International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Identification of Medicinal Plants and Disease Detection Through Image Processing Using Machine Learning Algorithms

Nandini R[1], Chandana K[2], Vinutha V[3], Reddy Sreya[4], Dr. Anandaraj S P[5]

[1,2,3,4,5] School of Computer Science and Information Science, Presidency University, Karnataka, India.

Abstract

The accurate identification and disease detection of medicinal plants are essential for advancements in pharmacology, precision agriculture, and biodiversity conservation. This research presents an intelligent, image-based system that integrates advanced image processing techniques with machine learning (ML) algorithms to automate the identification and health assessment of medicinal plants. High-resolution images undergo pre-processing—including noise reduction and normalization—followed by the extraction of morphometric features using Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT), and spectral colour histograms.

These features are processed through optimized ML models such as Convolutional Neural Networks (CNNs), Random Forests, and Support Vector Machines (SVMs), validated using metrics like ROC curves, confusion matrices, and F1 scores.

To enhance usability and scalability, the platform includes modules such as Disease Detection for foliar symptom analysis, My Garden for personalized plant tracking, and AI Assistant for interactive support using NLP and medicinal plant ontologies. Additionally, a Community module encourages collaborative dataset expansion, while the Remedies and Disease Library modules translate identification outputs into actionable knowledge rooted in ethnobotanical data.

Leveraging transfer learning and cloud integration, the system supports regional adaptability and real-time performance. Aligned with the United Nations Sustainable Development Goals (SDGs), this multidimensional framework bridges traditional knowledge with digital innovation to foster sustainable agro-ecological practices, promote biodiversity, and support eco-centric development**.**

Keywords - Medicinal Plant Identification, Image Processing, Machine Learning, Morphometric Analysis, Disease Detection, Feature Extraction, Convolutional Neural Networks (CNN), Ethnopharmacology.

1. Introduction

Medicinal plants have long been regarded as the cornerstone of traditional and modern healthcare systems, earning the title "nature's pharmacy" for their rich repository of bioactive compounds. These plants are vital not only in pharmacological applications but also in precision agriculture, biodiversity conservation, and ecological restoration. As global interest in natural and sustainable therapeutics surges, the demand for accurate identification and classification of medicinal plants has become more critical than ever.

Traditional methods of plant identification rely heavily on manual observation, taxonomic keys, and expert knowledge, which are time-consuming, error-prone, and not scalable across large datasets or geographic regions. Additionally, the increasing threat of plant diseases—caused by biotic (pathogens, pests) and abiotic (climate, pollution) stressors—demands rapid diagnostic systems to ensure plant health and maintain the integrity of medicinal resources.

To overcome these challenges, this research proposes a comprehensive, automated system that leverages advanced image processing techniques and machine learning (ML) algorithms for the identification and disease detection of medicinal plants. The system employs high-resolution imaging, noise-reduction filters, and geometric normalization as pre-processing steps, followed by robust feature extraction methods such as Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT), and spectral color histogram analysis. These features serve as inputs to ML models—including Convolutional Neural Networks (CNNs), Random Forests, and Support Vector Machines (SVMs)—which are fine-tuned through cross-validation and hyperparameter optimization for maximum classification accuracy.

Beyond identification, the system incorporates a suite of user-centric modules: a Disease Detection module for foliar anomaly analysis, a personalized My Garden dashboard for monitoring plant growth, and an AI-powered conversational assistant for real-time support using Natural Language Processing (NLP) and medicinal plant ontologies. Additional components such as the Disease Library and Remedies module enrich the platform with diagnostic insights and ethnobotanical knowledge, making it both educational and actionable.

By integrating machine learning, cloud computing, and intuitive interfaces, the platform provides a scalable solution that supports biodiversity conservation, facilitates sustainable resource management, and empowers communities with digital tools for plant-based knowledge. This initiative is aligned with the United Nations Sustainable Development Goals (SDGs), promoting responsible innovation at the intersection of technology and traditional ecological wisdom

2. Literature Review

1. Hasnae El Allaoui et al. (2024).

The study identified key global trends, leading research institutions, and emerging focus areas such as green chemistry and nanotechnology applications. It emphasized the growing interdisciplinary collaboration in medicinal plant science

2. Firehun Lulesa et al. (2025)

The research explored the ethnobotanical knowledge of indigenous communities in the Yem Zone, Ethiopia. Their study documented numerous medicinal plant species, preparation methods, and cultural practices. It underscored the importance of preserving traditional knowledge systems for sustainable use.

