International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Review on Deep Learning and Segmentation Techniques for Medicinal Plant Identification

R. Srinivasa Rao[1], V. Sowmya[2], K. Sai Yogitha[3], Mrs. A. Bhavani[4]

CSE, GMR Institute of Technology, Razam, Andhra Pradesh, India.
routhusrinivasarao2004@gmail.com, sowmyavegireddy@gmail.com, yogithkilari@gmail.com, bhavaniashapu@gmail.com

ABSTRACT:

The identification of medicinal plants plays a vital role in traditional medicine, pharmaceutical development, and biodiversity conservation. However, morphological similarities between plant species, environmental variations, and limitations of traditional manual identification methods have created challenges in accurate plant classification. In recent years, deep learning (DL) and image segmentation techniques have emerged as powerful tools to overcome these barriers. This review provides a comprehensive analysis of deep learning and segmentation-based approaches in medicinal plant identification. We explore various deep neural network architectures such as MobileNet, EfficientNet, DenseNet, NASNet, and Xception, along with segmentation models like U-Net and MSRF-Net. We further discuss the significance of curated datasets, preprocessing techniques, mobile and real-time applications, and hybrid learning methods. The challenges in dataset diversity, model generalization, feature extraction, and computational requirements are also addressed. Through this review, we aim to provide insights into the current state-of-the-art techniques, identify existing gaps, and propose future research directions for enhancing the reliability and scalability of automated medicinal plant identification systems.

Keywords: Medicinal Plants, Deep Learning, Image Segmentation, U-Net, CNN, MobileNet, Xception, Dataset Curation, Feature Extraction, Plant Identification

1. Introduction

Medicinal plants have played a vital role in human health for thousands of years, serving as foundational elements in traditional healing systems and continuing to contribute to modern pharmacological discoveries. Their natural therapeutic properties are used in the treatment and prevention of various diseases, making their accurate identification essential for ensuring safety, efficacy, and conservation. However, manual identification methods—based on morphology, expert observation, and herbarium samples—are often time-consuming, error-prone, and limited by the availability of botanical expertise. The visual similarity between different plant species, especially among those within the same genus, adds to the complexity of identification. Environmental factors such as lighting, background clutter, leaf damage, and growth stage variations can further reduce the effectiveness of traditional classification methods. These limitations highlight the need for robust, automated solutions that can provide consistent and accurate identification, even in uncontrolled or natural settings. In this context, artificial intelligence (AI) and computer vision have emerged as promising tools to bridge this gap.

Deep learning, a subset of AI, has shown remarkable success in image classification tasks due to its ability to learn complex hierarchical features from large datasets. In particular, Convolutional Neural Networks (CNNs) have become the cornerstone of image-based classification, including in the field of plant identification. CNNs automatically extract features such as shape, texture, and vein patterns from leaf images, eliminating the need for manual feature engineering. Their ability to generalize across variations in input data makes them suitable for real-world applications involving diverse plant species.

Alongside classification, image segmentation has also become a critical component in improving plant identification accuracy. Segmentation techniques help isolate the relevant parts of an image, such as the leaf, from the background. This is particularly important in field conditions where images are captured under inconsistent lighting and cluttered backgrounds. Deep learning models like U-Net and MSRF-Net have proven to be highly effective in medical and botanical image segmentation due to their ability to preserve fine spatial details.

Moreover, the growing availability of curated datasets, advances in model architectures, and the development of real-time, mobile-friendly solutions have significantly accelerated progress in this domain. Applications such as mobile plant recognition apps enable users, including farmers and healthcare professionals, to identify medicinal plants instantly, even in remote regions. These technologies not only enhance plant classification but also support conservation efforts, agricultural productivity, and the preservation of indigenous knowledge.

This review aims to provide a comprehensive overview of the recent advancements in deep learning and segmentation techniques for medicinal plant identification. It explores different CNN architectures, segmentation models, dataset resources, and real-time applications. The paper also discusses key challenges such as dataset diversity, model generalizability, and computational constraints, offering insights into future research directions that can further improve the scalability and reliability of these intelligent plant identification systems.

2. Literature Survey

Kavitha et al. (2023) designed a deep learning system centered around the MobileNet model for real-time medicinal plant detection. The model demonstrated high efficiency and accuracy (98.3%) in identifying six plant species using a balanced image dataset and was deployed in a mobile application to offer health information based on identified leaves.

Srivastava et al. (2021) introduced MSRF-Net, a segmentation framework that integrates multi-scale residual fusion to improve image segmentation outcomes. With strong performance on medical datasets, the architecture offers potential adaptability to plant-based tasks due to its modular shape stream and dense fusion techniques.

Geerthana et al. (2021) implemented a CNN-based classification model capable of recognizing five Indian medicinal plant types. Their model emphasized the significance of morphological traits like shape, texture, and leaf coloration, yielding a classification accuracy of 96.67% and highlighting the need for broader species coverage.

Roopashree and Anitha (2021) evaluated several deep CNN architectures and incorporated the Xception model into a mobile interface named HerbSnap. The resulting DeepHerb framework enabled effective recognition of Indian herbs in real time, achieving 97.5% accuracy with high reliability.

Pushpa and Rani (2023) curated the DIMPSAR dataset, designed to enhance machine learning models for Indian medicinal plant classification. The dataset, consisting of leaf images captured in diverse lighting, supported the training of a deep CNN called Ayur-PlantNet and emphasized its utility in plant disease prediction.

Dey et al. (2024) carried out a comparative evaluation of seven CNN variants using a dataset comprising over 5,800 images. DenseNet201 emerged as the top-performing model with 99.64% accuracy, underlining the importance of data augmentation in boosting generalization performance.

Bodhwani et al. (2019) employed a 50-layer ResNet framework to classify plant species and achieved over 93% accuracy. They enhanced training outcomes through augmentation tools and suggested extending the method to detect plant diseases and insect infestations.

Sarma et al. (2023) released MED 117, a dataset focused on medicinal plants in Assam, India. It featured leaf images with U-Net-based segmentation and provided botanical and vernacular naming conventions for better recognition accuracy.

