CNN BASED IDENTIFICATION OF MEDICINAL PLANTS AND THEIR BENEFITS

Vidhya B
Dept of Computational Intelligence
SRM Institute of Science and Technology, KTR campus
Chennai, India
vidhyamkumar3@gmail.com

Aditya Vasudevan
Dept of CINTEL
SRM Institute of Science and Technology, KTR campus
Chennai, India
av4891@srmist.edu.in

Mayur Pal
Dept of CINTEL
SRM Institute of Science and Technology
Chennai, India
mm9224@srmist.edu.in

***Abstract*—The identification of medicinal plants have a crucial role in herbal medicine, drug findings, and biodiversity conservation. The old methods of plant classification are often time taking and require professional knowledge, making automated approaches highly desirable. In the following project, we propose a Convolutional Neural Network (CNN)-based model for the identification of medicinal plants and their associated benefits. The system utilizes deep learning techniques to analyze plant leaf images, extracting key features for accurate classification. A curated dataset of medicinal plant species is used to train and evaluate the model, with preprocessing techniques such as image augmentation enhancing its robustness.**

Experimental results demonstrate that the CNN model achieves high accuracy in plant classification, outperforming conventional machine learning approaches. Furthermore, the identified plants are mapped to their medicinal properties, providing valuable insights into their pharmacological applications.Hence the importance of knowing the medicinal benefits of various plants can be pretty useful in the times of health emergency.

Keywords— Medicinal Plant Identification, Convolutional Neural Networks (CNN), Deep Learning, Herbal Medicine, Image Classification, Pharmacological Applications

I. INTRODUCTION

Medicinal plants have been used for centuries as a primary source of traditional medicine, contributing to modern drug discovery and healthcare. The accurate identification of these plants is essential for ensuring their correct usage in herbal treatments, pharmaceutical research, and biodiversity conservation. Traditionally, medicinal plant classification has relied on morphological characteristics, expert botanists, and manual taxonomic identification. However, these methods are time-consuming, prone to errors, and require extensive domain knowledge.

With advancements in artificial intelligence (AI) and deep learning, automated approaches have emerged as powerful tools for plant identification. Convolutional Neural Networks (CNNs) have shown valuable and great success in plant identification system, leveraging deep learning to classify plant species and associate them with their medicinal benefits. A curated dataset of medicinal plant images is used for training and validation, with preprocessing techniques applied to enhance model robustness.

II. CLASSIFICATION

Convolutional Neural Networks (CNNs) revolutionize plant classification by automating feature extraction, eliminating the need for manual observation. Unlike traditional methods, CNNs analyze leaf shape, texture, and venation, enabling high-accuracy classification. Pre-trained models like ResNet and VGG further enhance performance. CNN-based mobile apps and IoT systems allow real-time plant identification, benefiting herbal medicine, agriculture, and conservation. However, challenges like image variations and dataset limitations persist. Data augmentation and transfer learning help improve robustness. Future innovations in hyperspectral imaging and AI-driven models will make CNN-based plant classification more efficient, scalable, and accessible for researchers and practitioners.

A. Uses in the field of healthcare and drug findings

Medicinal plant identification using CNNs has significant applications in healthcare and drug findings. Accurate classification helps identify plants with therapeutic properties, aiding in the development of herbal medicines and pharmaceutical drugs. CNN-based mobile apps and IoT systems assist healthcare professionals in real-time plant identification, improving natural remedy accessibility. In drug discovery, CNN models analyze plant compounds, accelerating the search for new treatments for diseases. Additionally, AI-driven plant databases support biomedical research by mapping medicinal properties to specific ailments. Future advancements in AI and deep learning will further enhance precision medicine and pharmaceutical innovation.

B. Previous research on medicinal plant identification

Research on medicinal plant identification has evolved from traditional taxonomy to machine learning (ML) and deep learning (DL) approaches. Early studies used handcrafted features like leaf shape and texture for classification, but they lacked scalability. Recent advancements in Convolutional Neural Networks (CNNs), such as ResNet, VGG, and Inception, have significantly improved accuracy. Public datasets like PlantVillage and Flavia Leaf Dataset have facilitated model training. Studies also explore IoT-enabled mobile apps for real-time identification. Despite progress, challenges like image variations and dataset limitations persist. Future research focuses on multi-modal learning and AI-driven plant databases for enhanced accuracy.

