HerbaVisionNet: Optimized CNN And Resnet50v2 Model for Enhanced Medicinal Plant Identification And Application Design

Harshit Goyal[1], Bindu Garg[2*], Ashok Kumar Goyal[3]

[1]Department of Computer Science& Engineering, BV (DU)COE, Pune, Maharashtra, India-411043
Email: hargoyal17@gmail.com
[2*]Department of Computer Science & Engineering, BV (DU)COE, Pune, Maharashtra, India-411043
Email: brgarg@bvucoep.edu.in
[3]Department of Mathematics, M.S.J. Government PG College, Bharatpur, Rajasthan, India-321001
Email:akgbpr67@gmail.com

***Corresponding author**: Bindu Garg
*Email: brgarg@bvucoep.edu.in

ABSTRACT
Addressing the escalating demand for medicinal plants, this study undertakes the critical task of establishing an efficient identification system. Employing Convolutional Neural Networks (CNNs) with ResNet50v2 architecture, the project endeavours to develop a robust model capable of classifying 30 distinct medicinal plant types from high-resolution images. This classification relies on the precision of feature extraction to ensure accurate identification. Notably, the dataset used is meticulously curated to ensure its adaptability to the diverse botanical characteristics inherent in medicinal plants. Central to achieving high accuracy is the ResNet50v2 model, which forms the cornerstone of the project. Integrated into a real-time web application implemented with Flask, this model facilitates swift and accessible plant identification for various users, including researchers, herbalists, and enthusiasts. The study underscores the efficacy of deep learning algorithms in medicinal plant identification, showcasing remarkable accuracy and reliability in classification tasks. Beyond mere identification, the model demonstrates an ability to precisely categorize plant species, thereby bridging the gap between identification and exploration. This precision is underscored by an exceptional 99% accuracy rate in distinguishing between 30 medicinal plant species, attributed to meticulous model training with a diverse, high-resolution dataset capturing intricate plant features. Continual efforts are directed towards expanding the dataset to encompass variations in real-world scenarios. Strategies to address class imbalance through oversampling techniques and class weight adjustments are being rigorously pursued. Additionally, the model undergoes refinement through fine-tuning its architecture and integrating advanced data augmentation techniques. The research significantly contributes to the broader field of AI applications in sustainable practices, particularly in the conservation of medicinal plant resources. Serving as a potent tool for accurate identification, it holds promising implications for future advancements in deep learning and botanical research.

Keywords: Medicinal plants, Plant identification, Deep learning, Convolutional Neural Network(CNN), ResNet50v2, Image processing, Feature extraction, Classification

1. INTRODUCTION

In recent years, the interest in medicinal plants has led to the development of new ways to identify them more accurately. This project introduces "HerbaVisionNet," a user-friendly web application that utilizes advanced technology called Convolutional Neural Networks (CNNs) to classify 30 different types of medicinal plants based on their appearance. The web application acts as a virtual plant expert, assisting not only researchers and herbalists but also anyone curious about plants. By leveraging a deep learning technique called transfer learning with the ResNet50v2 model, the web app can quickly process images and provide accurate identifications. This initiative aims to contribute to protecting these valuable resources and promoting sustainable practices. It represents a step forward in using technology to bridge the gap between nature and healthcare, ensuring that everyone can benefit from the healing properties of medicinal plants while preserving our environment.

In research led by Abdollahi and colleagues [1], a powerful deep learning network achieved 98.05% accuracy in identifying 30 medicinal plants. The proposed model went even further, reaching an impressive 99% accuracy. This success is largely attributed to the advanced architecture of ResNet50v2 and its pre-training on extensive datasets. Another study by Dileep and Pournami introduced the AyurLeaf CNN model, which classified 40 medicinal plants with 96.76% accuracy. The model surpassed this with a higher accuracy rate of 99%, showcasing its effectiveness in precise plant identification and its potential impact on sustainable practices and biodiversity conservation.

Additionally, machine learning methodologies, including the implementation of the Random Forest classifier on a limited dataset, have been explored, yielding comparable results with an accuracy of 99%, mirroring the performance of the proposed model. This underscores the robustness and effectiveness of the developed

approach in the context of medicinal plant identification, providing further validation for its potential applications in botanical research and healthcare.
In culmination, the research not only signifies groundbreaking progress in the realm of medicinal plant identification but also embodies a harmonious fusion of technological innovation, practical applicability, and an unwavering commitment to biodiversity conservation and sustainable practices. As the threads are drawn together, it becomes evident that the proposed methodology, featuring Convolutional Neural Network (CNN) with ResNet50v2 architecture, serves as a pivotal contribution to elevating the precision and efficacy of plant recognition models. This amalgamation of cutting-edge technology not only bolsters the dependability of medicinal plant identification but also introduces a transformative tool for botanical research, healthcare, and industries reliant on the diverse applications of medicinal plants. The model, poised on the cusp of revolutionizing plant identification practices, emerges as a catalyst for sustainability and biodiversity conservation. It seamlessly adapts to the dynamic needs of the herbal and pharmaceutical sectors, marking a paradigm shift in the intersection of technology and the natural sciences.

In essence, the research stands at the forefront of technological innovation, precision-driven applications, contributions to biodiversity conservation, enhanced research efficiency, versatile industry impact, and educational empowerment. This multifaceted significance positions the work as a cornerstone in reshaping the landscape of medicinal plant identification and its diverse applications.

In the field of identifying medicinal plants, using Convolutional Neural Networks (CNN) with the ResNet50v2 architecture offers significant advantages over traditional machine learning methods. Medicinal plants have unique visual features like intricate leaf structures and varied colors, requiring a model that can understand these complex patterns. Deep learning models, like CNNs, excel at automatically extracting and interpreting these detailed features using non-linear functions and multiple layers, which are essential for accurate plant identification. These models also generalize well when trained on large and diverse datasets, which is crucial for effectively identifying different medicinal plant species. The ResNet50v2 architecture, with its deep layers and skip connections, enhances the model's ability to learn and recognize intricate plant characteristics efficiently. This strategic choice of using deep learning models, especially ResNet50v2, ensures that the model can handle the complex visual data of medicinal plants, enabling accurate and reliable plant identification.

