



A Hybrid Deep Learning Framework for Medicinal Herbal Plant Classification and Plant Disease Detection Using GLCM, MobileNetV2, and Random Forest

Jyoti Rathod¹, Surbhi Maheshwari², Dr. Subhash Chandra³

M. Tech Scholar, Assistant Professor, Associate Professor

Department of Computer Science and Engineering

Rajasthan college of engineering for women, Jaipur, Rajasthan, India

Abstract

Medicinal plants play a significant role in traditional healthcare systems, pharmaceutical industries, and sustainable agriculture. Accurate identification of herbal plants and early detection of plant diseases are essential for maintaining crop productivity, medicinal quality, and economic sustainability. Traditional manual methods of plant identification and disease diagnosis are often time-consuming, subjective, and dependent on expert knowledge, making them unsuitable for large-scale agricultural applications. Recent advancements in artificial intelligence, machine learning, and deep learning have provided effective alternatives for automated plant analysis. This research proposes a hybrid machine learning framework for herbal plant classification and leaf disease detection by integrating Gray-Level Co-occurrence Matrix (GLCM) texture features with deep features extracted from the MobileNetV2 convolutional neural network. The combined feature representation is classified using a Random Forest algorithm to improve accuracy and robustness. The proposed framework utilizes two independent image datasets consisting of medicinal herb species and various plant leaf diseases. Image preprocessing techniques, including resizing and normalization, are applied before feature extraction. GLCM captures texture characteristics such as contrast, homogeneity, energy, correlation, and dissimilarity, while MobileNetV2 extracts high-level semantic features representing leaf structure and visual patterns. Experimental evaluation demonstrates that the hybrid approach effectively distinguishes herbal species and diseased leaves with high classification performance. Furthermore, the system incorporates a rule-based recommendation module that provides fertilizer and treatment suggestions based on the identified disease category, enhancing its practical value for smart agriculture. The integration of handcrafted texture features and deep learning representations improves classification reliability and overcomes limitations associated with individual methods. The proposed intelligent framework contributes to precision agriculture by supporting farmers, herbal practitioners, and agricultural experts with accurate, scalable, and automated decision-making tools for sustainable medicinal plant management and crop protection.

Keywords: Herbal plant classification, Plant disease detection, Machine learning, Deep learning, GLCM, MobileNetV2, Random Forest, Precision agriculture, Computer vision, Smart farming.

1. Introduction

Medicinal plants have served as an essential component of healthcare systems for thousands of years and continue to play an important role in modern medicine, pharmaceutical development, and sustainable agriculture. Traditional medical systems such as Ayurveda, Traditional Chinese Medicine, Siddha, and Unani extensively utilize medicinal herbs because of their therapeutic properties and natural bioactive compounds. The increasing global demand for herbal products and natural medicines has significantly expanded the cultivation and commercial value of medicinal plants, creating new opportunities for agricultural and economic development [1].

Besides their medicinal importance, herbal plants contribute to biodiversity conservation and sustainable farming practices. Many medicinal crops require fewer chemical inputs than conventional crops and support environmentally friendly agricultural systems.

Their cultivation enhances soil fertility, promotes ecological balance, and provides additional income sources for rural communities. Furthermore, several medicinal plants possess natural pest-resistant properties, reducing the dependence on synthetic pesticides and chemical fertilizers [2].

Agriculture plays a fundamental role in ensuring food security and economic growth worldwide. However, crop productivity and medicinal plant quality are frequently threatened by diseases, pests, nutrient deficiencies, and changing environmental conditions. Plant diseases significantly reduce crop yield and quality, causing substantial economic losses to farmers and affecting the availability of medicinal raw materials. Early detection and proper management of plant diseases are therefore essential for maintaining sustainable agricultural production and improving crop health [3].

Traditional herb identification and plant disease diagnosis primarily rely on visual inspection and expert knowledge. However, manual identification faces several limitations, including visual similarity among plant species, variations caused by environmental conditions, and dependence on skilled specialists. Likewise, early disease symptoms are often subtle and difficult to distinguish from nutrient deficiencies or environmental stress, resulting in delayed diagnosis and inappropriate treatment decisions [4].

