Machine Learning Techniques for Medicinal Plants Identification Using Leaf Features: Review

Nilesh S. Bhelkar[1] and Swarnalata Bolavarapu[2]
[1]Assistant Professor, Artificial Intelligence and Data Science Department, Rajiv Gandhi Institute of Technology, Mumbai, India
[2]Assistant Professor, Computer Engineering Department, Mukesh Patel School of Technology Management and Engineering, NMIMS University, Mumbai, India

Cite this paper as:
Nilesh S. Bhelkar and Swarnalata Bolavarapu (2024). Machine Learning Techniques for Medicinal Plants Identification Using Leaf Features: Review. *Frontiers in Health Informatics,* 13(7), 520-526

Abstract
One of the most significant medical treasures is Ayurveda, which relies heavily on medicinal herbs to treat a wide range of illnesses or ailments. The only people who know about therapeutic plants are botanists and Ayurvedic practitioners. The following generation should be taught about medicinal plants. Experts tend to keep this knowledge to themselves and rarely communicate or publish it. The average person finds it quite challenging to identify the correct medicinal plant. In humans, incorrect identification resulted in unforeseen adverse effects. For the good of humanity, this life-saving information must thus be maintained through the use of modern technologies. An automated system for classifying and identifying items is being developed by a number of researchers for the benefit of humanity. In this paper, various machine learning classifier using leaf features are reviewed.

Keywords – *Medicinal plants, Classifier, Leaf features, Machine Learning.*

I. INTRODUCTION

There is no living being on the planet that does not use plants in some way. The protection of the environment, food, medicine, oxygen, and shelter they provide is one of their most important functions. Medicines can be found in a variety of plants due to their medicinal properties and active ingredients. As the mother of healing arts, Ayurveda uses medicinal plants found naturally on the Indian subcontinent to treat patients. The ancient Indian sages Charka, Sushruta, and Vaghbhata developed Ayurveda more than 5,000 years ago [30]. Herbal medicine contains therapeutic properties, curing diseases, feeding the mind, soul, and body, according to Acharya Charka.

The World Health Organization (WHO) estimates that 65 to 80% of people worldwide use medicinal plants as treatments for a variety of ailments. [29]. The plants are endangered and have become scarce due to environmental impacts and inadequate knowledge of therapeutic herbs among the general public. Botanists use a plant's biological properties to designate it as a therapeutic plant. Identification of plant species is a drawn-out process that takes more time because two plants may have identical physical traits. Misidentification could harm people's perceptions of Ayurveda treatments and result in unanticipated negative effects. Therefore it is necessary to create an effective system for classifying and recognizing medicinal plants in order to reduce human error and maximize humanity's benefits.

The computerized identification and classification of medicinal plants is essential for enhancing the cultivation of therapeutic plants. This will give farmers and members of the general public access to accurate information. Also, this system offers a species database and specifics on medical information to suppliers, representatives, pharmacy students, pharmaceutical businesses, research students, Ayurvedic practitioners and the cosmetic sector The characteristics of a plant, such as its height, growing region, and environmental conditions, as well as its leaves, flowers, fruits, roots, and stem, are used to identify it. Due to their two-dimensionality and accessibility at any time, many authors simply take plant leaves into account when identifying plants [31]. The related research on using computer vision, machine learning, and image pre-processing to identify and classify medicinal plants is described in the part that follows.

II. LITERATURE REVIEW

Several attempts to create an effective and efficient plant classification system have been performed by researchers during the past 20 years. The fascinating work on plant classification that has been done is briefly reviewed in this section. Characterization of plant leaves by extraction of distinguishing features is an significant task in plant classification.

Wu et al. work is among the most significant in the area of plant classification [24]. Five fundamental geometric elements were used to create twelve morphological features, which were then reduced in dimension by principal component analysis (PCA) so that a probabilistic neural network (PNN) could be employed with fewer inputs. They used the self-generated Flavia dataset, which had an average accuracy of 90.3%.

In [28], Hossain and Amin introduced a technique in which leaves are chopped across and parallel against major and minor axis respectively to obtain a distinctive collection of traits. The ratio between the lengths of the cuts and the leaf lengths (major axis) is then used to normalize the feature points. These characteristics are fed into the probabilistic neural network as inputs. The network was trained by 1200 basic leaves from 30 distinct plant species. The 91.41% average recognition accuracy was obtained using tenfold cross validation.

Using shape features of 20 different plants 400 images dataset, Du et al. (2007) achieved 93% accuracy with the k-nearest neighbor classification technique [25] and The author Du et al. (2009) applied a same classification technique on 2000 images dataset of 20 different leaf species and achieved 92.3% accuracy [26].