Sodjinou et al. (2022) presented a segmentation approach combining U-Net with K-means clustering to differentiate crops and weeds in complex agricultural environments. Their system attained a segmentation accuracy exceeding 99%, highlighting the benefits of hybrid models in complex image backgrounds.

Acharya et al. (2023) designed a system for medicinal plant recognition augmented with geolocation services. Leveraging an SVM classifier and a mapping interface, the tool allowed users to identify plants and visualize their native regions on a digital map.

Naeem et al. (2021) incorporated multispectral and textural image features to enhance machine learning-based classification of medicinal plants. Their model, trained on six species, surpassed 98% accuracy and demonstrated the benefits of multi-domain feature fusion.

Rajani and Veena (2019) compared conventional image segmentation methods, including edge detection and thresholding, to isolate plant features for classification. Though effective in simple environments, they noted challenges in complex scenes and advocated hybrid solutions.

Somasekhar et al. developed a CNN-based medicinal plant classifier that emphasized the importance of digital image processing and conservation. Although the model performed well, their work lacked comparative benchmarks against state-of-the-art alternatives.

Bose et al. (2021) examined machine learning applications for detecting leaf-based diseases in medicinal plants. Their findings emphasized the utility of morphological traits in classification and confirmed the superior performance of automated systems over manual inspection.

Anchitaalagammai et al. (2021) tested traditional and neural models for plant classification. Their experiments with Random Forest, KNN, Logistic Regression, and ANN highlighted the value of neural networks trained on color and edge histograms.

Jayalath et al. (2019) combined flower and leaf morphological features in a CNN-based classification framework, achieving 90% accuracy. They emphasized the added benefit of using multiple plant parts over relying solely on leaves.

Vayadande et al. (2024) developed AyurLeaf, a CNN-powered mobile platform for identifying Ayurvedic medicinal plants. The app classified raw images with high precision and supported traditional healthcare systems by offering identification and information simultaneously.

Yoo et al. (2020) proposed a deep learning framework that predicts the medicinal applications of plant-based chemical compounds. By leveraging chemical properties, molecular interaction data, and semantic knowledge from scientific literature, their in silico approach offers an accurate and large-scale alternative to conventional in vitro testing.

Kayhan and Ergün (2020) developed a deep learning-based classification technique focused on medicinal and aromatic plants, emphasizing the analysis of thousands of plant-derived compounds. Their model addresses data inconsistency while predicting medicinal value through computational simulation.

Pacifico et al. (2019) presented an image classification framework using machine learning models such as KNN, Random Forest, and Decision Trees to identify Brazilian medicinal plant species based on leaf texture and color. Their models exceeded 97% accuracy, showcasing the feasibility of ML-driven plant classification.

Ghosh et al. (2023) designed a hybrid classification model that combines Principal Component Analysis (PCA) and a VGG16 CNN to improve recognition accuracy. Evaluated on a dataset of 30 plant species, the model achieved 95.25% accuracy, outperforming traditional learning approaches.

Akter and Hosen (2020) created a CNN-based model for identifying Bangladeshi medicinal plants. Using a three-layer CNN with batch normalization and noise injection, they attained 71.3% accuracy and addressed overfitting issues through augmentation.

Uddin et al. (2023) addressed dataset limitations in Bangladesh by curating 5,000 leaf images from ten species. After applying various preprocessing steps and deep learning models, their ensemble-based approach achieved up to 99% classification accuracy.

Begue et al. (2017) proposed a plant identification system based on Random Forest classifiers trained on images of 24 medicinal plants. Their approach reached 90.1% accuracy and laid the foundation for future web and mobile deployment.

Duong-Trung et al. (2019) explored transfer learning-based CNNs for medicinal plant classification, supporting real-time mobile usage and future extension into recommendation systems.

Mulugeta and Sharma (2024) conducted a meta-review of 31 global studies, finding that over 80% utilized transfer learning, while two-thirds relied on CNN architectures. They emphasized the importance of open datasets and collaborative development.

Singh and Singh (2021) provided a comprehensive overview of machine learning techniques for plant classification, analyzing supervised, semi-supervised, and unsupervised methods. They highlighted the effectiveness of using geometric leaf features and recommended hybrid strategies.

Xue et al. (2019) applied morpho-colorimetric features and VIS/NIR spectroscopy to train ANN models for Chinese medicinal plants. Their classification model reached over 98% accuracy and suggested smartphone applications for affordable deployment.

Prasvita and Herdiyeni (2013) developed MedLeaf, an Android application that used LBP and Probabilistic Neural Networks for medicinal plant recognition in Indonesia, achieving moderate accuracy and encouraging future improvements for broader adoption.

Table 1: Literature survey Table

S.no	Title	year	Description	Limitations	Advantages	Performance metrics	Gaps
1	Medicinal Plant Identification in Real-Time Using Deep Learning Model	2023	The paper presents a vision-based smart method using a deep learning (DL) model for the real-time identification of medicinal plants.	The study's focus on only six medicinal plant species (betel, curry, tulsi, mint, neem, and Indian beech) from the Kaggle database limits the generalizability of its findings.	The use of the MobileNet DL model allows for efficient classification of plant species, as it uses fewer computing resources and can be deployed on low-powered devices.	The DL model used in the study is the MobileNet DL model, which achieved an accuracy rate of 98.3% in correctly identifying medicinal leaves.	The paper lacks a comparison of the MobileNet model's performance against other medicinal plant identification methods like ML classifiers or traditional image processing.
2	MSRF-Net: A Multi-Scale Residual Fusion Network for Biomedical Image Segmentation	2021	It discusses about a novel architecture called MSRF-Net for biomedical image segmentation, which uses multi-scale residual fusion to capture object variabilities and improve segmentation performance.	The paper uses four publicly available datasets for biomedical image segmentation, but they are relatively small and may not cover all the variations and challenges in real-world medical images.	The use of dual-scale dense fusion blocks facilitates information exchange across different resolution scales. By preserving both local and global features, the model achieves better segmentation performance.	The Dice Coefficient (DSC) achieved is 0.9217 on KvasirSEG, 0.9420 on CVC-ClinicDB, 0.9224 on the 2018 Data Science Bowl dataset, and 0.8824 on the ISIC-2018 skin lesion segmentation challenge dataset.	The paper lacks an ablation study or analysis of MSRF-Net's individual components (shape stream, triple attention, loss function), making it unclear how each contributes to performance and their associated trade-offs or limitations.
3	Medicinal Plant Identification Using Deep Learning	2021	This paper presents a system that can classify five different Indian medicinal plant species based on their leaf images using a convolutional neural network (CNN) model.	Using a pre-trained InceptionV3 model (314 layers, millions of parameters) as the base may cause overfitting, high computational cost, and low interpretability.	It utilizes a large and diverse dataset of 58,280 images collected by the authors, which can be useful for future research and applications.	The paper reports a 96.67% success rate in identifying the corresponding medicinal plant using a CNN model.	The paper claims a high accuracy of 96.67% but lacks details on its measurement, the metrics used, and how the model handles noise, occlusion, or image variations.
4	DeepHerb: A Vision Based System for Medicinal	2021S	It describes a vision-based system for medicinal plants using deep learning	The paper lacks details on image collection, annotation, and verification processes	It applies transfer learning on four pre-trained deep convolutional	The paper reports the highest accuracy of the proposed	The paper does not evaluate the DeepHerb model's