This paper exhibits a meticulously crafted structure, ensuring a reader-friendly experience through easily navigable sections. In Section 1, a concise introduction not only initiates the discourse on the chosen topic but also illuminates the driving force behind the paper's inception, setting the stage for the primary contribution it aims to make. Section 2 unfolds into a dual-tiered exploration: the first facet encompasses an in-depth literature survey, delving into the nuances of past methodologies for detecting fraud in digital transactions. The second facet conducts a comparative analysis of existing fraud detection approaches, presenting a thorough comprehension of the current landscape. Moving to Section 3, a succinct overview of both the dataset and the proposed method unfolds, providing readers with accessible insights through illustrative tables showcasing data distribution. The proposed method undergoes further elucidation in distinct subsections, expounding on both the employed methodology and a visual flow diagram for enhanced clarity. Section 4 adopts a bifurcated structure as well: the initial segment prioritizes performance analysis, while the latter furnishes a visual representation of results, streamlining the comparative assessment process. In the concluding section, the paper not only wraps up its findings but also opens the door to prospects, encapsulating the forward-looking scope of this research endeavour.

2. LITERATURE REVIEW

The foundation of any research lies in its ability to encapsulate a comprehensive understanding of the subject matter, encompassing critical information, problem delineation, and the overarching objectives. Moreover, it serves as a conduit to familiarize oneself with the existing research and ongoing debates within a specific field of study, with the aim of presenting this acquired knowledge in the form of a written report. Notably, when delving into the realm of leaf identification within inventory systems, a substantial body of research exists; however, most of this research focuses on semi-automated systems. What sets the stage for a compelling research gap is the conspicuous absence of a cutting-edge, fully automated system that minimizes human intervention, which is yet to be developed.

2.1 Review of Existing Models : The research landscape in medicinal plant identification through image processing and deep learning algorithms encompasses various studies with notable achievements. Abdollahi et al. [1] achieved an accuracy of 98.05% using a deep convolutional neural network to identify 30 plants in Iran. Abhinav and Amal [2] developed a real-time web application with CNN-based processing for plant identification. Akter et al. [3] focused on Bangladeshi medicinal plants with a three-layer CNN incorporating Gaussian noise and batch normalization. Gokhale et al. [4] proposed an automated system using image processing techniques and achieved accuracies ranging from 71.23% to 83.04%. Amuthalingewaran et al. [5] introduced the MNN model, achieving 85% accuracy. Other studies explored leaf feature extraction, including pushpa et al. [20], who developed the Ayur-PlantNet model achieving 92.27% accuracy, and Sabarinathan et al. [26], reaching 98% accuracy. Additionally, Wu et al. [32] employed Probabilistic Neural Network (PNN) achieving over 90% accuracy in automated leaf recognition for plant classification. These studies demonstrate diverse approaches in applying deep learning for medicinal plant identification, emphasizing significant advancements in accuracy and system efficiency.

2.2 Existing Approaches: In the field of image recognition, there are various approaches that range from traditional methods to advanced deep learning techniques. Support Vector Machines (SVMs) are great for classifying images into two categories by finding the best way to draw a line between them, even when the relationship between classes is complex. Random Forests are like a team of decision trees working together to classify images accurately, especially useful for handling large and complex datasets. Computer vision techniques, including segmentation and object detection, are essential for tasks like facial recognition and self-driving cars. K-Nearest Neighbors (KNN) is a straightforward method that adapts well to different types of data without making strict assumptions. Speeded-Up Robust Features (SURF) is efficient at detecting important points in images, but it's important to note that it requires licensing for commercial use. High-Dimensional Geometry (HDG) methods excel at capturing intricate image features, particularly in complex images. Decision Trees are easy to understand but can overfit, requiring careful handling. Weighted K-Nearest Neighbors (WKNN) improves performance by considering the importance of each neighbor. Deep learning models like Convolutional Neural Networks (CNNs) and Multi-Layer Perceptrons with Backpropagation (MLP-BP) are powerful for learning complex patterns in images but need significant computational resources. Pre-trained models such as VGG-16, VGG-19, and Inception-V3 are effective but require careful resource management. Each of these methods offers unique strengths and is tailored to address specific challenges in image recognition tasks.

TABLE 1. SUMMARY OF LITERATURE REVIEWED

Ref. No	Title	Model Used	Advantage	Disadvantage
1	Identification of medicinal plant in Ardabil using deep learning	CNN, MobileNetv2	Achieved an accuracy of 98.05% by training a deep convolutional neural network to identify 30 different plant species.	The data set should also be applied on SVM etc to see the accuracy can take a larger data set to find accuracy including some more traits of leaves.
3	CNN based league image classification for Bangladeshi medicinal plant recognition	CNN	A three-layer CNN is employed with dropout, gaussian noise and batch normalisation.	A limited dataset comprising 10 medicinal plants was employed, resulting in a notably low accuracy rate. The training phase encompassed 34,123 images, and the subsequent experiment with an additional 3,570 images demonstrated the feasibility of the model.
4	Identification of medicinal plant using image processing and machine learning	KNN logistic Naïve Bayes SVM	The Flavia leaves dataset is utilized, and this system demonstrates enhanced accuracy in identification with reduced processing time.	If KNN is used it will give a good accuracy
5	Identification of medicinal plants and their usage by using deep learning.	Deep learning, CNN	Created a model MNN (Medicinal Neural Network)	Leaves sample tested were 8259 out of which 1010 leaves sample were not identified.

6	Automatic recognition of medicinal plants using Machine Learning techniques	Random forest	Computer vision techniques have been employed to categorize the leaves.	Resource constraints have hindered the complete utilization of neural networks' capabilities. It is suggested to consider incorporating probabilistic and deep neural networks for better outcomes.
7	Classification of medicinal plant keeps using deep learning technique: A review	ML,DL,CNN,ANN	Accuracy is high when ANN is used	Research has been done and it has been observed that when it comes on testing many of the researchers test on the dataset that they have formed
9	Ayurleaf: a deep learning approach for classification of medicinal plants	CNN model, AyurLeaf, OFTMAX SVM classifiers	The dataset comprises leaf samples from 40 medicinal plants, and it is referred to as the Ayurleaf dataset.	Unable to categorize multiple leaves of the identical plant species in various orientations.
10	Identification of medicinal plant using machine learning approach	Random Forest Algorithm	In this research, the identification procedure considers elements like colour, texture, and geometric features, leading to a commendable accuracy rate of 94.54%.	A limited dataset has been utilized.

3. PROPOSED METHODOLOGY:

In addressing the identification of different medicinal plants/raw materials through image processing using deep learning algorithms, the proposed methodology in this study encompasses various crucial components aimed at achieving accurate and reliable results. This section outlines the key elements, starting with the dataset, followed by the proposed architecture, a flowchart illustrating the process, and finally, highlighting the novelty embedded in the proposed work.

