Machine Learning-Based Recognition of Indian Medicinal Plants Species Using Handcrafted Features

Gajanan Digambar Patil[1]* and Pritesh R. Gumble [2]

[1,2]Dept. of Electronics & Telecommunication Engineering
Sipna College of Engineering & Technology, Amravati, India

* gdpatil1203@gmail.com
[2] gpriteshg@gmail.com

Abstract. Recognizing and categorizing Indian medicinal plant species plays a vital role in safeguarding biodiversity and advancing pharmaceutical studies. Traditional manual classification methods take a lot of time and often involve mistakes. This paper introduces a machine learning-driven recognition system that automates the classification of Indian medicinal plants through handcrafted feature extraction methods. The method uses color and texture features to improve classification accuracy by analyzing the surface patterns and structural properties of plant leaves. Four machine learning algorithms classify the extracted features, and their performances are compared. The benchmark dataset includes images of various Indian medicinal plant species, which are processed and segmented before feature extraction. The experimental findings indicate that the Random Forest classifier surpasses other models in terms of accuracy, delivering superior classification performance. This approach offers a cost-effective and efficient solution for identifying plant species, helping botanists and researchers automate the classification of medicinal plants.

Keywords: Handcrafted Features, Indian Medicinal Plant, Machine Learning, Plant Species Classification.

1 INTRODUCTION

India is home to a wide variety of medicinal plant species that are widely used in Ayurveda, Siddha, and other traditional medicine practices. Identifying and classifying these plants is important for pharmaceutical research, conserving biodiversity, and herbal medicine. However, manual classification methods depend on botanical knowledge, making them time-consuming, subjective, and error-prone. With improvements in Artificial Intelligence and machine learning, automated plant species recognition has become a reliable and efficient alternative [1-2].

A. Jain and S. Gupta (eds.), *Proceedings of the International Conference on Policies, Processes and Practices for Transforming Underdeveloped Economies into Developed Economies (PPP-UD 2025)*, Advances in Economics, Business and Management Research 357,
https://doi.org/10.2991/978-94-6463-894-3_20

Machine Learning (ML) techniques allow for classifying plant species having features extracted from leaf images, which are key identifiers for plant species. Among the different feature extraction methods, handcrafted features such as color and texture analysis are especially useful for telling different plant species apart. Color features capture changes in pigmentation, while texture features look at the structural patterns on the leaf surface, making an accurate distinction possible. By combining handcrafted feature extraction with machine learning algorithms, this study aims to advance computational botany and medicinal plant research. The proposed framework offers a cost-effective, scalable, and efficient solution for identifying plant species. This can help researchers, herbalists, and agricultural experts automate plant classification tasks [3-4]. This study suggests a ML-based framework for recognizing Indian medicinal plant species using handcrafted feature extraction techniques. The primary contributions of this research are:

- Developed an efficient and automated classification model for Indian medicinal plant species.
- Analyzed the impact of handcrafted color and texture features on classification accuracy.
- Compared the performance of different machine learning classifiers in plant species recognition.

The remaining sections of this article are structured as follows: Section II provides a comprehensive literature review, highlighting existing research on medicinal plant classification. Section III details the materials and methodologies. Section IV presents the experimental setup, dataset, evaluation metrics, and results, followed by an in-depth discussion of model performance. Finally, Section V concludes the study, summarizing key findings, discussing implications, and suggesting future research directions.