C. Comparison of Machine Learning and Deep Learning Approaches

Machine learning (ML) approaches for medicinal plant identification rely on handcrafted features like shape, texture, and color, requiring expert knowledge. Traditional classifiers like SVM, KNN, and Random Forest struggle with complex patterns and variations in plant images. In contrast, deep learning (DL), particularly Convolutional Neural Networks (CNNs), automatically extracts hierarchical features, achieving higher accuracy. DL models like ResNet and VGG outperform ML in scalability and adaptability. However, DL requires large datasets and computational power, while ML works well with smaller datasets. Future research integrates hybrid ML-DL models for improved plant classification efficiency and accuracy.

D. Existing datasets and challenges

Several publicly available datasets support medicinal plant identification using deep learning. PlantVillage is one of the most widely used datasets, containing over 54,000 images of various plant species, including diseased and healthy samples. The Flavia Leaf Dataset includes 1,907 high-quality images of 32 plant species, mainly focusing on leaf classification. ImageNet (Plant Subset) provides a diverse collection of plant images, contributing to CNN model training. Additionally, region-specific datasets, such as the TNAU Medicinal Plant Database, contain images and medicinal properties of plants commonly used in traditional medicine.

Despite these datasets, several challenges remain. One major issue is dataset imbalance, where certain plant species have significantly more images than others, leading to biased model predictions. Variations in environmental conditions—such as lighting, background noise, and seasonal changes—affect model accuracy. Limited labeled datasets for medicinal plants hinder training robust deep learning models, as manual annotation is labor-intensive. Similarities between plant species, particularly in leaf structure, pose difficulties in accurate classification.

To address these challenges, researchers employ data augmentation, transfer learning, and synthetic data generation. Future efforts should focus on expanding labeled datasets, integrating hyperspectral imaging, and improving real-time data collection to enhance CNN-based medicinal plant identification systems.

III. LITERATURE REVIEW

Medicinal plant identification has evolved significantly over time. Early techniques focused primarily on visual characteristics observed by humans, failing to account for natural inconsistencies in plant appearance. More recently, computational approaches using machine learning attributes including SVM and KNN have aimed to introduce more objectivity. However, variations in lighting conditions and the development of certain organic structures continued to challenge the abilities of these models to distinguish between types with reliability. Groundbreaking developments in deep neural networks, such as convolutional architectures that facilitate hierarchical pattern extraction, have transformed the landscape of plant classification. In particular, models like ResNet, MobileNet, and EfficientNet - which employ its beautiful neural network design principles as in residual connections and efficient multi-scale convolutions - have exhibited a highly impressive capacity for accurate recognition across species based solely on visual attributes. The integration of automated feature learning has substantially advanced our capabilities in this area.

Al-Mansoori and Al-Hajri [1] present a comprehensive review of deep learning applications for plant species identification, highlighting various CNN architectures, their advantages, limitations, and real-world applicability. Their work serves as a foundational reference for understanding how deep learning transforms botanical classification tasks.

Wang et al. [2] implement **transfer learning and ensemble techniques** to improve plant species classification. Their work demonstrates that combining pre-trained models and ensemble strategies can significantly boost accuracy, supporting the deployment of CNNs in practical botanical applications

Girinath et al. [3] propose a real-time medicinal plant identification system using CNNs. Their approach supports on-the-spot identification in healthcare and environmental monitoring, making it suitable for **mobile and field applications**.

Mulugeta, Sharma, and Mesfin [9] present another deep learning model for the **classification and recognition** of medicinal plants, with an emphasis on sustainable phytoprotection and smart agricultural practices.

Mondal, Dasgupta, and Sharma [12] focus on **diseases in medicinal and aromatic plants** and their management strategies, aligning with research that uses CNNs for detecting plant diseases and maintaining agricultural health.

