Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 2 (2024)

Identification of Medicinal Plants by Visual Characteristics of Leaves

[1]Satyam Sangwan, [2]Kanupriya,[3]Amarpreet Kaur,[4]Vani Mittal,[5]Vijay Kumar
[1,2,3,4,5]*Department. of Computer Science Engineering Meerut Institute of Engineering and Technology, Meerut, Uttar Pradesh, India*

***Abstract*—**A vital ability for foresters and botanists is the ability to identify plant species from photos of leaves, bark, and needles. Plants are identified by human professionals using scent and visual traits. If therapeutic plants are incorrectly identified, there might be detrimental effects. Using visible morphological characteristics, such as leaf shape, color, and texture, plant identification may be done automatically. This paper demonstrates how image processing and machine learning techniques were used to accurately find rare medicinal plants. A unique featurecombination is used to classify the leaves. Validation rates ofup to 95% were obtained while testing more than 20 plants.

***Keywords*—**Medicinal Plant identification, Feature extraction, Morphological features, Classification,Indigenous medicine

I. Introduction

Currently, plants for therapeutic uses are primarily obtained from wooded regions or cultivated gardens. A significant obstacle comes, however, from the fact that persons participating in the gathering process frequently lack professional expertise in the precise identification of medicinal plants. This lack of competence has serious consequences for both pharmaceutical production units and customers, as it commonly leads to the procurement of inaccurate or replaced medicinal plants.

One of the main challenges is the lack of skilled quality control procedures inside industrial units. This deficit exacerbates the situation, since identifying the proper medicinal plant becomes difficult, especially when the plants are received dried. The lack of understanding among both buyers and sellers about the exact plants being traded exacerbates the situation, resulting to a loop of disinformationand potential mismanagement of these vital resources.

It is critical to recognise that improper usage of medicinal herbs can render Ayurvedic treatment inefficient and, in somesituations, even dangerous to one's health. As a result, correctidentification of medicinal plants emerges as a critical issue inthis setting. To address these issues, our proposed system includes a function that provides thorough descriptions of keymedicinal plants.

This contains information on the diseases for which each plant is historically employed, as well as the wide range of alternate names connected with these plants in various areas. Our approach intends to improve the precision of medicinal plant identification by providing a comprehensive and accessible database, so adding to the efficacy and safety of Ayurvedic therapy.

II. Related Works

Most of the studies has focused on the color and textureof the blossom as the attributes [1]. They used three steps in their methodology: picture capture, image processing, and neural network. The collection contains 15 floral picture categories. The most crucial activity that requires precision is flower identification and categorization. This phase includes four image processing steps: image filtering, picture segmentation, region detection, and feature extraction. CNN was used to categorize the photos based on color and texture. The overall outcomesvary depending on the type of flower used Some flowersare 50% accurate, while others are 99% accurate. They came to the conclusion that the quantity of flower photosdepended on the correctness of the training outcome.

For identification, Almogdady et al employed image processing techniques. Flowers may be identified by their color, texture, and form [2]. They have divided their approaches into four steps: image enhancement, segmentation, feature extraction, and classification. They used the Chan-vese image processing segmentation technique to separate the flowers from the rest of the image, simplifying and improving the feature extraction operation. To extract features, they used the HSV color descriptor, the Gray Level Co-occurrence Matrix (GLCM) texture descriptor, and the Invariant Moments (IM) form descriptor. The flower photographshave been trimmed to the suitable size. They cut a singlebloom from a group of flowers while leaving the complete bunch and structure intact. Categorization was accomplished using back propagation ANN. They covered all flower species with 20 images and attained an 81% accuracy rate.

CNN was utilized for classification, and most researchers extracted features from images using imageprocessing techniques. Three characteristics were used to identify flowers: color, texture, and shape[3]. Pre-processing, segmentation, hand-design feature extraction, and classification are the four processes in their technique. They used saliency-based algorithms tochoose the Region-Of-Interest (RIO) on flower images in order to separate the blossom from the backdrop. The mean-shift method, a common segmentation technique, was also used. The parameters were optimized using CNN. The overall result shows how effective CNN is for flower recognition. And the precision was 89%.

When we looked into leaf identification researchers, we discovered that some of them used edge detection to identify leaves. [4] They mostly employed existing algorithms to achieve their goals.. They have mostly concentrated on color mapping through the use of pre- processing methods and four primary steps.

