IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

AUTOMATED IDENTIFICATION OF MEDICINAL PLANTS USING MACHINE LEARNING

G. Priyanga [1], A. Aravindh[2], K.I. Ashwin[3], V. Gnana Shekar[4], M. Manjunathan[5].

[1] Assistant Professor, Adhiyamaan College of Engineering, Hosur, Tamil Nadu.

[2,3,4,5]Student's, Adhiyamaan College of Engineering, Hosur, Tamil Nadu.

Abstract: This paper presents a comprehensive review of machine learning techniques for the identification of medicinal plants. Despite several existing studies, there are still challenges in automating the identification of plant species accurately. To determine the model's effectiveness, it must first be trained, then validated, and ultimately tested. The ML model is evaluated using measures including accuracy, precision, and recall. For this reason, the model was able to correctly identify medicinal leaves at an accuracy rate of 98.3%. This system has been implementing a technique for medicinal plant identification using Fuzzy C means, a clustering machine learning algorithm based on color, texture and geometrical features. Then reduced feature vectors are inputted into the classification model. The proposed system utilizes adverse dataset of high-resolution leaf images representing various medicinal plant species like Image preprocessing techniques, including normalization and feature extraction. Several machine learning algorithms, such as Convolutional Neural Networks (CNNs) and Multi Logistic Regression, are investigated to discern their effectiveness in accurately identifying medicinal plant species.

I. INTRODUCTION

Medicinal plants have been utilized for centuries as a vital source of remedies for various ailments, contributing significantly to traditional medicine practices worldwide. However, the accurate identification and classification of medicinal plants are essential yet challenging tasks due to the vast diversity of plant species, variations in morphology, and the potential for misidentification. It has become common practice in recent years to use ML and DL techniques to classify plants based on images of their leaves; this article compares the accuracy and prediction abilities of these methods. A few classifiers used to identify leaves and extract important leaf attributes are described in this study's image-processing methodologies. There are millions of plant species on Earth; some are poisonous to humans, others are used in medicine, and yet others are on the verge of extinction. Plants are vital not just to human existence, but also to the entire stability of the food chain. Herbal plants are plants that can be used to treat illnesses naturally. The quality of the raw ingredients used to make Ayurveda treatments has come under fire as the Ayurvedic industry has become more commercialized. Women and children, who lack the specialized expertise necessary to identify the appropriate therapeutic herbs, now gather the plants from the wild. Incorrect or substitute medicinal plants are frequently delivered to manufacturing facilities. Most of these facilities lack proper quality control procedures to inspect these plants. Ayurveda therapy is useless when medicinal herbs are used improperly. Unanticipated consequences are also a possibility, manually identifying medicinal plants is equally difficult and time-consuming as identifying any other form of the plant. The study analyse the practicality of utilizing convolutional neural network (CNN)-based approaches to discern between several species of leaves.

II. LITERATURE SURVEY

Identification and Classification Methods for Medicinal Plants (March 2023): Shashank M. Kadiwal, Venkatesh Hegde, N. V. Shrivathsa, S. Gowrishankar, A. H. Srinivasa & A. Veena : The study includes the usefulness and reliability of many image processing algorithms, machine learning, and deep learning algorithms for plant classification based on the leaf images used in the recent years with their benefits and drawbacks. The effectiveness of these algorithms in recognizing leaf images based on plant characteristics like shape, grain, texture, and combination of many aspects is evaluated. Our paper looks at both publicly available and self-captured leaf datasets for automated plant identification, and it concludes with a summary of current research and areas for improvement.

III. PROPOSED SYSTEM

A. **Image Processing Techniques:**

- **Optimizing Footprints:** This likely refers to preprocessing the images to enhance their quality or reduce noise. Techniques such as resizing, denoising, or applying filters (e.g., Gaussian blur) may be used to optimize the footprint of the image, making it more suitable for further analysis.
- **Greyscale Conversion:** Greyscale conversion is a common preprocessing step in image processing tasks. It simplifies the image to a single channel (intensity) while retaining important visual information. This step is often performed to reduce computational complexity and focus on key features like edges and textures.
- **Edge Detection:** Edge detection algorithms identify and highlight boundaries or edges in an image. Techniques like the Canny edge detector or Sobel operator are commonly used for this purpose. Edge detection helps in segmenting objects in the image and extracting important structural information**.**

B. **Feature Extraction:**

- All images in the dataset are processed to extract texture features using Fuzzy C-means clustering.
- Fuzzy C-means clustering helps preserve texture features across different frequencies in the image, capturing important visual characteristics.

