Grenze International Journal of Engineering and Technology, June Issue

MP Detector: Intelligent Medicinal Plant Recognition using Machine Learning and Deep Learning

Tisha Murarka[1], Sanvi Bhelkar[2], Mayur Soni[3], Anurag Sinha[4], Bhagyashree Hambarde[5], Priya Parkhi[6] and Pournima Pande[7]
[1-4,6]Computer Science and Engineering, Ramdeobaba University, Nagpur, Maharashtra, India
Email: tishamurarka18@gmail.com, sanvibhelkar@gmail.com, mayursoniwork1@gmail.com, anuragksinha05102001@gmail.com, pariyaparkhi1@gmail.com
[5]Computer Science & Engineering (Data Science), St. Vicent Palloti College of Engineering, Nagpur, Maharashtra, India
Email: bhagyashreehambarde19@gmail.com
[7]Yeshwantrao Chavan College of Engineering, Wanadongari, Nagpur
Email: pournimapande@yahoo.co.in

***Abstract*—a groundbreaking approach aimed at revolutionizing plant identification, particularly in Ayurveda. Medicinal plants are fundamental resources in Ayurveda, emphasizing the need for accurate identification in medicine preparation. However, manual identification is labor-intensive and crucial for ensuring medicinal effectiveness and safety. Current challenges include adulteration and a lack of trust in the supply chain. This study implements image processing through machine learning technologies for automated, real-time plant recognition. By leveraging machine learning algorithms, this innovative solution enhances identification processes, ensuring the authenticity and safety of Ayurvedic treatments.**

***Keywords*— Ayurveda, Deep Learning, Machine Learning, Convolutional Neural Network, Logistic Regression, K-Nearest Neighbor, Decision Tree, Random Forest, Support Vector Machine.**

I. INTRODUCTION

Since ancient times, plants have been essential in Ayurvedic medicine. Correct identification is a critical step in preparing Ayurvedic remedies, which has traditionally been done manually [1]. However, many collectors and traders struggle with accurate identification, leading to substitution and adulteration, which compromises trust in Ayurvedic treatments. This study proposes a machine learning-based approach for real-time automated plant identification [2]. A trained model will be integrated into an application that enables real-time image analysis [15], ensuring authenticity and reducing adulteration.

II. LITERATURE REVIEW

Machine learning has recently gained prominence in medicinal plant image processing [5]. Various studies have explored methods to identify and classify medicinal plants based on visual features.

A. Existing Plant Identification Technologies

Pre-processing, segmentation, feature extraction, and classification methods are commonly employed.

Grenze ID: 01.GIJET.11.2.178_20

Studies have demonstrated the effectiveness of different models:

- A k-nearest neighbor (KNN) classifier achieved 93% accuracy in classifying 20 plant species.
- A probabilistic neural network (PNN) model reached 90.3% accuracy.
- Support Vector Machine (SVM) models achieved 90% accuracy [17].
- The Fast Region Convolutional Neural Network (FRCNN) demonstrated 92.5% accuracy in classifying leaves and flowers.
- A CNN-based system identified rare medicinal plants with 95%-99% accuracy on a test set of 500 images [23].

III. Methodology & Implementation

A. Data Collection

Data was sourced from the National Library of Government, which contains plant images from Karnataka and Kerala. The dataset consists of 5900 images of 40 plants and 6900 images of 80 plant leaves, captured in different seasons and varying backgrounds [6].

B. Image processing

Image preprocessing refers to a set of techniques used to prepare and enhance digital images before they undergo further analysis or processing [8]. It involves a series of operations aimed at improving the quality, usability, and interpretability of the images. Some common steps involved in image preprocessing are Resizing and scaling, normalization, noise reduction, background removal, cropping, rotating, etc.

Figure: 1: Dataset images [6]

Image Augmentation

Image augmentation is a technique used in machine learning, particularly in computer vision tasks like image classification or object detection. It involves applying a variety of transformations to existing images to create modified versions. Like rotating, flipping, zooming and scaling, cropping, and padding.