In their 2012 study [22], Herdiyeni and Wahyuni combined fuzzy local binary pattern and fuzzy color histogram. The author used Indonesian forests medicinal plants 2448 images dataset and achieved 74.5% classification accuracy with Probabilistic Neural Network.

For the identification and categorization of plant leaves, Du. et al. (2013) developed an strategy which is based on leaf vein and shape features of fractal dimension [20]. The author used 2422 images dataset and extracted 20 features. The 87.1% recognition rate is achieved using kNN classification method.

Backes et al. (2009) used a 2000 images dataset and extract a textural features from a leaf by applying fractal dimension volumetric approach which is a replacement for Fourier analysis and Gabor filters conventional methods. The 89.6% accuracy is attained with linear discriminant analysis (LDA) algorithm [27].

Munisami et al. (2015) used a smartphone camera to capture leaves images from 32 distinct medicinal plants and developed 640 leaves images dataset. They only took into account shape and colour information. The classification technique k-nearest neighbour (kNN) obtained 87.3% accuracy [16].

Hernandez-Serna and Jimenez-Segura (2014) [23] and Chaki et al. (2015) [17] both used Flavia dataset. The author [23] achieved 92.9 % accuracy with Artificial Neural Network classification technique and applied on 16 distinct features texture (8), geometrical (6) and morphological (2). The author [17] used texture and shape vector elements with neuro-fuzzy classification (NFC) technique to obtain 97.6% accuracy.

Siravenha and Carvalho [18] used solely shape features from the Flavia dataset to reach 97.5% accuracy, which is equivalent to Chaki et al [15]. To train and evaluate the ANN approach, the author employed a dataset of 1865 photos and tenfold cross validation.

Carranza-Rojas and Mata-Montero (2016) produced noisy dataset and noiseless dataset [19]. They used only texture and contour information for analysis. In the best case, accuracy achieved 87.2% with kNN classifier. The demonstration showed that photographs shot directly with a smartphone have a sufficient level of accuracy when compared to photographs that were manually edited in a laboratory before being categorised.

The author [21] created a 200 images dataset of medicinal plants from India's western ghat. It consist of 20 different classes each is having 10 leaves of different plants. The kNN classifier applied on HOG (vein features) and SURF (feature descriptor) experiment showed that accuracy achieved around 99.6%.

A different approach to classifying plant leaves is suggested in [11] using the Local Binary Patterns (LBP) method. The suggested method identifies plant leaves by utilising textural features that were collected from the leaves' surface. LBP, which gives the images' R and G colours. Also, the method's effectiveness is measured in comparison to Gaussian, pepper, and salt. Then, the suggested system's features are tested and classified using the Extreme Learning Machine (ELM) technique. This system makes use of the Flavia, Swedish, ICL and Foliage datasets. The acquired outcomes are contrasted to demonstrate that the suggested strategy can discriminate between noiseless and noisy photos. The Flavia, Swedish, ICL, and Foliage datasets (98.94%, 99.46%, 83.71%, and 92.92%) had the best accuracy rates.

The Multiscale Triangle Descriptor and the Local Pattern Histogram Fourier are the foundations of the classification method for plant leaves that the authors explain in [1]. Shape and texture are respectively described using the two ways. For Swedish, Flavia and MEW2012 datasets in their studies, the recognition accuracy was determined to be 98.4%, 99.1% and 95.6% respectively.

A novel method of detection based on Generalized Procrustes Analysis (GPA) is described in [2]. The technique classifies objects using contour features (shape). The main step of the procedure involves doing some math to figure out how far a group of contour points are from the contour's centre after applying specific alignments. For Leafsnap and Flavia datasets, recognition accuracy is 84.4% and 98.4%, respectively.

In [3], the Multiscale Sliding Chord Matching (MSCM) method is proposed to recognize soybean varieties based on common leaf patterns. Shape features are extracted using the MSCM technique. The test with 6000 representative photos reveals a 72.4% hit rate.

An important machine-learning method for classifying and identifying problems and learning from data is SVM. The study in [4] suggested using the centroid and leaf contour to develop leaf image recognition algorithms. The author extracted shape and geometrical properties of leaf using image processing technique from Flavia dataset. SVM classifier achieved 97.7% accuracy.

The author used Flavia dataset for leaf identification and classification in [5]. The 14 leaf features were extracted using a form detector and the result showed 90.9% maximum accuracy with SVM classifier.

