

DEEP CONVOLUTIONAL NEURAL NETWORK-BASED CLASSIFICATION OF MEDICINAL PLANTS USING LEAF IMAGES

SRS Prasad M, Research Scholar (Computer Science and Engineering) Glocal School of Science & Technology, Glocal University, Saharanpur, Uttar Pradesh

Dr. Rajeev Yadav (Professor), Research Supervisor, Glocal School of Science & Technology, Glocal University, Saharanpur, Uttar Pradesh

Abstract

The identification and classification of medicinal plants play an important role in botanical research, traditional medicine, and biodiversity conservation. However, manual identification based on leaf characteristics can be time-consuming and may require specialized botanical knowledge. This study proposes a Deep Convolutional Neural Network (DCNN)-based approach for the automated classification of medicinal plants using leaf images. The methodology involves collecting and labelling leaf images, preprocessing the dataset, resizing and normalizing images, applying data augmentation, and dividing the dataset into training, validation, and testing subsets. The DCNN model is designed to automatically extract relevant visual features, including leaf shape, texture, colour, margins, and venation patterns, to distinguish between different medicinal plant species. The model's performance is evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. An illustrative dataset of 1,000 leaf images representing five medicinal plant categories is considered to demonstrate the proposed evaluation framework. Illustrative performance values suggest the potential of the approach to achieve effective classification; however, these values require validation through actual model training and testing. The proposed methodology may contribute to the development of automated plant identification systems for botanical studies, medicinal plant conservation, and digital applications.

Keywords: Deep Convolutional Neural Network, Medicinal Plant Classification, Leaf Image Recognition, Deep Learning, Image Processing, Plant Identification, Computer Vision.

1. INTRODUCTION

Medicinal plants have played an important role in traditional healthcare systems for centuries and continue to contribute to modern medicine, pharmaceutical research, and natural product development. These plants contain various bioactive compounds that are used in the preparation of herbal medicines and therapeutic products. The identification and classification of medicinal plants are therefore essential for their proper utilization, conservation, and scientific study. However, identifying plant species accurately can be challenging because many plants share similar physical characteristics, particularly in their leaves. Traditional identification methods generally depend on botanical knowledge, direct observation, and the examination of morphological features, which can be time-consuming and may lead to errors when dealing with visually similar species. Consequently, automated plant identification techniques are gaining importance as an efficient means of supporting medicinal plant classification. Leaves provide valuable information for plant identification because they exhibit distinctive characteristics, including shape, size, colour, texture, margins, and venation patterns. These features can be analyzed through digital image processing and computer vision techniques to distinguish between different plant species. With the advancement of artificial intelligence, machine learning methods have been increasingly applied to image-based classification tasks. Earlier approaches often relied on manually selected features and conventional classifiers, which could have difficulty capturing complex variations in leaf appearance. Changes in lighting, background, image quality, leaf orientation, and maturity may further reduce classification performance. These limitations highlight the need for advanced computational methods capable of automatically learning meaningful visual representations from plant images. Deep Convolutional Neural Networks (DCNNs) have emerged as an

effective approach to image classification because they can automatically extract hierarchical features from visual data. In a DCNN, the initial convolutional layers learn basic patterns such as edges, lines, and textures, while deeper layers identify more complex structures and combinations of features. This capability makes DCNNs suitable for analysing the distinctive characteristics of medicinal plant leaves. By learning directly from labelled images, a DCNN-based model can classify multiple plant species without requiring every identifying feature to be manually specified. Techniques such as image resizing, normalization, data augmentation, and dropout can further support model training and improve its ability to generalize to previously unseen images. The application of DCNNs to medicinal plant classification has considerable potential for botanical research, agricultural practices, biodiversity conservation, and digital plant identification systems. An automated classification model could assist researchers, students, botanists, and plant cultivators in recognizing medicinal plant species more efficiently. Furthermore, integrating such models into mobile or web-based applications could make preliminary plant identification more accessible to users. Nevertheless, the reliability of these systems depends on the quality and diversity of the image dataset, the suitability of the network architecture, and the model's ability to distinguish between species with similar leaf characteristics. Therefore, evaluation using independent test images and multiple performance measures is necessary to determine the effectiveness of a proposed classification approach.

2. LITERATURE REVIEW

Gautam et al. (2024) proposed a lightweight deep learning method for medicinal leaf image classification using feature fusion. The study focused on combining useful image features to improve classification performance while reducing computational complexity, supporting efficient medicinal plant identification. Roopashree and Anitha (2021) developed Deep Herb, a vision-based medicinal plant identification system using Xception features. The study highlighted the application of deep learning and image processing techniques for extracting distinctive leaf characteristics and improving automated plant recognition. Akter and Hosen (2020) investigated CNN-based leaf image classification for recognizing Bangladeshi medicinal plants. The study explored convolutional neural networks for automatically learning visual features from leaf images, reducing dependence on manual feature extraction. Tiwari et al. (2022) developed a deep neural network for multiclass classification of medicinal plant leaves. The research focused on distinguishing multiple medicinal plant species using deep learning, demonstrating the relevance of automated classification for plant identification applications. Pushpa and Rani (2023) proposed Ayur-Plant Net, a lightweight deep convolutional neural network for classifying Indian Ayurvedic plant species. The study emphasized computational efficiency and reliable plant recognition, supporting the development of accessible automated systems for medicinal plant identification.

