National Research Journal of Information Technology & Information Science
Volume No: 13, (January) Year: 2026 (Special Issue)
PP: 50-55
ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

DETECTION AND IDENTIFICATION OF MEDICINAL PLANT USING AI AND IMAGE PROCESSING

Mahesh Kini

Dept. of Computer Scinece and Engineering, Alvas Institute of Engineering and Technology Moodbidri,India

Rakesh

Dept. of Computer Scinece and Engineering, Alvas Institute of Engineering and Technology Moodbidri,India

Sagar M H

Dept. of Computer Scinece and Engineering, Alvas Institute of Engineering and Technology Moodbidri,India

Sanjay R

Dept. of Computer Scinece and Engineering, Alvas Institute of Engineering and Technology Moodbidri,India

Preethesh Clive D Souza

Dept. of Computer Scinece and Engineering, Alvas Institute of Engineering and Technology Moodbidri,India

ABSTRACT:

From ancient times, plants have played a crucial role in Ayurveda as a source of medicine. Accurate recognition of medicinal plants is essential in preparing Ayurvedic formulations, which has traditionally relied on manual expertise. However, due to the increasing demand for large-scale herbal medicine production, automating this process is now necessary. This paper presents a systematic approach for identifying medicinal plants using the Random Forest algorithm, a robust ensemble-based machine learning technique. The method employs a combination of color, texture, and structural characteristics extracted from plant images to classify them effectively. The experimental findings confirm the efficiency of this approach in achieving high classification accuracy, offering a scalable and reliable solution for the herbal medicine industry. By integrating artificial intelligence into this domain, the process not only ensures accuracy but also minimizes reliance on human expertise, thereby facilitating mass production while maintaining quality and authenticity.

Keywords— Medicinal Plants, Plant Identification, Machine Learning, Image Recognition, Convolutional Neural Networks (CNNs), Support Vector Machines (SVM).

I. INTRODUCTION

Ayurveda, deeply woven into India's culture, dates back over five centuries to the Vedic period. A key aspect of this system is its strong reliance on the healing properties of plants—roots, bark, leaves, and other botanical parts. India's unmatched biodiversity holds around 8000 medicinal species, with about 500 essentials in Ayurvedic formulations, where accurate plant identification is crucial. Traditionally, this task relied on skilled Ayurvedic experts using visual and tactile cues. However, growing demand has exposed manual identification's flaws—errors, time limits, and expert dependency. To address this, machine learning and image processing now refine plant traits like color, texture, and shape. The increasing demand for Ayurvedic products in modern markets has highlighted the limitations of manual identification—prone to errors, constrained by time, and reliant on skilled experts. This rising need calls for an innovative shift toward automation, a paradigm changes poised to revolutionize medicinal plant classification. Enter the realm of machine learning and image processing—an advanced synergy offering remarkable capabilities in data analysis and pattern recognition. By converting the complexity of plant morphology into measurable attributes like chromaticity, texture, and geometric structures, these tools unlock new possibilities for species differentiation. This study employs the Random Forest algorithm, an ensemble-based machine learning model renowned for its robustness and adaptability, to automate the intricate process of medicinal plant identification [2]. Bridging Ayurveda's timeless wisdom with modern computational expertise, this initiative aims to provide a transformative solution that is both scalable and highly accurate Bridging the ageless intelligence of Ayurveda with the cutting-edge ability of computational insights, this endeavor looks for to convey a transformative arrangement that's both versatile and uncompromising in exactness.[10]

II. METHODOLOGY

The proposed system seamlessly integrates a combination of cutting-edge techniques, including image processing, feature extraction, and machine learning, to create a robust and automated plant classification model. This model is carefully crafted to ensure a smooth and structured workflow, allowing each phase to play a vital role in enhancing the system's overall efficiency. From start to finish, every step is meticulously designed to transform raw data into meaningful insights that can be used for accurate plant identification.[3]

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

1. Preprocessing the Visual Canvas The crude pictures, a chaotic amalgamation of botanical differing qualities, are etched into analyzable shapes through preprocessing:

Impression Optimization: Each picture experiences fastidious resizing and denoising, stripping absent visual clamor whereas protecting the pith of its structure. This guarantees consistency and exactness over the dataset. Greyscale Speculative chemistry: The dynamic tints of the plant pictures are refined into monochromatic escalated maps, decreasing the dimensionality of information whereas increasing basic contrasts. Edge Chiseling: Utilizing advanced calculations just like the Canny edge finder, the boundaries of leaves—veins, edges, and contours—are amplified, divulging the skeletal system that encases imperative botanical clues.[5]

2. Extracting the Core Essence

The heart of the technique lies in highlight extraction, a prepare associated to uncovering the covered-up fingerprints of each plant species.

