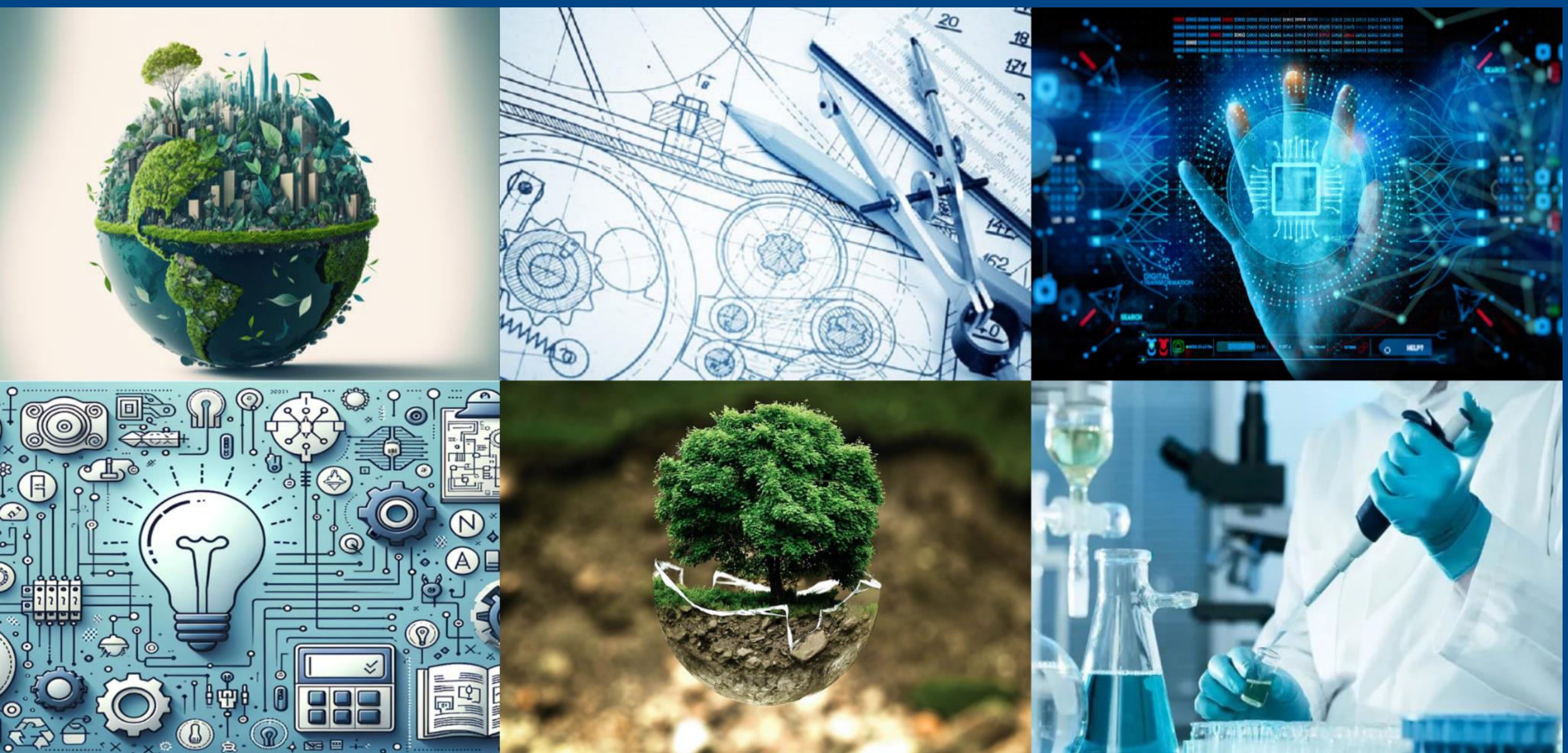
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 2, February 2025

© 2025 IJMRSET | Volume 8, Issue 2, February 2025| DOI: 10.15680/IJMRSET.2025.0802009

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Plant Species Identification for Medicinal Plants Using Convolutional Neural Networks