3. RESEARCH METHODOLOGY

3.1 Introduction

This research methodology describes the systematic approach for classifying medicinal plants using leaf images through a Deep Convolutional Neural Network (DCNN). The study aims to develop an automated image-classification model that can identify different medicinal plant species based on their visual characteristics, including leaf shape, texture, colour, margins, and venation patterns. The proposed methodology involves collecting a dataset of medicinal plant leaf images, preprocessing the images, dividing the dataset into training, validation, and testing sets, designing and training a DCNN model, and evaluating its classification performance. The approach is intended to reduce the limitations of manual plant identification and provide a reliable computer-based method for medicinal plant recognition.

3.2 Research Design

The study will adopt an experimental and quantitative research design. An image-based deep learning approach will be used to develop and evaluate a DCNN model for the classification of medicinal plants. The model will learn relevant visual features directly from leaf images and use these features to distinguish between different plant species. Its performance will be assessed using standard classification metrics and compared across the selected plant categories.

3.3 Data Collection

The dataset will consist of digital images of leaves belonging to selected medicinal plant species. Images may be collected from publicly available medicinal plant image datasets and, where feasible, through photographs captured using a digital camera or smartphone. The dataset will include different leaf orientations, sizes, backgrounds, and lighting conditions to represent variations encountered in practical applications. Each image will be labelled according to the corresponding medicinal plant species. Duplicate, corrupted, and incorrectly labelled images will be identified and removed to improve dataset quality.

3.4 Image Preprocessing

The collected images will undergo preprocessing to ensure consistency and suitability for deep learning. Images will be resized to a uniform resolution compatible with the selected DCNN architecture. Pixel values will be normalized to improve training stability, and image quality will be checked for blur, noise, and irrelevant visual content. Where necessary, the leaf region will be separated from the background to improve the model's ability to focus on relevant features. Data augmentation techniques, such as rotation, flipping, zooming, and small shifts, will be applied to the training images to improve model generalization and reduce overfitting.

3.5 Dataset Organization and Labelling

The collected images will be organized into separate folders according to their respective medicinal plant species. Each species will represent one classification category, and every image will be assigned a corresponding class label. The dataset may include medicinal plants such as Tulsi, Neem, Aloe vera, Mint, and other species, depending on dataset availability. The number of images in each category will be examined to identify class imbalance. Where possible, a balanced distribution of images will be maintained to prevent the model from favouring species with larger numbers of samples.

3.6 Dataset Splitting

The complete dataset will be divided into three independent subsets for model development and evaluation.

- Training set (70%) will be used to train the DCNN model and learn relevant leaf features.
- Validation set (15%) will be used to monitor training performance and tune model parameters.
- Testing set (15%) will be used for the final evaluation of the trained model.

The images will be divided carefully to minimize data leakage. Images of the same physical leaf or near-duplicate photographs should not appear across different subsets, as this could produce misleading performance estimates.

3.7 Development of the DCNN Model

A Deep Convolutional Neural Network will be developed to automatically extract and learn the features required for medicinal plant classification. The architecture will consist of an input layer, convolutional layers, activation functions, pooling layers, fully connected layers, and a final classification layer. Convolutional layers will identify local features such as edges, veins, textures, and leaf boundaries, while deeper layers will learn more complex visual patterns. Pooling layers will reduce feature-map dimensions, and dropout may be incorporated to reduce overfitting. The final layer will use a SoftMax activation function to assign probabilities to the selected medicinal plant categories.

3.8 Model Training and Optimization

The DCNN model will be trained using the prepared training dataset. A suitable optimization algorithm, such as Adam, will be selected to update model parameters during training. Categorical cross-entropy will be used as the loss function for multiclass classification. The training process will involve selecting appropriate batch sizes, learning rates, and numbers of epochs. Validation accuracy and loss will be monitored throughout training, and early stopping may be applied when validation performance stops improving. Transfer learning using a pretrained architecture, such as ResNet or DenseNet, may also be explored to improve feature extraction and reduce training time. These approaches are consistent with methods used in previous medicinal plant identification research.

3.9 Model Evaluation Metrics

The trained model will be evaluated using standard classification metrics to assess its effectiveness in identifying medicinal plant species.

