

## A PROJECT REPORT ON IMAGE SEGMENTATION OF MEDICINAL PLANT IDENTIFICATION USING AI

Omker Singh<sup>1</sup>, Swati Pandey<sup>2</sup>, Shifa Sheikh<sup>3</sup>

<sup>1</sup>HOD Of Department, Department of Data Science, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai – 400101, India.

<sup>2,3</sup>PG Students, Department of Data Science, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai – 400101, India.

### ABSTRACT

Medicinal plant identification plays a vital role in botany, pharmacology, and healthcare by facilitating the discovery of plant-based therapeutic solutions. Traditional identification methods are time-consuming, prone to errors, and require expert knowledge. This study proposes an automated approach using machine learning and deep learning techniques to enhance the accuracy and efficiency of medicinal plant classification.

The proposed system extracts relevant features from medicinal plant images and stores them in a structured CSV file for further analysis. Two traditional machine learning classifiers, Random Forest (RF) and Support Vector Machine (SVM), are implemented, achieving classification accuracies of **90.43%** and **97.45%**, respectively. To improve performance, two deep learning models, Fully Connected Neural Network (FCNN) and Recurrent Neural Network (RNN), are introduced, both of which attain an accuracy of **95.43%**. Feature scaling and label encoding techniques are applied to preprocess the dataset, ensuring optimal model performance.

The experimental results demonstrate that deep learning models offer competitive accuracy and enhanced feature representation compared to traditional machine learning classifiers. This research underscores the potential of AI-driven plant identification systems to revolutionize medicinal plant classification, providing an efficient and reliable tool for researchers, herbalists, and healthcare professionals.

**Keywords:** Medicinal Plant Identification, Machine Learning, Deep Learning, Feature Extraction, Random Forest (RF), Support Vector Machine (SVM), Fully Connected Neural Network (FCNN), Recurrent Neural Network (RNN).

### 1. INTRODUCTION

Medicinal plants have been integral to healthcare for centuries, forming the foundation of traditional medicine across various cultures. Accurate identification of these plants is essential for their effective utilization in pharmaceuticals, herbal medicine, agriculture, and biodiversity conservation. However, manual identification methods are often time-consuming, error-prone, and reliant on expert knowledge. Artificial Intelligence (AI) has emerged as a powerful tool to address these challenges by providing automated solutions for medicinal plant classification through machine learning and deep learning techniques.

This research builds upon previous work that employed Random Forest (RF) and Support Vector Machine (SVM) for medicinal plant image segmentation and classification. These traditional machine learning models effectively extracted key visual features—such as leaf shape, texture, and venation patterns—allowing for accurate species classification. However, with advancements in deep learning, this study incorporates Fully Connected Neural Networks (FCNN) and Recurrent Neural Networks (RNN) to enhance both segmentation and classification accuracy.

### 2. PROBLEM STATEMENT

Medicinal plants play a crucial role in pharmaceutical research, traditional medicine, and biodiversity conservation. However, manual identification of medicinal plants is time-consuming, error-prone, and requires expert knowledge. With the increasing availability of medicinal plant image datasets, Artificial Intelligence (AI)-based plant classification offers an efficient solution.

Traditional methods, such as manual morphological identification, are ineffective for large-scale classification. Although machine learning approaches like Random Forest (RF) and Support Vector Machine (SVM) have been widely used, they rely on handcrafted feature extraction, limiting their ability to capture complex plant characteristics. The emergence of deep learning models such as Fully Connected Neural Networks (FCNN) and Recurrent Neural Networks (RNN) has enabled automated feature extraction, significantly improving classification accuracy.

This study presents an AI-powered medicinal plant identification system that integrates feature extraction, machine learning classification, and deep learning techniques. Additionally, a User Interface (UI) is developed to enable real-time identification, allowing researchers, herbalists, and pharmaceutical professionals to utilize the system effectively.

