

REFINING THE DETECTION OF MEDICINAL PLANTS USING MACHINE LEARNING AND IMAGE ANALYSIS TECHNIQUES

Mr. A. Ramesh¹, Kokku Venkata Ganesh², Sangem Naresh³,
Potharaju Ajay⁴, Dyavari Paramesh⁵

¹Associate Professor, Computer Science and Engineering, ACE Engineering College, India.

^{2,3,4,5}Student, Computer Science and Engineering, ACE Engineering College, India.

ABSTRACT

India boasts a rich heritage of floral diversity, a vital source of medicinal plants integral to Ayurvedic Pharmaceuticals. However, the misidentification of these plants remains a critical issue, with numerous crude drugs sold under the same name in the market. This confusion stems from factors such as seasonal and geographical variations and similar morphological characteristics. Consequently, the demand for these resources has led to issues of impurity, substitution, and doubtfulness regarding their curative potential. In response, this project presents an innovative solution. It employs Image Processing and various Machine Learning Algorithms to develop software capable of identifying different medicinal plants and raw materials and listing out their uses. By doing so, it offers a transformative tool for stakeholders throughout the supply chain, from wholesalers to distributors, ensuring the authenticity and quality of Ayurvedic ingredients. This technology not only safeguards India's botanical heritage but also reinforces trust in the efficacy of traditional herbal medicine systems. The medicinal leaf dataset consists of 30 classes. Transfer learning approach was used to initialize the parameters and pretrain Neural networks namely MobileNetV2, InceptionV3, and ResNet50. These component models were used to extract features from the input images and the SoftMax layer connected to the Dense Layer was used as the classifier to train the models on the concerned dataset.

1. INTRODUCTION

Human beings rely heavily on plants for their survival. Other than providing us with food and oxygen, few of them also possess medicinal qualities which heal many diseases. Presence of antibacterial activity in 12 Indian Medicinal Plants provide an inspiration to develop drug compounds for human health which can be put into therapeutic uses like controlling diarrhea, nervous disorders, leprosy, and leukoderma among others. Medicinal plants have been an active topic of research for a long time now as the medicines made out of them are available at cheaper costs and have no possible side effects. Few plants also possess anti-colorectal cancer properties which tend to provide better outcome than the current chemotherapy treatments. Medicines obtained from plants also proved to be effective when battling with COVID-19 pandemic and with disease like obesity. Research on-going in the field of deriving medicines from medicinal plants, throws light on another important aspect which is the automatic detection of medicinal plants, that can make the common man easily identify plants growing in their vicinity and make utmost use of them for their health benefits. Though, plants can be identified using any plant part which includes stem, fruit, bark the leaf has been widely used due to its availability in all seasons. The feature vector can be constructed by exclusively considering shape, texture, venation, or using the combination of one or more Abstract The therapeutic nature of medicinal plants and their ability to heal many diseases raises the need for their automatic identification. Different parts of plants that help in their identification include root, fruit, bark, stem but leaf images have been widely used as they are an abundant source of information and are also easily available. This work explores the branch of Artificial Intelligence, called deep learning, and proposes an Ensemble learning approach to rapidly detect medicinal plants using the leaf image. The medicinal leaf dataset consists of 30 classes, Transfer learning approach was used to initialize the parameters and pretrain Neural networks namely MobileNetV2, InceptionV3, and ResNet50. These component models were used to extract features from the input images and the SoftMax layer connected to the Dense Layer was used as the classifier to train the models on the concerned dataset. The obtained accuracies were validated using threefold and fivefold cross validation The Ensemble Deep Learning- Automatic Medicinal Leaf Identification (EDL-AMLI) classifier based on the weighted average of the component model outputs was used as the final classifier. It was observed that the EDL-AMLI outperformed the state-of-the-art pre-trained models such as MobileNetV2, InceptionV3, and ResNet50 by achieving 99.66% accuracy on the test set and average accuracy of 99.9% using threefold and fivefold cross validation The Ensemble Deep Learning- Automatic Medicinal Leaf Identification (EDL-AMLI) classifier based on the weighted average of the component model outputs was used as the final classifier. It was observed that the EDL-AMLI outperformed the state-of-the-art pre-trained models such as MobileNetV2, InceptionV3, and ResNet50 by achieving 99.66% accuracy on the test set and average accuracy of 99.9% using threefold and fivefold cross validation Once the