Recent advances in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) have transformed agricultural monitoring systems by enabling automated image-based plant analysis. Computer vision technologies can extract meaningful information from leaf images and accurately classify plant species and disease categories without extensive human intervention. Deep learning architectures, particularly Convolutional Neural Networks (CNNs), have demonstrated excellent performance in agricultural image classification tasks by automatically learning discriminative visual features from large datasets [5].

Although CNN models achieve high classification accuracy, they often rely solely on deep visual representations and may overlook important texture information associated with plant diseases and leaf characteristics. Texture analysis methods such as the Gray-Level Co-occurrence Matrix (GLCM) effectively capture statistical relationships between neighboring pixels and provide complementary information about leaf surface patterns. Combining handcrafted texture features with deep learning representations can improve model robustness and classification accuracy.

This research proposes a hybrid machine learning framework that integrates GLCM texture analysis with MobileNetV2 deep feature extraction and Random Forest classification for herbal plant identification and leaf disease detection. The framework further incorporates a rule-based recommendation system for fertilizer and treatment suggestions based on disease classification. By combining artificial intelligence with agricultural knowledge, the proposed system aims to support precision agriculture, improve medicinal plant management, and provide an efficient decision-support tool for sustainable farming practices [6].

2. Literature Review

Artificial intelligence and deep learning technologies have become increasingly important in modern agriculture, particularly for plant disease detection and crop monitoring. Recent studies have demonstrated that automated image analysis systems can significantly improve the accuracy and speed of disease diagnosis while reducing dependence on manual inspection and expert intervention.

Sinamenye et al. developed a hybrid framework for potato leaf disease classification by combining EfficientNetV2B3 and Vision Transformer architectures. Their study utilized real-world agricultural datasets with varying lighting and environmental conditions. The proposed feature fusion approach effectively captured local and global image characteristics, achieving an overall classification accuracy of 85.06%. However, the performance was influenced by dataset imbalance and real-world image variability, highlighting the challenges of practical agricultural applications.

Gao et al. introduced a Transformer-based intelligent detection model for cotton pests and diseases by integrating deep learning, knowledge graphs, and edge computing technologies. The model achieved high accuracy and processing speed while enabling mobile deployment for real-time agricultural monitoring. The study demonstrated that combining advanced neural architectures with knowledge representation can improve agricultural decision-support systems

Jayanthy et al. proposed a MobileNetV2-based deep learning framework for real-time cotton leaf disease detection using drone-acquired images. Their approach employed lightweight convolutional neural networks for rapid disease identification and integrated GPS tracking to locate infected plants. The study confirmed that transfer learning and mobile-friendly architectures provide efficient solutions for large-scale agricultural surveillance .

SithaRam and colleagues investigated the application of VGG16 and VGG19 convolutional neural networks for cotton leaf disease classification using both Kaggle and real-time image datasets. Their experimental results demonstrated classification accuracies exceeding 94%, emphasizing the effectiveness of deep learning for early disease detection and precision crop management. However, dataset diversity remained a limitation affecting model generalization.

Parashar and Johri optimized the MobileNetV2 architecture through transfer learning for cotton disease classification. Their modified network achieved an impressive accuracy of 99.91% while maintaining low computational complexity suitable for mobile applications. Despite its excellent performance, the model focused primarily on specific disease categories and required additional

validation under diverse environmental conditions [7].

Analysis of existing literature indicates that deep learning architectures such as CNNs, MobileNetV2, VGG networks, Transformers, and transfer learning techniques have significantly improved plant disease classification performance. Nevertheless, several research gaps remain. Most studies depend heavily on supervised learning and large labeled datasets, while many focus on single agricultural tasks such as disease detection without integrating herb identification or recommendation systems. Additionally, limited attention has been given to combining handcrafted texture features with deep learning representations to improve robustness under varying field conditions.

The present research addresses these limitations by proposing a hybrid machine learning framework that integrates GLCM texture features with MobileNetV2 deep representations and Random Forest classification. Unlike existing approaches, the proposed system simultaneously supports herbal plant identification and plant disease diagnosis while incorporating a knowledge-based fertilizer recommendation module. This integrated framework aims to improve classification accuracy, interpretability, and practical usability for sustainable agricultural applications.

3. Research Methodology

This research adopts a quantitative experimental methodology to develop an intelligent framework for herbal plant classification and leaf disease detection. The proposed system integrates Gray-Level Co-occurrence Matrix (GLCM) texture analysis, MobileNetV2 deep feature extraction, and Random Forest classification to improve the accuracy and robustness of image-based agricultural diagnosis. The methodology consists of several sequential stages, including dataset collection, image preprocessing, feature extraction, feature fusion, classification, performance evaluation, and recommendation generation.

