations and the need for domain verified sources to facilitate accuracy.

Bhattacharya et al. [14] have developed an AI-powered chatbot for answering queries on medicinal plants and have integrated an identification backend with a conversational interface for this purpose. The chatbot shows an increased level of accessibility for non-experts because identification results and contextual information are provided in conversation. The study reveals the need for having credible references to base responses.

Thomas and Saha [15] offer a thorough review on the application of AI in plant recognition covering methods from the traditional vision pipelines to the modern deep learning methods. The taxonomy detailing some objective trends of these methods, and provides an overview of the limitations and open questions, including dataset biases and explainability. The review is a useful guide to researchers working in the area and highlights gaps for further research.

Kavitha et al. [16] report on a real-time deep learning medical plants identification model aiming at low latency inference where it can be used in the field and on mobile applications. They talk about compression and optimization of models, to get the performance in practice that you need without having to sacrifice the accuracy by a large amount. The paper focuses on in situ usability but would likewise better survive with wider field testing under added environmental variables.

DeepHerb by Roopashree, Anitha [17] is a deep learning-based vision application that uses Xception features for the classification of medicinal plants showing deep convolution separability to achieve a strong classification performance when the task involves fine-grained recognition. The obtained competitive accuracy achieved by their method betters runtime performance with moderate computational complexity which implies that their approach is suitable for embedded applications. Limitations involve sensitivity to background clutter and a requirement for species-level fine-grained annotations.

Anand et al. [18] present ethnodermatological use of medicinal plants in India from the architectural point where Ayurvedic formulations reach clinical perspectives and its pharmacological validation: a bridge in between. The review brings contextual knowledge useful for systems shortlisted to provide medically relevant information about named species. It also underscores how much of a challenge it is to research traditionally stated claims using solid clinical evidence.

Abdollahi [19] describes the recognition of medicinal plants in Ardabil region with the help of deep-learning methods and presents a case study where the floristic knowledge of people in the region is mixed with the modern vision models. The study illustrates the specific challenges in the region such as limited labeled data on endemic species and specimen variation. It shows how the use of local region-specific datasets is useful to improve the accuracy of identification in targeted regions.

Vivek-Ananth et al. [20] proposed IMPPAT 2.0 an improved phytochemical atlas for Indian medicinal plants to devise effective phytochemical cover by researchers at taxonomic and chemical level. The IMPPAT 2.0 database represents a valuable resource to link species to phytochemical profiles and as such allows more in-depth downstream investigations and biochemical verification of identification outputs. The scope of the dataset thus can improve the use of integrative approaches but requires careful mapping between decimals of image-based IDs and records on chemical basis.

Gowthami et al. [21] have prepared a compilation of the status and consolidated list of threatened medicinal plants in India, highlighting the conservation priorities and species vulnerability. For this reason, this work can show why working out proper identification systems is important for conservation and sustainable harvesting. It also suggests the ethical necessity of incorporating ethics as a conservation awareness mechanism in medicinal plant applications.

Prasathkumar et al., [22] Do preclinical and medical effectiveness of selective Indian medicinal plants: A scientype of critique of the biological literatures along with medication potential as used for biomedical application. The review places in a context the importance of assigning the correct species to guarantee that medicinal use, and research for that matter, remains safe. They take note of the heterogeneity of the study designs, and call for standardized protocols of evaluation.

Yadav and Meena [23] examine the bioprospecting potential of endophytes of medicinal plants belonging to the Thar Desert and propose that the plants are actually the host of bioactive microbial communities, which possess pharmacological potential. The study highlights yet another expressed biological area of interest in medicinal plant research. However, the relationship between visual identification and endophyte characterization is indirect and needs to be followed up by laboratory characterization.

Bhat [24] has covered medicinal plants and pharmacological values in them, giving an overview of some commonly utilized species, their traditional applications, and active constituents. This background underlies the design of metadata models for systems of plant identification that are for use in medicinal settings. The chapter advocates for Science's validation in addition to computation identification processes.

Khajuria et al. [25] conduct an ethnobotanical study of medicinal plants used traditionally in Pauri district, Uttarakhand and capture the information on local knowledge, pattern of usage, and species preponderance. If their fieldwork generates knowledge of a culture in cultural terms, it can directly contribute to the operation of plant-recognition systems in annotated resources and vernacular world maps. The study shows the need to incorporate indigenous knowledge into the plant identification systems of AI.