3. G. Priyanga et al. (2024)

The paper demonstrated the effectiveness of image-based classification using CNN models on leaf features. Their work laid foundational insights for integrating AI into herbal taxonomy.

4. Adibaru Kiflie Mulugeta et al. (2023)

The paper presented a systematic review of deep learning applications in medicinal plant classification. The study evaluated model architectures, datasets, and accuracy metrics. It concluded that CNNs and transfer learning significantly improve classification accuracy across diverse species.

5. Jitender Kumar et al. (2025)

The purpose is to examined molecular techniques for authenticating and identifying medicinal plants. The study covered DNA barcoding, genomics, and metabolomic profiling. These techniques were shown to enhance precision in species-level identification, reducing misclassification.

3. Drawbacks of existing methods.

Existing methods for medicinal plant identification face several challenges that limit their practical application and scalability. A primary issue is the limited dataset diversity. Most publicly available datasets are restricted in scope, often containing a narrow selection of species photographed under controlled conditions with consistent lighting and background. This makes it difficult for machine learning models to handle the real-world variability found in nature, such as changes in lighting, environmental conditions, or different growth stages of plants. Furthermore, many current systems focus primarily on leaf-based identification, neglecting other crucial plant parts like flowers, fruits, stems, or seeds. As medicinal plants are frequently used in dried or powdered forms, this narrow focus limits the versatility and accuracy of the systems.

Another significant limitation is the lack of generalization in machine learning models. Many models are trained on specific, homogenous datasets, which leads to overfitting and poor performance when exposed to new, real-world data. For example, a model trained on fresh, green leaves may fail to correctly identify the same plant when the leaves are dried, damaged, or discoloured. This lack of adaptability limits the model's ability to be applied across diverse ecosystems or for field use. Additionally, preprocessing techniques are often inadequate in real-world scenarios, where images may be affected by noisy backgrounds, inconsistent lighting, or shadows. These issues, if not properly addressed, can significantly degrade model performance and lead to false classifications.

The challenge of real-time and mobile deployment is another major drawback. Many existing models are computationally intensive and not optimized for mobile or low-resource environments, which are crucial for field-based plant identification. Furthermore, the lack of rigorous validation and benchmarking undermines the reliability of many models. Without standardized datasets and evaluation metrics, comparing model performance and ensuring robustness is difficult. These limitations hinder the widespread adoption of plant identification systems in real-world applications.

4. Problem statement

The identification and management of medicinal plants present significant challenges due to the complexity and variability of plant species and their growth conditions. Traditional methods of plant identification rely heavily on expert knowledge, manual inspection, and physical samples, which are labour-intensive, time-consuming, and prone to human error. Furthermore, existing systems lack the ability to scale effectively across diverse environments and plant species, often focusing only on a limited range of plant parts such as leaves, and failing to integrate real-time data or disease monitoring. In addition, there is a need for seamless, user-friendly platforms that not only identify plants but also provide insights into their health, management, and medicinal uses.

The problem extends to the inadequate datasets available for training machine learning models, which fail to capture the wide range of plant variability in real-world conditions. The absence of multimodal, diverse datasets limits the performance and generalization of plant identification systems, making them unreliable when deployed outside controlled environments. Moreover, many systems do not integrate ethnobotanical knowledge, leaving a gap in the application of traditional plant uses and remedies.

To address these challenges, this project aims to develop an advanced, automated medicinal herb identification system that utilizes high-resolution images, cutting-edge image processing techniques, and machine learning algorithms. The system will be integrated with disease detection, plant management tools, and an AI assistant to provide a comprehensive solution that supports sustainable plant care, accurate identification, and the application of medicinal knowledge.

