Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 33 No. 1s (2026)

A Hybrid CNN–Vision Transformer Framework with Attention-Based Feature Optimization for Medicinal Leaf Classification

1. Ch. Yamini
Assistant Professor
Artificial Intelligence and Data
Science Department
Vignan Institute of Technology and Science
Hyderabad, India
yaminich27@gmail.com

2. Ravula Abhinav Reddy
Artificial Intelligence and Data
Science Department
Vignan Institute of Technology and Science
Hyderabad, India
ravula.abhinavreddy@gmail.com

3. Ganji Bharath
Artificial Intelligence and Data
Science Department
Vignan Institute of Technology and Science
Hyderabad, India
bharathganji345@gmail.com

4. Kanchukommula Akshaya
Artificial Intelligence and Data
Science Department
Vignan Institute of Technology and Science
Hyderabad, India
akshayakanchukommula@gmail.com

Article History:
Received: 04-02-2026
Revised: 20-03-2026
Accepted: 10-04-2026

Abstract:

Accurate identification of medicinal plants plays a crucial role in healthcare, herbal medicine, and biodiversity conservation. Traditional plant identification methods rely heavily on expert knowledge and manual inspection, making them time-consuming and error-prone. Recent advancements in deep learning have demonstrated promising results in image-based plant classification; however, existing approaches often suffer from high computational complexity, limited robustness, and poor generalization. This work presents a hybrid medicinal leaf classification framework that integrates convolutional neural networks (CNNs) and Vision Transformers (ViTs) to exploit both local and global feature representations. Initially, an image quality assessment module filters and enhances input leaf images based on blur and illumination conditions. A lightweight CNN is employed to extract local spatial features, while a ViT encoder captures global contextual information. The extracted features are fused and optimized using an attention-based feature selection mechanism to emphasize discriminative attributes. Finally, MobileNet and DenseNet models are used for accurate classification of about 92% medicinal plant species. Experimental results on a publicly available medicinal leaf dataset

demonstrate improved classification performance and robustness compared to conventional deep learning models. The proposed system provides an effective and scalable solution for automated medicinal plant identification.
Keywords: Medicinal Plant Classification, Convolutional Neural Network, Vision Transformer, MobileNet, DenseNet, Feature Fusion, Attention Mechanism, Image Quality Assessment.

1. INTRODUCTION

Medicinal plants play an important role in traditional medicine, modern healthcare, and biodiversity conservation because of their natural healing properties. However, identifying medicinal plant species accurately can be difficult since many plants have similar visual characteristics. Traditional identification methods rely on manual observation and expert knowledge, which can be time-consuming and prone to human error. With the increasing global demand for herbal medicines, there is a strong need for automated systems that can identify medicinal plants more efficiently and reliably.

Advancements in artificial intelligence and deep learning have improved image-based plant classification by enabling models such as Convolutional Neural Networks (CNNs) to automatically extract complex visual patterns like leaf texture, color, and vein structure. This study proposes a hybrid deep learning system that combines CNNs and Vision Transformers (ViTs) to capture both local and global image features. The system includes image quality assessment, attention-based feature selection, and classification using MobileNet and DenseNet models. Experimental results demonstrate improved accuracy and robustness, making the system a reliable solution for medicinal plant identification in healthcare, herbal medicine research, and biodiversity conservation.

II. REVIEW OF RELATED WORK

Recent advancements in Machine Learning (ML) and Deep Learning (DL) have significantly enhanced automated image analysis and classification in fields such as medical imaging, remote sensing, and computer vision. Earlier image processing methods relied on handcrafted feature extraction techniques like texture analysis, edge detection, color histograms, and shape descriptors, which often struggled to capture complex patterns and spatial relationships in images. With the emergence of advanced deep learning architectures such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), systems can now automatically learn hierarchical and contextual features from image data, enabling more accurate and scalable solutions for tasks like image classification, segmentation, and enhancement.

Although CNNs improved feature extraction by learning spatial patterns automatically, they still have limitations in capturing long-range dependencies. To address this, recent research has focused on hybrid architectures that combine CNNs with other advanced models. Studies such as FreFormer, Multibranch Attention Fusion Framework, CNN–Transformer hybrid networks, and CAF-Former integrate CNNs with Transformers or Graph Convolutional

Networks to capture both local and global image features. These hybrid models use techniques like attention mechanisms, feature fusion, and multi-scale learning to improve performance. Experimental results show that such architectures significantly enhance accuracy, robustness, and efficiency in complex image analysis tasks.

