JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AN EXTENSIVE INVESTIGATION ON MEDICINAL PLANTS IN THE CONTEXT OF MACHINE LEARNING

[1]Sana Shaikh, [2]Dr. Vaishali A. Chavan

[1]Research Scholar, [2]Assistant Professor
[1]Computer Science & Information Technology, [2]Computer Science & Information Technology,
[1]Dr. G.Y. Pathrikar College of C.S. & I.T., MGM UNIVERSITY, Chh. Sambhajinagar, India
[2]Dr. G.Y. Pathrikar College of C.S. & I.T., MGM UNIVERSITY, Chh. Sambhajinagar, India

Abstract : India is home to a rich diversity of medicinal plants, which play a crucial role in traditional medicine systems such as Ayurveda, Siddha, and Unani. The automatic identification and classification of these medicinal plants using modern technologies like Machine Learning (ML) and Computer Vision (CV) have significant potential to enhance the conservation, utilization, and commercialization of these resources. This literature article examines 33 papers published between 2010 and 2023, exploring the key themes, methodologies, findings, and gaps in the existing research on this topic, focusing specifically on the different techniques used to identify and classify medicinal plants.

Keywords – Medicinal Plants, Identification, Classification, Leaf Recognition, Machine Learning, Deep Learning

I. INTRODUCTION

India has a rich heritage of medicinal plants. Its conservation as the COVID-19 pandemic happened, the world is now preferring immunity boosting herbal medicines to fight against viruses and diseases. The term 'medicinal plants' refer to a variety of plants that have medicinal properties. The World Health Organisation (WHO) defines traditional medicinal plants as natural plant materials which are used at least or in the absence of industrial processing for the treatment of diseases at a local or regional scale [1]. India is home to over 8,000 plant species that are known for their medicinal properties. Over 80% of plants used in ayurvedic formulations are collected from the forests and wastelands whereas the remaining are cultivated in agricultural lands [2]. Some claimed that the plants are collected by women and children from forest areas; those are not professionally trained in identifying correct medicinal plants [3].

1.1 Background

Ayurveda is India's traditional system of medicine. The term comes from two Sanskrit words – Ayur, meaning life and Veda, meaning science or knowledge [4]. It is important for both Ayurveda practitioners and traditional botanists to know how to identify and classify the medicinal plants through computers. Medicinal plants are classified according to both their internal and external characteristics. Their outward features, in particular, play an important role in identifying them. According to the plant's taxonomy, we find classification of plants based on the shapes of their leaves and flowers [5]. But the most popular part of the plant that has been used for plant recognition is the leaf because the leaf is available throughout the year and it is easily visible. Flowers are only available during specific seasons, while roots are often difficult to access. Mostly leaves are the most abundantly available data in botanical reference collections. Professionals who are working together with botany identify plants through leaves identification [6].

However, due to numerous challenges such as loss in biodiversity, climate change issues and the difficulties to monitor plants on the field, taxonomists are looking for a cheaper, more convenient, and practical method to identify plant species. This explains the interest of researchers to automate the recognition process based on the morphological and visible characteristics of plants [7]. Automatic identification and classification of medicinal plants will provide medicinal knowledge to common people and farmers which helps in increasing production of such essential plants. This automated classification system supports botanists, consumers, forestry departments, taxonomists, pharmaceutical companies, and Ayurveda practitioners in identifying and classifying medicinal plants – without the need for human intervention [8].

It is a desire to have an automated plant identification system that helps users without specialized knowledge and in-depth training in botany and plant systematics to find out the information of some herbal plants by taking pictures of the plants to feed into an automated plant recognition system [9]. Accurately identifying medicinal leaves helps botanists, taxonomists, and drug manufacturers produce high-quality medicines and minimize the risk of side effects from incorrect treatments [3]. Using the wrong plant can directly impact the effectiveness of a medicine and may cause side effects for consumers. To prevent this, it's essential to follow strict and accurate methods of plant identification [4].

Research in the field of Indian ayurvedic medicinal plant identification mainly focuses on single plant organ (leaf). Second, the researchers are creating custom dataset for their research as there is no standard dataset available for Indian medicinal plant organs [10]. Also, not to get confused with classification, some authors have defined 'classification' as, "Texture based image classification involves deciding the most pertinent texture category of the observed image. Classification here would mean the machine language classification of the various classes and not the Linnaean Taxonomic System of plant classification. When the prior knowledge of the established classes are available and the texture features are extracted, the given image could be classified to the appropriate class." [11].

In the following section explains the universal model for plant identification and classification using various techniques of Data Mining like Machine Learning, Deep Learning, Computer Vision, etc. followed by an overview of the related work in this field.

1.2 Standard Procedure

The following diagram shows the basic plant recognition techniques used by many researchers. It consists of image acquisition, preprocessing, feature extraction, classification, and the obtained results.

