Diverse Pre-processing Strategies for Enhancing the Performance in the Classification of Aromatic and Medicinal Plants Using Leaf Images

Shareena E M [1#*] , D. Abraham Chandy [2@]

[1] Research Scholar, Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India

[2]Associate Professor, Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India

[#] Assistant Professor, Department of Electronics, MES College Marampally, Aluva ,India

[*]E- mail id: shareenaem@mesmarampally.org

[@]E- mail id: abrahamchandy@karunya.edu

---**ABSTRACT**---

The work focuses on the essential aspects of pre-processing and feature extraction as a preliminary phase before implementing a deep learning classification model. Although deep learning models possess intrinsic feature learning capabilities, the preliminary extraction of essential features cannot be overlooked. This involves identifying salient characteristics, patterns, and structures within the data before feeding it into the deep learning model. By carefully preparing the data through techniques like normalization and augmentation, and by extracting key features from the dataset, the work aims to establish a solid groundwork that ensures the subsequent deep learning model can perform effectively and produce accurate results in classifying the data. The dataset comprises authentic images of plant leaves sourced from the Aromatic and Medicinal Plant Research Station, Odakkali, Kerala, and includes a diverse range of leaf images from 39 distinct aromatic and medicinal plant species. Considering the year-round availability of plant leaves, we give prominence to this essential botanical element. It acts as a cornerstone in our pursuit of attaining strong identification results that are dependable and accurate. In this study, we employ various approach that leverages shape, color, and texture features extracted from plant leaves. The work showcases the implementation of pre-processing techniques on a single sample leaf image of Ricinus Communis, the castor oil plant, sourced from the dataset.

Key words: **Aromatic and Medicinal Plants, Ricinus Communis, Morphological transformation, Minimum Area Rectangle**

I.INTRODUCTION

Plants play a crucial role in the natural ecosystem, contributing to essential functions such as drug synthesis, erosion prevention, and climate regulation. However, the continued existence of certain plant species is threatened, as they face the risk of extinction. Efforts have been undertaken by authorities to mitigate the decline of these plants. Among the endangered species, many aromatic and medicinal plants are of significant concern. So for the proper preservation of medicinal and ayurvedic plants an effective identification technique is needed. Nowadays personal opinion from experienced people is utilised in the identification process. In this process visual cues is mainly considered. Another resource for the identification procedure is the people living in forest areas. They collects the plant species for their livelihood along with urbanization, significantly affects the preservation of the valuable resources. Additionally ignorance of the modern population about the species strengthens the need of automatic identification. Also accurate identification is necessary for medicinal applications.

Medicinal and aromatic plants plays an important role in the preparation of antiviral drugs, cosmetics, flavour and fragrance providing components. So analysis using shape, color, texture features is an efficient way without the information about the chemical constituents.

This work focuses on the initial experiments for the implementation of medicinal and aromatic plant identification process using deep learning technique. Here the plant leaf is used for the identification process since the plant leaf is available through out the year. The parameters that we can analyse through leaves are leaf blade, texture, margin details such as apex, veins, midrib, base etc.

The following sections are organized as follows. First part includes the state of art regarding the leaf classification techniques followed by the discussion on various pre-processing steps. In this feature extraction using shape and texture is explained. Various edge detection algorithm is also presented.

II. LITERATURE REVIEW

In previous years various works contributed valuable insights in the domain of leaf identification and plant classification.