SVM and neural networks were utilised by Araujo et al. [6] as classifiers for the classification of leaf images. These classifiers trained four distinct features, including the Zernike Moments, speed of robust features and local binary pattern, using the histogram of gradients (HOG). According to the results, the system's use of numerous classifiers outperformed monolithic approaches and produced the best results. . For ImageCLEF 2011 and ImageCLEF 2012 datasets, recognition accuracy is 86.2% and 64.1%, respectively.

SVM was successful in detecting plants in an environment with significant overlapping and interference circumstances [7]. The scientists used 300 leaf photos from three different plant species in this experiment to identify the plants. The system's identification accuracy is 86.7%. By including more features and expanding the dataset used for the studies, the accuracy can be increased.

In this investigation [8], 180 images vein and shape characteristics were employed and achieved 94.4% accuracy utilizing ANN classifier. The author [9] with same classifier achieved same accuracy using colour, texture and shape features of 63 leaf images.

When thresholding was combined with the ANN as the classifier in [10], 97.3% accuracy achieved which is better than only with ANN classifier. The disadvantages of ANN include its high computing requirement and susceptibility for data overfitting.

The Cosine k-Nearest Neighbors (KNN) classifier and Principal Component Analysis (PCA) technique, according to the study in [12], outperform SVM and Patternnet neural networks. Using the ImageCLEF 2012, leafsnap, and Flavia datasets, their trials revealed recognition accuracy to be 88.80%, 74.50%, and 98.70%, respectively.

The authors of [13] improved leaf classification by combining edge and shape information with a KNN classifier. 32 plant species were provided for testing from the Flavia dataset. According to the findings, the suggested strategy raises the typical classification accuracy to 94.4%.

The authors in [15] discuss the problem of a poor recognition rate of plant recognition since the different classification criteria are applied to plants. This problem is addressed by PNN, which is used as a quick recognition technique on a variety of thirty broad-leaved trees. Broad-leaved trees' shape and textural traits are combined to create a synthetic feature vector, which is used for computer-assisted categorization of broad-leaved plants. The 93.70% recognition rate has been attained by the application of PNN.

In [14], a different PNN-based leaf classification technique is suggested. The smart procedure is utilized to detect the image edges after converting RGB image to its binary and sampling is used to determine the centroid distance. On the Flavia dataset, the technique has an average accuracy rate of 82.1%, while on the Swedish dataset, it has an average accuracy rate of 80.1%. Table 1 summarize the analysis details of classification.

Table I. Leaf Classification Analysis Summary

Author	Year	Dataset	Classifier	Features	Accuracy
Yang et.al. [1]	2021	Flavia *dataset* Swedish *dataset* MEW2012 *dataset*	MTD + LBP-HF	TextureShape	Flavia = 99.10% Swedish = 98.40% MEW2012 = 95.60%
Wang et.al.[3]	2020	6000 *images*	MSCM	Shape	72.40%
Turkoglu et.al.[11]	2019	Flavia *dataset* Swedish *dataset*ICL *dataset* Foliage *dataset*	LBP	Color Texture	Flavia = 98.94% Swedish = 99.46% ICL = 83.71% Foliage = 92.92%
Kherkhah. et.al.[12]	2019	ImageCLEF 2012 *dataset* Leafsnap *dataset* Flavia *dataset*	KNNs	Texture	ImageCLEF 2012 = 88.80% Leafsnap = 74.50% Flavia = 98.70%
Chaudhary et.al. [2]	2018	Leafsnap *dataset* Flavia *dataset*	GPA	Shape	Leafsnap = 84.40% Flavia = 98.40%
Srivastava et.al.[5]	2018	Flavia *dataset*	SVM	Shape	90.90%
Khmag et.al.[4]	2017	Flavia *dataset*	SVM	Shape	97.70%
Araujo et.al. [6]	2017	ImageCLEF 2011 *dataset* ImageCLEF 2012 *dataset*	SVM + Neural Network	Texture Shape	ImageCLEF 2011 = 86.20% ImageCLEF 2012 = 64.10%
Kumar et.al [13]	2016	Flavia *dataset*	KNNs	Shape	94.37%
Nesaratnam et. al.[7]	2015	300 *images*	SVM	Shape	86.70%
Munisami *et al*. [16]	2015	640	kNN	Shape,Colour	87.3