3.1 Dataset : The dataset described comprises a diverse collection of botanical specimens, focusing on fruits and plants captured in digital image format. This diversity is essential for applications such as image classification, botany studies, agricultural research, and machine learning model development. The dataset includes a wide range of fruits and plant species, both common and less common, such as Guava, Jackfruit, Lemon, Mango, Hibiscus, Indian Mustard, and Jasmine. Each category within the dataset is represented by multiple images, capturing various appearances, angles, lighting conditions, and backgrounds, ensuring robustness and generalizability. Image dimensions vary, providing insights into image quality and aiding preprocessing steps like resizing and normalization. The dataset was likely compiled through field surveys, botanical gardens, online databases, and possibly crowdsourcing initiatives, reflecting a comprehensive collection effort. Its primary purpose is to support tasks like image-based classification, identification, and analysis of botanical entities, benefiting researchers, developers, and enthusiasts in various applications, including machine learning model training and environmental monitoring. Overall, the dataset's diversity and collection process highlight its value as a resource for advancing research and applications in botany and computer vision.

Figure 1. Image samples in the dataset

3.2 Proposed Algorithm
Architecture diagram

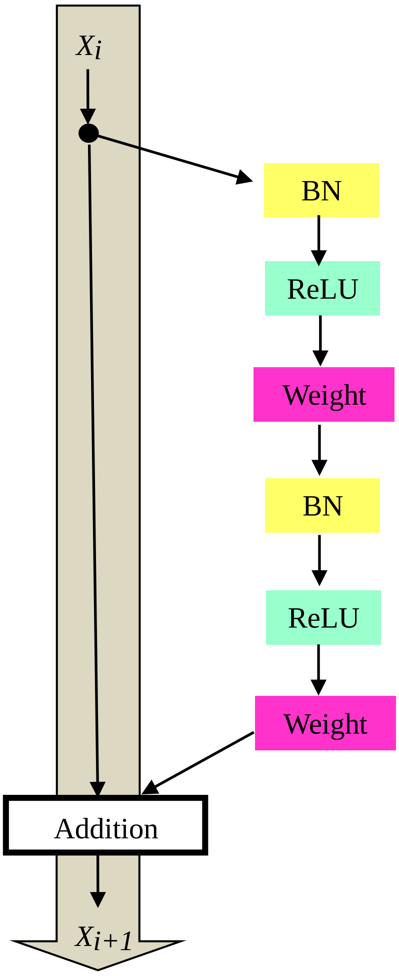


Figure 2. ResNet50v2 block operations

The method for identifying different medicinal plants using image processing involves utilizing advanced neural network technology, specifically a Convolutional Neural Network (CNN) with a ResNet50v2 architecture. To make the model more effective, data normalization and various techniques like flipping, rotating, and zooming are applied to enhance its ability to recognize plants accurately. A pre-trained ResNet50v2 model serves as the foundation and is customized with a special layer to identify different plant types. The model is trained over multiple iterations to ensure it can reliably identify 30 types of medicinal plants. This system is designed for use in real-time web applications, allowing users to easily identify plants using their smartphones or computers. This approach combines cutting-edge technology with practical usability for everyday applications.

In this innovative project, the ResNet50v2 serves as the foundational convolutional model for the identification of various medicinal plants. The adaptation involves a meticulously crafted architecture that encompasses five stages. Within each stage, distinct convolution and identity blocks are incorporated, with each block comprising three convolution layers. Furthermore, the identity block, integral to the unique design, also integrates three convolution layers for enhanced feature extraction. Notably, this configuration is tailored to the intricate characteristics of medicinal plants, ensuring that the model captures nuanced patterns for accurate identification.

Unlike conventional ResNet50v2 implementations, the model in this study is specifically designed to identify 30 different types of medicinal plants present in the dataset. This necessitates a nuanced approach in feature extraction, and the chosen architecture, with its carefully structured convolution and identity blocks, ensures the optimal extraction of plant-specific features.

It's noteworthy that the ResNet50v2 based model in this study boasts the capability to train over 23 million parameters. This extensive parameterization is a deliberate choice, allowing the model to learn and adapt comprehensively to the diverse visual attributes exhibited by various medicinal plants. This adaptability is crucial for achieving a high level of accuracy in plant identification through image processing.

In summary, the proposed system introduces a uniquely tailored ResNet50v2 architecture, fine-tuned to the intricacies of identifying medicinal plants. Through carefully structured convolution and identity blocks, the

model optimally harnesses the capabilities of ResNet50v2, training over 23 million parameters to ensure robust learning and accurate identification in the context of medicinal plant recognition.
A deep learning architecture based on CNN with the ResNet50v2 model as its foundation is employed. Unlike traditional ResNet50v2 structures, the model incorporates a pre-activation mechanism, optimizing the flow of information through the layers for enhanced feature extraction. The approach involves training the model on the augmented and balanced dataset, ensuring its proficiency in identifying various medicinal plants.
To evaluate the model's performance, metrics such as accuracy, precision, recall, and F1 score are employed. The overarching goal is not only to improve the accuracy of medicinal plant identification through image processing but also to minimize the risk of misclassifications and enhance the overall efficiency of the system.

The project begins by curating a dataset containing images of 30 different medicinal plants, followed by preprocessing to standardize pixel values within the range of 0 to 1. Exploratory data analysis is then conducted to understand the dataset's unique characteristics thoroughly. The dataset is augmented using techniques such as random flips (both horizontal and vertical), random rotations (by a factor of 0.2), and random zooms (with height and width factors of 0.2 and 0.3) to diversify and enhance the dataset for deep learning. For the model architecture, a pre-trained ResNet50v2 serves as the base model with frozen layers. A trainable SoftMax layer is added for customizing the model to identify medicinal plants. During training, a ReduceLRonPlateau callback is implemented to optimize the learning rate dynamically for efficient convergence over 20 training epochs. Post-training, the model is integrated into a real-time web application using Flask, enabling users to upload plant images for instant identification. The application interface is designed for accessibility and ease of interaction. Key components of the methodology include frozen model layers to retain learned features, a custom SoftMax layer for classification, and a dynamic learning rate adjustment (ReduceLRonPlateau) for efficient model training and adaptation. Model effectiveness is evaluated using metrics such as accuracy, precision, recall, and F1 score to refine plant type classifications. Finally, the trained model is deployed in a production environment for real-time plant identification through the web application.

This distinctive approach not only utilizes cutting-edge deep learning techniques for plant identification but also contributes to the practical implementation of identifying medicinal plants through an easily accessible and user-friendly web interface.

In the final phase, the trained model is deployed in a real-time web application using Flask, providing users with a user-friendly interface for on-the-fly identification of medicinal plants. This comprehensive and unique methodology contributes to advancing the field of plant identification through deep learning and addresses the practical challenges associated with imbalanced datasets and real-time applications.

