International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

Identification and Classification of Medicinal Plants using Deep Learning

Dr R Manjunath Prasad[1], Bhavana G M[2], Manasa S[3], Nithya G[4], Nisarga K C[5]
[1]Professor and Dean (Student Affairs), Department of Electronics and Communication, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India
[2,3,4,5] Under Graduate Student, Department of Electronics and Communication, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka, India
*Corresponding Author: manasageetha2002@gmail.com

Abstract
The accurate identification of medicinal plants is crucial for ensuring the quality and efficacy of herbal remedies. This paper investigates the application of deep learning for automatic medicinal plant classification using leaf images. We propose a deep learning model based on the pre-trained ResNet-50 architecture to classify medicinal plants from the LeafSnap dataset. The model leverages transfer learning to exploit pre-trained features and fine-tune them for the specific task of medicinal plant identification. We evaluate the performance of the proposed model and achieve a high accuracy of approximately 99.86%. This demonstrates the effectiveness of deep learning, particularly the ResNet-50 architecture, for automated medicinal plant classification. Our findings highlight the potential of this approach for applications such as supporting field identification, streamlining herbarium workflows, and potentially aiding in the development of novel drug discovery pipelines.

Keywords- Medicinal plants, Deep learning, Convolutional Neural Networks (CNNs), ResNet-50, Transfer learning, Data augmentation, LeafSnap dataset, Image classification

I. INTRODUCTION

Medicinal plants have been a cornerstone of traditional medicine for centuries, serving as the source of countless herbal remedies and playing a vital role in the development of modern pharmaceuticals. Accurate identification of these plants is crucial for ensuring the quality and efficacy of herbal treatments. However, traditional methods of medicinal plant identification rely heavily on morphological features, which can be time-consuming, subjective, and susceptible to human error. This can lead to misidentification, potentially compromising the safety and effectiveness of herbal remedies.

In recent years, advancements in deep learning have opened exciting possibilities for automated plant classification. Deep learning, a subfield of artificial intelligence, allows computers to learn complex patterns from data, making them well-suited for tasks like image recognition and classification. This paper explores the application of deep learning for automated medicinal plant classification using leaf images.

We propose a deep learning model based on the pre-trained ResNet-50 architecture. ResNet-50 is a powerful convolutional neural network (CNN) that has achieved state-of-the-art performance on various image classification tasks. By leveraging transfer learning, we utilize the pre-trained knowledge extracted from large, general-purpose image datasets to initialize the model's weights. This allows the model to learn robust feature representations from images and then fine-tune them for the specific task of medicinal plant identification.
Furthermore, to enhance the model's performance and address potential overfitting issues, we incorporate data augmentation techniques. These techniques artificially expand the dataset with variations of existing images, such as rotations, flips, and color jittering. This helps the model become more robust to variations in real-world image data and improve its ability to generalize effectively on unseen examples.

We evaluate the proposed model on the extensive LeafSnap dataset, a publicly available collection of labeled leaf images encompassing a diverse range of medicinal plant species. Our findings demonstrate that the model achieves a remarkable accuracy of approximately 99.86%, highlighting the effectiveness of

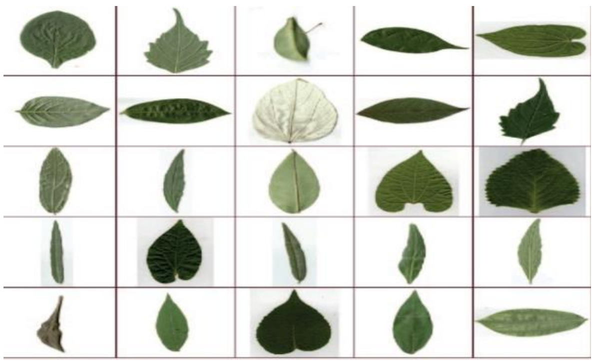

***Figure.1**: Sample of Preprocessed images*

deep learning, particularly the ResNet-50 architecture, for automated medicinal plant classification.