Sue Han Lee et al. (2015) [1] suggested an identification technique utilizing venation patterns. A curvature based shape features using CNN is implanted in this work. Adzkia Salima et al. (2015) [2] proposed a fast parallel thinning algorithm for segmentation using Hessian matrix.346 leaves comprising 55 classes is used for the experimentation. Naresh Y G et al. (2016) [3] tackled texture features through modified local binary pattern. They used texture features from 33 plant species. Intra cluster variation occurs due to texture variation is considered in this case. Trishen Munisami et al. (2015) [4] explored identification using color histogram and shape features. Using Matcher algorithm they demonstrated the significance of dataset and number of species on recognition accuracy. Guillermo L et al. (2016) [5] presented vein morphological patters using deep convolutional neural network. Kutha Krisnawijaya et al. (2017) [6] suggested image processing with parallel processing, using Fuzzy Local Binary Patterns for analysis. Shitala Prasad et al. (2017) [7] employed a higher-dimensional capturing device to capture shape and texture and using this images classification is done using PCA and the lαβ color space. H. X. Kan et al. (2017) [8] explored various shape and texture features using SVM classifiers. Multi-feature of leaves are extracted in this work. .Liwen Gao et al. (2019) [9] proposed watershed algorithm and Ostu method for vein enhancement and extraction. Luciano D. S. Pacifico et al. (2019) [10] compared different features of leaves utilizing the classifiers like K-Nearest Neighbors, Weighted K-Nearest Neighbors, Decision Trees, Random Forests, and Multi-Layer Perceptrons with Backpropagation.18 leaf features analysed in this work. Anu Paulson et al. (2020) [11] classified 64 species of medicinal plants using CNNs, VGG-16, and VGG-19, showing VGG-16 to provide the highest accuracy among the three algorithms. Dawei Li et al.(2019)[12] proposes a novel technique for individual leaf segmentation without overlaps in plant point clouds using 3D filtering and facet region growing. A 3D joint filtering operator is use for separating leaves with varying degrees of overlap. This operator combines a Radius-based Outlier Filter (RBOF) and a Surface Boundary Filter (SBF) Yang et al,(2020)[13] collected over 2500 leaf images with complex backgrounds and artificially labeling them with target and background pixels. Of these, 2000 images are utilized to train a Mask Region-based Convolutional Neural Network (Mask R-CNN) model for leaf segmentation. Additionally, a training set containing more than 1500 images from 15 species is employed to train a leaf classification model using a deep convolutional network with 16 layers (VGG16). Extensive parameter combinations are evaluated to identify the optimal hyperparameters for these methods. Hanno Scharr(2016)[14] utilized unique data set from phenotyping experiments for leaf segmentation.

Most of the studies mentioned the different techniques and methodologies for the identification of plants using leaf analysis. We can see the importance of different approaches to improve the efficiency and accuracy of identification using leaf characteristics .Above all some of the papers illustrate the advanced techniques to refine leaf segmentation and classification techniques. Through the extensive review, traditional image processing technique and complexities associated with leaf analysis for deeplearning techniques is explored. The results provide valuable insights for advancing the accuracy and efficiency of plant phenotyping and classification.

III.PROPOSED WORK

The conventional method of identification follows a step by step method involving image pre-processing, feature extraction and pattern classification. The first step is the pre-processing phase, which include the conversion from color to grayscale, filtering, segmentation and edge detection. This is followed by feature extraction which captures information about texture, shape and venation characteristics. This systematic sequence effectively helps in the identification of aromatic and medicinal plants through leaf analysis.

Usually fourier descriptors, invariant moments and other relevant shape attributes are used to extract shape features. In Fourier descriptors X and Y coordinates of the boundary is taken as real and imaginary components of complex numbers. Discrete Fourier Transform of these coordinates results in output coefficients which constitute the Fourier descriptors. The overall shape is determined by the initial coefficient , representing low frequencies. The fine details of the shape is provided by the high frequency components. This approach allows for the comprehensive characterization of shape attributes within the context of medicinal and aromatic plant identification

Taking coordinates as x,y

$$S(k)=X(k)+jY(k) \quad (1)$$

characteristics etc. Leaf blade includes important

The fourier descriptor $a(u)=\frac{1}{K}\sum_{k=0}^{K-1} s(k)e^{\frac{-2\pi juk}{K}}$ (2)

Reconstructed function $S(k)=\frac{1}{k}\sum_{u=0}^{K-1} a(u)e^{\frac{2\pi juk}{K}}$ (3)

Statistical moments are considered in integer orders where order represents the number of points in the data, order 1 represent the sum and it is used to find the average. Order 2 and order 3 represent variance and skew respectively. The nth moment

$$\mu_n(r) = \sum_{i=0}^{K-1}(r_i - m)^n g(r_i) \quad (4)$$

Where $m=\sum_{i=0}^{K-1} r_i g(r_i)$

Shape features can be extracted from image segmentation. Segmented image contains boundary as well as the region pixel surrounded by the boundary. So boundary based features and region based features can be extracted. Some of the shape features are aspect ratio, circularity, area convexity, perimeter convexity, rectangularity, compactness Hu's moment, irregularity, eccentricity, narrow factors etc.

