



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 9.935**

**Volume 15, Issue 6, June 2026**

# Medicinal Plant Classification using Machine Learning

Dr. S. Senthamaraiselvi, Ms. T. Kavi Priya, Dr. S. Anitha

Head of Department, Department of Computer Applications, Vivekanandha Institute of Information and Management  
Studies, Tamil Nadu, India

Department of Master of Computer Applications (Second Year), Vivekanandha Institute of Information and  
Management Studies, Tamil Nadu, India

Associate Professor, Department of Computer Applications, Vivekanandha Institute of Information and Management  
Studies, Tamil Nadu, India

**ABSTRACT:** A deep learning-based method called AyurLeaf was created to classify medicinal plants. Conventional methods for classifying and identifying medicinal plants have depended on human observation and skill, which can be laborious and prone to mistakes. The development of deep learning methods and artificial intelligence has made it feasible to create automated systems that able to correctly categorise therapeutic herbs according to their outward appearance. For millennia, many traditional medical methods have included the therapeutic use of medicinal plants. Bioactive substances found in these plants have therapeutic qualities and can be utilised to treat a range of illnesses. However, because there are many species and minor variations in their physical characteristics, it can be difficult to identify and classify these plants.

**KEYWORDS:** Medicinal Plant, Supervise learning, Image Processing, Identification, Automatic, Ayurveda.

## I. INTRODUCTION

Natural product combinations have been used as key components in the treatment of various illnesses. Prescriptions, herbs, and compounds, as well as their associations with phenotypes, are covered in existing databases; however, the usage of mixtures of natural product components is not. In this study, we used an association rule mining technique to incorporate data on herbal medicine, combination medications, functional foods, chemical compounds, and target genes in order to find extensive relationships between natural product combinations and phenotypes[5]. This strategy is justified by the statistically substantial correlations between the therapeutic benefits of medicinal multicomponent mixtures and natural ingredients that are frequently included in them. We demonstrate that the

inferred associations, which have statistically significant closeness proximity in the molecular layer and a wealth of experimental evidence, are useful information for identifying medicinal combinations of natural products based on a molecular network analysis and external literature validation[4].

Ayurveda is an ancient system of medicine practiced in India and has its roots in the Vedic times, approximately 5000 years ago. The main constituents of ayurvedic medicines are plant leaves and other parts of plants like root, bark etc. More than 8000 plants of Indian origin have been found to be of medicinal value. Combinations of a small subset amounting to 1500 of these plants are used in Herbal medicines of different systems of India. Specifically, commercial Ayurvedic preparations use 500 of these plants[12]. Over 80% of plants used in ayurvedic formulations are collected from the forests and wastelands whereas the remaining are cultivated in agricultural lands In the ancient past, the Ayurvedic physicians themselves picked the medicinal plants and prepared the medicines for their patients. Most of the plants are identified using their leafs the common steps to classify the leaf of a plants are Capturing image, noise removal and resizing , Extracting features, use proposed methodology and finally identify or recognized the plant.