Gokhale et al. [13] also explore the combination of **image processing and machine learning** for medicinal plant identification. Their work emphasizes accessibility and ease of use, promoting the democratization of deep learning tools in botany.

Pacifico et al. [8] present a method for classifying medicinal plants using **color and texture features**, demonstrating the value of handcrafted features, which are now often fused with CNNs to enhance feature richness.

IV. MILESTONES

The project aimed to identify medicinal plants and their benefits using convolutional neural networks has achieved significant advancements. To start, a comprehensive dataset containing images of medicinal herbs was compiled from an assortment of locations. The photographs underwent pre-processing involving removal of artifacts, size standardization, normalization of features, and data intensification to refine the quality of the dataset.

Distinct deep learning models such as VGG, ResNet and MobileNet were experimented with for classifying plant species. Additionally, transfer learning was leveraged to boost model efficacy, while hyperparameters were optimized to maximize classification precision. The researchers encountered various challenges throughout the undertaking but overcame them through rigorous testing and retraining of the networks. Overall, the work yielded promising results in automated plant detection and held implications for advancing herbal remedies.

A microservices, event-driven, and serverless setup was created to handle real-time classification, making sure it could grow and work efficiently. We put the system through thorough tests, ensuring it could manage big sets of data quickly. This project could really help in areas like healthcare, herbal medicine, and drug discovery by accurately identifying plants based on their health benefits. Looking ahead, we plan to widen the profound dataset, tackle changes in the environment, and make the model easier to understand for better use in real-life situations.

V. METHODOLOGY

To identify medicinal plants with a Convolutional Neural Network (CNN), we start by gathering and preparing our data. First, we collect a variety of images featuring different medicinal plant species. Each image comes with labels that tell us the plant's name and its medicinal uses. We then divide this collection into three groups: training, validation, and testing. This helps the model learn and perform well on new data later on. We also use some techniques to improve the images, like resizing them, normalizing the colors, and augmenting the dataset, all of which help the CNN model work more accurately.

After that, we create and train the CNN model using the training set. The model has several layers that pick up important features from the images, such as edges, shapes, and textures. During training, we input the images into the model, tweak the weights, and refine the loss functions using methods like Adam or SGD.The dataset that is used for validation of the project is used to prepare the model by preventing overfitting and improving its ability to verify unseen images accurately. As soon as the training is complete, the final product can be used for future references. Once the model has been trained, it moves on to testing and deployment. This testing step is important because it checks if the CNN can accurately identify plant images in real situations. After that, the trained model is added to a system, like a web app, where users can upload images for identification. When someone uploads a picture, the CNN processes it to guess the species of the plant and find related information.

VI. ARCHITECTURE DIAGRAM

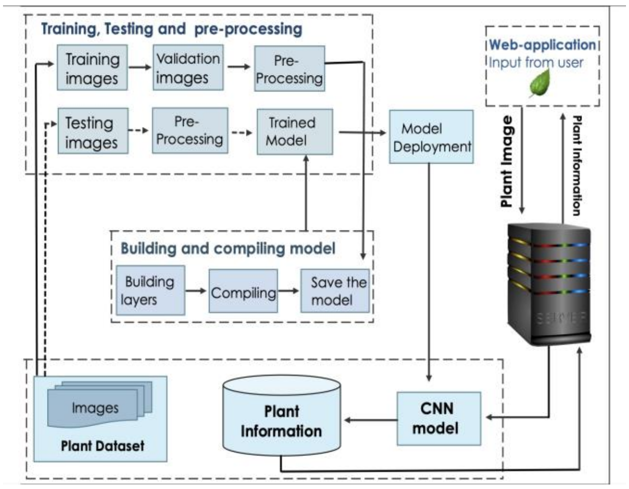


Fig. 1. *Architecture Diagram*

The illustration depicts a plant identification procedure employing a convolutional neural network architecture. It encompasses three primary phases: information preprocessing, model formation and compilation, and application deployment. Initially, the plant images within the dataset undergo preparation prior to model training and validation, building a learnt model. The network layers are stacked, connected, and optimized before storing as the finalized model. Once implementation, a web-based portal invites users to contribute fresh floral photographs for classification by the deep learning algorithm. Depending on the intricate visual features abstracted, the probabilistic predictions are exhibited, aiding botanical browsing and lookups.