Texture information provides the information about the spatial arrangement of colour or intensities in the image. Usually analysed texture descriptors are coarseness, contrast, directionality, line-likeness, regularity, roughness etc. Texture information features can be calculated using statistical method, structural method, modelling features and filter method. Statistical method is the most common approach utilized for texture analysis. It include general statistical parameters, autocorrelation features, texture energy features, co-occurrence matrix based features.

First order texture measures are statistics calculated from the original image values like variance and do not consider pixel neighbourhood relationships .First order measures are directly taken from the histogram. Some of the first order statistical measures are entropy, energy, third central moment etc.

Entropy $Ep =-\sum_{i=0}^{L-1} p_i \log_2 (p_i)$ (5)

Energy $E_n = \sum_{i=0}^{L-1} p_i^2$

Third central moment

$\mu_3=\sum_{i=0}^{L-1}(i-m)^3 p_i$

where $m=\sum^{L-1}$

to effectively separate occluded leaves. Kunlong more pixels. Calculation time and interpretation

difficulty is high in this analysis.

Leaf venation patterns are often analysed through edge detection algorithms. The process involves the removal of the leaf boundary, which is present in the edge-detected output but is not part of the venation pattern. This can be achieved by calculating the difference between the Canny edge-detected image and the leaf boundary. The extraction of curves is a critical step, as it separates the linear and curvilinear elements within a two-dimensional array. This extraction process considers connected pixels in eight directions. Any disconnected pixels indicate the presence of a venation curve. Hue normalization

is employed to enhance contrast, aiding in the differentiation between leaf veins and the rest of the leaf structure. Image fusion techniques are then used to distinguish the areas representing leaf venation patterns from other components of the leaf. Overall, these methods contribute to the accurate interpretation and analysis of various leaf venation types, providing valuable insights into the characteristics of medicinal and aromatic plants.

IV.RESULTS AND DISCUSSION

Pre-processing step include the conversion of RGB image to gray image. As the captured image is having noise due to camera sensor all the images is smoothened using gaussian filter. Open CV gaussian blur is selected. All the edges in the image are smoothened with minimum blur. Image is having different lighting condition in different area. A sample leaf image of Ricinus communis, the castor oil plant, sourced from the dataset, is taken for illustration. The image of the leaf is captured using a SONY ALPHA 7R III (ILCE-7RM3) camera, with a ZEISS BATIS 40 mm f/2 CF lens, a CANON EF 100 mm f/2.8 macro lens, and a SIGMA MC-11 mount converter on a white background.

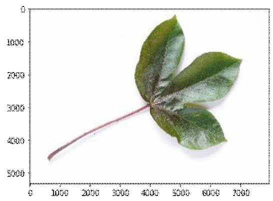


Fig.1. RGB Image

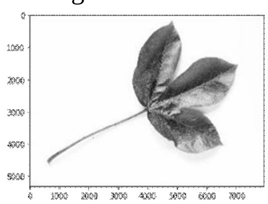
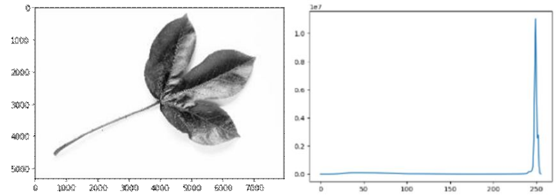


Fig.2. Gray Scale Image

(6)

(7)