- First, plants images are gathered from a pre-existing data set or by using HD cameras or even mobile phone cameras. These gathered images form a data set for the research.
- Second, the pre-processing is done for enhancing the image quality, removing background noise, filtering, segmenting, etc. This step is optional, as it depends upon the researchers whether to opt for pre-processing or not.
- Third, features are extracted using different techniques. Features include shape, texture, color, edge, morphology, geometry, etc.
- Fourth, the different classifiers classify the medicinal plants into respective classes.
- And lastly, the plant is recognized, that is, it is identified and classified accordingly.

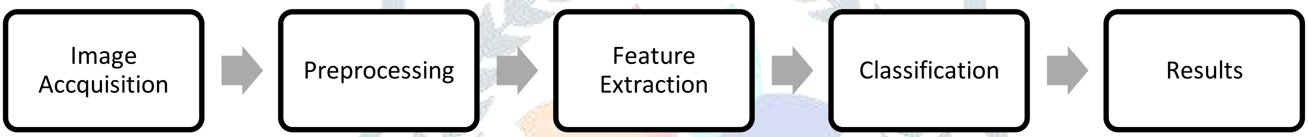

Figure 1 Standard procedure for plant identification and classification

II. LITERATURE SURVEY

The primary theme in this literature survey includes 'feature extraction' to identify and classify medicinal plants. The leaf is the most commonly used plant part as it is available throughout the year. Features like morphology, colour, texture, and area of leaf are mostly used to extract features for plant identification. Let us see these one by one.

2.1 Texture Features

Rad et al., describes how they used Image Processing technique to classify plants. The GLCM and PCA algorithms were used for leaf recognition. Their custom data set from Mysore consist of 13 plant species. The trained the algorithm on 390 images and tested it with 65 images. They achieved accuracy of 98% with PCA compared to 78% with GLCM. As they focused on specific plant species, in future, they expect the research would grow with bigger data set and other methods for comparison [12].

Here **Rashad et al.,** used CCLVQ for plant classification based on the characteristics of texture properties. They combined LVQ and RBF and achieved higher performance compare to other methods. They randomly customised data set with 128x128 resolution of images. They compared various methods like PCA, KNN, RBPNN, BPNN, SVM, etc. Their algorithm achieved 98.7%. Their work needs further optimisation of factors like number of epos. In future they will apply SPR methods which will take noise into consideration. They mentioned their work is limited to texture features and not on the colour [13].

FICUS is one of the largest genera in plant kingdom. To identify three species of FICUS which has similar leaf morphology, **Jye et al.,** developed an efficient baseline automated system using image processing and pattern recognition. They collected 54 leaf image samples from Malaysia, and also made the dataset available at https://data.mendeley.com/datasets/tvw4gy5ywy/draft?a=67c0cb84-80cb-4b41-a19d-38e29c3141b9. They extracted shape features, HU moment invariants, leaf texture features using GLCM analysis and HOG for leaf image analysis. They used ANN and SVM models and evaluated results with recognition accuracy of 83.3%. They said their system will not replace human taxonomic, but will provide a rapid and easy technique for plant identification. In future, they want to include more species of FIUS in order to improve system's robustness [14].

Authors **Ibrahim et al.,** did a comparative study on HOG, LBP, SURF, with SVM for identifying herbal plants of Malaysia. They investigated the application of texture features for leaf recognition by constructing a customized leaf data set from 10 different herbal plants. They also used Flavia dataset. The authors achieved up to 99% accuracy from HOG and LBP classifier on customised datasets and 97% accuracy on Flavia dataset. But these two outperformed SURF which achieved 74% accuracy and 63% accuracy on customised dataset and Flavia datasets respectively [6].

Arun et al., came up with an automated system for recognition of medicinal plant leaves and they also compared the different classifiers. They first collected the dataset of 5 species comprising of 250 leaf images. They used grey tone texture feature extraction method by using operators like GTSDM & LBP. They achieved 94.7% accuracy rate. They concluded that preprocessing is not required instead directly applying feature extraction method gives better results. They calculated recognition and error rate & observed if in future they increased features then accuracy will also increase [16].

2.2 Shape Features

In order to identify is 'shape' the key feature for automatic plant identification. **Jamil et al.,** used SIFT for shape feature extraction, colour moments for colour feature extraction and SFTA for texture feature extraction. They performed this experiment on 455 herbal medicinal plant leaves, they collected leaf plant images from Malaysia. They took 7 common herb species with 5 different leaves of each species and applied overall 13 transformations to it. Prior to plant leaf identification, two main processes were involved, the feature extraction and the training of Aboost classifier. They conducted seven experiments, to their surprise texture was the most discriminative feature, with single texture feature outperforming colour or shape with 92% identification rate but the fusion of all three achieved highest identification rate of 94%. Their experiment was limited to small collection of Malaysian medicinal plants and their future work is to apply same methodology on publicly available image data set and other feature extraction methods to verify results [17].