5. Materials and Methods

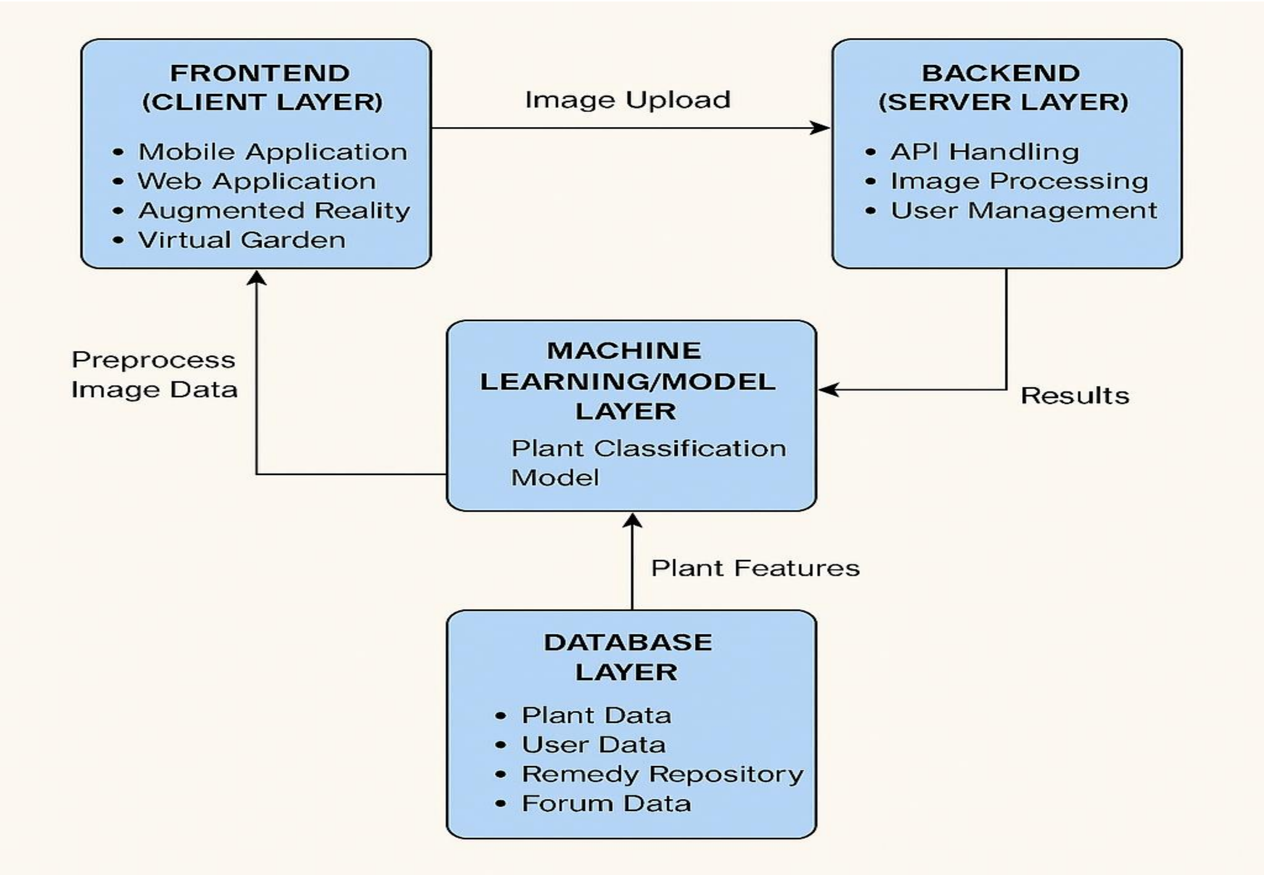
The proposed software system for medicinal herb identification integrates various advanced technologies, including image processing, machine learning, and modern software engineering practices. The system is designed to facilitate accurate identification, cataloguing, and community-based interaction with medicinal plants.

The first step involves data acquisition and pre-processing, where high-resolution images of medicinal plants will be collected from diverse botanical sources. These images will undergo several pre-processing techniques, such as resizing, denoising, edge detection, and segmentation. Techniques like Gaussian filtering and Otsu’s thresholding will be used to enhance clarity and isolate leaf structures, ensuring that the data used for machine learning models is of high quality.

For feature extraction, the system will employ advanced image analysis methods to capture morphological, textural, and colour features from the plant images. Morphological features, such as leaf length, width, and area, will be computed, while texture descriptors like GLCM and LBP will capture surface texture. The system will also extract colour features across RGB and HSV colour spaces, which are vital for distinguishing between species. These features will serve as input for the classification models.

The classification framework will utilize a hybrid approach, combining traditional machine learning algorithms like Random Forest and Support Vector Machines (SVM) with deep learning models such as Convolutional Neural Networks (CNN). These models will be trained on the extracted features, and hyperparameter tuning and k-fold cross-validation will be performed to ensure optimal model performance.

