Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 18 May 2026
Revised 22 June 2026
Accepted 10 July 2026

Natural Resources for Human Health 2026, Vol. 6 (6s): 263–271

KEYWORDS:
Medicinal plants, Deep learning, Image classification, MobileNetV2, Streamlit, CNN.

A Convolutional Neural Network Approach for Automated Medicinal Flora Recognition

Dr. Vinit Kotak[1], Dr. Swati Nadkarni[2], Ms. Nivedeeta Mukherjee[3], Dr. Anil Avhad[4]

[1]Department of Information Technology Shah & Anchor Kutchhi Engineering College, Mumbai. Email: vinit.kotak@sakec.ac.in
[2]Department of Information Technology Shah & Anchor Kutchhi Engineering College Mumbai. Email: swati.nadkarni@sakec.ac.in
[3]Department of Information Technology Shah & Anchor Kutchhi Engineering College, Mumbai. Email: nivedeeta.mukherjee@sakec.ac.in
[4]Department of Botany R. J. College of Arts, Science and Commerce, Mumbai. Email: anilavhad@rjcollege.edu.in

ABSTRACT: Medicinal plants are integral to both traditional healing practices and modern pharmacological research. However, accurate identification of these plants remains a persistent challenge, particularly for individuals without botanical expertise. This paper presents a deep learning-based system for automatic medicinal plant identification and information retrieval. The proposed solution employs a Convolutional Neural Network (CNN) architecture, specifically MobileNetV2, to classify plant species from user-submitted images with a classification accuracy of 75% with certain species reaching up to 93.5% accuracy. Upon successful identification, the system provides comprehensive medicinal information, including the plant's scientific name, therapeutic uses, and health benefits. A user-friendly web application is developed using Streamlit, enabling real-time interaction, while MongoDB is used for storing user interaction and metadata. This system is intended to assist students, researchers, farmers, and healthcare professionals by bridging the gap between plant identification and medicinal knowledge through an accessible, automated platform.

1 INTRODUCTION

Medicinal plants have long served as a foundation for both traditional healing systems—such as Ayurveda, Traditional Chinese Medicine, and indigenous remedies—and modern pharmacological discoveries. With an increasing interest in herbal and plant-based therapies worldwide, the demand for accurate and efficient identification of medicinal plant species has grown significantly [1, 2]. However, manual identification based on morphological characteristics remains a time consuming and error-prone process, especially for individuals lacking botanical expertise.

Recent advancements in computer vision and deep learning have opened new avenues for automating plant identification with high accuracy. Deep Convolutional Neural Networks (CNN) have proven particularly effective in recognizing plant species from leaf images, even in real-world, uncontrolled environments [3, 4, 5]. Several studies have developed mobile or web-based tools that utilize lightweight architectures like MobileNetV2 and Efficient Net for real-time classification and deployment on resource-constrained devices [6, 7, 8].

Despite the progress, most existing systems primarily focus on species classification without integrating domain specific medicinal information. This creates a practical gap, especially for users such as students, farmers, or healthcare professionals, who require not only plant identification but also details about its therapeutic uses and health benefits [9].

To address this, we propose a web-based application that combines plant species recognition with medicinal knowledge dissemination. The system utilizes the MobileNetV2 architecture for efficient and accurate classification of medicinal plants based on images captured or uploaded by the user. Once identified, the application provides essential information including the plant's scientific name, medicinal properties, and common uses. Developed using Streamlit for the front-end and MongoDB for data storage, this tool offers an interactive and informative platform tailored for educational, agricultural, and healthcare contexts.

2 LITERATURE REVIEW

Recent advancements in deep learning (DL) have significantly influenced the field of medicinal plant identification, offering reliable alternatives to traditional manual classification methods. The primary approach involves training CNN on plant image datasets to perform automated recognition based on leaf shape, color, and texture. Many researchers have adopted popular architectures such as MobileNet, ResNet, Inception, and EfficientNet due to their high classification accuracy and suitability for real-time applications.

A significant body of research demonstrates the success of deep CNN in real-time plant identification.

