International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 2s, 2025

Medicinal Herbs Identification in the wild by using CNN

Asad Hussain Faizan
ISL Engineering College,
Hyderabad, India

Mohd. Altaf Pasha
ISL Engineering College,
Hyderabad, India

Mohammed Waseem
ISL Engineering College,
Hyderabad, India

Dr. Mohammed Abdul Bari
Prof. CSE, ISL Engineering
College, Hyderabad, India.
ahfaizan04@gmail.com

***Abstract*—Many rural communities have a strong belief in plant diversity. They collect useful plants and herbs and use them with indigenous knowledge and customs. One such oldest system that results in the use of medicinal herbs is Ayurveda. Approximately 10,000 plants are used medicinally in India, but not all plants are included in the official Ayurvedic Pharmacopoeia. Before becoming part of Ayurvedic medicine, all plants need to be thoroughly studied. For this reason, identifying herbs is the most important step. Many of these identifications are fully supported by human perception, leaving room for error and misjudgement. Therefore, it is necessary to develop an efficient system using computer vision, pattern recognition, and image processing algorithms alongside the availability of various combinations of feature detection methods with different classifiers that are often utilized in building an automatic identification system for herbal leaves using leaf images and reveal its associated information to realize knowledge.**

Keywords—Classification, CNN, ANN

I. Introduction

Plants are always very important on earth. There are thousands of plant species on earth, many of which are medicinal, other plants are endangered, and others are harmful to humans. Plants are not only an essential resource for humans, but also serve as a muse for the entire food chain. Species knowledge is needed in different areas of work by different social groups such as forest managers, farmers, environmentalists and educators.

Therefore, species identification is an interdisciplinary field of study. In order to use and preserve plants, it is important to properly investigate and classify them. The identification of unknown plants relies heavily on the unique knowledge of experienced botanists. On the other hand, this can be a tedious and difficult task for the average person who is not familiar with plant terminology. Manual methods that support morphological features are successful methods for correctly identifying plants. As a result, many of the methods used to classify these plant species rely on human knowledge and skills. However, manual detection is often tedious and time consuming. On the other hand, machine learning and computer vision improvements are also available to make this process a little easier. As a result, a huge amount of research has been done to support automatic classification based on the physical properties of plants. However, there is not yet a system that can identify all plant species, whether they are endangered or not, record their efficacy, and provide prescriptions for the drug. The phases performed in this study are leaf collection, pre-treatment to detect unique properties, leaf classification, database construction, training for identification, and finally interpretation of the results. This project used data from a collection of leaves dataset containing 1822 images from 30 different species. Leaves are the most common way to identify plants, bud stems, flowers, petals, seeds, and even whole plants are used in automated processes. Automatic plant identification systems can be used by non-experts to easily and quickly identify plant species.

II. Literature Survey

A common issue in image processing may be the identification of plants using photographs of their leaves. For the goal of identifying leaves, those image processing algorithms often use shape-based digital morphological features. Even though there are many research on identifying plants based on their leaves, very few of them are for mobile devices. [1].

In the ayurvedic pharmaceutical industry, choosing the appropriate medicinal plants from which to make a medicine is crucial. A database is created from scanned images of front and back side of leaves. The classification of leaves is based on a special feature combination. When tested across a broad spectrum of classifiers, identification rates of up to 99% are obtained [2].

The study comprises of a system that uses image processing techniques to extract relevant information associated with leaves in conjunction with an artificial neural network to detect. Real samples of twelve different herbal medicine plant leaves are collected where each leaf is isolated in single image. Several features are extracted using techniques in image processing. With the synthetic neural network acting as autonomous brain network, the system can identify the species of the herbal medicine plant leaves being tested. The system can even provide more information about the leaves that helps to cure diseases [3].

ISSN 2347–3657

Volume 13, Issue 2s, 2025

In order to identify new or rare plant species and improve the ecosystem's balance, it is essential to have an honest understanding of plants. Plant identification refers to the process of matching a specimen plant to a recognised taxon, which implies classifying a chosen plant according to a known taxonomic category by comparing certain traits. This essay offers a variety of approaches from numerous authors who have investigated various methods for identifying plants. [4].