Zhao et al., proposed counting based shape descriptor for plant identification using shapes of the leaves. Their proposed novel feature captures biological information like overall shape, margin, type, leaflet, geometry, and arrangement. Their approach of I-IDSC feature based was experimented on five data sets, namely Swedish, ICL, Smith-Sonian, Plumbers-Island and their own customised data set, which consisted of 279 leaf images from 54 species from Hong Kong. Experimented results show their method performed better than state of the art methods. They also develop a demo system. In future they aim to include leaf texture and other features [18].

Yang et al., presented a new method for recognition of medicinal plant leaf by integrating shape and texture features. They proposed MTD for shape and LBP–HF for texture features extraction. They evaluated their approach on standard data sets like Flavia, Swedish and MEW2012 achieving accuracies 77.6%, 85.7% and 67.5% respectively. As this paper is limited to shape and texture, they insisted on adding more features and models like deep neural network to automate recognition process [19].

MedLeaf - An android mobile application was developed by **Prasvita et al.,** the application helps in identification of medicinal plants based on leaf images. LBPV was used as feature extractor and PNN was used as image classifier. They created their own dataset consisting of 1440 leaf images for 30 medicinal plant species. As the data set was created by capturing images from smartphone camera in Indonesia, the achieved accuracy was just 56.33%. A questionnaire was also used to evaluate satisfaction of the users which showed 50% as satisfied, 35% as quite satisfied and 15% as not satisfied. Although their application does document searching, they need to enhance their data set for better performance in future [20].

Here **Putri et al.,** developed a system for identifying medicinal plant leaves in Indonesia, using CNN (DL) by implementing back-propagation. The training dataset consisted of 180 leaves in total with 20 leaves of each 9 types of medicinal plants. The system successfully identified medicinal plants into 9 classes. Their system needs further validation & testing on larger datasets and improving the system performance [21].

2.3 Color Features

Dyrmann et al., represented a method using CNN for recognition of plant species in colour images. They built the entire network from scratch, training and testing with 10,000 images of 22 weed and crop species and achieving classification accuracy of 86.2%. All these images were of six different data sets. The authors mentioned, the classes with more species gave highest classification accuracy compared to the classes with few image samples [22].

Authors **Jeon et al.,** employed CNN Model and Google net for leaf classification on basis of leaf characteristics, such as texture, colour, contour. They experimented on Flavia data set and achieved recognition rate of 94% even with damaged leaves up to 30%. They aim to recognise leaves attached to branches for further system development [23].

The main objective of **Arun et al.,** was to identify medicinal plant leaves using textures and optimal colour spaces. The methodology involved the extraction of colour texture features using various colour spaces based on grey level, GTSDM and LBP operators. The research was also focused on utilising various classifiers like SGD, KNN, SVM–RBF, LDA and QDA for accurate classification of plant image based on the extracted features. The authors used a customised data set of 250 leaf images belonging to 5 different species of plants which are collected around the Western Ghats region of Karnataka Forest. They achieved identification rate of 98.7%. Their work is limited as their method works only with matured leaves of the plant [11].

Sivaranjani et al., used machine learning technique for real-time identification of medicinal parts. They created data set of 5 classes of medicinal plants with 20 images of each class. Their proposed work includes image segmentation, feature extraction and classification. They used RGB coloured image in ExG–ExR method, with WEKA classifier. They extracted features like colour and texture. They used LR classify in WEKA and attained accuracy of 93.3%. They aim to develop an automated system which identifies any plant species by analysing any part of the plant [15].

Vo et al., from Vietnam presented a herbal plant recognition system by using Deep Convolutional features. They also presented first herbal plant image dataset from Vietnam comprising of 10,000 images of 10 herbal plant species. This dataset is collected by capturing images using a mobile phone in natural daylight. They used 7 different classification methods and achieved highest accuracy of 93.6% with LightGBM model. As the work is limited to Vietnamese herbal plants, the authors aim in expanding the dataset and they also aim to enhance the performance of the model and compare LightGBM and other state of the art techniques [9].

Authors **Anami et al.,** combined colour, edge, and texture features to identify and classify 900 medicinal plants. They classified plants as herbs, shrubs and trees by using SVM and NN classifiers (RBENN). Individual accuracy achieved based on colour feature extraction is 74% whereas on edge texture feature extraction is 80%. But when combined the accuracy rate is improved to 90%. The authors concluded they achieved better results on tree compare to herbs and shrubs, so they came to work on it in near future. And they also concluded as mostly all leaves have green colour, classifying them on the basis of colour histogram does not give higher accuracy [5].

2.4 Morphological Features

Prasvita et al., first created a database by scanning front and back side of leaves of Ayurvedic medicinal plants. They collected 20 leaves each of 14 different plants species. Then the authors used features like morphology, colour, texture, Zernike movements, HU invariants and CR distance for plant identification. Then they applied classifiers like MLP, SVM, NB, K-star, FLR, KNN in WEKA classifier tool. As they identified both green and dry leaves, they achieved identification rate up to 99% and 94.5% respectively. They used 14 different classifiers in WEKA tool. They admit this method is not suitable for tiny leaves of plants. So they hope the future work would be to include an exhaustive database, add more features and to identify tiny leaves [2].