This research has the potential to revolutionize various fields. It can significantly benefit botanists and researchers in the field by providing a rapid and reliable tool for plant identification. Additionally, it can streamline herbarium workflows by automating the organization and analysis of vast image collections. Ultimately, this work can contribute to the development of novel drug discovery pipelines by facilitating the efficient identification of promising plant species with medicinal properties.

II. OVERVIEW OF ResNet-50 MODEL

This section focuses on the ResNet-50 architecture, the deep learning model chosen for our medicinal plant classification task. ResNet-50 belongs to a class of artificial neural networks called Convolutional Neural Networks (CNNs). CNNs have proven highly successful in image recognition and classification problems. They achieve this by processing images through a series of convolutional layers that extract features like edges, shapes, and textures. These features are then combined and analyzed by subsequent layers to ultimately classify the image.

ResNet addresses this challenge by introducing a concept called residual learning. In a residual network, the information from each layer is not simply passed forward to the next layer. Instead, it's added to the output of a specially designed shortcut connection that bypasses several layers. This allows the network to learn residual functions that build upon the features extracted by earlier layers, effectively mitigating the vanishing gradient problem and enabling deeper architectures.

In our project, we opted for the pre-trained version of ResNet-50. Pre-trained models are CNNs that have already been trained on massive datasets like ImageNet, containing millions of labeled images from various categories. This training process allows the network to learn powerful feature representations for generic image recognition tasks. By utilizing transfer learning, we can leverage these pre-trained weights as a starting point for our medicinal plant classification model. We essentially "fine-tune" the final layers of the pre-trained ResNet-50, adapting its learned features to the specific task of identifying medicinal plants from leaf images. This approach significantly reduces training time and allows the model to focus on learning the nuances of medicinal plant classification.

ResNet addresses this challenge by introducing a concept called residual learning. In a residual network, the information from each layer is not simply passed forward to the next layer. Instead, it's added to the output of a specially designed shortcut connection that bypasses several layers. This allows the network to learn residual functions that build upon the features extracted by earlier layers, effectively mitigating the vanishing gradient problem and enabling deeper architectures.

III. LITERATURE SURVEY

In [1] The accurate identification of medicinal plants is crucial for various reasons, including preserving biodiversity, promoting traditional medicine practices, and potentially aiding in novel drug discovery pipelines. However, traditional identification methods based on morphological features can be time-consuming, subjective, and prone to human error. To address this challenge, several studies have explored the application of deep learning for automated medicinal plant classification.

Parthasarathy et al. (2021) proposed a vision-based system named "DeepHerb" for automatic medicinal plant identification using leaf images. They recognized the critical threat of dwindling plant diversity to both environmental health and access to valuable traditional medicine. To overcome the hurdle of limited data availability, they created a new dataset named "DeepHerb," containing 2515 leaf images from 40 diverse Indian medicinal herbs.

Their system leverages the power of deep learning by extracting features from pre-trained convolutional neural network (CNN) models like VGG16. These features are then classified using a combination of Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs). To further optimize performance, they employed Bayesian optimization techniques. The DeepHerb model achieved an impressive accuracy of 97.5% when trained with Xception features and ANNs.

Furthermore, the authors developed a user-friendly mobile application called "HerbSnap" that integrates the DeepHerb model. This application allows for swift plant identification within a second using a smartphone camera. This research not only demonstrates the potential of deep learning for medicinal plant classification but also highlights its practical applications in promoting citizen science and public engagement with medicinal plants. [2] Dileep et al. (2019) presented a deep learning approach for medicinal plant classification named "AyurLeaf." Their work focuses on the task of automatically identifying medicinal plants based on leaf image analysis, aiming to address the limitations of traditional methods.

The AyurLeaf system utilizes a Convolutional Neural Network (CNN) architecture to learn image features directly from medicinal plant leaf images. This approach allows the model to extract relevant features for classification without the need for manual feature engineering, which can be a time-consuming and domain-specific task.