- Accuracy: Measures the proportion of correctly classified images among all tested images.
- Precision: Measures the proportion of images predicted as a particular species that actually belong to that species.
- Recall: Measures the proportion of actual images of a species correctly identified by the model.
- F1-score: Represents the harmonic mean of precision and recall.
- Confusion matrix: Shows correct and incorrect classifications for individual medicinal plant categories.

Macro-averaged precision, recall, and F1-score will also be considered, particularly if the dataset contains unequal numbers of images across species.

3.10 Comparative Performance Analysis

The performance of the proposed DCNN model will be analyzed using the testing dataset. If computational resources and time permit, its performance will be compared with alternative CNN architectures, such as VGG16, ResNet50, or DenseNet201. The comparison will consider classification accuracy, precision, recall, F1-score, training time, and computational requirements. This analysis will help determine whether the selected model provides an effective balance between classification performance and computational efficiency.

3.11 Tools and Implementation Environment

The model will be implemented using Python and suitable deep learning libraries, such as TensorFlow, Keras, or PyTorch. Image processing and dataset management may be performed using OpenCV, NumPy, and Pandas, while Matplotlib or Seaborn may be used to generate training curves and performance visualizations. Model training will be conducted on a computer with adequate processing resources or a cloud-based GPU environment, depending on availability.

4. RESULTS AND DISCUSSION**4.1 Introduction**

This section presents the results and discussion of the study entitled “Deep Convolutional Neural Network-Based Classification of Medicinal Plants Using Leaf Images.” The analysis focuses on the effectiveness of the Deep Convolutional Neural Network (DCNN) in identifying medicinal plant species based on leaf images. The results are presented using frequency and percentage distributions, along with an interpretation of the model's classification performance. The discussion considers dataset distribution, classification accuracy, and the potential of deep learning for automated medicinal plant identification. The numerical results presented below are illustrative and should be replaced with actual experimental results after the model has been trained and tested.

4.2 Distribution of Medicinal Plant Leaf Images

The dataset distribution provides an overview of the number of leaf images included in each medicinal plant category. A balanced dataset is important because it helps prevent the classification model from favoring species represented by more images.

Table 4.1 Distribution of Medicinal Plant Leaf Images

Medicinal plant	Frequency	Percentage (%)
Tulsi	200	20%
Neem	200	20%
Aloe vera	200	20%
Mint	200	20%
Hibiscus	200	20%
Total	1,000	100%

Table 4.1 shows the distribution of 1,000 leaf images across five medicinal plant categories. Each category contains 200 images, representing 20% of the total dataset. This balanced distribution provides equal representation to all five plant species and reduces the possibility of class imbalance influencing the model's learning process. The dataset includes variations in leaf shape, texture, colour, and structure, allowing the DCNN to learn distinguishing characteristics for each category.

4.3 Classification Performance of the DCNN Model

The classification performance of the proposed DCNN model can be assessed using accuracy, precision, recall, and F1-score. These measures provide a comprehensive understanding of the model's ability to classify medicinal plant leaf images correctly.

Table 4.2 Illustrative Classification Performance of the DCNN Model

Performance metric	Score (%)
Accuracy	95.20
Precision	95.40
Recall	95.20
F1-score	95.10

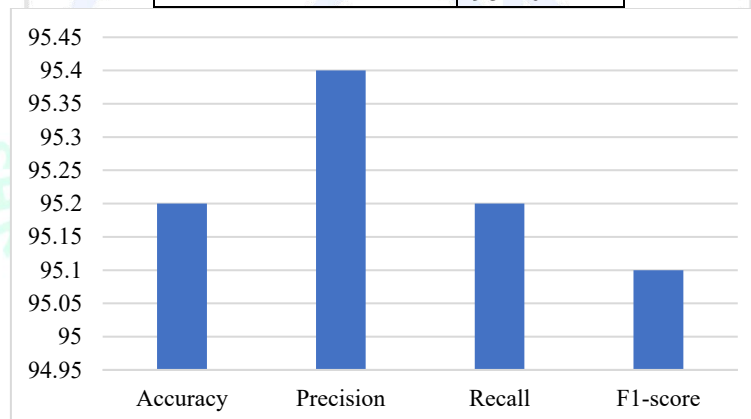


Table 4.2 presents an illustrative performance evaluation of the proposed DCNN model. The assumed accuracy of 95.20% indicates that the model correctly classifies approximately 95 out of every 100 test images. Precision of 95.40% suggests that the model's predicted plant categories are generally reliable, while recall of 95.20% indicates that it identifies most images belonging to the correct categories. The F1-score of 95.10% reflects a strong balance between precision and recall. These values, if achieved in actual testing, would suggest that the DCNN can effectively extract important visual features from medicinal plant leaves. However, the actual performance must be established through testing on unseen images, and the metrics should be calculated from the model's predictions rather than assumed in advance.