Abdollahi et al., built a medicinal plant data set of 30 different classes of plants, compiling of 3000 pictures in Ardabil, Iran. The database is accessible at https://github.com/Jafar-Abdollahi/medicinal-plants. Then he developed a model using CNN for identifying medicinal plant leaves later to be implemented in mobile based software. He extracted features like morphology, texture, and shape. He pre-processed the data, did data augmentation, trained CNN using transfer learning, extracted the features and then classified the medicinal plants. The performance matrix showed the model attained accuracy of 98.5%. The authors intent to expand the research by increasing the data set and developing deep learning (DL) model with comprehensive software [24].

Rao et al., proposed a method for identification of medicinal plants from images of front and back of leaves. They extracted unique features of texture, shape, and morphology. They used DenseNet type of CNN on self-collected database of 30 leaves of 50 different medicinal plant species. They implemented deep learning models by using epochs and back propagation. The entire software was developed in Python 3.9 version. Their work was limited to biggest size leaves. They aim to implement the algorithm on a standalone single board computer connected to a scanner [3].

Geerthana et al., exhibits how they developed medicinal plant identification system using deep learning. They focused on classifying 5 Indian medicinal plants, namely Pungai, Jamun, Jaterophar curas, Kuppa imeni & Basil. They utilised a ImageNet data set containing 58,280 images. They used features like leaf texture, shape, colour, physiology, and morphology. To achieve this a CNN classifier was used which gave 96.67% accuracy. Their studies limited to four classes only. They aim to increase the classes and solving the overfitting problem as future work [25].

This is another paper where the authors **Thanikkal et al.,** developed an android application for identifying immunity boosting medicinal plants using a leaf shape descriptor algorithm and deep learning model. The system was tested on Flavia & Swedish databases and achieved 96% accuracy exclusive for 64x64 sized images. The author mentioned other present mobile application for leaf identification, and also compared their work with them. They aim to combine complex leaf features & enhances the database in future [4].

2.5 Geometrical Features

The methodology of this paper was focused on image processing technique for plant identification. The authors **Kumar et al.,** wanted to facilitate the recognition and prevention of medicinal plants through computer-based identification methods. Leaf images of 9 medicinal plants were used. Leaf features such as area, colour, histogram, and edge histogram were mostly utilised for plant identification. This proposed methodology showed successful identification of medicinal plants with some exceptions like Tulsi being wrongly identified as Mint. The authors mentioned the algorithm was restricted to mature leaf images and also require a white background for better accuracy. In future they expect of overcoming miss identification [26].

Authors **Zhang et al.,** developed a novel plant recognition method based on GLMMDP to enhance discriminant performance of training samples. They combined global and local information for improved classification performance. The GLMMDP was based upon MNMDP. The GLMMDP outperformed the other algorithms. SGLPP and GLMMDP were better than other combined methods. They used available ICL leaf data set yielding 95% recognition accuracy, and Leafsnap data set yielding more than 90% recognition accuracy. In future they hope to extend the method to semi supervised case for improving classification performance [27].

Here authors **Preethi et al.,** used machine learning techniques to identify and classify rare medicinal plants. The algorithms used were RF, KNN, LR for classifying rare medicinal plants from non-medicinal plants. They claim their method is the fastest identification method without any manual support. They gave pictorial representation of the propose system which included in putting a leaf image, then segmentation, then feature extraction, then classification, then testing and finally the plant leaf is identified, as either medicinal or non-medicinal. They mentioned their future work is to include variety of leaves, other parts of plants and using artificial neural network (ANN) for advance identification [28].

Leaves from 24 different medicinal plant species were collected and later photographed for automatic medicinal plant identification by **Begue et al.** Computer vision and machine learning techniques were applied for extracting leaf features like length, width, perimeter, area, number of vertices, area of hull and colour. They created first ever image data set on the Island of Mauritius and achieved accuracy of 90.1% using random forest classifier out of the five machine learning classifiers. To check performance of the system, WEKA machine learning workbench was used. They did not do any pre-processing but post processing operations were automatically performed. As a future work, they aim on implementing ANN and DLNN [29].

Authors **Kansara et al.,** followed a standard process for creation of image dataset of Indian Ayurvedic medicinal plant organs and also brilliantly developed a mobile based tool for acquiring structured plant organ images. Upon capturing, this tool stores the geographic location and time of the year. They mentioned there was not standard plant image organ dataset available. The standardisation process included dataset structure, naming convention and other parameters. The paper mentioned other plant image dataset which do not belong to India. This paper was published in 2021 and the authors promised releasing the dataset for future research [10].