2 RELATED WORK

Numerous current studies have concentrated on the automated classification of medicinal plants through a variety of computational methods that are covered in more detail below. A Random Forest (RF) classifier with a promising accuracy of 75.46% on the RTP40/80 real-time datasets and 94.54% on the GSL100 leaf dataset is used in the new hierarchical classification framework to classify one hundred Indian medicinal plant species [5]. Using pretrained models VGG16 and VGG19 on 64 medicinal plants, Deep Learning (DL) techniques, and Convolutional Neural Networks (CNN) are used to identify the native ayurvedic medicinal plant species [6]. All piper plant types are classified using the hyperparameter tuned random forest algorithm because of its effective optimal hyperparameter tuning with Deep CNN feature extraction, which yields an accuracy value of 0.88 [7]. The medicinal plant portion is based on multi-label categories determined by using a sigmoid classifier as the final layer of a CNN with Mobile Net [8]. DL based algorithms based on Mendeley Data, which includes 30 different medicinal plant species, are used to identify and classify medicinal plants optimally [9]. The leaf-based automatic plant categorization system is explored and compared to two distinct deep learning algorithms, DNN and Deep CNN [10]. A collection of screening methods for spectrum features relevant to plant categorization and status monitoring based on ISODATA and JM distance, as well as an investigation into the effectiveness of several ML classifiers in the spectral library application [11]. A chi-square-based feature selection approach was employed to develop an ML-driven classification method utilizing run-length matrix, texture, and multispectral data [12]. Many ML techniques are used to identify medicinal plants more accurately by extracting features and applying a model.[13] The plant leaf species recognition was created using SMO and SVM on a random dataset [14]. Radial Basis Function Neural Networks (RBFNN) are applied to fit the leaf boundary points into a continuous contour for determining the leaf shape's centroid, while the Salp Swarm Algorithm is employed to optimize the SVM [15]. The neural network approach used for the Indian medicinal plants classification [16]. A detailed evaluation of ML algorithms for identifying medicinal plants is offered, including Fuzzy C means, a clustering ML algorithm based on color, texture, and geometrical aspects [17]. A hybrid approach is proposed that addresses this challenge by integrating two types of classifiers, leveraging CNN and SVM to accurately identify medicinal plants [18]. The effectiveness of these DL classifiers in detecting leaf images based on typical plant qualities such as shape, vein, texture, and a variety of other factors is categorized, and the results provide information on the therapeutic benefits of the plants [19]. The venous phenotypic extraction and multi-omics association analysis utilizing machine learning technologies offered a theoretical foundation for optimizing the vein network architecture to improve crop productivity [20]. Pre-processing, contour-based leaf feature extraction for plant species classification and classification using several ML approaches [21]. A mobile-based tool is used to capture structured images, reducing data cleaning needs and enabling accurate classification with ANN, PNN, or CNN [22]. ML was used to extract attributes from the images for the medicinal plants classification, such as leaf texture, shape, and color [23]. A two-stage DL based method was designed to merge NEON remote-sense data from three sensors for individual plant species and genera classification [24]. The use of an RF classifier to identify and detect medicinal plant species is a novel approach [25]. Plant species are identified from leaf samples using the improved vegetation index, ExG-ExR, to extract richer vegetative details [26]. Hyperspectral leaf data are classified with various ML models, enhanced through feature reduction, selection techniques, and Bayesian optimization [27].

3 METHODOLOGY

The proposed medicinal plant species classification methodology consists of multiple stages, including dataset collection, image preprocessing, handcrafted feature extraction, classification using machine learning models, and performance evaluation. The workflow is structured as shown in Fig. 1.

3.1 Indian Medicinal Plant Datasets

The proposed experimentation used Indian Medicinal plant datasets [28], which consist of images of medicinal plants. The pictures are taken with different backgrounds and no limits on the surroundings. This dataset consists of medicinal leaf images. It comprises 80 distinct Indian leaf varieties renowned for their potent medicinal properties and offers a rich opportunity for advancing healthcare, botanical studies, and machine learning applications.

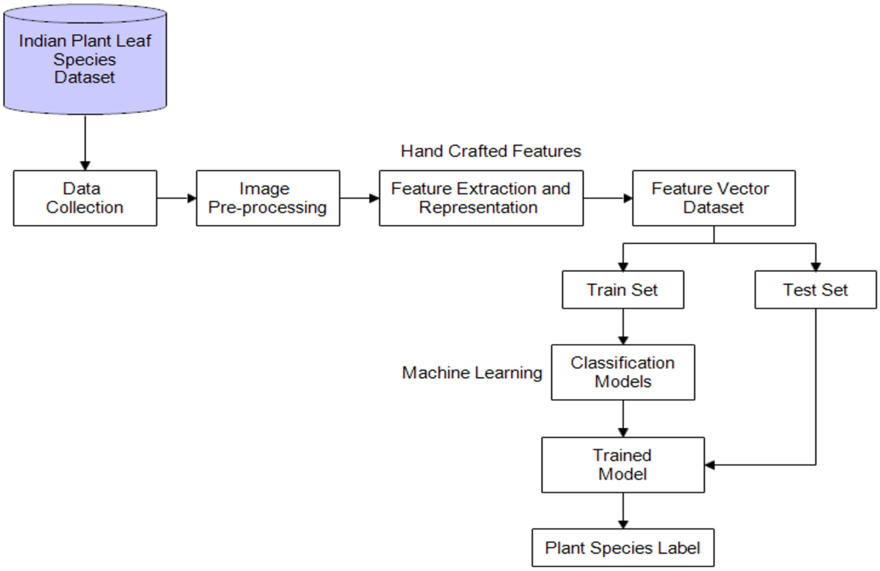

Fig.1. Indian Medicinal Plant Species Classification Framework

3.2 Image Preprocessing

Pre-processing plays a crucial role in improving image quality and maintaining consistency throughout the dataset. The all images were resized to a uniform resolution as 256×256 pixels for standardization, applying Gaussian or median filters to remove noise, and enhancing visual clarity using histogram equalization.