Chaki *et al.* [17]	2015	930	NFC	Shape, Texture	97.6
Siravenha et.al. [18]	2015	1865	ANN	Shape	97.5
Carranza et.al.[19]	2016	2345	kNN	Curvature, Texture	87.2
Mahdikhanlou et.al.[14]	2014	Flavia *dataset* Swedish *dataset*	PNN	Shape	Flavia = 82.01% Swedish = 80.01%
Du *et al.* [20]	2013	2422	kNN	Curvature, Veins	87.1
Sabu et.al. [21]	2017	200	kNN	Vein, feature descriptor	99.6
Herdiyeni et.al.[22]	2012	2448	PNN	Texture, Colour	74.5
Hernandez-Serna et.al. [23]	2014	1800	ANN	Shape, Texture	92.9
Wu et.al. [8]	2006	180 *images*	ANN	Shape Vein	94.40%
Janani et.al.[9]	2013	63 *images*	ANN	ShapeColor Texture	94.40%
Fu et. al.[10]	2007	2940 *images*	ANN	Color Vein	97.33%
Wu *et al.* [1]	2007	1800 *images*	PNN	TextureColor Shape	90.3
Du *et al.* [25]	2007	400	kNN	Shape	93
Du *et al.* [26]	2009	2000+	kNN	Shape	92.3
Backes *et al.* [27]	2009	2000	LDA	Texture	89.6
Hossain et.al. [28]	2010	1200	PNN	Shape	91.4
Huang et. al. [15]	2008	900 *images*	PNN	Shape Texture	93.70%

III. ANALYSIS AND FUTURE SCOPE

Our analysis found that the classification methods are concentrates on a number of issues, most widely used issues are leaf features, the testing datasets, type of classifiers and how they impact accuracy rate. Researchers have included the ensuing qualities into their techniques: Shape, colour, vein, and texture. Also, we discovered several studies integrate many parameters to increase the accuracy ratio. These characteristics are thoroughly addressed and examined, and their effectiveness in improving the recognition and categorization process is demonstrated.

The accuracy of manual identification processes is heavily reliant on human skill. Manual identification of medicinal plants is time-consuming and prone to human mistake. Automatic plant identification could overcome these issues, but developing such a system would necessitate a significant time and financial investment, as well as a detailed grasp of plant morphology. The majority of recent research on autonomous plant identification systems evaluates their effectiveness by examining pre-existing datasets generated in a controlled environment. As a result, more research into photos shot in different lighting conditions and against complicated backgrounds is required.

IV. CONCLUSION

In our specific analysis, we looked at and addressed a number of aspects that could have an impact on classification accuracy. These aspects include characteristics, classifiers, and testing datasets. According to our findings, adding additional features improves the categorization process. Furthermore, using a sizable dataset is advised to enhance training. The accuracy of the developed identification method would consequently rise. Furthermore, the study discovered that using a variety of elements, such as shape, vein, color, and texture, had a substantial impact on an object's potential classification effectiveness. Because of improved accuracy, the use of medicinal plants in the medical industry may rise, and environmental preservation would be extremely advantageous.

REFERENCES

[1] C. Yang, "Plant leaf recognition by integrating shape and texture features", Pattern recognition, vol.112, p. 107809, 2021.

[2] S. Choudhury, J. Yu and A. Samal, "Leaf recognition using contour unwrapping and apex alignment with tuned random subspace method", Biosystems engineering, vol. 170, pp. 72-84, 2018.

[3] B. Wang, Y. Gao, X. Yuan and S. Xiong, "From species to cultivar: Soybean cultivar recognition using joint leaf image patterns by multiscale sliding chord matching", Biosystems engineering, vol. 194, pp. 99-111, 2020.

[4] A. Khmag, S. Al-Haddad and N. Kamarudin, "Recognition system for leaf images based on its leaf contour and centroid," 2017 IEEE 15th student conference on research and development, pp. 467–472, 2017.

[5] V. Srivastava and A. Khunteta, "Comparative analysis of leaf classification and recognition by different SVM classifiers," 2018 International conference on inventive research in computing applications, pp. 626–631, 2018.

[6] Araujo, A. Britto, A. Brun, A. Koerichm and R. Palate, "Multiple classifier system for plant leaf recognition," 2017 IEEE international conference on systems, man and cybernetics, pp. 1880–1885.

[7] R. Nesaratnam and C. Murugan, "Identifying leaf in a natural image using morphological characters," International conference on innovations in information, embedded and communication systems, 2015.

[8] Q. Wu, C. Zhou and C. Wang, "Feature extraction and automatic recognition of plant leaf using artificial neural network," Research on computing science, vol. 20, pp. 3–10, 2007.

[9] R. Janani and A. Gopal, "Identification of selected medicinal plant leaves using image features and ANN," 2013 International conference on advanced electronic systems, pp. 238–242, 2013.

[10] H. Fu and Z. Chi, "Combined thresholding and neural network approach for vein pattern extraction from leaf images," IEE Proc. - Vision, image signal process, vol. 153, pp. 881–892, 2007.