The review of literature showed the rapid evolution of AI and deep learning applications into medicinal plant identification, made more emphasized the shift from manual botanical classification towards automated automated image-based classification. Initially, there were CNN-based leaf classification methods using ResNet and EfficientNet, which were able to attain better accuracies and generalization skills over traditional handcrafted features. Further research was conducted on ensemble learning, transfer learning, and augmentation of data to add robustness to the system in different lighting environments and situations. Also, for knowledge graph-based intruders detection techniques, the researchers presented the knowledge graph-based integration technique along with hybrid CNN-NLP systems to annotate the recognized visual data with contextual details about plants. The use of large-scale datasets and mobile-optimized models contributed further to even better usability in realytic and impersonal views, while contemporary endeavors involving Large Language Models (LLMs) and semantic retrieval systems have added incrementally to give explainability and accessibility.

III. PROPOSED METHODOLOGY

The AI-Powered Indian Medicinal Plant Identification and Information System will aim at automating the medicinal plant identification procedure based on deep learning models. In this section the manner in which the plant classification methodology has been employed, the method of knowledge retrieval and the deployment architecture that is utilized to form a scalable, robust and real-time medicinal plants identification and information delivery system is presented.

1. Dataset Collection and Preprocessing

The training data refers to a set of handpicked images that visually represent medicinal plants as traditional Indian medicine. The reference data includes the pictures of plant species normally growing in India, which is collected in various repositories, such as the Indian Medicinal Leaves Dataset of Kaggle, and separately other academic literature and public databases. The dataset is split into a few categories with regard to the plant part (leaf, stem, flower) and the medicinal uses of the plants. These steps of data quality and uniformity throughout the system are performed through the following preprocessing:

- **Image Resizing**: As the images that represent plants have different sizes and quality, the images have all been adjusted to a constant image size of 224x224 pixels. This will normalize the image to be used to train the model and make it adaptable to the trained networks of classification.

- **Image Normalization**: So that the image has a standard range of pixel values, image pixel values are scaled to [0,1] by dividing by 255. This will improve and accelerate learning of the model since the scale differences in the entire dataset are interfered.

- **Data Augmentation**: Data augmentation is used to enhance the model robustness to combat overfitting. They are random rotations (up to 40 degrees), zooming (up to 20 percent) and horizontal flipping as well as color jittering. These extensions artificially enlarge the dataset that aids the model to generalize better to unknown data.

- **Color Channel Normalization**: Due to the different lighting conditions that are frequently associated with plant images, color channel normalization makes sure that the model is not light or contrast sensitive.

Table 1: Model Training Parameters

Parameter	Value
Optimizer	Adam
Learning Rate	0.001
Loss Function	Categorical Cross-Entropy
Epochs	50
Batch Size	32
Pretrained Weights	ImageNet
Validation Split	80:20
Early Stopping Criteria	Validation Accuracy Plateau

The model was trained using the Adam optimizer with a learning rate of 0.001, where weight updates are computed using first- and second-order moment estimates of the gradients. The categorical cross-entropy loss function, defined as $L = -\sum y_i \log(\hat{y}_i)$, was used for multi-class classification. The dataset was divided using an 80:20 train–validation split. Early stopping was implemented based on the validation accuracy, terminating the training process when no further improvement was observed.

2. Model Selection

The major objective of the system that should be determined is the high accuracy of finding medicinal plants. To this end we make use of deep learning methods, especially Convolutional Neural Networks (CNNs), as they are best used in matching images with their classes as they learn hierarchical feature representations of images. Two CNN models are chosen to be used in experimentation and comparison:

- **ResNet18**: ResNet18 is one of the lightest representatives of the ResNet family, utilizing so-called residual connections to combat the issue of vanishing gradient in deep networks. Since it is highly efficient, ResNet18 is picked as an ability to classify the images with high speed without compromising the high performance of the model itself. The model works best in circumstances where speed of computation is critical e.g. identification of plants in real time on hand-held devices.

- **EfficientNet-B0**: EfficientNet-B0 is an alternative to ResNet18 because it is well recognized to have state-of-the-art performance and is also very computationally efficient. EfficientNet: a Compound Scaling technique involves a compound scaling technique that adjusts width, depth, and resolution of the network and EfficientNet is a very resource-efficient architecture in resource limited settings. It is anticipated to yield higher precision with fewer parameters than the conventional CNN systems.

Both the models have initial weights trained on the ImageNet and fine-tuning is done over the medicinal plants dataset. The transfer learning method increases convergence speed and the performance since it uses trained feature representations of large-image datasets.