3.3 Feature Extraction

Handcrafted feature extraction involves computing statistical and texture-based attributes from leaf images. Two types of features are extracted Color and Texture:

Color features provide essential information about the pigmentation and chemical composition of plant species. The following statistical measures are computed for each channel (R, G, B):

Mean: Average intensity of each channel.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Standard Deviation: Measure of colour variation across pixels.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2}$$

Skewness: Asymmetry of the colour distribution.

$$\gamma = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i - \mu}{\sigma}\right)^3$$

Kurtosis: Sharpness of the colour distribution curve.

$$\kappa = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i - \mu}{\sigma}\right)^4$$

Texture features capture the surface patterns and structural characteristics of leaves. The following GLCM-based features are extracted:

Contrast: Measures the intensity difference between a pixel and its neighbour.

$$Contrast = \sum_{i,j} P(i,j)(i-j)^2$$

Correlation: Evaluates the dependency between neighbouring pixels.

$$Correlation = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i \sigma_j}$$

Energy: Represents textural uniformity (higher energy means less variation.

$$Energy = \sum_{i,j} P(i,j)^2$$

Homogeneity: Measures closeness of pixel distributions.

$$Homogeneity = \sum_{i,j} \frac{P(i,j)}{1+|i-j|}$$

Additionally, Entropy (randomness in the image) and Mean Pixel Intensity (average gray scale value) are computed. Entropy measures how random or unpredictable the image is quantifies the randomness or unpredictability present in the image. The mean pixel intensity is the average gray scale value of all pixels in the picture.

$$Entropy = -\sum_{i,j} P(i,j) \log P(i,j)$$

$$I_{mean} = \frac{1}{N} \sum_{i=1}^{N} I_i$$

3.4 Classification Using Machine Learning Models

The collected features after feature extraction process, serve as input for four ML classifiers, RF, GB, SVM and KNN.

Random Forest (RF): An ensemble method that constructs numerous decision trees and combines their outputs to obtain reliable and accurate categorization.

Gradient Boosting (GB): A boosting strategy that develops models progressively to rectify earlier errors while reducing the loss function via gradient descent.

Support Vector Machine (SVM): A supervised learning approach that identifies the best hyper plane for class separation. Nonlinear classification is handled with a Radial Basis Function (RBF) kernel.

k-Nearest Neighbours (KNN): Using the Euclidean distance metric to evaluate similarity, this distance-based classifier assigns labels based on the majority class among the nearest neighbours.

4 EXPERIMENTAL RESULTS AND DISCUSSION

The proposed plant species recognition reorganization was evaluated experimentally using MATLAB R2021a on a system with the following configuration: Intel Core i5, 16 RAM, Windows OS. This approach allowed for the efficient execution of image processing, feature extraction, and classification tasks by leveraging MATLAB's built-in functions for texture analysis, statistical computations, and machine learning implementation. To guarantee that all plant species are represented fairly, the feature dataset splitted into train and test sets as 80-20% ratio using stratified sampling. The training set is used to train each classifier, which is then tested on new data. The metrics used to evaluate performance are as follows:

Accuracy: It represents the overall percentage of plant species samples that are correctly classified.

Precision: This measures the proportion of correctly classified plant species among all samples predicted as that species.

Recall: It indicates the proportion of correctly identified samples for each plant species among all actual samples of that species.

F1-score: It offers a fair assessment of categorization ability since it is the harmonic mean of precision and recall.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The benchmark dataset was classified into multiple categories. The cross-validation technique was used to determine the best accuracy among four machine learning classifiers, and the results are reported. Fig.2. depicts the sample test photos from the dataset utilized in the testing. There are 40 species categories used for classification.