[11] M. Turkoglu and D. Hanbay, "Leaf-based plant species recognition based on improved local binary pattern and extreme learning machine," Physical A: statistical mechanics and its applications, vol. 527, p. 121297, 2019.

[12] F. Kheirkhah and H. Asghari, "Plant leaf classification using GIST texture features," IET computer vision, vol. 13, p. 369, 2018.

[13] P. Kumar, K. Rao, A. Raju, and D. Kumar, "Leaf classification based on shape and edge feature with k-NN classifier," 2016 2nd Int. Conf. Contemp. Comput. Informatics, pp. 548–552, 2016.

[14] K. Mahdikhanlou, and H. Ebrahimnezhad, "Plant leaf classification using centroid distance and axis of least inertia method," 2014 22nd Iran. Conf. Electr. Eng., pp. 1690–1694, 2014.

[15] L. Huang, and P. He, "Machine recognition for broad-leaved trees based on synthetic features of leaves

Using Probabilistic Neural Network," Proc. - Int. Conf. Comput. Sci. Softw. Eng. CSSE 2008, vol. 4, pp. 871– 877, 2008.

[16] Munisami, T., Ramsurn, M., Kishnah, S. and Pudaruth, S., 2015. Plant leaf recognition using shape features and colour histogram with k- nearest neighbour classifiers. Procedia Computer Science, 58, 740-747.

[17] Chaki, J., Parekh, R. and Bhattacharya, S, 2015. Plant leaf recognition using texture and shape features with neural classifiers. Pattern Recognition Letters, 58, 61-68.

[18] Siravenha, A.C.Q. and Carvalho, S.R., 2015. Exploring the use of Leaf Shape Frequencies for Plant Classification. 28th SIBGRAPI Conference on Graphics, Patterns and Images, Salvador, Brazil, 297-304.

[19] Carranza-Rojas, J. and Mata-Montero, E., 2016. Combining Leaf Shape and Texture for Costa Rican Plant Species Identification. CLEI Electronic Journal, 19(1), Paper 7.

[20] Du, J.X., Zhai, C.M. and Wang, Q.P., 2013. Recognition of plant leaf image based on fractal dimension features. Neurocomputing, 116, 150- 156.

[21] A.Sabu, K. Sreekumar, and R. R. Nair. Recognition of ayurvedic medicinal plants from leaves: A computer vision approach. In 2017 Fourth International Conference on Image Information Processing (ICIIP), pages 1–5, Dec 2017

[22] Herdiyeni, Y. and Wahyuni, N.K.S., 2012. Mobile Application for Indonesian Medicinal Plants Identification using Fuzzy Local Binary Pattern and Fuzzy Color Histogram. International Conference on Advanced Computer Science and Information Systems (ICACSIS), West Java, Indonesia, 301-306.

[23] Hernandez-Serna, A. and Jimenez-Segura, L.F., 2014. Automatic Identification of species with neural networks. PeerJ 2:e563; doi:10.7717/peerj.563.

[24] Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F. and Xiang, Q.L., 2007. A Leaf Recognition Algorithm for Plant Classification using Probabilistic Neural Network. 7th IEEE International Symposium on Signal Processing and Information Technology, Giza, Egypt, 11-16.

[25] Du, J.X., Wang, X.F. and Zhang, G.J., 2007. Leaf shape based plant species recognition. Applied Mathematics and Computation, 185, 883- 893.

[26] Du, M., Zhang, S. and Wang, H., 2009. Supervised Isomap for Plant Leaf Image Classification. 5th International Conference on Emerging Intelligent Computing Technology and Applications, Ulsan, South Korea, 627-634.

[27] Backes, A.R., Casanova, D. and Bruno, O.M., 2009. Plant Leaf Identification based on Volumetric Fractal Dimension. International Journal of Pattern Recognition and Artificial Intelligence, 23(6), 145-1160.

[28] Hossain, J. and Amin, M.A., 2010. Leaf Shape Identification Based Plant Biometrics. 13th International Conference on Computer and Information Technology, Dhaka, Bangladesh, 458-463.

[29] Naresh, Y. G., & Nagendraswamy, H. S. (2016). Classification of medicinal plants: An approach using modified LBP with symbolic representation. Neurocomputing, 173, 1789–179.

[30] Zhang, S., Lei, Y.-K., Dong, T., and Zhang, X.-P. (2013). Label propagation based supervised locality projection analysis for plant leaf classification. Pattern Recognition, 46(7):1891{189

[31] Zhu, S.-P., Du, J.-X., and Zhai, C.-M. (2013b). Content-based diversifying leaf image retrieval. In ICIC (2), pages 39{46.