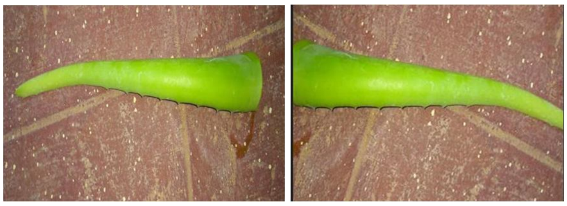

Figure: 2: Augmented images

Background Removing

Background removal is a technique used to eliminate or isolate the background from the main subject in an image, which can aid in reducing noise or distractions.

While it might not directly address all forms of noise, removing the background can help simplify an image, making it easier for subsequent processing steps to focus on the main subject.

Figure: 3 Masked Images

C. Training and Testing data

In machine learning, especially in tasks like image classification for medicinal plant identification, splitting the dataset into training and testing sets is crucial for model development and evaluation. Divide the dataset into 70-30% for further model evaluations.

D. Model Selection

To classify the plant Supervised Classifier of Machine Learning and CNN of deep learning were used[18]. Each model is explained below:

- Logistic Regression: A simple yet effective statistical model used for binary classification. It predicts the probability of a plant being medicinal based on input features and uses a sigmoid function to map predictions between 0 and 1. Simple, effective for binary classification problems.
- K-Nearest Neighbor (KNN): A non-parametric algorithm that classifies plants by comparing them to their k nearest neighbors in the feature space. It is highly effective in cases with clear distinctions between categories. Effective for distinguishing categories.
- Decision Tree: A hierarchical model that makes decisions by splitting data at different nodes based on feature importance. It is easy to interpret and works well with categorical and numerical data.
- Simple, interpretable, works well for both categorical and numerical data.
- Random Forest Classifier: An ensemble learning method that builds multiple decision trees and averages their predictions to improve accuracy and reduce overfitting. It provides high performance and robustness. Reduces overfitting, better performance.
- Support Vector Machine (SVM): A powerful classification algorithm that finds the optimal hyperplane to separate different classes. It is particularly effective in high-dimensional spaces. Effective for high-dimensional datasets.
- Naïve Bayes: A probabilistic classifier based on Bayes' theorem that assumes feature independence. Despite its simplicity, it works well for text classification and certain structured data. Fast, handles structured data despite assumptions.
- Convolutional Neural Network (CNN): A deep learning model that automatically extracts features from images. It consists of convolutional, pooling, and fully connected layers that process image data efficiently. Powerful for automatic image feature extraction.

IV. Results

To train the tree, examples of plants with known outcomes are presented, and the tree learns by identifying the best questions to ask at each node. The tree splits criteria by selecting those that effectively separate plants into different categories. Multiple Decision Trees can be combined into an ensemble to enhance performance and mitigate overfitting, creating a Random Forest. The accuracy came as 80.39%.

A. Random Forest Classifier

Each tree in the forest makes a random decision[9] based on a subset of plant characteristics. The algorithm gains strength from the diversity of trees, acknowledging that individual trees may not be perfect on their own. The algorithm identifies the most important features of plants, contributing to understanding what makes a plant medicinal. For medicinal plant identification, the Random Forest learns from a dataset containing known medicinal and non-medicinal plants, predicting the medicinal status of a given plant based on this learning. The accuracy of the model from this random forest classifier came to be 88.23 %.

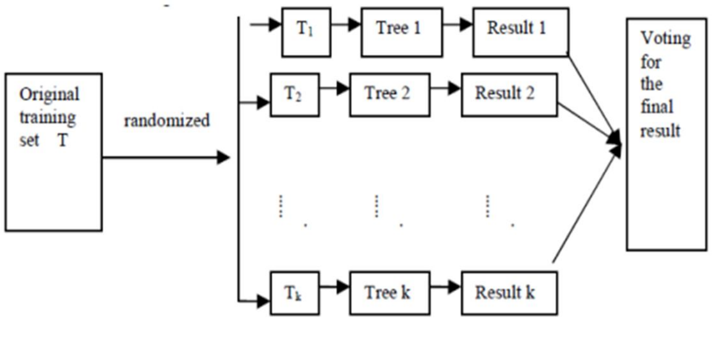

Figure: 4 Random Forest [9]

B. Support Vector Machine

Support Vector Machine (SVM)[10] is a supervised machine learning algorithm used for both classification and regression tasks. SVM is particularly effective in high-dimensional spaces and is well-suited for tasks where clear decision boundaries are needed[16]. The main idea behind SVM is to find the hyperplane that best separates the data into different classes. In precis, SVM is a flexible and powerful algorithm for class tasks, especially when coping with high-dimensional records or when a clear separation among training is desired. Accuracy came as 81.04%.

