ARTICLE IN PRESS

[Short title + Author Name - P&H title] xxx (xxxx) xxx

HOSTED BY

Contents lists available at ScienceDirect

Journal of the Saudi Society of Agricultural Sciences

journal homepage: www.sciencedirect.com

Medicinal plant species detection by comparison review

Beulah Jabaseeli N.[a,*], Umanandhini D.[a]

[a] Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Tamilnadu, India

ARTICLE INFO

Keywords:
Therapeutic plant
Ayurvedic
Pharmaceutical
Hygiene
Illnesses
Gastrointestinal
Excretory

ABSTRACT

The detection of therapeutic plant species is critical in today's environment. Though many valuable medicinal plants are essential for survival, everyone needs Ready-to-use chemical medicines instead of ayurvedic ones. Because herbal medicines are prepared on time, they cannot be stored with preservatives like pharmaceutical ones. Also, identifying correct medicinal plants is essential; it can be achieved by classifying herbs by leaf shape, colour and texture of both sides using the established parameters of various methods. Recent utilization of traditional medicinal plants has increased, but identifying correct medicinal plants is more complicated. All herbal plants are natural resources that are reliable and healthy for consumption. Medicinal plants are used to treat heart problems, respiratory problems, reproductive problems, gastrointestinal problems, joint problems, skin illnesses, excretory problems, etc. Because of the necessity for mass manufacturing, identifying these plants is urgent. Detecting and classifying medicinal species is more critical to provide better treatment. Due to the deficiency of knowledgeable persons, effective detection and categorization of edible herbal plants are difficult. As a result, a completely computerized approach to identifying edible herbal plants is an essential key factor of the research. In this paper, various plant detection techniques like Convolution Neural Network (CNN), pertained models VGG16, K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), Support Vector Machine (SVM) and Multi-Layer Perception (MLP) are taken and compared, and then finally concluded with the best results based on accuracy rate.

1. Introduction

Most valuable plants are becoming endangered and threatened due to the lack of awareness and environmental issues. Employ cutting-edge methods for precise identification, such as image processing and DNA barcoding, to preserve threatened medicinal plants. Create customized growing procedures that regulate environmental variables and imitate the plant's native habitat. Create botanical gardens and seed banks with the help of the local community to guarantee their preservation in their natural habitats. Through DNA barcoding and sophisticated imaging techniques, modern science and technology improve the identification and conservation of endangered medicinal plants. Accurate multiplication of species and their preservation for future generations are guaranteed by modern cultivation techniques such as tissue culture and controlled environment agriculture. Identifying correct and exact medicinal plants is quite difficult for humans, though the knowledge of Ayurveda is inherited from our elders by oral or general notes. So, recalling and identifying all medicinal plants is difficult because each has unique and identical features. If we choose incorrect herbs, then it will create some wrong impressions on ayurvedic medicines and rarely may cause some side effects. The transition from traditional Ayurvedic knowledge and practice to chemical medications has resulted in a change in perception, a decrease in practice, and an economic impact on the cultivation of medicinal plants. The dwindling usage of traditional plant-based treatments puts them at risk of disappearing. Conventional approaches are taught to fewer practitioners. Because of this problem, a vital requirement arises to develop a computerized detection and categorization of medicinal plants for the better benefit of future generations. It will educate and give essential information to all people about medicinal plants of all aspects. Computerized identification and categorizations of medicinal plants requires reference systems and databases. They store pictures, morphological data, chemical compositions, and conventional knowledge to allow algorithms to compare characteristics from fresh samples. This advances both theoretical and practical aspects of herbal medicine research. Hybrid databases can effectively integrate traditional knowledge of medicinal plants with contemporary identification procedures by adding ethnobotanical data and working with conventional healers on research initiatives to evaluate and record

* Corresponding author.
E-mail addresses: nbeulahajabaseeli@veltech.edu.in (N. Beulah Jabaseeli), drumanandhini@veltech.edu.in (D. Umanandhini).

https://doi.org/10.1016/j.jssas.2024.09.005
Received 15 July 2024; Received in revised form 3 September 2024; Accepted 16 September 2024
Available online 24 September 2024

their expertise using contemporary scientific methodologies.

The plant leaf plays a significant part in plant identification since it is freely available throughout the year (Azadnia and Kheiralipour, 2021). When distinguishing between species with similar forms but varied colourations, colour, texture, and vein pattern traits are crucial. Texture features are retrieved using statistical techniques such as co-occurrence matrices, whereas colour features are computed by summing up the distribution, mean, and variance. Leaf skeletonization is used to extract details from the vein pattern. Plants are identified by leaf factors like vein, colour, structure, edges and length structures (Yousefi et al., 2017) (Lukic et al., 2017). Several algorithms have been proven to discover accurate medicinal plants using only their leaves, flowers, roots, and bark. Still, among the various components of the plants, the leaf plays a critical role (Fekri-Ershad, 2020)(Zandi, 2017). As a result, we considered this and conducted a comparative assessment of how medicinal plants are discovered from their leaves by describing sufficient traits with the highest accuracy rate. As a result, this research compares and contrasts various Machine Learning techniques for improved detection of accurate medicinal plants.

2. Survey of species detection

Ayurveda is a traditional Indian medical system that dates back to the Vedic period, some 5000 years ago(Rajani and Veena, 2018). Rather than a Plant's leaf, other parts like stems, branches, and roots are the most critical factors for herbal medicines. The therapeutic qualities of plant stems, branches, and roots are the main emphasis of traditional Indian medicine, especially in Ayurveda. Bioactive substances, including alkaloids, flavonoids, tannins, and glycosides, are present in these sections and have medicinal properties. While stems and branches maintain and improve the body's structural integrity, roots store and transport nutrients. To increase the plant's overall efficacy, these portions are utilized in decoctions, powders, and extracts to cure various ailments. The plan seeks to improve public awareness of the advantages of herbal medicines, standardize formulations, develop user-friendly formats, and create regulatory frameworks to guarantee quality, safety, and efficacy. Ultimately, it hopes to foster consumer confidence and incorporate herbal medicines into contemporary healthcare practices. Over 8000 medicinal plants from India have been discovered. Herbal remedies from various Indian systems use combinations of a tiny subset of these plants, totalling 1500 (Gavhale and Thakare, 2020). Various Chronic diseases like cardiac-related diseases, respiratory-related diseases, fertility-related diseases, gastrointestinal diseases, bone issues, skin diseases, excretory disorders, and different diseases are treated with medicinal herbs. The medicinal plants Adhatoda and Tulsi are used to treat skin diseases, respiratory disorders, chronic illnesses, and problems with fertility. These benefits stem from their bioactive constituents, anti-inflammatory qualities, adaptogenic and reproductive health-boosting qualities, hormone balance, and enhancement of female reproductive health. As a result, adequate detection and preservation of medicinal plants must be considered (Figs. 1–7).