leaf image features are extracted and the relevant feature vector is obtained, it is input into the classifier to classify the image into its species. Different feature extraction and classification techniques used in recent years is discussed. Another aspect of Artificial Intelligence, called deep learning imitates the human brain to recognize patterns in images and to distinguish 1 class of images from another. In this paper, we have explored transfer learning and ensemble learning to distinguish medicinal leaf images into 30 classes. The transfer learning approach makes use of existing knowledge to solve problems in other fields. It has been widely used in image recognition, where the Convolutional Neural Network models pre-trained on high-end GPUs can classify objects into 1000 classes. This knowledge attained by these networks can be transferred for medicinal leaf image recognition. We have trained 3 models namely MobileNetV2, Inception V3, and ResNet50 on the medicinal leaf dataset and evaluated their efficiency, and then finally used ensemble learning which is based upon using the weighted averages of the component models.

2. LITERATURE SURVEY

As part of the Literature Survey, we have referred few project papers and the findings from them are:

Real-Time Identification of Medicinal Plants using Machine Learning Techniques: Sivaranjani.C, Lekshmi Kalinathan, Amutha.R [1]

The lighting condition of the environment is uncontrolled, so the segmentation of a leaf from the background is considered as a complex task. Here we propose a system which can identify the plant species based on the input leaf sample. An improved vegetation index, ExG-ExR is used to obtain more vegetative information from the images. The reason here is, it fixes a built-in zero threshold and hence there is no need to use or any threshold value selected by the user. In spite of the existence of more vegetative information in ExG with otsu method, our ExG-ExR index works well irrespective of the lighting background. Therefore, the ExG-ExR index identifies a binary plant region of interest. The original color pixel of the binary image serves as the mask which isolates leaves as sub-images. The plant species are classified by the colour and texture features on each extracted leaf using Logistic Regression classifier with the accuracy of 93.3%. In this work, we addressed the problem of identifying the medicinal plant species by the analysis of leaf images obtained directly from their habitat and irrespective of lighting conditions. The fixed zero threshold, ExG-ExR vegetative index is successfully tested for image dataset. The result shows that the algorithm can adequately segment the leaf region. This method worked well in images with reflection. The feature extraction based on the colour and texture features is done. Identification of Medicinal Plants using Deep Learning: R. Upendar Rao, M. Sai Lahari, K. Pavana Sri and team. [2]

Identification of the correct medicinal plants that goes in to the preparation of a medicine is very important in ayurvedic, folk and herbal medicinal industry. The main features required to identify a medicinal plant is its leaf shape, color and texture. Color and texture from both sides of the leaf contain deterministic parameters to identify the species. In this project we explore feature vectors from both the front and back side of a green leaf along with morphological features to arrive at a unique optimum combination of features that maximizes the identification rate. A database of medicinal plant leaves are created from scanned images of front and back side of leaves of commonly used medicinal plants. The leaves are classified based on the shape and dimension combination. It is expected that for the automatic identification of medicinal plants this system will help the community people to develop their knowledge on medicinal plants, help taxonomists to develop more efficient species identification techniques and also participate significantly in the pharmaceutical drug manufacturing. Detection and Quantification of Root-Knot Nematode (*Meloidogyne* Spp.) Eggs from Tomato Plants Using Image Analysis, Top Bahadur Pun, Arjun Neupane, Richard Koech and Kirsty J. Owen 2022. [3]

Root-knot nematodes (*Meloidogyne* spp.; RKN) are major plant-parasitic nematodes that cause significant loss to agricultural production. An accurate assessment of the RKN population density at a field level is crucial for decisions about the application of control measures to minimize yield losses. Traditionally, RKN populations are identified and counted by nematologists using a microscope. This method is a specialized, time-consuming process and prone to errors. In our study, we investigated three semi-automated methods to detect and count RKN eggs using image analysis: contour arc (CA), skeleton structure (SS), and extreme point (EP). These methods were used to automate the length measurement of RKN eggs, and the results were compared with traditional methods of quantification. The EP method produced the highest correlation with the manual length measurement of RKN eggs. Further, these methods were used to detect and count RKN eggs to quantify low and highly cluttered images. We estimated the optimal range of the ratio of each method to detect and count RKN eggs. Textural Analysis for Medicinal Plants Identification Using Log Gabor Filters, Frimpong Twum, Yaw Marfo Missah, Stephen Opoku Oppong and Najim Ussiph 2022. [4]