C. **Classification Model:**

- Initially, a binary CNN classifier is employed for identifying medicinal plants, designed for distinguishing between two classes (e.g., medicinal plant vs. non-medicinal plant).

D. **Extension to Multinomial Logistic Regression:**

- Since the classification task involves multiple classes of medicinal plants, the binary CNN model is extended to handle multinomial logistic regression.
- Multinomial logistic regression allows the model to estimate probabilities for each class, accommodating the complexity of multiple class distinctions.

E. **Combination of Binary CNN and Multinomial Logistic Regression:**

- The proposed system combines the strengths of the binary CNN model's feature extraction capabilities with the multinomial logistic regression's ability to handle multiple classes.
- Multiple groups are compared using multinomial logit, enhancing the model's accuracy in classifying different types of medicinal plants based on extracted features.

IV. MODEL BUILDING

A. Dataset Preparation:

Include botanical gardens, herbarium collections, online databases (such as USDA Plants Database, Tropicos, or iNaturalist), research publications, and field surveys. Collecting high-quality images with clear visibility of plant features and minimal distortion or noise. Preferably, choosing images that are well-lit, properly focused, and captured from multiple angles for identification of medicinal plants.

fig 1: medicinal plants data

B. Model Configuration

A Convolutional Neural Network (CNN) classifier, initially designed for binary classification, is extended to handle multi-class classification using Multinomial Logistic Regression (MLR). The CNN component of the model learns hierarchical features from the images.

C. Model Accuracy

An accuracy of 98.3% means that the model correctly identifies medicinal plants in images about 98 times out of every 100 images. This high accuracy indicates that the model is very good at distinguishing between different plant species based on their visual features. It's like having a highly accurate expert who can recognize medicinal plants with great precision, making the model reliable for practical use in plant identification tasks.

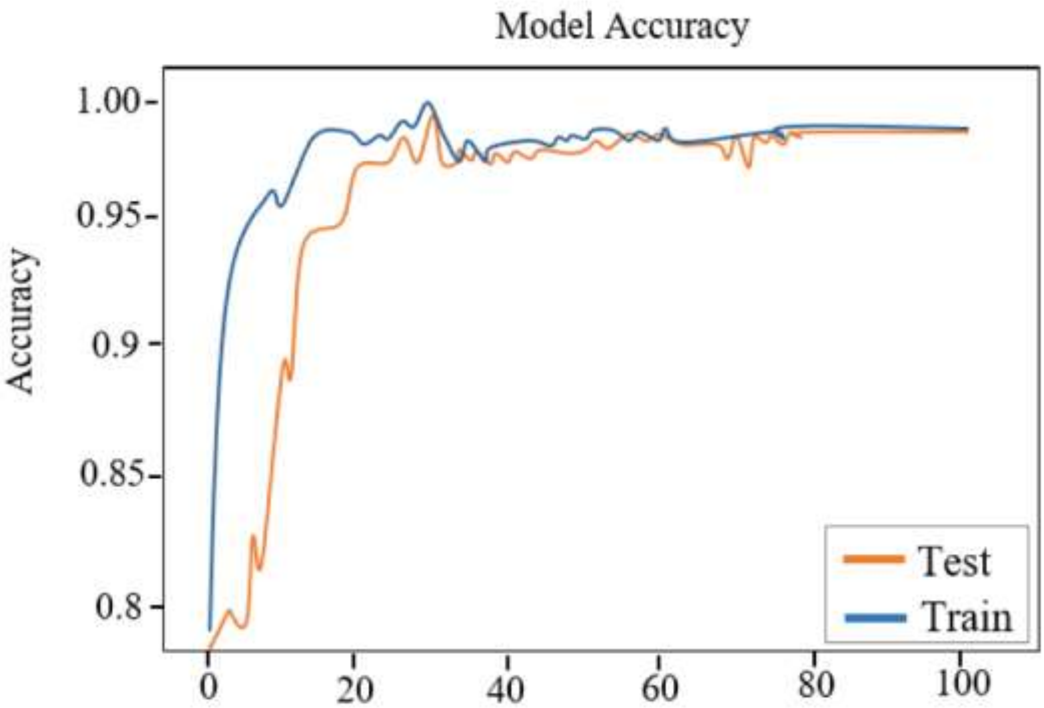

fig 2: accuracy

D. Model Evaluation:

Model evaluation involves assessing the performance and reliability of a machine learning model, particularly in the context of identifying medicinal plants from images.