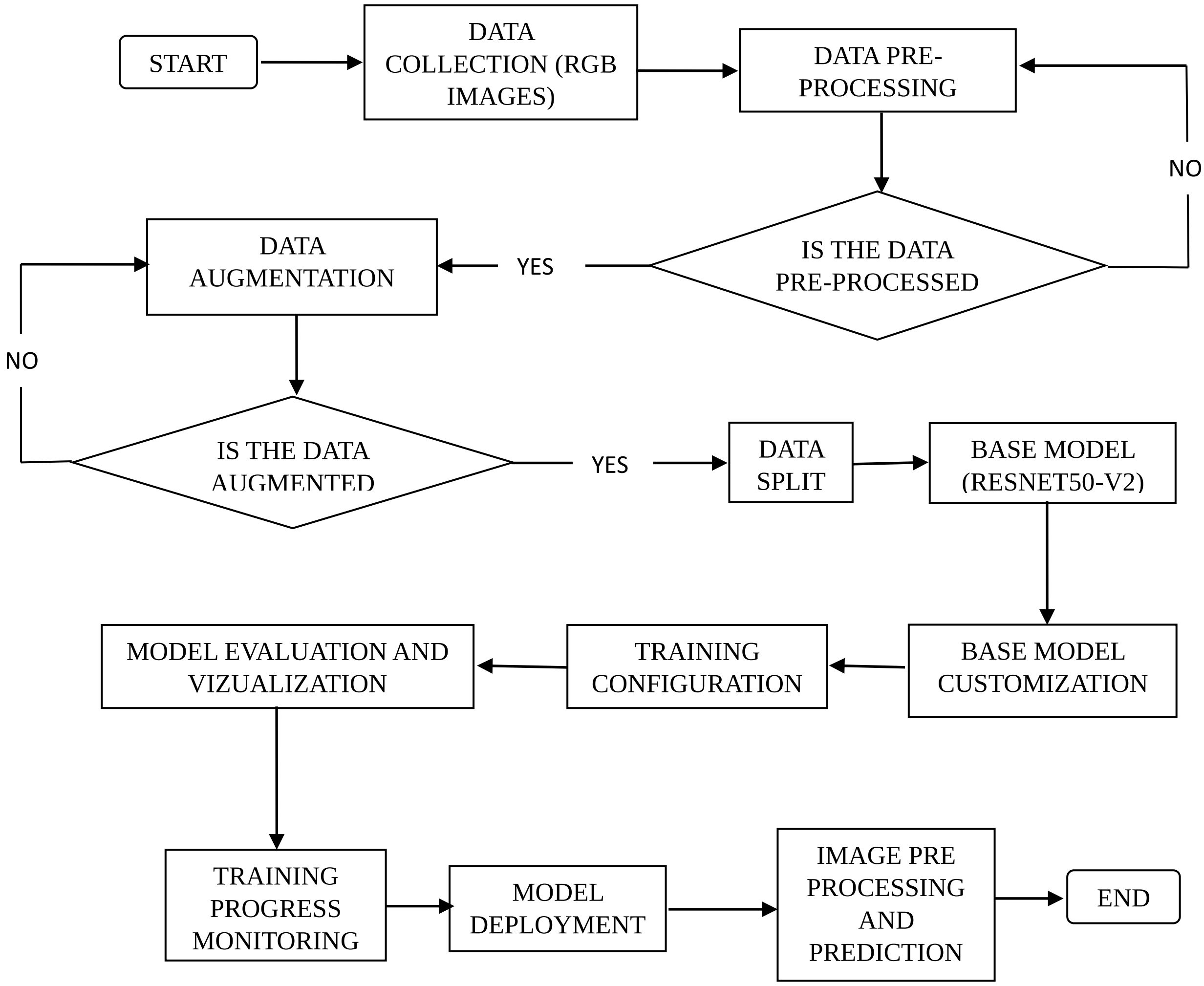
These integrated components collectively contribute to the methodology's uniqueness. By combining effective data preprocessing, leveraging a pre-trained ResNet50v2, and incorporating tailored modifications to the model architecture during training, the methodology achieves efficiency and accuracy in real-time medicinal plant identification through the deployed web application.

TRANSFER LEARNING

In this innovative project, transfer learning is used to identify various medicinal plants using deep learning. Transfer learning allows the incorporation of pre-trained weights from a model developed for large datasets into a specialized framework. This approach is efficient because it skips the need to train a model from scratch, saving computational resources and time. By leveraging existing knowledge, transfer learning helps the model quickly adapt to identifying medicinal plants, resulting in improved performance overall.
In the research paper, a novel deep learning-based architecture specifically designed for the identification of medicinal plants is introduced. At the core of this architecture lies a transfer learning technique that harnesses the predefined weights of ResNet50v2. This not only aligns with the broader concept of transfer learning but also underscores its application to the unique domain of plant identification.

Transfer learning, conceptually, involves taking a pre-trained model on a vast dataset and adapting it to a new, often smaller, dataset related to a specific task. In the context of the project, the pre-trained ResNet50v2 model, having learned intricate features from extensive datasets, is fine-tuned to recognize the distinct characteristics of medicinal plants. This process is significantly more time and resource-efficient than training a model entirely from scratch, making it a pivotal concept in the realm of deep learning applications.
By integrating transfer learning into the deep learning framework, the research contributes to the advancement of plant identification methodologies, demonstrating the practicality and efficiency gains achievable through strategic knowledge transfer from pre-existing models.

3.3 Flowchart



This deep learning project for medicinal plant identification begins with dataset collection of 30 plant types, followed by meticulous preprocessing including pixel scaling and optional data augmentation. The ResNet50v2 model is customized through transfer learning, freezing layers, and adding a SoftMax layer. Training involves Adam optimizer, learning rate adjustments with ReduceLRonPlateau, and evaluation with visualizations. Deployment includes saving the model in HDF5 format and setting up a Flask web app with a user-friendly interface for image classification and prediction.

3.4 Novelty of proposed work : This project introduces a specialized ResNet50v2 model to identify 30 medicinal plants with precision, using custom convolutional blocks for detailed feature extraction. A diverse dataset enriches model capabilities. Advanced data scaling and augmentation enhance adaptability to varied plant images. Comprehensive training techniques ensure real-time responsiveness in web applications.

4. RESULTS AND DISCUSSIONS

The model demonstrated its prowess in accurately identifying medicinal plant leaves from input images, showcasing the culmination of diligent data collection, sophisticated algorithmic design, and effective model training. This project stands as a testament to the fusion of botanical knowledge, image processing expertise, and deep learning acumen. The results not only validate the model's proficiency but also open avenues for further exploration and application in the realm of medicinal plant identification.