III. WORKING PRINCIPLE

The proposed system uses ML and DL techniques for the analysis phase [5]. Colour, shape and sizes are the features of recognising leaves. The steps we partitioned are enhancement of images, segmentation, attribute extraction and classification [6]. In this study, we have used Artificial Neural Network and Convolutional Neural Network for image processing segmentation technique and created two data Models for segmenting leaves. For feature extraction, we used RGB to grayscale conversion, smoothing with Gaussian filter, adaptive image threshold processing by Otsu's method, hole closure by morphological conversion, and finally boundary extraction by contour [7]. The flower photo has been cropped to remind the proper size. After training and testing using the CNN and ANN algorithms, the system classifies plant species and returns their species name, medical description, habitat with some prescriptions, making this method different from other existing systems [8].

A. Tools And Software Used

- Jupyter notebook
- Pycharm
- Python
- Tensorflow
- Keras
- Numpy
- Flask framework

IV. MODULES

A. MODULE 1: Image Capture

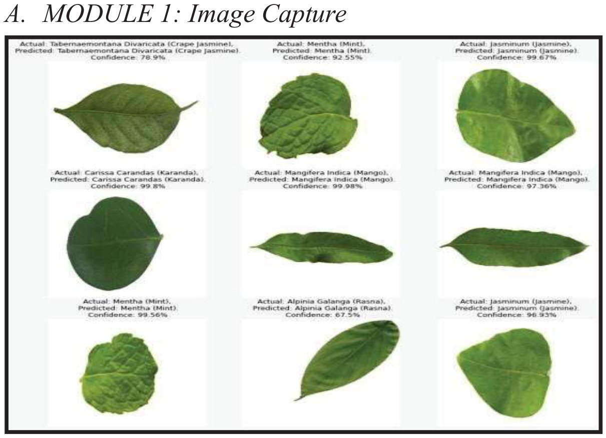

Fig. 1. Plant leaves database

The initial phase in the recognition method is to collect images of the plants. The image is first retrieved from an external source using this method, and it is then used for such future operations. The dataset contains 1822 images of 30 different plant species with each image labels [9]. Plant leaves database is shown in Fig. 1.

1) Image Pre-Processing: Pre-processing an image can be a crucial step because it helps to enhance the image quality for subsequent processing. This phase is very important because the image is inherently noisy. This can reduce the accuracy of the classification. This is done to avoid noise that interferes with the identification process and to handle degraded data in the system.

- RGB to grayscale conversion
- Smoothing with a Gaussian filter
- Adaptive image binarization using Otsu's binarization method
- Hole closure by morphological transformation
- Boundary extraction by contour.

B. MODULE 2: Image Segmentation

The purpose of image segmentation is to distinguish between foreground and background data. The ROI is obtained and used as a unique feature. Segmentation is a time-consuming, multi-step process. The process of defining the spatial area of an image is called "segmentation" [10]. First, in order to extract the shape features from the photo, we need to create a binary image that contains only the edges of the leaves. The image is then converted to grayscale and then binarized using Otsu's method. A rectangular smoothing filter from the 3x3 kernel is used to minimize noise. Leaf boundaries parameters is shown in Fig. 2. To find the leaf boundaries, convolve the image using the 3x3 kernel with the following parameters:

0	1	0
1	-4	1
0	1	0

Fig. 2. Leaf boundaries parameters

The final image of the shape feature extraction contains the 1-pixel width boundary of the leaf image.

1) Feature Extraction: We can classify leaf photos by color, shape, texture, or a combination of these elements. Take advantage of all three acceptable qualities with a photo of leaves with a plain background. These features are further analyzed before being used for classification. Shape features, also called morphological features, usually refer to the shape of the leaves. It is assumed that features of various shapes will be extracted for additional processing, leaving the best features for the final feature set.

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

a) Diameter: The substantial interspace between any two points on the sheet edge of the sheet is the diameter of the sheet.

b) Physiological Length: Physiological length is defined as the distance between the endpoints of two leaves. Humans identify these ends by clicking on the main vein at the end of the petiole and the tip of the end of the leaf.

c) Leaf Area: Counting the number of pixels contained in a smoothed leaf image is an easy way to calculate leaf area.

d) Leaf Perimeter: Sheet circumference can be calculated simply by enumerating the number of PPI that make up the border of the sheet.

e) Burnish Factor: The burnish factor is used to assess the impact of noise on the image area. The binary leaf image is smoothed with a 5*5 smoothing filter and a 3*3 smoothing filter, and the smooth factor is the difference between the areas of both processed images.

f) Aspect Ratio: This is called the ratio of corporeal length to the corporeal width, and the difference between the circle with the diameter of the blade and the blade itself is represented by the form factor.