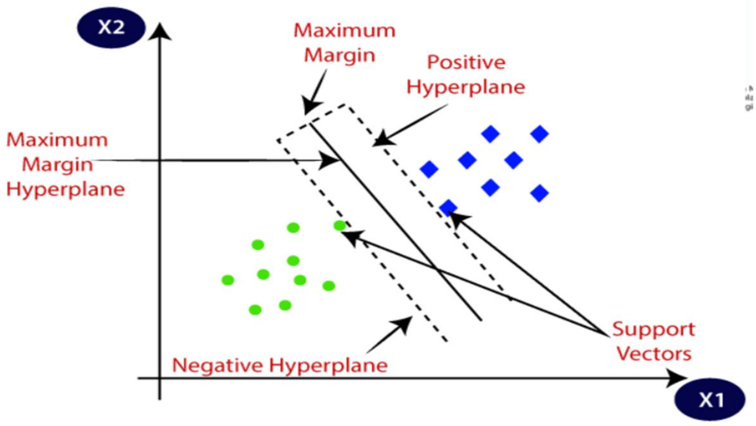

Figure: 5 Support Vector Machine

C. Naïve Bayes

Bayes' theorem is the foundation of the Naive Bayes classifier. It mathematically expresses the relationship between the probability of a hypothesis (or class) given evidence (features) and the probability of that evidence given the hypothesis

P(Class|Features) = P(Features)P(Class) x
P(Features|Class) (1)

Clarify how data for trend evaluation was collected and the criteria used for selecting trends. Ans: mentioned features we are considering

The "naive" assumption in Naive Bayes is that the features used for classification are conditionally independent given the class label. It allows us to decompose the likelihood term into the product of individual feature likelihoods:

P (Features Class) =P (Feature1|Class) × P (Feature2|Class) ×…× P (Feature2|Class) (2)

Conditional Probability Estimation: For each feature and each class, calculate the conditional probability of observing that feature given the class. Given a set of features for a new instance, the Naive Bayes classifier predicts the most likely class using Bayes' theorem.

D. CNN Layers

Convolutional Neural Networks (CNNs), a subset of deep learning models, have demonstrated high effectiveness in tasks like image recognition, segmentation, and object detection[7]. Automatic Feature Extraction: In contrast to traditional methods relying on handcrafted features, deep learning allows models to automatically learn hierarchical representations, capturing intricate patterns and relationships within the data[20].
Transfer Learning's Benefits: The technique of transfer learning within deep learning enables the reuse of pre-trained models on large datasets, proving highly advantageous for image processing, especially with limited labelled data.
Continued Technological Advancements[14]: As technology advances, the integration of deep learning continues to shape the landscape of image processing, offering innovative solutions and pushing the boundaries of what is achievable in the field[19]. By performing deep learning the accuracy is 80.21 %.

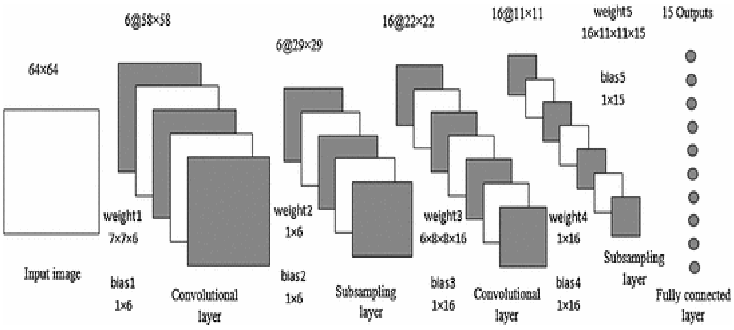

Figure: 6 CNN layers

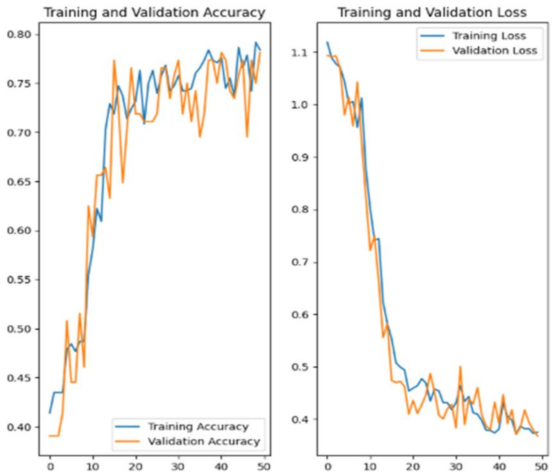

Figure: 7 Deep learning accuracy

Accuracy-Calculated as the sum of true positives and true negatives divided by the total number of samples[21].

