



# MEDICINAL PLANTS IDENTIFICATION USING MACHINE LEARNING CLASSIFICATION TECHNIQUES

<sup>1</sup>R.Kiruthika, <sup>2</sup>M.Kousalya, <sup>3</sup>. Mr.C.Rathnakumar

<sup>1</sup>II MCA, Paavai Engineering College, Namakkal

<sup>2</sup> II MCA, Paavai Engineering College, Namakkal

<sup>3</sup> Professor, Department of MCA, Paavai Engineering College, Namakkal

**Abstract:** In the traditional Indian medical sector, it is crucial to identify the proper medicinal plants used in the manufacturing of a drug. A medicinal plant can be recognized primarily by its leaf shape, color, and texture. Deterministic parameters to identify the species are present in the color and texture of the leaf on both sides. To identify the optimal mix of features that maximizes the recognition rate, this research investigates feature vectors from both the front and back sides of a green leaf in addition to morphological data. Images of the front and back of the leaves of frequently used ayurvedic medicinal plants are used to build a database of medicinal plant leaves. Using a special feature combination, the leaves are categorized.

**Keywords:** Machine Learning, Support vector machine, Decision Tree, MLP, Bagging with J48

## I. INTRODUCTION

More than 80% of plants utilized in ayurvedic definitions are gathered from the woods and Badlands though the excess is developed in rural terrains [1]In the antiquated past, the Ayurvedic doctors themselves picked the restorative plants and arranged the prescriptions for their patients. Today a couple of experts follow this training. The assembling and promoting of Ayurvedic drugs have turned into a flourishing industry. The quantity of authorized Ayurvedic medication producers in India effectively surpasses 100 billion. This business in Ayurvedic area has acquired concentrated a few inquiries concerning the nature of unrefined substances utilized for Ayurvedic prescriptions. Today the plants are

gathered by ladies and kids from woodland regions; those are not expertly prepared in recognizing

The right restorative plants. Fabricating units frequently get inaccurate or subbed restorative plants. The majority of these units need satisfactory quality control components to screen these plants. Likewise, disarray because of varieties in the nearby name is additionally widespread. A few plants show up in dried structure and this makes the manual recognizable proof errand substantially more troublesome. Erroneous utilization of restorative plants makes the Ayurvedic medication incapable. It might create capricious side results moreover. In this present circumstance, severe measures for quality control should be authorized on Ayurvedic medications and unrefined substances involved by the business to support the current development of the industry by keeping up with the viability and validity of prescriptions [2]

A prepared Botanist searches for every one of the accessible elements of the plants, for example, leaves, blossoms, seeds, roots and stem to distinguish plants. Aside from the leaf, all others are 3D items and increment the intricacy of examination by PC. Notwithstanding, plant passes on are 2D articles and convey adequate data to distinguish the plant. Leaves can be gathered effectively and picture obtaining might be done utilizing economically advanced cameras, cell phones, or record scanners. It is accessible whenever of the year rather than blossoms and seeds. Leaves obtain a particular tone, surface, and shape when it develops and these progressions are somewhat irrelevant. Plant acknowledgment in light of leaves relies upon tracking down

precise descriptors and separating the component vectors from them. Then the element vectors of the preparation tests are contrasted and the component vectors of the test to find the level of comparability utilizing a proper classifier.

## II. RELATED WORKS

Parag Bhandarkar, Rizwan Ahmed, et al. [3] deteriorated the morphology of leaf edges utilizing predefined primary components and separated an underlying mark which measures the leaf shape highlight. They utilized the root mean square mistake between the component vectors of the information picture and the picture in the data set for processing the character. The information base made by the creators comprises 40 leaf tests of 10 distinct species. They accomplished a general characterization pace of 67.5%, which is free of leaf size and direction. The recognizable proof rate is relatively low to be useful in pragmatic executions.

T. Sathwik, R. Yaraswini, et al. [4] fostered a plant identification strategy utilizing surface highlights alone. They separated 10 textural highlights from GLCM of the leaf picture and utilized these for characterization utilizing the least disparity technique. The framework was prepared with 63 and tried with 33 leaves and accomplished an exactness of 95%. At the point when they utilized a blend of converse contrast second, entropy, totally normal, and different change, a precision of 91% is accomplished. The technique was tried exclusively on a set number of tests and was not completely invariant to the turn of leaf.

NurilAslina, Nursuriati Jamil, et al. [5] utilized Scale Invariant Feature Transform (SIFT) as a shape descriptor and variety of moments. The picture is disintegrated into HSV planes and each plane is partitioned into 9 networks. Variety minutes are determined for every matrix of every plane and utilized as a component vector. The least Euclidean distance among test and preparing sets are utilized for recognizable proof. The data set is made by the creators by procuring 40 leaf pictures of Malaysian spices from a regular environment in normal light. A precision of 87.5% is gotten autonomous of scaling and revolution of pictures. The filter is computationally serious when used to separate central issue highlights.