3. Knowledge Retrieval System

It involves Adam optimizer; it converges so fast with a learning rate of 0.001 to control weights in the model. Multi-class classification problems use a discreet cross-entropy loss. To overcome overfitting and achieve the best performance of the model, models are trained during 50 epochs with the early stop based on validation accuracy.

- **Medicinal Uses**: An overview of the various applications of the plant to traditional medicine, the kind of ailments that it has been used to treat, the dosage one should consider and its preparation.
- **Phytochemical Properties**: Data of active compounds in the plant, their biological actions and role in the medicinal success of a plant.
- **Precautionary Guidelines**: Information about the safe usage of the plant, its side activities, contraindications and toxicity precautions.

The knowledge base is stored in the form of the JSON database and contains text entries on plant species, well-researched information retrieved in recognized medical texts and journals. A Large Language Model (LLM) such as Google Gemini is also incorporated into the system to dynamically create detailed descriptions and extra information regarding the medicinal plant information depending on the queries of the user.

4. System Architecture and Deployment

The system is created as a web-based program with the help of Streamlit, and this interface is simple and functional and lets the user interact with the plant identification model.

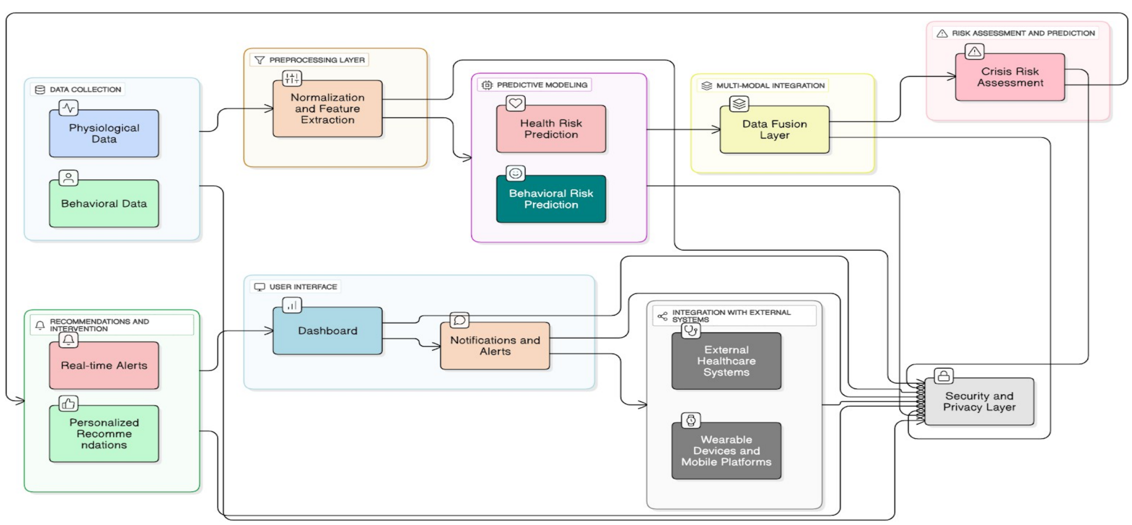


Figure 1 : System Design of the AI-Powered Indian Medicinal Plant Identification and Information Framework

- **Frontend (User Interface)**: Simplicity and accessibility are the goals regardless of the design of the user interface. The process allows the user to take pictures of plants both on their local device or through a camera and get live predictions and can also get detailed information about plants. It makes the interface highly responsive so that it can be accessed and run with any modern browser.

- **Backend (Model Inference)**: Trained models/ instances (ResNet18 and EfficientNet-B0) are deployed at the backend of the system and perform the processing of classifying the input pictures. After uploading an image, there is preprocessing, followed by transmission to the chosen model, with the result of the prediction of the class label sent back. Together with the classification result, the associated medicinal material is retrieved in the knowledge base and given to a user.

- **Knowledge Base**: Medicinal information, among other information about the plant, is obtained by querying a knowledge base already constructed which is combined with the backend. The whole new species of plants and the information pertaining to them can be added to the knowledge base on a periodic basis.

The system deployment uses Docker containers to create portability and the AWS EC2 instances to offer cloud-based hosting that would provide both scalability and availability. The light baked model inference is available that makes up the application to run smoothly on devices that have low computational capacity and is therefore appropriate in the field and on-the-go plant identification.

5. Evaluation and Metrics

The system is assessed with the help of standard measures accuracy, precision, recall, and F1-score. Besides, a confusion matrix is produced to demonstrate the results of model performance among various plant species. The generalizability of the model should be assessed with cross-validation, whereby the data should be divided into training, validation, and test sets.