	Plants Using Xception Features		techniques and a custom dataset of Indian herbs.	for accuracy and consistency, and it doesn't mention any data cleaning or quality control procedures.	neural network architectures (VGG16, VGG19, InceptionV3 and Xception) and compares their performance with two machine learning classifiers (ANN and SVM).	models as 97.5% on the DeepHerb dataset (using DeepHerb model: Xception + ANN) and 97.22% on the Flavia dataset (using the same model).	robustness to variations in leaf features (shape, size, orientation, color, texture, background) or challenges like compound leaves, occlusion, and damage.
6	Assessing deep convolutional neural network models and their comparative performance for automated medicinal plant identification from leaf images.	2024	The study evaluated seven deep learning algorithms (VGG16, VGG19, DenseNet201, ResNet50V2, Xception, InceptionResNetV2, InceptionV3) for automated medicinal plant identification using leaf images from public (PI) and blended (PFI) datasets.	The paper did not address the challenges of identifying medicinal plants from other plant parts, such as flowers, fruits, stems, or roots, which may have different features and characteristics than leaves.	It evaluates the performance of the models on two types of data: public data with plain backgrounds and field data with complex backgrounds, reflecting the real-world scenarios of plant identification.	The paper found that DenseNet201 outperformed all other models, achieving 99.64 % accuracy and 98.31 % precision for PI, and 97 % accuracy and 96.8 % precision for PFI.	The paper lacks a comparison of the DCNN models' performance with other existing medicinal plant identification methods or tools, potentially limiting the generalizability and applicability of its findings.
7	Deep Residual Networks for Plant Identification	2019	The paper proposes a 50-layer deep learning model using convolutional neural networks (CNN) and residual networks (ResNet) to classify plant species from leaf images.	The authors designed a 50-layer deep residual network, which may require a lot of computational resources and time to train and test. A simpler or more efficient model may achieve similar or better results with less cost.	It uses a deep residual network to learn discriminative features for plant identification from leaf images, which is a challenging task in computer vision and botany.	The paper reports an accuracy of 93.09% on the test set, demonstrating that deep learning is a promising technology for large-scale plant identification in the natural environment.	The paper highlights the n The paper does not justify the choice of parameters (filters, learning rate, batch size, epochs) for the ResNet50 model, making it unclear how these parameters affect the model's accuracy and efficiency, and whether they are optimal.
8	MED 117: A dataset of medicinal plants mostly found in Assam with their leaf images, segmented leaf frames and name table	2023	A dataset of 117 mostly Assamese medicinal plant species is presented, featuring leaf images, segmented frames (using U-Net), and a name table, collected from medicinal gardens and processed from	The data analysis, centered on U-Net image segmentation, neglects other plant aspects (medicinal properties, uses, phylogeny) and is constrained by computational resources and chosen parameters/algorithms.	The dataset covers 117 medicinal plant species (botanical, common, Assamese names, family, sample size) and provides both raw and U-Net segmented leaf images for tasks like classification,	The paper applies U-Net segmentation model to extract the leaf regions from the raw images. The paper reports an average training accuracy of 96.33% and an	The paper omits discussion of U-Net segmentation's limitations (need for large/diverse data, hyperparameter sensitivity, computational

			leaf videos.		segmentation, or feature extraction.	average validation accuracy of 95.49% for the U-Net model.	complexity) and potential solutions or improvements.
9	A deep semantic segmentation-based algorithm to segment crops and weeds in agronomic color images	2022	The paper proposes a segmentation method for separating crops and weeds in agricultural color images using semantic segmentation (U-Net) followed by K-means clustering initialized with subtractive clustering.	The paper uses only 200 agronomic color images from two databases, which may not capture the full variability of crops and weeds. A larger, more diverse dataset could enhance segmentation robustness and generalization.	The paper proposes a novel crop/weed segmentation method combining semantic segmentation (using the powerful and efficient U-Net) and K-means algorithms for pixel-level prediction in color images.	The proposed algorithm achieved a maximum accuracy of 99.19%, with a positive rate of 0.9952, negative rate of 0.8985, and segmentation scores of 0.472 (base case) and 0.602 (model-based simulator).	The paper lacks comparison with leading semantic segmentation methods like DeepLab, PSPNet, or Mask R-CNN, potentially limiting performance evaluation and validation.
10	Automated Plant Recognition System with Geographical Position Selection for Medicinal Plants	2023	The paper focuses on designing an automated plant recognition system based on the best recognition algorithm and the Google platform to locate all plant locations on a map.	The paper does not provide detailed information about the specific recognition algorithm used for plant recognition, making it difficult to assess the effectiveness and accuracy of the system.	The paper addresses the challenge of recognizing plant leaf images by designing an automated plant recognition system, which can handle massive data and provide intermediate analysis.	The paper offers valuable insights into smart farming's role in sustainable rice production, emphasizing its potential to boost productivity and reduce environmental impact.	The absence of information on evaluation metrics makes it difficult to assess the system's performance and compare it to existing methods.
11	The Classification of Medicinal Plant Leaves Based on Multispectral and Texture Feature Using Machine Learning Approach	2021	This paper proposes a machine learning-based classification method for medicinal plant leaves using multispectral and texture features.	The study only focuses on six specific varieties of medicinal plant leaves, which may not be representative of all medicinal plants.	The paper proposes a machine learning-based method for classifying medicinal plant leaves, aiding in the identification and categorization of plant species for various applications.	The paper achieves a relatively high accuracy of 99.01% for the classification of medicinal plant leaves using the multi-layer perceptron classifier.	The paper highlights smart farming's potential to enhance sustainable rice production by improving productivity and minimizing environmental impact.
12	Medicinal Plants Segmentation using Thresholding and Edge based Techniques.	2019	The paper explores image processing techniques for identifying and classifying medicinal plants, focusing on segmentation to isolate the object of interest from the background.	The paper does not provide a comprehensive evaluation of all available segmentation techniques, as it only focuses on edge detection and thresholding methods.	Provides insights into the use of image processing techniques for the identification and classification of medicinal plants.	The paper provides valuable insights into the use of image processing techniques for the identification and classification of medicinal plants .	The paper provides valuable insights into the use of image processing techniques for the identification and classification of medicinal plants.
13	Identification of medicinal plants by image	2023	The paper proposes a method for classifying medicinal plants	The paper does not provide details about the specific image processing algorithms	Provides a method for classifying medicinal plants using image	The paper does not provide specific accuracy metrics or	The paper lacks specific accuracy metrics and

