International Journal of Scientific Research in Engineering and Management (IJSREM)
Special Edition - Volume 09 SPEJSS -2025 SJIF Rating: 8.586 ISSN: 2582-3930

MEDICINAL PLANT IDENTIFICATION USING CNN

SHILPA R.V, GURUPRASAD M.K, DILEEP KUMAR H.M

[1]*Assistant Professor, Computer Science Department, JSS College of Arts Commerce and Science Ooty-Road, Mysuru.*

[2]*U.G. Student, Computer Science Department, JSS College of Arts Commerce and Science Ooty-Road, Mysuru*

[3]*U.G. Student, Computer Science Department, JSS College of Arts Commerce and Science Ooty-Road, Mysuru*

Abstract: —Accurate identification of medicinal plants is essential for ensuring the efficacy and safety of herbal treatments, particularly in regions where traditional medicine plays a vital role in primary healthcare. This study presents a deep learning-based approach for the automatic identification of medicinal plant species using Convolutional Neural Networks (CNNs). The proposed system has the potential to assist botanists, pharmacists, and traditional medicine practitioners in reliable and efficient plant identification, thereby reducing the risks associated with misidentification and promoting the sustainable use of medicinal flora. Future work includes expanding the dataset, incorporating image data from various environmental conditions, and developing a mobile application for real-time field use.

Key Words: **— Medicinal Plants, Leaf, CNN, Identification, artificial intelligence, Deep learning (ML)**

I. INTRODUCTION

Medicinal plant identification is essential in many fields, such as pharmacology, traditional medicine, and biodiversity conservation, were accurate. Many scientific, medicinal, and ecological endeavors need the identification of plant species. Nonetheless, the extreme variety of plant species and their subtle visual traits provide significant challenges to manual identification techniques. These methods are labor intensive and prone to error because they rely on human. Observation and expertise, especially considering how subjective botanical knowledge is. In the past, obtaining precise identification frequently required in-depth training or the ad vice of specialized specialists, which prolonged the quick developments in deep learning and artificial intelligence. Through the application of (CNNs), among other computer vision techniques, researchers are now able to automate and improve the identification of medicinal plants with previously unheard-of levels of accuracy and efficiency. These sophisticated algorithms can recognize complex patterns in large databases of plant photos, analyze them, and learn to differentiate between different species according to their distinctive traits. By repeatedly training CNN models on carefully selected datasets, the system can improve its plant classification skills to the point where it can match or even exceed human specialists in terms of accuracy. This groundbreaking technology offers a systematic and objective way to identify medicinal plants while streamlining the identification process and reducing the subjectivity included in conventional methods. Furthermore, there is great potential for improving our knowledge of plant biodiversity, speeding up the process of finding new drugs, and encouraging sustainable herbal medicine practices through the incorporation of artificial intelligence into botanical research and conservation initiatives. Therefore, the combination of artificial intelligence (AI) and plant identification is a big step forward in using technology to improve human health and environmental stewardship.

II. LITRATURE SURVEY

The literature survey for this project involves a review of several research studies and techniques that have been used to identify Medicinal Plant with the help of Deep learning (ML). These studies highlight different approaches, tools, and levels of effectiveness in identification. Below is a summary of a few key works:

2.1 A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network

1. Authors: Forrest Sheng Bao; **Eric You-Xu; Yu-Xuan Wang; Yi-Fan Chang; Qiaolian Xiang Methodology**: In this paper, we employ probabilistic neural network (PNN) with image and data processing techniques to implement a general-purpose automated leaf recognition for plant classification. 12 leaf features are extracted and orthogonalized into 5 principal variables which consist of the input vector of the PNN. The PNN is trained by 1800 leaves to classify 32 kinds

of plants with an accuracy greater than 90Compared with other approaches, our algorithm is an accurate e artificial intelligence approach which is fast in execution and easy in implementation.

