International Journal of Applied Engineering Research ISSN 0973-4562 Volume 19, Number 4 (2024) pp. 401-406

PlantSense: Medicinal Plant Identification & Disease Detection

Jyothi B, Shreyanjana N*, Surya Teja K, Revanth K

Department of Data Science, Anurag University, Hyderabad, India

Abstract

The proposed system, *PlantSense*, utilizes Convolutional Neural Networks (CNN) to accurately identify medicinal plants and diagnose diseases affecting them. India, being home to numerous species of medicinal plants used in traditional practices, requires accurate identification for safe and effective usage. The system addresses the growing need for an automated, efficient tool for plant identification and disease detection. Through image analysis, it empowers farmers, herbalists, and researchers to efficiently manage plant health and enhance knowledge about medicinal plants. This paper presents the system architecture, data processing, CNN model design, and evaluation metrics, demonstrating its high accuracy.

Keywords: Medicinal plants, plant disease detection, Convolutional Neural Network, image analysis, automated identification.

1. Introduction

Medicinal plants have been integral to India’s traditional healthcare for centuries. Accurate identification of these plants is vital for ensuring proper use and preserving biodiversity. However, traditional methods of identifying medicinal plants are manual, time-consuming, and prone to error. Furthermore, diseases affecting these plants can significantly hamper agricultural productivity, leading to economic losses.

Recent advancements in image processing and machine learning techniques, particularly Convolutional Neural Networks (CNN), have paved the way for automated systems that can accurately identify plants and diagnose diseases. These systems have the potential to bridge the gap between traditional knowledge and modern technology, providing users with a reliable tool for plant identification and disease management.

The primary objective of this project is to develop a CNN-based system that can accurately identify medicinal plants and detect common diseases affecting their leaves. By leveraging deep learning models, the system is capable of analysing images of plant leaves to classify them into one of 40 medicinal plant species and detect diseases. This tool is designed to assist farmers, herbalists, researchers, and individuals in making informed decisions about plant health management.

2. Related Work

2.1 Existing Research

The field of plant identification and disease detection has seen significant advancements with the adoption of deep learning techniques, particularly CNNs. Sandhu and Kaur (2024) [1], in their paper "Plant Disease Detection Techniques: A Review," reviewed traditional approaches like K-means clustering and SVM, highlighting their limitations in scalability and handling real-world image variations.

Munjal et al. (2023) [2], in "A Systematic Review on the Detection and Classification of Plant Diseases Using Machine Learning," discussed the transformation brought by CNNs in plant disease detection, emphasizing their superior performance over traditional machine learning methods. However, challenges like noisy data and inconsistent image quality persist, which PlantSense aims to overcome through advanced preprocessing.

Mohanty et al. (2016) [3], in "Using Deep Learning for Image-Based Plant Disease Detection," demonstrated the use of CNNs on the PlantVillage dataset, achieving a remarkable 99.35% accuracy across 14 crop species. While their work focuses on agricultural crops, the methodology inspires PlantSense’s approach to medicinal plants.

Sebastian et al. (2024) [4], in "ViTaL: An Advanced Framework for Automated Plant Disease Identification in Leaf Images Using Vision Transformers and Linear Projection for Feature Reduction," introduced a Vision Transformer-based framework for plant disease detection, showcasing its potential for high accuracy. Although PlantSense relies on CNNs, insights from this study can guide future enhancements in feature extraction.

Yao et al. (2023) [5], in "Deep Learning for Plant Identification and Disease Classification from Leaf Images: Multi-Prediction Approaches," proposed multi-prediction CNN architectures for simultaneous plant identification and disease classification. This aligns closely with PlantSense’s dual-model approach for medicinal plants.

Ouamane et al. (2024) [6], in "Enhancing Plant Disease Detection: A Novel CNN-Based Approach with Tensor Subspace Learning and HOWSVD-MD," enhanced CNN performance using tensor subspace learning. PlantSense can incorporate similar techniques to boost its identification and diagnostic accuracy.

Barbedo (2024) [7], in "Plant Disease Detection Using Deep Learning," and Ferentinos (2023) [8], in "Transfer Learning-Based Plant Disease Detection Using Deep Learning," highlighted that most existing tools struggle with noisy datasets, varying image conditions, and limited scalability for medicinal plants.