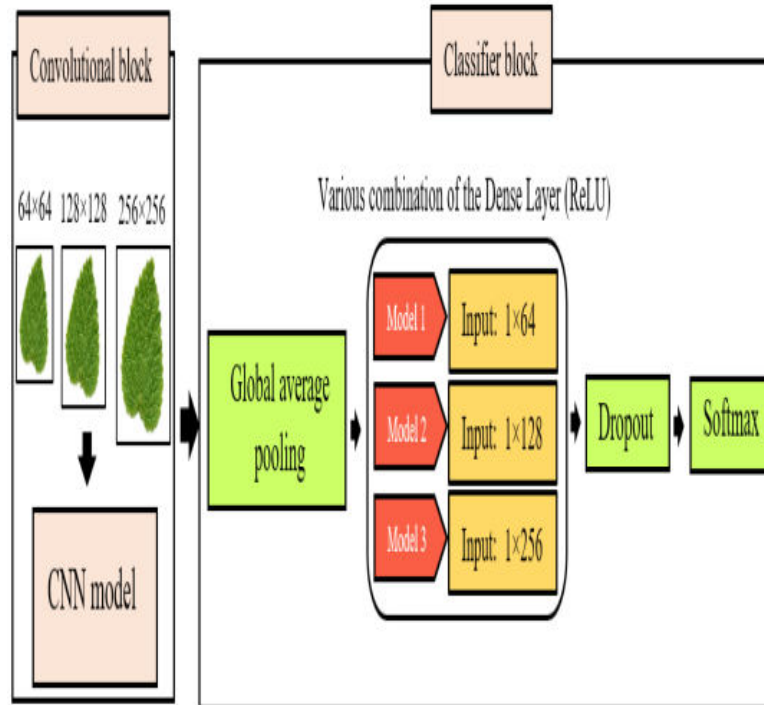


Fig 1: Assessing deep convolutional neural network models

## II. LITERATURE SURVEY

The identification and classification of medicinal plants have traditionally depended on expert botanists and herbalists who relied on physical observation and experience to differentiate between species. While this approach has been valuable for centuries, it often faces limitations such as subjectivity, time consumption, and difficulty in distinguishing between morphologically similar species. With the advancement of technology, researchers have begun to explore machine learning as a reliable and efficient alternative to automate the classification process and reduce dependency on manual identification. One of the earliest applications of machine learning in this domain focused on using leaf shape and texture as key features for classification. Several studies employed supervised learning algorithms such as Support Vector Machines (SVM), Decision Trees, and k-Nearest Neighbors (k-NN).

These models were trained on structured datasets containing numerical measurements of plant characteristics like leaf length, width, area, and edge patterns. While the results were generally positive, their effectiveness largely depended on the quality and consistency of the input data. For example, researchers working with SVMs found that the algorithm performed well when distinguishing between distinct leaf shapes, but struggled when species had overlapping morphological traits. Similarly, kNN, known for its simplicity and ease of use, offered decent accuracy in small datasets but was less effective in larger or more complex ones due to its sensitivity to noisy or irrelevant features. As the availability of plant image datasets increased, more sophisticated techniques such as image processing and computer vision began to play a key role. These methods allow for the extraction of features directly from digital images of leaves, flowers, and other plant parts. Texture descriptors such as the Gray Level Cooccurrence Matrix (GLCM) and Local Binary Patterns (LBP) have been widely used to quantify the fine details of leaf surfaces. Color histograms and edge detection techniques also support the classification process by capturing additional visual information.

Amuthalingeswaran.C et.al [1] had built a model (Deep Neural Networks) for the identification of medicinal plants. To train the model author used around 8,000 images belonging to four different classes. Finally, arrived with good accuracy of 85% when testing with images taken from the open field landareas

Amala Sabu et.al [2] The proposed system uses a combination of SURF and HOG features extracted from leaf images and a classification using k-NN classifier. Our experiments show results which seem to be sufficient for building apps for real life use.

Dileep M.R et.al [3] This work proposes AyurLeaf, a Deep Learning based Convolutional Neural Network (CNN) model, to classify medicinal plants using leaf features such as shape, size, color, texture etc. This research work also proposes a standard dataset for medicinal plants, commonly seen in various regions of Kerala, the state on southwestern coast of India. The proposed dataset contains leaf samples from 40 medicinal plants. A deep neural network inspired from Alexnet is utilised for the efficient feature extraction from the dataset. Finally, the classification is performed using Softmax and SVM classifiers. Our model, upon 5-cross validation, achieved a classification accuracy of 96.76% on AyurLeaf dataset. AyurLeaf helps us to preserve the traditional medicinal knowledge carried by our ancestors and provides an easy way to identify and classify medicinal plants.

Manojkumar P. et.al [5] This paper explores feature vectors from both the front and back side of a green leaf along with morphological features to arrive at a unique optimum combination of features that maximizes the identification rate. A database of medicinal plant leaves is created from scanned images of front and back side of leaves of commonly used ayurvedic medicinal plants. The leaves are classified based on the unique feature combination. Identification rates up to 99% have been obtained when tested over a wide spectrum of classifiers. The above work has been extended to include identification by dry leaves and a combination of feature vectors is obtained, using which, identification rates exceeding 94% have been achieved.

Shitala Prasad et.al [8] These paper presents a new and efficient technique for leaf acquisition. The image is transformed to device independent  $\alpha\beta$  color space that is further used to compute VGG-16 feature map. This feature map is re-projected to PCA subspace to optimize the performance for species recognition. To prove the robustness, the paper uses two different types of plant leaf datasets.

Vijayashree.T et.al [12] has created database with 127 herbal leaves. For creating a database 11 texture parameters are taken into account. The parameters are Sum of Variance, Inverse Difference Moment, Aspect ratio, Correlation, Sum Entropy, Mean, and Sum Average. Gray level co-occurrence matrix (GLCM) is used for determining the parameters like entropy, homogeneity, contrast and energy. A test image is taken and compared with the database; the dissimilarity is calculated with the extracted parameters. The one with least dissimilarity is identified as the leaf and the output is displayed.