Malik et al. [1] presented a CNN-based model capable of real-time identification of medicinal plants in natural environments, focusing on species from the Borneo region. Their work highlighted the robustness of CNN in handling noisy, real-world image data. Similarly, Kavitha et al. [3] employed MobileNetV2 for real-time identification, achieving high classification accuracy and demonstrating the efficiency of light weight models for mobile deployment.

Dubey, S. et al. [4] introduced a lightweight, Python-based leaf detection system using traditional image processing methods and is implemented as a web application on a PC.

S Prajapati et al. [5] also explored CNN-based models using data augmentation to avoid overfitting.

Transfer learning has emerged as a crucial strategy in overcoming the challenges posed by limited and imbalanced datasets.

Studies like those by Sahib et al. [10] and Ghosh et al. [11] leveraged deep transfer learning models such as ResNet50 and hybrid models to improve classification accuracy across diverse plant species.

Similarly, Khaira et al. [7] and Musyaffa et al. [8, 12] developed mobile applications incorporating MobileNetV2 and other pre-trained models to identify Indonesian medicinal plants with high precision.

Applications like MedicPlant [6] and IndoHerb [8, 12] demonstrate the practical implementation of DL models immobile environments, offering user-friendly interfaces and real-time performance. These systems can identify multiple medicinal plants and providing relevant medicinal information, making them valuable tools for public health

awareness and educational purposes.

Other notable contributions include federated learning approaches proposed by Viet et al. [13], which facilitate collaborative model training across decentralized data sources while preserving privacy.

Additionally, optimized CNN-SVM hybrids [14] and ensemble methods like HybNet [15] have been shown to enhance model robustness and reduce overfitting.

Several papers emphasized the importance of real-time deployment and user-centric design. For instance, Sugiarto et al. [16] developed a mobile app using CNN for on-the-go identification of plant leaves, while Thapa et al. [9] created a CNN-powered mobile application tailored for local Nepalese flora.

In addition to visual-based classification, efforts have been made to systematically review the existing literature, such as Mulugeta et al. [2], who analysed the performance of various DL models and highlighted gaps in dataset quality, class imbalance, and lack of medicinal metadata integration.

While most studies focus solely on plant recognition, few integrate medicinal context. This work bridges that gap by combining deep learning-based identification with detailed medicinal information, making it more useful for education, traditional medicine, agriculture, and public health.

3 SYSTEM ARCHITECTURE

The architecture consists of the following layers:

- **Frontend:** Built using Streamlit for ease of use and rapid development. Users can log in and upload an image.
- **Backend:**

– **Model:** A pretrained MobileNetV2 CNN fine-tuned on a custom dataset of 80 medicinal plant species.

– **Database:** MongoDB is used to store user pro-files and prediction history.

– **Middleware:** PyMongo handles all communication between the backend and the database.

4 METHODOLOGY

4.1 Dataset

For this project, we created a custom dataset consisting of over 80 medicinal plant species, with 100 images per species, resulting in a total of approximately 8,000 images. The images were collected from reliable online sources and field photography. Each image was labeled with the plant's common name, and medicinal uses. This tailored dataset enabled better control over image quality, diversity, and labeling accuracy, contributing to improved model performance.

4.2 Preprocessing

The collected images had varying resolutions, backgrounds, and lighting conditions. To standardize and optimize model performance, images were resized to 224×224 pixels, and pixel values were normalized to a 0–1 scale. Image augmentation techniques like rotation, flipping, zooming, and bright-ness adjustment were applied to improve generalization and prevent overfitting. Label encoding was used to convert plant names into numerical format. These preprocessing steps ensured a high-quality dataset for training the deep learning model.

4.3 Model Training

The MobileNetV2 model was fine-tuned using TensorFlow and Keras in a Jupyter Notebook. It was trained for 25 epochs with an 80:20 training-validation split. The Adam optimizer and categorical cross-entropy loss function were used. Model accuracy is shown in Table 1 and Table 2 with the accuracy graphs in Figure 1 and Figure 2. The comparative analysis of the work is presented in Table 3. The classification report for the trained model can be seen in Figure 3.