They change the picture components to create three different images from the identical test image. In their study article, we can see that they turned the photos into grayscale after considering Hue, Saturation, and Intensity. Following that, using a process known asOtsu's algorithm, all three photos were transformed to binary images.

They employed morphological traits to specify the precise leaf while identifying it. Initially, in this study work [5,] they extracted leaf attributes and utilized machine learning techniques to build a model to classify leaves. They employed Shallow Neural Network (SN), Radial Basis Function (RBF), and Deep Belief Neural Network (DBN) in their research.

Supervised learning is the process of training a model with the aforementioned approaches.The model is trained on a named or labeled dataset. composed of several types of leaf pictures in this technique. They concentrated on numerous aspects of a leaf picture. Theytook into account the colour, shape, and texture of the leaf. Once the test picture is ready, the first stage in image pre-processing is noise reduction. The goal of noise reduction is to improve feature extraction accuracy. Color-based feature extraction has been completed by taking into account pixel values in the image's Red, Green, and Blue channels. The authors focused on the border of the leaf to extract shape-based characteristics. They developed a method for determining the form of the leaf by extracting the border.

Proposed Work Plan

III. Materials And Methods

A. Data Collection

Two separate training datasets were created and distributed: the "trusted" dataset and the "web" dataset.

"Trusted" training dataset: This training dataset is based on a highly vetted collection of approximately 2000 photos and collected primarily by GBIF (Global Biodiversity Information Facility). We kept the amount of photographs per plant to around 100.

The "web" training dataset, on the other hand, was built using a collection of web photos gathered from search engines like as Google and Bing. The "web" dataset eventually contained around 1000 million photos.

Test set: For the evaluation of the models, a separate test set was created for the models' assessment, utilising multi-image plant observations acquired on the Pl@ntNet platform throughout 2021, ensuring that they were not present in the training datasets. The test set included approximately 300 photos.

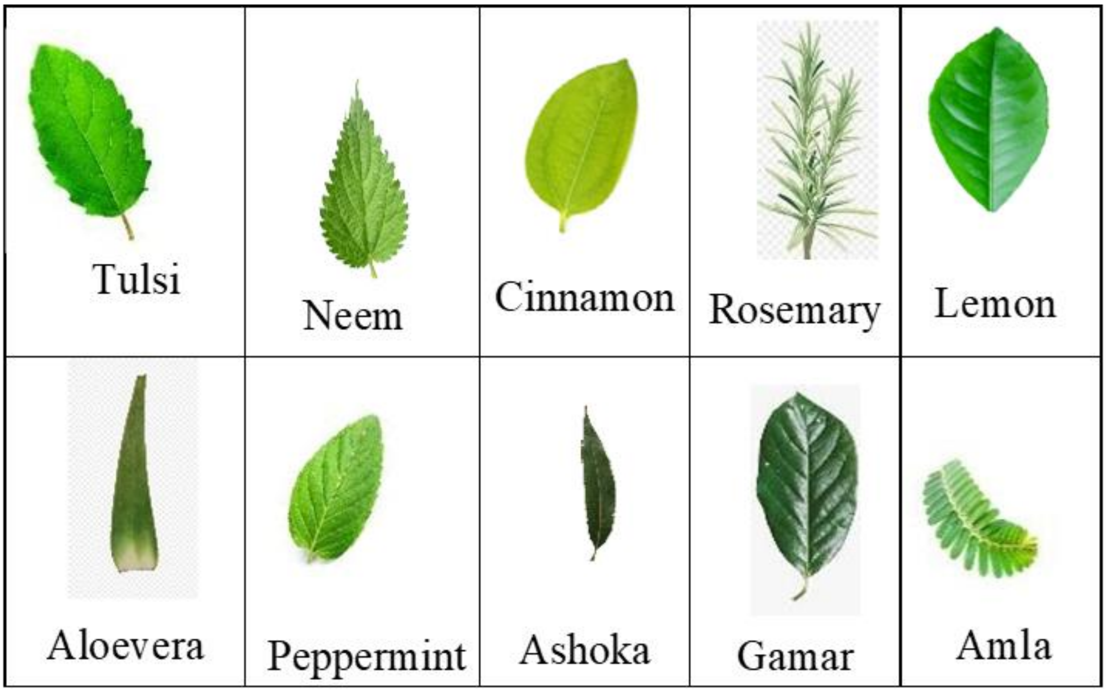


Fig. 1: Leaf images

B. Analysis

Manually analysing big volumes of data is difficult. More time and effort needed to analyse data.. To address these issues, a technology known as MachineLearning was developed.