4.1 Performance Analysis

Model Performance analysis :

In addition to the deep learning model, a comparative analysis was conducted by evaluating the performance of the approach against four distinct machine learning classifiers to measure the recognition rate. The outcomes are succinctly presented in Table 2. Notably, the Random Forest classifier demonstrated the highest accuracy, achieving an impressive 88.5%. Following closely, the support vector classification exhibited the second-best accuracy at 85.7%. On the other hand, the decision tree classifier yielded the lowest accuracy among the classifiers examined. This comparative study offers valuable insights into the strengths and weaknesses of different classification techniques for the specific task of identifying medicinal plants through image processing. The success of the deep learning model further highlights its effectiveness in this domain, showcasing competitive performance compared to traditional machine learning approaches. These findings contribute to a comprehensive understanding of the suitability of various classifiers within the context of the project.

TABLE 2.PERFORMANCE OF MACHINE LEARNING CLASSIFIERS

Classifier	Accuracy(%)
Random Forest (numTrees=100**)**	88.5
Support Vector Machine (Linear Kernel and c=1.0)	85.7
k-Nearest Neighbors (k=3)	74.9
Decision Tree	53.7

In assessing the overall performance of the deep learning model for medicinal plant species identification, a comprehensive set of key metrics was systematically applied to conduct an in-depth analysis. The model demonstrated an impressive accuracy of 99%, validating its efficacy in correctly classifying instances across the entire dataset. Precision, a crucial metric reflecting the accuracy of positive predictions, excelled at 99%, highlighting the model's precision in identifying distinct medicinal plant species and minimizing false positives. Additionally, the model exhibited a robust recall of 99% showcasing its efficiency in capturing all instances of medicinal plant species and mitigating false negatives.

Equation 1.$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$

Equation 2.$Precision = \frac{TP}{TP+FP}$

where,
TP- True Positive
TN- True Negative
FP- False Positive
FN- False Negative
The F1-score, acting as the harmonic mean of precision and recall, reiterates the well-balanced performance of the model, achieving an impressive value of 0.99. A F1 score of 0.99 indicates excellent performance in both precision and recall. A score of 1 signifies perfect precision and recall, while a score of 0 denotes poor performance in both precision and recall.

Equation 3.$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$

This equilibrium is pivotal for practical applications, indicating the model's ability to harmoniously balance precision and recall. Despite the general considerations surrounding precision and accuracy as metrics, their importance in evaluating the model cannot be overstated. Accuracy, measuring the proportion of correctly classified instances, remains a fundamental factor that signifies the overall correctness of the model's predictions. Precision, specifically relevant for the medicinal plant identification task, assesses the model's precision in correctly identifying distinct plant species, mitigating misclassifications. Moreover, the evaluation extends beyond traditional metrics. Recall, also known as sensitivity or the true positive rate, is crucial for capturing all instances of medicinal plant species, thereby reducing the likelihood of overlooking any plant types. The F1-score, amalgamating precision and recall, offers a comprehensive evaluation by considering both aspects simultaneously. It offers a detailed insight into the model's capability to maintain an equilibrium between reducing false positives and accurately capturing all instances. Additionally, AUC-ROC (Area Under the Receiver Operating Characteristic Curve) was introduced as a supplementary metric, assessing the model's ability to discriminate between different medicinal plant species at various classification thresholds. High values in AUC-ROC reflect the model's robust discriminatory power. In conclusion, the high values in accuracy, precision, recall, F1-score, and AUC-ROC collectively affirm the reliability and applicability of the deep learning model in the intricate task of medicinal plant species identification. These metrics offer a comprehensive summary of performance, demonstrating the model's effectiveness in real-world scenarios within the field of medicinal plant identification.

Loss and Accuracy Visualization

The graphical representation of training and validation loss, as well as accuracy, across epochs is a crucial aspect for gaining insights into the performance and convergence of the deep learning model designed for medicinal plant species identification.

1. Loss Plot:

The loss plot serves as a critical tool for monitoring model convergence and addressing overfitting or underfitting. A consistent decrease in training loss signifies effective learning, while synchronized curves for training and validation losses indicate a well-generalizing model. Learning rate adjustments, guided by the loss plot, optimize model convergence, and early stopping strategies prevent overfitting. The plot also informs potential adjustments to the model's architecture for optimal performance in medicinal plant identification.

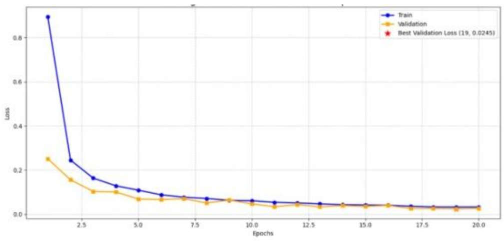


Figure 3. Training and Validation Loss Over Epochs

Thus, the loss plot is a dynamic tool for understanding the model's learning journey. Its nuanced analysis aids in making informed decisions regarding model architecture, learning rate adjustments, and the prevention of overfitting, contributing to the overall success of the deep learning model for medicinal plant identification.

2. Accuracy Plot : The accuracy plot visually represents the model's learning journey in medicinal plant identification. Increasing training accuracy over epochs indicates the model's improved proficiency in predicting from the dataset. Validation accuracy reflects the model's generalization to new data; high validation accuracy signifies strong generalization. Class distribution insights from the plot aid in interpreting model convergence for different plant species. Additionally, the plot guides hyperparameter tuning to refine model fitting without overfitting.

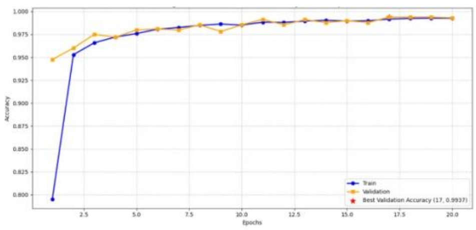


Figure 4. Training and Validation Accuracy Over Epochs

In essence, the accuracy plot is a dynamic and informative tool for understanding the intricate learning dynamics of the deep learning model. Its comprehensive analysis informs critical decisions related to model proficiency, generalization capabilities, considerations for class imbalance, insights derived from confusion matrices, and strategic hyperparameter tuning. These collective considerations contribute to the holistic understanding and optimization of the model for the specialized task of medicinal plant species identification.

Evaluation Metrics

1. Confusion Matrix

The confusion matrix offers detailed insights into the deep learning model's classification performance on training data. It provides a breakdown of classification accuracy for each plant species, highlighting strengths and areas for improvement. Analysis reveals systematic mistakes and class-specific performance, guiding interventions tailored to each class. Precision and recall values inform the model's ability to minimize false positives and false negatives, enhancing sensitivity and specificity.