	processing of leaf samples		using image processing of leaf samples.	used for feature extraction and picture improvement, which may limit the reproducibility and comparability of the results.	processing of leaf samples, which can be a time-consuming and labor-intensive task for humans.	comparative results with other methods, making it difficult to assess the accuracy of the proposed method.	comparative results, making it difficult to evaluate the proposed method's performance and effectiveness.
14	Survey and Analysis of Medicinal Plant Identification via Image Processing and Machine Learning Techniques	2023	The objective of the paper is to enhance the accuracy and efficiency of identifying medicinal plants in Ayurvedic medicine by incorporating advanced image processing and machine learning algorithms.	The paper does not provide a detailed analysis of the limitations or challenges faced during the implementation of the advanced image processing and machine learning algorithms for medicinal plant identification.	Leveraging these technologies improves medicinal plant identification accuracy, aiding the Ayurvedic supply chain by reducing market confusion and ensuring quality and authenticity.	The trained neural network model achieved a test accuracy of 90% using TensorFlow on a self-created dataset .	The paper does not discuss the limitations or challenges of implementing image processing and machine learning for medicinal plant identification.
15	Identification and classification of rare medicinal plants using machine learning techniques	2022	The paper focuses on using machine learning algorithms to identify and classify rare medicinal plants based on visual morphological characteristics like shape, color, and texture of leaves and flowers.	The paper does not mention the specific dataset used for training and testing the machine learning algorithms, which could affect the generalizability of the results.	Machine learning algorithms can save time and effort in identifying and classifying rare medicinal plants, compared to manual identification by experts.	The highest accuracy achieved in a similar study was 90.1% using the random forest classifier.	The paper lacks details about the dataset used for training and testing the machine learning algorithms, which is crucial for assessing the results' generalizability.
16	Identification of Medicinal Plants by Visual Characteristics of Leaves and Flowers	2019	Image processing and machine learning techniques are applied to create a database of scanned images of leaves and flowers of rare medicinal plants used in Sri Lankan Ayurveda medicine.	The paper lacks detailed information on the image processing and machine learning techniques used for plant identification, hindering replication and further exploration of the methodology.	The paper presents a method for automated medicinal plant identification using visual morphological characteristics, offering greater efficiency and accuracy than manual identification by experts.	The paper achieves identification rates up to 98% when tested over 10 rare medicinal plants, indicating a high level of accuracy.	The paper does not mention the specific limitations or gaps in the proposed method for automated identification of medicinal plants using visual characteristics.
17	Identification of Ayurvedic Medicinal Plant Using Deep Learning	2023	The paper discusses using convolutional neural networks (CNN) and leaf images to identify Ayurvedic medicinal plant leaves, employing computer vision and image classification techniques for accurate recognition	The research focuses on the identification of Ayurvedic medicinal plant leaves using convolutional neural networks, but it does not discuss the potential limitations of this approach or any potential drawbacks of using deep learning for leaf classification.	The new technique that uses convolutional neural networks (CNN) and leaf images to accurately identify the leaves of Ayurvedic medicinal plants, providing a more accurate classification	the paper highlights that the deep neural network method yields a more accurate classification compared to traditional methods, indicating that accuracy could	Limited Plant Variety: The system is currently able to identify 30 different plant groupings2. Expanding the dataset to include more plant species could improve