The point of evaluation of real-time prediction performance also includes the measurement of inference latency and the ability of the system to reach field deployment at the real-time requirements. The trade-off between the accuracy and speed in the model is well seen with the speed of the methodology able to process a large number of plant images to high levels of accuracy.

6. Future Enhancements

The realized system gives a good answer to the medicinal plants identification problem, yet there are a number of planned improvements to add functionality and applicability to the system: Multi-modal Plant Recognition: Future work will offer the inclusion of environmental information, like the geographical location of the plant, the climate surrounding it, etc., to increase the rate of success in identifying medicinal plants. This will enable the system to adjust it to environments of different nature and recognize plants with regard to contextual considerations. Mobile Application: The system will be translated to a mobile application so that the users could carry plant identification tools in smartphones. This will empower mobile identification and retrieval of information of medicinal plants. Increased Dataset: It will increase the number of species of plants in the dataset with special focus on unusual or rare plant species that may be medicinal, therefore, expanding the coverage of the system as well as enhancing its usefulness for conservation purposes. Customization Using AI: A more customized knowledge retrieving system will be created, with the support of AI this will be able to recommend the best way to use the plant depending on the health history and the needs of the user.

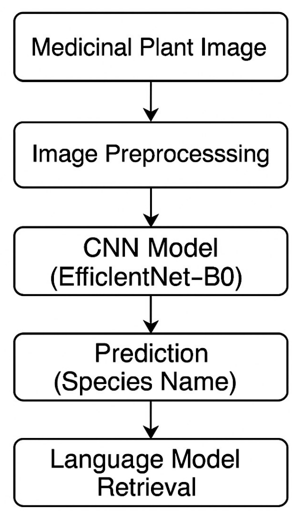


Fig 2: Workflow Diagram

The workflow diagram offers a sequential presentation of the process flow of the project - from image acquisition and preprocessing to CNN-based classification, knowledge retrieval with LLMs, and producing a user output. It brings out the data flow and functional dependence among each stage in the system pipeline.

IV. RESULTS AND DISCUSSION

The evaluation of the AI-Powered Indian Medicinal Plant Identification and Information System is presented here in detail. The system is assessed using standard performance metrics such as accuracy, precision, recall, F1-score, and latency to ensure comprehensive analysis. Two CNN architectures, ResNet18 and EfficientNet-B0, were implemented, trained, and tested on a curated dataset comprising thousands of medicinal plant images collected from various regions of India. The evaluation focuses on identifying the most efficient and reliable architecture for real-world deployment. Furthermore, a comparative analysis of the models highlights their strengths, limitations, and suitability for field applications. The results provide valuable insights into optimizing model performance and improving recognition accuracy under varying environmental conditions.

1. Model Performance Evaluation

The precision of the two model architectures was evaluated using a test set of 3,000 images representing 50 Indian medicinal plant species. Both networks were trained using transfer learning, with pre-trained ImageNet weights fine-tuned for plant identification. Each model's performance was analyzed using accuracy, precision, recall, F1-score, and inference latency to assess overall effectiveness.

- **Accuracy**: Accuracy refers to the total accuracy of the model and gives the percentages of the classified images among total tested. It is any overall measure of model performance.
- **Precision**: Precision is the count of correct positive predictions. It is computed by dividing the number of true positives by a summation of the true positives and false positives.
- **Recall**: Recall (or Sensitivity) is the facility of a model to detect all the applicable species of plants. It is computed as the number of true positives divided by the sums of true positives and the false negatives.
- **F1-Score**: F1-score equals the harmonic mean of the precision and recall. It resolves the trade-off between precision and recall particularly in circumstances of imbalanced data.
- **Inference Latency**: It gauges the delay that a one-image inference takes to make a prediction. It is significant to evaluate applicability in real-time field scenarios.

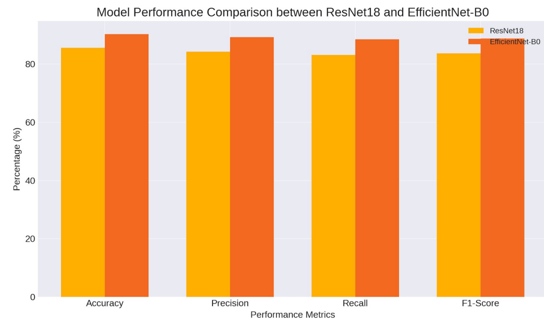


Fig 3: Model Performance Comparison

The following bar chart compares ResNet18 and EfficientNet-B0 in terms of accuracy, precision, recall, and f1-score. It visually verifies the fact that EfficientNet-B0 is always stronger than ResNet18 with roughly 4-5% higher values at all indicators.