2.2 Vnplant-200–a public and large-scale of Vietnamese medicinal plant images dataset

2. **Author: Trung Nguyen Quoc and Vinh Truong Hoang**

Methodology: Plant identification is an essential topic in computer vision with various applications such as agronomy, preservation, environmental impact, discovery of natural and pharmaceutical Product. However, the standard and available dataset for medicinal plants have not been widely published for research community. This work contributes the first large, public and multi class dataset of medicinal plant images. Our dataset consists of total20,000 images of 200 different labeled Vietnamese medicinal plant (VNPlant200). We provide this dataset into two versions of size 256 × 256 and 512 × 512 pixels. The training set consists of 12,000 images and the remainder are used for testing set. We apply the Speed-Up Robust Features (SURF) and Scale Invariant Feature Transform (SIFT) for extracting features, and the Random Forest (FR)classifier is associated to recognize plant.

The experimental results on the VNPlant200 have been shown the interesting challenge task for pattern recognition.

2.3 Local binary pattern based on image gradient for bark image classification

3. Authors: Tuan Le-Viet and Vinh Truong Hoang

Methodology: In this work, we present a discriminative and effective local texture descriptor for bark image classification. The proposed descriptor is based on three factors, namely, pixel, magnitude and direction value. Unlike most other de scriptors based on original local binary pattern, the proposed descriptor is conducted the changing of local texture of bark image. The performance of the proposed descriptor is evaluated on three benchmark datasets. The experimental results show that our approach is highly effective.

2.4 Study on Identification and Classification of Medicinal Plants

4. Author: Rajani S and Veena

Methodology: This study offers information on the various preprocessing methods used to extract characteristics from a leaf image and the process of identifying medicinal plants based on those features. Among the many attributes removed from each leaf were its length, depth, boundaries, dimension, shade, rectilinear, and the concept of circularity. It is expected that a website-based or mobile computing platform for automatic recognition of therapeutic plants will contribute significantly to the development of pharmaceuticals, raise public awareness of these plants, and help taxonomic experts create more accurate methods to detect species .

2.5 An AI Based Approach for Medicinal Plant Identification Using Deep CNN Based on Global Average Pooling .

III. PROJECT IMPLEMENTATION/PROPOSED METHODOLOGY

Deep learning has revolutionized the ability of computer systems to learn from data and make predictions without the need for explicit programming. In this project, Medicinal plant identification is primarily based on Convolutional Neural Networks (CNNs). The architecture diagram below provides an overview of the system components and their interactions. It outlines the major steps involved in the process, which are:

- **Architecture Overview:** The architecture defines the flow of the process, starting with collecting the raw data and ultimately using it to predict Plants usage and properties. This step sets the foundation for the entire model development process.

- **Data Preprocessing:** The raw data collected is not always in a format suitable for deep learning models. Preprocessing transforms this data into a clean and understand able format, making it ready for analysis and model training.

- **Model Training:** Once the data is ready, we split it into two parts: a training dataset and a testing dataset. The training dataset is used to train the machine learning models. CNN the primary model are applied on the preprocessed image for the classification accuracy of each image is evaluated.

- **Testing the Models:** After the models have been trained, they are tested using the same algorithms

to assess how well they can predict Medicinal Plant's on new, unseen data.

- **Comparison of Results**: Finally, the results from the model are compared based on their classification ac curacy. This helps to determine the perfect accuracy predicition of the images used

METHODOLOGY

The proposed approach uses logistic regression models combined with other machine learning techniques to predict Medicinal Plant and its medicinal properties. The approach aimed to identify the ridge patterns, color and texture to accurately predict the plant name. The database of plants which contain unique bioactive compounds like alkaloids, flavonoids, tannins, and terpenoids are used to classify.

The data is preprocessed and analyzed to highlight important patterns. Along with
Convolutional Neural Networks (CNNs), Support Vector Machines (SVM), K-Nearest Neighbors (KNN) are trained using 75% of the dataset. These models aim to classify whether a plant or leaf sample belongs to a Medicinal Plant or not. The remaining 25% of the data is used for testing and evaluation. Model performance is assessed using metrics such as accuracy, precision, sensitivity, confusion matrix[22], and ROC-AUC score.