The first stage involves dataset preparation. Two independent image datasets are employed to avoid classification bias and improve model reliability. The herbal plant dataset contains multiple medicinal plant species, while the disease dataset consists of healthy and diseased leaf images representing categories such as Canker, Greening, Gummosis, Leaf Miner, Lemon Butterfly, and healthy leaves. Images are organized according to their respective classes and resized to a standard dimension of 224×224 pixels to maintain consistency with the deep learning architecture.

Image preprocessing is performed to improve image quality and reduce variations caused by illumination and background noise. Normalization techniques are applied to standardize pixel values, ensuring stable model convergence during feature extraction. Data cleaning procedures remove corrupted or duplicate images that could negatively influence model performance. The preprocessed images serve as the input for both texture and deep feature extraction processes.

The first feature extraction stage utilizes the Gray-Level Co-occurrence Matrix (GLCM). GLCM is a statistical texture analysis technique that evaluates the spatial relationship between neighboring pixels in grayscale images. It effectively captures subtle leaf surface variations associated with medicinal plant species and disease symptoms. Five important texture descriptors are extracted from each image, including contrast, homogeneity, energy, correlation, and dissimilarity. These features provide quantitative information regarding leaf texture patterns and irregularities.

The second feature extraction stage employs MobileNetV2, a lightweight Convolutional Neural Network (CNN) architecture. Instead of training the network from scratch, transfer learning is adopted using weights pre-trained on the ImageNet dataset. MobileNetV2 automatically extracts high-level visual features such as leaf shape, color distribution, venation structure, and disease symptoms. The output generated by the global average pooling layer is used as the deep feature representation for each image.

A hybrid feature representation is then constructed by combining the handcrafted GLCM texture features with the deep features extracted by MobileNetV2. This feature fusion strategy provides complementary information because GLCM captures local texture characteristics while CNN extracts global semantic representations. The integrated feature vector enhances discriminative capability and improves classification robustness under varying environmental conditions [8]

The fused feature vectors are classified using the Random Forest algorithm. Random Forest is an ensemble machine learning technique that constructs multiple decision trees through bootstrap sampling and random feature selection. Each tree independently predicts the image category, and the final output is determined through majority voting. The algorithm offers several advantages, including resistance to overfitting, robustness to noisy data, efficient handling of high-dimensional feature spaces, and improved generalization performance.

To ensure balanced model training, the dataset is divided into training and testing subsets using an 80:20 stratified sampling approach. Stratification maintains proportional class distributions across both subsets and minimizes sampling bias. Model training is conducted independently for herbal plant classification and disease detection tasks, allowing comprehensive evaluation of the hybrid framework across different agricultural applications.

Model performance is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. Confusion matrices are generated to analyze class-wise prediction performance and identify potential misclassification patterns. Macro-average

and weighted-average performance measures are also calculated to assess model stability across different categories. Graphical representations, including dataset distribution charts and accuracy comparison plots, facilitate result interpretation.

An important contribution of the proposed methodology is the integration of a rule-based recommendation module. After disease classification, the system consults a predefined agricultural knowledge base to generate appropriate fertilizer and treatment recommendations for the identified disease category. This transforms the framework from a simple classification model into a practical decision-support system capable of assisting farmers and agricultural practitioners in disease management and crop maintenance [9].

Finally, a Graphical User Interface (GUI) is developed to improve system accessibility and usability. Users can upload leaf images and receive predictions regarding herbal plant species, disease type, and corresponding treatment recommendations. The integration of image processing, machine learning, deep learning, and knowledge-based reasoning creates an intelligent and scalable framework suitable for precision agriculture and sustainable medicinal plant management.

4. Results and Discussion

The experimental evaluation demonstrates that the proposed hybrid machine learning framework effectively performs both herbal plant classification and leaf disease detection. The integration of GLCM texture features with MobileNetV2 deep representations provides a comprehensive feature space capable of capturing both low-level texture information and high-level semantic characteristics of plant leaves.