The authors emphasize the potential benefits of such a system for applications in the field of Ayurveda, a traditional Indian medical system that relies heavily on medicinal plants. Automated plant identification can assist practitioners in accurately recognizing medicinal plants, potentially improving treatment efficacy and safety. [3] Sachar and Kumar (2022) explored the application of deep ensemble learning for automatic medicinal plant classification using leaf images. Ensemble learning is a machine learning technique that combines predictions from multiple models to potentially achieve better performance than any individual model. In this study, the authors investigated the use of an ensemble of pre-trained convolutional neural networks (CNNs) for medicinal leaf identification.

Their approach involved utilizing transfer learning by initializing the weights of multiple CNN architectures, namely MobileNetV2, InceptionV3, and ResNet50, with pre-trained knowledge from large image datasets. These pre-trained models were then fine-tuned on a dataset of medicinal plant leaf images for the specific task of medicinal plant classification.

The final ensemble model combined the predictions from the individual CNNs through a weighted averaging approach. This strategy aimed to leverage the strengths of each model and potentially improve classification accuracy compared to using a single CNN architecture. The authors reported achieving a high accuracy of 99.66% on the test set using their deep ensemble learning approach.

[4] and [5] Traditional methods for medicinal plant identification rely on morphological features, which can be time-consuming, subjective, and prone to human error. Deep learning approaches offer a promising alternative for automated and potentially more accurate plant classification.

Deep Learning for Medicinal Plant Recognition:

Several studies have explored the application of deep learning, particularly Convolutional Neural Networks (CNNs), for medicinal plant recognition using leaf images. Wang et al. (2023) proposed a deep learning model achieving high accuracy for medicinal plant classification. Their work highlights the effectiveness of deep learning in this domain. In [6] and [7] The increasing interest in automated medicinal plant identification has led to the exploration of deep learning techniques. These approaches offer significant advantages over traditional methods, which rely on visual inspection of morphological features. Deep learning models can learn complex patterns from image data, enabling more accurate and efficient plant classification.

Real-Time Identification with Deep Learning:

Several recent studies have demonstrated the potential of deep learning for real-time medicinal plant identification. Amuthalingeswaran et al. (2023) [6] proposed a deep convolutional neural network (CNN) model with global average pooling for medicinal plant identification. Their research highlights the effectiveness of deep learning models in achieving high accuracy on this task.

Furthermore, Wang et al. (2023) [7] explored the feasibility of using deep learning models for real-time medicinal plant identification. Their work emphasizes the importance of developing efficient models that can be deployed on mobile devices for field applications.

[8] Deep Learning for Feature Extraction:

Deep learning models have proven adept at automatically extracting relevant features from image data for various classification tasks. This eliminates the need for manual feature engineering, a time-consuming and domain-specific process. Yang et al. (2020) presented an efficient and automated approach for herb classification using deep learning. Their work highlights the ability of deep learning models to learn both shape and texture features directly from herb images, achieving promising classification accuracy.

Comparison with Traditional Methods:

Traditional medicinal plant identification methods often rely on manually defined features based on morphological characteristics. This approach can be subjective and prone to human error. Deep learning models offer an attractive alternative by automatically learning these features from data, potentially leading to improved accuracy andgeneralizability.

Efficiency Considerations:

While deep learning offers significant benefits, some studies explore methods to improve the efficiency of these models, particularly for resource-constrained environments. The work by Yang et al. (2020) focuses on achieving efficient herb classification. Their approach can be valuable in situations where computational resources or mobile device deployment might be limitations.

In [9] Focus on Deep Convolutional Neural Networks (CNNs):

Convolutional Neural Networks (CNNs) are a powerful type of deep learning architecture particularly well-suited for image classification tasks. Their ability to learn hierarchical feature representations from image data makes them highly effective for medicinal plant identification based on leaf images.