- **Precision:** The ratio of true positive predictions to the total number of positive predictions. Precision measures the model's ability to avoid false positive predictions.
- **Recall (Sensitivity):** The ratio of true positive predictions to the total number of actual positive instances in the dataset. Recall measures the model's ability to capture all positive instances.
- **F1-score:** The harmonic mean of precision and recall. It provides a balanced measure of a model's performance, especially in situations where there is an imbalance between classes.

- **Confusion Matrix:** A matrix that summarizes the counts of true positive, true negative, false positive, and false negative predictions, providing insights into the model's performance for each class.
- **ROC Curve and AUC:** Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) measure the trade-off between true positive rate and false positive rate.

E. Architecture

1. **Input Layer:**
- The Architecture diagram starts with an input layer that accepts grayscale or RGB images of medicinal plants. The input images are typically pre-processed to optimize footprints, convert to grayscale, and apply edge detection.
2. **Convolutional Neural Network (CNN):**
- The CNN component comprises multiple convolutional layers responsible for feature extraction. Each convolutional layer uses filters to detect patterns and features at different spatial scales.
3. **Pooling Layers:**
- Pooling layers follow the convolutional layers, reducing spatial dimensions and retaining important features. Pooling helps in capturing invariant features and reducing computational complexity.
4. **Fuzzy C-means Clustering (Texture Feature Extraction):**
- As part of preprocessing and feature extraction, Fuzzy C-means clustering is applied to extract texture features from the images. This step helps in preserving texture information across different frequencies.
5. **Fully Connected Layers:**
- Fully connected layers follow the flattened layer, performing high-level feature learning and classification. Dropout regularization may be included to prevent overfitting by randomly dropping connections during training.
6. **Multinomial Logistic Regression (MLR):**
- The output layer of the model is configured with multiple nodes corresponding to different classes of medicinal plants. MLR is used for multi-class classification, computing class probabilities and predicting the most likely class for a given input image.

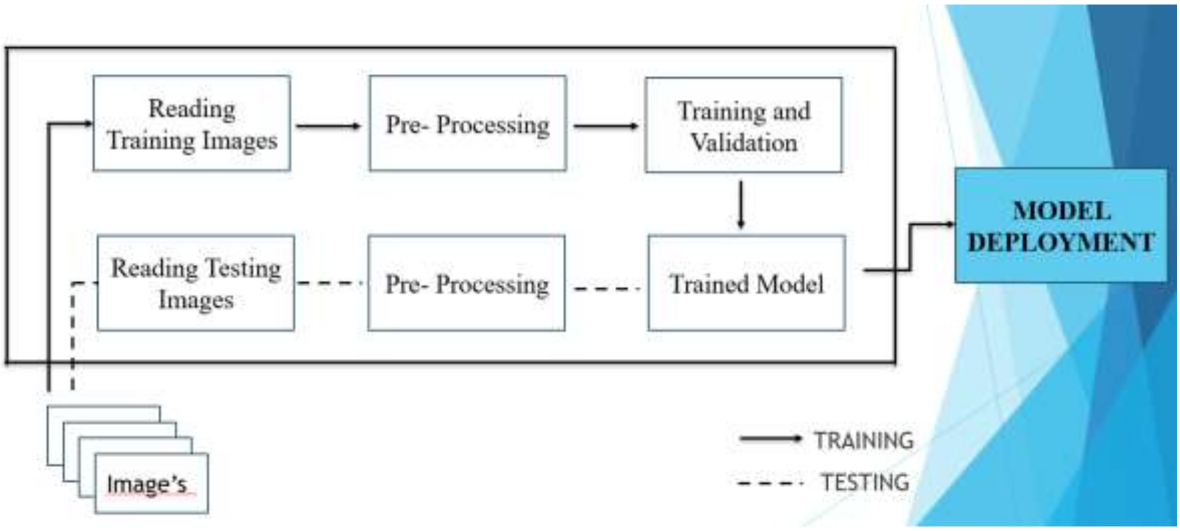

fig 3: architecture

V. CONCLUSION:

The proposed system integrates image processing techniques, Fuzzy C-means clustering, CNN classifier, and multinomial logistic regression to create a robust framework for identifying medicinal plants from images. Here's a conclusion for this proposed system:

The combination of image processing techniques such as optimizing footprints, grayscale conversion, and edge detection, followed by feature extraction using Fuzzy C-means clustering, lays a strong foundation for capturing essential texture features of medicinal plants. This preprocessing step enhances the quality of input data and ensures that crucial visual characteristics are preserved.

