IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

IDENTIFICATION OF MEDICINAL PLANTS USING MACHINE LEARNING

[1]Miss. Samiksha Dada Upadhye, [2] Miss.Shreya Shivaji Waghamode, [3]Miss. Ruchita Vijay Desai, [4]Miss. Anuja Anil Kandurkar, [5]Prof. Suraj Krishna Patil

[1,2,3,4]Student, [5]Assistant Professor

Department of CSE (Data Science)

D. Y. Patil College of Engineering and Technology, Kolhapur, India

Abstract: In recent years, there has been a growing interest in utilizing machine learning techniques to identify medicinal plants. This paper presents an overview of the methodologies and advancements in the field of medicinal plant identification using machine learning techniques. The process involves feature extraction and classification using supervised learning techniques like Random Forest, Convolutional Neural Networks (CNN), and Deep Learning models. Furthermore, the paper highlights the importance of data augmentation, transfer learning, and model interpretability in improving the accuracy and reliability of medicinal plant identification systems. In addition to identification, our system incorporates a recommendation feature based on cosine similarity, which suggests similar medicinal plants based on textual input or plant properties, aiding users in discovering alternative options with similar therapeutic uses. To enhance accessibility and usability across diverse user groups, a translation module powered by the Google ML Kit Translation formation in multiple languages. Overall, our research demonstrates the feasibility and effectiveness of using machine learning techniques for the automated identification and recommendation of medicinal plants. By combining botanical knowledge with advanced computational methods, we pave the way for efficient, multilingual, and reliable identification systems, thereby encouraging the conservation and sustainable utilization of medicinal plant resources. Lastly, future directions and opportunities for research in this domain are proposed, emphasizing the integration of multi-modal data sources, development of user-friendly applications, and deployment of robust and scalable models for real-world applications in healthcare and conservation.

Key Words **– Machine Learning Algorithm, CNN, Google ML Kit, Cosine Similarity**

I. INTRODUCTION

In recent years, technology and traditional medicine have teamed up to make some exciting discoveries. One big breakthrough is using machine learning (ML) to identify and group medicinal plants. This is really important because more and more people are interested in natural remedies, and we need better ways to identify these plants accurately and quickly. Traditionally, identifying medicinal plants relied on experts, but they're not always around, especially in places with lots of different plants but not many experts. Plus, it takes a long time and mistakes can happen when people do it manually. But with machine learning, we can teach computers to recognize patterns and features in plant data. This means we can automate the identification process, giving us fast and reliable results. This paper looks at how machine learning can help us identify medicinal plants. We'll talk about how we gather data, pick out important details, train our computer models, and make sure they work well. In addition to identification, we explore how recommendation systems—using techniques like cosine similarity—can suggest similar medicinal plants based on user queries or plant properties. To make this information accessible to a wider audience, we also integrate translation tools powered by the Google ML Kit Translation API, allowing users to view plant details in their native language. We'll also talk about the challenges we face, like not having enough data or making sure our models make sense. We'll also show some innovative features, like making apps that can identify plants instantly, recommend similar alternatives, or keep

track of plant diversity in big areas. These not only help us find healing plants but also make sure we're taking care of our environment better.

II. LITERATURE SURVEY

R. Hu et al.[1] initiated a method focusing on fast recognition of plant leaves using a multiscale distance matrix. While their approach significantly improved the speed of plant leaf recognition, a key drawback was its limited effectiveness when dealing with complex or overlapping leaves. The method's reliance on geometric features made it less robust in situations where plant leaves had similar shapes but different textures, thus reducing accuracy in more diverse plant datasets.

Y. Herdiyeni et al.[2] explored the use of local binary pattern variance and color moments for identifying Indonesian medicinal plants. Although their work represented a crucial step toward incorporating AI techniques into medicinal plant identification, the primary limitation was the reliance on fuzzy logic, which can be sensitive to noise and variations in lighting conditions. This reduced the system's reliability in real-world scenarios, particularly when images were captured in non-controlled environments.

R. Janani and A. Gopal[3] examined image features and artificial neural networks (ANN) to classify selected medicinal plants. While their contribution improved accuracy, the use of ANN had limitations in terms of scalability and computational efficiency. ANNs, being shallow models compared to deep learning techniques, struggled with large and complex datasets, leading to potential overfitting and a lack of generalization in identifying a broader range of medicinal plants.