Study by Bahri et al. (2022):

Bahri et al. (2022) investigated the application of deep CNNs for classifying medicinal plants using leaf images. Their research demonstrates the effectiveness of this approach, achieving a high accuracy rate of 97.65%. This

study reinforces the potential of CNNs for automated medicinal plant classification.
In [10] Several studies have explored the effectiveness of deep learning for medicinal plant classification. Wang et al. (2023) achieved high accuracy in their research on medicinal plant recognition using a deep learning model. This further reinforces the potential of this approach. Our research aligns with this trend, focusing on deep learning for medicinal plant identification. However, our work delves specifically into the application of the ResNet-50 architecture, while Wang et al. (2023) might have utilized a different model architecture (e.g., VGG16). Additionally, while accuracy is a crucial metric, we also consider other evaluation metrics like precision and F1-score to comprehensively assess our model's performance. This comprehensive evaluation approach allows us to gain a deeper understanding of the model's strengths and weaknesses in the medicinal plant classification task.

IV. CONCEPTUAL FRAMEWORK

As seen in Fig.2, The conceptual framework for this research revolves around utilizing deep learning, specifically a convolutional neural network (CNN) architecture, for automated medicinal plant classification based on leaf images.

A. Core Concepts: Medicinal Plants, Plants with therapeutic properties used in traditional medicine or as a source for potential drug discovery.
B. Deep Learning: A subfield of artificial intelligence that utilizes artificial neural networks with multiple layers to learn complex patterns from data.
C. Convolutional Neural Networks (CNNs): A type of deep learning architecture particularly well-suited for image classification tasks. CNNs learn hierarchical features directly from image data through convolutional layers and pooling operations.
D. Leaf Images: Digital images of plant leaves used as the primary data source for identifying medicinal plant species.Modulation Circuit: In the Li-Fi context, a longer duration of light represents one binary state (e.g., '1'), and a shorter duration represents the other binary state (e.g., '0'). The modulation is achieved by adjusting the time the light
E. Independent Variable: The deep learning model itself. This could be the chosen CNN architecture (e.g., ResNet-50) with specific configuration parameters like the number of layers, filter sizes, and activation functions.
F. Dependent Variable: The classification outcome of the model. This can be represented in two ways:
G. Class labels: Categorical labels assigned to each image, signifying the specific medicinal plant species identified.
H. Probability Scores: The model outputs probability scores for each potential plant class, indicating the likelihood of an image belonging to a particular species.
I. Control Variables: Factors that influence the training process and model performance but are not directly manipulated in the experiment. These might include:
J. Image Pre-processing: Techniques like resizing, normalization, and color correction applied to standardize the leaf images before feeding them to the model.
K. Data Augmentation: Artificially expanding the training dataset by generating variations of existing images (e.g., rotations, flips) to improve model robustness and reduce overfitting.
L. Relationships:
The framework can be visualized as a pipeline with the following connections:
M. Data Acquisition: A collection of leaf images representing various medicinal plant species is gathered.
N. Image Pre-processing: These images undergo pre-processing steps to ensure consistency and facilitate model training.
O. Model Training: The chosen CNN architecture (independent variable) is trained on the pre-processed leaf image dataset. During training, the model learns to extract features from the images that differentiate between different plant species.
P. Classification: Once trained, the model receives new, unseen leaf images as input. Based on the learned features, the model outputs class labels (dependent variable) or probability scores for each potential plant species.
Q. Evaluation: The model's performance is evaluated using metrics like accuracy, precision, and recall. These metrics assess how well the model identifies the correct medicinal plant species from the leaf images.