0
1000
2000
3000
4000
5000
0 1000 2000 3000 4000 5000 6000 7000
1e7
1.0
0.8
0.6
0.4
0.2
0.0
0 50 100 150 200 250

entropy value lies near to the middle of entropy

Second order measures the relationship between the group of two pixels in the original image. Gray level occurrence matrix is the most commonly used texture analysis measure. Third and higher order measures considers the relationship among three or

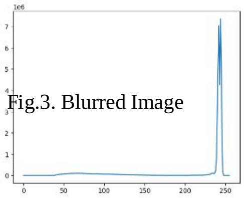


Fig.3. Blurred Image Fig.4. Histogram of Orignal Image

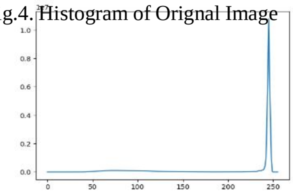


Fig.5. Histogram of RGB Fig.6. Histogram of Gray Scale

Image Type	Entropy Value
RGB Image	4.38802
BGR Image	4.22744
Gray Scale Image	4.185144

Table:.1. Entropy value obtained from histogram

A commonly used tool for image analysis and processing is histogram, which provides the information regarding the pixel intensity distribution of an image. For dark images histogram is skewed to the left side, ie lower intensity value. All the histogram shows above skewed to high intensity values which shows the sample image is sufficiently bright. Also unexpected spike in the histogram represents noise. Here we can notice the the random distribution is less in gray scale image.

For original image, the entropy is 4.22744 which is higher than gray scale image and lower than RGB image. RGB image is having entropy 4.38802, which represents high complexity, randomness and variation in pixel intensity. Gray scale image is having entropy value 4.185144 which represents less variability than color images. The variation in entropy values arises from the amount of color and intensity information retention or loss in various images. Most of the leaf images have shadows, textures and subtle color differences. Some of these complexity reduced in gray scale image. Depending upon the color variation, texture and lighting condition the entropy values are reasonable. Highest entropy 4.38802 indicates the variations of color in the leaf image due to gradients in the green or other color tones. Since leaf is having subtle color variation, this value is reasonable. 4.22744 being slightly lower than RGB could be due to how OpenCV stores and processes the image in BGR format, though the difference is small. Entropy 4.185144 for gray scale being the lowest is expected since the conversion from color to grayscale simplifies the image by removing color information and representing it with pixel intensities only. The gray scale image is having entropy 4.185144 suggests that the image has a reasonable amount of detail and variation. It indicates that the image consists of some degree of textures which is due to the presence of veins or surface irregularities. The scale, ie image is not uniform and does not contain severe randomness or noise.

First method of image segmentation is done using adaptive image thresholding which converts gray scale image to binary image by finding the threshold of a small region so that different threshold for different region which gives better results. Ostu's thresholding is done on the regions of the image. As real time images of leaves are captured most of the leaves is having some kind of black points, it is compensated using morphological transformation. These pre-processed images are used for feature extraction.

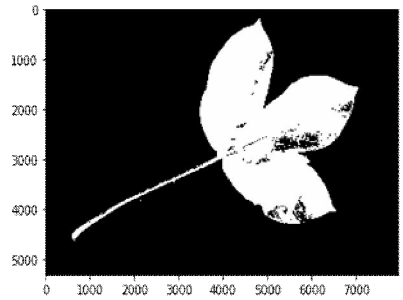


Fig.7. Ostu Threshold Image

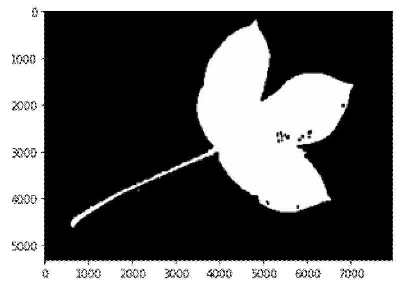


Fig.8. Segmented Image after Morphological Transformations

The optimal threshold value obtained for the given image is 127, which indicates all the pixels with value greater than 127 are classified as foreground and less than 127 classified as background. From these values important features such as area, perimeter, solidity, aspect ratio etc. can be explored, which is inevitable for leaf segmentation.