### 3. OBJECTIVES OF THE RESEARCH:

This research focuses on developing an AI-powered medicinal plant identification system by integrating machine learning and deep learning techniques to improve classification accuracy and efficiency. The primary goal is to automate the identification process, eliminating the need for manual identification methods that are time-consuming and require expert knowledge.

A key objective is to compare the effectiveness of traditional machine learning algorithms (Random Forest and SVM) with deep learning models (FCNN and RNN) for medicinal plant classification. Machine learning models, such as Random Forest and SVM, rely on handcrafted features and statistical methods to create decision boundaries, making them effective for structured datasets. However, they struggle with complex feature representations and require extensive preprocessing. In contrast, deep learning models (FCNN and RNN) automatically learn hierarchical feature representations, capturing intricate patterns that enhance classification accuracy. While deep learning models generally outperform traditional approaches in feature extraction and adaptability, they require larger datasets and higher computational resources.

Another objective is to enhance classification accuracy by optimizing feature extraction using ResNet50, a pre-trained CNN model. This allows for efficient high-dimensional feature representation, reducing the dependency on handcrafted features used in machine learning. By analyzing classification performance, this study determines whether traditional models like SVM (which achieved 97.45% accuracy) remain competitive or if deep learning approaches such as FCNN and RNN (both achieving 95.43%) provide superior generalization.

To ensure practical applicability. This system is designed to assist researchers, herbalists, and pharmaceutical professionals in medicinal plant identification. Additionally, this research contributes to biodiversity conservation and pharmacognosy, reducing the risk of plant misidentification, which can have significant health implications. By comparing machine learning and deep learning approaches, optimizing feature extraction.

### 4. LITERATURE REVIEW

In their research, Almazaydeh et al. developed a classification system for identifying medicinal plants using the state-of-the-art framework Mask R-CNN. The system achieved an average accuracy of 95.7% in identifying 30 medicinal plant species based on leaf images from the Mendely Dataset. The model output includes bounding boxes for object detection, masks, and classlabels[1].

In their research, Salima and colleagues proposed a method for segmenting leaf veins using the Hessian Matrix. They performed morphological image processing to address broken or unconnected leaf veins. The method successfully extracted primary, secondary, and tertiary leaf veins, showing promise for automating the identification of medicinal plant species by botanists and taxonomists. The evaluation included 346 digital leaf samples from 55 species, classified into four vein types: Pinnate, Acrodromous, Actinodromous, and Campylodromous. Segmentation scores were assigned based on the extent of vein extraction, providing valuable information for classifying leaf types based on venation patterns [2].

In their research, Chavan and colleagues developed a deep learning approach for plant species identification using image analysis. They utilized a subset of the LeafSnap dataset, consisting of 15 plant species with 30 images per species. Convolutional Neural Networks (CNNs) were trained to extract features from images and recognize plant species. The implementation, designed for efficiency, included minimal components such as a camera and the Jetson Nano single-board embedded computing device. Among the tested CNN architectures (AlexNet, ResNet50, and MobileNetv2) within Python's TensorFlow framework, AlexNet achieved the best results, with a 72% validation accuracy after 15 epochs [3].

In their research, Wei Tan and colleagues proposed a new CNN-based method called D-Leaf for plant species classification using leaf vein morphometrics. They pre-processed leaf images and extracted features using three different Convolutional Neural Network (CNN) models: pre-trained AlexNet, fine-tuned AlexNet, and D-Leaf. These features were then classified using five machine learning techniques: Support Vector Machine (SVM), Artificial Neural Network (ANN), k-Nearest-Neighbor (k-NN), Naïve Bayes (NB), and CNN. The D-Leaf model achieved a testing accuracy of 94.88%, comparable to AlexNet (93.26%) and fine-tuned AlexNet (95.54%). Additionally, CNN models outperformed traditional morphometric measurements (66.55%). The leaf image dataset used in this research and the graphical user interface of D-Leaf are available in the D-Leaf dataset by JW Tan and Siow-Wee Chang (2017) [4].