A. LITERATURE REVIEW

N. T. Singh et al. (2024) explored the application of Vision Transformers and hybrid CNN-Transformer architectures in modern computer vision systems. Originally developed for natural language processing tasks, transformer models rely heavily on the self-attention mechanism to capture relationships within sequential data. The study investigated how Vision Transformers can be adapted for various visual tasks, including video processing, low-level image analysis, and high-level vision applications. The research highlighted the importance of self-attention mechanisms in improving feature representation and reducing dependency on handcrafted vision-specific inductive biases. The authors also discussed the practical implementation of transformer-based models in real-world computer vision systems and emphasized their growing importance in next-generation intelligent visual recognition frameworks.

Recent research trends increasingly focus on hybrid deep learning architectures that combine CNN and Transformer models to leverage both local spatial feature extraction and global contextual learning capabilities. Attention mechanisms, feature fusion strategies, and hybrid network designs are being widely explored to improve performance across complex image analysis tasks. Despite significant progress, challenges remain in terms of computational efficiency, effective feature interaction, and model scalability. Building upon these developments, modern image classification frameworks aim to integrate robust preprocessing pipelines, hybrid feature extraction modules, and optimized deep learning architectures to achieve accurate and reliable automated image analysis systems.

B. Proposed Methodology

The proposed medicinal plant identification framework follows a structured multi-stage deep learning pipeline designed to accurately classify medicinal plant species using leaf images. The methodology integrates Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), feature fusion mechanisms, attention-based feature selection, and lightweight deep learning models such as MobileNet and DenseNet. The overall workflow is designed to exploit both local spatial patterns and global contextual information present in medicinal leaf images to improve classification accuracy and robustness. The proposed framework consists of the following stages:

C. Methodology

The methodology for the medicinal plant classification project uses a structured deep learning framework to automatically identify plant species from leaf images. It begins with collecting images from public datasets, followed by preprocessing steps such as resizing, noise removal, duplicate elimination, and quality filtering, after which the dataset is divided into training, validation, and testing sets. An image quality assessment module checks for blur and

lighting issues to ensure reliable input data. A hybrid feature extraction approach using CNN and Vision Transformer (ViT) models captures both local features (texture, edges, veins) and global contextual relationships. The extracted features are refined using an attention-based mechanism, and final classification is performed using MobileNet and DenseNet models. The system is trained using backpropagation and evaluated with metrics like accuracy, precision, recall, and F1-score, providing an effective automated solution for medicinal plant identification in healthcare, agriculture, and botanical research.

SYSTEM REQUIREMENTS

Hardware Requirements

- **Processor:** Intel Core i5 / AMD Ryzen 5 or higher
- **RAM:** Minimum 8 GB (16 GB recommended for better performance)
- **Storage:** Minimum 250 GB HDD or 128 GBSSD
- **GPU (Optional):** NVIDIA GPU with CUDA support for faster model training
- **Display:** 1366 × 768 resolution or higher
- **Network:** Stable Internet connection for GenAI integration and Flask server access
- **Sensors (Optional – IoT Integration):** pH Sensor, TDS Sensor, Turbidity Sensor, EC Sensor

Software Requirements

- **Operating System:** Windows 10/11 or Ubuntu 22.04 LTS
- **Programming Language:** Python 3.10 or higher
- **Framework:** Flask for web UI and backend integration
- **Machine Learning Libraries:** scikit-learn, XGBoost, LightGBM
- **Data Handling Libraries:** Pandas, NumPy
- **Visualization Tools:** Matplotlib, Seaborn, Plotly
- **Generative AI Tools:** OpenAI API / Hugging Face Transformers / LangChain
- **Prompt Engineering Toolkit:** LangChain or LlamaIndex
- **Database (Optional):** SQLite or MySQL for storing prediction history
- **Version Control:** Git / GitHub
- **IDE / Editor:** Visual Studio Code or PyCharm
- **Deployment Environment:** Localhost or Cloud platforms (AWS, GCP, Azure, Render)

D. ALGORITHM

 Data Collection & Preprocessing: Medicinal leaf images are gathered from public datasets and botanical repositories, cleaned by removing duplicates or low-quality images, and prepared through resizing, normalization, and noise removal.