There are various classes in Machine Learningmethods. Tensorflow is one of them. It is a machine learning open source. It includes all of the libraries and resources needed to perform data categorization [7].

C. Utilizing Recurrent Neural Network [RNN] and Convolutional Neural Network [CNN] to improve Plant Recognition

Recurrent neural networks (RNNs) are artificial neural networks that can handle sequential data, such as time series or text. RNNs have a cyclic structure, allowing them to remember past inputs and outputs. This makes them suitable for modeling dynamic systems and predicting future outcomes based on historical data.

CNNs are an additional type of neural network that can extract high-level features. from images, such as edges, shapes, and textures. The proposed method for medicinal plant detection involves using a combination of convolutional neural networks and recurrent neural networks

The CNN component is responsible for extracting features from the input images, capturing important patterns and characteristics that are indicative of plant diseases. These features are then fed into the RNN component, which utilizes sequential information to make predictions and classify the diseases. This combined CNN-RNN approach allows for holistic analysis of the image data, taking into account both spatial and temporal information. By combining RNNs and CNNs, these methods can achieve high accuracy and efficiency in identifying medicinal plants.

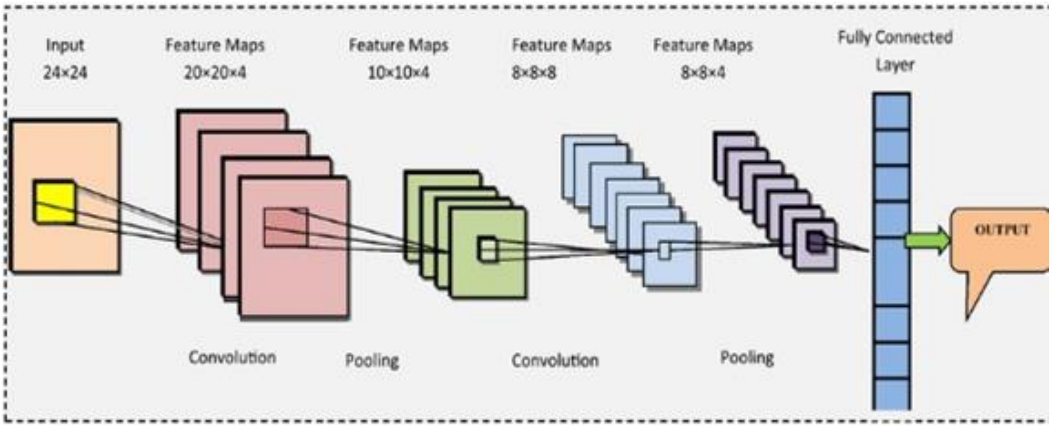


Fig. 2: Convolutional layers

A. Model for the proposed system

In the pursuit of automating the detection and classification of diseases in medicinal plants, a comprehensive framework leveraging both Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) was meticulously developed. The approach capitalizes on the inherent strengths of each neural network architecture to extract meaningful features from image data and classify them with high accuracy.

The initial phase involved the creation of a customized CNN architecture tailored specifically for the task at hand. This architecture comprised multiple convolutional layers, interspersed with pooling layers, designed to progressively extract hierarchical features from input images of medicinal plants. To get the desired accuracy, the mask of 3x3 form is taken as shown in Figure 4.

The dataset used for training the CNN model was carefully curated to encompass a wide array of plant species, growth stages, and environmental conditions. By incorporating diverse images of both healthy and diseased plants, the dataset facilitated the robust learning of discriminative features by the CNN model.

During the training process, the CNN model was trained using advanced deep learning frameworks like TensorFlow and PyTorch. Techniques like dataaugmentation were applied to augment the dataset artificiallyand enhance the model's ability to generalize to unseen data. The CNN model learned to identify intricate patterns, textures, and spatial configurations indicative of plant health or disease, thus serving as a powerful feature extractor.

Following feature extraction, the extracted features underwent a transformation process to prepare them for classification by the RNN component. High-level features extracted from the final convolutional layer of the CNN model were reshaped and encoded into a format suitable for sequential processing by the RNN. This semanticrepresentation ensured that the extracted features encapsulated relevant information crucial for accurate disease classification.

