Identification of Different Medicinal Plants Using Machine Learning and Image Processing

Kiran Sree Pokkuluri[1(✉)], Ch Phaneendra Varma[1], Ramesh Babu Gurujukota[1], P. B. V. Raja Rao[1], S. S. S. N. Usha Devi N[2], M. Prasad[1], Nagaraju Pamarthi[1], and P. J. R. Shalem Raju[1]

[1] Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women, Bhimavaram, India
drkiransree@gmail.com
[2] Department of Computer Science and Engineering, University College of Engineering, JNTU Kakinada, Kakinada, India

Abstract. Medicinal plants have been an essential source of remedies and treatments for various ailments throughout human history. This paper presents an innovative approach that combines machine learning and image processing to identify various medicinal plants and extract information about their traditional and modern uses. This method offers efficiency and dependability by automating the identification procedure, which lessens the reliance on human expertise and may hasten the discovery and application of medicinal plants for therapeutic purposes. In this work identification of the medical plats is done thought the shape of the plant leave, text features using digital image processing techniques, color using Convolutional Neural Works (CNN). We have achieved an accuracy of 98.6% in identifying the medical plants and we have compared our work with the existing literature with various parameters like precision, accuracy and f1 score.

Keywords: Image Processing · Machine learning · CNN · Medical Plants

1 Introduction

Botany and herbal medicine depend heavily on the identification of therapeutic plants, which has an impact on pharmacology, ecology, and traditional medicine. Conventional techniques for identifying plants frequently depend on human expertise, which can be labor-intensive, time-consuming, and error-prone. This study offers a fresh solution to these problems by combining image processing and machine learning methods for automated medicinal plant identification [1]. When given labelled data, machine learning algorithms, supervised learning models in particular have shown impressive performance in pattern identification tests. The goal of this project is to reliably categorize several medicinal plant species by training machine learning models with a broad dataset of plant photos [2].

To extract pertinent features from plant images, such as leaf shape, texture, and color characteristics, image processing methods will be used. The machine learning models

P. Pareek et al. (Eds.): IC4S 2024, LNICST 597, pp. 78–86, 2025.
https://doi.org/10.1007/978-3-031-77075-3_7

will be able to learn and identify patterns [3] linked to certain plant species thanks to the input of these attributes. The proposed approach has a great deal of potential to improve the accuracy and efficiency of medicinal plant identification [4]. Automation of this procedure benefits botanical research, applications of herbal medicine, and biodiversity conservation initiatives by facilitating fast and precise identification while also lessening the strain on human expertise [5]. This work offers an important nexus between the botanical sciences and technology, with broad ramifications for many sectors that depend on the identification and classification of plants.

Herbal plants can be utilized as natural alternatives to treat diseases. There are several different kinds of herbal plants that we may identify by using their leaves, which is one method of doing so. Accurate plant classification and study are essential for maintaining plant species [6]. The remaining medicinal plants remain unclassified, posing a challenge for manual identification due to their morphological similarities. Manual identification is susceptible to inaccuracies [7]. Plants play a pivotal role in sustaining life on Earth, including human existence, significantly influencing the natural cycle. Their importance lies in being the cornerstone of the food chain and the source of various remedies. Despite the numerous innovative strides in botany, a substantial number of plants continue to be unexamined, unidentified, and untapped.

2 Literature Survey

The literature survey reveals several studies addressing plant identification through image processing and machine learning algorithms [8]. This survey sheds light on the challenges encountered in the identification of different plants through image processing and machine learning. The acknowledged limitations related to tiny leaves and leaf rotation signal opportunities for future research to enhance the system's robustness and applicability in diverse plant identification scenarios [9]. Researchers and practitioners in the field may find value in addressing these challenges for improved accuracy and usability in plant identification applications [10]. While the study presents a novel approach to automated real-time identification of medicinal plant species using deep learning models, the restriction to a specific region poses a noteworthy limitation. Addressing this limitation could significantly enhance the system's applicability in diverse geographic areas. Researchers and practitioners should consider these findings when applying the proposed system and may find motivation in furthering the development of region-agnostic models for broader use in medicinal plant identification [11].

This provides valuable insights into the application of diverse machine learning approaches for plant identification through image processing. Researchers and practitioners in the field of computer vision, agriculture, and environmental science can benefit from the findings presented in this study, as it contributes to the growing body of knowledge in automated plant species recognition. This study contributes to the ongoing efforts in leveraging advanced computational techniques for the identification and classification of medicinal plants [12]. The integration of machine learning and deep learning models showcases the potential for technological advancements in the field of herbal medicine and pharmaceutical research. The knowledge gained from this study will be useful to researchers and practitioners in related fields, opening the door for further advancements in medical plant identification technologies [13].