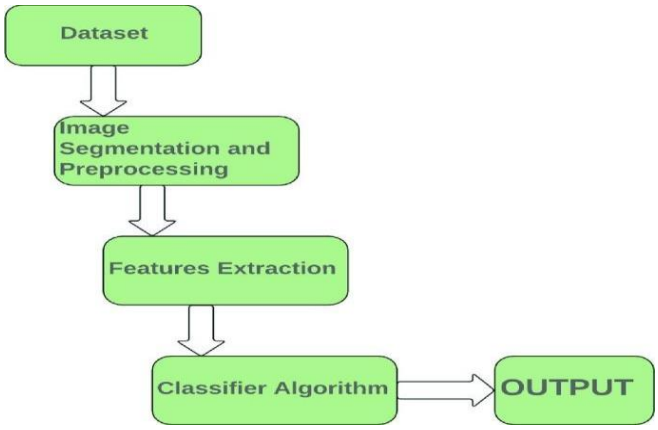

Fig. 1: Architecture diagram of the proposed system

A. System Architecture

The system architecture provides a clear and organized f low of how Medicinal Plant Identification is carried out using Deep learning. It starts from collecting relevant plant's data and continues through stages of cleaning, training, testing, and evaluating the results of different models.

This approach helps ensure that the models are well prepared to identify patterns associated with plants or leafs. By structuring the process into separate stages, we can better manage and analyze each step, leading to more accurate and reliable predictions.

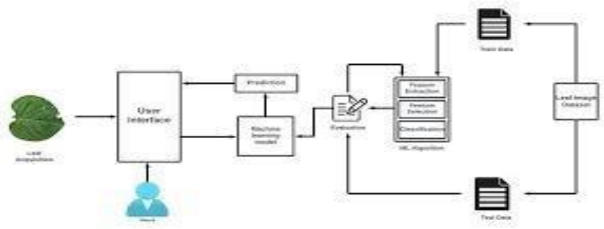

Fig .2: System Architecture Diagram

The architecture includes the following stages: –

Data Collection: Images of medicinal plants are cap tured using cameras or smartphones. Additional data such as leaf texture, shape, color, and environmental details are recorded. Historical and ethnobotanical information may be included to enrich identification.

Preprocessing: Raw data is cleaned, normalized, and transformed into a format suitable for machine learning algorithms.

Model Training: The processed data is used to train different algorithms, including CNN, Random Forest, and KNN.

Model Testing: The trained models are tested on a separate portion of the data to measure their performance.

Result Evaluation: The accuracy and other evaluation metrics are used to compare and analyze the results from each model.

B. Data Design

The dataset used in this study was sourced from Kaggle, containing images and spectral data of various medicinal plant species. Each row represents one of 195 plant samples, and each column corresponds to specific mor Pho logical and spectral features such as leaf texture, vein patterns, and color indices.

The dataset's main goal is to differentiate between different medicinal plant species, with each species assigned a unique label. Before training and testing the models, the raw data underwent preprocessing, which included cleaning, handling missing values, normalization, and feature extraction. The result is a

standardized dataset ready for ma chine learning applications.

V RESULTS

The Medicinal Plant Identification System utilizes advanced machine learning models to classify and recognize medicinal plants based on image and morphological data provided by the user. These models are pre-trained on ex tensive botanical datasets and integrated into the Streamlit framework, enabling real-time identification and classification. The system's performance is evaluated using key metrics such as accuracy, precision, recall, and F1 score, ensuring reliable identification of medicinal plants. Additionally, the identified plants are cross-referenced with established herbal medicine databases to validate their medicinal properties and potential applications.

Leaves Image	Name	Medicinal Uses
	Mentha (Mint)	• Easing queasy stomach. • Mask Bad Breath
	Ocimum Tenuiflorum (Tulsi)	• Cure fever, common cold, sore throat and kidney stones.
	Psidium Guajava (Guava)	• Used for stomach and intestinal condition, pain.
	Azadirachta Indica (Neem)	• Used for eye disorders, bloody nose, stomach upset and liver problem.