$$Accuracy=(NumberofCorrectPredictions)/(TotalNumberofPredictions) \quad (3)$$

Precision-Calculated as the number of true positives divided by the sum of true positives and false positives.
Precision = True positives/ (True positives + False positives) (4)
Recall-Calculated as the number of true positives divided by the sum of true positives and false negatives [22].
Recall = True Positives / (True Positives + False Negatives) (5)
F1 Score = 2 * (Precision * Recall) / (Precision + Recall (6)

V. RESULT

In closing, strong classification abilities and thorough process create meaningful change in plant recognition. Its precision delivers results dependably alongside flexibility for many uses involving botanical matters. The model excellently recognized different types of plants, reaching a great accuracy level across various plant kinds.
Points like entropy, uniformity, contrast, and activity were skilfully resolved, permitting correct identification. It has proven its ability to efficiently manage a substantial database, emphasizing its capacity for growth and increased usage. Individuals can depend on the tool for swift and trusted plant identification, establishing it as a helpful resource for botanists, scientists [12], and interested people.

TABLE I: RESULT

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistics Regression	79.08	87.75	62.31	72.88
k -nearest neighbor	82.35	90.38	68.11	77.68
Decision Tree	80.39	85.45	68.11	
SVM	81.04	81.63	66.63	73.39
Random Forest	88.23	92	76.66	83.63
Naive Bayes	64.0	56.81	41.66	48.07

The Table I shows the comparative results for Logistic Regression, K-Nearest Neighbor, Decision Tree SVM, Random Forest and Naïve Bayes Algorithms. The Random Forest algorithm shows most promising results. It shows 88% Accuracy, 92% precision, 76% Recall and 83% F1-Score.

VI. CONCLUSION

The study confirms that the Random Forest algorithm is the most effective for medicinal plant identification. The integration of machine learning and deep learning ensures a reliable, automated system that enhances the credibility of Ayurvedic medicine.

FUTURE WORK

Future research will focus on expanding the dataset with more plant species, developing a mobile application for real-time classification, and incorporating explainable AI techniques to enhance model interpretability.

REFERENCES

[1] Asmtia Manna, Rishi Shinde, Vishal Supekar, Mihir Salvi, Kunal Yeole, "Identification of Medicinal Herbs using Modified Xception Architecture", 2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA), pp.1-5, 2024.

[2] Bhagyashree Hambarde, Priya Parkhi, Prashant Govardhan, Amruta Tapas, Poorva Agrawal, Gagandeep Kaur, Pournima Pande; Detection of eye defect with the aid of answerable AI. AIP Conf. Proc. 10 December 2024; 3188 (1): 080005. https://doi.org/10.1063/5.0242547

[3] Christine Kikome, Geoffrey Okumu, Emmanuel Wagisha, Daudi Jjingo, John Kizito, Ggaliwango Marvin, "Transparent Machine Vision Techniques For Medicinal Plant Species Identification", 2024 5th International Conference on Image Processing and Capsule Networks (ICIPCN), pp.135-142, 2024.

[4] Revathy Prasannan, Preethy Roy, Jithi P V, Amritha R, Anjana V Nair, Ann Mariya Jose, "Medicinal Plants Identification Using Deep Learning: A Survey", 2024 IEEE Recent Advances in Intelligent Computational Systems (RAICS), pp.1-5, 2024.

[5] Nora Naik, Snehal Bhogan, "Machine Learning Approach for Classification of Herbal Plant Leaves", 2024 1st International Conference on Trends in Engineering Systems and Technologies (ICTEST), pp.1-5, 2024.

[6] Rasendram Muralitharan, Nilani Inuvaisubramaniam, Upul Jayasinghe, Roshan G Ragel, Aarumuham Rajeevan, Gramma Niladhari, "Plant Leaves Classification System", 2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), pp.1-6, 2023.