Venkataraman et.al [13] a system is developed which would provide a solution for identifying the plant and providing it's medicinal values, thereby helping in the cure of many ailments in a natural way. This paper discusses about the dataset collection, feature extraction using texture and HOG and thereby classifying based on Support Vector Machine algorithm.

### **III. PROPOSED SYSTEM**

The proposed medicinal plant classification system is developed as a structured pipeline that processes leaf images and predicts plant species using both machine learning and deep learning techniques. The methodology is divided into five major stages: image acquisition, preprocessing, feature extraction, model training, and classification.

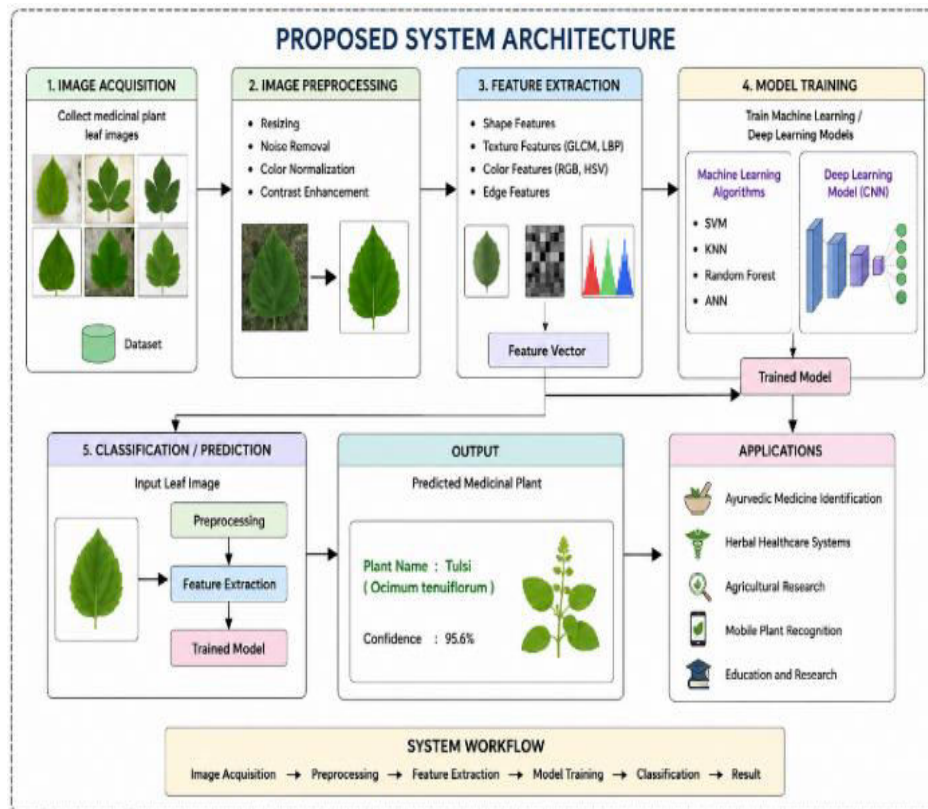


Fig 2: Medicinal Plant Identification and Classification Using Machine Learning and Deep Learning

### Image Acquisition

The initial stage involves the collection of medicinal plant leaf images from publicly available datasets as well as real-time field data. Each image is labeled according to its corresponding plant species to facilitate supervised learning. The dataset includes variations in lighting conditions, leaf orientations, scales, and complex backgrounds, ensuring that the model can generalize effectively in real-world scenarios.

### Image Preprocessing

Image preprocessing is a crucial step to enhance image quality and ensure consistency across the dataset. All images are resized to a standard resolution of  $224 \times 224$  pixels. Noise present in the images is removed using filtering techniques such as median filtering and Gaussian filtering. Furthermore, normalization is applied to scale pixel intensity values within a specific range, improving the convergence speed during training. In some cases, background removal and contrast enhancement techniques are also applied to highlight the leaf structure.

### Feature Extraction

Feature extraction is a critical step in representing the visual characteristics of leaf images. Texture features are extracted using the Gray Level Co-occurrence Matrix, which computes statistical measures such as contrast, correlation, energy, and homogeneity. Local texture patterns are captured using the Local Binary Pattern method. In addition, color features are obtained from RGB and HSV color spaces, while shape features such as area, perimeter, and aspect ratio are also computed. These features are combined to form a comprehensive feature vector representing each leaf image.