			and classification.		compared to traditional methods.	be a potential performance metric.	the system's robustness and applicability
18	A Deep Learning-Based Approach for Identifying the Medicinal Uses of Plant-Derived Natural Compounds	2020	The paper presents a deep learning approach to identify the medicinal uses of plant-derived compounds, addressing the limitations of in vitro methods by utilizing extensive, heterogeneous drug and compound data.	The study highlights the time and cost challenges of in vitro and in vivo approaches and the issues with in silico analysis due to incomplete data. It also notes that the limited number of compounds used in the deep learning model impacts prediction reliability.	Predicts natural compound efficacies with high accuracy, reducing time and cost, and aids in screening candidates for medicinal use.	The deep learning model predicted natural compound efficacies with high accuracy, achieving an average AUROC of 0.900 ± 0.040 across 15 diseases, demonstrating strong predictive capability.	Traditional methods face challenges with incomplete natural compound data, while in vitro methods are costly and time-consuming. In silico approaches have low coverage due to hidden information.
19	Medicinal and Aromatic Plants Identification Using Machine Learning Methods	2020	The study focuses on the importance of medicinal plants and their extracts, particularly natural compounds like phytochemicals, antioxidants, vitamins, and minerals, in promoting health and preventing disease.	The approach may struggle with incomplete, heterogeneous data, and integrating complex information is challenging. Additionally, a partially connected deep neural network requires a large training dataset, which may be hard to obtain.	Uses 10 features, including shape, gray, and fractal features, to classify plants. Compares machine learning methods like NBC, CART, KNN, and PNN, highlighting NBC's effectiveness for this application based on leaf characteristics.	High AccuracyUtilizes Naive Bayes Classifier (NBC) to achieve high classification rates, with training data at 98.39% and test data at 98.00%	The study focuses on NBC, CART, KNN, and PNN. Future research could explore hybrid methods and the impact of additional features for improved accuracy.
20	Automatic Classification of Medicinal Plant Species Based on Color and Texture Features	2019	The paper presents an automatic system for classifying medicinal plant species using color and texture features, testing five machine learning classifiers to find the best algorithm.	The dataset includes 15 Brazilian medicinal plant species and 1,148 leaf segments. Expansion is needed for more species and features like shape, venation, and leaf tooth.	Creation of a new dataset with features extracted from 1148 leaf image segments.Empirical analysis complemented by hypothesis testing to evaluate classifier performance.	High Accuracy Implementation of five machine learning classifiers with the best achieving over 97% accuracy on the dataset	A global, public dataset for country-specific medicinal plants is needed, and the reliability of deep learning approaches for plant classification and recognition requires further investigation
21	Identification of medicinal plant using hybrid transfer learning technique	2023	The paper discusses using a hybrid transfer learning technique to identify medicinal plants, focusing on improving plant identification for Ayurvedic treatments. It introduces a novel method combining deep learning and PCA for classification.	The paper does not discuss the specific limitations or challenges faced during the implementation of the proposed learning to prediction approach.	The study uses PCA and VGG16 for classification, with a dataset of 30 medicinal plant species, addressing the challenge of plant identification for Ayurvedic treatments.	The research demonstrates the effectiveness of the proposed PCA-VGG16 hybrid deep learning architecture, achieving a test accuracy of 95.25%.	The dataset includes 30 medicinal plant species, and expanding it with live samples could improve generalizability and real-world applicability.
22	CNN-based	2020	The study compared	This paper deals with	The proposed	This research	Future research

	Leaf Image Classification for Bangladeshi Medicinal Plant Recognition.		the proposed methodology with existing works and found it to be reliable, efficient, and better in terms of medicinal plant leaves classification.	Hand-crafted features affected by noise and environmental changes. And Natural background in images makes classification challenging in real-world scenarios.	machine learning-based methodology for classifying medicinal plant leaves is reliable, efficient, and accurate, outperforming existing approaches.	achieved a 71.3% accuracy rate in plant classification, with 2,550 accurately classified images and 1,020 misclassified.	could explore hybrid methods and the impact of additional features for improved accuracy.
23	Deep-Learning-Based Classification of Bangladeshi Medicinal Plants Using Neural Ensemble Models	2021	This paper addresses the lack of Bangladeshi medicinal plant datasets, using deep learning models and novel neural network architectures to improve classification accuracy.	The dataset is limited to Bangladeshi medicinal plants, and the models' generalization ability for unseen images is not thoroughly discussed.	The paper presents a dataset of 5,000 images of ten Bangladeshi medicinal plant species and applies preprocessing techniques to improve image quality and classification accuracy.	The models achieved high accuracy, with DenseNet201 at 85.28%, DRD at 95%, DRCD at 97%, and the soft ensemble technique achieving the highest accuracy of 99%.	There is a lack of discussion on the deployment of these models in real-world scenarios and how they would perform in less controlled environments.
24	Automatic Recognition of Medicinal Plants using Machine Learning Techniques	2017	The paper presents a fully automated method for recognizing medicinal plants using computer vision and machine learning techniques.	The paper uses a limited dataset of 24 Mauritius plant species and smartphone-collected leaf images, which may introduce variations in quality and lighting, affecting accuracy.	Leaves from 24 medicinal plant species were photographed with a smartphone in a lab, and features like length, width, perimeter, area, color, and number of vertices were extracted.	The random forest classifier achieved 90.1% accuracy with 10-fold cross-validation, outperforming other methods like k-NN, naive Bayes, SVM, and neural networks.	The paper overlooks challenges like the need for diverse datasets, potential biases, and generalization to unseen species in plant recognition.
25	A Combination of Transfer Learning and Deep Learning for Medicinal Plant Classification	2016	The paper focuses on classifying medicinal plants using transfer learning and deep learning, proposing an adapted architecture and exploring transfer learning for real-world classification.	Lack of hardware infrastructure and financial support for model training. Overfitting observed in the model during testing.	The paper explores transfer learning, deep learning, and medicinal plant classification, proposing a novel method and collecting a benchmark dataset to address hardware limitations.	Achieved classification accuracy of 98.7% on test set. Lowest accuracy score reported was 93.2%.	The paper does not provide detailed information about the specific deep learning architecture used for medicinal plant classification.
26	Deep learning for medicinal plant species classification and recognition: a systematic review	2024	The paper outlines deep learning techniques for classifying medicinal plants, covering feature detection, image classification, object detection, segmentation, and identification characteristics.	Lack of globally available public dataset for medicinal plants. Trustworthiness of deep learning for plant classification and recognition.	The paper reviews studies on deep learning for medicinal plant classification, identifying trends, gaps, and challenges, with a focus on recent advancements in plant recognition.	CNN achieved 98.4% precision, 98.5% recall, and 98.45% F1-Score, while MobileNetV2 and Mask-RCNN reached 98.3% and 98.7% accuracy, respectively. VGG19 and Logistic Regression also showed high performance.	Scarcity of datasets and geographical disparities are significant research gaps. Lack of globally available public dataset for medicinal plants is a gap.