g) Rectangularity: The difference between a leaf and a rectangle of the same dimensions is described by rectangularity. It's computed using the formula L*W/A, where A is the leaf's area, L is the leaf's physiological length, and W is the leaf's physiological breadth.

h) Width: This factor also accounts for the leaf's spread by taking into consideration both physiological length and width.

i) Vein Features: Vein features are generated using morphological techniques in grayscale images containing disc-shaped elements with radii 1, 2, 3, and 4. Edges are removed from the image produced by the morphological process to preserve the venous structure of leaf. Feature set data is shown in Fig. 3.

	area	perimeter	physiological_length	physiological_width	aspect_ratio	rectangularity	circularity	mean_r	mean_g	mean_b	stddev_r	stddev_g	stddev_b	contrast
0	197484.0	3479.036038	1416	759	1.865613	5.442183	61.289480	6.395667	13.643413	4.388007	24.025329	40.200931	21.448412	12.639656
1	101248.0	2490.381812	1190	130	9.153846	1.527931	61.255546	7.049316	9.232018	10.876066	33.816205	37.382221	46.479230	8.137424
2	86570.5	2290.683327	1095	119	9.201681	1.505189	60.612219	3.434303	6.371511	2.644757	19.975699	29.057374	19.275050	8.553729
3	190214.0	2856.479353	1318	254	5.188976	1.759976	42.896287	7.670415	13.303599	6.049157	28.822885	40.221845	26.948604	8.440041
4	227727.0	2917.248904	1324	286	4.629371	1.662798	37.370804	8.992028	16.671171	6.294281	30.967158	45.040192	28.595326	8.641438

Fig. 3. Feature set data

2) Feature Normalization: Because features are extracted using multiple procedures and their values (ranges) can vary greatly, normalization of features is required before dimensionality reduction or classification. In the absence of a normalization phase, the features with higher values will have a stronger influence on the classifier, and the classification results will be dominated by them. Each feature's maximum and lowest values will be checked, and its value will be normalized to create a uniform range of values.

C. MODULE 3: Classification

The work's ultimate stage is categorization, which defines the entire model. This section discusses various classification techniques that can be used to categorize plant species using extracted features. In this situation, we used the common precision linear classifiers CNN and ANN.

1) Testing & Training : Train/Test is a technique for determining the model's accuracy. A workout data and a testing data were created from the data set. Training accounts for 80% of the budget, while testing accounts for 20%. We used CNN and ANN Machine Learning Algorithms to practice the model. Then we use the testing set to put the model to the test.

There are 1822 pictures in the Flavia dataset, representing 30 different plant species. The classifier was trained using these photos. We choose 5 species of leaves from testing sets for each type of plant in the flavia dataset, which were then utilized to evaluate the proposed algorithm's accuracy and execution time. Training and Testing analysis graph is shown in Fig. 4.

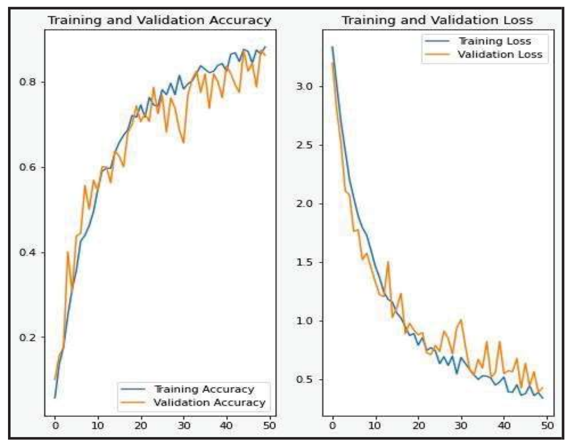

Fig. 4. Training and Testing analysis graph

V. CONCLUSION

This study proposes an approach that uses machine learning techniques to automatically identify plant species that classify plant leaf photographs. The study was divided into four phases: image preprocessing, image segmentation, feature extraction, and image classification. This design combines texture and color. Plant Leaf Species Identification Supported by Machine Learning Algorithms This effort presented a reliable method for identifying anomalous medicinal plants using CNN. Using TensorFlow with the dataset I created, I was able to achieve a test accuracy of 93.02%. To maintain this immunity, important properties of leaf images have been extracted. Due to the popularity of statistical models, the samples were created using neural networks. The number of training sessions or epochs will improve accuracy. Higher accuracy can be achieved by increasing the number of epochs.

VI. RESULT

In this study, it is robust and computationally efficient approach is described that takes into account the leaf's colour and teeth properties, as well as its form aspects.