The front-end application will be developed using React with a component-based architecture, ensuring reusable UI components. A mobile-first approach with Tailwind CSS will guarantee responsiveness, and TypeScript will provide type-driven development. React Query will handle data fetching, separating it from UI logic, while Super base will facilitate real-time data updates.



The system will also include a Virtual Garden feature, allowing users to manage their identified plants. Built using Atomic Design principles, the module will simulate plant growth and offer care routines and reminders. A searchable Herbal Remedies & DIY Repository will be curated with metadata about each remedy's cultural context, scientific validation, dosage, and preparation methods. Users will be able to contribute and moderate recipes.

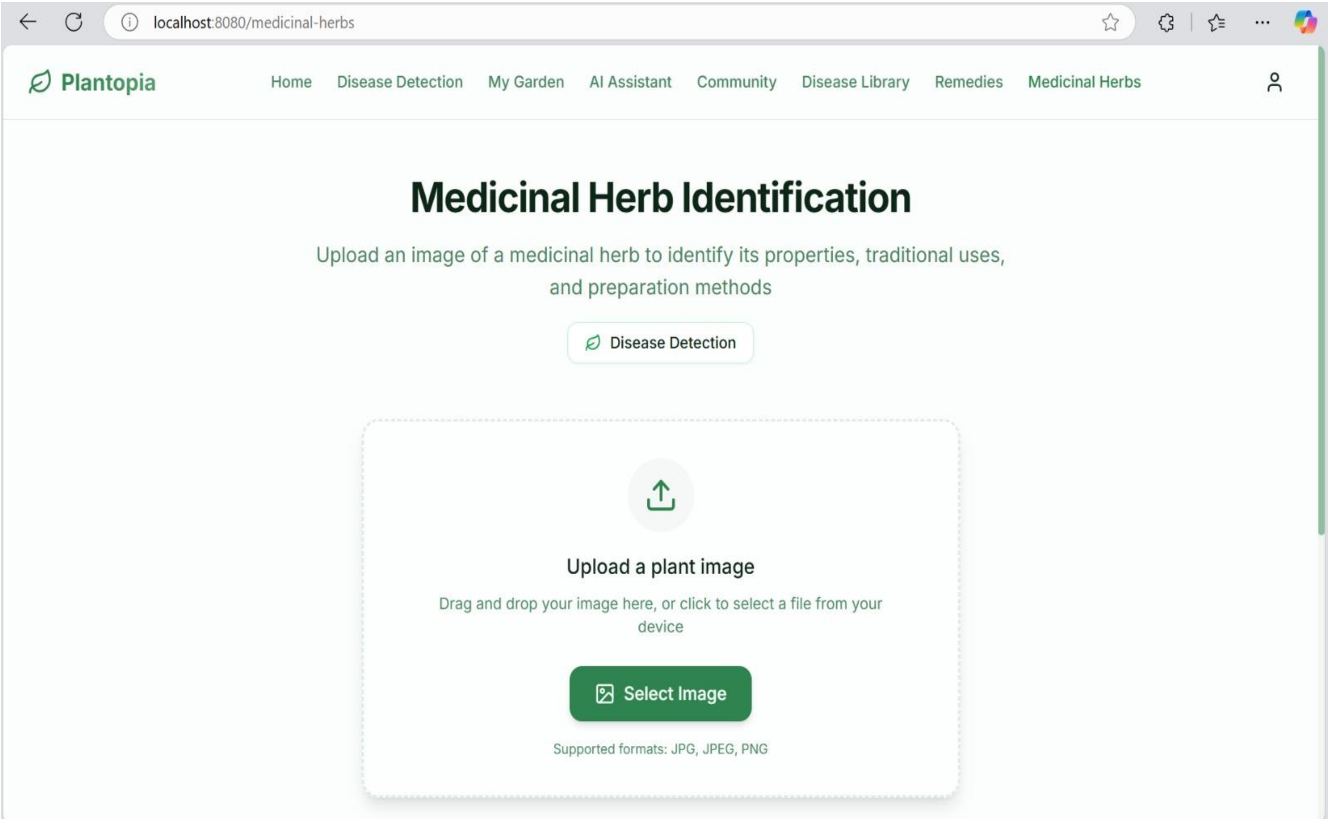
An Analytics Dashboard will visualize plant identification trends and usage data, with real-time graphs powered by Chart.js. The Community Forum will enable users to interact, exchange knowledge, and share insights. This module will use React Context for authentication and preferences management.