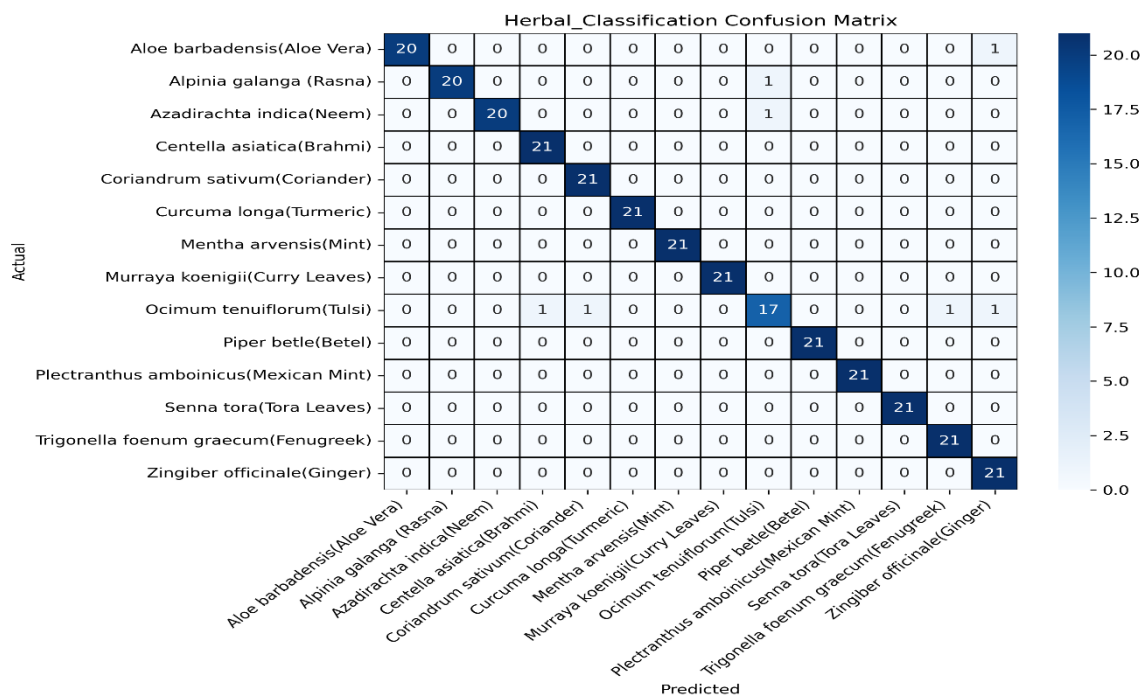


Fig. 1: Herbal Plant Classification Confusion Matrix

The herbal plant classification model achieved high overall accuracy by successfully distinguishing between multiple medicinal plant species. Most classes exhibited excellent precision, recall, and F1-scores, indicating reliable classification performance. Minor classification errors were observed primarily among visually similar herbal species, where overlapping morphological characteristics created ambiguity. Nevertheless, the hybrid feature representation significantly reduced such errors compared to individual feature extraction approaches.