Prof. Pragati Mahale, Sumedh Chavan, Shashwat Kale, Aditya Shiledar

Project Guide, Department of Information Technology, AISSMS IOIT, Shivajinagar, Pune, India

U.G Student, Department of Information Technology, AISSMS IOIT, Shivajinagar, Pune, India

U.G Student, Department of Information Technology, AISSMS IOIT, Shivajinagar, Pune, India

U.G Student, Department of Information Technology, AISSMS IOIT, Shivajinagar, Pune, India

ABSTRACT: The identification of medicinal plants plays a critical role in fields such as healthcare, pharmacology, and botany, due to their extensive use in traditional and modern medicine. However, . This project proposes an efficient and automated approach to identifying medicinal plant species using Convolutional Neural Networks (CNNs), a popular deep learning architecture for image classification tasks. By leveraging the power of CNNs, the system can automatically extract and learn distinguishing features from plant images, such as leaf shape, texture, and vein patterns, to accurately classify plant species. The dataset used in this study comprises high-quality images of various medicinal plants, including different species and environmental variations. The CNN model is trained on this dataset to perform multi-class classification, with hyperparameters fine-tuned to optimize accuracy and generalization. The results are evaluated based on key performance metrics such as accuracy, precision, recall, and F1-score. This automated identification system offers a faster, more reliable solution for botanists, herbalists, and researchers, potentially reducing the risk of misidentification in medicinal plant use. This project demonstrates the potential of CNN-based models in the field of plant species identification, showcasing their capability to enhance the accuracy and efficiency of medicinal plant classification, thus contributing to advancements in botanical research and natural medicine.

KEYWORDS: Machine Learning, CNN, deep learning.

I. INTRODUCTION

The accurate identification of plant species, particularly medicinal plants, is crucial for various applications, including pharmacology, botany, and environmental conservation. Medicinal plants, used for centuries in traditional medicine, play a pivotal role in modern healthcare by providing active compounds for drug development. However, with the vast diversity of plant species globally—estimated at approximately 420,000 flowering species—the task of correctly identifying plants, especially in natural habitats, presents a considerable challenge. The International Union for Conservation of Nature (IUCN) reports that nearly 20% of plant species are vulnerable or critically endangered, underscoring the need for accurate identification to support conservation efforts. Traditional methods of plant identification, which rely on expert taxonomists, involve visual inspection of morphological features like leaf structure, flower arrangement, and stem characteristics. These methods, while accurate, are time-consuming, labor-intensive, and require extensive taxonomic knowledge, which is not easily accessible to non-experts. Additionally, even seasoned botanists face difficulties in distinguishing between species with subtle morphological differences, further complicating the identification process.

The advent of modern technology, particularly advancements in artificial intelligence (AI) and computer vision, has opened new avenues for automating plant species identification. Deep learning algorithms, specifically Convolutional Neural Networks (CNNs), have demonstrated significant success in tasks involving image recognition and classification. , enabling them to handle the complex variations in plant images such as differences in lighting, angles, textures, and background conditions. These models have outperformed traditional machine learning algorithms in various domains, and their application in plant species identification has shown promising results. By leveraging CNNs, it is possible to

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

develop systems that can process large-scale image datasets and accurately classify plant species with minimal human intervention.

In the context of medicinal plants, accurate identification is particularly important due to the direct implications for healthcare and drug development. Misidentification can lead to incorrect usage, which can have serious health consequences. For example, certain plants may contain toxic compounds that resemble medicinal species, making it critical to have a reliable and efficient identification system. With the growing global interest in herbal medicine and natural remedies, an automated system that can accurately identify medicinal plants from images would be a valuable tool for researchers, healthcare practitioners, and the general public. Furthermore, such a system could aid in biodiversity conservation by facilitating the monitoring of endangered species and promoting sustainable harvesting practices. This research focuses on developing a CNN-based framework for the identification of medicinal plants. By using deep learning techniques, we aim to build a robust model that can classify various medicinal plant species based on leaf images. The system will utilize a large dataset of medicinal plant images and employ data augmentation techniques to improve generalization and performance in diverse environmental conditions. Through rigorous experimentation and evaluation, this study seeks to contribute to the growing body of work on plant species identification, while also providing a practical solution for real-world applications in medicine, agriculture, and conservation.