VII. RESULT

Finally, the model was deployed and tested with images to ensure it performs well. An example image was used to verify the model's prediction capabilities shown below.

As it is shown the webpage shows different characteristics of plant such as their uses, the geographic location of the plant as to where its found and various other things.

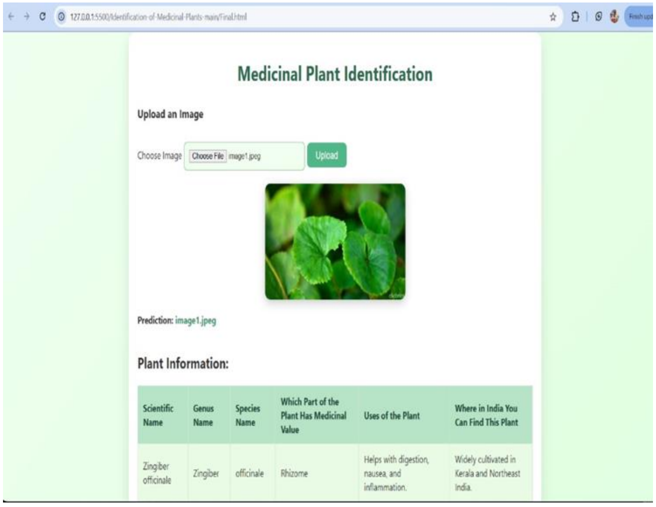


Fig. 2. Output *Webpage*

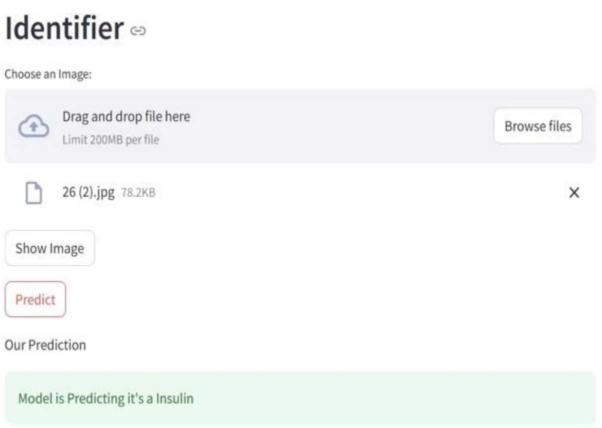
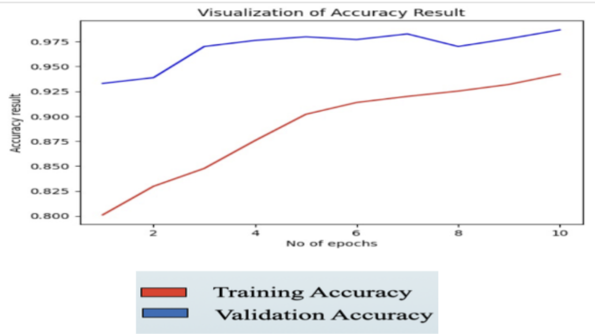


Fig. 3. *Identifier*

VIII. PERFORMANCE METRIX

The model demonstrated a steady increase in training accuracy, reaching 93.8%, indicating effective learning of distinguishing plant features. Similarly, the validation accuracy improved consistently, achieving 98.6%, showing that the model generalizes well to unseen data and maintains strong performance beyond the training dataset.



IX. CONCLUSION

Convolutional Neural Networks (CNNs) for the identification of medicinal plants provides a fast and precise method of classifying plant species using image recognition. This approach utilizes advanced deep learning methods that can capture subtle characteristics of plant images for more accurate identification than traditional approaches. These elements working in sync from data acquisition, preprocessing, model training, and eventual deployment create a seamless and automated workflow that simplifies and scales the unexpected challenges involved in classifying plants.