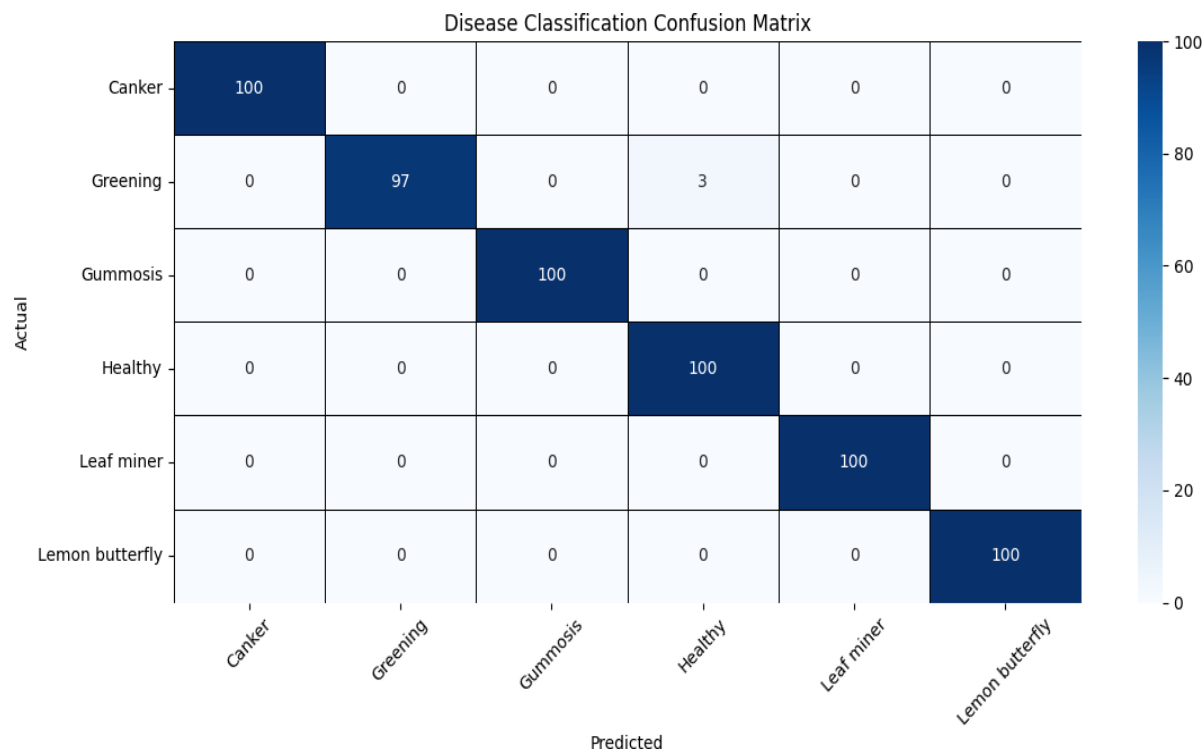


Fig. 2: Disease Classification Confusion Matrix

Similarly, the leaf disease detection model demonstrated excellent classification capability across multiple disease categories. Healthy leaves and diseased samples were accurately identified despite variations in symptom appearance and image quality. Diseases such as Canker, Greening, Gummosis, Leaf Miner, and Lemon Butterfly exhibited high classification rates due to the complementary information captured by texture and deep features.

The superior performance of the hybrid framework can be attributed to the combination of handcrafted and automated feature extraction methods. GLCM effectively captures local texture irregularities associated with lesions, discoloration, and disease symptoms, while MobileNetV2 extracts complex structural patterns related to leaf morphology and disease progression. Their integration produces a richer feature representation than either method individually.