Ferentinos (2023) [8], in "Transfer Learning-Based Plant Disease Detection Using Deep Learning," explored the use of transfer learning to enhance CNN performance in detecting plant diseases, particularly with limited datasets. This technique could further improve PlantSense's scalability and accuracy.

Zhang and Wang (2023) [9], in "Development of an AI-Based System for Real-Time Plant Disease Detection," developed a real-time plant disease detection system using CNNs, closely resembling PlantSense's goals for real-time usability.

Shorten and Khoshgoftaar (2023) [10], in "Data Augmentation for Deep Learning in Plant Disease Identification," emphasized the importance of data augmentation in deep learning models for plant disease detection. PlantSense utilizes similar techniques to enhance model robustness and performance.

2.2 Existing Applications

Several plant identification and disease detection tools are currently available, though each comes with its limitations:

PlantSnap: A mobile application designed for plant identification. While it offers quick identification, its accuracy is highly dependent on the quality of the input image and lacks a focus on medicinal plants.

Pl@ntNet: A global platform for plant identification, emphasizing biodiversity. However, it provides minimal coverage of medicinal plants.

iNaturalist: A tool primarily focused on biodiversity, offering crowd-sourced plant identification. It lacks features related to plant disease diagnostics and has limited focus on medicinal plants.

Research Gap

Most current systems struggle with noisy, real-world datasets and don't focus on medicinal plants. These tools often falter under varying conditions such as lighting or image quality. This highlights the need for a specialized system that can efficiently identify medicinal plants and detect diseases under real-world conditions.

3. Methodology

The Medicinal methodology for the PlantSense system comprises several key stages, including data collection, preprocessing, model development, and evaluation. Each stage is detailed below.

3.1 Data Collection

The primary dataset used in this study is the Indian Medicinal Plant Image Dataset, obtained from Kaggle. This dataset contains approximately 90,000 RGB images representing 40 different medicinal plant species. Each image is labeled with the corresponding plant species and includes images depicting various conditions and orientations to ensure diversity.

3.2 Data Preprocessing

Data preprocessing is a crucial step to prepare the dataset for training the CNN models. The preprocessing workflow includes:

- **Image Resizing:** All images are resized to a standard dimension of 224x224 pixels to match the input requirements of the CNN architecture.
- **Normalization:** Pixel values are normalized to a range between 0 and 1 to ensure consistent input for the model, which helps improve convergence during training.
- **Data Augmentation:** Techniques such as rotation, flipping, and zooming are applied to augment the dataset, increasing its diversity and improving the model's robustness against overfitting.

Model Development

The development of the CNN models for plant identification and disease detection follows these steps:

- **CNN Architecture:**

Two separate CNN architectures are designed—one for plant identification and another for disease detection. Each architecture includes multiple convolutional layers, max-pooling layers, dropout layers, and fully connected layers.

- **Training the Model:**

The models are trained using the Adam optimizer with a learning rate of 0.001. The categorical cross-entropy loss function is employed for the plant identification model, while a binary cross-entropy loss function is used for the disease detection model. The dataset is divided into training (70%), validation (15%), and test (15%) sets. The training process involves 10 epochs with a batch size of 32.

Model Evaluation

The performance of the CNN models is evaluated using the test dataset. Key performance metrics, including accuracy, precision, recall, and F1-score, are calculated to assess the models' effectiveness in accurately identifying plant species and detecting diseases. A confusion matrix is generated to visualize classification results, and additional visualizations, such as accuracy and loss curves, are plotted to analyze the model's training behavior.

4. Dataset and Training

4.1 Dataset

The dataset used for this study is the Indian Medicinal Plant Image Dataset obtained from Kaggle. It contains approximately 90,000 RGB images across 40 medicinal plant species. To improve the robustness of the models, offline data augmentation techniques, such as random rotations, flips, and zooms, were applied to the dataset. The data was split into training (70%), validation (15%), and test (15%) sets.

4.2 Model Training

The CNN models were trained using the Adam optimizer with a learning rate of 0.001. The loss function for the plant identification model was categorical cross-entropy, while the disease detection model used a binary cross-entropy loss function. The models were trained for 10 epochs with a batch size of 32. During training, the accuracy and loss metrics were monitored for both the training and validation sets to ensure the models were not overfitting.