Table 1: Model Training Performance

Model Accuracy			
Accuracy	Precision	Recall	F1 Score
0.7543	0.78124	0.7535	0.7482

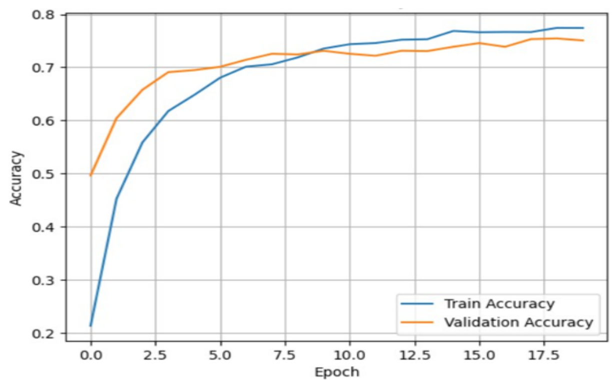

Figure 1: Model Accuracy

Natural Resources for
Human Health
eISSN: 2583-1194

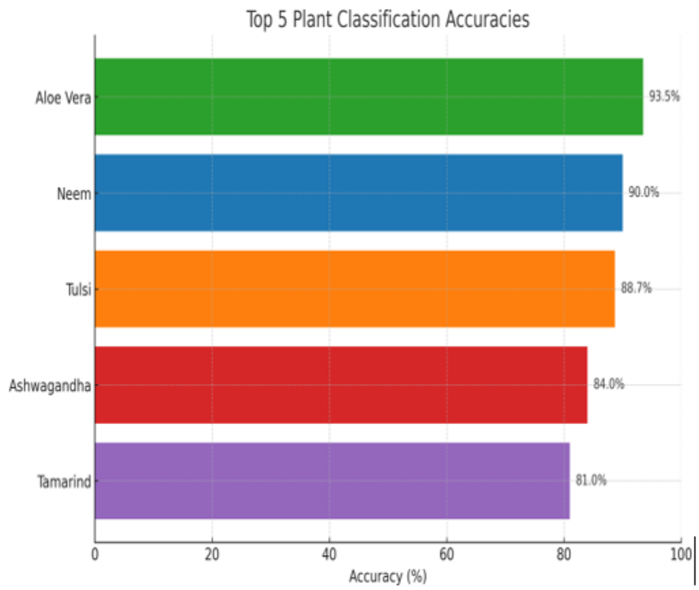

Figure 2: Top Plant Classification Accuracies

Table 2: Accuracy of specific medicinal plants using the model

Plant No.	Plant Name	Accuracy (%)
1	Aloe Vera	93.5
2	Neem	90.0
3	Tulsi	88.7
4	Ashwagandha	84.0
5	Tamarind	81.0

Table 3: Comparative Analysis of MobileNetV2 Models from Various Studies

Study	Dataset	Accuracy (%)
Malik O A et al., 2022 [1]	Field images from Borneo	75.4
Kavitha S et al., 2024 [3]	Real-time image capture	80.1
Khaira U et al., 2024 [7]	4,000+ Indonesian plant images	79.8
Sugiarto D et al., 2023 [16]	Leaf images from Indonesian flora	78.5
Tejaswini, N. P et al., 2025[21]	3027 images of different categories	88.0
Proposed System 2026	80 species of medicinal plants	93.5