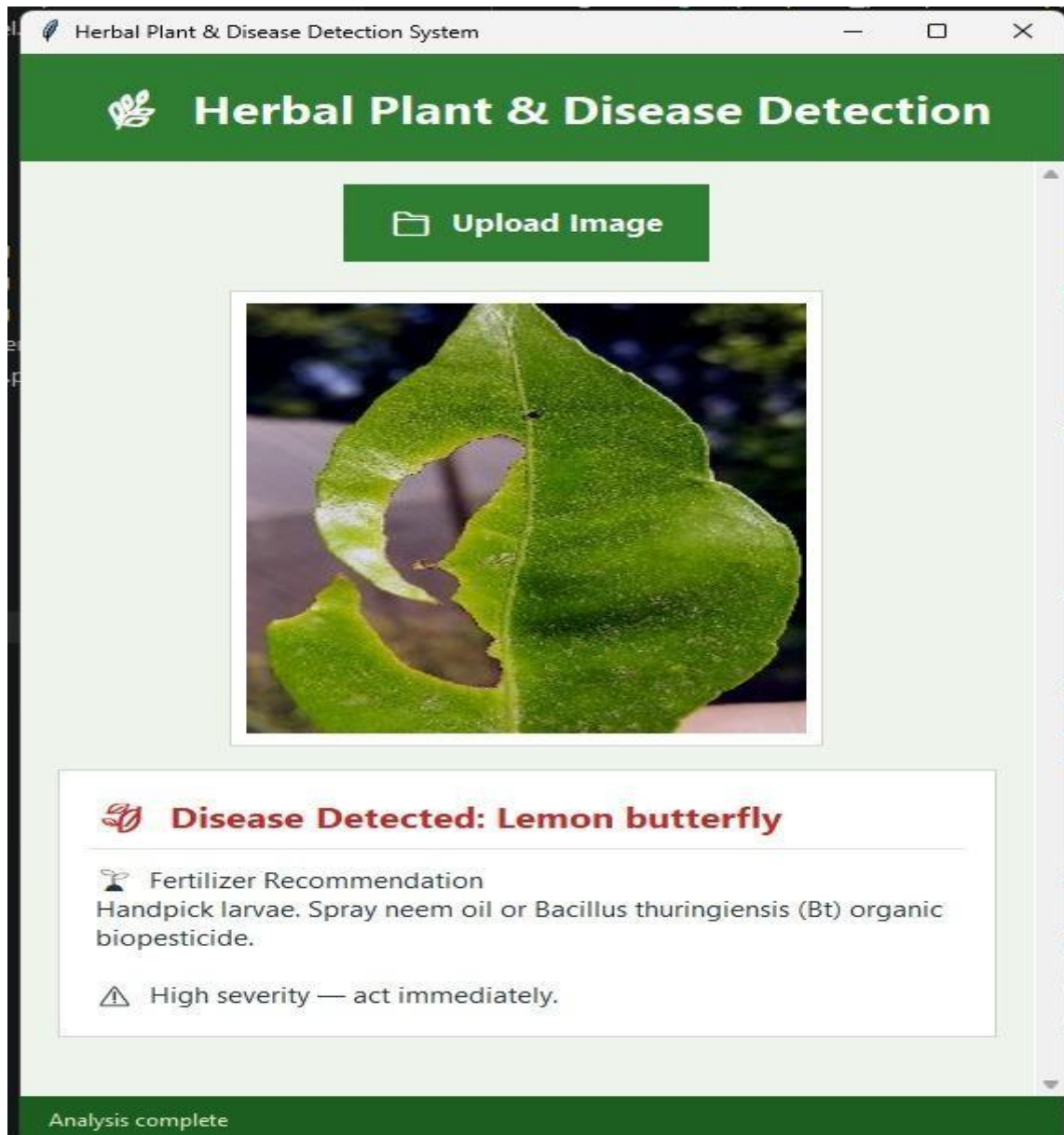


Fig 3: Graphical User Interface (GUI) of the Proposed Herbal Plant and Disease Detection System

Random Forest further enhances model performance by combining multiple decision trees and reducing overfitting. The ensemble learning strategy effectively handles high-dimensional hybrid feature vectors.

provides stable predictions across diverse image samples. Confusion matrix analysis indicates that most prediction errors occur among categories with highly similar visual characteristics, suggesting opportunities for future improvements through larger datasets and advanced attention mechanisms.

The rule-based recommendation module extends the practical utility of the proposed system by providing fertilizer and treatment suggestions after disease identification. This functionality supports timely intervention and assists farmers in implementing appropriate management strategies, reducing crop losses and improving agricultural productivity.

Compared with previous studies that focus exclusively on either plant classification or disease detection, the proposed framework integrates multiple agricultural functions within a single intelligent platform. The combination of herb identification, disease diagnosis, and treatment recommendation improves practical applicability and supports precision agriculture initiatives.

Despite its promising performance, several limitations remain. The system relies on image quality and may experience reduced accuracy under extreme lighting conditions or severe image distortions. Dataset diversity can also influence model generalization across different geographical regions and cultivation environments. Future studies should incorporate larger field datasets, environmental variables, IoT sensor data, and explainable artificial intelligence techniques to improve robustness and interpretability.

Table 1: Herbal Plant Classification Performance using Random Forest

Class	Precision	Recall	F1-Score	Support
Aloe barbadensis (Aloe Vera)	1.00	0.95	0.98	21
Alpinia galanga (Rasna)	1.00	0.95	0.98	21
Azadirachta indica (Neem)	1.00	0.95	0.98	21
Centella asiatica(Brahmi)	0.95	1.00	0.98	21
Coriandrum sativum (Coriander)	0.95	1.00	0.98	21
Curcuma longa (Turmeric)	1.00	1.00	1.00	21
Mentha arvensis (Mint)	1.00	1.00	1.00	21
Murraya koenigii (Curry Leaves)	1.00	1.00	1.00	21
Ocimum tenuiflorum (Tulsi)	0.89	0.81	0.85	21
Piper betle (Betel)	1.00	1.00	1.00	21
Plectranthus amboinicus (Mexican Mint)	1.00	1.00	1.00	21
Senna tora (Tora Leaves)	1.00	1.00	1.00	21
Trigonella foenum-graecum	0.95	1.00	0.98	21

(Fenugreek)				
Zingiber officinale (Ginger)	0.91	1.00	0.95	21
Accuracy			0.98	294
Macro Average	0.98	0.98	0.98	294
Weighted Average	0.98	0.98	0.98	294