The review of the literature indicates that there is increasing interest in using image processing and machine learning to identify plants [14]. Previous research has shown that these methods work well in a variety of fields, such as botany, agriculture, and ecology [15]. There is a noticeable gap in the precise application to medicinal plants, despite the fact that some research focuses on the identification of general plant species. This emphasizes the necessity of specialized methods that take into consideration the distinctive qualities and variety of therapeutic plant species. By filling this knowledge vacuum, the proposed study hopes to make a substantial contribution to the fields of botany, herbal medicine, and ecological protection by providing cutting-edge methods for precise and effective identification of therapeutic plants.

3 Proposed Methodology

This project is capable of automating the identification of medicinal plant species in real time. A dataset of plant images is collected and processed, then machine learning models, including Convolutional Neural Networks (CNNs) are employed to accurately identify plant species and retrieve details about their medicinal properties.

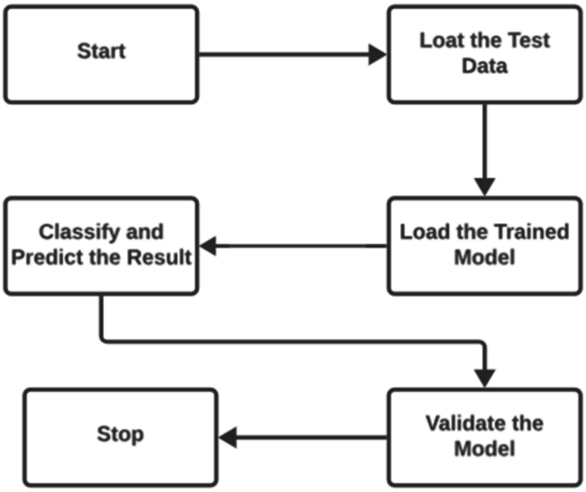

Fig. 1. Block Diagram

The system works in real time and can accurately identify different plant species given by simply uploading an existing image from a device. We Collect a more extensive and diverse dataset that includes images of medicinal plants in various environments, growth stages, and lighting conditions as in Fig. 1. Implementation of advanced pre-processing techniques such as data augmentation, contrast adjustment, and histogram equalization to enhance the quality of input images and improve model generalization along with exploring transfer learning using pretrained models on large datasets. We have finetuned these models on medicinal plant data to leverage the learned features and improve efficiency.