Natural Resources for
Human Health
eISSN: 2583-1194

4.4 Evaluation Matrix.

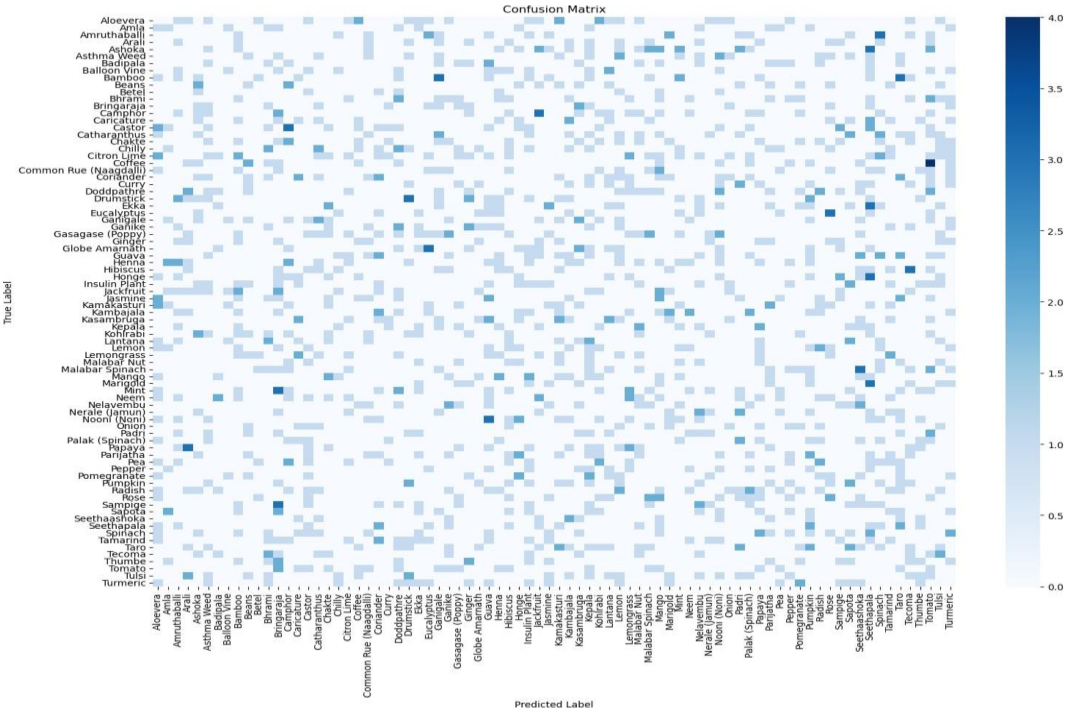

Figure 3: Confusion Matrix of Plant Species Classification Model on Multi-Class Dataset

5. Implementation

5.1 Web Interface

Users access the system via a Streamlit-powered web app. After logging in, users upload a plant image. The backend preprocesses the image and passes it to the model, which predicts the species, as shown in Figure 4.

5.2 Information Retrieval

Once the species is identified, the app queries a medicinal plant database (stored in MongoDB) and retrieves the plant's name, uses, chemical constituents, and applications in traditional medicine, as shown in Figure 5.

5.3 Technologies Used

The following technologies were used in the system:

1. **Jupyter Notebook**: Model training environment
2. **MobileNetV2**: CNN architecture
3. **MongoDB & PyMongo**: Database and interaction
4. **Pandas**: Dataset manipulation
5. **Pillow**: Image preprocessing
6. **Streamlit**: Web UI
7. **TensorFlow & Keras**: Model creation and training
8. **VS Code**: Development environment