Summary of Pixel Distribution in Sample Image	
Foreground Pixel Count	6,997,955
Background Pixel Count	35,179,453
Total Pixel Count	42,177,408
Percentage of Foreground (representing the leaf area)	16.6%
Percentage of Background	83.4%

Table:2. Pixel Distribution

6,997,955 pixels are identified as foreground, ie leaf after applying Ostu's thresholding.This constitute approximately 16.6 % of the total image area of 42,177,408 pixels.Remaining 35,179,453 pixels constitutes the background which is 83.4 % of the image. These numerical figures provide a strong quantitative assessment of the valid segmentation process and offer intuitions into the features of the image being analyzed.

Metric	Before Closing	After Closing	Change (Absolute)	Change (%)
Fore ground Pixel Count	6,997,955	7,136,78 2	+138,827	+1.98%
Back ground Pixel Count	35,179,45 3	35,040,6 26	-138,827	-0.39%
Total Pixel Count	42,177,40 8	42,177,4 08	-	-

Table:3. Comparison of Foreground and Background Pixel Counts Pre- and Post-Closing Operation

The table 3 shows an increase of **138,827** foreground pixels approximately **1.98%** representing the effectiveness in filling gaps by transformation and improves the continuity of the leaf structures. Similarly, the background pixel count reduced by **138,827** pixels approximately **0.39%**, reflecting a reduction in small artifacts and noise.

These results proves that morphological closing can significantly enhance the quality of leaf images by refining the representation of the foreground objects, which is important for subsequent analysis such as object counting or feature extraction.

Second method of segmentation utilizes edge detection technique where the difference in the intensities at the border is taken in to account. Both gradient based sobel edge detection and Laplacian based edge detection are applied to the data set. This operation reduces the number of pixel in the image but retains the structural aspect of the image analysed. Sobel edge detection is a gradient based method which uses first order derivative of the image along x axis and y axis of the image separately. Convolution operation is done with kernel on the image.