[7] N. Shobha Rani, Akshatha Prabhu, Abhiram M P, Adarsh V, "Bacterial Wilt Detection from Okra Leaf Using Mask-RCNN", 2023 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), pp.153-158, 2023.

[8] Sudharshan Duth P, Vedavathi S, Roshan S, "Herbal Leaf Classification using RCNN, Fast RCNN, Faster RCNN", 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), pp.1-8, 2023.

[9] Veena Tapale, Rutuja Rekhawar, Arpita Savagonmath, Sainath Habre, Savita Adhav, "Phyto Leafy Recognition Using Deep Learning", 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), pp.1-5, 2023.

[10] Swati G. Kale, Shoaib Ansari, Tawfeek Khan, Mansi Chitriv, Harshit Sathone, Hariyali Mansata, P.G. Jaiswal, N. R. Wankahde, Roshan Umate, "Identification of Ayurvedic Leaves using Deep Learning", 2023 International Conference on Communication, Circuits, and Systems (IC3S), pp.1-6, 2023.

[11] Savita Adhav, Rutuja Rekhawar, Veena Tapale, Sainath Habre, Arpita Savagonmath, "Survey on Healing Herbs Detection using Machine Learning", 2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON), pp.1-5, 2023.

[12] Achanta Lakshmi Kalyani, Rajesh Singamaneni, Greeshmanth Penugonda, "Remembrance of Monocotyledons Using Residual Networks", 2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), pp.1-6, 2022.

[13] R, P. B., & Rani, N. S. (2023). DIMPSAR: Dataset for Indian medicinal plant species analysis and recognition. Data in Brief, 49, 109388.

[14] Kovvuri Uday Surya Deveswar Reddy, Ayesha Shaik, A Balasundaram, Mattapally Sai Nithin, Lakshmi Sai Ram Kakarla, Sk Noor Mahammad, "Classification of Indian Medicinal Leaves using Transfer Learning based Convolutional Neural Networks", 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC), pp.1050-1058, 2022.

[15] Vina Ayumi, Ermatita Ermatita, Abdiansah Abdiansah, Handrie Noprisson, Mariana Purba, Marissa Utami, "A Study on Medicinal Plant Leaf Recognition Using Artificial Intelligence", 2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS, pp.40-45, 2021.

[16] Savitha Patil, M. Sasikala, "An Automated System for Identification of the Medicinal Leaf using MKSVM", 2020 IEEE International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET), pp.1-6, 2020.
[17] Soumya S Kiranagi, Savitha Patil, "Medicinal Plant Classification Using SVM", 2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON), pp.1-5, 2024.
[18] Vaishnavi N, Sanjana M Jannu, Sharanya M D, Harini S, "Comparative Study of Different ML And DL Algorithms For Indian Medicinal Leaf Detection", 2023 International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems (ICAECIS), pp.616-621, 2023.
[19] Barinderjit Singh, Rashmi S. Deshpande, Mohd. Shaikhul Ashraf, Shaik Vaseem Akram, "Leaf Disease Detection Using Deep Neural Network", 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), pp.1-6, 2022.
[20] Vinay Kukreja, Rishabh Sharma, Amanpreet Kaur, Ravi Kumar Sachdeva, Vikas Solanki, "Deep Neural Network for Multi-Classification of Parsley Leaf Spot Disease Detection", 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp.1378-1382, 2022.
[21] Gracelin S, Kumudha Raimond, "Deep Learning based Indigenous Herbal Medicinal Plants Recognition: A Comprehensive Review", 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), pp.1412-1418, 2022.
[22] Kale, Swati G., et al. "Identification of Ayurvedic Leaves using Deep Learning." 2023 International Conference on Communication, Circuits, and Systems (IC3S). IEEE, 2023.
[23] Kalaiselvi P, Esther C, Aburoobha A, Nishanth J, Gopika S, Kabilan M, "Deep Learning Technique for Medicinal Plant Leaf Identification: Using Fine-Tuning of Transfer Learning Model", 2024 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), pp.1-6, 2024.
[24] Sai Subramanian Srambickal, Aman Srivastava, S. K. Somasundaram, "Medical Plant Detection using Transfer learning Coupled with K-Nearest Neighbour Algorithm", 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT), pp.1-6, 2024.