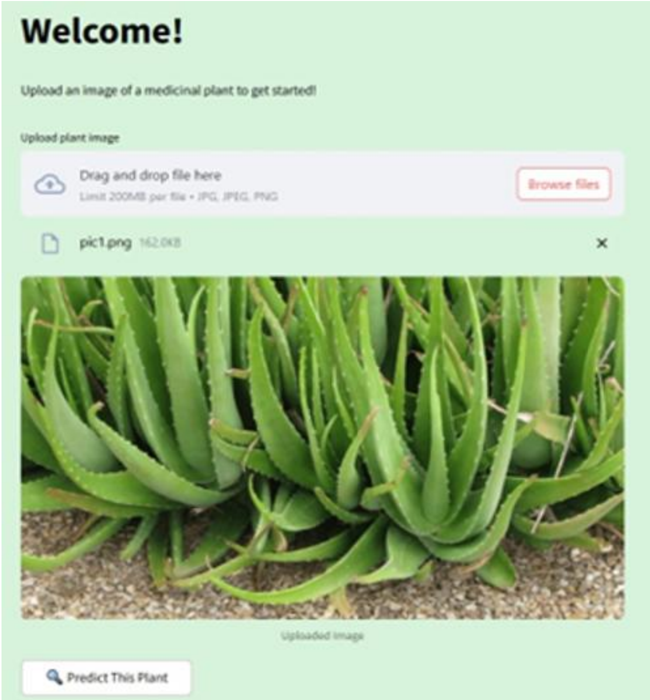

Figure 4: Image File Upload

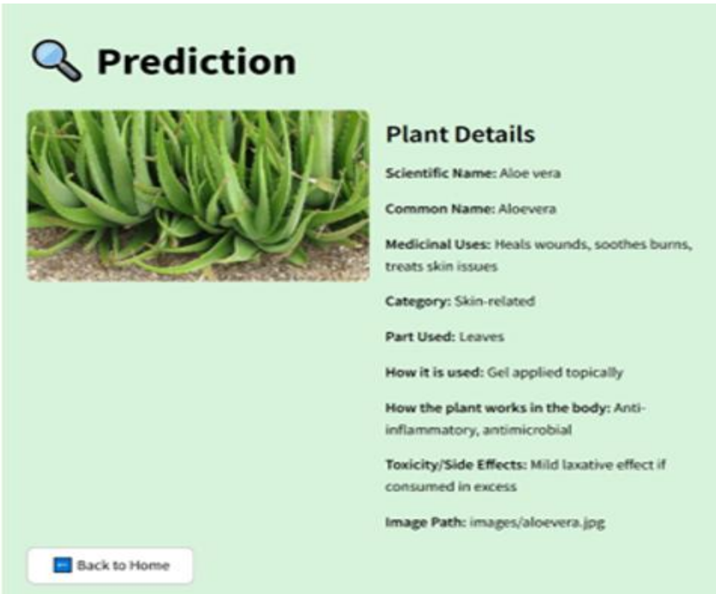

Figure 5: Model Prediction

5.4 Challenges Faced

During the development of the Medicinal Plant identification, several challenges were encountered. One major issue was gathering a balanced dataset with diverse images, especially for lesser-known medicinal plants. The model initially showed signs of overfitting, performing well on training data but poorly on validation data.

To address this, image augmentation and dropout layers were introduced to improve generalization. Another challenge was bias towards certain plant species, likely due to class imbalance or data distribution issues. This became evident when evaluating the model's performance with a confusion matrix, which showed significant misclassifications between species.

Figure 3 displays the confusion matrix for the plant species classification model. Analyzing this helped fine-tune the model's learning rate and training parameters, improving fairness and prediction balance. These adjustments led to more stable and reliable predictions, effectively addressing overfitting and class imbalance.

6. RESULTS & DISCUSSIONS

The system was tested on a separate validation set and showed strong performance, with high Accuracy, Precision, Recall, and F1 Score. Users found the interface intuitive and appreciated how quickly the system produced results. However, some false positives were observed, primarily due to poor image quality or background noise, which interfered with the model's ability to make accurate predictions, as shown in Figure 6. This highlights the need for improved preprocessing, such as image enhancement, noise reduction, or background removal, to refine the results and minimize errors. Addressing these issues would likely improve overall robustness and reliability, particularly in real-world applications where image quality may vary. These observations are further supported by Table 3, which compares the accuracy of MobileNetV2 models across various studies, emphasizing the importance of preprocessing and data quality. Future work should focus on fine-tuning these preprocessing steps to ensure the system can handle a wider range of input conditions.