Fig. 3: Data set images sample

VII. DISCUSSION

Using machine learning (ML) algorithms for medicinal plant identification provides an efficient and accurate method for distinguishing plant species based on their morphological features. This is particularly valuable for herbal medicine, botanical research, and biodiversity conservation. In this project, we applied three ML algorithms—Convolutional Neural Networks (CNN), Random Forest, and Support Vector Machines (SVM)—to develop models that identify medicinal plants from images and feature data.

Among the models tested, the CNN-based approach demonstrated the best performance, achieving an accuracy of 97 in classifying medicinal plant species. This result aligns with CNN's ability to automatically extract hierarchical image features, making it particularly effective for plant identification tasks. The Random Forest model also performed well, leveraging ensemble learning to classify plants based on morphological characteristics like leaf shape and texture. Meanwhile, SVM proved useful for high-dimensional feature classification, helping refine the identification process, especially when differentiating visually similar species.

By integrating botanical databases and expert validation, the system ensures reliability in identifying medicinal plants and linking them to their therapeutic applications. The results highlight machine learning's potential in automating plant identification, supporting herbal research, and preserving traditional medicinal knowledge.

VII. COCLUSION AND FUTURE WORK

We have developed an effective approach to create an ac curate predictive model for Medicinal Plant Identification using CNN, Random Forest, and KNN classifiers. This method successfully classifies various medicinal plant species with an accuracy ranging from 90 to 95 percent.

Our in-depth study shows that leaf morphology and spectral characteristics contain enough information to reliably identify medicinal plants. In future research, exploring different feature selection or reduction techniques could help further enhance the classification accuracy.

FUTURE WORK

– In the future, these models can be trained with different datasets that include more features, which could potentially improve prediction accuracy.

– If the accuracy rate increases, these models could be used in Pharmaceutical Industry and Herbal Supplements Nutraceuticals to easily predict plant and its medicinal properties.

– These models could also be applied to different and Healthcare Medicinal field, expanding their utility.

– A future direction could be to extend the work by developing a hybrid model that can predict more than one Plant, using roots and using an accurate dataset that includes common features from both diseases.

VIII. REFERENCES

1) Alexis Joly, Herve Go ´ eau, Pierre Bonnet, Vera Baki ¨ c, Julien Barbe,Souheil Selmi, Itheri Yahiaoui, JenniferCarre, Elise Mouysset, Jean- ´ FrancoisMolino, Nozha Boujemaa, and Daniel Barthel´ emy. Interactive plant identification based on social image data. Ecological Informatics, 23:22–34, September 2014.

2) Fateme Mostajer Kheirkhah and Habibollah As ghari. Plant leaf classification using GIST texture features.IET Computer Vision, 13(4):369– 375, June 2019.

3) Tuan Le-Viet and Vinh Truong Hoang. Local binary pattern based on image gradient for bark image classification. InKezhi Mao and Xudong Jiang, ed itors,Tenth International Conference on SignalPro cessing Systems, page 39, Singapore, Singapore, April 2019. SPIE.

4) Shanwen Zhang, Chuanlei Zhang, Zhen Wang, and Weiwei Kong. Combining sparse representation and singular value decomposition for plant recognition Applied Soft Computing, 67:164–171, June 2018.

5) Shanwen Zhang, Chuanlei Zhang, and Wenzhun Huang. Integrating leaf and flower by local discrim inant CCA for plant species recognition. Computers and Electronics in Agriculture, 155:150–156, De cember 2018.

6) Tuan Le-Viet and Vinh Truong Hoang. Local binary pattern based on image gradient for bark image classification. In Kezhi Mao and Xudong Jiang, editors,Tenth International
Conference on Signal Processing Systems, page 39, Singapore,Singapore, April 2019. SPIE.