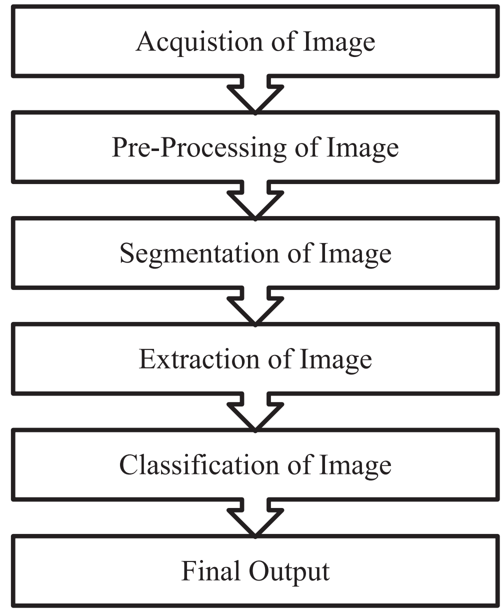

Fig. 1. Image Processing Steps.

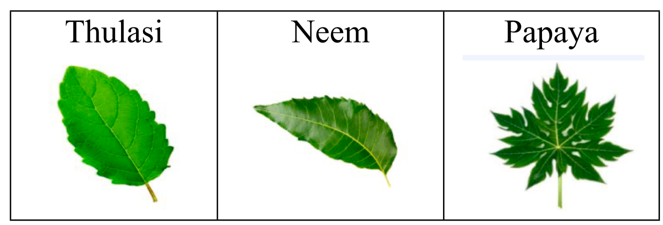

Fig. 2. Sample leaves image (front side).

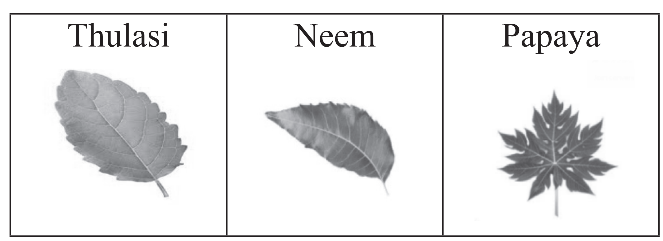

Fig. 3. Gray-scale image.

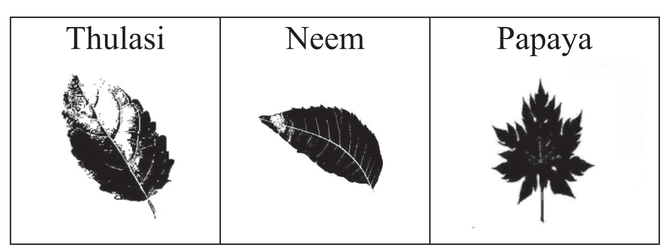

Fig. 4. Binary image.

3. Issues on plant species detection

- Manual identification takes time and can be approximate or inaccurate (Pushpanathan, 2021).
- Most of these therapeutic plants are also found in regions where human beings cannot access them with their hands.(Raghukumar and Narayanan, 2020)
- Edible/ poison plant species detection is also problematic (Priya et al., 2020). There are drawbacks to the conventional approaches to plant species identification, such as manual identification and machine learning models like KNN and SVM. Large datasets and high-dimensional data render KNN inefficient, but SVM exhibits more robustness at the cost of processing power. Because of the possibility of misidentification and the requirement for quick identification and categorization in massive datasets, computerized identification of edible herbal plants is crucial. Due to their accuracy and feature extraction, CNNs are perfect for identifying different plant species; nevertheless, they need a lot of labelled data for training and a substantial amount of processing power. Insufficient data can also cause them to overfit. Regarding large or complicated datasets, KNNs

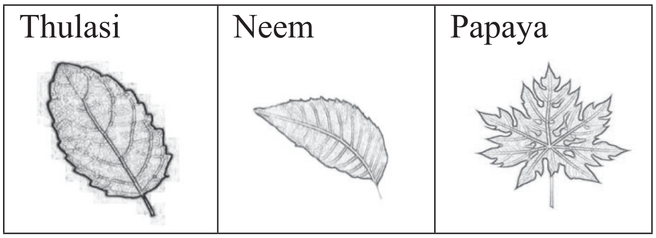

Fig. 5. Image outline.

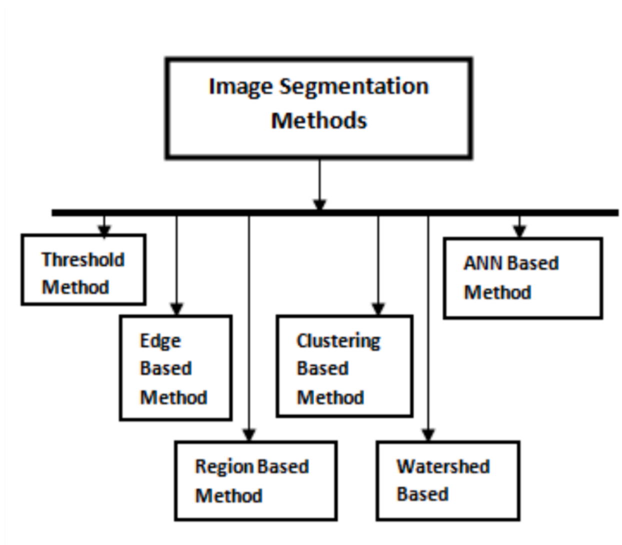

Fig. 6. Image Segmentation Methods.