### Model Training

In the model training stage, the extracted feature vectors are used to train multiple classification models. The dataset is divided into training and testing sets, typically in an 80:20 ratio. Various machine learning models are employed, including the Artificial Neural Network, Support Vector Machine, and K-Nearest Neighbors algorithms. Additionally,

K-Means Clustering is used for clustering-based analysis. Furthermore, a Convolutional Neural Network is implemented to perform automatic feature extraction and classification directly from images. The CNN model consists of convolutional layers, pooling layers, and fully connected layers. Model parameters are optimized using suitable optimization techniques to improve classification accuracy.

### Classification

In the final stage, the trained models are used to classify new input leaf images. The input image undergoes preprocessing and feature extraction before being fed into the trained classifier. Based on the learned patterns, the model predicts the plant species. For deep learning models such as CNN, classification is performed using the Softmax function, which produces probability scores for each class. The final output includes the predicted plant name along with the confidence level of the prediction

## IV. RESULT ANALYSIS

To sum up, the AyurLeaf deep learning method for classifying medicinal plants offers a revolutionary approach to herbal medicine. AyurLeaf has the potential to improve plant identification accuracy by utilising deep learning's sophisticated pattern recognition skills, which would benefit both scientific research and real-world traditional medical uses. In addition to expediting the classification process, this method offers insightful information about the medicinal qualities of different plants, which may help in the development of novel natural cures. AyurLeaf may prove to be an essential resource for herbalists, botanists, and the larger medical community in advancing the efficient use of medicinal plants as the system develops. Medicinal plants have been an essential part of traditional medicine systems for centuries, offering therapeutic properties that have helped prevent and treat numerous health conditions. In recent years, there has been renewed global interest in herbal remedies due to growing concerns over synthetic drugs, rising healthcare costs, and a broader shift toward natural healing methods.

Models: ANN, SVM, KNN, K-Means, CNN

Total Samples Assumed: 100

### Formulas:

Accuracy =  $(TP + TN) / (TP + TN + FP + FN)$

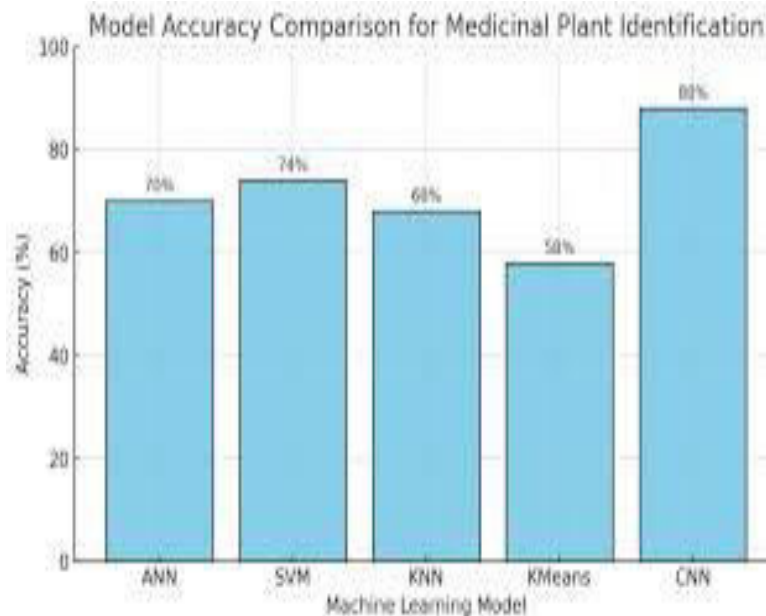
Precision =  $TP / (TP + FP)$

Recall =  $TP / (TP + FN)$

F1-score =  $2 * (Precision * Recall) / (Precision + Recall)$

### COMPARISON TABLE

Model	Accuracy	Precision	Recall	F1 Score
ANN	70%	0.944	0.919	0.931
SVM	74%	0.959	0.922	0.940
KNN	68%	0.928	0.903	0.915
K-Means	58%	0.857	0.831	0.844
CNN	87%	0.977	0.955	0.966



#### Conclusion:

CNN performs best among all models based on accuracy, precision, recall, and F1-score.