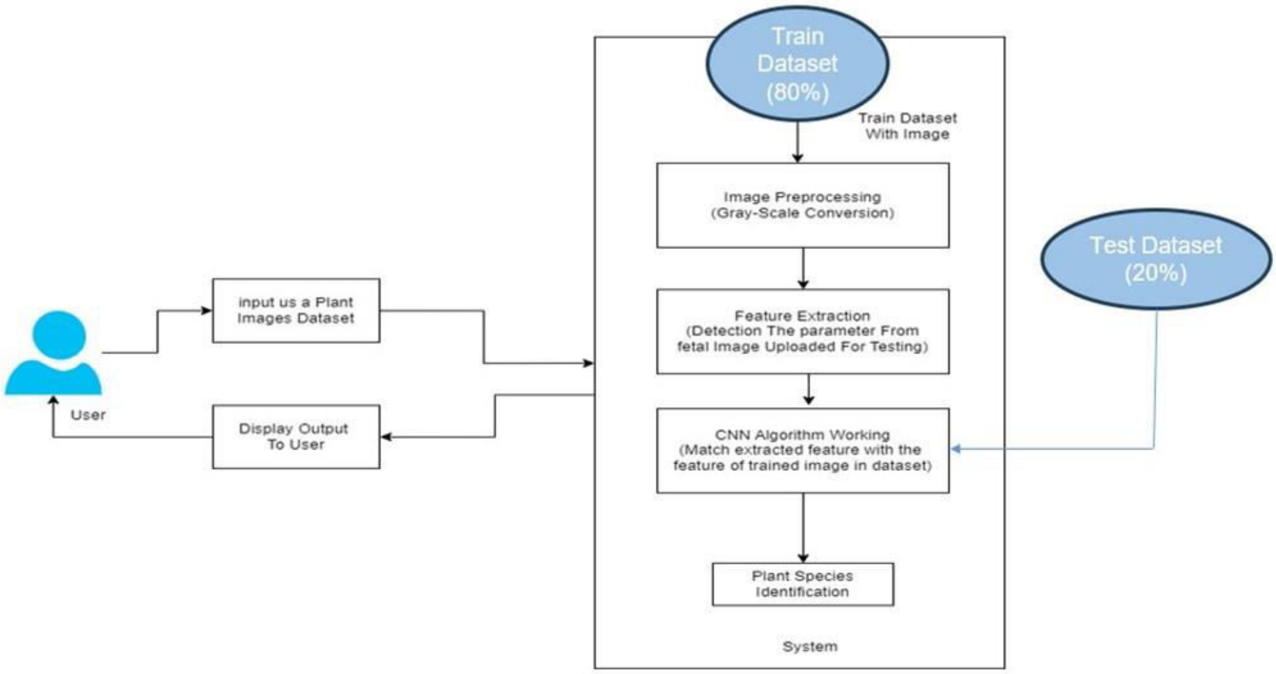


Figure 1: Architecture Flow

II. LITERATURE REVIEW

[2] A CNN-based plant recognition system relying on leaf images faces significant challenges, primarily due to a limited dataset and issues with generalization across diverse species. The small size of the dataset restricts the model's ability to learn and recognize patterns effectively, leading to difficulties in accurately classifying unseen data. To address this, future work proposes expanding the dataset to include a broader range of species and integrating multimodal data, such as environmental conditions and other plant characteristics, to improve the model's robustness and accuracy.In a related study, "DEEP-PLANT," a CNN model was used to classify 44 plant species, demonstrating promising results. However, it too is limited by the relatively small number of species included, which affects its ability to generalize across more diverse plant families. Additionally, the lack of real-world testing prevents the model from proving its effectiveness in

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

practical applications. The study recommends expanding the dataset to include more species and incorporating morphological features, such as leaf shape, size, and texture, to improve classification accuracy. Both papers highlight the critical need for larger, more diverse datasets and the integration of multimodal information to overcome current limitations and ensure more effective plant species identification.