ItheriYahiaoui, OlfaMzoughi, et al. [6] utilized five geometrical highlights and a limit descriptor named Directional Fragment Histogram for distinguishing proof. They utilized 3070 filtered and 897 sweep-like pictures of plant leaves dataset for a try. For examined pictures, a precision of 77.83%, and for a filter like pictures, an exactness of 67.47% is acquired.

Pavan Kumar Mishra, Sanjay Kumar Maurya, et al. [7] utilized four mathematical elements i.e., convexity, robustness, whimsy, and circularity alongside three RGB variety highlights as redness, greenness, and blueness lists for the experiment. The information base gathered by the creators, comprises of 1000 leaf pictures. A three-phase correlation of element vectors is utilized to speed up the recognizable proof interaction. They accomplished a distinguishing proof pace of 85%.

Gopal, S. Prudhveswar Reddy, et al. [8] utilized four mathematical highlights, HU invariant minutes and Polar Fourier change coefficients as shape descriptors for distinguishing proof. They utilized mean, standard deviation, skewness, and kurtosis of RGB variety planes as variety descriptors. The framework was prepared with 100 and tried with 50 checked leaf images. By and large grouping proficiency of 92% is gotten. The strategy is computationally concentrated and isn't completely invariant to pivot.

S. Harish, Aditi Hedge, et al. [9] utilized viewpoint proportion, structure factor, rectangularity, edge proportion of distance across, robustness, convexity, and abnormality as mathematical highlights, and Zernike minutes as a shape descriptors. The examinations were directed on Flavia's information base and a data set made by the creators yielding an exactness of 89%. They analyzed the presentation of four unique classifiers given mathematical and Zernike second capabilities independently and arranged the outcomes. It is seen that Zernike minutes gave better exactness in every one of the classifiers when contrasted with mathematical descriptors. Gullible Bayes classifier, K-NN, support vector machine, and PNN classifiers were utilized in the analysis. From the past work, mathematical, variety, surface highlights of the leaf are utilized to distinguish the plant.

### III. MATERIALS AND METHODS

No standard information bases of Ayurvedic therapeutic plant leaves were accessible for directing the trials. A bunch of therapeutic plant leaf pictures was gathered from a confidential natural garden. 20 leaves were gathered irregularly from 40 different plant species utilized for Ayurvedic, homegrown, and society meds. The leaves were gathered from their normal living space also, and the determination of leaves and plants was very arbitrary. The leaves were fanned out on a common report scanner and filtered with a most noteworthy conceivable goal. Both the front and rears were filtered and the pictures were put away in jpeg design for additional handling. The pictures of individual leaves were then isolated, changed to a uniform goal of 1600×1200, and organized into isolate sets.

#### III.I Classification

**Decision Tree:** Data mining as often as possible utilizes the choice tree (DT), a famous learning procedure. At the point when we discuss DT, we're discussing a progressive, prescient model that utilizes the thing's perceptions as branches to get to the thing's objective worth in the leaf. DT is a tree having leaf hubs that mirror the choice and choice hubs that have a few branches.

**Bagging:** In request to build the exactness of unsteady arrangement frameworks, Breiman [10] made sacking, alongside bootstrap and accumulation draws near. For packing, a choice tree is developed utilizing  $X$  examples and  $X$  bootstrap datasets with  $X$  haphazardly picked models and substitution from  $Y$ . By a greater part vote, the normal new example is not entirely settled. The results of contrasting new cases with  $X$  choice trees are reported. Albeit a solitary choice tree's clear understanding is lost, packing builds the accuracy of classification rules. In this examination, choice stump and sacking with J48 are used "Boosting" is utilized to build a specific learning calculation's exactness. An AI strategy called "supporting" distinguishes and joins free guidelines to get an exact classification. The supporting strategy utilizes an assortment of preparing set subsets from the base figuring out how to find governs more than once.

**AdaBoost:** The main compelling supporting strategy for parallel order was called AdaBoost, short for Adaptive Boosting. It is a calculation

for regulated AI and is utilized to work on the presentation of all AI calculations. Like choice trees, it works best with hesitant understudies. These are models whose order exactness is somewhat better compared to arbitrary possibility.

**Support vector machine:** Support vector machines contain a calculation, which reasoning a capability from given information. This strategy is utilized for grouping and returning investigation of the information. Choice limits are characterized by choice planes. These planes separate arrangements of items in various classes. SVM is working in light of these planes. They are playing out the straight arrangement. The principal property of SVM is Duality.