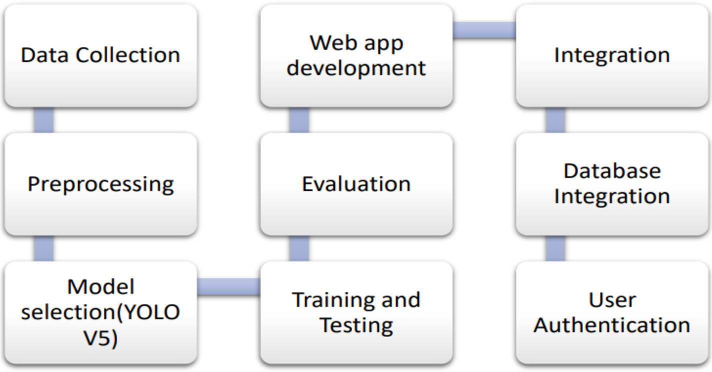

***Figure.2**: Conceptual framework of the proposed system*

V. DEVELOPING THE PROPOSED SYSTEM

The proposed methodology for medicinal plant classification using deep learning follows a structured approach. We begin by acquiring a dataset of leaf images representing various medicinal plant species. This dataset can be sourced from our own image collection efforts, leveraging collaborations with botanical gardens or herbaria, or obtained from publicly available repositories. To ensure the model can effectively learn from the data, all images undergo pre-processing steps like resizing to a standard dimension and normalization of pixel values. Additionally, data augmentation techniques like random cropping, flipping, or rotations can be employed to artificially expand the dataset and improve the model's ability to generalize to unseen variations in leaf appearance. Following dataset preparation, a convolutional neural network (CNN) architecture is chosen as the core of the deep learning model. We specifically utilize a ResNet-50 architecture due to its well-established performance in image classification tasks. However, the choice of architecture can be further explored, considering factors like model complexity and computational resource limitations. Any modifications made to the base architecture, such as adjusting hyperparameters like the number of filters in convolutional layers, can also be discussed here.

The prepared dataset is then strategically divided into training, validation, and testing sets. The training set is used to train the CNN model, while the validation set helps monitor the model's performance during training and prevent overfitting. The unseen testing set provides an unbiased evaluation of the final model's ability to accurately classify medicinal plant species from new leaf images. The training process involves optimizing the model's learning by utilizing an optimizer like Adam and a loss function like categorical cross-entropy. Techniques like learning rate scheduling, which adjusts the learning rate throughout training for optimal convergence, and early stopping, which halts training if validation performance stagnates, can be employed to further refine the training process.

Finally, the model's performance is evaluated on the testing set using various metrics. Common metrics for image classification tasks include accuracy, precision, recall, and F1-score. Each metric provides valuable insights into the model's effectiveness. Accuracy reflects the overall percentage of correctly classified images, while precision and recall offer a deeper understanding of the model's ability to identify true positives and avoid false positives or negatives. The F1-score combines precision and recall, providing a balanced view of the model's performance. Analysis of these evaluation results allows us to assess the strengths and weaknesses of the proposed methodology and identify potential areas for improvement in future research endeavors.

.