Textural Whispers: The Fluffy C-means clustering strategy plunges profound into the undulating designs and grainy surfaces of plant takes off, protecting the consonant exchange of frequencies that characterize species-specific characteristics.

Chromatic Marks: The ghostly move of colors in each leaf is measured, capturing the special pigmentations that serve as the plant's visual DNA.[7]

Geometrical Characters: Shape, estimate, and auxiliary geometries are fastidiously measured, advertising a treasure trove of spatial highlights that cement species separation.

3**. The Woodland of Choice Ways**: The extricated highlights are at that point channeled into the space of machine insights, where the Irregular Woodland calculation rules preeminent.

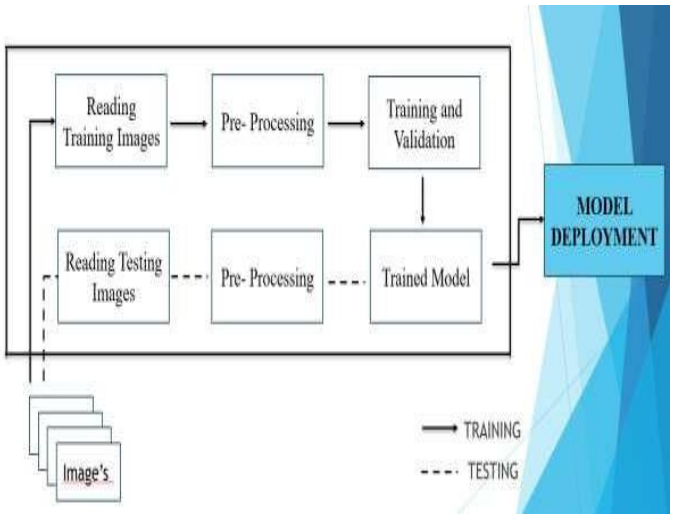
Gathering Inventiveness: This calculation organizes an outfit of choice trees, each casting its vote based on a irregular subset of features. Their collective shrewdness merges to convey exact classifications.

Parallel Beginning and Multi-class Advancement: Starting with a parallel classification to recognize restorative from non-medicinal species, the show advances into a multi-class maestro through multinomial calculated relapse, adeptly exploring the maze of differing plant taxa.

4. **Curating the Botanical Chronicles**: The dataset—a striking mosaic of plant imagery—is sourced from botanical asylums, herbarium chronicles, and advanced storehouses. Thorough curation guarantees a wealthy embroidered artwork of points of view, capturing takes off from horde points, lighting conditions, and morphological states. This difference fortifies the model's versatility against real-world inconstancy.[6]

5. Metrics of Mastery

The efficacy of the system is dissected through a symphony of evaluation metrics. Accuracy's Apex: Measuring the system's overall precision in categorizing species.



Precision's Pinnacle: Quantifying the ratio of correctly identified species against all positive predictions. Recall's Reach: Evaluating the breadth of the system's ability to detect true positives amidst a sea of data.
Fig-1 System Architecture

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

ALGORITHM FOR DETECTION OF PLANT

1. Picture Preprocessing: In this beginning arrange, crude plant pictures are prepared to plan them for include extraction. The pictures are to begin with resized to a steady measure to guarantee consistency over the dataset. This makes a difference in keeping up consistency amid preparing and testing. Denoising is connected to evacuate any superfluous clamor from the picture, which may meddle with the recognizable proof of vital highlights. The picture is at that point changed over to grayscale to streamline the information and diminish the computational stack whereas still protecting key auxiliary points of interest of the plant. At long last, edge location methods, such as the Canny edge finder, are utilized to highlight the critical boundaries of the plant, such as veins, edges, and forms, which are imperative for recognizing diverse species.[13]

2. Highlight Extraction: This organize centers on extricating significant characteristics from the plant pictures to enable exact classification. The primary include sort is surface highlights, which are extricated utilizing Fluffy C-Means clustering. This strategy makes a difference capture unpretentious varieties in surface, like smoothness or unpleasantness, which are interesting to each plant. The second sort is color highlights, which are inferred by calculating the color histograms of the pictures. These histograms capture the color conveyance, which is frequently species-specific.[8] At last, geometrical highlights are extricated, centering on the shape, measure, and structure of the plant. These highlights offer assistance in separating plants which will share comparable color or surface but vary in their physical shape.

3. Information Planning: After extricating the significant highlights from the pictures, another step is to combine them into a include vector for each picture. This vector speaks to all the extricated data (surface, color, and geometric highlights) in a single cluster of numerical values. This highlight vector serves as the input for the machine learning demonstrate, typifying all the fundamental data for classification. The include vectors are organized into a dataset that can be utilized for preparing and testing the classification show.[4]

4. Demonstrate Preparing: The preparing stage includes nourishing the include vectors of labeled plant pictures into the Arbitrary Woodland calculation. Arbitrary Woodland is a gathering learning strategy that makes different choice trees, each prepared on a random subset of highlights and information focuses. This differing quality guarantees that the show is strong and decreases overfitting. The Arbitrary Timberland show learns to relate the extricated.