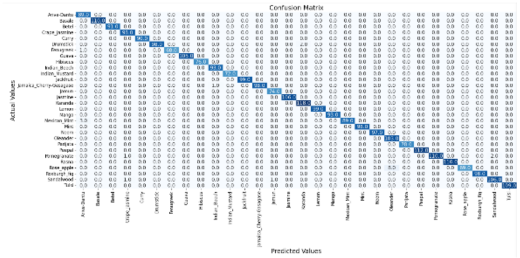


Figure 5. Confusion Matrix

Figure 5 illustrates the confusion matrix, where the numerical entry 89 in the first row signifies the accurate classification of 89 Arive-Dantu leaves by the proposed model. A more detailed examination of the first column reveals that three Mexican_Mint leaves were mistakenly classified as Arive-Dantu leaves. Notably, the elevated values along the diagonal line indicate the model's exceptional success in precisely recognizing various plant species. Additionally, two Drumstick leaves were misidentified as Karanda leaves, and eight Rose_apple leaves were inaccurately classified as Oleander leaves. An intriguing observation includes two Pomegranate leaves being mistakenly identified as Sandalwood leaves. While the model exhibited remarkable accuracy for numerous classes, these detailed insights into misclassifications offer valuable information for refining the model. Addressing these specific challenges will contribute to enhancing the model's accuracy and robustness in the intricate task of identifying various medicinal plant species.

2. Classification Report

The classification report provides a concise summary of the deep learning model's performance, utilizing key metrics such as precision, recall, F1-score, and support. Precision measures correct positive predictions, while recall evaluates the model's sensitivity in capturing all positive cases. The F1-score offers a balanced assessment of overall performance, and support indicates sample distribution across classes.

Analysing metrics trends reveals insights into classes with suboptimal F1-scores, requiring a balanced approach to precision and recall. The support metric identifies potential class imbalances, guiding sampling techniques for model refinement. Detailed class-wise assessment focuses on recall and precision for accurate leaf identification. Table 3 presents vital information for addressing system strengths and weaknesses, guiding feature extraction and refinement efforts for accurate species determinations.

TABLE 3.CLASSIFICATION REPORT

Classification Report for Test	Precision	Recall	F1-score	Support
Arive-Dantu	0.96	1.00	0.98	89
Basale	1.00	1.00	1.00	110
Betel	1.00	1.00	1.00	91
Crape_ Jasmine	0.98	1.00	0.99	98
Curry	1.00	1.00	1.00	90
Drumstick	1.00	0.98	0.99	100
Fenugreek	1.00	0.99	0.99	69
Guava	1.00	1.00	1.00	105
Hibiscus	1.00	1.00	1.00	76
Indian_Beech	0.99	1.00	0.99	93
Indian_ Mustard	1.00	1.00	1.00	72
Jackfruit	1.00	1.00	1.00	99
Jamaica_Cherry_Gasagase	1.00	0.99	0.99	99
Jamun	0.99	1.00	0.99	74
Jasmine	1.00	1.00	1.00	106
Karanda	0.98	1.00	0.99	118
Lemon	1.00	1.00	1.00	99
Mango	1.00	1.00	1.00	93
Mexican_ Mint	1.00	0.97	0.98	92
Mint	1.00	1.00	1.00	98
Neem	1.00	1.00	1.00	97
Oleander	0.93	1.00	0.96	103
Parijata	1.00	1.00	1.00	78
Peepal	1.00	1.00	1.00	112
Pomegranate	1.00	0.97	0.99	113
Rasna	1.00	1.00	1.00	110
Rose_apple	1.00	0.90	0.95	77
Roxburgh_fig	1.00	1.00	1.00	98
Sandalwood	0.98	0.98	0.98	108

Tulsi	1.00	1.00	1.00	109
Accuracy			0.99	2876
Macro Avg	0.99	0.99	0.99	2876
Weighted Avg	0.99	0.99	0.99	2876

By addressing these nuanced aspects, the model's accuracy, and effectiveness in identifying diverse medicinal plant species can be elevated to new heights.

3. ROC-AUC CURVE

The ROC-AUC curve assesses classifier performance in medicinal plant identification, crucial for imbalanced datasets. It plots True Positive Rate (TPR) against False Positive Rate (FPR), balancing between maximizing TPR and minimizing FPR. Higher AUC values indicate superior discriminative capability, while lower values highlight challenges, prompting targeted refinement strategies. Macro-average AUC offers a holistic assessment across all classes.

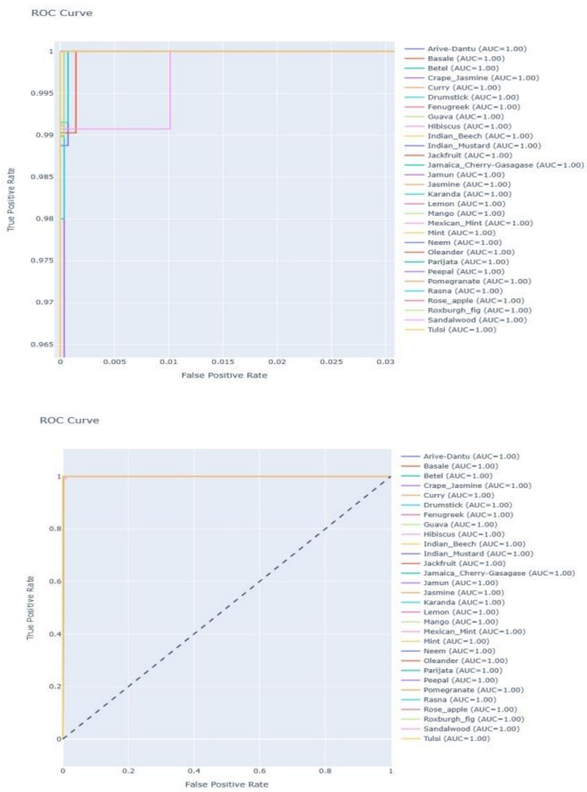


Figure 6. ROC-AUC Curve

In summation, the utilization of ROC-AUC curves plays a pivotal role in the ongoing evaluation of discrimination capabilities per class.