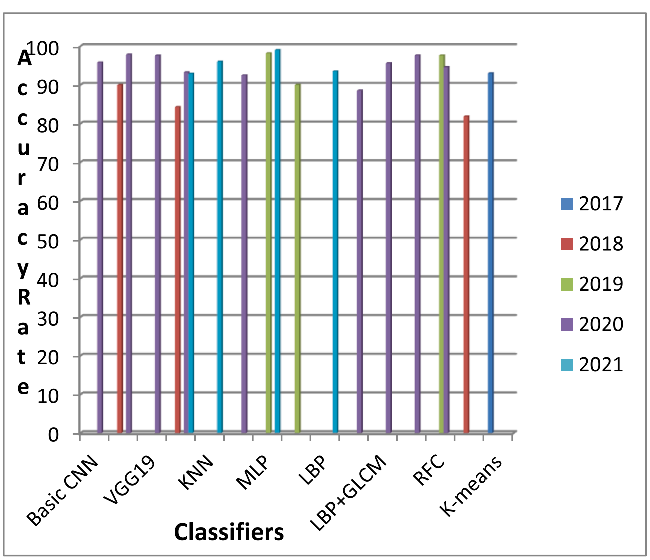

Fig. 7. Bar-Chart-Accuracy Rate.

are less effective and more straightforward, whereas SVMs perform better on smaller datasets but need more precise tuning. Because they reduce human error, handle enormous datasets rapidly, handle large-scale studies, and speed up research by analyzing and comparing data from various sources, computerized approaches increase the accuracy of identifying edible herbal plants.

- Not able to maintain proper documentation from manuscripts, ancestors and Botanist regarding medicinal plants(Sujith and Aji, 2020)
- Plant disease detection and pest range evaluation, among other things, are complex manual tasks in the field of agricultural plant protection (Li et al., 2021).
- Humans have a hard time identifying the right therapeutic herbs (Akter et al., 2020).
- Some people will find the process laborious because they are unfamiliar with these plants (Amuthalingeswaran, 2019).
- Various existing plant leaf approaches are standardized and followed by some scale format to recognize them, which generates poor results (Hu, 2018).
- Difficult to find out the exact region of the plant(Sachar and Kumar, 2021)
- The current rate of species extinction has prompted a slew of efforts to safeguard and maintain biodiversity (Wäldchen, 2018).

The three primary methods for processing images are outline extraction, binary image extraction, and grey-scale conversion. Greyscale conversion simplifies photos, extracts plant characteristics from binary images, and detects plant forms and edges via outline extraction. We have taken a leaf as a sample to detect the correct medicinal plants and followed the following steps (Tables 1–3).

3.1. Acquisition of image

It is the act of retrieving an image/capturing an unprocessed leaf image by external sources like cameras, sensors, handheld computer devices, and other datasets. The image transmission protocol guarantees that data is sent without loss or corruption and is essential for preserving the integrity of leaf pictures throughout the transfer. Methods like error detection, compression, and secure protocols reduce noise and maintain image quality, particularly in distant or cloud-based systems. While collecting raw images, the sample size, illumination of light effects, and

Table 1
Segmentation Methods Analysis.

S/No	Segmentation Method	Function	Merits	Demerits
1	Thresholding Method	Histograms of images are used to detect peak values and find related pixels.	Not difficult to use, didn't required costlier pre-processing	Threshold-related mistakes are acceptable. Some timings essential elements may be deleted.
2	Edge Based Method	Discontinuity detection is adapted rather than similarity detection for detecting colour, brightness, contrast, saturation	It's best for images with more contrast between objects.	Noisy images are not recommended
3	Region-Based Method	based on dividing a picture into homogenous areas	Noisy images are acceptable	Takes more time and memory
4	Clustering Based Method	Splitting images into k numbers of homogenous pixels to form objects	Enhanced for fuzzy logic.	Difficult for cost function determination for minimization.
5	Watershed Method	based on image boundary topological interpretation	Segments are stabilized for detecting borders.	Ridge gradient calculation is difficult
6	Neural Networks	based on deep learning algorithms – CNN	Uses regular Python ready-made libraries	It takes time and money to train the model for custom and business images

Table 2
Comparative Analysis.