27	A Survey on Different Methods for Medicinal Plants Identification and Classification System	2021	Medicinal plant identification methods and leaf-based automated classification are compared, highlighting the advantages of herbal treatments for chronic illnesses over synthetic options.	SVM not suitable for large data. CNN needs more tropical plant species for flexibility. K-NN classifier time-consuming for real-world applications.	Automated medicinal plant identification via machine learning on leaf images is discussed, reviewing methods like deep CNNs and emphasizing accuracy and performance.		This paper does no explain about the classification that much efficiently. There are some portions of this paper needs to be improvised.
28	Automated Chinese medicinal plants classification based on machine learning using leaf morpho-colorimetry, fractal dimension and visible/near infrared spectroscopy	2019	Models based on leaf morpho-colorimetry and VIS/NIR spectral analysis. ANN models achieved high accuracy for plant classification. Models A and B developed for leaf classification. Models can be implemented through smartphone applications	Describes complexity of techniques and expensive instrumentation. Limited number of species studied due to cost prohibitive NIR instrument.	Automated classification of Chinese medicinal plants using machine learning models. Morpho-colorimetric and VIS/NIR spectral analysis for plant identification. Multivariate data analysis for relationship between morphometric parameters and plants.	Model A accuracy: 98.3% in leaf classification. Model B accuracy: 92.5% in leaf classification	Lack of detailed description on machine learning modelling techniques. Limited information on specific machine learning algorithms used in the study. Absence of discussion on potential challenges faced during model development.
29	MedLeaf: Mobile Application for Medicinal Plant Identification Based on Leaf Image	2013	MedLeaf, a mobile app, identifies medicinal plants and aids discovery via leaf texture analysis (Local Binary Pattern) and image classification (Probabilistic Neural Network).	This paper does not address the effectiveness of the mobile application. Also does not specify the particular learning algorithms responsible for the cost effectiveness.	MedLeaf mobile app for medicinal plant identification is the focus, including user satisfaction with its results, and its use of Local Binary Pattern and Probabilistic Neural Network for classification.	User satisfaction evaluated through questionnaire with heuristic evaluation. The accuracy of medicinal plant identification based on leaf texture is 56.33%.	High user satisfaction in plant ID contrasts with complex botanical garden use, challenging this model's proposal.
30	A Group Labelled Classification Model for Accurate Medical Plant Detection Used in Drug Preparation	2021	Research on accurate medical plant detection for drug preparation. Utilizes group labelled classification model for leaf feature classification. Focuses on optimizing recognition rate of green leaves for accuracy.	Limited discussion on real-world application scenarios. Lack of detailed comparison with other state-of-the-art models.	Focus on medicinal plant classification using leaf features. Mention of hierarchical clustering and leaf shape recognition techniques. Comparison with traditional methods showing improved performance.	The proposed model exhibits lower classification times and reduced labeling mark time intervals. Furthermore, its accuracy levels surpass existing methods, and feature optimization contributes to enhanced overall performance.	The paper lacks specifics on the features used for classifying medicinal plant leaves and omits details about the training/testing dataset, raising concerns about the generalizability of the model's results.

3. Methodology

This section outlines the variety of methodologies reported in the literature for medicinal plant identification using deep learning and segmentation techniques.

3.1 Feature Engineering and Preprocessing

Preprocessing is a critical phase in any deep learning pipeline as it directly influences the learning capacity and performance of the models. Standard preprocessing steps include image resizing, normalization, and enhancement to ensure uniformity and quality across datasets. Data augmentation techniques—such as horizontal/vertical flipping, rotation, scaling, cropping, color jittering, and Gaussian noise—are applied to increase the diversity of training data without collecting new samples. These augmentations make models more robust to variations in orientation, lighting, and occlusion.

Normalization of pixel intensities typically rescales image data to a fixed range (such as 0 to 1 or -1 to 1), allowing the models to converge faster during training. Advanced preprocessing also includes contrast-limited adaptive histogram equalization (CLAHE) for enhancing the visibility of leaf veins and textures.

Beyond traditional image manipulation, Naeem et al. (2021) explored the combination of multispectral imaging and texture analysis, which provides a richer feature space by capturing information beyond visible light. These multispectral features, coupled with statistical texture descriptors, help in classifying morphologically similar species that may be indistinguishable by standard RGB images.

3.2 Segmentation Techniques

Segmentation is a foundational step in isolating plant structures—such as leaves, flowers, or fruits—from noisy backgrounds. This step enhances model focus by filtering irrelevant regions and enabling precise feature extraction.

The U-Net model, with its symmetric encoder-decoder structure and skip connections, remains one of the most effective segmentation techniques in plant and biomedical domains. Sarma et al. (2023) demonstrated its use on the MED 117 dataset to generate clean, grayscale leaf masks from raw images. The network's capability to preserve spatial information is especially beneficial in distinguishing overlapping leaves or uneven boundaries.

MSRF-Net (Srivastava et al., 2021) brings enhanced performance through multi-scale feature extraction and residual connections. This allows it to capture finer details across varying resolutions, improving segmentation accuracy in diverse conditions. In agricultural contexts, Sodjinou et al. (2022) proposed a hybrid model combining U-Net and K-means clustering to improve leaf-background separation in cluttered field images.

Traditional segmentation approaches—like Sobel edge detection, Canny filters, Otsu's thresholding, and morphological operations—still find relevance in simple setups but lack the adaptability needed for complex outdoor imagery. These classical methods may serve as lightweight alternatives in resource-constrained scenarios or pre-segmentation pipelines.

3.3 Datasets for Medicinal Plant Classification

The success of any deep learning model heavily depends on the availability and quality of datasets. In the field of medicinal plant identification, several datasets have been introduced to address the diversity and specificity of regional flora.

Pushpa and Rani (2023) introduced the DIMPSAR dataset, comprising high-resolution images of Indian medicinal plants captured under varying lighting and environmental conditions. This dataset supports training of robust models adaptable to real-world variability.

The MED 117 dataset (Sarma et al., 2023) contains over 117 distinct plant species found in Assam. It includes raw RGB images and U-Net segmented grayscale versions, each annotated with botanical and common names, thus facilitating multi-label and supervised learning strategies.

Uddin et al. (2023) developed a comprehensive dataset focused on Bangladeshi medicinal plants, featuring 5,000 leaf images across 10 categories. They employed preprocessing techniques like background removal and histogram equalization to enhance visual clarity. Their evaluation of five DL models and a custom Dense Residual CNN (DRCD) revealed that model ensembles could achieve up to 99% accuracy.