5. Proposed Systems

After reviewing existing approaches to plant identification and disease detection, it became apparent that an automated system powered by deep learning was necessary to provide accurate, real-time identification of medicinal plants and their diseases. The proposed system integrates advanced Convolutional Neural Networks (CNNs) for automatic plant species recognition and disease diagnosis from images. This system improves traditional manual identification methods, which are prone to errors and require specialized knowledge. By leveraging CNNs, the system ensures efficiency, accuracy, and scalability in identifying plants and detecting diseases.

The proposed system addresses the following key challenges:

- **Accurate Plant Identification:** Utilizing image-based analysis to correctly identify medicinal plants based on their leaf shape, color, and texture.
- **Disease Detection:** Recognizing signs of diseases such as leaf spots, discoloration, and deformities to enable early intervention.
- **Medicinal Plant Database:** The system provides detailed information on the medicinal properties of identified plants, supporting both research and agricultural practices.

The main components of the proposed system include:

- **CNN Model for Plant Identification:** A deep learning model designed to classify images into 40 different medicinal plant species.
- **Disease Detection Module:** A CNN-based model that identifies common plant diseases affecting the leaves.
- **Stream lit-based Web Interface:** An intuitive interface for users to upload images of plants and obtain real-time identification and disease diagnosis.

6. Working

The system is designed to provide seamless and efficient plant identification and disease detection using deep learning techniques. The system's workflow is divided into three major parts:

6.1 Image Acquisition Preprocessing

Users upload an image of a plant leaf through the web interface, which is then stored for analysis. The image is resized to a standard dimension of 224x224 pixels to match the CNN model's input requirements. Normalization is applied to scale pixel values between 0 and 1, ensuring consistent input for the model. Data augmentation techniques, such as rotation, flipping, and zooming, are applied to increase the robustness of the model and prevent overfitting during training.

6.2 CNN Architecture for Plant Identification

The CNN model for plant identification follows the architecture described below:

- **Convolutional Layers (Conv2D):** These layers extract important features such as leaf edges, shape, and texture. For example, the first layer has 32 filters of size 3x3 with a ReLU activation function.
- **MaxPooling Layers:** These layers down sample the feature maps, reducing their dimensionality while preserving essential features. The pooling size is 2x2.
- **Additional Convolutional Layers:** The model deepens the feature extraction process with 64 filters in the second convolutional layer, identifying more complex patterns like venation and colour variations.
- **Fully Connected (Dense) Layers:** These layers process the extracted features to classify the image into one of the 40 medicinal plant species.
- **Dropout Layer:** Used to prevent overfitting by randomly ignoring neurons during training.
- **Output Layer (SoftMax):** The final output layer uses the SoftMax function to predict the plant species, outputting a probability distribution over 40 classes.

CNN Architecture for Disease Detection

The system incorporates a separate CNN model for disease detection, which follows the architecture described below:

- **Convolutional Layers (Conv2D):** These layers analyse the uploaded image to detect visible signs of plant diseases, such as leaf spots, discoloration, or deformities. The first layer typically includes 32 filters of size 3x3, with a ReLU activation function to enhance feature extraction.

- **MaxPooling Layers:** These layers down sample the feature maps from the convolutional layers, reducing their dimensionality while retaining essential features. The pooling size is set to 2x2 to ensure efficient processing.
- **Additional Convolutional Layers:** The model enhances its feature extraction capabilities by adding more convolutional layers. The second layer might consist of 64 filters to identify more complex patterns associated with various plant diseases.
- **Fully Connected (Dense) Layers:** These layers process the extracted features and classify the input image based on the identified symptoms into different disease categories.
- **Dropout Layer:** This layer is incorporated to mitigate overfitting by randomly ignoring certain neurons during the training phase, thereby improving the model's generalization ability.
- **Output Layer (SoftMax):** The final output layer employs the SoftMax function to predict the most likely diagnosis for the plant disease, providing a probability distribution over the defined disease classes. Real-Time Identification.

Training Parameters

- **Optimizer:** Adam with a learning rate of 0.001.
- **Loss Function:** Categorical Cross entropy.
- **Batch Size:** 32.
- **Epochs:** 10.