**Multilayer perceptron:** A Multilayer perceptron (MLP) is a sort of feed-forward counterfeit brain organization. It courses the info information in one set to suitable results in one more set. Coordinated diagram comprising numerous hubs. An MLP contains many layers of this hub. The speciality of these layers is that they are associated with each other. We can think about every hub in a layer as a neuron. For preparing the organization MLP uses a back engendering procedure and this strategy is a regulated learning technique. Information that is not directly detachable can be recognized by MLP[11]

### IV. PROPOSED SYSTEM

The framework has two phases of activity: the preparing stage and the testing stage. In the preparation stage, the pictures in the leaf data set are given as a contributions to the framework individually. The pre-handling step guarantees that every one of the pictures has the same standard goal. The following stage is to change over the variety of picture to grayscale and afterward to parallel pictures. The twofold picture goes through a morphological disintegration and widening process that eliminates little defects like specks and breaks. The biggest part of the double picture is chosen for tracking down morphological highlights. The element extraction stage separates morphological, variety, and surface highlights from parallel and variety pictures individually and stores these qualities in another data set. This is rehashed for every one of the leaves in the leaf data set. A classifier is prepared with

chosen include values from the component data set.

In the testing stage, a question picture is introduced to the framework on a white foundation. It goes through every one of the stages that each preparing picture has gone through and chosen highlights are separated from it. At the point when these qualities are given as a contributions to the classifier, it settles on a choice whether the test coordinates with any of the classes with which it was prepared before. The name of the class in that the leaf has a place is the result of the classifier. The highlights are examined with the assistance of the Weka classifier device to figure out the ideal mix of mathematical, variety, and surface elements that boosts the

recognizable proof rate utilizing green leaves. Likewise, the least list of capabilities that amplifies the ID rate utilizing dry leaves is additionally completed.

At first, features are extracted from the front and posterior of each leaf picture. The examination is led utilizing the pre-handling and visualization devices accessible in Weka on the mathematical, variety, and surface highlights to figure out the general significance of every one of the elements on the recognizable proof cycle. The examination of the 16 variety features from the front side brought about a decreased list of capabilities of 8. The viable elements for separating a leaf are the mean and standard deviation of HSV variety planes.



Fig: 1 Front and back side of the leaf images

**V. RESULTS AND CONCLUSION**

In this proposed strategy, 300 examples are taken for preparing and testing. The following classifiers give the better results

Classifier Used	Classification Accuracy %
Bagging with J48	89%
Multilayer Perceptron	92%
Support Vector Classifier	91%
Decision Tree	93%
Adaboost	94%

The adequacy of individual capabilities alongside their mixes is then broke down utilizing Adaboost classifiers to find the ideal blend, which is then tried with various classifiers. Various classifiers were utilized to distinguish the plants utilizing dry and green leaves. Among this Adaboost Classifier achieve more accuracy.

**References**

[1]D. G. Adler and C. J. Gostout, “Wireless capsule endoscopy”, *Hospital Physician*, pp. 14-22, 2003.

[2]L. Baopu and M. Q. H. Meng, “Computer-aided detection of bleeding regions for capsule endoscopy images”, *IEEE Trans. Biomedical Engineering*, vol. 56, no. 4, pp. 1032-39, Apr. 2009.

[3]S. Liangpunsakul, L. Mays and D. K. Rex, “Performance of given suspected blood indicator”, *American Journal of Gastroenterology*, vol. 98, no.12, pp. 2676-8, Jan. 2004.

[4]J. M. Buscaglia et. al., “Performance characteristics of the suspected blood indicator feature in capsule endoscopy according to indication for study”, *Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association*, vol. 6, no. 3, pp. 298-301, Mar. 2008.

[5]GuolanLv, Guozheng Yan and Zhiwu

- Wang, "Bleeding detection in wireless-capsule endoscopic images based on color invariants and spatial pyramids using support vector machines", *33rd annual international conference of IEEE EMBS Boston, USA*, pp. 6643-46, 2011.
- [6] Y. Fu, M. Mandal and G. Guo, "Bleeding region detection in we images based on color features and neural network", in *Proc. IEEE 54th Int. Midwest Symp. Circuits Syst.*, pp. 1-4, Aug. 2011.
- [7] Jie Li, Jinwen Ma, TammamTillo, Bailing Zhang and Eng Gee Lim, "A training based support vector machine technique for blood detection in wireless capsule endoscopy images", *IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 826-830, Dec. 2012.
- [8] S. Sainju, F. M. Bui and K. Wahid, "Bleeding detection in wireless capsule endoscopy based on color features from histogram probability", in *Proc. of CCECE.*, pp. 1-4, 2013.
- [9] Yanan Fu, Wei Zhang, Mrinal Mandal and Max Q.H.Meng, "Computer-aided bleeding detection in WCE video", *IEEE journal of biomedical and health informatics*, vol. 18, no. 2, Mar. 2014.
- [10] Leo Breiman, (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.
- [11] Sasi Kumar. M and Y.S. Kumaraswamy, "Medical Image Retrieval System using an Improved MLP Neural Network", *European Journal of Scientific Research*, Vol.66, No.4, pp. 532-540, 2011.