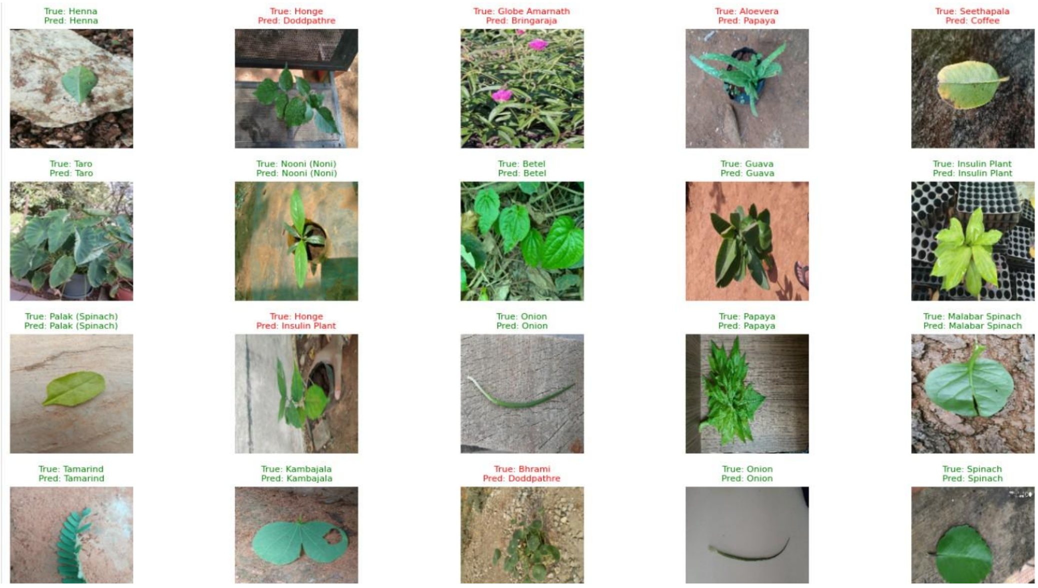

Figure 5. Medicinal Plant Species considered for proposed work.

7. CONCLUSION

This paper presents an intelligent and user-friendly medicinal plant identification system that integrates computer vision and deep learning techniques. By leveraging models such as MobileNetV2, the system offers a powerful solution for automatic plant species classification, even for non-experts, with an accuracy of 93.5 %. Several studies have highlighted the effectiveness of deep learning in plant recognition tasks, demonstrating its potential for real-world applications such as biodiversity monitoring and healthcare [17]. The approach combines traditional knowledge with advanced technology, enabling users to identify plants and access essential information in a simplified manner [20]. Our proposed system also offers a platform for future expansion, including the inclusion of additional plant species and improvements in model accuracy and efficiency. Moreover, the system is designed with scalability in mind, allowing easy integration with web-based and mobile platforms, as seen in recent works on real time plant identification applications.

8. FUTURE WORK

Future work will focus on enhancing both the system's capabilities and its accessibility. First, the dataset will be expanded by incorporating more plant species and increasing the number of images per class. This will not only improve model accuracy but also ensure better generalization across various environmental conditions, as demonstrated in studies that used diverse datasets for robust model training. Additionally, the system will be extended to support multi-modal data, such as audiobased plant identification or text descriptions, to complement the image-based recognition process.

A critical next step will involve the development of a dedicated mobile application that allows real-time plant identification directly from smartphones. This will enhance the system's portability and make it more practical for use in everyday environments, especially in rural or field settings where access to traditional plant identification resources may be limited. Real-time identification, supported by on-device processing, will be explored to ensure seamless functionality without requiring constant internet connectivity [18].

Furthermore, integrating community-driven platforms where users can contribute data and feedback will help expand the system's coverage and foster collaborative learning, as indicated by recent initiatives in crowdsourced plant identification applications [19].

REFERENCES

[1] Malik O A, Ismail N, Hussein B R, and Yahya U, 2022. Automated real-time identification of medicinal plant species in natural environment using deep learning models—A case study from Borneo region. *Sadhana*, 47(1):1–12.

[2] Mulugeta A K, Sharma D P, and Mesfin A H, 2024. Deep learning for medicinal plant species classification and recognition: a systematic review. *Frontiers in Plant Science*, 14: 1286088.

[3] Kavitha S, Satish Kumar T, Naresh E, Kalmani V H, Bamane K D, and Pareek P K, 2024. Medicinal plant identification in real-time using deep learning model. *SN Computer Science*, 5: 73.

[4] Dubey, S., Patel, K., Walele, T., Gala, C., Nadkarni, S., & Kotak, V. (2026, May). Leaf Disease Detection using Classical Image Processing and Rule-based Analysis. In 2026 7th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV) (pp. 1621-1626). IEEE.