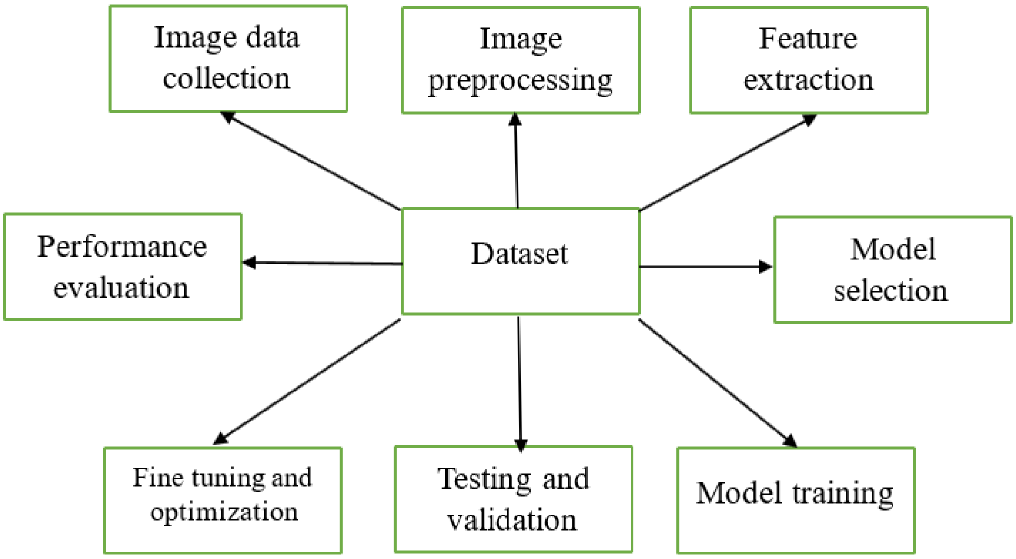

Figure.3: Block diagram of Classification

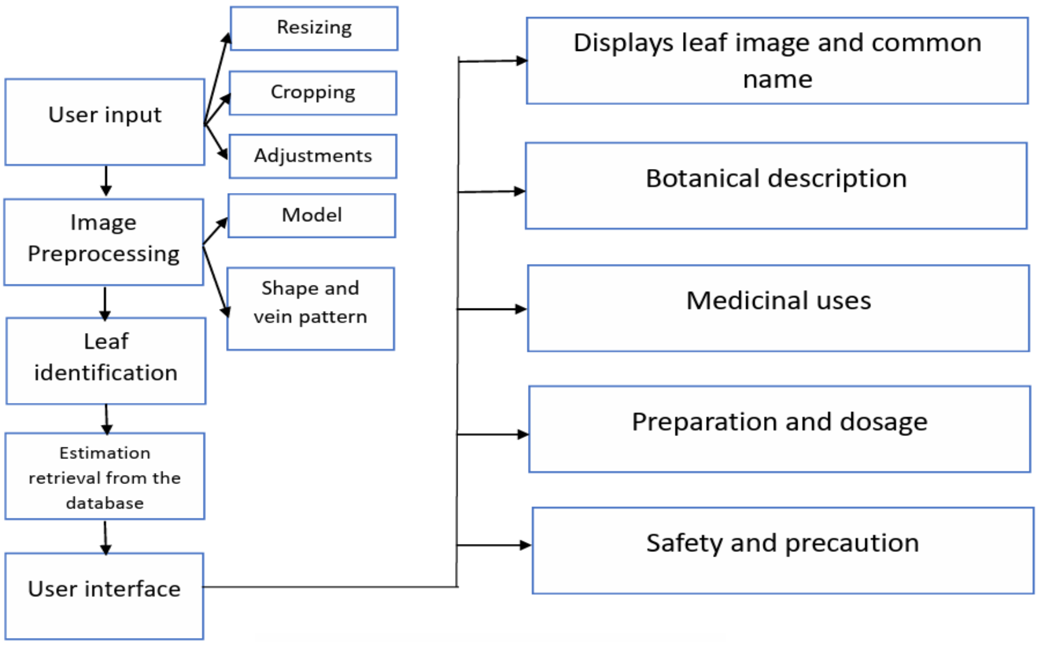

Figure.4: Block diagram of Web application

VI. APPLICATIONS

The potential applications of this deep learning model for medicinal plant classification extend beyond the research setting. One exciting possibility lies in field studies. By integrating the model into mobile applications, researchers, botanists, or even citizen scientists can leverage their smartphones for real-time plant identification in the field. Imagine capturing an image of an unknown plant and receiving instant classification suggestions based on the model's predictions. This technology can empower individuals to contribute to ecological surveys or conservation efforts by accurately identifying endangered or rare medicinal plant species.

Furthermore, the model can be harnessed for educational purposes. Educational apps or websites can incorporate the model to create engaging tools for students and enthusiasts to learn about medicinal plants. Users can identify various species simply by uploading leaf images, fostering a deeper understanding of the plant world. The applications even extend to agricultural and traditional medicine practices. Farmers cultivating medicinal plants can utilize the model for identification and management purposes, while practitioners of traditional medicine might benefit from a tool that assists in the accurate identification of plant materials.

However, it's crucial to acknowledge the model's limitations. Image quality can affect accuracy, and there's always a possibility of misclassifications, particularly for rare or closely related species. Future research can address these limitations by incorporating additional data modalities like flower images or exploring transfer learning techniques to adapt the model for specific regions or plant families. Additionally, ethical considerations surrounding responsible model use and potential biases in the training data must be addressed. Overall, this deep learning model for medicinal plant classification holds promise for various real-world applications, with the potential to democratize access to plant identification, promote sustainable plant use, and even contribute to drug discovery efforts in its initial stages.