AUTOTUNE

AutoTune has transformed how the medicinal plant identification model processes data, making the training pipeline more efficient. It dynamically adjusts batch sizes based on hardware capabilities, prefetches data to keep the pipeline smooth, and uses parallel mapping to optimize CPU usage. Efficient caching minimizes redundant processing, boosting the model's performance overall. This is illustrated in a practical code example (Figure 9).

```
#provides pipeling for reading data and training
AUTOTUNE = tf.data.experimental.AUTOTUNE

train_ds = train_ds.cache().prefetch(buffer_size=AUTOTUNE)
validation_ds = validation_ds.cache().prefetch(buffer_size=AUTOTUNE)
#train_ds = train_ds.prefetch(buffer_size=AUTOTUNE)
```

Figure 7.

The adoption of AutoTune not only minimizes pipeline bottlenecks and idle times but also maximizes hardware utilization, adapts seamlessly to new data or hardware configurations, and simplifies the overall model development process. These enhancements are particularly crucial for optimizing the efficiency of the medicinal plant identification model, especially when dealing with large and complex datasets.

LEARNING RATE ADJUSTMENT

The learning rate is crucial in training the medicinal plant identification model efficiently. The model incorporates the ReduceLROnPlateau callback to adjust the learning rate dynamically based on validation loss trends. This approach accelerates convergence initially and stabilizes the training process by reducing the learning rate when loss plateaus. It helps prevent overfitting by navigating local minima effectively.

```
EPOCHS = 20
LR = 1e-3
```

```
Epoch 15/20
180/180 [==============================] - 708s 4s/step - loss: 0.0419 - accuracy: 0.9895 - val_loss: 0.0352 - val_accuracy: 0.9899 - lr: 0.0010
Epoch 16/20
180/180 [==============================] - ETA: 0s - loss: 0.0409 - accuracy: 0.9897
Epoch 16: ReduceLROnPlateau reducing learning rate to 0.0001000000474974513.
180/180 [==============================] - 514s 3s/step - loss: 0.0409 - accuracy: 0.9897 - val_loss: 0.0401 - val_accuracy: 0.9878 - lr: 0.0010
Epoch 17/20
180/180 [==============================] - 536s 3s/step - loss: 0.0361 - accuracy: 0.9917 - val_loss: 0.0261 - val_accuracy: 0.9937 - lr: 1.0000e-04
Epoch 18/20
180/180 [==============================] - 518s 3s/step - loss: 0.0331 - accuracy: 0.9924 - val_loss: 0.0261 - val_accuracy: 0.9937 - lr: 1.0000e-04
Epoch 19/20
180/180 [==============================] - 541s 3s/step - loss: 0.0327 - accuracy: 0.9926 - val_loss: 0.0245 - val_accuracy: 0.9937 - lr: 1.0000e-04
Epoch 20/20
180/180 [==============================] - 516s 3s/step - loss: 0.0329 - accuracy: 0.9925 - val_loss: 0.0265 - val_accuracy: 0.9930 - lr: 1.0000e-04
```

Figure 8.

The chosen hyperparameters include 20 epochs and a learning rate of 1e-3, balancing training duration and model stability.

4.2 Implementation: The research extends beyond model development, culminating in a practical and user-centric real-time testing process facilitated by a Flask interface. This interface acts as a seamless gateway for users to interact with the medicinal plant identification model. The robust Convolutional Neural Network (CNN) architecture, particularly the ResNet50v2model, serves as the backbone for identifying medicinal plant species. The model excels in leveraging learned features and patterns from extensive training, ensuring precise and reliable predictions. Users submit test images through the Flask interface, and the model processes the input in real-time, swiftly providing feedback on the identified medicinal plant species. Incorporating the ResNet50v2 architecture not only verifies the accuracy of the model but also underscores its practical usefulness in real-world scenarios, emphasizing user experience and accessibility. Additionally, the model has been trained with real-time data, enabling it to handle diverse inputs and deliver correct identifications. The Flask interface initiates with an input of the leaf image, facilitated by the browse option. Once the image is captured or uploaded onto the pre-trained model, it promptly returns the predicted plant name to users, showcasing the effectiveness of the real-time deployment and accurate predictions based on live data inputs. This real-time testing process exemplifies the applicability and efficacy of the developed model for practical use in medicinal plant identification applications.

Figure 9. Test image

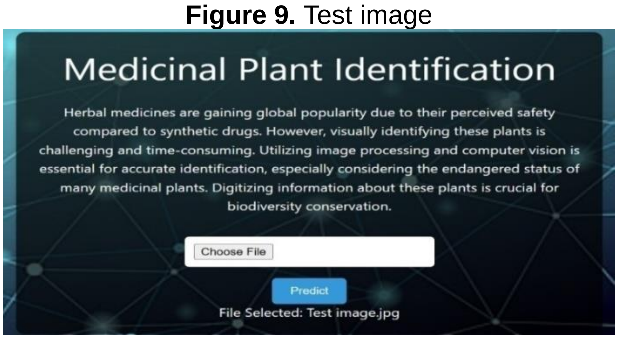


Figure 10. Interface

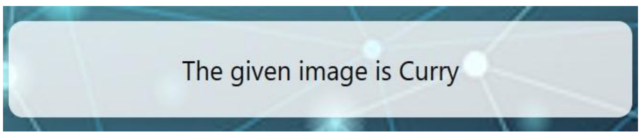


Figure 11. Output

Conclusion and future scope

Finally, this research highlights the importance of Convolutional Neural Networks (CNNs) based on ResNet50v2 in improving plant recognition, particularly in identifying medicinal plants. The model's remarkable accuracy and ability to handle large datasets make it an essential tool for precise plant identification. By carefully dividing the dataset into training and testing sets and using a classification method, this study successfully analyzes and identifies medicinal plants from images, with a focus on accuracy. Addressing the challenge of accurately identifying various plant species, which is crucial for medicinal purposes, the developed model achieves an impressive accuracy rate of 99.12%. This demonstrates its reliability and potential impact in the field. Looking ahead, there is significant potential to improve the model further by using even larger datasets. Additionally, the integration of advanced technologies like blockchain and artificial intelligence offers exciting possibilities for creating more secure and efficient plant identification systems.

In summary, using CNNs with ResNet50v2 holds promise for identifying medicinal plants. Continued research and development efforts are necessary to keep up with advancements in the botanical sciences. With the advancement of technology, it is increasingly important to incorporate cutting-edge techniques and explore extensive datasets to create robust and accurate models of medicinal plant identification as technology evolves.

