Vivekanand Education Society's Institute of Technology

Department of Computer Engineering

Group No.: 30
Date :- 02/08/2024

Project Synopsis (2024-25) - Sem VII

MedLeaf: AI-based Identification and Medicinal Value Assessment of Flora

Mrs. Priya R. L
Assistant Professor, Department of Computer Engineering

Sanika Hadap
V.E.S.I.T
2021.sanika.hadap@ves.ac.in

Tanvi Naik
V.E.S.I.T
2021.tanvi.naik@ves.ac.in

Kevin Patel
V.E.S.I.T
d2021.kevin.patel@ves.ac.in

Abstract

The identification of medicinal plants and the analysis of their benefits have significant implications for traditional medicine, agricultural development, and pharmaceutical research. This project leverages the power of artificial intelligence and computer vision to create a comprehensive system for identifying medicinal plants and analyzing their medicinal properties. By utilizing convolutional neural networks (CNNs), we aim to accurately classify various medicinal plants based on images. The system will be trained using publicly available datasets, data scraped from reliable websites, and data collected from trusted sources in the field. Additionally, the medicinal benefits will be curated from official websites, authoritative books, and verified practitioners of plant-based medicine. The project is designed to assist Ayurvedic practitioners, medicinal plant harvesters, and students by providing a reliable tool for plant identification and finding its benefits. One of the primary challenges addressed in this project is the management and accuracy of a large training dataset to ensure high precision in plant identification and finding its benefits.

Introduction

Medicinal plants have been central to traditional medicine and continue to be crucial for modern pharmacology. Accurate identification of these plants is vital for their effective use, ensuring the correct utilization of medicinal plants and preventing the use of incorrect or harmful species[1].

A large number of higher plants have been used as a source of drugs by mankind for several thousand years. It is estimated that 35,000 to 70,000 plant species have at one time or another been used in some culture for medicinal purposes[2].

Traditional methods of plant identification are time-consuming and require extensive expertise. This project utilizes computer vision to enhance plant identification and finding medical benefits. Using convolutional neural networks (CNNs), we aim to classify medicinal plants accurately based on images[3]. The models will be trained on datasets compiled from publicly available sources, web scraping, and reliable field data.

Beyond identification, the project will analyze and present medicinal benefits using information from official websites, authoritative literature, and verified practitioners. This tool will assist Ayurvedic practitioners, medicinal plant harvesters, and students, providing a reliable resource for plant identification and benefit analysis.

Problem Statement

Current methods for identifying medicinal plants are often labor-intensive and require specialized expertise, presenting significant barriers for practitioners, harvesters, and students. Traditional identification processes, which rely on detailed botanical knowledge and manual examination of plant characteristics, are both time-consuming and complex. This complexity limits accessibility and efficiency, making it challenging for non-experts and professionals alike. The growing shortage of skilled taxonomists and the broader "taxonomic crisis" further exacerbates these difficulties. Additionally, existing tools fall short in integrating detailed medicinal benefits information, which hinders users' ability to access comprehensive and relevant data effectively. With the rapid decline in biodiversity and increasing demand for accurate species identification, there is an urgent need for an integrated, user-friendly solution that bridges the gap between precise plant identification and extensive medicinal data[4]. This solution should leverage modern technologies, such as computer vision and digital image processing, to enhance accessibility and efficiency in medicinal plant research and application, thus supporting both traditional practices and contemporary research.

Proposed Solution

The proposed solution is a comprehensive AI-driven application that utilizes convolutional neural networks (CNNs) for the accurate identification of medicinal plants through image classification, complemented by natural language processing (NLP) techniques to analyze and present their medicinal benefits[5]. The application will be developed by collecting a diverse dataset of plant images and their corresponding medicinal properties from reliable sources, followed by data preprocessing to enhance model training. A user-friendly interface will be designed to facilitate seamless interaction, allowing users—such as Ayurvedic practitioners, plant harvesters, and students—to easily identify plants and access detailed information on their medicinal uses. By integrating precise plant identification with extensive medicinal data, the solution aims to improve accessibility, efficiency, and support for traditional practices and contemporary research in herbal medicine.