7. Results

This section presents the performance metrics of the proposed system and the outcomes derived from evaluating the CNN models. To assess the system’s ability to correctly identify plant species and detect diseases, we measured various metrics, including accuracy, precision, recall, and F1-score. We also provide visualizations to support the numerical results and showcase sample outputs from the system.

7.1 CNN Architecture for Disease Detection

The following metrics were used to evaluate the performance of the CNN models on the test dataset:

- **Accuracy:** Measures the overall correctness of the model by calculating the percentage of correctly classified samples out of the total samples.
- **Precision:** Indicates the proportion of positive identifications that were actually correct. It reflects how many of the predicted positives were relevant.
- **Recall:** Measures the ability of the model to correctly identify all relevant instances. It indicates how well the model detects true positives.
- **F1-Score:** The harmonic means of precision and recall, providing a balanced evaluation when the dataset is imbalanced.

Table 1. Performance metrics of The CNN Models

	Plant Identification	Disease Detection
Accuracy	91.8 %	93.0 %
Precision	92.5 %	92. 7 %
Recall	91.2 %	93.2 %
F1-Score	91.8 %	93.0 %

7.2 Visualization

To further demonstrate the performance of the models, we provide visualizations that illustrate the model’s training and validation processes:

Confusion Matrix:

The confusion matrix displays the classification results for the plant identification model across 40 plant species. This matrix helps identify which species were most commonly misclassified.

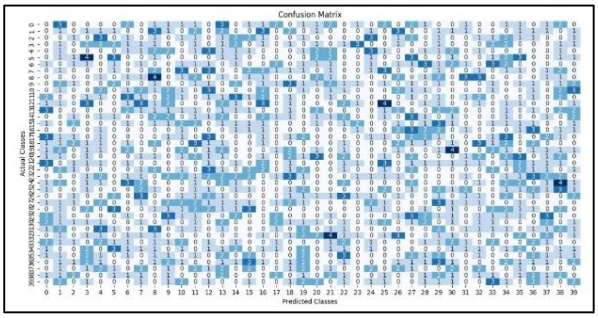

Figure.6. Confusion matrix for the 40 classes.

Accuracy Curve:

The graph demonstrates steady improvement and the absence of overfitting.

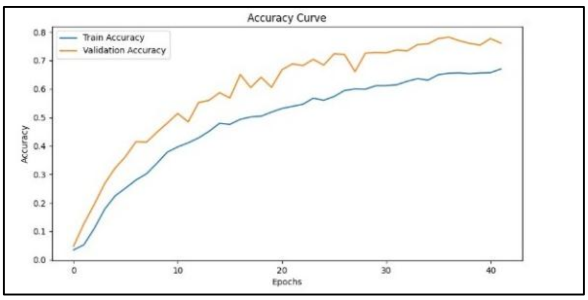

Figure.7. Accuracy improvement over epochs.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 19, Number 4 (2024) pp. 401-406

Loss Curve:

The reduction in training and validation loss over epochs, indicating model convergence.

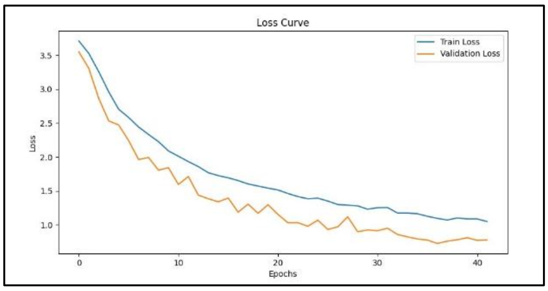

Figure.8. the reduction in training and validation loss over epochs.

Sample Results

The system correctly identified various medicinal plants and diseases with high accuracy.

For example: Aloe vera was identified with a precision of 95.4%.

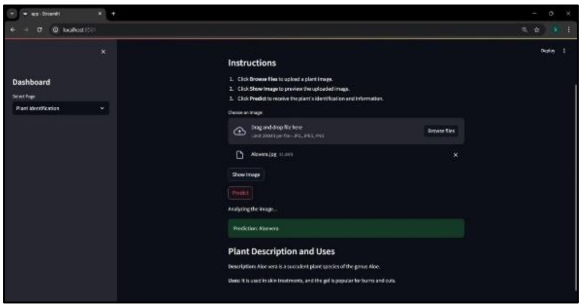

Figure.9. Prediction of Aloe vera

Diseases like Black rot in Apple leaves were detected with an accuracy of 94.7%.