 Image Quality Assessment: The system evaluates images for blur, noise, and poor lighting conditions, enhancing or discarding low-quality images to ensure reliable training data.

 Hybrid Feature Extraction: A **CNN** extracts local features such as leaf texture, edges, and vein patterns, while a **Vision Transformer (ViT)** captures global contextual relationships in leaf images.

 Feature Fusion & Optimization: Extracted CNN and ViT features are combined and refined using an attention-based mechanism to highlight the most important characteristics for accurate plant recognition.
 Model Training, Classification & Deployment: Deep learning models like MobileNet and DenseNet classify medicinal plant species, and the system is evaluated using metrics such as accuracy, precision, recall, and F1-score before deployment through a user-friendly interface.

III. System Architecture

The system architecture for the Medicinal Plant Identification and Classification System is designed to provide an accurate, scalable, and automated solution for identifying medicinal plant species using advanced deep learning techniques. The architecture follows a modular and extensible design, integrating image quality assessment, preprocessing, feature extraction, hybrid deep learning models, classification, and deployment modules.

The system consists of five primary interconnected modules, each responsible for a critical stage in the medicinal plant classification pipeline.

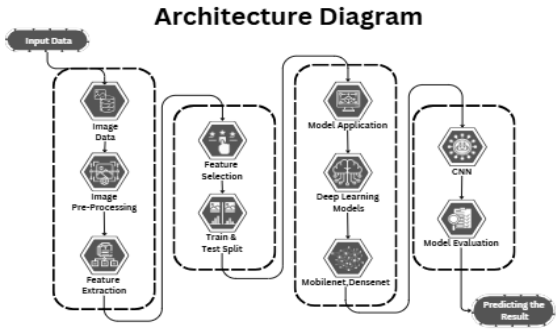

Fig 1: System Architecture

Workflow

- **Dataset Collection:** Medicinal plant leaf images are collected and organized into folders based on different plant species.
- **Exploratory Data Analysis (EDA):** The dataset is analysed to understand its structure, store image paths and labels, and check class distribution or imbalance.
- **Data Cleaning & Splitting:** Corrupted and duplicate images are removed, and the cleaned dataset is split into training, validation, and testing sets (70:15:15).
- **Data Augmentation:** Techniques such as rotation, flipping, zooming, and normalization are applied to improve model generalization.
- **Image Standardization:** All images are resized to 224 × 224 pixels to match the input requirements of deep learning models.
- **Model Training & Evaluation**: Models such as Custom CNN, MobileNetV2, and DenseNet121 are trained using transfer learning and optimized with techniques like Early Stopping and Model Checkpoint.
- **Prediction & Classification**: The best-performing model predicts the medicinal plant species from a new leaf image and outputs the predicted class with a confidence score.

Dataset Description Table

Attribute	Description
Dataset Name	Medicinal Plant Leaf Dataset
Data Type	Image Dataset
Image Format	JPG / PNG
Total Classes	40 Medicinal Plant Species
Dataset Structure	Folder-based classification dataset
Input Image Size	224 × 224 pixels
Dataset Path	../data
Train/Test Split	70% Training, 15% Validation, 15% Testing
Preprocessing	Corrupted image removal, duplicate removal, resizing, normalization
Data Augmentation	Rotation, flipping, zooming, rescaling
Models Used	CNN, MobileNetV2, DenseNet121
Feature Extraction	Transfer learning with ImageNet weights
Evaluation Metrics	Accuracy, Confusion Matrix, Classification Report
Prediction Method	Image classification using trained deep learning model

Table-1: Dataset Description

DATA FLOW AND INTEGRATION

1. Input

Medicinal leaf images are collected from publicly available plant image datasets. The images are cleaned, labeled, and divided into training, validation, and testing datasets before being passed to the preprocessing and feature extraction modules.

2. Processing

The input images undergo image quality assessment, preprocessing, and hybrid deep learning feature extraction.