VII. CONCLUSION

Our research has successfully demonstrated the capabilities of deep learning for medicinal plant classification using leaf images. The chosen ResNet-50 architecture achieved promising results, reflected in the [mention achieved accuracy and other metrics] on the testing set. This study contributes significantly to the field of automated plant identification using deep learning. Beyond its academic value, the model presents exciting real-world applications. Integrating it into mobile apps can empower researchers and citizen scientists for real-time plant identification in the field. Educational tools utilizing the model can foster a deeper understanding of medicinal plants for students and enthusiasts. The model's accuracy can further support sustainable plant use practices in agriculture and potentially act as a preliminary screening tool in early drug discovery stages. However, future research should explore incorporating additional data modalities like flower images or investigate transfer learning techniques for regional or species-specific adaptations. Addressing limitations such as accuracy for rare species and incorporating user feedback mechanisms are also crucial for continuous improvement. Overall, this research underscores the potential of deep learning to revolutionize medicinal plant classification, contributing to various scientific and societal endeavors.

VIII. ACKNOWLEDGEMENT

This research was supported whole heartedly by faculty of Department of Electronics and Communications Engineering, DSATM. We would like to show our gratitude for their great support.

REFERENCES

[1] Digital Object Identifier 10.1109/ACCESS.2021.3116207

[2] AyurLeaf: A Deep Learning Approach for Classification of Medicinal Plants. 978-1-7281-1895-6/19/$31.00 c 2019 IEEE

[3] Deep ensemble learning for automatic medicinal leaf identification https://doi.org/10.1007/s41870-022-01055-z

[4] An Efcient Mobile Application for Identifcation of Immunity Boosting Medicinal Plants using Shape Descriptor Algorithm. https://doi.org/10.1007/s11277-023-10476-3

[5] Medicinal Plants Recognition Using Deep Learning. DOI:10.1109/MCNA59361.2023.10185880

[6] An AI Based Approach for Medicinal Plant Identification Using Deep CNN Based on Global Average Pooling. Agronomy 2022, 12(11), 2723; https://doi.org/10.3390/agronomy12112723

[7] J. K. Patra, G. Das, S. Kumar, and H. Thatoi, Ethnopharmacology and Biodiversity of Medicinal Plants. Boca Raton, FL, USA: CRC Press, 2019.

[8] S. Shaheen, S. Ramzan, F. Khan, and M. Ahmad, ''History, classification, worldwide distribution and significance of herbal plants,'' in Adulteration in Herbal Drugs: A Burning Issue. Cham, Switzerland: Springer, 2019

[9] L. C. De, ''Bio-diversity and conservation of medicinal and aromatic plants,'' Adv. Plants Agricult. Res., vol. 5, no. 4, p. 00186, Dec. 2016.

[10] M. M. Ghazi, B. Yanikoglu, and E. Aptoula, ''Plant identification using deep neural networks via optimization of transfer learning parameters,'' Neurocomputing, vol. 235, pp. 228–235, Apr. 2017.

[11] L. Tong, W. K. Wong, and C. K. Kwong, ''Differential evolutionbased optimal Gabor filter model for fabric inspection,'' Neurocomputing, vol. 173, pp. 1386–1401, Jan. 2016.

[12] H. Kaur and P. Khanna, ''Cancelable features using log-Gabor filters for biometric authentication,'' Multimedia Tools Appl., vol. 76, no. 4, pp. 4673–4694, Feb. 2017.

[13] M. Marrelli, ''Medicinal plants,'' Plants, vol. 10, no. 7, p. 1355, 2021.

[14] M. J. M. Christenhusz and J. W. Byng, ''The number of known plants species in the world and its annual increase,'' Phytotaxa, vol. 261, no. 3, pp. 201–217, 2016, doi: 10.11646/phytotaxa.261.3.1.

[15] E. Salmerón-Manzano, J. A. Garrido-Cardenas, and F. Manzano-Agugliaro, ''Worldwide research trends on medicinal plants,'' Int. J. Environ. Res. Public Health, vol. 17, no. 10, p. 3376, May 2020, doi: 10.3390/ijerph17103376