S/No	Paper	Year	Objective	Algorithm	Novelty	Evaluation Metric	Accuracy
1	(Paulson and Ravishankar, 2020)	2020	For recognizing and classifying Images	Basic CNN	Reduce computation without losing the essence of data	Texture	95.79
2	(Paulson and Ravishankar, 2020)	2020	Using the lowest inception filter of 3x3 convolution filters can achieve superior performance.	VGG16	High potentials in classifications.	Shape, Color, Texture	97.8
3	(Abas and Haqmi, , 2018)	2018	VGGnet, a CNN version, was trained to identify and classify the plants using flower images.	VGGnet (VGG16)	Works on a combination of fine-tuning and transfer learning of VGG16 to recognize plant species from flower images	Flower	89.96
4	(Paulson and Ravishankar, 2020)	2020	Focusing on deep layers from 16 to 19 is a common way to achieve prior-art setups.	VGG19	higher accuracy, faster training speed, and fewer training samples each time	Shape, Color, Texture	97.6
5	(Raghukumar and Narayanan, 2020)	2020	Hyperplanes separate every class to find the exact location.	SVM	Ability to generalize well	Shape and colour	93.23
6	(Dahigaonkar and Kalyane, 2018)	2018	Identify hyperplanes that find data points in N-dimensional space.	SVM	Both classification and regression are possible	Shape, Color, Texture	84.25
7	(Sachar and Kumar, 2021)	2021	Least Square Support vectorMachine (LSSVM) Sparseness necessitates the use of pruning procedures.	LS-SVM	Fast and not complex	Shape, Color, Texture	92.88
8	(Thella and Ulagamuthalvi, 2021)	2021	The labels will be classified based on the neighbour's samples.	KNN	There is no need for much training.	Texture	96
9	(Huixian, 2020)	2020	A non-linear projection is created by connecting various fundamental units. Mainly used in pattern recognition	ANN	Less formal statistical training is required.	Shape and Texture	92.47
10	(Qadri, 2019)	2019	MLPs categorize the input signals split as space	MLP	applied for non-linear problems	Spatial texture, Gray Level Run-Length Matrix (GLRLM)	**98.14**
11	(Naeem, 2021)	2021	Adapting fused aspects used to classify medicinal plants by Machine learning techniques.	MLP	Identify the correct medicinal plants	Multispectral + Texture Features	**99.01**
12	(Qadri, 2019)	2019	It is a radial base network	PNN	Provide precise target probability scores that are projected.	Digital Morphological Features (DMFs)	90
13	(Naeem, 2021)	2021	Used for invariant classification of rotational texture.	LBP	resistance to monotonous gray-scale shifts	Texture	93.5
14	(Sujith and Aji, 2020)	2020	Gradient (vertical and Horizontal images) are used for calculating histograms in radians	LBP+HOG	Combined principles of LBP and HOG	Shape and Texture	88.49
15	(Sujith and Aji, 2020)	2020	Used in feature extraction by calculating spatial relationships in pixels with exact value	LBP+GLCM	Combined principles of LBP and GLCM	Shape and Texture	95.57
16	(Sujith and Aji, 2020)	2020	Combination of LBP, HOG and GLCM	LBP+HOG+GLCM	Gives better results than LBP	Shape and Texture	97.63
17	(Gavhale and Thakare, 2020)	2020	The technique used features and geometrical aspects to identify the correct medicinal plant species.	RFC	Both categorical and continuous values perform well	Colour, Texture, Geometric	94.54
18	(Pacifico, 2019)	2019	Used for the automatic classification system in case of performance and speed	RFC	Improve decision tree accuracy by reducing overfitting.	Colour and Texture	97.61
19	(Tan, et al., 2018)	2018	Used to predict the class of unknown samples by the statistical classifier.	NB	With the amount of predictors and data points, it is highly scalable.	vein	81.86
20	(Lukic et al., 2017)	2017	Grouped unlabeled datasets into k unique clusters that are combined by identical attributes.	K-means	Adapts for size and shapes to form a cluster	Shape, Color, Texture	93

colour factors are to be considered for less information loss (Azadnia and Kheiralipour, 2021). Raw leaf photos for analysis depend highly on lighting conditions during image acquisition. While uneven illumination can change colour and obfuscate texture and vein patterns, it can also cause erroneous feature extraction. Low light levels can increase noise, which lowers the efficacy of feature identification algorithms. Poor lighting can cause a loss of clarity in leaf edges and tiny structures.

3.2. Pre-Processing of image

Once the leaf image is collected, it is given to the pre-processing layer. The noise source will arise along with the acquisition and transmission. During image transmission, some environmental factors like CCTV cameras, light levels, and sensor temperature can corrupt the quality of an image, leading to increased noise levels in the resulting image. Control lighting, use consistent, non-reflective backdrops, ensure leaves are clean and clear of contaminants, and utilize tripods or image stabilization equipment to minimize environmental influence by preventing movement-induced blurring in leaf images. Because leaf images might have a wide range of shapes and sizes, various angles and vein factors must be processed initially. By forcing particular criteria for normalizing these elements, the classification of medicinal plants is done by pre-processing techniques (Chaki et al., 2020). The pre-processing layer is divided into four steps (grey-scale conversion, binary image extraction, outline extraction, and filtering result (Paulson and Ravishankar, 2020).

Table 3
Year-Wise Analysis.

Classifier	2017	2018	2019	2020	2021
Basic CNN				95.79	
VGG16		89.96		97.8	
VGG19				97.6	
SVM		84.25		93.23	92.88
KNN					96
ANN				92.47	
MLP			**98.14**		**99.01**
PNN			90		
LBP					93.5
LBP+HOG				88.49	
LBP+GLCM				95.57	
LBP+HOG+GLCM				97.63	
RFC			97.61	94.54	
NB		81.86			
K-means	93				

3.3. Gray-scale conversion

Three pictures can be seen in one RGB image (RED, GREEN, BLUE SCALE IMAGE). Whereas Gray-scale images are single-layered images. As a result, grey-scale images require less memory to store, and segmentation is easier with grey-scale images (Chaki, 2019). Grey-scale conversion keeps important information while simplifying visual data. Better analysis is made possible by binary image extraction, which separates items of interest from the background. Plant shape and structure are highlighted by outline extraction, also known as edge detection, which locates object boundaries. These methods are essential for identifying and categorizing species according to their physical characteristics.

3.4. Binary image conversion

Binary images can be created by thresholding a grey image. A grey image has only a pair of intensities per pixel, like black and white. The intensity of white ranges from 1 to 255, and the intensity of black is 0. These values are referred to as the thresholding value, and they are used to determine masking (Das, 2020).

3.5. Outline extraction

The outline of an image shows its boundaries; it is used to identify the leaf effectively. Edges play an important role in identifying the correct leaf image by referring to its vein arrangement and shape (Paulson and Ravishankar, 2020).

After this extraction, the image is allowed for noise filtering. During image capture, noise is added due to various environmental factors, affecting the image's processing quality. Noise reduction is essential before moving on to the next step. We can employ averaging low pass or Gaussian filters depending on whether the noise is linear or non-linear. Averaging filters are recommended for reducing noise from photographs, and averaging filters and median filters are helpful for salt and pepper noise reduction. By substituting each pixel's median value for its neighbouring pixels, maintaining edges, eliminating noise, and maintaining picture integrity, median filters successfully minimize salt and pepper noise. While median filters are better at reducing noise, average filters distort the image and lose information.

4. Segmentation of image

Segmentation of an image is the division of the digitalized image into smaller groups to simplify the analysis and process of the image. Labeling pixels is the process of segmentation (Tan, et al., 2018). There are several methods for image segmentation, as follows:

4.1. Similarity approach (Region Approach)

Analyzing the matching between the pixels of images using a threshold to create a segment. Image segmentation is mainly used for machine learning algorithms like clustering.