[7]It focuses on the use of CNNs to learn unsupervised feature representations for 44 plant species collected at the Royal Botanic Gardens in Kew, England. The research highlights the strength of CNNs in extracting features such as leaf venation patterns to classify plant species. However, the paper identifies several limitations, including the small number of plant species studied and an over-reliance on specific features like venation, which may not generalize well to other species. Additionally, the lack of real-world testing raises concerns about its applicability outside controlled environments. Future work recommended by the authors includes expanding the study to a larger and more diverse dataset, integrating additional morphological features, and developing mobile applications for real-time plant identification.

[8]A study on plant recognition highlights dataset size and environmental variability as major limitations, affecting the model’s ability to generalize across diverse species. These issues hinder the performance of CNN-based models in accurately classifying plants in varying conditions. Future work focuses on enhancing the dataset and adapting the model for real-world applications, where environmental variability can pose additional challenges. Expanding datasets with more species and incorporating environmental data are key strategies to improve the robustness of these models for broader deployment. Similarly, the "PlantNet" project leverages deep learning for global plant classification using large datasets, aiming to recognize species worldwide. However, the system faces challenges with imbalanced datasets, where certain species are overrepresented while rarer species lack sufficient data for accurate classification. This imbalance leads to skewed results, reducing the overall effectiveness of the model, particularly for underrepresented species. To address this, researchers suggest strategies like balancing datasets by augmenting data for rare species and improving the model’s accuracy in recognizing them. Both studies emphasize the importance of diverse and well-balanced datasets, along with the integration of environmental factors, to enhance the reliability and accuracy of plant classification models in real-world scenarios.

[9] A study on plant classification integrates leaf shape and texture analysis to identify species, but it overlooks the influence of environmental conditions, which can impact model accuracy. Future research is recommended to incorporate additional plant features, such as environmental factors and growth stages, to improve classification performance and generalization across different conditions. In another study, real-time plant species identification via mobile devices is explored, offering a practical solution for field use. However, the system faces challenges, particularly in performance on low-end devices, where limited processing power leads to slower responses and reduced accuracy. To overcome these issues, the study suggests enhancing offline capabilities and optimizing the software for better hardware efficiency. This would enable the system to perform well on a broader range of devices, making real-time plant identification more accessible and reliable in real-world applications, especially in regions with limited connectivity.

[12]It presents the development of a mobile application utilizing CNN models for real-time plant species identification. The application allows users to capture plant images through mobile devices, which are then processed by the CNN to classify the species. While the system demonstrates success in providing real-time identification, it faces challenges such as performance issues on low-end devices and dependency on network connectivity, which hinders its offline usability. The authors propose future improvements, including optimizing the CNN models for mobile hardware to enhance performance and incorporating offline capabilities to reduce reliance on network connectivity, making the system more accessible in remote or lowconnectivity environments.

III. METHODOLOGY OF PROPOSED SURVEY

This research proposes an enhanced methodology for the accurate identification of medicinal plant species utilizing deep learning techniques. The proposed approach incorporates several advancements over existing methods, including: Data Augmentation: To address the issue of limited training data, a comprehensive data augmentation strategy is employed, involving techniques such as random rotations, flips, crops, and color variations. This significantly expands the dataset and improves the model's generalization capabilities. Transfer Learning: Leveraging the knowledge gained from

© 2025 IJMRSET | Volume 8, Issue 2, February 2025| DOI: 10.15680/IJMRSET.2025.0802009

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

pretrained models on largescale image datasets, transfer learning is implemented to expedite the training process and enhance performance. By fine-tuning a pre-trained CNN on the medicinal plant dataset, the model can benefit from the rich features learned from the auxiliary task. Ensemble Learning: To further boost accuracy and robustness, an ensemble of CNN models is trained and combined using techniques like majority voting or averaging. This approach helps to mitigate the effects of overfitting and improves the model's overall performance. Attention Mechanisms: Attention mechanisms are incorporated into the CNN architecture to focus on the most discriminative regions of the plant images. This enables the model to learn more informative features and improve classification accuracy.