Texture plays a crucial role in computer vision, providing valuable information about image regions. Log Gabor filters that mimic the human eyes visual cortex are used as feature extractors to identify medicinal plants based on the leaf

textural features. This method was tested on a dataset developed from the Centre of Plant Medicine Research, Ghana, consisting of forty-nine (49) plant species as well as the Flavia and Swedish Leaf datasets, which are benchmark datasets. The Log Gabor filter outperformed the Gabor filters, which have been extensively used in this area when tested on nine supervised classifiers (K-Nearest Neighbor, Support Vector Machine, Naive Bayes, Logistic Regression, and Decision tree, Random Forest, Multilayer Perceptron, Gradient Boosting and Stochastic Gradient Descent) with 10-fold cross validation. The Support Vector Machine and Multilayer Perceptron were the best performing classifiers for both Log Gabor filter and Gabor filter in terms of accuracy, precision, true positive rate, F1 score and false positive rate. The Log Gabor filters highest accuracy was 79% for My dataset, 97% for Flavia, and 98% for the Swedish Leaf dataset while the Gabor filters highest accuracy was 66% for Dataset, 92% for Flavia and 96% for the Swedish Leaf dataset. Gabor filters which have been used extensively in literature for texture extraction were compared with the Log Gabor filter proposed in this study. Automatic Plant Counting and Location Based on a Few-Shot Learning Technique Azam Karami, Melba Crawford and Edward J. Delp. [5]

Plant counting and location are essential for both plant breeding experiments and production agriculture. Stand count indicates the overall emergence of plants compared to the number of seeds that were planted, while location provides information on the associated variability within a plot or geographic area of a field. Deep learning has been successfully applied in various application domains, including plant phenotyping. This article proposes the use of deep learning techniques, more specifically, anchor-free detectors, to identify and count maize plants in RGB images acquired from unmanned aerial vehicles. The results were obtained using modified CenterNet architecture, with validation performed against manual human note. Experimental results demonstrated an overall precision $>95\%$ for examples where training and testing were performed on the same field. Few-shot learning was also explored, where the trained network was 1) directly applied to the fields in other geographic areas and 2) updated using small quantities of training data from the other locations. Instance Segmentation and Classification Method for Plant Leaf Images Based on ISC-MRCNN and APS-DCCNN: Xiaobo Yang, Aibin Chen, Guoxiong Zhou, Jianwu Wang, Wenjie Chen, Yuan Gao and Rundong Jiang. [6]

This paper proposes an instance segmentation and classification method for plant leaf images based on ISC-MRCNN and APS-DCCNN. Firstly, an ISC-MRCNN based plant leaf image segmentation method is proposed, which is used to preprocess the plant leaf images to remove the background interference. In the ISC-MRCNN method, the following three problems have been solved. Deep ensemble learning for automatic medicinal leaf identification: Silky Sachar, Anuj Kumar. [7]

Automatic detection of medicinal plants opens new doors for the development of medicines to cure diseases that have not yet been cured by allopathy. It will allow the layman to be aware of the plants growing in their surroundings and make utmost use of them to cure common ailments with no possible side effects. Artificial Intelligence makes this purpose even more achievable. We proposed an Ensemble of deep learning models to automatically detect medicinal plants. The medicinal leaf images were obtained from a medicinal leaf dataset published in Mendeley. By employing Transfer learning. DeepHerb: A Vision Based System for Medicinal Plants Using Exception Features, S. Roopashree and J. Anitha 2021. [8]