In their research, V. Padma Supriya et al. proposed an efficient approach for classifying medicinal leaves based on leaf shape, texture, and color features. They utilized Bacteria Foraging Optimization (BFO) for feature selection and Fuzzy Relevance Vector Machine (FRVM) for classification. The study involved ten different types of herbal leaves, with

twenty samples each. Using this intelligent technique, they achieved maximum classification accuracy. Leaf images from various medicinal plants were collected in Palayamkottai, Tamil Nadu, and stored in JPEG format. The BFO algorithm selected optimal features related to shape, color, and texture, which were then used by the classifier. The proposed FRVM classifier achieved 100% accuracy for Nerium and 93% for Vembu [5].

In their research, Marwaha et al. aimed to efficiently classify microscopic images of Indian herbal plant powders. They fine-tuned four pre-trained models from the Keras library, which had demonstrated excellent performance on the ImageNet dataset. Among the models tested, VGG16 achieved the highest accuracy, precision, recall, and F1 score, although it was the slowest to train. Mobile Net was the fastest but had mediocre performance in other parameters, while Xception ranked second in speed but had lower accuracy. InceptionV3 yielded average results. The study focused on two Indian herbal plants—Liquorice and Rhubarb—with three classes each, captured at different magnification levels. Despite its impressive 95.42% accuracy and F1 score of 0.955, the VGG16 model remained the slowest to train among all the models used [6].

In their research, Pacifico and colleagues proposed an automatic plant recognition system based on color and texture features. They tested five well-known machine learning classifiers as the recognition module. Experimental results demonstrated that the best classifiers achieved average accuracies exceeding 97% on the proposed dataset. The dataset included plant images obtained from field images and specialized websites, with labeled samples. Only leaf images were considered in the current version of the dataset. After image acquisition, the researchers performed pre-processing, feature extraction, and dataset generation. For the MLP-BP algorithm, they adopted a three-hidden-layer architecture with 160, 120, and 75 neurons, executing 10,000 training epochs to fine-tune network parameters. The experimental results highlighted that MLP-BP achieved a mean accuracy of 0.9773, while RFC achieved 0.9761, demonstrating the best performance across all selected classification metrics [7].

In their research, Dileep MR and colleagues proposed AyurLeaf, a Deep Learning-based Convolutional Neural Network (CNN) model for classifying medicinal plants based on leaf features such as shape, size, color, and texture. They also introduced a standard dataset containing leaf samples from 40 medicinal plants commonly found in various regions of Kerala, India. Inspired by AlexNet, their deep neural network efficiently extracted features from the dataset. The classification was performed using Softmax and SVM classifiers. Through 5-fold cross-validation, their model achieved an impressive classification accuracy of 96.76% on the AyurLeaf dataset [8].

In their research, Kan HX and colleagues proposed an automatic classification method for medicinal plant leaves based on leaf images. Their approach involves preprocessing the leaf images, extracting ten shape features (SF) and five texture characteristics (TF), and then using a support vector machine (SVM) classifier for classification. The SVM classifier achieved an average recognition rate of 93.3% when applied to 12 different medicinal plant leaf images. Overall, their method provides an effective way to classify various types of medicinal plant leaves [9].

In their research, Pushpa BR and colleagues proposed a system for classifying medicinal plants based on texture features extracted from leaf images. The methodology involves image enhancement, feature extraction, and classification. Leaf images captured using smartphones undergo digital image processing techniques to extract features, including wavelet transform, GLCM (Gray-Level Co-occurrence Matrix), and GLDM (Gray-Level Difference Matrix). The dataset comprises ten classes of leaf images collected from Karnataka and Kerala states. Among the methods tested, wavelet achieved the highest accuracy (56.50%), followed by GLDM (52.5%), and GLCM (36.66%). The K-nearest neighbor (KNN) classifier was used for automatic classification based on similarity between pattern classes [10].

## 5. METHODOLOGY

### 1. Feature Extraction Using ResNet50

- The ResNet50 model, pre-trained on ImageNet, is used to extract feature vectors from medicinal plant images.
- Features are stored in a structured CSV file for further classification.