entropy value lies near to the middle of entropy

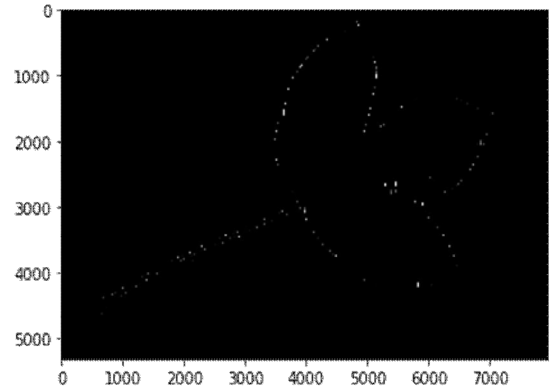


Fig.9. Image after applying Sobel Kernel

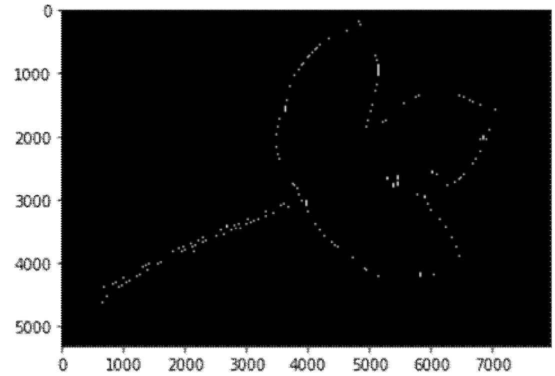


Fig.10. Image after Binary Thresholding

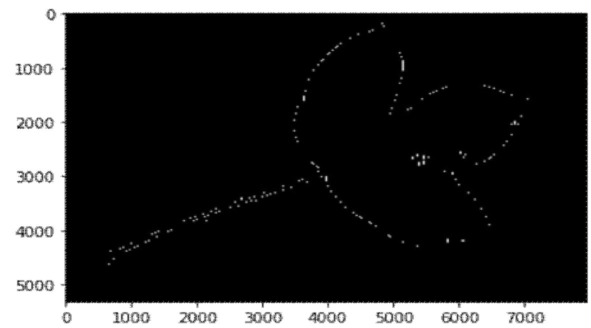


Fig.11.Image after first level of Morphological Transformation

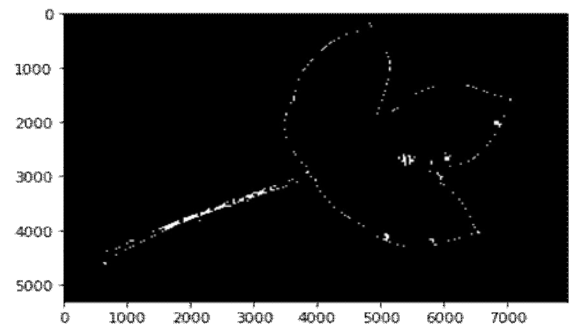


Fig.12.Image after second level of Morphological Transformation

Another method used is contouring. It forms a curve joining the continuous point having same parameters such as color or intensity. This technique is useful in shape detection and segmentation. For contouring to attain high accuracy, first threshold the image so as to get white object with black background. From contour analysis we can find various features such as moments, area, perimeter, bounding rectangle, minimum enclosing area, ellipse fitting etc.

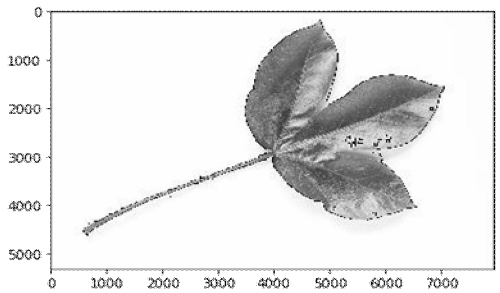


Fig.13. Image after Contouring

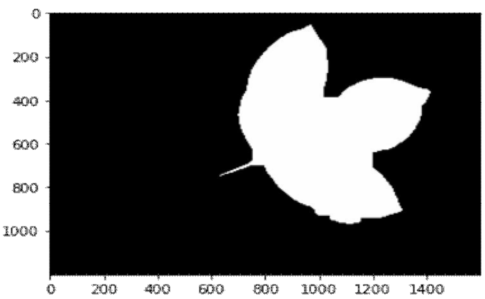


Fig.14. Minimum Area Rectangle

V.CONCLUSION

The research presented in this paper addresses the critical need for effective leaf classification in the context of aromatic and medicinal plants. The work emphasizes the importance of pre-processing and feature extraction as vital preparatory steps for accurate deep learning classification models, although the deep learning models itself have feature learning capability, the initial feature extraction is essential to enhance the performance of the model.

The study illustrates the advantage of various pre-processing techniques. All these techniques improves the quality of images, which provides accurate feature extraction. By utilizing shape, color, and texture features extracted from plant leaves along with effective pre-processing, classification accuracy is improved.

The experimentation is carried out using an authentic dataset collected from the Aromatic and Medicinal Plant Research Station in Odakkali, Kerala, which includes various aromatic and medicinal plant species, reinforces the credibility of the findings. Various techniques, including gradient-based edge detection, adaptive thresholding, and contour analysis, shows the flexibility of techniques employed for precise leaf segmentation and classification.

VI.FUTURE SCOPE

Deep learning technique on the real time data set improves the classification efficiency without the need of pre-processing and sophisticated learning. Image augmentation provides a dataset comprising wider range of variations which provides better accuracy to the model. By combining shape, color, texture features a more accurate classification outcome can be obtained. Undertaking a thorough comparative analysis of various classification algorithms, encompassing both traditional machine learning and deep learning models, would provide insights into the most effective techniques for this specific application. Real world validation is possible and thereby plant identification and conservation process can be enhanced. Data set can be enhanced by including various species and thereby improves the efficiency of classification model. Collaboration with botanist, pharmacology, experts in conservation of plants could lead to an integrated system for identification, preservation and medicinal application.