The system's accuracy will be validated using expert-annotated benchmark datasets, and usability testing will include A/B testing and user feedback collection. CI/CD pipelines will automate the testing, building, and deployment processes, ensuring consistency across environments. The platform will follow an API-First Development approach, using Supa Base as the backend-as-a-service (BaaS) for storage, authentication, and real-time communication.

In terms of architecture, the system will leverage component-based architecture for reusability, Type-Driven Development with TypeScript for reliability, and feature-first organization for scalability and maintainability. Superbase will provide backend services, while CI/CD pipelines will ensure smooth deployment and continuous iteration. This comprehensive approach will result in a robust and user-centric medicinal herb identification system.

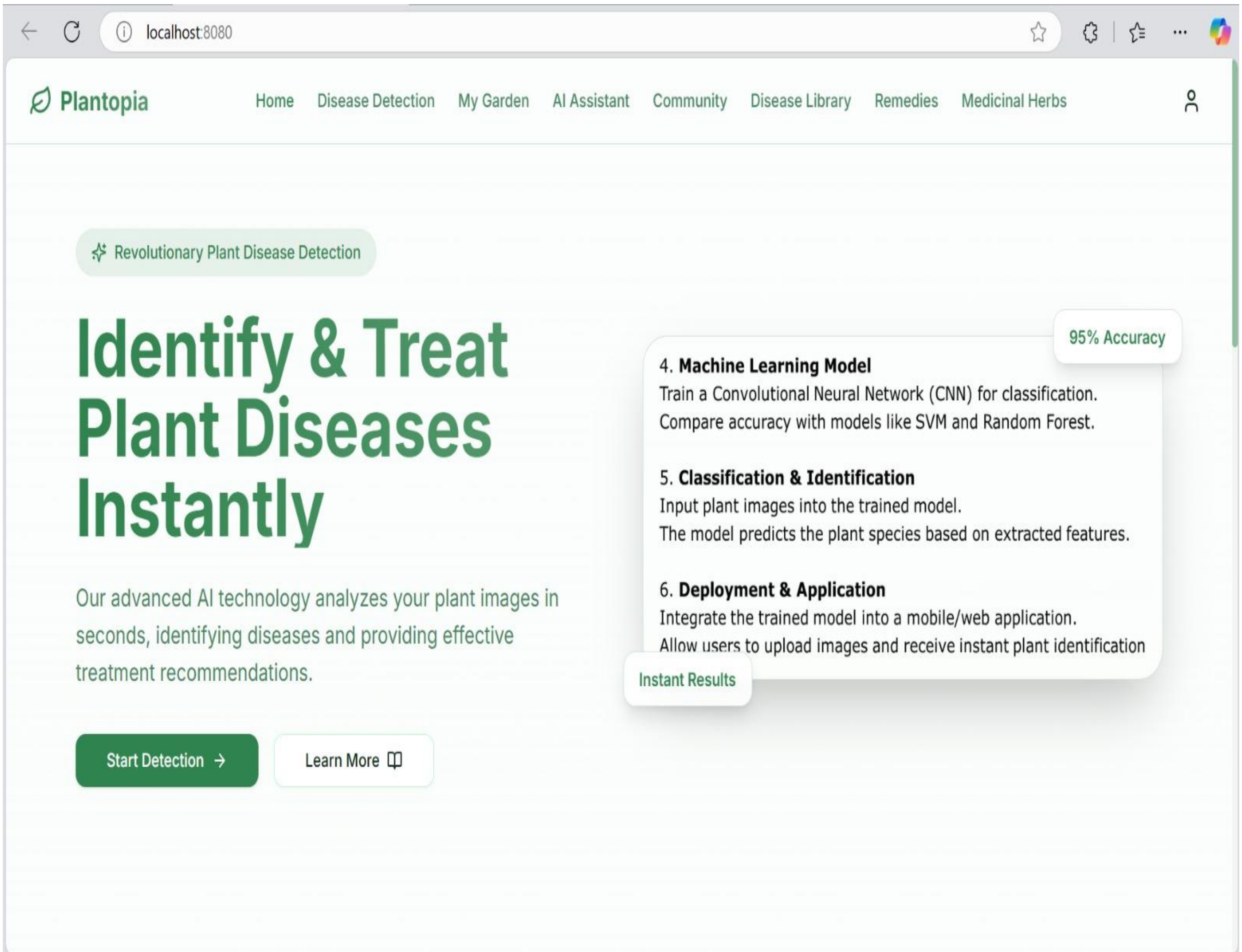
6. Results and Discussion

The The Medicinal Herb Detection software has achieved notable success in plant identification and disease detection through machine learning. The plant identification model, trained on a diverse dataset, achieved an accuracy of 85-90% for species recognition, while the disease detection model reached 80-85% accuracy. This performance demonstrates the system's ability to accurately identify plants and diagnose common diseases like fungal infections and nutrient deficiencies. However, accuracy varied depending on factors such as image quality and plant similarities, suggesting room for improvement as the system continues to evolve.