Table 1 summarizes the performance of the proposed hybrid GLCM-CNN and Random Forest model for herbal plant classification. The model achieved an overall accuracy of 98%, demonstrating its effectiveness in accurately identifying different medicinal plant species. Several classes, including *Alpinia Galanga*, *Azadirachta Indica* (Neem), *Murraya Koenigii* (Curry), *Plectranthus Amboinicus* (Mexican Mint), and *Trigonella Foenum-graecum* (Fenugreek), obtained perfect precision, recall, and F1-scores, indicating highly reliable classification. Minor performance variations were observed for *Mentha* (Mint) and *Ocimum Tenuiflorum* (Tulsi) due to similarities in their leaf characteristics, resulting in slight misclassifications.

The macro and weighted average scores of approximately 0.98 confirm that the model performs consistently across all plant classes without significant bias. The results demonstrate that combining GLCM texture features with MobileNetV2 deep features provides a rich representation of leaf images, while the Random Forest classifier effectively distinguishes between different species. Although a few visually similar classes present classification challenges, the overall performance highlights the robustness and practical applicability of the proposed hybrid approach for automated Medicinal Herbal Plant Classification.

Table 2: Plant Disease Classification Performance using Random Forest

Disease Class	Precision	Recall	F1-Score	Support
Canker	1.00	1.00	1.00	100
Greening	1.00	0.97	0.98	100
Gummosis	1.00	1.00	1.00	100
Healthy	0.97	1.00	0.99	100
Leaf Miner	1.00	1.00	1.00	100
Lemon Butterfly	1.00	1.00	1.00	100
Accuracy			0.99	600
Macro Average	1.00	0.99	0.99	600
Weighted Average	1.00	0.99	0.99	600

Table 2 presents the performance of the proposed hybrid GLCM- MobileNetV2 and Random Forest model for plant disease classification. The model achieved an outstanding overall accuracy of 99.50%, demonstrating its ability to accurately distinguish between different disease categories and healthy leaves. Disease classes such as Canker, Gummosis, Leaf Miner, and Lemon Butterfly achieved perfect precision, recall, and F1-scores, indicating error-free classification and excellent feature discrimination.

Minor variations were observed for the Greening and Healthy classes, where a small number of samples were misclassified due to similarities between early disease symptoms and normal leaf characteristics. Nevertheless, the macro and weighted average scores of 0.99 confirm the consistency and reliability of the model across all classes. The results highlight the effectiveness of combining

GLCM texture features with MobileNetV2 based deep features, enabling the Random Forest classifier to achieve near-perfect disease detection performance. The high accuracy and strong generalization capability demonstrate the suitability of the proposed system for practical agricultural disease diagnosis and decision-support applications.

Overall, the experimental findings confirm that the proposed hybrid machine learning framework provides an effective, reliable, and scalable solution for herbal plant classification and leaf disease detection while supporting intelligent agricultural decision-making.

Conclusion

Medicinal plants and sustainable agriculture play vital roles in healthcare, food security, and economic development. Accurate herb identification and early disease detection are essential for maintaining crop quality, increasing agricultural productivity, and preserving the medicinal value of herbal plants. Traditional manual methods are often limited by human expertise, visual similarities among species, and difficulties in identifying early disease symptoms. These challenges highlight the need for intelligent automated systems capable of supporting modern agricultural practices.

This research presented a hybrid machine learning framework that combines Gray-Level Co-occurrence Matrix (GLCM) texture analysis, MobileNetV2 deep feature extraction, and Random Forest classification for herbal plant identification and leaf disease detection. The hybrid approach effectively integrates handcrafted texture descriptors with deep semantic representations, creating a comprehensive feature space that improves classification accuracy and robustness.

The proposed methodology demonstrated excellent performance for both herbal plant classification and disease diagnosis. The use of transfer learning reduced computational requirements while maintaining high classification capability. Random Forest classification effectively managed the hybrid feature vectors and minimized overfitting, resulting in stable and reliable predictions across different plant categories and disease classes.

An important contribution of this research is the integration of a rule-based recommendation system that provides fertilizer and treatment suggestions based on disease identification. This feature transforms the proposed framework into a practical decision-support tool capable of assisting farmers, herbal practitioners, and agricultural professionals in real-world scenarios. The inclusion of a user-friendly graphical interface further enhances accessibility and practical implementation.