In conclusion, the research provides a strong foundation for automatic leaf classification in aromatic and medicinal plants through pre-processing techniques. By improving the accuracy of feature extraction, the study makes a valuable insight to the larger domains of plant identification, conservation, and medicinal research. Future work includes the inclusion of deep learning, expansion of datasets, and real-world implementation to further enhance the accuracy and practical significance of the proposed techniques.

References

1. S. H. Lee, C. S. Chan, P. Wilkin and P. Remagnino, "Deep-plant: Plant identification with convolutional neural networks," IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 2015, pp. 452-456, doi: 10.1109/ICIP.2015.7350839. (2015)

2. A. Salima, Y. Herdiyeni and S. Douady, "Leaf vein segmentation of medicinal plant using Hessian matrix,"IEEE International Conference on Advanced Computer Science and Information

Systems (ICACSIS), Depok, Indonesia, pp. 275-279, doi: 10.1109/ICACSIS.2015.7415152. (2015)

3. Naresh, Y., & Nagendraswamy, H. Classification of medicinal plants: An approach using modified LBP with symbolic representation. Neurocomputing, 173, 1789-1797. https://doi.org/10.1016/j.neucom.2015.08.090(2016)

4. Munisami, T., Ramsurn, M., Kishnah, S., & Pudaruth, S. Plant Leaf Recognition Using Shape Features and Colour Histogram with K-nearest Neighbour Classifiers. Procedia Computer Science, 58, 740-747. https://doi.org/10.1016/j.procs.2015.08.095(2015)

5. Guillermo L. Grinblat, Lucas C. Uzal , Mónica G. Larese, Pablo M. Granitto, Deep learning for plant identification using vein morphological patterns, Computers and Electronics in Agriculture**,**Volume 127, Pages 418-42(September 2016)

6. N.N. Kutha Krisnawijaya1, Yeni Herdiyeni2 & Bib Paruhum Silalahi, Parallel Technique for Medicinal Plant Identification System using Fuzzy Local Binary Pattern, J. ICT Res. Appl., Vol. 11, No. 1, 77-90(2017)

7. Shitala Prasad, Pankaj P Singh, Medicinal Plant Leaf Information Extraction Using Deep Features, Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8(2017)

8. H. X. Kan, L. Jin, and F. L. Zhou, Classification of Medicinal Plant Leaf Image Based on Multi-Feature Extraction, ISSN 1054-6618, Pattern Recognition and Image Analysis, Vol. 27, No. 3, pp. 581–587. © Pleiades Publishing, Ltd.(2017)

9. Liwen Gaoa , Xiaohua Linb, Fully automatic segmentation method for medicinal plant leaf images in complex background, Computers and Electronics in Agriculture,Volume 164, , 104924 (September 2019)

10. Luciano D. S. Pacifico, Larissa F. S. Britto, Emilia G. Oliveira, Teresa B. Ludermir, Automatic Classification of Medicinal Plant Species Based on Color and Texture Features, 8th Brazilian Conference on Intelligent Systems (BRACIS)(2019)

11. Anu Paulson, Ravishankar S, AI Based Indigenous Medicinal Plant Identification, 978-1-7281-6453-3/20/$31.00 © IEEE(2020)

12. D. Li *et al.*, "An Overlapping-Free Leaf Segmentation Method for Plant Point Clouds," in IEEE Access, vol. 7, pp. 129054-129070, , doi: 10.1109/ACCESS.2019.2940385.(2019)

13. Yang, Kunlong & Zhong, Weizhen & Li, Fengguo. Leaf Segmentation and Classification with a Complicated Background Using Deep Learning. Agronomy. 10. 1721. 10.3390/agronomy10111721. (2020).

14. Scharr, H., Minervini, M., French, A.P. et al. Leaf segmentation in plant phenotyping: a collation study. Machine Vision and Applications **27**, 585–606 https://doi.org/10.1007/s00138-015-07(2016)