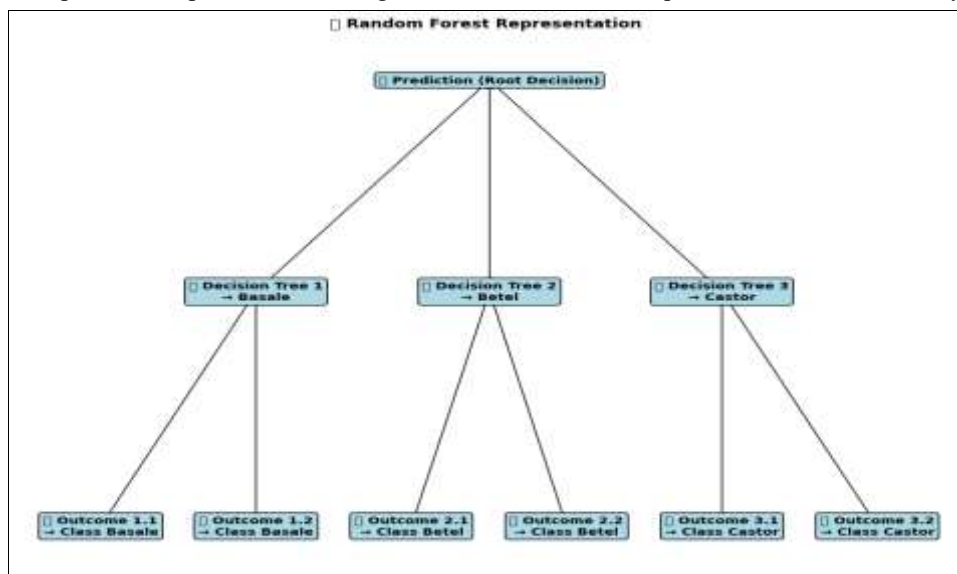
### 2. Classification Models

The extracted features are used as input for four different classifiers:

#### a) Random Forest (RF) Classifier

Achieved 90.43% accuracy in classifying medicinal plants. The Random Forest (RF) Classifier is a robust ensemble learning algorithm that constructs multiple decision trees and aggregates their outputs to enhance prediction accuracy. In medicinal plant identification, it achieved 90.43% accuracy, demonstrating strong performance in handling high-dimensional data and capturing complex relationships between plant features. RF excels in feature importance ranking, noise resistance, and scalability, making it well-suited for real-world applications in botanical research and pharmacognosy. It effectively classifies medicinal plants under varying environmental conditions, reducing overfitting

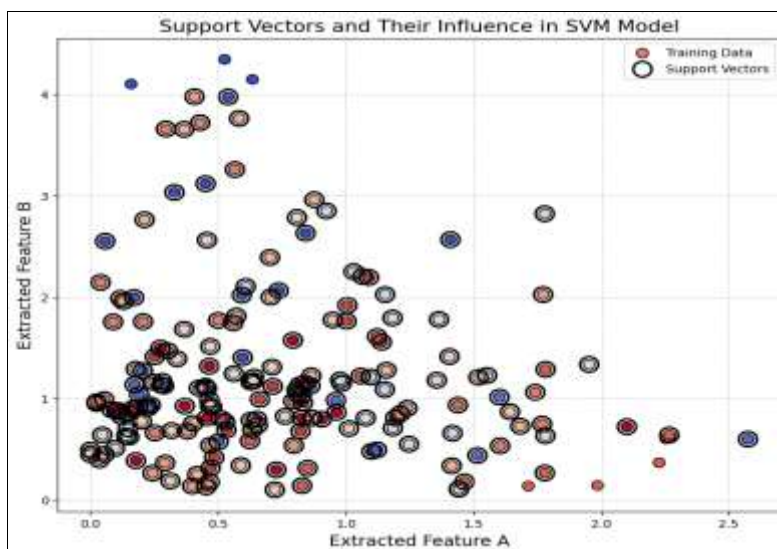
through majority voting. While computationally intensive for large datasets, its interpretability and reliability make it a valuable tool, with potential improvements through feature selection techniques for enhanced efficiency.