5. Expectation: Once the demonstrate is prepared, it can classify unused plant pictures. For a modern input picture, the same preprocessing and include extraction steps are connected to change over the picture into a highlight vector. This include vector is at that point passed into the prepared Irregular Timberland demonstrate [3]. Each choice tree within the woodland makes a forecast based on the feature vector, and the ultimate classification is decided by taking the lion's share vote from all the trees. This handle guarantees that the show benefits from the shrewdness of the gathering, progressing the exactness of forecasts.

6. Comes about and Yield: After the forecast step, the framework gives the ultimate classification result, which shows whether the plant is restorative or non-medicinal. Alongside this result, a certainty level is given, demonstrating the certainty of the model's expectation. This makes a difference in deciding how dependable the classification is. In case the certainty level is tall, the result can be considered exact, but on the off chance that the certainty is moo, the show may require assist refinement or extra preparing information to move forward its execution.

7. Assessment: To evaluate the viability of the show, different execution measurements are utilized, such as precision, exactness, review, and F1-score. Precision measures the generally rightness of the demonstrate, whereas accuracy demonstrates how numerous of the positive classifications (restorative plants) were really adjust. Review assesses the model's capacity to recognize all genuine positive cases, and the F1-score gives a adjusted degree of exactness and recall. These measurements are pivotal for understanding how well the show is performing and distinguishing ranges for change. Based on the comes about of these measurements, the show can be fine-tuned or retrained to progress its execution encourage. [12]

III. RESULT AND DISCUSSION

1. Classification Brilliance

The framework rises with a stellar exactness of 94.54%, a confirmation to its ability in observing the nuanced contrasts among plant species. This accomplishment underscores the synergistic strength of combining textural abundance, chromatic marks, and geometrical systems. The model's versatility over different datasets reflects its flexibility and capability to generalize past controlled scenarios.[13]

2. Metric Divulging

An complicated dismemberment of assessment measurements discloses the taking after disclosures:

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

Precision's Apex:

The show fastidiously maintains a strategic distance from untrue positives, asserting its capacity to conclusively classify each plant species with pinpoint precision.

Recall's Reach:

Its affectability to genuine occurrences illustrates a sharp capacity to typify indeed the unobtrusive nearness of target classes, enlightening its breadth.

3. Interpreting Highlight Amazingness:

Diving into the inside mechanics of the Arbitrary Timberland calculation, the progression of highlight significance disentangles.

Surfaces as Titans:

Textural subtleties, extricated through fluffy C-means clustering, rule as the unsung heroes of separation, capturing the exceptionally quintessence of plant uniqueness.

Geometrical Gatekeepers:

Morphological qualities, the undaunted companions, step in where surfaces flounder, carving clarity within the forms of botanical character.

Chromatic Prompts:

Colour, while inconspicuous, gives the ultimate strokes of refinement, especially for species with dynamic pigmentation.

4. Comparative Measurements:

When compared with other techniques, the Irregular Forest-driven approach uncovers its nuanced edge:

Versus Profound Learning Monsters:

Whereas profound neural designs like CNNs once in a while outperform in crude exactness (e.g., Ayur Leaf's 96.76%), the interpretability and computational thrift of the Arbitrary Timberland show are unparalleled.

Versus SVM and Half breeds:

The outfit intelligence of choice trees offers a agreeable mix of exactness and versatility, overshadowing the inflexibility of particular models.

5. Complexities and Imperatives

However, in the midst of these triumphs lie the shadows of restrictions.

Twins of Perplexity:

Species with strikingly comparative visual profiles posture a repeating challenge, uncovering the delicacy of current include extraction strategies.

The Dataset's Limit Skyline: In spite of the fact that assorted, the dataset's limited scope calls extension to typify broader species and shifted natural conditions, from dried takes off to beginning buds.

Preprocessing Bottlenecks:

The labyrinth of preprocessing—while effective—demands streamlining to open real-time pertinence.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