Block Diagram

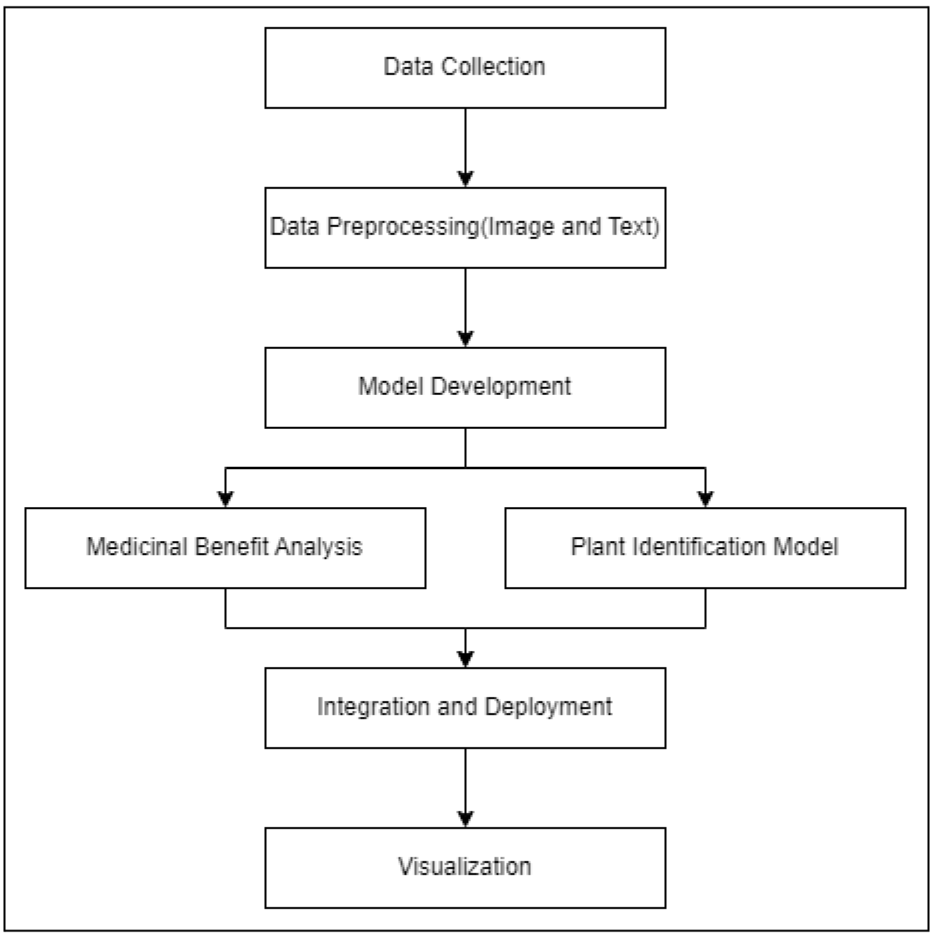

Fig: Block Diagram

1. **Data Collection and Preparation**
    a. **Image Dataset Collection**: Gather images of various plants from reliable sources such as plant databases, botanical gardens, and online repositories.
    b. **Medicinal Information Collection**: Compile a database of plants with their medicinal properties and associated raw materials
    c. **Data Annotation**: Label the collected images with plant species names and annotate the medicinal benefits in the database.
2. **Data Preprocessing**
    a. **Image Preprocessing**: Resize and normalize images, and augment the dataset with transformations
    b. **Text Preprocessing:** Clean the text data Model Development
3. **Model Development**
    a. **Plant Identification Model:**
        i. Model Selection: ResNet, Inception, or EfficientNet architectures to perform accurate image classification and transfer learning with pre-trained models such as VGG16, ResNet50, or MobileNet.
        ii. Training: Data augmentation techniques to enhance dataset diversity, k-fold cross-validation for robust model evaluation, and optimizers like Adam or RMSprop, experimenting with learning rate schedules and early stopping for optimal performance.
    b. **Finding Medicinal Benefits:**
        i. Model Selection: NLP models like BERT, GPT, or RoBERTa to extract and comprehend contextual information from text data and Named Entity Recognition (NER) to identify and categorize key medicinal entities and use relation extraction techniques to understand relationships between entities.
        ii. Training: Annotating datasets where medicinal information is labeled for training purposes. Fine-tuning pre-trained NLP models on the annotated dataset to adapt them to the specific domain of medicinal plants and utilizing knowledge graphs to represent relationships between plants and their medicinal benefits, facilitating efficient information extraction and organization.
4. **Integration and Deployment:** Connecting the frontend with the backend to ensure seamless data flow and real-time response to user queries, then deploying the application.
5. **Visualization:** Create visual representations of the plant identification results, including images of identified plants and their medicinal benefits.

Hardware, Software and tools Requirements

Hardware Requirements:

- **Processor:** Multi-core processor (Intel i7 or AMD Ryzen 7 and above)
- **Memory:** Minimum 16 GB RAM
- **Storage:** SSD with at least 500 GB storage for fast data access