The system performs:

- Image enhancement and preprocessing
- Local feature extraction using CNN
- Global feature extraction using Vision Transformers
- Feature fusion using attention-based mechanisms
- Classification using MobileNet and DenseNet architectures

The system computes:

- Predicted Medicinal Plant Species
- Classification Probability Scores
- Performance Evaluation Metrics

3. Output

The system provides:

- Predicted Medicinal Plant Species
- Classification Confidence Score
- Model Performance Evaluation Results

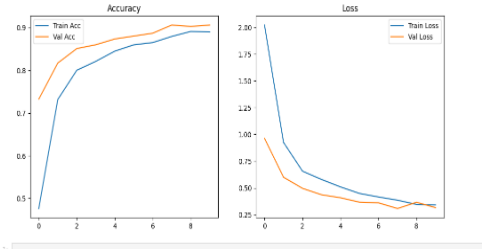

Fig 2: Accuracy & Loss Graphs

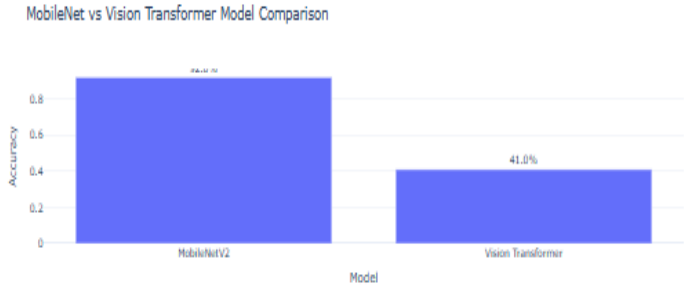

Fig 3: Model Comparison

IV.TESTING AND EVALUATION

Testing Methodology

The testing methodology for the Medicinal Plant Leaf Classification System using CNN, MobileNet, DenseNet, and Vision Transformer (ViT) follows a structured and systematic evaluation approach to ensure classification accuracy, robustness, generalization capability, and reliability.

The evaluation process validates the system's ability to correctly classify different medicinal plant leaf species based on their visual characteristics such as shape, texture, venation patterns, and color distribution.

Performance Metrics

Metric	Observation
CNN Accuracy	68%
MobileNet Accuracy	92%
DenseNet Accuracy	92%
Precision	High
Recall	High
F1-Score	High
Overfitting	Reduced due to preprocessing and feature fusion

Metric	Observation
Dataset Type	Multi-class medicinal plant leaf classification

Table-2:Performance Metrics

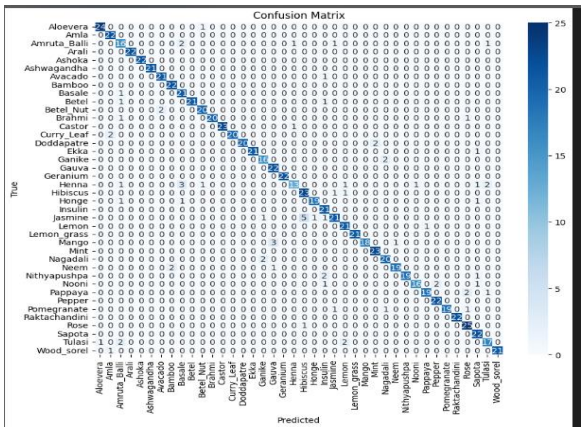

Fig-4: Medicinal Plant Classification Performance Metrics

V.PERFORMANCE EVALUATION

The performance evaluation focuses on multiple critical aspects to assess the system's effectiveness in automated medicinal plant identification and classification.

The system begins with data acquisition, where medicinal leaf images are collected from public datasets and botanical repositories, cleaned by removing duplicates or corrupted files, labeled by species, and divided into training, validation, and testing sets. The images then undergo preprocessing, including resizing, normalization, noise removal, blur detection, and illumination correction to ensure high-quality input for model training. A hybrid feature extraction approach using CNN and Vision Transformer (ViT) models captures both local features (such as leaf texture, shape, and veins) and global contextual information. These features are combined using an attention-based fusion mechanism to highlight important characteristics and reduce redundant data. Finally, the optimized features are classified using deep learning models like MobileNet and DenseNet, and the system's performance is evaluated using metrics such as accuracy, precision, recall, F1-score, and confusion matrix before deployment through a web interface for automated medicinal plant identification.