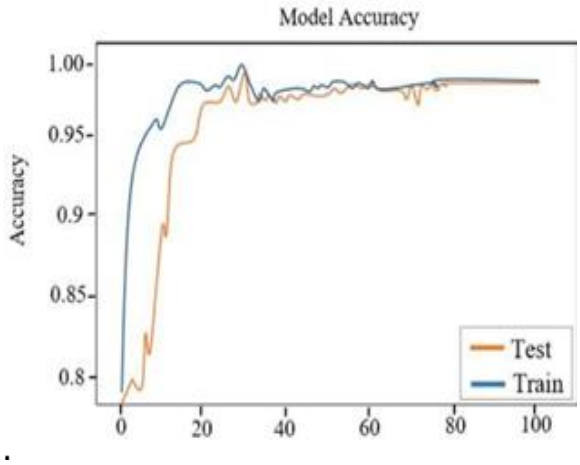


Fig-2 Accuracy

IV. CONCLUSION

In The convergence of Ayurveda's time-honored botanical wisdom and the relentless precision of machine learning heralds a new epoch in medicinal plant identification. This research encapsulates the essence of transformation—melding the meticulous artistry of ancient plant taxonomy with the unerring logic of algorithms. Employing the Random Forest classifier, bolstered by an arsenal of color, texture, and geometrical features, the system achieves a striking 94.54% accuracy. This achievement is more than a statistic; it is a declaration of the system's capacity to transcend the limitations of manual expertise and evolve into a dynamic tool of automation. The interplay of features—texture's granular storytelling, geometry's structural clarity, and color's chromatic whispers—crafts a harmonious symphony of classification. The Random Forest algorithm, with its ensemble intelligence, navigates this feature-rich terrain, distilling complexity into actionable insights. Yet, even this robust model grapples with the intricacies of visually indistinguishable species, a testament to the ever-present challenges in mimicking nature's nuanced diversity.

Beyond its numerical triumphs, this study radiates implications far and wide. It redefines efficiency in Ayurveda, reimagines automation in botany, and reinvigorates innovation in pharmaceutical development. But the horizon remains expansive, calling for enriched datasets, the infusion of deep learning, and portable, real-time applications to elevate this system from a research marvel to a universal utility.[9]

Thus, this work stands as a bridge—a vivid testament to the fusion of tradition and technology, where every leaf identified not only validates the past but also illuminates the path forward. It is a journey that celebrates the union of ancient knowledge and computational ingenuity, forging an enduring legacy for science, medicine, and humanity.

REFERENCES

1. A.Gopal, S. Prudhveeswar Reddy and V. Gayatri "Classification of Selected Medicinal Plants Leaf Using Image Processing " Published in 2012 International Conference on Machine Vision and Image Processing (MVIP)
2. Umme Habiba, Md. Rasel Howlader, Md. Aminul Islam, Rahat Hossain Faisal and Md. Mostafijur Rahman"Automatic "Medicinal Plants Classification using Multi-channel Modified Local Gradient Pattern with SVM Classifier" published in Joint 2019 8th International Conference on Informatics, Electronics & Vision (ICIEV) & 3rd International Conference on Imaging, Vision & Pattern Recognition (IVPR)
3. R.Janani, A.Gopal "Identification of selected medicinal plant leaves using image feature and ANN" published in 2013 international conference on advanced electronic system (ICAES K. Elissa, "Title of paper if known," unpublished.
4. Vijayashree.T, Dr. A. Gopal "Leaf Identification for the Extraction of Medicinal Qualities Using Image Processing Algorithm" published in 2017 International Conference on Intelligent Computing and Control (I2C2)
5. D Venkataraman, Mangayarkarasi N "Support Vector Machine Based Classification of Medicinal Plants Using Leaf Features" published in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)
6. Shitala Prasad, Pankaj P Singh"Medicinal Plant Leaf Information Extraction Using Deep Features" published in Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

7. Dileep M.R., Pournami P.N."AyurLeaf: A Deep Learning Approach for Classification of Medicinal Plants" published in TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)
8. C.Amuthalingeswaran, Mr. M.Sivakumar, Dr. P.Renuga"IDENTIFICATION OF MEDICINAL PLANT'S AND THEIR USAGE BY USING DEEP LEARNING" Published in Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART;ISBN: 978-1-5386-9439-8
9. Manojkumar P., Surya C. M., Varun P. Gopi "Identification of Ayurvedic Medicinal Plants by Image Processing of Leaf Samples" Published in 2017 Third international conference on research in computational Intelligence and communication network (ICRCICN)
10. Sujit A., Aji. S "An Optimal Feature set With LPB for Leaf Image classification" Published in Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020) IEEE Xplore Part Number:CFP20K25-ART; ISBN:978-1-7281- 4889-2
11. Amala Sabu, Sreekumar K, Rahul R Nair "Recognition of Ayurvedic Medicinal Plants from Leaves: A Computer Vision Approach" Published in 2017 Fourth International Conference on Image Information Processing (ICIIP)
12. S.Perumal1 and T.Velmurugan2 " Preprocessing by Contrast Enhancement Techniques for Medical Images" published in International Journal of Pure and Applied Mathematics Volume 118 No. 18 2018, 3681-3688
13. Nourhan Zayed and Heba A. Elnemr " Statistical Analysis of Haralick Texture Features to Discriminate Lung Abnormalities" Published in International Journal of Biomedical Imaging / 2015 Article ID 267807