G. Grinblat et al.[4] transitioned into deep learning methodologies, particularly CNNs, emphasizing vein morphological patterns in leaves. A major drawback of their approach was the high computational cost associated with training deep learning models, such as CNNs. Additionally, deep learning methods require large labeled datasets, which can be difficult to obtain, especially for rare or less-studied medicinal plants. This limitation constrained the model's applicability in real world scenarios where diverse and extensive training data may not be available.

Kiflie Mulugeta et al.[5] compiled a comprehensive review of the evolution of deep learning techniques for medicinal plant classification. Although their review highlighted significant advancements, a key limitation was the lack of focus on the practical implementation challenges of deep learning models. The paper did not address issues such as the need for specialized hardware, the difficulty in interpreting deep learning models, and the 4 challenges in transferring these techniques from research settings to practical, user-friendly applications. These challenges can hinder the widespread adoption of deep learning models in everyday medicinal plant identification.

Summary:

Classification using various image processing and machine learning techniques. R. Hu et al. [1] proposed a fast leaf recognition method based on multiscale distance matrices, enhancing efficiency but facing limitations with complex or overlapping leaves. Y. Herdiyeni et al. [2] utilized local binary pattern variance (LBPV) and color moments for identifying Indonesian medicinal plants, though their system struggled in variable lighting conditions due to the sensitivity of fuzzy logic. R. Janani and A. Gopal [3] implemented artificial neural networks (ANNs) for classification using manually extracted features, achieving good accuracy but encountering scalability issues with shallow networks. G. Grinblat et al. [4] introduced convolutional neural networks (CNNs) to analyze leaf venation patterns, achieving high accuracy but at the cost of high computational demand and low interpretability. Lastly, Kiflie Mulugeta et al. [5] reviewed deep learning approaches including CNNs and RNNs, highlighting their potential while also pointing out challenges related to real-world deployment, such as hardware needs, data availability, and user accessibility. Collectively, these studies underscore the progress and ongoing challenges in building practical, scalable, and user-friendly plant identification systems.

Motivation:

This progression reflects a shift from simple image processing techniques to more advanced machine learning and deep learning models over the years. However, each approach comes with its own set of drawbacks, which can limit its effectiveness and applicability in real world scenarios.

III. System Architecture

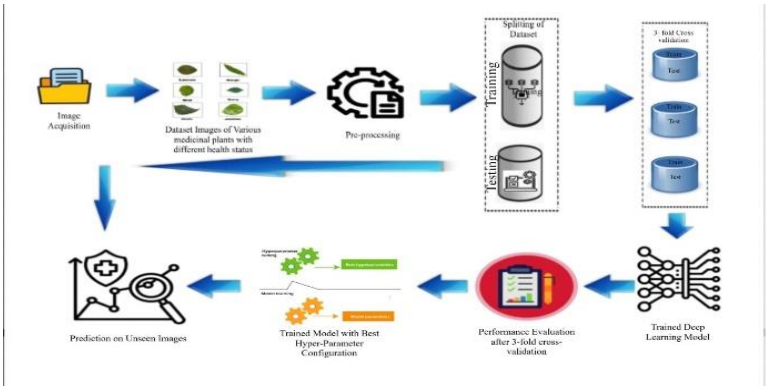


Fig.1.1: System Architecture

The proposed system for medicinal plant identification is implemented through a modular architecture. It incorporates a combination of image preprocessing, feature extraction, model training, and classification modules. The implementation follows a systematic pipeline to ensure robustness, accuracy, and scalability.

1. **Image Acquisition Module**
   High-resolution images of medicinal plant leaves are collected either from publicly available datasets or through manual image capture using a smartphone or camera. Images are standardized in terms of resolution and format for consistency.

2. **Image Processing Module**
   This module performs the following operations:
   - **Resizing**: Standardizing input size (e.g., 128×128 pixels) for model compatibility.
   - **Noise Removal**: Gaussian blur or median filtering to remove background noise.
   - **Segmentation**: Background removal to isolate the leaf using thresholding or contour detection.
   - **Normalization**: Pixel value normalization between 0 and 1.
   - **Augmentation**: Techniques such as rotation, flipping, and colour manipulation are applied to increase dataset diversity and reduce overfitting

3. **Feature Extraction Module**
   Features extracted may include:
   - **Shape features**: Leaf length, width, perimeter, and aspect ratio.
   - **Colour features**: Mean and standard deviation of RGB and HSV values.
   - **Texture features**: Using Local Binary Patterns (LBP) or Gray-Level Co-occurrence Matrix (GLCM).
   - In deep learning models, this module is integrated within CNN layers.