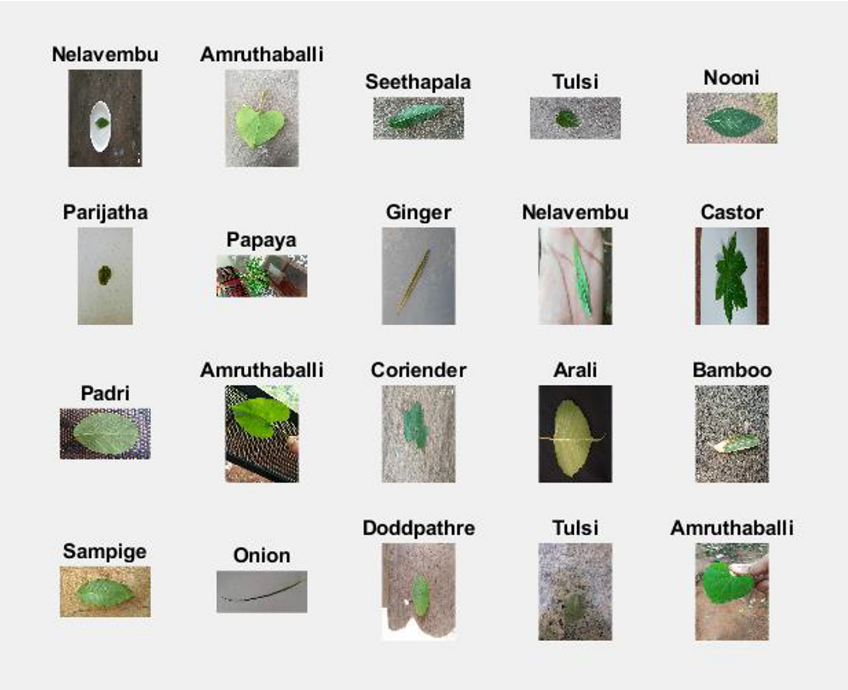

Fig.2. Sample test image of Indian Medicinal Plant Leaf Species

Table 1. displays the computational time required to train and test machine learning classifiers. It demonstrates that RF and GB classifiers take longer to train and test the system, as illustrated in Fig.3.

Table 1. Computation Time Analysis

Classifiers	**Computation Time**	
	Training Time (sec)	**Testing Time (sec)**
RF	580	251
GB	615	284
SVM	558	265
KNN	595	210

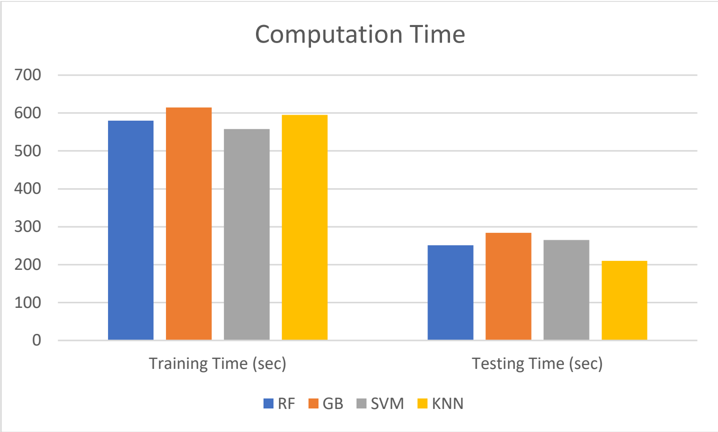

Fig.3. Analysis of computation time in sec

Table 2. Displays the overall classification performance of all four machine learning classifiers, including accuracy, precision, recall, and f1-score values. It demonstrates that RF outperformed other machine learning classifiers with a high accuracy of 91.2%, as illustrated in Fig.4.

Table 2. Classification Performance Analysis

Classifiers	**Overall Performance Measures**			
	Accuracy (%)	**Precision (%)**	**Recall (%)**	**F1-score (%)**
RF	91.2	90.5	91.5	90.2
GB	88.5	88.1	89.4	88.4
SVM	86.8	86.9	86.1	86.2
KNN	85.4	85.5	84.9	85.1