4.2. Discontinuity approach (Boundary Approach)

The image's pixel intensity levels are discontinuous. Line, Edge, and Point detection techniques use this kind of approach to obtain intermediate segmentation results that can subsequently be processed to get the final segmented image.

Making an image easy to examine is the goal of image segmentation. An image can be divided into segments with similar features using image segmentation. The split portions of the image are called image Objects. The following are some image segmentation techniques (Pukhrambam and Rathna, 2021);

Two techniques are employed in plant identification: the Similarity Approach and the Discontinuity Approach. The Similarity Approach combines pixels with similar characteristics, such as colour, intensity, or texture, to separate plants from their backdrop. When distinguishing between various species, the Discontinuity Approach is essential for spotting edges with notable variations in colour or intensity. Segmenting an image into meaningful sections corresponding to distinct items or components of an entity is the basic idea behind segmentation techniques. Segmentation is used in medicinal plant identification to separate the plant from its surroundings or its many anatomical components for additional examination. Ensuring the accuracy and dependability of the following feature extraction and categorization depends on this stage.

5. Feature extraction of image

The most essential elements retrieved from an image are used to create leaf recognition models. Discovering the features is one of the most challenging aspects of feature extraction. With an adequate method, the image's uniqueness may be preserved. Features like shape, texture, leaf colour, and veins are characteristics needed to find the correct leaf (Pushpa and Athira, 2021).

5.1. Color features

Color is an essential aspect of image representation. Colour moments are examined by statistical functions like mean, median, and standard deviation, which are calculated on the three RGB channels (red, green, and blue) (Sivaranjani, 2019). Consider each leaf picture segment to be a grey image I of dimensions in MN, with pixels represented by p (i, j) (0, L-1), Whereas L is the No. of Grey levels in I (Pacifico, 2019)(Singh1, 2019)(Wang, 2017).

*5.2. **Shape** feature*

The image's shape is a crucial aspect of its description. It is made up of the boundary and the pixels in the region that are encircled by the boundary. A limit is required when an object is composed of an image, and a locality is defined by a pixel limit (MostajerKheirkhah and HabibollahAsghari. , 2019)(Pukhrambam and Rathna, 2021)(Sabu and Sreekumar, 2017)(Kan et al., 2017)(Yu, 2021)(Durgadevi, 2021). Zernike, Fourier descriptors, or contour detection can extract shape characteristics from plant outlines. Moments are vital for identifying plant species because they offer a condensed and consistent depiction of the morphology of the plant, which is critical for differentiating between species with comparable textures or colours but distinct forms.

5.3. Vein feature

Vein pattern is a significant property for identifying plants since each leaf has a distinct vein pattern. The characteristics of different plant species include colour, texture, vein pattern, and form. Leaf vascular system analysis determines vein pattern, Gabor filters or Haralick features determine texture, colour is retrieved using pixel values in various channels, and form is obtained using contour detection and shape descriptors. It's calculated as the total amount of white pixels divided by the leaf's area.

5.4. Texture features

The image's texture specifies characteristics such as smoothness, sparsity, etc. Three types of texture description methods are often used: statistics, structure, and frequency spectrum (Singh1, 2019)(Sabu and Sreekumar, 2017)(Muthevi and Uppu, 2017)(Wang, 2017). The GLCM approach is used to retrieve the leaf's texture features (Vijayashree and Gopal, 2017).

5.5. Additional Features

5.5.1. SURF- speeded up robust features

Used for feature detecting and classifying of objects by feature detector and descriptor. The summing responses of the 2D Haar wavelet serve as the foundation for the SURF algorithm (Sabu et al., 2017).

5.5.2. SIFT—Scale invariant feature transform

This method detects and describes an image's local properties. As the name suggests, the detection of features consistently recovers many frames from an image, even in the presence of some variations in light effects, angle views, and other viewing methods. The descriptor establishes a robust and compact identification of the locations by connecting a signature with them.

5.5.3. HoG –Histogram of oriented Gradients

For object detection, exactors are employed to count orientation in specific image regions. This kind of approach is comparable to the scale feature, but it is computed with grid cells for normalization in case of overlapping to get highly accurate results (Chaki et al., 2020)(Sabu et al., 2017)(Sachar and Kumar, 2021). A feature descriptor called Histogram of Orientated Gradients (HOG) is used in object recognition to identify an item's form or structure by examining the gradient orientations in some regions of an image. Offering a solid depiction of plant forms, particularly those with distinct edges, increases the accuracy of plant species identification.

5.5.4. Zernike Moments (ZM)

ZM is a polynomial moment set-based region-based shape descriptor that uses shape information most and enables accurate recognition. Zernike Moments (ZMs) are robust image-processing techniques that are especially helpful for identifying and categorizing therapeutic herbs. They can locate species regardless of orientation because they can record the form and texture of plants in a rotationally invariant manner. ZMs offer a strong characteristic description that simplifies distinguishing various species, even ones with similar morphologies. ZM descriptors provide a thorough representation of the plant's appearance, capturing tiny differences in form and texture essential for differentiating between species and considerably improving the accuracy of medicinal plant species recognition. These polynomials are defined in polar coordinates, which makes rotation invariance easier to achieve. The feature vector in the ZM descriptor is made up of 37 floating point characteristics (Chaki et al., 2020) (Dahigaonkar and Kalyane, 2018) (Araujo, 2017). Different methods for detecting plants require different amounts of processing power; Zernike Moments requires the greatest since it involves intricate mathematical computations. While classic techniques like K-Nearest Neighbour and Support Vector Machines may struggle with large-scale data and complicated characteristics, CNNs are scalable and efficient for massive datasets. Erroneous classification of edible and toxic plants can result in serious health hazards, such as poisoning or even death, and financial damages, especially for sectors like agriculture and medicines that depend on precise plant species identification. The vital shape information of the plant species is captured in the feature vector obtained from Zernike Moments, which is essential for differentiating between plant classes. The ZM feature vector offers a strong representation that improves the classification accuracy of medicinal plants by capturing rotationally invariant properties.