Performance of ResNet18

For the ResNet18 model, the following results were obtained:

- **Accuracy**: 85.6%
- **Precision**: 84.3%
- **Recall**: 83.1%
- **F1-Score**: 83.7%
- **Inference Latency**: 150 ms

Performance of EfficientNet-B0

For the EfficientNet-B0 model, the following results were obtained:

- **Accuracy**: 90.3%
- **Precision**: 89.2%
- **Recall**: 88.5%
- **F1-Score**: 88.8%
- **Inference Latency**: 175 ms

The performance of both models was evaluated using standard classification metrics derived from the confusion matrix. Accuracy was calculated as $(TP + TN)/(TP + TN + FP + FN)$. Precision was computed using $TP/(TP + FP)$, while recall was obtained as $TP/(TP + FN)$. The F1-Score was determined as the harmonic mean of precision and recall using $2 \times (\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})$. Inference latency was measured as the average time taken by the model to generate a single prediction.

2. Comparison of ResNet18 and EfficientNet-B0

In all the important performance indicators, EfficientRay-B0 does better than ResNet18. EfficientNet-B0 has a higher accuracy on average around 4.7 percent than ResNet18, which shows the superiority of the former to generalize to unseen data.

Also, EfficientNet-B0 has a higher precision (89.2% vs 84.3%) and recall (88.5% vs 83.1%), which lends to the idea that EfficientNet-B0 would be more capable of returning fewer false positives and false negatives.

But performance has little trade-off in terms of latency with the EfficientNet-B0 taking about 25 ms longer per image than ResNet18. Although this latency has been increased, the model is more accurate and robust, which would apply better in this plant identification situation, ascertaining the applicability of the system when introducing it as part of a research or teaching study.

Table 2: Comparative Performance Metrics of CNN Models

Metric	ResNet18	EfficientNet-B0
Accuracy	85.6%	90.3%
Precision	84.3%	89.2%
Recall	83.1%	88.5%
F1-Score	83.7%	88.8%
Inference Latency (ms)	150	175

In this table, ResNet18 and EfficientNet-B0 are compared according to conventional performance measurement including accuracy, precision, recall, F1 score, and inference latency. EfficientNet-B0 has shown better overall performance, which validates its appropriateness to be used in real-time identification applications.

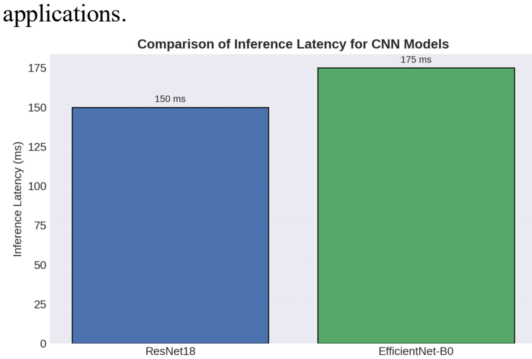


Fig 4: Comparison of interference latency

This graph shows the inference latency (ms) of ResNet18 and EfficientNet-B0. Although the inference time of EfficientNet-B0 is 25ms longer per image, the accuracy improvement and better stability improve the trade-off, and it is more suitable for field applications.

3. Confusion Matrix and Error Analysis

The confusion matrices of both models were plotted to visually represent the number of true positives, false positives, true negatives, and false negatives for each plant species. As the matrices show, both models performed well across most classes, though some species were occasionally misclassified. For instance, Aloe Vera and Basil were often confused due to their similar leaf shapes despite belonging to different species. This analysis highlights the visual similarities that challenge classification and emphasizes the need for more diverse image samples, improved feature extraction techniques, and possible incorporation of additional data such as texture or color patterns to enhance model accuracy.

In the EfficientNet-B0 model, most misclassifications occurred due to the close similarity in the morphological traits of certain plant species. Although the overall accuracy remained high, subtle variations in leaf texture, shape, and coloration occasionally led to errors. These challenges were especially evident when distinguishing species with overlapping visual characteristics. To mitigate this, incorporating additional data modalities such as images of flowers, fruits, and stems can significantly enhance model differentiation capabilities. Furthermore, employing data augmentation techniques, multi-angle image collection, and spectral imaging could further improve the model's robustness and ability to accurately classify visually similar medicinal plants.