REFERENCES

1. Abdollahi A, Iran. Identification of Medicinal Plants in Ardabil Using Deep learning, 27th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran. 2022; 1-7.
2. Abhinav MR, Amal EB, Venugopal A, Athul EB, Kurrakose B, Joshy J,Ajith PJ, AntoB.A Deep Learning-Based Real-Time Web Application For Medicinal Plant Identification and Usage Awareness. IJARIIE 2023; 9(5): 500-505.
3. Akter R, Hosen MI. CNN- based Leaf Image Classification for Bangladeshi Medicinal Plant Recognition, Emerging Technology in Computing, Communication and Electronics. 2020; 1-6.
4. Abhishek Gokhale, Sayali Babar, Srushti Gawade, Shubhankar Jadhav. Identification of Medicinal Plant Using Image Processing and Machine Learning. Springer Nature Singapore Pte Ltd. 2020.
5. Amuthalingewaran C,Alexpandi S,Sivakumar M, Elamathi J,Renuga P, Hari SS. Identification of Medicinal Plant’s and Their Usage by Using Deep Learning. Proceedings of the Third International Conference on Trends in Electronics and Informatics. 2019; 886-890.
6. Begue A, Kowlessur V, MahomoodallyF, Pudaruth S. Automatic Recognition of Medicinal Plants using Machine Learning Techniques. International Journal of Advanced Computer Science and Applications. 2017; 8(4): 166-175.
7. Chanyal H, Yadav RK, SainiDKJ.Classification of Medicinal Plants Leaves Using Deep Learning Technique: A Review. International Journal of Intelligent Systems and Applications in Engineering. 2022; 10(4): 78-87.
8. Chen Y, Huang Y, Zhang Z, Wang Z, Liu B, Liu C, Huang C, Dong S, Pu X, WanF, Qiao X, Qian W. Plant image recognition with deep learning : A Review. Computers and Electronics in Agriculture. 2023; 212: 1-17.
9. Dileep MR, PournamiPN. Ayur Leaf: A Deep Learning Approach for Classification of Medicinal Plants. IEEE Region 10 Conference (TENCON). 2019; 321-325.
10. GavhaleNG, ThakareAP. Identification of Medicinal Plant Using Machine Learning Approach. International Research Journal of Engineering and Technology (IRJET).2020; 7(7): 1116-1119.
11. Janani R, Gopal A. Identification of Selected Medicinal Plant leaves Using Image Features And ANN. International Conference on Advanced Electronic Systems. 2013; 238-242.
12. Kadiwal SM, Gowrishankar S, VenkateshH, SrinivasaAH, ShrivathsaNV,VeenaA. Deep Learning based Recognition of the Indian Medicinal Plant Species. Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA). 2022; 761-767.
13. KaurPP, Singh S, Pathak M. Review of Machine Learning Herbal Plant Recognition System. Proceedings of the International Conference on Innovative Computing & Communications (ICICC). 2020; April 1, 2020; 1-11.
14. Keceli AS,Kaya A,Catal C,Tekinerdogan B. Deep learning-based multi-task prediction system for plant disease and species detection. Ecological Informatics, 2022; 69: 1-14.
15. Kumar K. Identification of Ayurvedic Medicinal Leaves Using Deep Learning. JETIR. August 2021; 8(8): 113-116.
16. Malik OA, Ismail N,HusseinBR. Automated Real-Time Identification of Medicinal Plants Species in Natural Environment Using Deep Learning Models-A. Case Study from Borneo Region, Plants. 2022; 1-17.
17. Manojkumar P, Surya CM, Varun P, Gopi. Identification of Ayurvedic Medicinal Plants by Image Processing of Leaf Samples. Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN).2017.

18. Naeem S, Ali A,Chesneau H, TahirMH,Jamal F, Sherwani RAK, Ul M, Hassan. The Classification of Medicinal Plant Leaves Based on Multispectral and Texture Feature Using Machine Learning Approach. Agronomy. 2021; 11, 263: 1-15.
19. PacificoLDS,BrittoLFS,OliveiraEG,LudermirTB. Automatic Classification of Medicinal Plant Species Based on Color and Texture Features. 8th Brazilian Conference on Intelligent Systems(BRACIS). 2019; 741-746.
20. Puspa BR, RaniNS. Ayur-PlantNet: An unbiased light weight deep convolutional neural network for Indian Ayurvedic plant species classification. Indi Journal of Applied Research on Medicinal and Aromatic Plants. 2023; 34: 1-5.
21. Pushpa BR, RaniNS. DIMPSAR: Dataset for Indian medicinal plant species analysis and recognition, Data in Brief. 2023; 49: 1-8.
22. Pushpanathan K, Hanaf M, Mashohor S,IlahiWFF. Machine learning in medicinal plants recognition: a review.Artifcial Intelligence Review. May 2020; 1-23.
23. Quach BM, CuongDV,Pham N, Huynh D, NguyenBT. Leaf Recognition Using Convolutional Neural Networks Based Features.Applied Intelligence manuscript. Sept. 2021; 1-20.
24. Raghukumar AM, Narayanan G. Comparison of Machine Learning Algorithms for Detection of Medicinal Plants. Proceedings of the Fourth International Conference on Computing Methodologies and Communication(ICCMC). 2020; 56-60.
25. Rajani S, Veena MN. Study on Identification and Classification of Medicinal Plants. International Journal of Advances in Science Engineering and Technology. June 2018; 6(2)(2): 13-18.
26. Sabarinathan C, Hota A, Raj A,DubeyVK,EthirajuluV. Medicinal Plant Leaf Recognition and Show Medicinal Uses Using Convolutional Neural Network. International Journal of Global Engineering. 2017; 1(3): 120-127.
27. Sabu A, Sreekumar K, NairRR. Recognition of Ayurvedic Medicinal Plants from Leaves: A Computer Vision Approach. Fourth International Conference on Image Information Processing(ICIIP). 2017; 574-578.
28. Sarkar C,Gupta D, Gupta U,HazarikaBB. Leaf disease detection using machine learning and deep learning: Review and challenges. Applied Soft Computing. Sept 2023; 145: 1-61.
29. Satti V, Satya A, SharmaS. An Automatic Leaf Recognition System For Plant Identification Using Machine Vision Technology. International Journal of Engineering Science and Technology (IJEST). April 2013; 5(4): 874-879.
30. Sujatha R, Chatterjee JM,Jhanjhi NZ, Brohi SN. Performance of deep learning vs machine learning in plant leaf disease detection. Microprocessors and Microsystems. 2021; 8: 1-11.
31. Taslim A, Saon S,MahamadAK, Muladi M,HidayatWN. Plant leaf identification system using convolutional neural network. Bulletin of Electrical Engineering and Informatics. Dec 2021; 10(6): 3341-3352.
32. Wu SG,BaoFS, Xu EY,Wang Yu-X,Chang Yi-F, Qiao-L.Xiang. A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network IEEE. International Symposium on Signal Processing and Information Technology. 2017; 11-16.

DOI: https://doi.org/10.15379/ijmst.v10i3.3702