6. Classification of image

6.1. Convolutional Neural network (CNN)

Based on Deep Learning, a Convolutional Neural Network (CNN) model classifies medicinal plants according to the characteristics of their leaves, including form, size, colour, texture, and so on. Convolutional layers are used in plant species detection to extract attributes from pictures; pooling layers are used to minimize spatial dimensions; and fully connected layers are used to predict plant species while preserving model performance at various scales and image variances. Combining several aspects, including colour, texture, form, and vein patterns, improves the accuracy of identifying different plant species by giving a complete picture of the morphology of the plant. This lessens misclassification by helping to differentiate between species that may seem similar in some ways but differ in others. Since its neurons are composed of learnable weights and biases, it functions similarly to a feed-forward neural network. An input layer, an output layer, and multiple hidden layers between the input and output rectified linear unit (ReLU), convolution, and fully connected layer, known as the classification layer. A convolution layer carries out a convolution operation, which uses a set of convolution filters to extract features from input images and produce feature maps. By using non-linear down-sampling, the pooling layer allows only the lowest number of parameters for feature maps (Paulson and Ravishankar, 2020)(Dileep and Pournami, 2019) (Jayalath, 2019)(Sabri, 2018). Several obstacles to accurately recording herbal medications include data paucity, morphological similarity, and inconsistent terminology. Regional differences in naming traditions and classifications might result in discrepancies, making manual identification challenging and prone to mistakes. By reducing map dimensionality and concentrating on essential features, pooling layers improve feature extraction. By lessening the model's sensitivity to feature location, they help identify different plant species by enhancing identification in all kinds of lighting and background situations.

Pretrained models VGG16 is a 16–layer Convolutional Neural Network architecture called VGG16. Among the layers are softmax, fully connected, convolutional, and max pooling layers. Overall, there are 138 million parameters in VGG1. There are 16 wt layers out of 21, including three fully connected, five max pooling, and 13 convolutional layers. We have a softmax output of one of 1000 classes and 4096 fully connected layers (FC). 1000 classes and more than 14 million images were used to train the VGG16 network. (Paulson and Ravishankar, 2020)(Quoc and Hoang, 2020)(Kaya, 2019)(Abas and Haqmi, 018) (Prasad and Singh, 2017).

6.2. VGG19

With its 19 wt layers, VGG19 is capable of classifying images into 1000 distinct object varieties. Overall, there are 144 million parameters in VGG19. Three fully connected layers and sixteen convolutional layers total nineteen weight layers. Lastly, there are two FC layers, each with 4096 units, and a third layer with 100 units that can be classified in 1000 distinct ways (one for each class). A Softmax layer is also available (Paulson and Ravishankar, 2020).

6.3. Support vector Machine (SVM)

SVM(Support Vector Machine) is a linear classifier that is not random. SVM is a kernel-based supervised learning technique that uses supervised learning to split data into two or more classes. The original data must be moved into high dimensions, whereas hyperplanes separate as classes from data. Hyperplanes are used by the support vector machine algorithm, a machine learning system, to divide each class. The hyperplanes are precisely located using the support vectors (Raghukumar and Narayanan, 2020)(Bambil, 2020)(MostajerKheirkhah and HabibollahAsghari. , 2019)(Dileep and Pournami, 2019). The non-parametric, adaptable, and computationally costly K-Nearest Neighbour (KNN) technique is used to identify different plant species. New data points are categorized according to the predominant class of their k-nearest neighbours. A model-based learning technique called Support Vector Machine (SVM) manages non-linear decision boundaries, maximizes the margin between plant species classes, and uses less memory than KNN. Machine learning models have greatly enhanced the recording and preservation of medicinal plants by processing massive datasets, automating species identification, and lowering human error. By evaluating enormous databases, they support research, data management, and conservation initiatives by discovering new species or undiscovered therapeutic qualities.

6.4. K- Nearest Neighbor(KNN)

The K-nearest neighbour algorithm, or KNN algorithm, is based on the neighbours who are closest to you. Supervised KNN is produced by applying the labels to the training data set. Predicts output based on the label available on the image from its neighbour.(Thella and Ulagamuthalvi, 2021) (Applalanaidu and Kumaravelan, 2021)(Raghukumar and Narayanan, 2020)(MostajerKheirkhah and HabibollahAsghari. , 2019). CNNs are deep learning models that use hierarchical characteristics they have learnt from raw pictures to classify images accurately. KNNs are straightforward, non-parametric algorithms that are computationally costly and less accurate for big datasets. ANNs are more versatile neural networks assessed using comparable criteria and utilized for various tasks, including classification. Reducing variability and boosting dependability, a computerized system automates recognizing and classifying edible herbal plants. Applying uniform criteria to every sample through sophisticated algorithms integrates data from many sources to provide a thorough knowledge of plants' properties and possible medical applications.

6.5. Artificial Neural network (ANN)

Artificial Neural Networks analyze data in a manner like the human brain(Azadnia and Kheiralipour, 2021). Multiple designs of ANN models were constructed by varying the number of neurons in hidden layers and testing them to get the optimum classification model. In each developed architecture, the input, hidden and output layers had three digits of neurons.(Huixian, 2020).

6.6. Multi-Layer perception (MLP)

A multi-layer perception is a kind of ANN that employs feed-forward learning. It links the relevant outputs in one group of input data to those present in another directed graph with many nodes. These nodes have many layers that are unique from others in that they are interconnected to form MLP. Neurons comprise the MLP network and are present in every layer. It is trained using the supervised backpropagation approach, which is typically used for non-linear problems (Durgadevi, 2021)(Qadri, 2019)(Dahigaonkar and Kalyane, 2018).

6.7. Probabilistic Neural Network(PNN)

It is a feed-forward neural network designed to address problems related to pattern recognition and classification. It has four layers: input layer, output layer, summation layer, and pattern layer. The PNN uses the training data to calculate most of the words. The only thing that has to be adjusted is the smoothing factor, which sets a limit on Gaussian function deviations. It's a classifier that separates input patterns according to various categorization thresholds. It can be forced to approximate a more general function (Qadri, 2019) (Zhang et al., 2018).