- **Graphics Card:** Dedicated GPU (NVIDIA GTX 1060 or above) for training CNNs

Software Requirements:

- **Development Environment:** Windows 10/11
- **Programming Languages:** Python
- **IDE/Editor:** PyCharm, VS Code, or Jupyter Notebook

Libraries:

- TensorFlow or PyTorch for deep learning
- OpenCV for image processing
- NLTK or spaCy for NLP tasks

Proposed Evaluation Measures

- Identification Accuracy: Confusion Matrix, Precision and Recall, F1 Score
- Classification Performance: Top-K Accuracy, K-fold cross-validation
- Benefit Analysis Accuracy: Entity Recognition Accuracy, Information Retrieval Precision, Relation Extraction Accuracy

Conclusion

Our project effectively harnesses the capabilities of artificial intelligence and computer vision to transform how medicinal plants are identified and analyzed. Through sophisticated image classification and natural language processing, our tool delivers precise plant identification and comprehensive medicinal insights, streamlining processes that previously required extensive expertise. This innovation provides a practical, accessible resource that enhances the efficiency and accuracy of medicinal plant research, offering significant benefits to practitioners, harvesters, and researchers alike.

References

1. "Plant leaf Disease detection, classification, and diagnosis Using Computer Vision and Artificial intelligence: A review," IEEE Journals & Magazine | IEEE Xplore, 2024. https://ieeexplore.ieee.org/document/10458943

2. "International trade in non-wood forest products: An overview - XI. Medicinal plants." https://www.fao.org/4/x5326e/x5326e0e.htm

3. J. Wäldchen and P. Mäder, "Plant Species Identification Using Computer Vision Techniques: A Systematic Literature review," *Archives of Computational Methods in Engineering*, vol. 25, no. 2, pp. 507–543, Jan. 2017, doi: 10.1007/s11831-016-9206-z. https://link.springer.com/article/10.1007/s11831-016-9206-z

4. P. K. Pang and K. H. Lim, "Review on automatic plant identification using Computer Vision approaches," *IOP Conference Series Materials Science and Engineering*, vol. 495, p. 012032, Jun. 2019, doi: 10.1088/1757-899x/495/1/012032. https://iopscience.iop.org/article/10.1088/1757-899X/495/1/012032/pdf#:~:text=Computer%20vision%20consists%20of%20two,the%20accuracy%20of%20the%20classification

5. "Plant Species Identification Based on Plant Leaf using Machine Learning Techniques," *IEEE Conference Publication | IEEE Xplore*, May 09, 2023. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10166735&tag=1