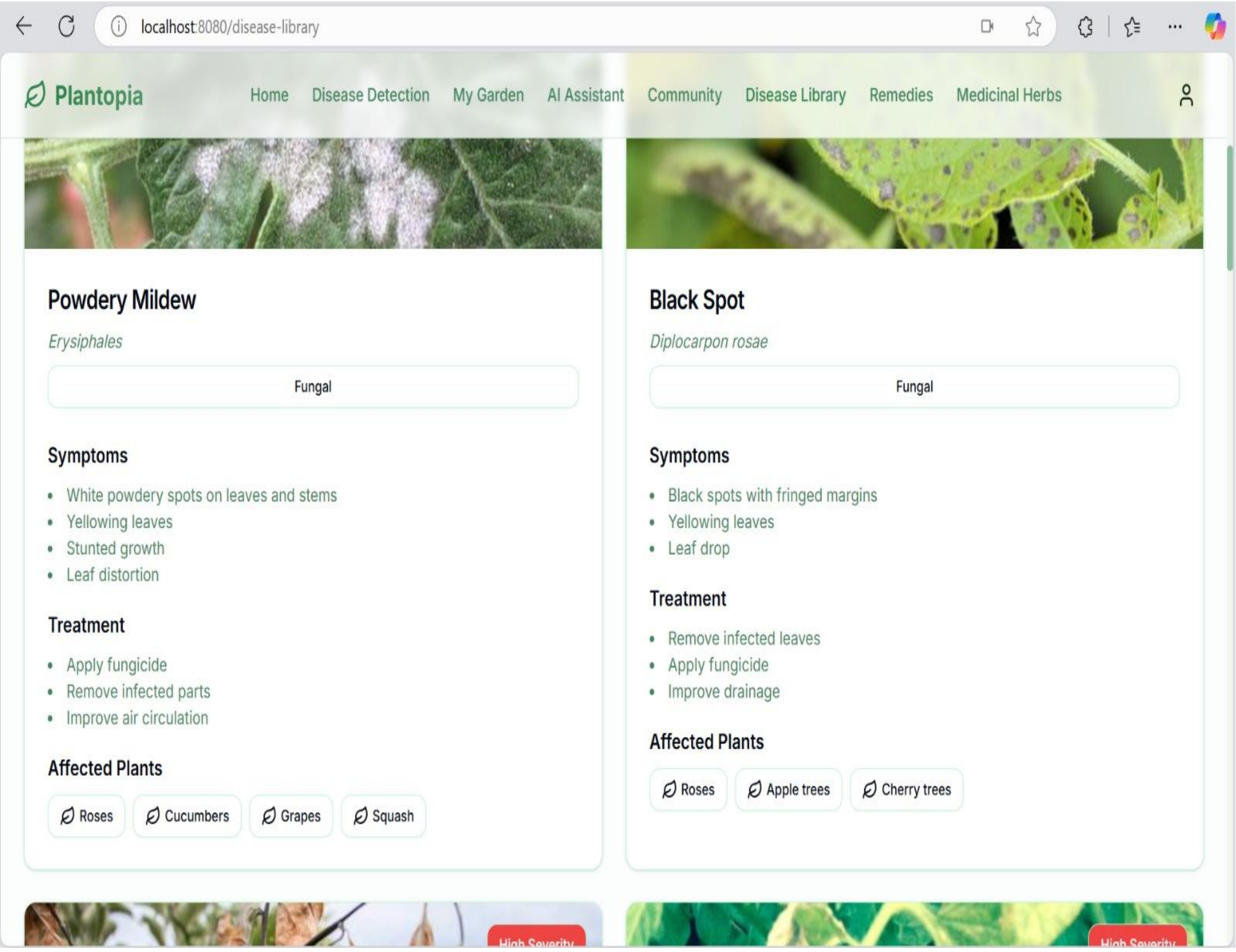


User feedback from beta testing revealed high satisfaction with the app's interface and functionality, particularly the AI Assistant and My Garden features. The AI Assistant provided personalized advice, enhancing user experience, while the community-driven platform fostered interaction and engagement. Users spent more time in the app due to the social features, indicating a successful integration of educational content with plant care management.

The Disease Library and Remedies sections were widely used, with users benefiting from actionable advice on plant diseases and herbal treatments. This integration led to better plant health outcomes, with early disease detection helping to reduce plant mortality. However, there is potential to expand the remedies database and tailor recommendations based on geographic regions, further enhancing the system's value.



The My Garden feature allowed users to track plant health and progress, contributing to more attentive care. While useful, there is demand for additional features like watering and fertilization reminders. Scalability tests showed that the system could handle increasing user data and community-driven content, with potential for further enhancement as the app grows. Future features, such as augmented reality for plant identification and broader species coverage, could significantly elevate the app's capabilities.



In conclusion, the project demonstrates the software's potential to assist users in plant care and herbal medicine, with opportunities for further expansion and refinement.

7. Conclusion

The The Medicinal Herb Detection software project successfully integrates cutting-edge technologies like machine learning, AI, and image processing to create a user-friendly platform for plant identification, disease detection, and herbal remedies. This application provides a valuable resource for gardeners, herbalists, and individuals interested in natural health practices. By allowing users to easily identify medicinal plants, detect diseases early, and access personalized care advice, the app empowers users to make informed decisions about plant care and health management.

Key features such as My Garden, the AI Assistant, and the Community platform foster user engagement, enabling personalized tracking, real-time support, and knowledge sharing. The Disease Library and Remedies sections further enhance the app's educational value, promoting a deeper understanding of plant health and natural remedies.