A vision-based automatic medicinal plant identification system is proposed using different neural network techniques in computer vision and deep learning. The challenge lies in the unavailability of the medicinal herb dataset. The paper showcases a novel medicinal leaf dataset entitled DeepHerb dataset comprising of 2515 leaf images from 40 varied species of Indian herbs. The efficacy of the dataset is revealed by comparing pre-trained deep convolution neural network architectures such as VGG16, VGG19, InceptionV3 and Xception. The work concentrates on adopting the transfer learning technique on the pre-trained models to extract features and classify using Artificial Neural Network (ANN) and Support Vector Machine (SVM). The SVM hyper parameters are tuned further by Bayesian optimization to achieve a better performance model. The proposed DeepHerb model learned from Xception and ANN outperformed by 97.5% accuracy. A Comparative Analysis on Machine Learning Models for Accurate Identification of Medical Plants Prabhat Kumar Thella1, V. Ulagamuthalvi2. [9]

This research examines methods used in identifying and classifying medicinal plants as well as the medicinal properties of plants that have become increasingly relevant in the recent past. There is a vital importance placed on identifying the suitable medicinal plants in the creation of an ayurvedic medication. In order to identify a medicinal plant, look for these three features: leaf form, colour, and texture. From the both sides of the leaf, there are both deterministic and nondeterministic factors that identify the species. In leaf identification, rates as high as 99% have been found when tested on a wide range of classifiers. Extending the prior work by using dried leaves and feature vectors results in identification using which identification rates of 94% are possible. This paper explores feature vectors from both the

front and back side of a green leaf along with morphological features to arrive at a unique optimum combination of features that maximize the identification rate. A database of medicinal plant leaves is created from scanned images of front and back side of leaves of commonly used ayurvedic medicinal plants. The leaves are classified based on the unique feature combination. Identification rates up to 99% have been obtained when tested over a wide spectrum of classifiers.

3. COMPARISION ANALYSIS

S. No	Title	Methodology	Outcome	Drawback
1	Real-Time Identification of Medicinal Plants using Machine Learning Techniques	ExG-ExR, Logistic Regression classifier.	An improved vegetation index, ExG-ExR is used to obtain more vegetative information from the images.	1.Accuracy is based on the light on the plant or leaf 2.Difficulties to obtain better performance
2	Detection and Quantification of Root-Knot Nematode (Meloidogyne Spp.) Eggs from Tomato Plants Using Image Analysis	Contour arc (CA), Skeleton structure (SS), and Extreme point (EP) methods	RKN populations are identified and counted by nematologists using a microscope.	1.Narrowly specialized knowledge 2.Additional configuration is required 3.It is not an easy-to-use method
3	Textural Analysis for Medicinal Plants Identification Using Log Gabor Filters	K-Nearest Neighbor, Support Vector Machine, Nave Bayes, Logistic Regression	Human eyes visual cortex are used as feature extractors to identify medicinal plants based on the leaf textural features	1.Difficulties to obtain better performance 2.Poor Application Performance
4	DeepHerb: A Vision Based System for Medicinal Plants Using Xception Features,	Artificial Neural Network (ANN) and Support Vector Machine (SVM)	Medicinal leaf dataset entitled DeepHerb dataset comprising of 2515 leaf images from 40 varied species of Indian herbs	1.Unsuitable for large scale scenarios. 2.Generally have high polynomial running times. 3.May take huge time and cost to construct
5	Automatic Plant Counting and Location Based on a Few-Shot Learning Technique	CenterNet architecture	This article proposes the use of deep learning techniques, more specifically, anchor-free detectors	1.Resulting data errors. 2.Have not been investigated thoroughly 3. Difficulties to obtain better performance
6	Instance Segmentation and Classification Method for Plant Leaf Images Based on ISC-MRCNN and APS-DCN	ISC-MRCNN and APS-DCCNN	Preprocess the plant leaf images to remove the background interference	1.Narrowly specialized knowledge 2.Significantly increases capital and operating expenditures 3.They are hard to maintain
7	Identification of Medicinal Plants using Deep Learning	CNN Layer, DenseNet	We explore feature vectors from both the front and back side of a green leaf along with morphological features	1.Geographical features were not included 2. May take huge time and economic cost to construct
8	Deep ensemble	Ensemble Deep	The plants growing in their	1. Less accuracy

	learning for automatic medicinal leaf identification	Learning	surroundings and make utmost use of them to cure common ailments with no possible side -effects	2. Geographical features were not included
9	A Comparative Analysis on Machine Learning Models for Accurate Identification of Medical Plants	computer vision and machine learning	The leaves were selected due to the fact that hypertension is considered a significant medical condition.	1.Poor Application Performance 2.Narrowly specialized knowledge

4. FUTURE SCOPE

Future directions for the project include enhancing the accuracy of identification algorithms through continuous refinement and incorporation of diverse datasets. Expansion of the database to include newly discovered species and additional information about existing plants will be essential for keeping the model relevant and reliable.