6.8. Local binary patterns (LBP)

It is utilized for feature excision; the provided leaf image 's 3X3 mask is employed to calculate the LBP code. A matrix representation of the pixel intensity will be used (Naeem, 2021). There is a set of neighbours, P, for every pixel, C. A histogram h is defined by the variations in texture intensity between C and its neighbours in P. An attractive property of this descriptor is that changes in the average intensity of the central pixel remain constant when compared to its neighbours. This work uses non-uniform patterns with LBP; the histogram for an 8-pixel neighbourhood has 59 bins. The photos are split into four equal areas and normalized to 256X256 pixels to extract distinct attributes from each image area (Araujo, 2017).

6.9. Random forest Classifier(RFC)

This technique follows the decision tree concept to get predicted outputs by the voting process. From the training data, RFC (Pacifico, 2019).

6.10. Naive Bayes classifier (NBC)

The most straightforward way to predict the accurate output. Being a probability classifier, predictions are based on an object's probability. Naïve means one feature prediction is related to another feature's prediction, and Bayes means following Bayes Theorem concepts (Tan, et al., 2018)(Mittal et al., 2018).

6.11. K-Means classifier (K-Means)

The K-average or K-means algorithm is frequently utilized. The clustering algorithm is representative of the clustering subgroup as a whole; it is the average of all sampled data. The fundamental concept is that by breaking up a massive data set into smaller groups through an iterative process, the clustering criterion evaluation function is optimized to get the best results, resulting in compact findings for each cluster. Although not intended for discrete data processing, clustering impacts continuous data and offers specific benefits (Wang, 2017) (Begue, 2017).

7. Conclusion

According to Ayurveda's history, every plant has therapeutic worth, thus determining what appropriate portion contains the medicinal value used to treat human diseases. Many diseases are treated using medicinal plant components like roots, stems, flowers, leaves, fruits, seeds, oil extracts, etc. Botanists and herbal practitioners manually identify therapeutic plants using these components, which requires more time. The research aims to produce completely computerized medicinal plant identification using image processing techniques, thereby decreasing labour costs and boosting crop productivity. According to the literature review, most researchers classified medicinal plants based on leaf features, comparatively with less study on other parts of plants. From our research, we have concluded, based on the comparison table, that Multi-Layer Perception (MLP) is the simplest classification method with high accuracy. The demerits of this MLP are time-consuming.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Funding

Authors did not receive any funding.

Authors' contributions

Beulah Jabaseeli, N, is responsible for designing the framework, analyzing the performance, validating the results, and writing the article. Umanandhini D, is re- sponsible for collecting the information required for the framework, provision of software, critical review, and administering the process.

CRediT authorship contribution statement

N. Beulah Jabaseeli: Conceptualization. **D. Umanandhini:** Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.
No datasets were generated or analyzed during the current study.

References

Abas, Mohamad Aqib Haqmi, et al. "VGG16 for plant image classification with transfer learning and data augmentation." Int. J. Eng. Technol. (UAE) 7.4 (2018), 90–94.

Akter, Raisa, and Md Imran Hosen. "CNN-based Leaf Image Classification for Bangladeshi Medicinal Plant Recognition." 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE). IEEE, 2020.

Amuthalingeswaran, C., et al., 2019. Identification of Medicinal Plant's and Their Usage by Using Deep Learning. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI).

Applalanaidu, M.V., Kumaravelan, G., 2021. A Review of Machine Learning Approaches in Plant Leaf Disease Detection and Classification. 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV).

Araujo, Voncarlos, et al. "Multiple classifier systems for plant leaf recognition." *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2017.

Azadnia, R., Kheiralipour, K., 2021. Recognition of leaves of different medicinal plant species using a robust image processing algorithm and artificial neural networks classifier. J. Appl. Res. Med. Aromat. Plants 100327.

Bambil, D., et al., 2020. Plant species identification using colour learning resources, shape, texture, through machine learning and artificial neural networks. Environm. Syst. Decis. 40 (4), 480–484.

Begue, A., et al., 2017. Automatic recognition of medicinal plants using machine learning techniques. Int. J. Adv. Comput. Sci. Appl. 8 (4), 166–175.

Chaki, J., et al., 2019. Fragmented plant leaf recognition: Bag-of-features, fuzzy-colour and edge-texture histogram descriptors with multi-layer perceptron. Optik 181, 639–650.

Chaki, J., Parekh, R., Bhattacharya, S., 2020. Plant leaf classification using multiple descriptors: A hierarchical approach. J. King Saud Univ. Comput. Inform. Sci. 32 (10), 1158–1172.

Dahigaonkar, T.D., Kalyane, R., 2018. Identification of ayurvedic medicinal plants by image processing of leaf samples. Int. Res. J. Eng. Technol. (irjet) 5, 351–355.

Das, D., et al., 2020. Leaf Disease Detection using Support Vector Machine. 2020 International Conference on Communication and Signal Processing (ICCSP).

Dileep, M.R., Pournami, P.N., 2019. Ayurleaf: A deep learning approach for classification of medicinal plants. TENCON 2019–2019 IEEE Region 10 Conference (TENCON).

Durgadevi, M., 2021. Generative Adversarial Network (GAN): a general review on different variants of GAN and applications. 2021 6th International Conference on Communication and Electronics Systems (ICCES).

Fekri-Ershad, S., 2020. Bark texture classification using improved local ternary patterns and multi-layer neural network. Expert Syst. Appl. 158, 113509.

Gavhale, Nayana G., and A. P. Thakare. "Identification of Medicinal Plant Using Machine Learning Approach." (2020).

Hu, J., et al., 2018. A multiscale fusion convolutional neural network for plant leaf recognition. IEEE Signal Process Lett. 25 (6), 853–857.

Huixian, J., 2020. The analysis of plants image recognition based on deep learning and artificial neural network. IEEE Access 8, 68828–68841.