The RNN component, often implemented using Long Short-Term Memory (LSTM) networks, wasresponsible for classifying the sequential feature representations derived from the CNN. The LSTM network leveraged its ability to capture long-range dependencies and temporal dynamics to discern subtle patterns indicative of different plant diseases. Each feature vector from the CNN served as an input timestep to the LSTM network, allowing it to learn sequential dependencies and make informed predictions.

Throughout the model development process, rigorous training and evaluation procedures were employed to evaluate the model's performance and ensure its generalization to unseen data. Techniques like cross- validation and rigorous testing on diverse datasets were instrumental in validating the efficacy of the proposed framework.

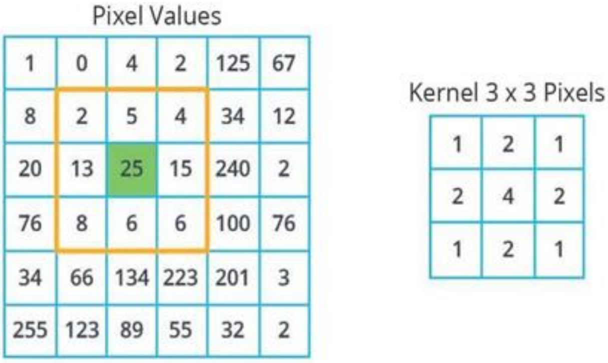


Fig. 4: 3x3 mask for CNN

IV. RESULTS

The trained neural network model is evaluated on the test set to measure classification accuracy. The regularisation parameters are adjusted in order to extract the optimal model features from the data. The table below shows the categorization accuracies.

The model's accuracy on the training set was 99.9%. This precision is determined by the number of epochs committed. When evaluated on 300 photos of 20 uncommon medicinal plants, the test set accuracies ranged from 97% to 99%. Figure 4 depicts the accuracy gained after training the photos with the model. Figure 5 depicts the losses or mistakes that happened during training. The loss is seen to diminish in this graph. That indicates there is a high level of precision throughout training The validate accuracy is depicted in Figure 6. The validation was performed as a result of picture training overfitting and underfitting. Because after the training is completed, there is the best chance of achieving the maximum accuracy on the training set.

Table 2: Classification Accuracies (100%)

	Plants name	Training Accuracy (%)	Test Accuracy (%)
1	Tulsi	100	99
2	Neem	100	98
3	Cinnamon	100	98
4	Rosemary	99	97
5	Lemon	100	98
6	Aloe Vera	100	99
7	Mint	100	98
8	Black Pepper	99	97
9	Ashoka	100	99
10	Gamar	100	98
11.	Amla	100	98
12.	Lavender	100	98
13.	Ashwagandha	100	97
14.	Chamomile	100	99
15.	Garlic	100	99
16.	Ginger	100	98
17.	Turmeric	100	99
18.	Ginseng	100	98
19.	Fengureek	100	97
20.	Hibiscus	100	99