Developing a user-friendly mobile application could democratize access to plant identification, allowing users to identify medicinal plants conveniently through their smartphones. Integration of geolocation data could provide users with information about plant prevalence and harvesting seasons in their region, aiding in sustainable foraging practices.

Collaboration with botanists, herbalists, and other experts will ensure the accuracy of plant identification and provide valuable insights into the medicinal properties and uses of identified plants. Furthermore, partnerships with research institutions and organizations can lead to the development of new features based on the latest research findings.

5. CONCLUSION

In conclusion, this project leverages image processing and machine learning to classify and detect unique medicinal plants with remarkable accuracy. This tool can be utilized in the future to not only enhance our understanding of rare and diverse flora but also to aid in the early identification and treatment of rare diseases. By continuously refining our models and expanding our databases, we ensure the system remains current and effective. The forthcoming development of a user-friendly mobile application and our ongoing collaborations with botanical experts will further democratize access to this valuable resource, making a significant difference in both educational and clinical settings globally.

6. REFERENCES

- [1] Parekh J, Chanda S (2010) Antibacterial and phytochemical studies on twelve species of Indian medicinal plants. *Afr J Biomed Res* 10:175–181.
- [2] Benarba B, Pandiella A (2018) Colorectal cancer and medicinal plants: principle findings from recent studies. *Biomed Pharmacother* 107:408–423. <https://doi.org/10.1016/j.biopha.2018.08.006>
- [3] Adhikari B, Marasini BP, Rayamajhee B et al (2021) Potential roles of medicinal plants for the treatment of viral diseases focusing on COVID-19: a review. *Phyther Res* 35:1298–1312. <https://doi.org/10.1002/ptr.6893>
- [4] de Freitas Junior LM, de Almeida EB (2017) Medicinal plants for the treatment of obesity: ethnopharmacological approach and chemical and biological studies. *Am J Transl Res* 9:2050–2064
- [5] Wäldchen J, Mäder P (2018) Plant species identification using computer vision techniques: a systematic literature review. *Arch Comput Methods Eng* 25:507–543. <https://doi.org/10.1007/s11831-016-9206-z>
- [6] Sachar S, Kumar A (2020) Survey of feature extraction and classification techniques to identify plant through leaves. *Expert Syst Appl*. <https://doi.org/10.1016/j.eswa.2020.114181>
- [7] Pan SJ, Yang Q (2010) A survey on transfer learning. *IEEE Trans Knowl Data Eng* 22:1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- [8] Lee SH, Chan CS, Mayo SJ, Remagnino P (2017) How deep learning extracts and learns leaf features for plant classification. *Pattern Recognit* 71:1–13. <https://doi.org/10.1016/j.patcog.2017.05.015>
- [9] Grinblat GL, Uzal LC, Larese MG, Granitto PM (2016) Deep learning for plant identification using vein morphological patterns. *Compute Electron Agric* 127:418–424. <https://doi.org/10.1016/j.compag.2016.07.003>
- [10] Gogul I, Kumar VS (2017) Flower species recognition system using convolution neural networks and transfer learning. 2017 4th Int Conf Signal Process Commun Networking. *ICSCN* 2017:1–6. <https://doi.org/10.1109/ICSCN.2017.8085675>
- [11] Liu Z, Zhu L, Zhang XP et al (2015) Hybrid deep learning for plant leaves classification. *Lect Notes Compute Sci (Include Subser Lect Notes ArtifIntell Lect Notes Bioinform)* 9226:115–123. https://doi.org/10.1007/978-3-319-22186-1_11