Manoharan et al., used two stage authentication procedure by implementing image segmentation and machine learning classifier for detection of herbal plant leaves. Their method address and answer many drawbacks which are associated with state-of-the-art methods where only machine learning is used for herbal plant leaf detection. The two-stage authentication consist of edge based herbal plant detection in phase-I and classification based herbal plant detection in phase-II. The operators used in phase-I were Prewitt, Canny, Laplace and Sobel. Then in phase-II Chi-Square technique was used to extract features like colour, shape, length and width followed by CNN classifiers for classification of herbal plant leaves. They got 92% accuracy recognition rate but faced computational time and storage space problems. They aim to work in future on advance Digital Image Processing and statistical approach of Support Vector Machine [30].

2.6 Other Features

Achieving our recognition rate of 91.78% on self-acquired data set of 10,000 images of 100 ornamental plant species in Beijing. The authors **Sun et al.,** also designed a 26-layer deep learning model consisting of eight residual building blocks called ResNet26 for large scale plant identification. Their data set is named BJFU100 and is first plant image data set collected by mobile phone in natural environment. This data set is accessible at http://pan.baidu.com/s/1jILsypS. And they aim to improve by increasing the data set and deep learning model in future [31].

The authors **Suh et al.,** came up with the idea of transfer learning for classifying sugar beet and volunteer potato. They emphasised how transfer learning is better option than deep learning as it requires less amount of time for training the network. They evaluated transfer learning on AlexNet along with six network architectures. They used their customised ImageNet data set to pre-train the network. The highest classification accuracy achieved is 98.7%. Their future work would consist of implementing a complete pipeline for weed detection to enhance overall performance [32].

The authors **Pudaruth et al.,** from the Republic of Mauritius developed a mobile application for recognition and classification of 70 different medicinal plants. They used Tensor Flow framework as base for CNN classification model. Their application is called MedicPlant and it is an offline software. It does not involve any pre-processing steps. Any part of the plant can be photographed and the application software will classify it. This software was developed on deep learning architecture and was trained on 100 images of 70 different medicinal parts. This real-time mobile application gave 95% recognition accuracy and every time a plant is identified, it displays information like scientific name, English name, common name, etc. Their future work was to increase the data sets and provide more information of the plants [7].

Roopashree et al., develop data set called DeepHerb consisting of 2515 leaf images of 40 different species of Indian Ayurvedic herbs in different regions of Karnataka. They also develop a vision based automatic medicinal plant identification system. This paper also uses transfer learning in deep learning. The work was divided into four phases of data sampling, image pre-processing and segmentation, features extraction and classification. Feature extraction was done using CNN and they used models like VGG16, VGG19, InceptionV3 and Xception. Classification was done using machine learning techniques like ANN and SVM. There DeepHerb model was finally integrated into the cross platform HerbSnap mobile application, which is available in both android and iOS and was developed using flutter. They achieved 97.5% average accuracy. They also built TMH database. Their future is to increase both the data set and increase efficiency of the entire system [33].

Barradas et al., studied medicinal plants found in Borneo region. They created an automated real time identification system which composed of computer vision system training and testing the deep learning model, a knowledge-base (KB) that is the database for storing plant images, and a frontend mobile application for user interaction. They combined public and private datasets, PlantCLEF2015 & UBD Botanical Garden dataset with over 25,000 images and implement an Efficient Net-B1 based deep learning model. As the size of customised dataset was insufficient, the authors opted for transfer learning with ImagNet pre-trained weights. The android mobile application supports API level 16 and above. The user uploads an image and the web server retrieves a list of predictive species. The model also provides a feedback mechanism to report the misclassification of the plant species. The proposed model improved 10% accuracy as compared to other baseline models. But the performance was dropped when tested on actual samples, so they intended to improve the system's performance in near future [34].

An automated plant identification system of Vietnamese medicinal plants without using a dedicated database is presented by **Nguyen et al.,** they combined Deep Learning, Transfer Learning, and Crowd Sourcing. They proposed an approach which involved four steps stated as, plant organ detection, plant image collection, data validation and plant identification. The authors explicitly explained the benefits of using transfer learning and crowd sourcing. They used OrangeNet and VNPlantNet for organ detection and plant identification. They achieved accuracy of 87.18% and they aim to improve the performance by increasing the dataset [35].

Table 1 Overview of previous research

Sr No.	Features	Classifier	Accuracy Rates (%)	Dataset	Year of Publish	References
1	Color, edge, and texture features	SVM, RBENN	90.00%	Self - collected 900 images of medicinal plants	2010	[5]
2	Texture features	GLCM & PCA	GLCM = 78% & PCA = 98%	Self-collected dataset with 390 training images and 65 test images	2010	[12]
3	Texture features	LVQ + RBF	98.70%	Randomly took 30 blocks of each texture as a training set & another 30 blocks as a testing set	2011	[13]
4	Leaf area, color histogram, and edge histogram; boundary pattern and vein pattern	Canny edge detection algorithm	Out of 9 only 2 plants were wrongly identified	Self-collected dataset of 9 medicinal plant species	2012	[26]
5	Shape, color, texture	LBPV, PNN	56.33%	Self - collected 30 species of Indonesian medicinal plants consisting of 1440 leaf images	2013	[20]
6	Texture analysis with features such as grey textures, GTSDM & LBP	SGD, KNN, SVM, Decision Tree, Extra Trees, Random Forest	94.70%	Self - collected 250 leaf images from five species	2013	[16]
7	Shape; global & local shape information	I-IDSC	Proposed method outperformed other methods	Swedish, ICL, Smithsonian, Plumbers Island, and the researchers' own dataset of 279 leaf images	2015	[18]
8	Shape, color & texture features	SFTA, TTBD, SIFT, AdaBoost classifier	Single texture feature: 92%; Fusion of all three: 94%	Self - collected 455 herbal medicinal plant leaves dataset	2015	[17]
9	Shape & color	Deep CNN, SIFT, SURF, TLR	86.20%	Self - collected 10413 images containing 22 weed & crop species from 6 different dataset	2016	[22]