Fig.5: Accuracy graph on training set

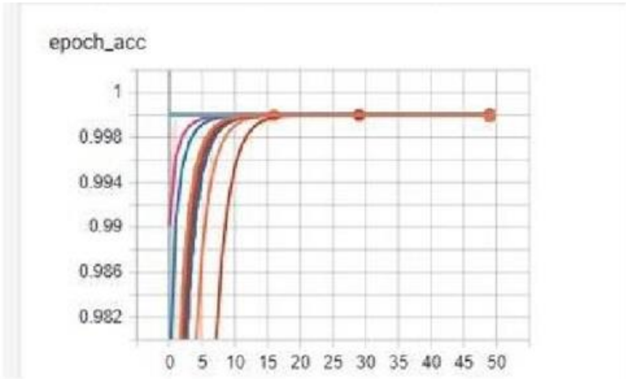


Fig.6: Validate-Accuracy graph on training set

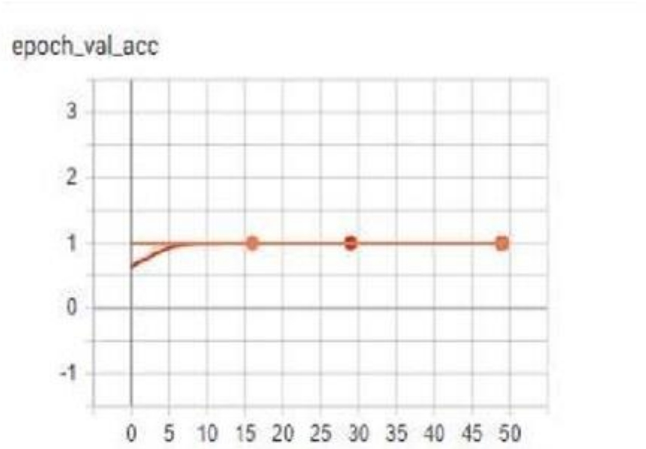
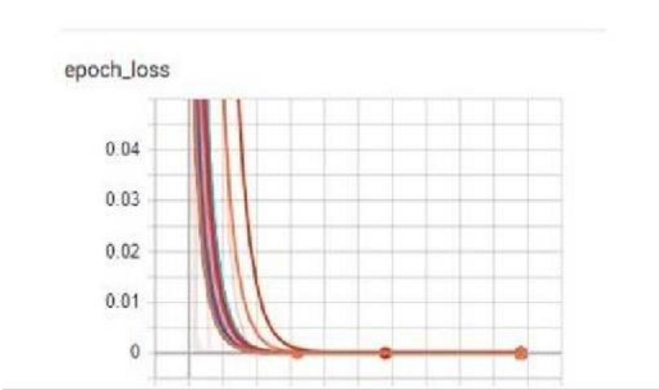


Fig.7: Loss graph on training set



V. Conclusion

The paper provided a comprehensive approach for identifying uncommon medicinal plants that makes use of RNN. Using TensorFlow on the dataset that we constructed, we attained a test accuracy of 99%. This immunity was achieved by extracting the relevant properties of the leaf picture. The accuracy is improved by the quantity of training epochs. Whenthe number of epochs is increased, it can achieve excellent precision.

References

[1] A. Jayalath (2019). Identification of Medicinal Plants by Visual Characteristics of Leaves and Flowers Available at: https://www.connectedpapers.com/main/d5dbb8016823 8951fb62a53811acc4ca0aab9cbd/Identification-of- Medicinal-Plants-by-Visual-Characteristics-of-Leaves-and-Flowers/graph [Accessed 8 August. 2023].

[2] M. Dileep, P. Pournami (2019). AyurLeaf: A Deep Learning Approach for Classification of Medicinal Plants. Available at: https://www.semanticscholar.org/paper/AyurLeaf%3A- A-Deep-Learning-Approach-for-of-Medicinal-Dileep- Pournami/2472861ab43409cf261ce7a5b2fbfdd7b7bd8ad 5 [Accessed 8 Aug. 2023]

[3] Md.Khairul Islam (2019).Bangladeshi Plant Recognition using Deep Learning based Leaf Classification. Available at: https://www.semanticscholar.org/paper/Bangladeshi- Plant-Recognition-using-Deep-Learning-Habiba- Islam/421723e833f0ad5da359d3dfefd0f3c25dbf3715 [Accessed 8 Aug. 2023]

[4] Ashutosh Bhawsar, (2020) AYURVEDIC HERB DETECTION USING COMPUTER VISION Available: https://www.semanticscholar.org/paper/AYURVEDIC-HERB- DETECTION-USING-COMPUTER-VISION-Bhawsar- Suryavanshi/64bd36f383ce56ebff73cf310d4990ec2ab9834e [Accessed: 07- Aug. 2023].

[5] Gurdit Singh(2020), Plant Identification Using Leaf Specimen. Available at : https://www.semanticscholar.org/paper/Plant- Identification-Using-Leaf-Specimen-Singh-Aggarwal/dce92a38b5e78c14943799b62609581d783ef9d9 [Accessed: 07- Aug. 2023].

[6] Keshav Kumar (2022), IDENTIFICATION OF AYURVEDIC MEDICINAL LEAVES USING DEEP LEARNING Available at: https://www.semanticscholar.org/paper/IDENTIFICATION- OF-AYURVEDIC-MEDICINAL-LEAVES-USING- Kumar/d999f521e485eab2098a42c9057dd54f9c69eec3 [Accessed: 07- Aug. 2023].

[7] A Comparative Analysis on Machine Learning Models for Accurate Identification of Medical Plants. Available at: https://www.semanticscholar.org/paper/A- Comparative-Analysis-on-Machine-Learning-Models-Thella- Ulagamuthalvi/1edeaff4d13a71bd395b1d5a9e21526b5f29e425 [Accessed 10 Aug. 2023]

[8] Introduction to Deep Learning – Deep learning basics with Python, Tensorflow and KerasAvailable at: https://pythonprogramming.net/introduction- deep- learning-ython-tensorflow-keras/[Accessed 9 Mar.2019]