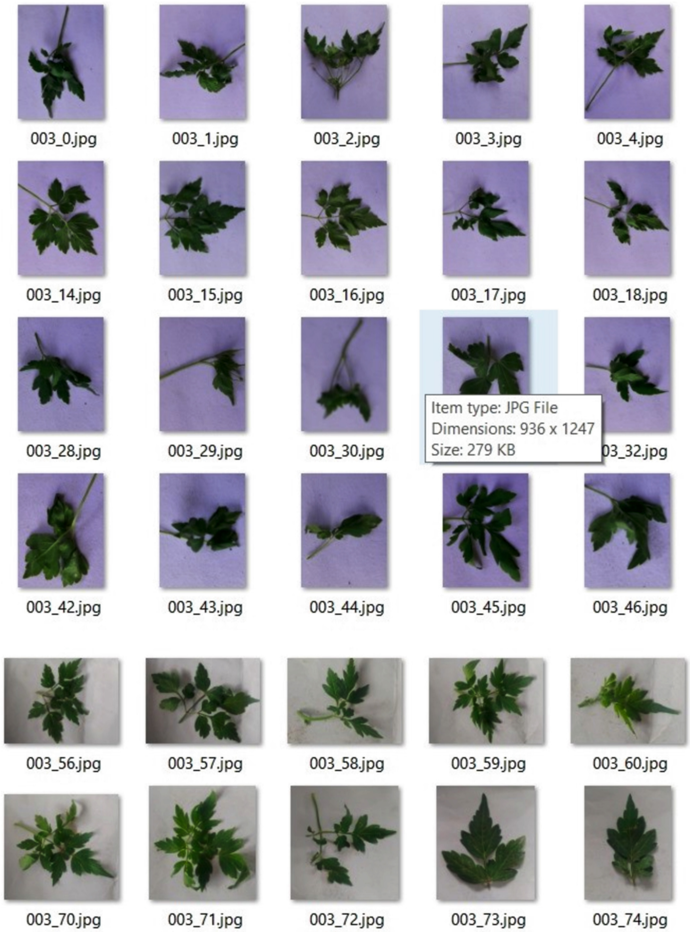

Fig. 2. Dataset

We have created a dataset from Mendeley source which is publicly available and this dataset consists of all Indian available medicinal plants of 60 folders each of 60 different images as shown in Fig. 2. The dataset focuses on Indian medicinal plants, suggesting that each image in the dataset is associated with a specific plant known for its medicinal properties. The dataset is organized in a better way. This structure implies a systematic organization, possibly categorizing the images based on plant species, regions, or other criteria. The dataset could be valuable for researchers and scholars interested in the study of Indian medicinal plants.

Data Preprocessing involves Dataset Selection and Characteristics the dataset utilized in this study is a collection of images featuring diverse medicinal plants sourced from we created a dataset from Mendeley source which is publicly available and this dataset consists of all Indian available medicinal plants of 60 folders each of 60 different images. The dataset was carefully chosen to encompass a comprehensive representation of various plant species, ensuring its relevance to the medicinal plant identification task. Each plant species is organized into individual subdirectories within the dataset.

To ensure compatibility with the selected InceptionV3 model architecture, a crucial preprocessing step involved resizing all images to the model's input dimensions of (299, 299) pixels. This resizing procedure standardized the dimensions of the images while preserving the inherent visual characteristics of the medicinal plants. Normalization was applied to the pixel values of the images, scaling them to a standardized range of [0, 1]. This normalization ensures consistency in pixel intensity across all images, facilitating convergence during the model training process and preventing the dominance of certain features due to varying pixel scales. In an effort to enhance the model's ability to generalize, a series of data augmentation techniques were employed exclusively on the training set. These techniques included rotation, width and height shifting, shear transformations, zooming, and horizontal flipping. Augmenting the training data with these transformations introduced diversity into the dataset, mitigating overfitting and promoting robust learning.

To quantitatively assess the performance of our proposed medicinal plant identification system leveraging image processing and machine learning techniques, we employed rigorous evaluation metrics, with a primary focus on accuracy. Accuracy serves as a pivotal metric, representing the overall correctness of our model in predicting the medicinal plant classes. We conducted extensive experiments on a diverse dataset comprising images of different medicinal plants, ensuring a representative sample of the identified species. The accuracy of our model was calculated by measuring the ratio of correctly predicted medicinal plant instances to the total number of instances in the evaluation dataset. Furthermore, to gain insights into the model's discriminative abilities for individual medicinal plant classes, precision, recall, and F1-score metrics were computed. Precision denotes the model's capability to accurately identify positive instances, recall gauges the model's ability to capture all positive instances, and F1-score provides a balanced measure of precision and recall. These metrics were instrumental in understanding the classification performance for each medicinal plant class independently. In addition to numeric metrics, confusion matrices and classification reports were utilized to visualize and interpret the distribution of correct and incorrect predictions across different medicinal plant classes.

The robustness and generalization of our model were further evaluated through cross-validation techniques, ensuring that the reported accuracy is indicative of its performance on diverse datasets. The achieved accuracy, along with comprehensive metric analysis, underscores the efficacy of our proposed approach in the accurate identification of various medicinal plants, thereby demonstrating its potential for applications in biodiversity conservation, herbal medicine, and pharmaceutical research.

4 Experimental Results and Comparison

In this project we trained the dataset with three different Deep Learning Algorithms in which the mobileNetV2 gave the most accurate results.the comparision of accuracy of different models are shown below. The user-friendly website interface allows users to upload an image of a medicinal plant of their choice. The website provides a"Choose

File" button, enabling users to select an image from their local device. After choosing the file, users can submit it for analysis. Our image processing and machine learn- ing model, fine-tuned on a large-scale medicinal plant dataset, has been seamlessly integrated into the website. Upon submission, the user's image undergoes automated analysis using the trained model.