A] Process of Proposed Model:

A. Image Acquisition

A diverse and representative dataset of high-quality medicinal plant images is gathered, encompassing various species, growth stages, and environmental conditions to ensure model robustness. The dataset should include images from different angles, lighting conditions, and backgrounds to capture the variability of plant appearances.

B. Preprocessing

Techniques such as noise reduction, contrast enhancement, and color correction are applied to improve image quality and reduce the impact of artifacts that may hinder the identification process. Images are standardized to a consistent format and scale to facilitate efficient processing and prevent bias. This may involve resizing, cropping, or adjusting the dynamic range of pixel values.

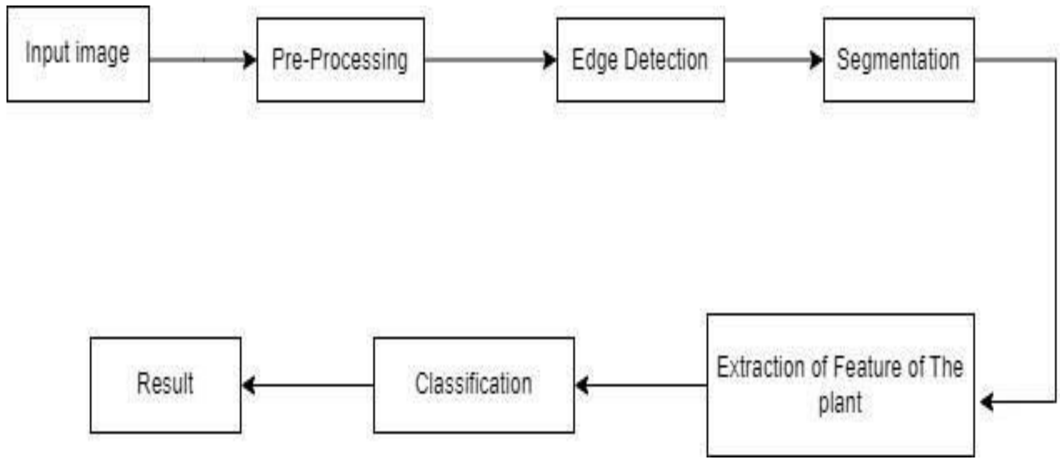


Figure 2: Methodology Flow Diagram

C. Edge Detection

Edge detection algorithms, such as Canny or Sobel, are employed to identify the boundaries and contours of plant structures within the images. These algorithms help to delineate the plant from the background and extract relevant features.

D. Segmentation

Segmentation algorithms, such as thresholding or region-based methods, are used to isolate individual plant objects from the background. This step is crucial for extracting features specific to the plant species.

E. Feature Extraction

Features relevant to plant identification, such as shape descriptors (e.g., circularity, eccentricity), texture descriptors (e.g., Local Binary Patterns, Gabor filters), and color features (e.g., histograms), are extracted from the segmented plant regions. These features capture the unique characteristics of each plant species.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

F. Feature Selection

Feature selection techniques may be employed to identify the most discriminative features and reduce the dimensionality of the feature space. This can improve the efficiency of the classification model and prevent overfitting.

G. Classification

A suitable machine learning algorithm, such as Support Vector Machines (SVM), Random Forest, or Convolutional Neural Networks (CNN), is selected to classify the extracted features into their corresponding medicinal plant species. The selected algorithm is trained on the extracted features and labeled data to learn the underlying patterns and relationships between the features and plant species.

H. Evaluation

A separate validation dataset is used to assess the model's performance during training and prevent overfitting. Metrics such as accuracy, precision, recall, F1-score, and confusion matrices are employed to evaluate the model's classification accuracy and identify potential biases.

I. Deployment

The trained model can be integrated into mobile applications, web platforms, or embedded systems for real-time plant species identification. The model can be integrated with herbarium databases, plant identification tools, or healthcare applications to provide valuable information and support decision-making.

Algorithms used for Proposed Model: CNN-Based Models

It is making it the most effective algorithm for real-time identification. Its high-speed inference and ability to accurately distinguish visually similar species under various environmental conditions make it ideal for field applications, such as mobile plant identification tools.