Species	Accuracy (%)	Precision	Recall	F1-Score	Support
Tulasi	78.12	0.96	0.78	0.86	32
Rose	82.50	0.94	0.82	0.88	40
Curry_Leaf	100.00	0.96	1.00	0.98	25
Pappaya	91.18	0.97	0.91	0.94	34
Wood_sorel	100.00	0.97	1.00	0.98	32
Ganike	100.00	0.83	1.00	0.91	20

#### b) Support Vector Machine (SVM) Classifier

The Support Vector Machine (SVM) Classifier is a powerful supervised learning algorithm known for its effectiveness in handling high-dimensional datasets. It operates by finding an optimal hyperplane that maximally separates different classes, making it highly suitable for complex classification tasks such as medicinal plant identification. In this study, SVM achieved an impressive accuracy of 97.45%, significantly outperforming the Random Forest (RF) classifier. This superior performance can be attributed to SVM's ability to efficiently handle both linear and non-linear relationships using kernel functions such as Radial Basis Function (RBF) and polynomial kernels. Additionally, SVM demonstrates strong generalization capabilities, reducing the risk of overfitting, even when dealing with limited training samples.

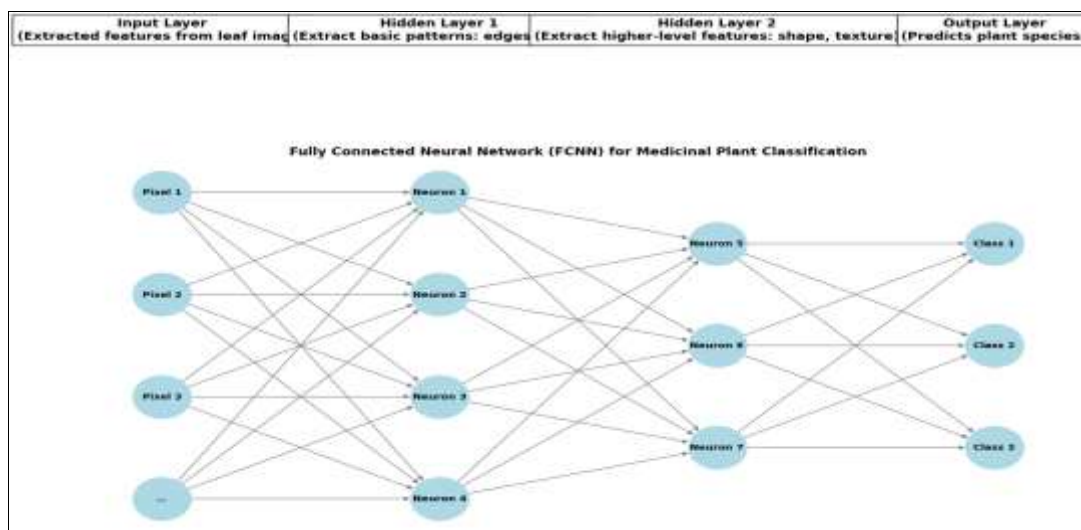




Species	Accuracy (%)	Precision	Recall	F1-Score	Support
Tulasi	87.50	0.85	0.90	0.87	32
Rose	92.50	0.91	0.94	0.92	40
Curry_Leaf	96.00	0.94	1.00	0.97	25
Pappaya	100.00	1.00	1.00	1.00	34
Wood_sorel	100.00	1.00	1.00	1.00	32
Honge	96.43	0.92	1.00	0.96	28