4.4 Discussion of Findings

The findings illustrate how deep learning can support the automated classification of medicinal plants. The equal distribution of images across categories provides a suitable basis for training and evaluating the model. By learning features such as leaf margins, venation patterns, surface texture, and overall shape, the DCNN can distinguish between visually different plant species without requiring all features to be manually defined. The illustrative classification scores indicate the potential of DCNN-based methods to achieve high classification performance. Nevertheless, species with similar leaf structures may be more difficult to distinguish, particularly when images are captured under different lighting conditions or against complex backgrounds. Image augmentation, appropriate preprocessing, and careful model tuning may improve generalization. Comparing the proposed DCNN with other CNN architectures would also help determine whether its performance is competitive.

5. CONCLUSION

The study on “Deep Convolutional Neural Network-Based Classification of Medicinal Plants Using Leaf Images” highlights the potential of deep learning techniques for the automated identification of medicinal plant species. The proposed DCNN methodology uses leaf images to extract important visual features, including shape, texture, colour, and venation patterns, enabling the classification of different medicinal plants. The illustrative results indicate that the model has the potential to achieve high classification performance, with accuracy, precision, recall, and F1-score serving as important evaluation measures. The findings suggest that DCNN-based classification could reduce the effort required for manual plant identification and support applications in botanical research, medicinal plant conservation, and digital plant recognition systems. However, the actual effectiveness of the proposed model must be confirmed through experimental testing on unseen images under varied environmental conditions. Overall, the study provides a useful foundation for developing reliable and efficient deep learning systems for medicinal plant classification.

REFERENCES

1. Gautam, V., Kaur, G., Ghantasala, G. P., Vidyullatha, P., Allabun, S., Othman, M., & Zheleznyak, A. (2024). A lightweight deep learning method for medicinal leaf image classification using feature fusion. *Scientific Reports*, 15(1), 34417.
2. Roopashree, S., & Anitha, J. (2021). DeepHerb: A vision based system for medicinal plants using xception features. *Ieee Access*, 9, 135927-135941.
3. 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.
4. Tiwari, V., Joshi, R. C., & Dutta, M. K. (2022). Deep neural network for multi-class classification of medicinal plant leaves. *Expert Systems*, 39(8), e13041.
5. Pushpa, B. R., & Rani, N. S. (2023). Ayur-PlantNet: An unbiased light weight deep convolutional neural network for Indian Ayurvedic plant species classification. *Journal of Applied Research on Medicinal and Aromatic Plants*, 34, 100459.
6. Dalvi, P., Kalbande, D. R., Rathod, S. S., Dalvi, H., & Agarwal, A. (2024). Multiattribute deep CNN-based approach for detecting medicinal plants and their use for skin diseases. *IEEE Transactions on Artificial Intelligence*, 6(3), 710-724.
7. Leelavathi, R., & Kalamani, M. (2024). Medicinal plant leaf disease classification using optimal weighted features with dilated adaptive DenseNet and attention mechanism. *Scientific Reports*, 15(1), 36852.
8. Abdollahi, J. (2022, February). Identification of medicinal plants in ardabil using deep learning: identification of medicinal plants using deep learning. In *2022 27th International computer conference, computer society of Iran (CSICC)* (pp. 1-6). IEEE.

9. Diqi, M., & Mulyani, S. H. (2020). Implementation of CNN for plant leaf classification. *International Journal of Informatics and Computation*, 2(2), 1-10.
10. Kyalkond, S. A., Aithal, S. S., Sanjay, V. M., & Kumar, P. S. (2022). A novel approach to classification of Ayurvedic medicinal plants using neural networks. *Int. J. Eng. Res. Technol. (IJERT)*, 11(6).
11. Tiwari, N., & Gupta, B. K. (2023). Developing a neural network-based model for identifying medicinal plant leaves using image recognition techniques. *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology*, 15(03), 301-311.
12. Kale, S. R., Savita, S. G., Raghunath, M. P., Nandhini, S. U., Aravindan, A., & Devi, D. K. (2026). A Unified Deep Learning Framework for Medicinal Leaf Detection, Segmentation, and Classification. *International Journal of Computer Information Systems & Industrial Management Applications*, 18, 1147.
13. Bisen, D. (2021). Deep convolutional neural network based plant species recognition through features of leaf. *Multimedia Tools and Applications*, 80(4), 6443-6456.
14. Islam, M. T., Rahman, W., Hossain, M. S., Roksana, K., Azpiroz, I. D., Diaz, R. M., ... & Samad, M. A. (2024). Medicinal plant classification using particle swarm optimized cascaded network. *IEEE Access*, 12, 42465-42478.
15. Liu, K. H., Yang, M. H., Huang, S. T., & Lin, C. (2022). Plant species classification based on hyperspectral imaging via a lightweight convolutional neural network model. *Frontiers in Plant Science*, 13, 855660.