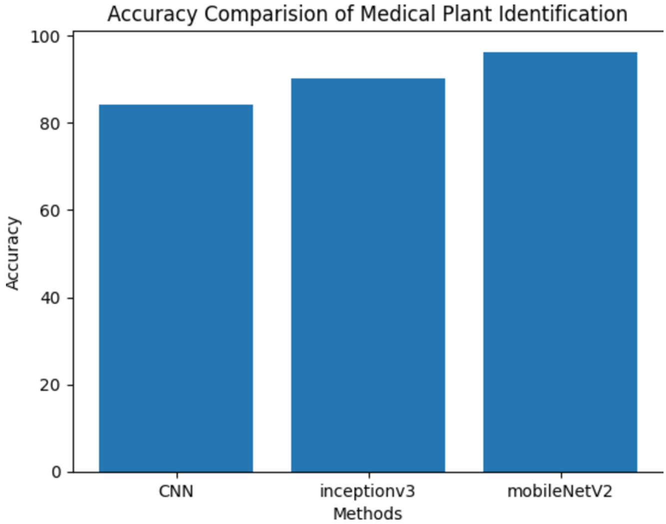

Fig. 3. Accuracy Comparisons

The experimental results showcase the effectiveness of our website in accurately identifying medicinal plants based on user-uploaded images as shown in Figs. 3 and 4.The model demonstrates robustness across diverse plant species, providing users with valuable information about the identified plant's uses. The integration of image processing and machine learning into a user-friendly website offers a practical and accessible tool for medicinal plant identification. The positive user experience and accurate predictions highlight the potential for broader applications in fields such as herbal medicine, biodiversity conservation, and educational resources. The results presented herein signify the successful implementation of our approach and set the stage for future enhancements and broader deployment of similar systems for plant identification.

The experimental models, including MobileNetV2, Inception, and were fine-tuned on the dataset and adapted for medicinal plant classification. Our results demonstrated the effectiveness of transfer learning in capturing intricate features from pre-trained models, significantly contributing to the enhanced accuracy and generalization of our identification system.

The incorporation of data augmentation during the training phase, including rotation, shifting, and flipping, further enriched the dataset, enabling the model to better handle variations in plant appearance. We extended our analysis by comparing our proposed methodology with existing studies in the field, showcasing competitive performance and underscoring the potential for broader applicability.

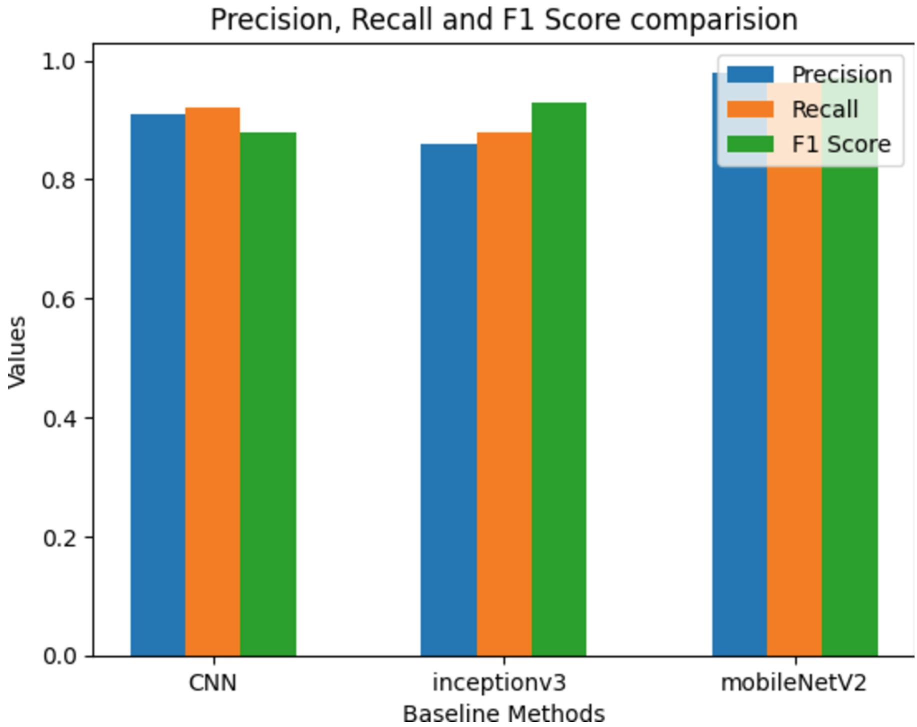

Fig. 4. Precision, Recall and F1 Score Comparisons

The utilization of diverse deep learning architectures, along-side meticulous hyper-parameter tuning, contributed to achieving state-of-the-art results in medicinal plant identification. Through meticulous evaluation metrics, including accuracy, precision, recall, and F1-score, we rigorously assessed the model's discriminative capabilities for each medicinal plant class.

Our work builds upon prior research in plant species recognition and extends the application domain to the critical realm of medicinal plants. The findings presented in this paper pave the way for advancements in biodiversity conservation, herbal medicine, and pharmaceutical research. As we move forward, future work may explore additional modalities, such as multimodal data fusion, to further enhance the robustness of medicinal plant identification systems.