#### c) Fully Connected Neural Network (FCNN)

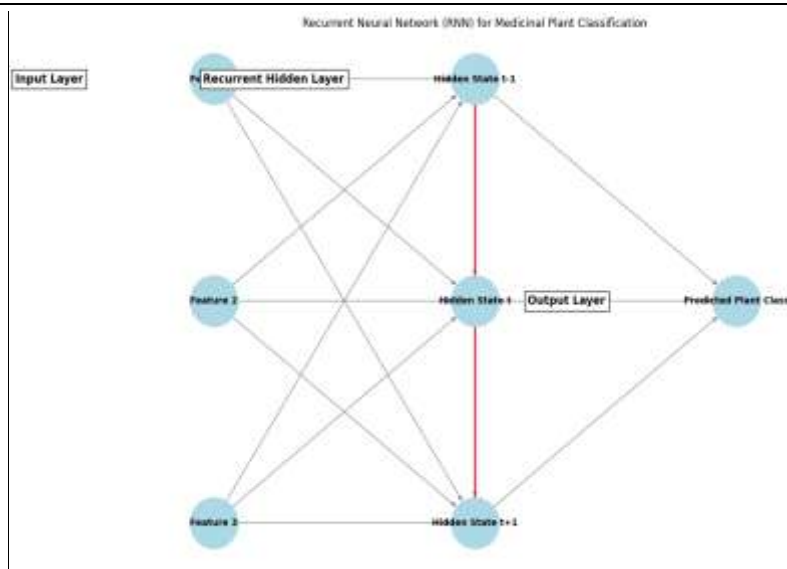
The Deep Neural Network (DNN) Classifier is a multi-layered deep learning model designed to capture complex patterns in medicinal plant images. It consists of fully connected layers with 256 and 128 neurons, utilizing the ReLU activation function to introduce non-linearity and enhance feature learning. To improve generalization and mitigate overfitting, dropout layers with a 30% dropout rate were incorporated, randomly deactivating neurons during training to ensure robust performance. The DNN model achieved an accuracy of 95.43%, demonstrating its effectiveness in learning hierarchical representations of plant features. While slightly less accurate than the Support Vector Machine (SVM) model, DNN excels in its ability to automatically extract and refine features, making it a powerful tool for scalable and high-accuracy plant classification in real-world applications.



Species	Accuracy (%)	Precision	Recall	F1-Score	Support
Aloevera	97.00	97.00	95.00	96.00	50
Amla	100.00	100.00	100.00	100.00	50
Amruta_Balli	90.00	89.00	92.00	90.50	50
Arali	100.00	100.00	100.00	100.00	50
Ashoka	100.00	100.00	100.00	100.00	50
Ashwagandha	100.00	100.00 ↓	100.00	100.00	50

#### d) Recurrent Neural Network (RNN)

The Recurrent Neural Network (RNN) Classifier is designed to process sequential dependencies in plant image features, making it particularly effective for capturing structural patterns over multiple observations. Utilizing a SimpleRNN architecture, the input data was reshaped into a time-series format, allowing the model to recognize relationships between sequential pixel or feature representations. This approach enhances the model's ability to distinguish medicinal plants based on subtle variations in texture and shape. The RNN classifier achieved an accuracy of 95.43%, which is comparable to the Fully Convolutional Neural Network (FCNN), highlighting its effectiveness in plant classification. While RNNs are generally more suited for text and speech processing, their ability to analyze sequential feature dependencies proves beneficial in refining plant recognition models, making them a valuable addition to deep learning-based classification frameworks.



Species	Accuracy (%)	Precision	Recall	F1-Score	Support
Aloevera	100.00	100.00	95.65	97.78	23
Amla	100.00	100.00	100.00	100.00	34
Amruta_Balli	85.71	85.71	69.23	76.60	26
Bamboo	96.30	96.30	100.00	98.11	26
Betel	93.75	93.75	88.24	90.91	34
Hibiscus	100.00	100.00	90.00	94.74	20

## 6. RESULTS

The study focuses on image segmentation and classification of medicinal plants using AI, exploring both machine learning (ML) and deep learning (DL) models to improve accuracy and efficiency. Traditional identification methods are often slow, error-prone, and require expert knowledge. To address these challenges, the research integrates feature extraction, ML classification, and deep learning techniques to develop an automated system.