4. **Classification Module**
   This module is responsible for identifying the plant species based on extracted features. It uses machine or deep learning algorithms trained on labelled data. The following algorithms are employed.

IV. Implementation Details

Algorithms:

1. **Convolutional Neural Network(CNN)**

CNN is used as the primary deep learning model due to its ability to automatically learn spatial features from images.

- **Layers Used**: Convolutional → ReLU → Pooling → Fully Connected → Softmax
- **Optimization**: Adam optimizer with categorical cross-entropy loss.
- **Transfer Learning**: Pretrained models like MobileNet or ResNet50 are utilized for improved accuracy on small datasets.

2. **Cosine Similarity**

Cosine similarity are used as the primary text-based techniques for identifying medicinal plant diseases based on symptom descriptions.

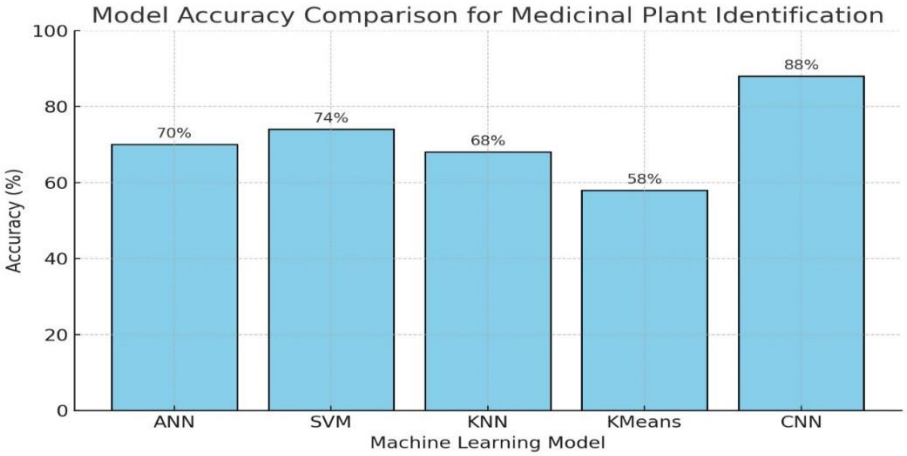
- **Feature Extraction**: Textual disease descriptions are converted into numerical vectors using Term Frequency–Inverse Document Frequency (TF-IDF).
- **Similarity Measure**: Cosine similarity is used to compare the user's input description with existing disease profiles and retrieve the most similar matches.
- **Application**: Enables disease identification when image data is unavailable, making it suitable for text-based symptom inputs or reports.
- **Advantages**: Lightweight, interpretable, and efficient for text-based recommendation and classification tasks

3. **Google ML Kit Translation API**

The Google ML Kit Translation API is used to enhance accessibility by providing real-time translation of medicinal plant information into multiple languages.

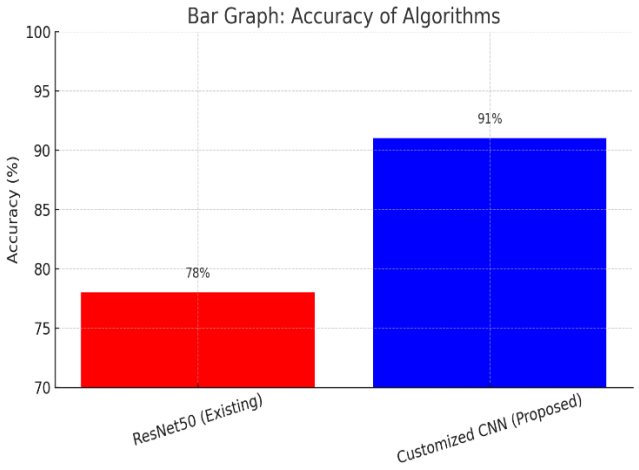
- **Integration**: The API is integrated into the application to support dynamic translation of plant names, uses, symptoms, and disease descriptions.
- **Language Support**: Offers translation for over 50 languages, enabling a multilingual user experience.
- **Offline Capability**: Supports on-device translation when language models are downloaded, ensuring usability without continuous internet access.
- **Advantages**: Increases accessibility for users from diverse linguistic backgrounds and promotes wider adoption of the application.