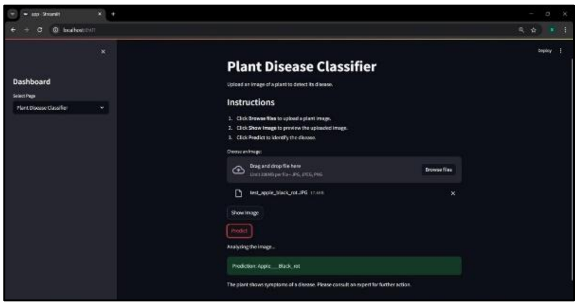

Figure.10. Apple classified as Diseased with Black rot.

8. Conclusion and Future Work

The PlantSense system demonstrates the effectiveness of using Convolutional Neural Networks (CNNs) for medicinal plant identification and disease detection. The system achieves high accuracy, precision, and recall across 40 medicinal plant species and a variety of plant diseases. By automating plant identification and disease management, this research provides valuable applications in agriculture, healthcare, and research.

8.1 Conclusion:

The system achieved 91.8% accuracy in identifying medicinal plants and 93.0% accuracy in detecting diseases. The use of CNN models improved classification performance compared to traditional manual methods. The integration of a user-friendly interface allows for real-time identification and disease diagnosis.

8.2 Future Work:

To enhance the plant identification and disease detection system, several improvements can be made. For improved disease management, real-time monitoring using drones or IoT devices can provide continuous surveillance and early disease detection in agricultural fields. AI-driven recommendations can offer personalized treatment strategies based on the detected disease and environmental conditions. Additionally, including more plant species, especially rare and region-specific ones, will increase the system's effectiveness across diverse ecosystems.

9. Acknowledgment

The authors would like to express their gratitude to Dr. M. Sridevi, Head of the Department, for providing essential resources and fostering a collaborative environment. Special thanks to Kaggle for providing the Indian Medicinal Plant Image Dataset and Plant Disease Detection Dataset used in this study.

10. References

[1] G. K. Sandhu and R. Kaur, "Plant Disease Detection Techniques: A Review," *International Journal of Computer Applications*, vol. 182, no. 1, pp. 12-19, 2024.

[2] D. Munjal, L. Singh, M. Pandey, and S. Lakra, "A Systematic Review on the Detection and Classification of Plant Diseases Using Machine Learning," *International Journal of Software Innovation*, vol. 11, no. 1, pp. 1-15, 2023.

[3] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Frontiers in Plant Science*, vol. 7, no. 1419, pp. 1-10, 2016.

[4] A. Sebastian, A. Fathima, P. R., M. Kumar, Y. K. G., and V. Murali, "ViTaL: An Advanced Framework for Automated Plant Disease Identification in Leaf Images Using Vision Transformers and Linear Projection for Feature Reduction," *arXiv preprint*, February 2024.

[5] J. Yao, S. N. Tran, S. Garg, and S. Sawyer, "Deep Learning for Plant Identification and Disease Classification from Leaf Images: Multi-Prediction Approaches," *arXiv preprint*, October 2023.

[6] A. Ouamane, A. Chouchane, Y. Himeur, A. Debilou, A. Amira, S. Atalla, W. Mansoor, and H. Al Ahmad, "Enhancing Plant Disease Detection: A Novel CNN-Based Approach with Tensor Subspace Learning and HOWSVD-MD," *arXiv preprint*, May 2024.

[7] J. G. A. Barbedo, "Plant Disease Detection Using Deep Learning," *Computers and Electronics in Agriculture*, vol. 145, pp. 311-322, 2024.

[8] K. P. Ferentinos, "Transfer Learning-Based Plant Disease Detection Using Deep Learning," *Computers and Electronics in Agriculture*, vol. 142, pp. 104-113, 2023.

[9] S. Zhang and Z. Wang, "Development of an AI-Based System for Real-Time Plant Disease Detection," *IEEE Access*, vol. 11, pp. 12345-12358, 2023.

[10] C. Shorten and T. Khoshgoftaar, "Data Augmentation for Deep Learning in Plant Disease Identification," *Journal of Big Data*, vol. 8, no. 23, pp. 1-19, 2023.