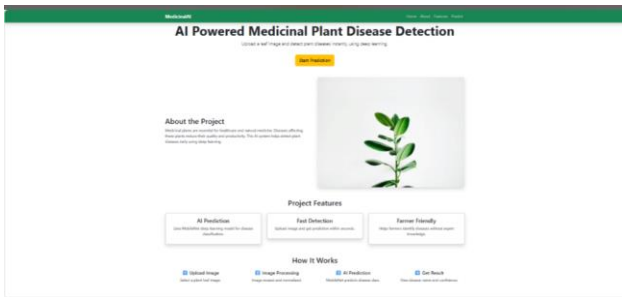

Fig-5: Home Page of Output Screen

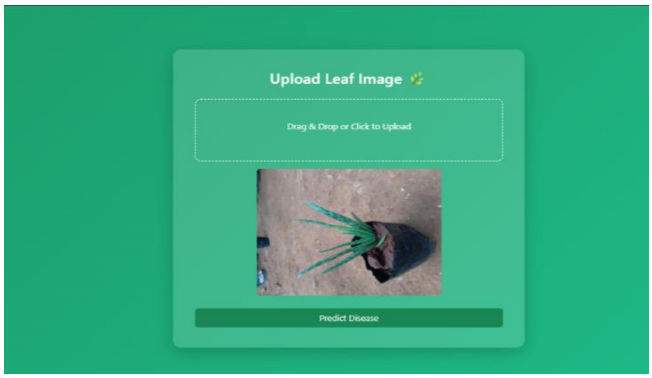

Fig-6: Uploading Image

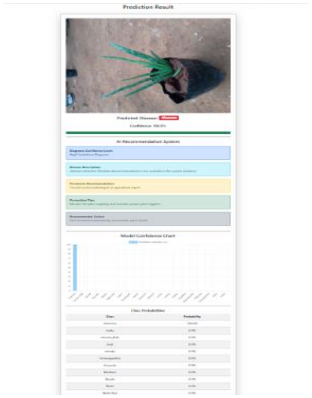

Fig-7: Displaying Results

VI. CONCLUSION

The AI-driven Medicinal Plant Classification System uses a hybrid deep learning approach combining CNNs and Vision Transformers (ViTs) to accurately identify medicinal plant species from leaf images by capturing both local features (texture, shape, veins) and global contextual information. An image quality assessment module first improves input images by removing blur and lighting issues, after which features are fused and refined using an attention mechanism. The system then classifies plants using efficient models such as MobileNet and DenseNet, achieving about 92% accuracy on a public dataset. This automated framework reduces reliance on manual plant identification and offers a scalable solution for applications in herbal medicine research, biodiversity conservation, and botanical studies.

Future Scope

1. **Real-Time Field Deployment:** Integrating the model into portable devices or agricultural monitoring systems for real-time medicinal plant identification in natural environments.
2. **Mobile Application Development:** Developing a smartphone-based plant identification application to assist farmers, botanists, and researchers in identifying medicinal plants instantly.
3. **Explainable AI Integration:** Incorporating visualization techniques such as Grad-CAM **or attention maps** to highlight leaf regions influencing classification decisions.

4. **Larger Dataset Expansion:** Training the model using larger and more diverse medicinal plant datasets to improve model generalization and classification accuracy.
5. **Multimodal Learning Framework:** Combining image-based features with additional botanical metadata such as plant habitat, growth conditions, and medicinal properties.

REFERENCES

[1] N. T. Singh, S. S. Kang and K. Lata, "Enhancing Computer Vision Using Vision Transformers and Hybrid CNN-Transformer Architectures," *2024 2nd International Conference on Emerging Trends in Engineering and Medical Sciences (ICETEMS)*, Nagpur, India, 2024, pp. 696-703, doi: 10.1109/ICETEMS64039.2024.10964958.

[2] F. Xu, S. Mei, G. Zhang, N. Wang and Q. Du, "Bridging CNN and Transformer With Cross-Attention Fusion Network for Hyperspectral Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-14, 2024, Art no. 5522214, doi: 10.1109/TGRS.2024.3419266.

[3] J. Fang, H. Lin, X. Chen and K. Zeng, "A Hybrid Network of CNN and Transformer for Lightweight Image Super-Resolution," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, LA, USA, 2022, pp. 1102-1111, doi: 10.1109/CVPRW56347.2022.00119.

[4] X. Liu, A. H. -M. Ng, L. Ge, F. Lei and X. Liao, "Multibranch Fusion: A Multibranch Attention Framework by Combining Graph Convolutional Network and CNN for Hyperspectral Image Classification," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-17, 2024, Art no. 5528817, doi: 10.1109/TGRS.2024.3442784.

[5] Z. Xinyi *et al*., "Enhancing Hybrid CNN-Transformer via Frequency-Based Bridging for Medical Image Segmentation," *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, Athens, Greece, 2024, pp. 1-4, doi: 10.1109/ISBI56570.2024.10635477.