Jayalath, A.D.A.D.S., et al., 2019. Identification of Medicinal Plants by Visual Characteristics of Leaves and Flowers. 2019 14th Conference on Industrial and Information Systems (ICIIS).

Kan, H.X., Jin, L., Zhou, F.L., 2017. Classification of medicinal plant leaf image based on multi-feature extraction. Pattern Recognit Image Anal. 27 (3), 581–587.

Kaya, Aydin, et al. "Analysis of transfer learning for deep neural network based plant classification models." Comput. Electron. Agricult. 158 (2019): 20-29.[11].

Li, L., Zhang, S., Wang, B., 2021. Plant disease detection and classification by deep learning—a review. IEEE Access 9, 56683–56698.

Lukic, M., Tuba, E., Tuba, M., 2017. Leaf recognition algorithm using support vector machine with Hu moments and local binary patterns. 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI).

Mittal, Parul, Manie Kansal, and Hardeep kaurJhajj. "Combined classifier for plant classification and identification from leaf image based on visual attributes." *2018 International Conference on Intelligent Circuits and Systems (ICICS)*. IEEE, 2018.

MostajerKheirkhah, Fateme, and HabibollahAsghari. "Plant leaf classification using GIST texture features." *IET Computer Vision* 13.4 (2019): 369–375.

Muthevi, A., Uppu, R.B., 2017. Leaf classification using completed local binary pattern of textures. 2017 IEEE 7th International Advance Computing Conference (IACC).

Naeem, S., et al., 2021. The classification of medicinal plant leaves based on multispectral and texture feature using machine learning approach. Agronomy 11 (2), 263.

Pacifico, L.DS., et al., 2019. Automatic classification of medicinal plant species based on colour and texture features. 2019 8th Brazilian Conference on Intelligent Systems (BRACIS).

Paulson, Anu, and S. Ravishankar. "AI-Based Indigenous Medicinal Plant Identification." *2020 Advanced Computing and Communication Technologies for High-Performance Applications (ACCTHPA)*. IEEE, 2020.

Prasad, S., Singh, P.P., 2017. Medicinal plant leaf information extraction using deep features. TENCON 2017–2017 IEEE Region 10 Conference.

Priya, K., Ranjana, S., Manimegala, R., 2020. Medicinal plant identification using andr oid application based on leaf image. Europ. J. Mol. Clin. Med. 7 (9), 1496–1506.

Pukhrambam, B., Rathna, R., 2021. A Smart Study on Medicinal Plants Identification and Classification using Image Processing Techniques. 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV).

Pushpa, B.R., Athira, P.R., 2021. Plant Species Recognition Based on Texture and Geometric Features of Leaf. 2021 3rd International Conference on Signal Processing and Communication (ICPSC).

Pushpanathan, K., et al., 2021. Machine learning in medicinal plants recognition: a review. Artif. Intell. Rev. 54, 1.

Qadri, S., et al., 2019. Machine vision approach for classification of citrus leaves using fused features. Int. J. Food Prop. 22 (1), 2072–2089.

Quoc, T.N., Hoang, V.T., 2020. Medicinal Plant identification in the wild by using CNN. 2020 International Conference on Information and Communication Technology Convergence (ICTC).

Raghukumar, A.M., Narayanan, G., 2020. Comparison of machine learning algorithms for detection of medicinal plants. 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC).

Rajani, S., Veena, M.N., 2018. Study on Identification and Classification of Medicinal Plants. Int. J. Adv. Sci. Eng. Technol. 6 (2), 13–18.

Sabri, N., et al., 2018. "Comparing convolution neural network models for leaf recognition". Int. J. Eng. Technol. 7 (3), 141–144.

Sabu, A., Sreekumar, K., Nair, R.R., 2017. Recognition of ayurvedic medicinal plants from leaves: A computer vision approach. 2017 Fourth International Conference on Image Information Processing (ICIIP).

Sabu, A., Sreekumar, K., 2017. Literature review of image features and classifiers used in leaf-based plant recognition through image analysis approach. 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT).

Sachar, S., Kumar, A., 2021. Survey of feature extraction and classification techniques to identify plant through leaves. Expert Syst. Appl. 167, 114181.

Singh[1], Sukhdev, and Priya Pinder Kaur. "A study of geometric features extraction from plant leafs." (2019).

Sivaranjani, C., et al., 2019. Real-Time Identification of Medicinal Plants using Machine Learning Techniques. 2019 International Conference on Computational Intelligence in Data Science (ICCIDS).

Sujith, A., Aji, S., 2020. An Optimal Feature Set with LBP for Leaf Image Classification. 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC).

Wei Tan, Jing, et al. "Deep learning for plant species classification using leaf vein morphometric." *IEEE/ACM transactions on computational biology and bioinformatics* 17.1 (2018): 82-90.

Thella, P.K., Ulagamuthalvi, V., 2021. A Brief Review on Methods for Classification of Medical Plant Images. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC).

Vijayashree, T., Gopal, A., 2017. Leaf identification for the extraction of medicinal qualities using image processing algorithm. 2017 International Conference on Intelligent Computing and Control (I2C2).

Wäldchen, J., et al., 2018. Automated plant species identification—Trends and future directions. PLoS Comput. Biol. 14 (4), e1005993.

Wang, Z., et al., 2017. Review of plant identification based on image processing. Arch. Comput. Meth. Eng. 24 (3), 637–654.

Yousefi, E., Baleghi, Y., Sakhaei, S.M., 2017. Rotation invariant wavelet descriptors, a new set of features to enhance plant leaves classification. Comput. Electron. Agric. 140, 70–76.

Yu, J., et al., 2021. Progress in the use of DNA barcodes in the identification and classification of medicinal plants. Ecotoxicol. Environ. Saf. 208, 111691.

Zandi, P., et al., 2017. Fenugreek (Trigonella foenum-graecum L.): An Important Medicinal and Aromatic Crop. In: Active Ingredients from Aromatic and Medicinal Plants, pp. 207–224.

Zhang, S., Zhang, C., Huang, W., 2018. Integrating leaf and flower by local discriminant CCA for plant species recognition. Comput. Electron. Agric. 155, 150–156.