10	Geometric (morphology), Zernike moments, HU-Invariant, CR Distance, Centroid radii distances, color features, texture features	Multilayer Perceptron Neural Network Support Vector Machine Random Committee Random Forest Bagging Classifier J48 Tree Classifier	Green leaves = 99.00%; Dry leaves = 94.5%	20 leaves were collected in a random manner from 40 different plant species	2017	[2]
11	Length, width, perimeter, area, number of vertices, color, area of hull	Random Forest, MPNN, SVM, Naïve Bayes, KNN	90.10%	Self - collected dataset of 24 medicinal plant species	2017	[29]
12	Natural environment	ResNet26 model	91.78%	Self - collected BJFU100 dataset with 10,000 images of 100 plant species	2017	[31]
13	Leaf color, contour, texture & shape	CNN, GoogleNet, MMC, HOG, SIFT, softmax classifier	94%	Flavia	2017	[23]
14	Color texture analysis	SVM-RBF, SGD, KNN, LDA, QDA, GTSDM, LBP	98.70%	Self - collected dataset 250 leaf images belonging to five different species	2017	[11]
15	Texture features	HOG, LBP, SURF, SVM	HOG = 99%; SURF = 74%	Self - collected, Flavia	2018	[6]
16	Features extracted from FC6 and FC7	SVM, Random Forest, LDA,; comparison between AlexNet, VGG-19, GoogleNet, ResNet-50, ResNet-101, Inception-V3	98.70%	ImageNet dataset; Self - collected 1100 plant images dataset	2018	[32]
17	Shape, texture, histogram of oriented gradient	ANN, SVM, HOG, GLCM	83.30%	54 leaf image samples of 3 different species	2018	[14]
18	Color, texture	WEKA, ExG-ExR, Logistic Rrgression	93.30%	Dataset contains 5 classes of medicinal plants with 20 images each	2019	[15]

19	Shape, color, texture	Deep Learning, LightGBM, Random Forest, KNN, SVM, AdaBoost, Logistic Regression XGBoost	LightGBM = 93.6%; SVM = 90.8%	Self - collected Vietnamese herbal plant dataset with 10 species and 10,279 images	2019	[9]
20	Organ detection, plant identification	Deep Learning, Transfer Learning, Crowd Sourcing, OrangeNet, VnPlantNet	OrangeNet = 87.18%; VnPlantNet = 81.58%	No dedicated dataset; PlantClef2015, ImageNet	2019	[35]
21	Texture, shape, color, local & global structure	GLMMDP, MNMDP, SIFT, PNN	ICL dataset = 95%; Leafsnap dataset = 90%	ICL, Leafsnap	2020	[27]
22	Color, shape, area, perimeter of herbal plants and texture traits of leaves.	CNN, Derwitt, Canny Edge Detector, Sobel Edge Operator, Lapace & Proposed TSA algorithm	92%	250 leaf samples	2021	[30]
23	Picture of part of the plant or the whole plant	CNN based on TensorFlow; InceptionV3	95%	Self - collected thousands of images of seventy different medicinal plant species	2021	[7]
24	Features extracted from pretrained convolution layers using Transfer Learning	VGG16, VGG19, InceptionV3, and Xception models. ANN and SVM classifiers. DeepHerb model	97.50%	DeepHerb dataset with 2515 leaf images from 40 Indian herb species; TMH; ImageNet & Flavia dataset	2021	[33]
25	Leaf texture, shape, color, physiological, and morphological features	CNN, InceptionV3	96.67%	Self- collected dataset contains 58,280 images with 10,000 images per species; ImageNet dataset	2021	[25]
26	Color, size, texture & shape	CNN, ANN, KNN, LBP, SVM, MLP	highest accuracy achieved	Self - collected 9 classes of medicinal leaves datasets	2021	[21]
27	Geometric, texture, shape & color	Isolation Forest algorithm	88.79%	Creation of Indian medicinal plant organ image dataset considering 50 medicinal plants	2021	[10]