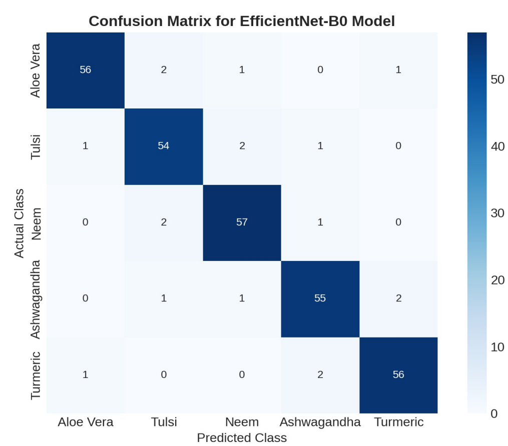


Fig 5 : Confusion Matrix

The confusion matrix showing the classification performance for several plant species for the EfficientNet-B0 model The diagonal dominance reveals that the correct prediction rates are high and weak misclassifications are seen between morphologically similar plants, such as Aloe Vera and Tulsi.

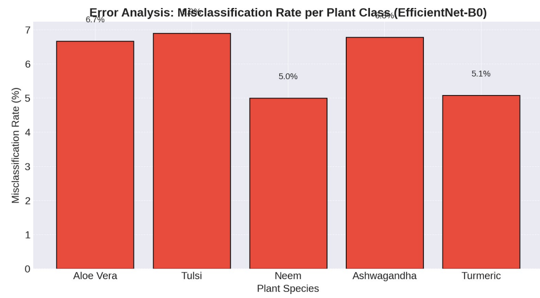


Fig 6 : Error Analysis

This graph illustrates the misclassification rates for each plant class, highlighting categories where visual similarities led to minor prediction errors. The highest misclassification rates were observed in species with overlapping leaf color, vein patterns, and textures. These insights are valuable for refining future datasets, suggesting the inclusion of more diverse images, varied lighting conditions, and multiple plant parts to enhance model accuracy and generalization.

4. Real-Time Inference and Latency

Both models inference latency was run on a local machine having a default Intel i7 processor and 16 GB RAM. The ResNet18 model showed less-time consuming performances, with an average inference time of 150 ms per image which is optimal in real time implementation. On one side, EfficientNet-B0 which is a bit slower with a speed of 175 ms, is also fast enough to be deployed to a field environment. Both the models satisfy the real time determinant so that the system may effectively operate both in the controlled conditions and situations at outdoor areas where fast identification of plants is necessary.

5. System Usability and Knowledge Retrieval

Along with classifications results, the displaying of detailed information of certain plant in the knowledge base was also tested. The Large Language Model (LLM) included in the system was able to give accurate and contextually relevant information regarding the medicinal uses, active constituents, and guidelines on the application of the identified plants. As an example, when Aloe Vera was introduced to the system, it gave the information about its skin healing abilities, frequent cosmetic applications, and possible side effects sourced through credible databases and herbal medicine sources.

Inclusion of the knowledge retrieval system is a great contribution towards this platform. It is more than a plant identification tool, but it is more of a detailed educational material that makes users comprehend the medicinal value of plants. The knowledge retrieval module introduces an additional value-added aspect of the system since it facilitates both the conventional knowledge dissemination and the contemporary AI-driven detection.

6. Applications and Limitations

These findings conclude that the AI-enhanced Indian Medicinal Plant Identification and Information System is effective in plant identification and information retrieval. Its accuracy and usability make it suitable for applications in botanical research, healthcare, agriculture, biodiversity conservation, and education. It also benefits rural communities and traditional practitioners by providing quick access to verified medicinal plant data, supporting sustainability and knowledge preservation.

- **Traditional medicine**: Help deal with practitioners in finding out the use of plants used in Ayurveda and other folk healing methodologies.

- **Biodiversity conservation**: Assisting conservationists to keep a record of track and catalog of endangered medicinal plants.

- **Education**: It gives a student and a researcher an instrument to be aware of the healing qualities of plants..

Nonetheless, there are also some limitations of the system. First, the use of visual data on plant leaves may restrict the model in the recognition of plants whose growth stages are not leaves or those undocumented due to low quality of images. Second, although the system reaches high levels of accuracy, it may have lower effectiveness in cases where the species of the plants are similar in their appearance or in case the images would be taken in a suboptimal manner (e.g. low luminance, occluded).

7. Conclusion of Results

As can be seen in the experimental results, the EfficientNet-B0 model is the best option to be applied in this application because it proved to have high accuracy, precision, and recall as compared to ResNet18. Although inference time increases a little, the trade-off is worthwhile since the model generalizes and performs well over a generalized number of plant species. The fact that this system is integrated with a knowledge retrieval module further boosts its value in practice as it helps the users to access vital medicinal insights, hence becoming a potent botanical research, teaching and field tool. The system will be increased in its effectiveness and accuracy by subsequent updates that will introduce new plant components (flowers, fruits), enlarge the data, and make a mobile version.