Finally, we used a combination of colour, form, and morphological characteristics. The system was evaluated using two classifiers on the Flavia dataset, and the results were acceptable, as shown in the experimental results. The suggested work is expanded to include spot complex images with petiole and clustered leaves, as well as real-time leaf images.

VII. Future Scope

We were able to improve segmentation using thresholding, region expanding, and clustering-based approaches. To the feature vector, better form features and some texture characteristics have been added. Weighted normalization is a technique that involves assessing the retrieval accuracy of individual features and applying additional weight to them in the feature vector. Other classifiers can be evaluated, or a binary tree-based SVM can be used. Select the top five classifier o utputs and weight them suitably to generate the confidence-based class label to reduce inaccurate classification.

Trace the outline of the leaf on the digital screen. Example: A camera can be used to detect it. The trace drawn with the user's finger to trace the image of the leaves is linked to a preset algorithm, much like the Swype keyboard works on mobile phones. The paths are mapped, the equivalent paths are extracted from the dataset, and the leaves are recognized after the finger is lifted off the screen. In addition, a sheet of similar shapes and path maps is provided to avoid mistakes. There is a debate about input differences as a result of user shifts. However, the uniqueness of digital fingerprints and the fixed default method (using Python) will almost certainly stabilize problems for different users.

We can use digital fingerprints to identify the leaves. This approach works like a media detection app. We can mark positions at various depths on the sheet and combine them with images that can be plotted against the chart using laser scanning. The plotted areas form a unique digital fingerprint to identify the sheet.

References

[1] "Computer assisted plant identification system for Android," 2015 Moratuwa Engineering Research Conference (MERCon), Moratuwa, 2015, pp. 148-153.

[2] P. M. Kumar, C. M. Surya and V. P. Gopi, "Identifi cation of ayurvedic medicinal plants by image processing of leaf samples," 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, 2017, pp. 231-238.

[3] R. G. de Luna et al., "Identification of philippine herbal medicine plant leaf using artificial neural network," 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Manila, 2017, pp

[4] Rafael Rojas-HernÃ¡ndez and AsdrÃºbal LÃ³pez-Chau, "Plant identification using new geometric features with standard data mining methods", Networking Sensing and Control (ICNSC) 2016 IEEE 13th International Conference on, pp. 1-4, 2016.

[5] Wang, X., D. Huang, J. Du, H. Xu, & L. Heutte, 2008. Classification of plant leaf images with complicated background.

[6] N. Deepak and R. Ahmed, "Neural networks based leaf identification using shape and structural decomposition - IEEE Conference Publication", Ieeexplore.ieee.org, 2016.

[7] Backes, A.R., D. Casanova, & O.M. Bruno, 2009. Plant leaf identification based on volumetric fractal dimension. International Journal of Pattern Recognition and Artificial Intelligence, 23(6): 1145-1160.

[8] Edi, L., A. Susanto, Kadir, A. Nugroho, & P.I. Santosa, 2011. Leaf Classification Using Shape , Color , and Texture Features. International Journal of Computer Trends and Technology, pp: 225-230.

[9] Gao, T., H. Yao, J. Song, C. Liu, Y. Zhu, X. Ma, X. Pang, H.C.S. Xu, 2010. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. Journal of Ethno pharmacology.

[10] Kadir, A., L.E. Nugroho, A. Susanto, & P.I. Santosa, 2011. Foliage Plant Retrieval Using Polar Fourier Transform, Color Moments And Vein Features. International Journal of Signal & Image Processing, 2(3): 1-13.

[11] Nausheen Fathima, Dr. Mohd Abdul Bari , Dr. Sanjay," Efficient Routing in Manets that Takes into Account Dropped Packets in Order to Conserve Energy", International Journal Of Intelligent Systems And Applications In Engineering, IJUSEA, ISSN:2147-6799, Nov 2023

[12] Afsha Nishat, Dr. Mohd Abdul Bari, Dr. Guddi Singh," Mobile Ad Hoc Network Reactive Routing Protocol to Mitigate Misbehavior Node", International Journal Of Intelligent Systems And Applications In Engineering, IJUSEA, ISSN:2147-6799, Nov 2023

[13] Ijteba Sultana, Dr. Mohd Abdul Bari ,Dr. Sanjay," Routing Performance Analysis of Infrastructure-less Wireless Networks with Intermediate Bottleneck Nodes", International Journal of Intelligent Systems and Applications in Engineering, ISSN no: 2147-6799 IJISAE,Vol 12 issue 3, 2024, Nov 2023.