[5] Prajapati, S., Qureshi, S., Rao, Y., Nadkarni, S., Retharekar, M., & Avhad, A. (2023, May). Plant disease identification using deep learning. In 2023 4th International Conference for Emerging Technology (INCET) (pp. 1-5). IEEE.

[6] Pudaruth S, Mahomoodally M F, Kissoon N, and Chady F, 2021. MedicPlant: A mobile application for the recognition of medicinal plants from the Republic of Mauritius using deep learning in real-time. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 10(4): 938–947.

[7] Khaira U, Saputra E, and Adriadi A, 2024. I-NusaPlant Apps: Indonesian medical plants identification using convolutional neural network with pre-trained model MobileNetV2. *Jurnal Ilmu Komputer dan Agri-Informatika*, 11(2): 185–194.

[8] Musyaffa M S I, Yudistira N, Rahman M A, Basori A H, Mansur A B F, and Batoro J, 2024. IndoHerb: Indonesia medicinal plants recognition using transfer learning and deep learning. *Heliyon*, 10(23): e40606.

[9] Thapa P, Sharma M, Nyachhyon J, and Pandeya Y R, 2025. Local herb identification using transfer learning: A CNN-powered mobile application for Nepalese flora. *arXiv preprint*, arXiv:2505.02147.

[10] Sahib K A, Oleiwi B K, and Nasser A R, 2024. Medicinal plants recognition using deep transfer learning models. *International Journal of Design & Nature and Ecodynamics*, 19(5): 1–10.

[11] Ghosh S, Singh A, and Kumar S, 2023. Identification of medicinal plant using hybrid transfer learning technique. *Indonesian Journal of Electrical Engineering and Computer Science*, 31(3): 1605–1615.

[12] Musyaffa M S I, Yudistira N, Rahman M A, and Batoro J, 2023. IndoHerb: Indonesia medicinal plants recognition using transfer learning and deep learning. *arXiv preprint*, arXiv:2308.01604.

[13] Viet K L D, Ha K L, Quoc T N, and Hoang V T, 2024. Medicinal plants identification using federated deep learning. *Procedia Computer Science*, 234: 247–254.

[14] Diwedi H K, Misra A, and Tiwari A K, 2023. CNN-based medicinal plant identification and classification using optimized SVM. *Multimedia Tools and Applications*, 83(11): 33823–33853.

[15] Pushpa B R, Jyothsna S, and Lasya S, 2024. HybNet: A hybrid deep model for medicinal plant species identification. *MethodsX*, 14: 103126.

[16] Sugiarto D, Siswantoro J, Naufal M F, and Idrus B, 2023. Mobile application for medicinal plants recognition from leaf image using convolutional neural network. *Indonesian Journal of Information Systems*, 5(2): 43–56.

[17] Hieu N V and Hien N L H, 2020. Automatic plant image identification of Vietnamese species using deep learning models. *International Journal of Engineering Trends and Technology (IJETT)*, 68(4): 25–31.

[18] Gokhale A, Babar S, Gawade S, and Jadhav S, 2020. Identification of medicinal plant using image processing and ma chine learning. In: Iyer B, Rajurkar A M, and Gudivada V (eds.), *Applied Computer Vision and Image Processing*, pp. 272–282. Springer.

[19] Tiwari N, Gupta B K, Prakash A, Tiwari K, Alshmrany S, Ali A, Husain M, and Singh D, 2023. Developing a neural network-based model for identifying medicinal plant leaves using image recognition techniques. *Journal of Advanced Zoology*, 44(S5): 1564.

[20] Patil S S, Patil S H, Azfar F N, Pawar A M, Kumar S, and Patel I, 2023. Medicinal plant identification using convolutional neural network. *AIP Conference Proceedings*, 2890(1):020023.

[21] Tejaswini, N. P., Karthikeya, M. S., Keerthi, M., Sowmya, M., & Babu, P. K. (2025, May). Deep Learning Based Medicinal Plants Identification. In 2025 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC) (pp. 1-7). IEEE.