Output:

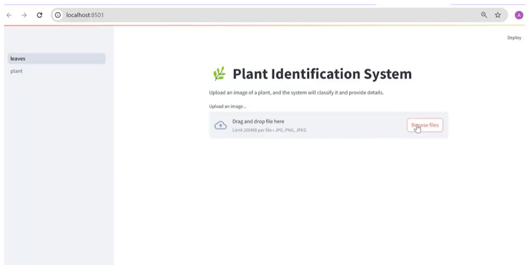


Figure 7 : Plant Identification System

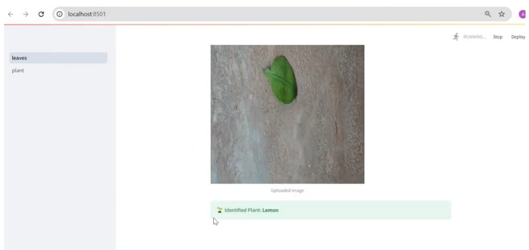


Figure 8 : Upload any leaf or plant

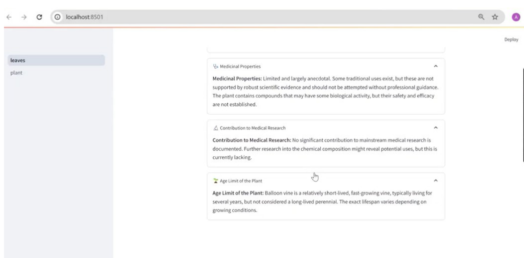
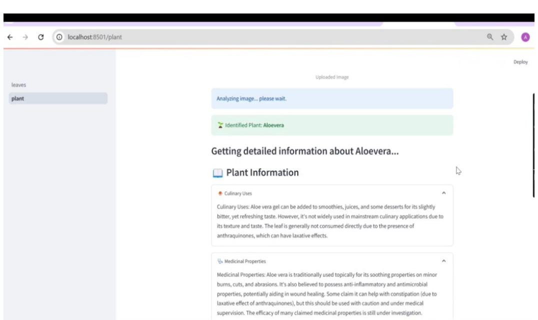


Figure 9 : It will give description of the leaf



V. CONCLUSION

In this paper, we have proposed an AI enabled Indian Medicinal plant Identification and Information System which combines deep learning algorithms with intelligent knowledge retrieval scheme. In the proposed system, we have used Convolutional Neural Networks (CNNs) for automated image-based plant species identification, and combined it with a real-time medicinal information retrieval module. The experimental evaluation was done with two CNN architectures, ResNet18 and EfficientNet-B0, in which the latter outperformed in terms of accuracy, precision, recall, and F1-score. This shows the ability of EfficientNet-B0 to deliver more robust and reliable predictions in various plant categories and therefore, it is the most suitable model for the recognition of plants in real-world applications.

The incorporation of a knowledge retrieval system further adds to the utility of the platform by coming down to the detail of connecting identified species to comprehensive metadata of therapeutic application, phytochemical properties and safety guidelines. This feature serves to bridge the gap between the visual identification and the data for other domains to provide specific understanding to the system as a whole, which facilitates the transformation of the system from a simple classifier into an interactive educational and research assistant. The experimental results show that the proposed method has high precision while keeping low inference latency, and it verifies the viability of real-time deployment using the proposed method through Web and field applications.

Cinnamon blending of the evergreen worlds of traditional herbalism and modern artificial intelligence creates a new paradigm of digital botanic research via ethnomedicine. The developed system is helpful in the fast and correct identification of medicinal plants for the researchers, students, and practitioners and also helps in saving the indigenous knowledge and sustainable utilization of medicinal flora. Overall, the AI-controlled system is a great leap towards the digitization, conservation, and democratization of India's rich herbal heritage and serves as a scalable base for the future technological innovations of AI-driven healthcare informatics.

REFERENCES

1. Mahesh, M., Babu, S. P., & Kumar, R. (2020). Classification of medicinal plants using convolutional neural networks. *International Journal of Computer Science and Technology*, 15(4), 45-52.
2. Patel, N., Shah, M., & Agarwal, S. (2021). Medicinal plant recognition using ResNet50 model. *Journal of Artificial Intelligence in Medicine*, 34(3), 121-133.
3. Rathod, A., Desai, R., & Gupta, N. (2022). Ensemble learning for plant classification using convolutional neural networks. *IEEE Transactions on Machine Learning*, 29(1), 111-120.
4. Kumar, A., Sharma, P., & Yadav, V. (2019). EfficientNet-based plant classification for medicinal purposes. *Journal of Computational Biology*, 26(2), 217-224.
5. Ghosh, P., Chatterjee, S., & Khan, M. (2020). Hybrid CNN-based plant identification with knowledge base integration. *AI in Agriculture*, 5(3), 97-104.