28	Shape & texture	MTD, LBP-HF	Flavia = 99.28%; Swedish = 98.4%; MEW2012 = 95.6%	Flavia, Swedish & MEW2012 leaf datasets	2021	[19]
29	Shapes, sizes, texture, edges of plants; area, eccentricity, major axis, minor axis, etc	Random forest, KNN, Logistic regression	highest accuracy achieved	Not mentioned	2022	[28]
30	Morphological features, texture & shape features	CNN, Mobile Net V2	98.05%	Dataset includes 30 classes of medicinal plants with 3000 photos	2022	[24]
31	Shape, color, texture & morphological features.	Densenet121, CNN	Not mentioned	Self-collected dataset includes 3777 images, 30 leaves of 50 different medicinal plant species	2022	[3]
32	Natural environment	EfficientNet-B1 deep learning model, CNN	87% to 84%	PlantCLEF 2015, UBD Botanical Garden, merged dataset	2022	[34]
33	Morphological shape features	Deep CNN, SDAMPI, HOG, SIFT, SURF	96.00%	Self - collected dataset with 1000 samples, Flavia & Swedish	2023	[4]

III. CONCLUSION

India is a rich source of medicinal plants. The traditional medicine systems such as Ayurveda, Siddha, and Unani are widely used in India. There is a constant need to identify and classify medicinal plants for pharmaceutical industries and to protect them from extinction. The automatic identification and classification of medicinal plants using Data Mining techniques like Machine Learning is a promising field with significant advancements in feature extraction, model development, and real-time applications. However, there are notable gaps in dataset availability, environmental robustness, and interdisciplinary collaboration. Many researchers have created their own datasets for experiments and research. The future research should focus on addressing these gaps to develop more comprehensive and practical solutions. We conclude, an automated medicinal plants identification system will be useful to common man, farmers, botanist, taxonomist, and researchers.

REFERENCES

[1] Jamshidi-Kia F, Lorigooini Z, Amini-Khoei H. Medicinal plants: past history and future perspective. *J HerbMed Pharmacol*. 2018;7(1):1–7. doi:10.15171/jhp.2018.01

[2] Prasvita DS, Herdiyeni Y. MedLeaf: Mobile application for medicinal plant identification based on leaf image. *Int J Adv Sci Eng Inf Technol*. 2013;3(2):103.

[3] Rao RU, Lahari MS, Sri KP, Srujana KY, Yaswanth D. Identification of medicinal plants using deep learning. *Int J Res Appl Sci Eng Technol*. 2022;10(4):306–322. doi:10.22214/ijraset.2022.41190

[4] Thanikkal JG, Dubey AK, Thomas MT. An efficient mobile application for identification of immunity boosting medicinal plants using shape descriptor algorithm. *Wirel Pers Commun*. 2023;131(2):1189–1205. doi:10.1007/s11277-023-10476-3

[5] Anami BS, Nandyal SS, Govardhan A. A combined color, texture, and edge features-based approach for identification and classification of Indian medicinal plants. *Int J Comput Appl*. 2010;6(12).

[6] Ibrahim Z, Sabri N, Mangshor NNA. Leaf recognition using texture features for herbal plant identification. *Indones J Electr Eng Comput Sci*. 2018;9(1):152–156. doi:10.11591/ijeecs.v9.i1.pp152-156

[7] Pudaruth S, Mahomoodally MF, Kissoon N, Chady F. Medicplant: A mobile application for the recognition of medicinal plants from the Republic of Mauritius using deep learning in real-time. *IAES Int J Artif Intell*. 2021;10(4):938–947. doi:10.11591/ijai.v10.i4.pp938-947

[8] Rajani S, Veena MN. Study on identification and classification of medicinal plants. *Int J Adv Sci Eng Technol*. 2018;6(2). [Internet]. Available from: http://iraj

[9] Vo AH, Dang HT, Nguyen BT, Pham VH. Vietnamese herbal plant recognition using deep convolutional features. *Int J Mach Learn Comput*. 2019;9(3):363–367. doi:10.18178/ijmlc.2019.9.3.811

[10] Kansara M, Parikh A. Indian Ayurvedic plant identification using multi-organ image analytics: Creation of image dataset of Indian medicinal plant organs. *3rd Int Conf Innov Comput Commun*. [Internet]. Available from: https://ssrn.com/abstract=3563074

[11] Arun CH, Durairaj DC. Identifying medicinal plant leaves using textures and optimal colour spaces channel. *J Ilmu Komput Inf*. 2017;10(1):19. doi:10.21609/jiki.v10i1.405

[12] Rad AE, Ehsanirad A. Plant classification based on leaf recognition. *Int J Comput Sci Inf Secur*. 2010;8(4). [Internet]. Available from: http://sites.google.com/site/ijcsis/

[13] Rashad MZ, el-Desouky BS, Khawasik MS. Plants images classification based on textural features using combined classifier. *Int J Comput Sci Inf Technol*. 2011;3(4):93–100. doi:10.5121/ijcsit.2011.3407

[14] Jye KS, Manickam S, Malek S, Mosleh M, Dhillon SK. Automated plant identification using artificial neural network and support vector machine. *Front Life Sci*. 2017;10(1):98–107. doi:10.1080/21553769.2017.1412361

[15] Sivaranjani C, Kalinathan L, Amutha R, Kathavarayan RS, Jegadish Kumar KJ. Real-time identification of medicinal plants using machine learning techniques. In: *Proceedings of the 2nd International Conference on Computational Intelligence in Data Science (ICCIDS 2019)*. 2019. doi:10.1109/ICCIDS.2019.8862126

[16] Arun CH, Emmanuel WRS, Durairaj DC. Texture feature extraction for identification of medicinal plants and comparison of different classifiers. *Int J Comput Appl*. 2012;62(12):1–9.