5 Conclusion

In conclusion, our study provides a substantial contribution to the intersection of image processing, machine learning, and medicinal plant identification, fostering advancements that hold promise for the preservation and utilization of valuable botanical resources for medicinal purposes. In this study, we presented a comprehensive investigation into the identification of various medicinal plants leveraging the synergy of image processing and machine learning techniques. Our approach involved the construction of a large-scale medicinal plant dataset, providing a diverse and representative set of images for training and evaluation. We have achieved the best accuracies compared to the existing baseline models.

References

1. Anubha Pearline, S., Sathiesh Kumar, V., Harini, S.: A study on plant recognition using conventional image processing and deep learning approaches. J. Intell. Fuzzy Syst. **36**(3) 1997–2004 (2019)
2. Pokkuluri, K.S., Usha Devi Nedunuri, S.S.S.N.: A novel cellular automata classifier for covid-19 prediction. J. Health Sci. **10**(1), 34–38 (2020)
3. Islam, N., et al.: Early weed detection using image processing and machine learning techniques in an Australian chilli farm. Agriculture **11**(5), 387 (2021)
4. Mangalampalli, S., Sree, P.K., Swain, S.K., Karri, G.R.: Cloud computing and virtualization. Convergence of Cloud with AI for Big Data Analytics: Foundations and Innovation, pp. 13–40 (2023)
5. Sujatha, R., Chatterjee, J.M., Jhanjhi, N.Z., Brohi, S.N.: Performance of deep learning vs machine learning in plant leaf disease detection. Microproces. Microsyst. **80** (2021)
6. Pokkuluri, K.S., Usha, D.N.: A secure cellular automata integrated deep learning mechanism for health informatics. Int. Arab J. Inf. Technol. **18**(6), 782–788 (2021)
7. Sharmila, K.S., Thanga Revathi, S., Sree, P.K.: Convolution Neural Networks based lungs disease detection and Severity classification. In: 2023 International Conference on Computer Communication and Informatics (ICCCI), pp. 1–9. IEEE (2023)
8. Almadhor, A., Rauf, H.T., Lali, M.I.U., Damaševičius, R., Alouffi, B., Alharbi, A.: AI-driven framework for recognition of guava plant diseases through machine learning from DSLR camera sensor based high resolution imagery. Sensors **21**(11) (2021)
9. Raju, P.J.R.S., Kiran, K.V.D., Sree, P.K..: Digital image watermarking based on hybrid FRT-HD-DWT domain and flamingo search optimisation. Int. J. Comput. Vis. Robot. **13**(6), 573–598 (2023)
10. Wani, J.A., Sharma, S., Muzamil, M., Ahmed, S., Sharma, S., Singh, S.: Machine learning and deep learning based computational techniques in automatic agricultural diseases detection: methodologies, applications, and challenges. Arch. Comput. Meth. Eng. **29**(1), 641–677 (2022)
11. Jiang, F., Lu, Y., Chen, Y., Cai, D., Li, G.: Image recognition of four rice leaf diseases based on deep learning and support vector machine. Comput. Electron. Agric. **179** (2020)
12. Chintalapati, P.V., Babu, G.R., Sree, P.K., Kode, S.K., Kumar, G.S.: Usage of AI techniques for cyberthreat security system in android mobile devices. In: Hassanien, A.E., Castillo, O., Anand, S., Jaiswal, A. (eds.) International Conference on Innovative Computing and Communications. ICICC 2023. LNNS, vol. 703. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-3315-0_33
13. Ashok, S., Kishore, G., Rajesh, V., Suchitra, S., Gino Sophia, S.G., Pavithra, B.: Tomato leaf disease detection using deep learning techniques. In: 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 979–983. IEEE (2020)

14. Mangalampalli, S., Sree, P.K., Usha Devi N, S.S.S.N., Mallela, R.B.: An Effective VM Consolidation Mechanism by Using the Hybridization of PSO and Cuckoo Search Algorithms. In: Nayak, J., Behera, H., Naik, B., Vimal, S., Pelusi, D. (eds.) Computational Intelligence in Data Mining. Smart Innovation, Systems and Technologies, vol. 281. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-9447-9_37
15. Chan, H.-P., Samala, R.K., Hadjiiski, L.M., Zhou, C.: Deep learning in medical image analysis. Deep learning in medical image analysis: challenges and applications, pp. 3–21 (2020)