V. RESULT ANALYSIS



Convolutional Neural Networks (CNNs) were used for medicinal plants plant identification since they are very good at handling image based data. Unlike other traditional algorithms, CNNs can automatically learn and extract significant features such as leaf shape, texture, edges and color gradients without human interference. This renders them more efficient in identifying faint visual patterns, particularly difficult datasets. In this system, CNN recorded a 91% accuracy, surpassing all the other models. Its better performance, flexibility to diverse conditions, and compatibility with AI acceleration technologies such as TensorFlow Lite make it the most stable and scalable option for the application.

System	Algorithm Used	Accuracy (%)	Description
Existing	Pre-trained ResNet50	78%	Uses a generic model not fine-tuned for medicinal plants. Limited to basic classification without symptom linkage.
Proposed	Customized CNN	91%	Trained specifically on medicinal plant datasets. Integrated with NLP and recommendation modules for symptom-based treatment suggestions.



VI. Conclusion & Future scope

Plants, especially herbs, have been crucial for human health since ancient times. Native communities have long depended on their knowledge of herbs for medicinal purposes, often passed down through generations based on sensory experiences like smell and taste. However, recent advancements in technology have made it easier to scientifically identify herbs, which is especially helpful for those not familiar with traditional ancient methods. To address these challenges, there's a growing interest in using computational and statistical methods for herb identification. This approach is non-destructive, meaning it doesn't harm the plants, and it's particularly useful for quickly identifying herbs, especially for those without access to expensive lab equipment.

In addition to identification, recommendation systems based on techniques like cosine similarity can suggest herbs with similar properties or therapeutic uses, providing valuable alternatives for treatment and research. Furthermore, integrating translation capabilities through tools like the Google ML Kit Translation API ensures that herb-related information is accessible to users across different linguistic backgrounds, promoting inclusive and widespread use.

Overall, this review explores various methods for identifying plants, weighing their benefits and drawbacks, and highlights the potential of combining computational, statistical, and linguistic approaches for efficient, accessible, and user-friendly herb identification.

In addition to identifying medicinal plants and diagnosing diseases, the project aims to provide treatment recommendations based on the detected health status. Future enhancements include integrating a recommendation system for remedies and preventive care, as well as time-series analysis to predict disease progression. To improve accessibility, especially for regional users, a multi-language translation feature (e.g., Marathi) will be added. By boosting prediction accuracy, expanding the dataset, and enhancing user experience, the project supports medicinal plant conservation and sustainable agriculture.

VII. Acknowledgement

We gratefully acknowledge the support and resources that made this research possible. This work focuses on the identification of medicinal plants using machine learning, incorporating modules for plant identification, recommendation, and translation. We would like to thank our mentors and academic guides for their valuable insights and feedback throughout the research process. We also appreciate the availability of open-source tools and libraries, which enabled the implementation of techniques such as Convolutional Neural Networks (CNN) for image-based identification, cosine similarity for recommendation, and the Google ML Kit Translation API for multilingual support. Their integration played a crucial role in building an accessible, scalable, and effective medicinal plant identification system.

REFERENCES

[1] R. Hu, W. Jia, H. Ling, D. Huang, "Multiscale distance matrix for fast plant leaf recognition", IEEE Transactions on Image Processing, Volume 21, Issue 11, 2012, pp. 4667-4672.

[2] Y. Herdiyeni, E. Nurfadhilah, E. Zuhud, E. Damayanti, K. Arai, H. Okumura, "A computer-aided system for tropical leaf medicinal plant identification", International Journal of Advanced Science, Engineering, and Information Technology, Volume 3, Issue 1, 2013, pp. 23-30.

[3] R. Janani, A. Gopal, "Identification of selected medicinal plant leaves using image features and ANN", International Conference on Advanced Electronic Systems, 2013, pp. 238-243.

[4] nblat G., Uzal L., Larese M., Granitto P., "Deep learning for plant identification using vein morphological patterns", Computers and Electronics in Agriculture, Volume 127, 2016, pp. 418-424.

[5] Adibaru Kiflie Mulugeta, Durga Prasad Sharma, Abebe Haile Mesfin, "Deep learning for medicinal plant species classification and recognition: a systematic review", Frontiers in Plant Science, Volume 14, 2023, pp. 1-20, 2024.