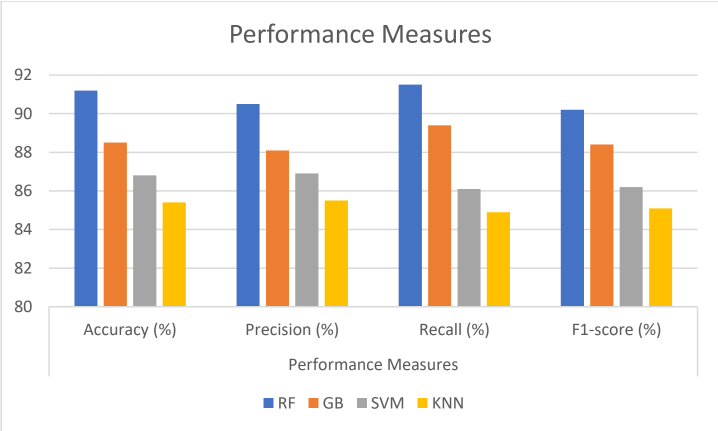

Fig. 4. Analysis of performance measures

5 CONCLUSIONS

This paper demonstrated the use of machine learning and handmade feature extraction techniques to identify Indian medicinal plant species. The method successfully distinguished between plant species by using texture and color characteristics. Several ML classifiers were used to classify the collected features; Random Forest had the highest accuracy of 91.2%. Handcrafted characteristics offer a computationally efficient and interpretable solution for plant species classification. Furthermore, Random Forest's ensemble learning capabilities allowed it to beat other classifiers by efficiently addressing differences in leaf texture and pigmentation. Some issues still exist despite encouraging results, such as classification errors brought on by identical leaf appearances, sensitivity to changes in lighting, and dataset limits. In the future, the dataset will be expanded to include additional plant species, real-time mobile or web-based applications will be developed for practical deployment, and DL algorithms will be integrated for automatic feature extraction.

References

1. S Sridevy, M Nirmala Devi, M Sankar, Nivas Raj Moorthi. Image pre-processing techniques utilized for the plant identification: A review. Pharma Innovation 2023; 12(6):320-324.
2. P. Dalvi and D. R. Kalbande, "A Comprehensive Review of Plant Recognition Approaches: Techniques, Challenges, and Future Direction", ssgmjse, vol. 1, no. 1, pp. 1–5, Jun. 2023.
3. Thyagharajan, K.K., Kiruba Raji, I. A Review of Visual Descriptors and Classification Techniques Used in Leaf Species Identification. Arch Computat Methods Eng 26, 933–960 (2019). https://doi.org/10.1007/s11831-018-9266-3.
4. Nilesh S. Bhelkar, Dr. Avinash Sharma, "Study on Identification and Classification of Medicinal Plants Using Machine Learning and Deep Learning," International Journal of Engineering Technology and Management Sciences Website: ijetms.in Issue: 3 Volume No.7 May - June – 2023, DOI:10.46647/ijetms.2023.v07i03.109.
5. Pushpa, B.R., et al. "On the Importance of Integrating Convolution Features for Indian Medicinal Plant Species Classification Using Hierarchical Machine Learning Approach." Ecological Informatics, vol. 81, 2024, p. 102611, https://doi.org/10.1016/j.ecoinf.2024.102611.
6. A. Paulson and S. Ravishankar, "AI Based Indigenous Medicinal Plant Identification," 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), Cochin, India, 2020, pp. 57-63, doi: 10.1109/ACCTHPA49271.2020.9213224.
7. A Pravin and Dr. C.Deepa, "Piper Plant Classification using Deep CNN Feature Extraction and Hyperparameter Tuned Random Forest Classification", tjes, vol. 13, Aug. 2022.
8. M. Widneh, A. Workneh, and A. Alemu, "Medicinal Plant Parts Identification and Classification using Deep Learning Based on Multi Label Categories", EJSSD, vol. 8, no. 2, pp. 96-108, Sep. 2021.
9. Ghosh, Sukanta & Singh, Amar & Kumar, Shakti. (2023). Identification of medicinal plant using hybrid transfer learning technique. Indonesian Journal of Electrical Engineering and Computer Science. 31. 1605. 10.11591/ijeecs.v31.i3.pp1605-1615.
10. Tiwari, Shamik. "A Comparative Study of Deep Learning Models With Handcraft Features and Non-Handcraft Features for Automatic Plant Species Identification," International Journal of Agricultural and Environmental Information Systems (IJAEIS) 11, no.2: 44-57. https://doi.org/10.4018/IJAEIS.2020040104.
11. Zhang, Jingcheng, etal. "Machine Learning-Based Spectral Library for Crop Classification and Status Monitoring." Agronomy, vol. 9, no. 9, 2019, p. 496, https://doi.org/10.3390/agronomy9090496.
12. Naeem, Samreen, et al. "The Classification of Medicinal Plant Leaves Based on Multispectral and Texture Feature UsingMachineLearningApproach."Agronomy,vol.11,no.2,2021,p.263, https://doi.org/10.3390/agronomy11020263.
13. Meshram, R.S., Patil, N. (2022). Classification of Medicinal Plants Using Machine Learning. In: Reddy, V.S., Prasad, V.K., Mallikarjuna Rao, D.N., Satapathy, S.C. (eds) Intelligent Systems and Sustainable Computing. Smart Innovation, Systems and Technologies, vol 289. Springer, Singapore. https://doi.org/10.1007/978-981-19-0011-2_24.