[17] Jamil N, Hussin NAC, Nordin S, Awang K. Automatic plant identification: Is shape the key feature? *Procedia Comput Sci*. 2015;76:436–442. doi:10.1016/j.procs.2015.12.287

[18] Zhao C, Chan SSF, Cham WK, Chu LM. Plant identification using leaf shapes - A pattern counting approach. *Pattern Recognit*. 2015;48(10):3203–3215. doi:10.1016/j.patcog.2015.04.004

[19] Yang C. Plant leaf recognition by integrating shape and texture features. *Pattern Recognit*. 2021;112:107809. doi:10.1016/j.patcog.2020.107809

[20] Prasvita DS, Herdiyeni Y. MedLeaf: Mobile application for medicinal plant identification based on leaf image. *Int J Adv Sci Eng Inf Technol*. 2013;3(2):103. doi:10.18517/ijaseit.3.2.287

[21] Putri YA, Djamal EC, Ilyas R. Identification of medicinal plant leaves using convolutional neural network. *J Phys Conf Ser*. 2021;1845(1). doi:10.1088/1742-6596/1845/1/012015

[22] Dyrmann M, Karstoft H, Midtiby HS. Plant species classification using deep convolutional neural network. *Biosyst Eng*. 2016;151:72–80. doi:10.1016/j.biosystemseng.2016.08.024

[23] Jeon WS, Rhee SY. Plant leaf recognition using a convolutional neural network. *Int J Fuzzy Log Intell Syst*. 2017;17(1):26–34. doi:10.5391/IJFIS.2017.17.1.26

[24] Abdollahi J. Identification of medicinal plants in Ardabil using deep learning. In: *Proc 27th Int Comput Conf, Comput Soc Iran (CSICC 2022)*. 2022. doi:10.1109/CSICC55295.2022.9780493

[25] Geerthana R, Nandhini P, Suriyakala R. Medicinal plant identification using deep learning. *Int Res J Adv Sci Hub (IRJASH) Special*. 2021. Available from: www.rspsciencehub.com

[26] Kumar SE. Leaf color, area and edge features based approach for identification of Indian medicinal plants. [Internet]. [n.d.]. Available from: [source not provided]

[27] Zhang S, Zhang C, Wang X. Plant species recognition based on global–local maximum margin discriminant projection. *Knowl-Based Syst*. 2020;200:105998. doi:10.1016/j.knosys.2020.105998

[28] Preethi S, Princy J, Gowda A, Priyadarshini A. Identification and classification of rare medicinal plants using machine learning techniques. *Int Res J Eng Technol*. 2022. Available from: www.irjet.net

[29] Begue A, Kowlessur V, Singh U, Mahomoodally F, Pudaruth S. Automatic recognition of medicinal plants using machine learning techniques. *Int J Adv Comput Sci Appl*. 2017;8(4). Available from: www.ijacsa.thesai.org

[30] Manoharan JS. Flawless detection of herbal plant leaf by machine learning classifier through two stage authentication procedure. *J Artif Intell Capsule Netw*. 2021;3(2):125–139. doi:10.36548/jaicn.2021.2.005

[31] Sun Y, Liu Y, Wang G, Zhang H. Deep learning for plant identification in natural environment. *Comput Intell Neurosci*. 2017. doi:10.1155/2017/7361042

[32] Suh HK, IJsselmuiden J, Hofstee JW, van Henten EJ. Transfer learning for the classification of sugar beet and volunteer potato under field conditions. *Biosyst Eng*. 2018;174:50–65. doi:10.1016/j.biosystemseng.2018.06.017

[33] Roopashree S, Anitha J. DeepHerb: A vision-based system for medicinal plants using Xception features. *IEEE Access*. 2021;9:135927–135941. doi:10.1109/ACCESS.2021.3116207

[34] Barradas A, Marques Da Silva J, Mariano P, Ahn TH, Koubouris G, Malik OA, et al. Automated real-time identification of medicinal plants species in natural environment using deep learning models – A case study from Borneo region. *Plants*. 2022. doi:10.3390/plants

[35] Nguyen TTN, Le TL, Vu H, Hoang VS. Towards an automatic plant identification system without dedicated dataset. *Int J Mach Learn Comput*. 2019;9(1):26–34. doi:10.18178/ijmlc.2019.9.1.761