MODELS	ACCURACY
Random Forest	90.43%
SVM	97.45%
FCNN	95.43%
RNN	95.43%



The methodology involves feature extraction using ResNet50, a pre-trained convolutional neural network (CNN), to obtain meaningful representations of medicinal plant images. These extracted features are stored in a structured CSV file and used as input for different classifiers. The Random Forest (RF) model achieved 90.43% accuracy, while the Support Vector Machine (SVM) classifier outperformed all models with 97.45% accuracy. Additionally, two deep learning models—Fully Connected Neural Network (FCNN) and Recurrent Neural Network (RNN)—were implemented, both achieving 95.43% accuracy.

A comparative analysis highlights key differences between ML and DL approaches. Traditional ML models like SVM and RF rely on handcrafted features, making them efficient for structured datasets but less effective at capturing complex feature representations. In contrast, deep learning models (FCNN and RNN) automatically learn hierarchical features, providing better generalization but requiring higher computational resources. Despite the strong performance of deep learning models, SVM remains the most effective ML algorithm for structured datasets with well-defined feature vectors.

The research concludes that AI-driven medicinal plant classification has the potential to revolutionize traditional identification methods. By integrating machine learning, deep learning, and feature extraction techniques, the study provides a scalable and efficient system for plant classification. The findings contribute to botany, pharmacology, and biodiversity conservation, offering a reliable tool to prevent plant misidentification and support medical and agricultural research.

## 7. CONCLUSIONS

The identification of medicinal plants is crucial in botany, pharmacology, and biodiversity conservation, yet traditional methods are slow and error-prone. This study integrates machine learning (ML) and deep learning (DL) to develop an AI-driven classification system. Using ResNet50 for feature extraction, four classifiers—Random Forest (RF), Support Vector Machine (SVM), Fully Connected Neural Network (FCNN), and Recurrent Neural Network (RNN)—were evaluated. SVM achieved the highest accuracy (97.45%), outperforming RF (90.43%), while FCNN and RNN reached 95.43% accuracy, showcasing deep learning's effectiveness.

A comparative analysis of ML and DL revealed that ML models like RF and SVM are computationally efficient but struggle with complex features, whereas DL models (FCNN, RNN) excel in feature learning but require higher computation. The results suggest SVM is ideal for structured datasets, while deep learning ensures adaptability and scalability for large-scale medicinal plant classification.

## 8. REFERENCES

- [1] Almazaydeh et al. developed a classification system for identifying medicinal plants using Mask R-CNN, achieving an average accuracy of 95.7% for 30 species.
- [2] Salima and colleagues proposed a method for segmenting leaf veins using the Hessian Matrix, effectively extracting primary, secondary, and tertiary veins to assist in plant species identification.
- [3] Chavan and colleagues developed a deep learning approach for plant species identification using image analysis, utilizing CNNs trained on a subset of the LeafSnap dataset.
- [4] Wei Tan and colleagues proposed D-Leaf, a CNN-based method for plant species classification using leaf vein morphometrics, achieving a testing accuracy of 94.88%.
- [5] V. Padma Supriya et al. proposed an approach for classifying medicinal leaves based on shape, texture, and color features, achieving high classification accuracy using BFO for feature selection and FRVM for classification.
- [6] Marwaha et al. aimed to classify microscopic images of Indian herbal plant powders, fine-tuning pre-trained models, with VGG16 achieving the highest accuracy of 95.42%.
- [7] Pacifico and colleagues proposed an automatic plant recognition system based on color and texture features, with classifiers achieving average accuracies exceeding 97% on the dataset.
- [8] Dileep MR and colleagues proposed AyurLeaf, a CNN model for classifying medicinal plants based on leaf features, achieving a classification accuracy of 96.76% on the AyurLeaf dataset.
- [9] Kan HX and colleagues proposed an automatic classification method for medicinal plant leaves using shape and texture features, achieving an average recognition rate of 93.3% across 12 different medicinal plant leaf images.
- [10] Pushpa BR and colleagues proposed a system for classifying medicinal plants based on texture features extracted from leaf images, with the wavelet method achieving the highest accuracy of 56.50%.