This methodology has notable ramifications in medicine, drug development, and plant research. The model can be useful for scientific studies, it may help to preserve traditional knowledge and may contribute to development of herbal medicine, because it gives a handy indication to formulary the medicinal plants.Future enhancements, such as expanding the dataset, optimizing model performance, and integrating user feedback, can further improve the system's effectiveness. Ultimately, CNN-based medicinal plant identification contributes to technological advancements in biodiversity conservation and healthcare application.

X. REFERENCES

[1] H. M. Al-Mansoori and M. A. Al-Hajri, "Deep Learning for Plant Species Identification: A Review", IEEE Access, vol. 8, pp. 217552-217567, 2020.

[2] L. Wang et al., "Plant Species Classification Using Transfer Learning and Ensemble Learning Techniques", *IEEE Access*, vol. 9, pp. 101132101141, 2021.

[3] S. Girinath, P. Neeraja, M. S. Kumar, S. Kalyani, B. L. Mamatha and N. R. T. GruhaLakshmi, "Real-Time Identification of Medicinal Plants Using Deep Learning Techniques", *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies*, pp. 1-5, 2024.

[4] S. Begum et al., "Prediction of herbs with its benefits using deep learning techniques", *Proc. Int. Conf. Sustain. Comput. Data Commun. Syst. (ICSCDS)*, 2022.

[5] K. Priya, S. Ranjana and R. Manimegala, "Medicinal Plant Identification using Android Application based on leaf image", *Eur. J. Mol. Clin. Med.*, vol. 07, no. 09, pp. 2515-8260, 2020.

[6] Boran Sekeroglu and Inan Yucel, "Leaves recognition system using a neural network", *Proceedings of 12th International Conference on*

Application of Fuzzy Systems and Soft Computing, pp. 578-582, 29–30 August, 2016.

[7] S. Shaheen, S. Ramzan, F. Khan, M. Ahmad, S. Shaheen, S. Ramzan, et al., "History classification worldwide distribution and significance of herbal plants", *Adulteration in Herbal Drugs: A Burning Issue*, pp. 35-49, 2019.

[8] L. D. Pacifico, L. F. Britto, E. G. Oliveira and T. B. Ludermir, "Automatic classification of medicinal plant species based on color and texture features", *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, pp. 741-746, 2019.

[9] A. K. Mulugeta, D. P. Sharma and A. H. Mesfin, "A deep learning for classification and recognition of medicinal plant species", *Sustainable and Intelligent Phytoprotection*, vol. 14, 2024.

[9] S. Girinath, P. Neeraja, M. S. Kumar, S. Kalyani, B. L. Mamatha and N. R. T. GruhaLakshmi, "Real-Time Identification of Medicinal Plants Using Deep Learning Techniques", *2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies*, pp. 1-5, 2024.

[10] J. Abdollahi, "Identification of medicinal plants in Ardabil using deep learning", *2019 27th International Computer Conference Computer Society of Iran (CSICC)*, pp. 9780493, 2022.

[11] D.V. Khoa, V.K. Hai, T.Q. Nguyen and V.T. Hoang, *Identification of medicinal plants using federated deep learning Procedia Computer Science*, 2024

[12] G. Mondal, B. Dasgupta and R. Sharma, "Disease of medicinal and aromatic plants and their management", pp. 251-283, 2018.

[13] A. Gokhale, S. Babar, S. Gawade and S. Jadhav, "Identification of Medicinal Plants Using Image Processing and Machine Learning", *Applied Computer Vision and Image Processing*, vol. 1, pp. 272-282, 2020.

[14] S. Sachar and A. Kumar, "Survey of feature extraction and classification techniques to identify plant through leaves", *Expert Syst. Appl.*, vol. 167, 2021.

[15] J. Wäldchen and P. Mäder, "Plant identification using deep neural networks via optimization of transfer learning parameters", *Journal of theoretical biology*, vol. 457, 2018.