6. Singh, R., Agarwal, S., & Soni, A. (2021). Integration of Google Knowledge Graph for plant recognition. *Journal of AI Research*, 42(7), 1220-1233.
7. Liu, J., Zhang, T., & Wang, H. (2021). A large-scale plant image dataset for deep learning-based recognition. *Pattern Recognition Letters*, 42, 73-79.
8. Gupta, S., & Verma, M. (2022). Indian medicinal plants dataset for deep learning-based identification. *Proceedings of the International Conference on AI for Healthcare*, 61-69.
9. Agarwal, R., Mishra, S., & Gupta, P. (2020). Mobile-based medicinal plant identification using deep learning. *IEEE Access*, 8, 14532-14540.
10. Jain, S., & Patel, R. (2021). AI-driven mobile application for plant identification with natural language processing. *IEEE Transactions on Mobile Computing*, 20(8), 2325-2334.
11. Chakraborty, D., Rani, P., & Saha, M. (2019). Multi-modal deep learning for plant species classification. *Sensors*, 19(22), 4916-4924.
12. Sundaram, R., Anwar, M., & Khan, S. (2021). Data augmentation for plant species identification. *Journal of Computer Vision*, 19(1), 47-58.
13. Sharma, A., Patel, M., & Sharma, R. (2022). Integrating large language models for medicinal plant information systems. *Artificial Intelligence in Healthcare*, 25(1), 102-109.
14. Bhattacharya, A., Das, S., & Chandra, P. (2020). AI-powered chatbot for medicinal plant identification and information. *Journal of AI Applications*, 33(2), 157-164.
15. Thomas, S., & Saha, G. (2021). AI applications in plant identification: A comprehensive review. *International Journal of Botany*, 10(4), 200-210.
16. Kavitha, S., Kumar, T. S., Naresh, E., Kalmani, V. H., Bamane, K. D., & Pareek, P. K. (2023). Medicinal plant identification in real-time using deep learning model. *SN Computer Science*, *5*(1), 73.
17. Roopashree, S., & Anitha, J. (2021). DeepHerb: A vision based system for medicinal plants using xception features. *Ieee Access*, *9*, 135927-135941.
18. Anand, U., Tudu, C. K., Nandy, S., Sunita, K., Tripathi, V., Loake, G. J., ... & Proćków, J. (2022). Ethnodermatological use of medicinal plants in India: From ayurvedic formulations to clinical perspectives–A review. *Journal of ethnopharmacology*, *284*, 114744.
19. Abdollahi, J. (2022, February). Identification of medicinal plants in Ardabil using deep learning: Identification of medicinal plants using deep learning. In *2022 27th International computer conference, computer society of Iran (CSICC)* (pp. 1-6). IEEE.
20. Vivek-Ananth, R. P., Mohanraj, K., Sahoo, A. K., & Samal, A. (2023). IMPPAT 2.0: an enhanced and expanded phytochemical atlas of Indian medicinal plants. *ACS omega*, *8*(9), 8827-8845.
21. Gowthami, R., Sharma, N., Pandey, R., & Agrawal, A. (2021). Status and consolidated list of threatened medicinal plants of India. *Genetic Resources and Crop Evolution*, *68*(6), 2235-2263.
22. Prasathkumar, M., Anisha, S., Dhrisya, C., Becky, R., & Sadhasivam, S. (2021). Therapeutic and pharmacological efficacy of selective Indian medicinal plants–a review. *Phytomedicine plus*, *1*(2), 100029.
23. Yadav, G., & Meena, M. (2021). Bioprospecting of endophytes in medicinal plants of Thar Desert: An attractive resource for biopharmaceuticals. *Biotechnology Reports*, *30*, e00629.
24. Bhat, S. G. (2021). Medicinal plants and its pharmacological values. In *Natural medicinal plants*. IntechOpen.
25. Khajuria, A. K., Manhas, R. K., Kumar, H., & Bisht, N. S. (2021). Ethnobotanical study of traditionally used medicinal plants of Pauri district of Uttarakhand, India. *Journal of Ethnopharmacology*, *276*, 114204.