The utilization of a CNN classifier, initially designed for binary classification but extended to handle multinomial logistic regression, demonstrates the model's adaptability and scalability for classifying multiple classes of medicinal plants. This approach leverages the hierarchical feature learning capabilities of CNNs, allowing the model to differentiate between diverse plant species based on learned patterns and features.

The extension to multinomial logistic regression enables the model to estimate probabilities for each class, facilitating multi-class classification and enhancing the accuracy of plant identification. By comparing multiple groups through multinomial logit, the model can make more informed and precise predictions, contributing to reliable and trustworthy results.

Overall, the proposed system offers a comprehensive and effective solution for automated medicinal plant identification. Its ability to preprocess images, extract texture features, and classify multiple plant species accurately makes it a valuable tool for researchers, botanists, and herbalists in various fields. The system's adaptability, accuracy, and scalability make it well-suited for practical applications and contribute to advancements in plant identification and biodiversity conservation efforts.

VI. REFERENCE

[1] Shobha Rani N., Pushpa B.R. A simple and efficient technique for leaf extraction in complex backgrounds of low-resolution mobile photographed images. J. Intell. Fuzzy Syst. 2022

[2] Rao R.U., Lahari M.S., Sri K.P., Srujana K.Y., Yaswanth D. Identification of medicinal plants using deep learning. *Int. J. Res. Appl. Sci. Eng. Technol.* 2022;10:306–322.

[3] Sachar S, Kumar A. Deep ensemble learning for automatic medicinal leaf identification. Int J Inf Technol. 2022;14(6):3089–97.

[4] Chanyal H, Yadav RK, Saini DKJ. Classification of medicinal plants leaves using deep learning technique: a review. Int J Intell Syst Appl Eng. 2022;10(4):78–87.

[5] Rao RU, Lahari MS, Sri KP, Srujana KY, Yaswanth D. Identification of medicinal plants using deep learning. Int J Res Appl Sci Eng Technol. 2022;10:306–22.

[6] Valdez DB, Aliac CJG, Feliscuzo LS. Medicinal plant classification using convolutional neural network and transfer learning. In: 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology.

[7] Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. 2018, 145, 311–318.

[8] Bisen, D. Deep convolutional neural network based plant species recognition through features of leaf. Multimed. Tools Appl. 2021, 80, 6443–6456.

[9] Chukwuma, E.C.; Soladoye, M.O.; Feyisola, R.T. Traditional medicine and the future of medicinal Plants in Nigeria. J. Med. Plants Stud. 2015, 3, 23–29.

[10] Russel, N.S.; Selvaraj, A. Leaf species and disease classification using multiscale parallel deep CNN architecture. Neural Comput. Appl. 2022, 34, 19217–19237.

[11] Muneer, A.; Fati, S.M. Efficient and automated herbs classification approach based on shape and texture features using deep learning. IEEE Access 2020, 8, 196747–196764.

[12] Mahendran, G.; Rahman, L.U. Ethnomedicinal, phytochemical and pharmacological updates on Peppermint (Mentha × piperita L.)—A review. Phytother. Res. 2020, 34, 2088–2139.

[13] Roopashree S., Anitha J., Mahesh T.R., Kumar V.V., Viriyasitavat W., Kaur A. An IoT based authentication system for therapeutic herbs measured by local descriptors using machine learning approach. Measurement. 2022;200.

[14] Thanikkal J.G., Dubey A.K., Thomas M.T. Deep-Morpho Algorithm (DMA) for medicinal leaves features extraction. Multimedia Tools Appl. 2023:1–21.

[15] Divyasree, G., & Sheelarani, M. (2022). Ayurvedic leaf identification using deep learning model: VGG16.