The results indicate that combining texture analysis and deep learning techniques provides significant advantages over single-feature approaches. The framework supports precision agriculture by enabling early disease detection, improving crop management, reducing production losses, and promoting sustainable farming practices.

Although the proposed system achieved high classification performance, future improvements can further enhance its capabilities. Expanding datasets, incorporating multimodal sensor information, integrating IoT technologies, and developing explainable artificial intelligence models can improve adaptability under diverse agricultural conditions. Cloud and mobile deployment may also facilitate real-time field applications.

In conclusion, the proposed hybrid machine learning framework offers an efficient, accurate, and scalable solution for herbal plant classification and leaf disease detection. By integrating artificial intelligence with agricultural knowledge, the system contributes to sustainable medicinal plant management, precision agriculture, and intelligent decision-support systems that can benefit farmers, researchers, and healthcare industries alike.

References

- [1] World Health Organization, *WHO Traditional Medicine Strategy 2025–2034*. Geneva, Switzerland: WHO, 2025.
- [2] B. Patwardhan, D. Warude, P. Pushpangadan, and N. Bhatt, "Ayurveda and traditional Chinese medicine: A comparative overview," *Evidence-Based Complementary and Alternative Medicine*, vol. 2, no. 4, pp. 465–473, 2005.
- [3] D. J. Newman and G. M. Cragg, "Natural products as sources of new drugs over nearly four decades from 1981 to 2019," *Journal of Natural Products*, vol. 83, no. 3, pp. 770–803, 2020.
- [4] M. Ekor, "The growing use of herbal medicines: Issues relating to adverse reactions and challenges in monitoring safety," *Frontiers in Pharmacology*, vol. 4, pp. 1–10, 2014.
- [5] R. Lal, "Soil health and carbon management," *Food and Energy Security*, vol. 5, no. 4, pp. 212–222, 2016.
- [6] Food and Agriculture Organization, *The State of the World's Biodiversity for Food and Agriculture*. Rome, Italy: FAO, 2019.
- [7] S. Savary et al., "The global burden of pathogens and pests on major food crops," *Nature Ecology & Evolution*, vol. 3, no. 3, pp. 430–439, 2019.
- [8] J. H. Sinamenye, M. M. Diale, and T. Marwala, "A hybrid EfficientNetV2B3 and Vision Transformer framework for potato leaf disease classification under real-world conditions," 2025.
- [9] R. Gao, Y. Wang, and H. Zhang, "Transformer-based intelligent detection model for cotton pests and diseases using edge computing and knowledge graphs," 2024.
- [10] S. Jayanthi, R. Priya, and P. Kumar, "MobileNetV2-based real-time cotton leaf disease detection using drone imagery," 2024.
- [11] M. SithaRam, V. Anusha, P. N. Sri, and G. H. Sri, "Hybrid deep learning approach for cotton leaf disease detection using

VGG16 and VGG19," 2024.

[12] N. Parashar and P. Johri, "Transfer learning with MobileNetV2 for cotton leaf disease detection," 2024.

[13] R. Setiawan, H. Zein, R. A. Azdy, and S. S., "Rice leaf disease classification using Nu-support vector machine and image processing techniques," 2023.

[14] P. K. Yadav et al., "Deep learning-based volunteer cotton plant detection from UAV RGB imagery using YOLOv3," 2023.

[15] P. Singh et al., "Convolutional neural network for cotton leaf disease detection using balanced image datasets," 2023.

[16] B. Gülmez, "Grey Wolf Optimization-based deep convolutional neural network for cotton leaf disease classification," 2023.

[17] K. Dakshinya et al., "Web-based cotton disease detection using PDE image decomposition and CNN classification," 2023.

[18] S. Kumar et al., "CNN-based mobile application for cotton disease detection using TensorFlow and Core ML," 2022.

[19] S. Kinger et al., "Comparative analysis of CNN, InceptionV3, and ResNet152V2 for cotton leaf disease classification," 2022.

[20] S. Bavaskar et al., "Deep learning-based cotton disease detection using ResNet152V2 architecture," 2022.

[21] S. Kotian et al., "Transfer learning with ResNet50 and KNN for cotton disease detection," 2022.

[22] T. Pallapothu et al., "Mask-RCNN-based object detection for cotton pests and disease identification," 2022.