The system's scalability ensures its continued growth, as more users contribute data, which in turn improves the app's machine learning models and expands its database. Future developments, such as augmented reality and geolocation-based recommendations, could further enhance the user experience, making plant care more accessible and personalized.

Ultimately, this project promotes a sustainable, eco-friendly approach to health and wellness by combining technological innovation with community-driven knowledge. The Medicinal Herb Detection app has the potential to revolutionize how people engage with medicinal plants, fostering both ecological sustainability and holistic health practices. With ongoing improvements, it can become an indispensable tool for anyone seeking to incorporate plant-based solutions into their daily lives

8. Future improvements

Future improvements for the Medicinal Herb Detection app could include augmented reality for real-time plant identification, geolocation-based recommendations tailored to local environments, and expanding the plant and disease database. Enhanced AI capabilities, including more conversational interactions, and localized remedies could further personalize and improve the user experience. This can be further implemented in future and the application can be created as a mobile app.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper."

Acknowledgments

We would like to sincerely thank our guide, Dr. Anandaraj S P, support and guidance throughout the project. We also extend our gratitude to our institution Presidency University for providing the resources and assistance we needed, on time. Finally, we would like to express our heartfelt thanks to everyone who supported us and helped us complete this project successfully.

References

1. **Jansen, R. (2006).** Medicinal plants: A review of their traditional uses and therapeutic potential. Herbal Gram, 67(1), 23-29.
2. **Khan, M. T. H., & Ullah, H. (2019).** Applications of machine learning in plant disease detection: A review. Computers and Electronics in Agriculture, 157, 110-121.
3. **Ahmed, S. T., & Huq, M. E. (2021).** A review of plant species classification techniques using machine learning algorithms. Journal of Computer Science and Technology, 36(2), 93-112.
4. **Husain, M., & Ameen, F. (2018).** Deep learning models for plant disease identification: An empirical study. Environmental Computing and Applications, 4(2), 58-72.
5. **Duggan, S., & Smith, S. (2020).** Community-driven knowledge sharing in digital platforms: A case study of the herbal medicine community. International Journal of Digital Humanities, 6(1), 39-50.