14. Sahil Vyas, Manish Kumar Mukhija, Satish Kumar Alaria; Design and analysis of new algorithm for plant leaf species identification using by SVM and SMO. AIP Conf. Proc. 15 June 2023; 2782 (1): 020120. https://doi.org/10.1063/5.0154299.
15. Ahmed A, Hussein SE (2020) Leaf identification using radial basis function neural networks and SSA based support vector machine. PLOS ONE 15(8): e0237645. https://doi.org/10.1371/journal.pone.0237645.
16. G. Murali et.al. “Machine Learning Based Categorization of Medicinal Plants,” International Research Journal of Modernization in Engineering Technology and Science, Volume:06/Issue:03/March-2024.
17. G. Priyanga et.al. “Automated Identification of Medicinal Plants Using Machine Learning,” 2024 IJCRT | Volume 12, Issue 3 March 2024 | ISSN: 2320-2882, pp. 718-722.
18. M. Srinivasa Rao, S. Praveen Kumar, and K. Srinivasa Rao, “Classification of Medical Plants Based on Hybridization of Machine Learning Algorithms”, IJISS, vol. 13, no. 2, pp. 14–21, Nov. 2023.
19. V. Kokila et.al. “Analysis of Medicinal Herbs Using Machine Learning,” 2024, International Journal of Research Publication and Reviews, Vol 5, no 3, pp 3257-3261.
20. Zhang, Yubin et.al. “Machine learning for image-based multi-omics analysis of leaf veins.” Journal of experimental botany vol. 74,17 (2023): 4928-4941. doi:10.1093/jxb/erad251.
21. Malarvizhi K et.al. “Machine Learning for Plant Species Classification using Leaf Vein Morphometric,” International Journal of Engineering Research & Technology (IJERT), Vol. 10 Issue 04, April-2021, pp. 513-516.
22. C. S., Udaya, and Usharani M. "Medicinal Plant Identification Using Machine Learning Techniques: Automatic Recognition of Medicinal Plants." In Handbook of Research on Advances in Data Analytics and Complex Communication Networks, edited by P. Venkata Krishna, 120-130. Hershey, PA: IGI Global, 2022. https://doi.org/10.4018/978-1-7998-7685-4.ch008.
23. T. Vigneswari et.al. “Medicinal Plant Identification using Machine Learning,” International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), Volume 4, Issue 7, April 2024, pp. 394-400.
24. Scholl, Victoria M.,etal. "Fusion Neural Networks for Plant Classification: Learning to Combine RGB, Hyperspectral, and Lidar Data." PeerJ, vol. 9, 2021, p. e11790, https://doi.org/10.7717/peerj.11790.
25. R. Nidhya, K. Imambee, M. Geethanjali, M. L. Anusha and I. Vyshnavi, "Real Time Medicinal Plant Identification & Classification Using Random Forest Algorithm," 2024 10th International Conference on Communication and Signal Processing (ICCSP),Melmaruvathur,India,2024,pp.1499-1504,doi: 10.1109/ICCSP60870.2024.10543388.
26. C. Sivaranjani, L. Kalinathan, R. Amutha, R. S. Kathavarayan and K. J. Jegadish Kumar, "Real-Time Identification of Medicinal Plants using Machine Learning Techniques," 2019 International Conference on Computational Intelligence in Data Science(ICCIDS),Chennai, India, 2019, pp. 1-4, doi: 10.1109/ICCIDS.2019.8862126.
27. Song, Guangman, and Quan Wang. "Species Classification from Hyperspectral Leaf InformationUsingMachineLearningApproaches." EcologicalInformatics, vol. 76, 2023, p.102141, https://doi.org/10.1016/j.ecoinf.2023.102141.
28. B R, Pushpa; Rani, Shobha (2023), “Indian Medicinal Leaves Image Datasets”, Mendeley Data, V3, doi: 10.17632/748f8jkphb.3.