Pacifico et al. (2019) provided a dataset centered on Brazilian medicinal flora, using hand-segmented leaf images to evaluate KNN, Random Forest, and Decision Tree classifiers. Despite being smaller in scale, this dataset demonstrated the efficacy of color and texture features in traditional ML pipelines.

3.4 Deep Learning-Based Classification Approaches

The backbone of intelligent medicinal plant identification systems lies in robust classification architectures. Convolutional Neural Networks (CNNs) dominate this space due to their high performance in visual recognition tasks.

Kavitha et al. (2023) designed a MobileNet-based system tailored for real-time plant classification. Its lightweight design ensures fast inference, making it ideal for mobile deployment. With 98.3% accuracy on six Indian herbs, it stands out in real-world usability.

Roopashree and Anitha (2021) implemented the Xception model in the DeepHerb platform. Xception's depthwise separable convolutions reduce model size without sacrificing performance, achieving 97.5% accuracy in identifying Ayurvedic plants.

Dey et al. (2024) conducted a comparative study involving VGG, ResNet, DenseNet, and Inception models on a dataset with 5,878 images. DenseNet201, due to its efficient feature reuse and gradient flow, emerged as the top performer, with 99.64% accuracy on public data and 97% on mixed field data.

Other classical models like VGG16, InceptionV3, and ResNet50 have also shown potential, albeit with higher computational demands. Ghosh et al. (2023) proposed an innovative hybrid transfer learning model combining Principal Component Analysis (PCA) with VGG16. Their model achieved a test accuracy of 95.25%, emphasizing the value of dimensionality reduction in enhancing classification performance.

4. Results and Discussion

This section presents a comparative analysis of various models and techniques based on their performance metrics, as reported in the literature. The metrics commonly used for evaluation include Accuracy, Precision, Recall, and F1-Score. These metrics provide insights into how well a classification or segmentation model performs in identifying medicinal plant species.

Among the reviewed models, DenseNet201 and Xception emerged as top performers. DenseNet201, evaluated by Dey et al. (2024), achieved a classification accuracy of 99.64% using public datasets and 97% with mixed field data. This performance is attributed to its dense connectivity pattern, which promotes feature reuse and strengthens gradient flow. Similarly, Roopashree and Anitha (2021) reported that the Xception model achieved 97.5% accuracy within the DeepHerb mobile system. Its use of depthwise separable convolutions made it computationally efficient without compromising accuracy.

The MobileNet architecture, implemented by Kavitha et al. (2023), showed promising results for mobile deployment, reaching 98.3% accuracy in real-time leaf classification. Meanwhile, hybrid approaches like the PCA-VGG16 model by Ghosh et al. (2023) demonstrated how dimensionality reduction could enhance performance while maintaining a lighter computational footprint.

Segmentation techniques have also shown varied success across datasets. U-Net remains a staple for leaf segmentation, offering consistent accuracy due to its skip-connection-based architecture. Sarma et al. (2023) applied U-Net effectively on the MED 117 dataset, enhancing classification outcomes through clean region-of-interest extraction. MSRF-Net, tested by Srivastava et al. (2021), provided superior boundary accuracy in complex scenarios by utilizing multi-scale residual feature fusion.

Despite these advancements, certain gaps in existing literature were observed. Most studies lacked ablation analysis or component-wise performance breakdown, limiting the understanding of which specific layers or features contributed most to accuracy. In addition, consistent benchmarking across models is rare—making it difficult to perform direct comparisons. Very few studies evaluated traditional machine learning models like SVMs or Decision Trees alongside deep learning counterparts under the same settings.

Furthermore, real-world applicability remains a challenge. Many high-performing models are evaluated in controlled environments or curated datasets, which may not fully represent the variability encountered in the wild. Environmental factors like lighting, shadow, leaf occlusion, and overlapping vegetation remain underexplored in large-scale deployment scenarios. Therefore, while accuracy values above 97% are encouraging, there is a strong need for robustness evaluation across diverse real-world conditions.

5. Conclusion

This review underscores the transformative potential of deep learning and image segmentation in medicinal plant identification. Models such as DenseNet201, Xception, and MobileNet have shown remarkable classification accuracy, particularly when enhanced by preprocessing techniques and robust segmentation models like U-Net and MSRF-Net. The inclusion of curated datasets such as DIMPSAR and MED 117 has further enabled these models to be trained and validated effectively. Collectively, these advancements indicate that deep learning is capable of automating plant classification with a high degree of precision.

However, challenges remain that hinder real-world application. These include a lack of standardized and diverse datasets, environmental variability in image capture, and overreliance on leaf-based features. Furthermore, many high-performing models demand significant computational resources, limiting their use in mobile or rural contexts. Future research should prioritize the development of lightweight, interpretable, and offline-capable models that integrate data from multiple plant parts. Encouraging cross-disciplinary collaboration and fostering open data initiatives will be key to creating scalable and accessible systems for global medicinal plant recognition and conservation.

REFERENCES

1. Kavitha, S., Kumar, T. S., Naresh, E., Kalmani, V. H., Bamane, K. D., & Pareek, P. K. (2023). Medicinal Plant Identification in Real-Time Using Deep Learning Model. SN Computer Science, 5(1), 73.
2. Srivastava, A., Jha, D., Chanda, S., Pal, U., Johansen, H. D., Johansen, D., ... & Halvorsen, P. (2021). MSRF-Net: a multi-scale residual fusion network for biomedical image segmentation. IEEE Journal of Biomedical and Health Informatics, 26(5), 2252-2263.
3. Geerthana, R., Nandhini, P., & Suriyakala, R. (2021). Medicinal Plant Identification Using Deep Learning. International Research Journal on Advanced Science Hub, 3, 48-53.
4. Roopashree, S., & Anitha, J. (2021). DeepHerb: a vision based system for medicinal plants using xception features. IEEE Access, 9, 135927-135941.
5. Pushpa, B. R., & Rani, N. S. (2023). DIMPSAR: Dataset for Indian medicinal plant species analysis and recognition. Data in Brief, 49, 109388.
6. Dey, B., Ferdous, J., Ahmed, R., & Hossain, J. (2024). Assessing deep convolutional neural network models and their comparative performance for automated medicinal plant identification from leaf images. Heliyon, 10(1).
7. Bodhwani, V., Acharjya, D. P., & Bodhwani, U. (2019). Deep residual networks for plant identification. Procedia Computer Science, 152, 186-194.
8. Sarma, P., Boruah, P. A., & Buragohain, R. (2023). MED 117: A dataset of medicinal plants mostly found in Assam with their leaf images, segmented leaf frames and name table. Data in Brief, 47, 108983.
9. Sodjinou, S. G., Mohammadi, V., Mahama, A. T. S., & Gouton, P. (2022). A deep semantic segmentation-based algorithm to segment crops and weeds in agronomic color images. Information Processing in Agriculture, 9(3), 355-364.