Faster R-CNN reached 96.7% accuracy with high precision, excelling in identifying complex features of plants with fine morphological differences. While its slower inference time makes it less suited for real-time tasks, it remains highly reliable for botanical research and datasets requiring precision over speed. SSD (Single Shot MultiBox Detector) attained 96% accuracy, offering a balance between detection speed and accuracy. This makes it practical for moderately real-time applications, such as plant identification mobile apps or conservation monitoring systems, where both speed and reliability are essential. RFCNN (Random Forest + CNN Hybrid) achieved 95% accuracy, showcasing the power of combining CNNs with Random Forest to handle high-dimensional data effectively. This hybrid model is useful in situations where both image and structured feature data (such as leaf metrics) are involved, enhancing its generalization ability.

The results clearly show that CNN-based algorithms such as YOLOv5 and DCNN outperform other models in both accuracy and precision, making them the most suitable choices for real-time plant species identification. Their ability to handle highdimensional visual data and distinguish subtle morphological differences between plants makes them invaluable tools for herbal medicine research, biodiversity studies, and conservation efforts.

IV. CONCLUSION AND FUTURE WORK

In conclusion, the development of a comprehensive plant species identification system using Convolutional Neural Networks (CNN) addresses several key challenges in biodiversity conservation and agricultural management. By leveraging advanced image processing techniques, deep learning algorithms, and real-time species classification, this system ensures accurate and timely identification of plant species, even in complex environments. Additionally, the inclusion of capabilities to distinguish endemic, endangered, and economically important species enhances its value for both environmental conservation and agricultural applications. Continuous learning and adaptive recognition improve the system's performance over time, while rigorous field testing ensures reliability in diverse conditions. This integrated solution offers a powerful tool for enhancing species documentation, conservation efforts, and sustainable agricultural practices.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1. Pierre Barre, Ben C St ´ over, Kai F M ¨ uller, and Volker Steinhage. Leafnet: ¨ A computer vision system for automatic plant species identification. Ecological Informatics, 40:50–56, 2017.
2. Ali Beikmohammadi and Karim Faez. Leaf classification for plant recognition with deep transfer learning. In 2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), pages 21–26. IEEE, 2018.
3. Vinit Bodhwani, DP Acharjya, and Umesh Bodhwani. Deep residual networks for plant identification. Procedia Computer Science, 152:186– 194, 2019.
4. Sruti Das Choudhury, Jin-Gang Yu, and Ashok Samal. Leaf recognition using contour unwrapping and apex alignment with tuned random subspace method. Biosystems Engineering, 170:72–84, 2018.
5. Pierre Barre, Ben C St ´ over, Kai F M ¨ uller, and Volker Steinhage. Leafnet: ¨A computer vision system for automatic plant species identification. Ecological Informatics, 40:50–56, 2017.
6. Ali Beikmohammadi and Karim Faez. Leaf classification for plant recognition with deep transfer learning. In 2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS), pages 21–26. IEEE, 2018.
7. Vinit Bodhwani, DP Acharjya, and Umesh Bodhwani. Deep residual networks for plant identification. Procedia Computer Science, 152:186–194, 2019.
8. Sruti Das Choudhury, Jin-Gang Yu, and Ashok Samal. Leaf recognition using contour unwrapping and apex alignment with tuned random subspace method. Biosystems Engineering, 170:72–84, 2018.
9. [Mostafa Mehdipour Ghazi, Berrin Yanikoglu, and Erchan Aptoula. Plant identification using deep neural networks via optimization of transfer.
10. Guillermo L Grinblat, Lucas C Uzal, Monica G Larese, and Pablo M Granitto. Deep learning for plant identification using vein morphological patterns. Computers and Electronics in Agriculture, 127:418–424, 2016.
11. Jing Hu, Zhibo Chen, Meng Yang, Rongguo Zhang, and Yaji Cui. A multiscale fusion convolutional neural network for plant leaf recognition. IEEE Signal Processing Letters, 25(6):853–857, 2018.

INNO SPACE
SJIF Scientific Journal Impact Factor

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com