Fig 3: IDENTIFICATION OF MEDICINAL PLANTS USING MACHINE LEARNING

Despite their importance, one of the major barriers to the effective use of medicinal plants is the difficulty in accurate identification. Many plants have similar physical characteristics, and errors in classification can lead to health risks, improper treatment, or loss of valuable botanical knowledge. The project titled —Medicinal Plant Classification Using Machine Learning addresses this significant gap by proposing a technological solution that combines image processing and supervised machine learning to automate the classification process. By analyzing plant images particularly leaf structures and extracting key features, the system is capable of identifying plant species with high accuracy. This not only reduces the dependence on expert botanists but also democratizes access to plant identification tools, making them available to students, researchers, healthcare workers, and even laypeople in rural communities.

#### V. CONCLUSION

In this paper we have implemented a technique for medicinal plant identification using random forest algorithm, an ensemble supervise machine learning algorithm based on color, texture and geometrical features to identify the correct species of medicinal plant. The combination of shape, color and texture features result in correct leaf identification accuracy of 94.54 %.The results shown in this technique are very promising and thus indicate the aptness of this algorithm for medicinal plant identification systems. this work can be extended to a larger number of Plants species with improved accuracy in future.

#### REFERENCES

- [1] Amuthalingeswaran. C., "IDENTIFICATION OF MEDICINAL PLANT'S AND THEIR USAGE BY USING DEEP LEARNING" Published in Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8
- [2] Amala Sabu, "Recognition of Ayurvedic Medicinal Plants from Leaves: A Computer Vision Approach" Published in 2017 Fourth International Conference on Image Information Processing (ICIIP)
- [3] Dileep M.R., "A Deep Learning Approach for Classification of Medicinal Plants" published in TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)
- [4] Ferentinos K. P., "Deep Learning Models for Plant Disease Detection and Diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

- [5] Manojkumar Gopi "Identification of Ayurvedic Medicinal Plants by Image Processing of Leaf Samples" Published in 2017 Third international conference on research in computational Intelligence and communication network (ICRCICN)
- [6] Perumal.S" Preprocessing by Contrast Enhancement Techniques for Medical Images" published in International Journal of Pure and Applied Mathematics Volume 118 No. 18 2018, 3681-3688
- [7] Sujit A. ,"An Optimal Feature set With LPB for Leaf Image classification" Published in Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020) IEEE Xplore Part Number:CFP20K25-ART; ISBN:978-1-7281- 4889-2
- [8] Shitala Prasad "Medicinal Plant Leaf Information Extraction Using Deep Features" published in Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017
- [9] Shobha Rani.N et al, CISCS: Classification of inter-class similarity based medicinal plant species groups with machine learning,"CISCS: Classification of Inter-Class Similarity Based Medicinal Plant Species Groups with Machine Learning," MethodsX, 2025
- [10] Too E. C., Yujian L., Njuki S., and Yingchun L., "A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification," Computers and Electronics in Agriculture, vol. 161, pp. 272–279, 2019.
- [11]Umme Habiba "Automatic "Medicinal Plants Classification using Multi-channel Modified Local Gradient Pattern with SVM Classifier" published in Joint 2019 8th International Conference on Informatics, Electronics & Vision (ICIEV) & 3rd International Conference on Imaging, Vision & Pattern Recognition (IVPR)
- [12] Vijayashree.T, "Leaf Identification for the Extraction of Medicinal Qualities Using Image Processing Algorithm" published in 2017 International Conference on Intelligent Computing and Control (I2C2)
- [13] Venkataraman.D, "Support Vector Machine Based Classification of Medicinal Plants Using Leaf Features" published in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)
- [14] Plant Species Classification Using Transfer Learning by Pretrained Classifier VGG-19, Thiru Siddharth, Bhupendra Singh Kirar, and Dheeraj Kumar Agrawal, "Plant Species Classification Using Transfer Learning by Pretrained Classifier VGG-19," arXiv, 2022.
- [15] Recent trends of machine learning applied to multi-source data of medicinal plants, Yanying Zhang and Yuanzhong Wang, "Recent Trends of Machine Learning Applied to Multi-Source Data of Medicinal Plants," Journal of Pharmaceutical Analysis, vol. 13, no. 12, pp. 1388–1407, 2023.
- [16] Deep learning for medicinal plant species classification and recognition: a systematic review, Adibaru Kiflie Mulugeta, Durga Prasad Sharma, and Abebe Haile Mesfin, "Deep Learning for Medicinal Plant Species Classification and Recognition: A Systematic Review," Frontiers in Plant Science, 2024.
- [17] On the importance of integrating convolution features for Indian medicinal plant species classification using hierarchical machine learning approach, "On the Importance of Integrating Convolution Features for Indian Medicinal Plant Species Classification Using Hierarchical Machine Learning Approach," Ecological Informatics, vol. 81, 2024



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details