10. Acharya, B., Ghosh, A., Panda, S., & Hu, Y. C. (2023). Automated Plant Recognition System with Geographical Position Selection for Medicinal Plants. Advances in Multimedia, 2023.
11. Naeem, S., Ali, A., Chesneau, C., Tahir, M. H., Jamal, F., Sherwani, R. A. K., & Ul Hassan, M. (2021). The classification of medicinal plant leaves based on multispectral and texture feature using machine learning approach. Agronomy, 11(2), 263.
12. Rajani, S., & Veena, M. N. (2019). Medicinal plants segmentation using thresholding and edge based techniques. International Journal of Innovative Technology and Exploring Engineering, 8(6S4), 71-76.
13. Somasekhar, T., Krishna, K. R., Sushmitha, K., Tejaswini Nayaka, S., & Thirumal, R. IDENTIFICATION OF MEDICINAL PLANTS BY IMAGE PROCESSING OF LEAF SAMPLES.
14. Bose, P., Dutta, S., Goyal, V., & Bandyopadhyay, S. (2021). Leaf Diseases Detection of Medicinal Plants based on Image Processing and Machine Learning Processes.
15. Anchitaalagammai, J. V., Revathy, J. S. L., Kavitha, S., & Murali, S. (2021, November). Factors influencing the use of Deep Learning for Medicinal Plants Recognition. In Journal of Physics: Conference Series (Vol. 2089, No. 1, p. 012055). IOP Publishing.
16. Jayalath, A. D. A. D. S., Amarawanshaline, T. G. A. G. D., Nawinna, D. P., Nadeeshan, P. V. D., & Jayasuriya, H. P. (2019, December). Identification of medicinal plants by visual characteristics of leaves and flowers. In 2019 14th Conference on Industrial and Information Systems (ICIIS) (pp. 125-129). IEEE.
17. Vayadande, K., Ghadekar, P. P., Sawant, A., Shelke, M. P., Shirsath, B., Bhande, B., ... & Samgir, S. (2024). Identification of Ayurvedic Medicinal Plant Using Deep Learning. International Journal of Intelligent Systems and Applications in Engineering, 12(15s), 678-693.
18. Yoo, S., Yang, H. C., Lee, S., Shin, J., Min, S., Lee, E., ... & Lee, D. (2020). A deep learning-based approach for identifying the medicinal uses of plant-derived natural compounds. Frontiers in Pharmacology, 11, 584875.
19. KAYHAN, G., & ERGÜN, E. (2020). Medicinal and aromatic plants identification using machine learning methods. Balkan Journal of Electrical and Computer Engineering, 8(1), 81-87.
20. Pacifico, L. D., Britto, L. F., Oliveira, E. G., & Ludermir, T. B. (2019, October). Automatic classification of medicinal plant species based on color and texture features. In 2019 8th Brazilian Conference on Intelligent Systems (BRACIS) (pp. 741-746). IEEE.
21. Ghosh, S., Singh, A., & Kumar, S. (2023). Identification of medicinal plant using hybrid transfer learning. Indonesian Journal of Electrical Engineering and Computer Science, 31(3), 1605-1615.
22. Akter, R., & Hosen, M. I. (2020, December). CNN-based leaf image classification for Bangladeshi medicinal plant recognition. In 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE) (pp. 1-6). IEEE.
23. Uddin, A. H., Chen, Y. L., Borkatullah, B., Khatun, M. S., Ferdous, J., Mahmud, P., ... & Por, L. Y. (2023). Deep-learning-based classification of Bangladeshi medicinal plants using neural ensemble models. Mathematics, 11(16), 3504.
24. Begue, A., Kowlessur, V., Singh, U., Mahomoodally, F., & Pudaruth, S. (2017). Automatic recognition of medicinal plants using machine learning techniques. International Journal of Advanced Computer Science and Applications, 8(4), 166-175.
25. Duong-Trung, N., Quach, L. D., Nguyen, M. H., & Nguyen, C. N. (2019, February). A combination of transfer learning and deep learning for medicinal plant classification. In Proceedings of the 2019 4th International Conference on Intelligent Information Technology (pp. 83-90).
26. Mulugeta, A. K., & Sharma, D. P. (2024). Deep learning for medicinal plant species classification and recognition: a systematic review. Frontiers in Plant Science, 14, 1286088.
27. Singh, M. M., & Singh, T. R. (2021). A Survey on Different Methods for Medicinal Plants Identification and Classification System. Revista Geintec-gestao Inovacao e Tecnologias, 11(4), 3191-3202.
28. Xue, J., Fuentes, S., Poblete-Echeverria, C., Viejo, C. G., Tongson, E., Du, H., & Su, B. (2019). Automated Chinese medicinal plants classification based on machine learning using leaf morpho-colorimetry, fractal dimension and visible/near infrared spectroscopy. *International Journal of Agricultural and Biological Engineering*, *12*(2), 123-131.
29. Prasvita, D. S., & Herdiyeni, Y. (2013). MedLeaf: mobile application for medicinal plant identification based on leaf image. International Journal on Advanced Science, Engineering and Information Technology, 3(2), 5-8.
30. Thella, P. K., & Venugopal, U. (2021). A Group Labelled Classification Model for Accurate Medical Plant Detection Used in Drug Preparation. Revue d'Intelligence Artificielle, 35(2).