IDENTIFICATION OF DIFFERENT MEDICINAL PLANTS THROUGH IMAGE PROCESSING USING MACHINE LEARNING ALGORITHMS(MEDISCAN)

Submitted for partial fulfillment of the requirements

for the award of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE ENGINEERING – ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

by

D. Santhi	-	**20BQ1A4214**
Ch. Harsha Vardhan	-	**20BQ1A4210**
G. Kishore	-	**20BQ1A4219**
J. Sai Krishna	-	**20BQ1A4224**

Under the guidance of

N. Nalini Krupa, M.Tech

Assistant Professor

VVIT

VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING – ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY

Permanently Affiliated to JNTU Kakinada, Approved by AICTE

Accredited by NAAC with ‘A’ Grade, ISO 9001:2008 Certified

NAMBUR (V), PEDAKAKANI (M), GUNTUR – 522 508

Tel no: 0863-2118036, url: www.vvitguntur.com

April 2024

VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY
Permanently Affiliated to JNTUK, Kakinada, Approved by AICTE
Accredited by NAAC with ‘A’ Grade, ISO 9001:20008 Certified
Nambur, Pedakakani (M), Guntur (Gt) -522508

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING - ARTIFICIAL INTELLIGENCE & MACHINE LEARNING

CERTIFICATE

This is to certify that this **Project Report** is the bonafide work of **Ms. D. Santhi, Mr. Ch. Harsha Vardhan, Mr. G. Kishore, Mr. J. Sai Krishna**, bearing Reg. No. **20BQ1A4214, 20BQ1A4210, 20BQ1A4219, 20BQ1A4224** respectively who had carried out the project entitled **“Identification Of Different Medicinal Plants Through Image Processing Using Machine Learning Algorithms”** under our supervision.

Project Guide	**Head of the Department**
(N. Nalini Krupa, Assistant Professor)	(Dr. K. Suresh Babu, Professor)

Submitted for Viva voce Examination held on _____________________________

Internal Examiner	**External Examiner**

DECLARATION

We, Ms. D. Santhi, Mr. Ch. Harsha Vardhan, Mr. G. Kishore, Mr. J. Sai Krishna hereby declare that the Project Report entitled **“ Identification of different medicinal plants through image processing using machine learning algorithms ”** done by us under the guidance of Ms. N. Nalini Krupa, Assistant Professor, Computer Science Engineering - Artificial Intelligence & Machine Learning at Vasireddy Venkatadri Institute of Technology is submitted for partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science Engineering - Artificial Intelligence & Machine Learning. The results embodied in this report have not been submitted to any other University for the award of any degree.

DATE :

PLACE : Nambur

SIGNATURE OF THE CANDIDATES

D. Santhi

Ch. Harsha Vardhan

G. Kishore

J. Sai Krishna

ACKNOWLEDGEMENT

We take this opportunity to express my deepest gratitude and appreciation to all those people who made this project work easier with words of encouragement, motivation, discipline, and faith by offering different places to look to expand my ideas and helped me towards the successful completion of this project work.

First and foremost, we express my deep gratitude to **Mr. Vasireddy Vidya Sagar**, Chairman, Vasireddy Venkatadri Institute of Technology for providing necessary facilities throughout the B.Tech program.

We express my sincere thanks to **Dr. Y. Mallikarjuna Reddy**, Principal, Vasireddy Venkatadri Institute of Technology for his constant support and cooperation throughout the B.Tech program.

We express my sincere gratitude to **Dr. K. Suresh Babu**, Professor & HOD, Computer Science Engineering - Artificial Intelligence & Machine Learning, Vasireddy Venkatadri Institute of Technology for his constant encouragement, motivation and faith by offering different places to look to expand my ideas.

We would like to express my sincere gratefulness to our Guide Ms. **N. Nalini Krupa**, Assistant Professor, Computer Science Engineering - Artificial Intelligence & Machine Learning for her insightful advice, motivating suggestions, invaluable guidance, help and support in successful completion of this project.

We would like to express our sincere heartfelt thanks to our Project Coordinator **Mr. N. Balayesu**, Assistant Professor, Computer Science Engineering - Artificial Intelligence & Machine Learning for his valuable advices, motivating suggestions, moral support, help and coordination among us in successful completion of this project.

We would like to take this opportunity to express my thanks to the **Teaching and Non-Teaching** Staff in the Department of Computer Science Engineering - Artificial Intelligence & Machine Learning, VVIT for their invaluable help and support.

Names of Students
D. Santhi
Ch. Harsha Vardhan
G. Kishore
J. Sai Krishna

TABLE OF CONTENTS

LIST OF FIGURES

NOMENCLATURE

ML	Machine Learning
DL	Deep Learning
AI	Artificial Intelligence
CNN	Convolution Neural Network
PC	Personal Computer
VGG	Visual Geometry Group
ER	Entity Relationship
DFD	Data Flow Diagram

ABSTRACT

The identification of medicinal plants and raw materials plays a crucial role in various fields, including healthcare, pharmaceuticals, and environmental conservation. However, manual identification methods are often time-consuming and require expert botanical knowledge. In this project, we propose a novel approach to automate the identification process using image processing and machine learning algorithms.

We present a comprehensive study on the identification of different medicinal plants and raw materials through image processing techniques. Our methodology involves the collection of a diverse dataset containing images of various plant species and raw materials. We preprocess the dataset to standardize the images and remove noise, ensuring the robustness of our models.

We employ state-of-the-art machine learning models, including InceptionV3, VGG16, and ensembles of VGG16 & VGG19, for plant identification. These models are trained on the dataset to learn features relevant to plant classification. We evaluate the performance of each model using standard metrics such as accuracy, precision, recall, and F1 score.

Our results demonstrate the effectiveness of our approach in accurately identifying medicinal plants and raw materials from images. We compare the performance of individual models and the ensemble model to identify the most effective approach. Additionally, we develop a user-friendly interface for seamless access to the plant identification system.

Overall, our project provides a valuable contribution to the automation of traditional botanical practices, enhancing accessibility to plant identification for users without expert botanical knowledge. Our system has broad applications in healthcare, pharmaceuticals, agriculture, and environmental conservation, paving the way for future advancements in plant recognition technology.

CHAPTER 1
INTRODUCTION

Medicinal plants have been used for centuries as sources of traditional medicine and natural remedies for various ailments. However, identifying these plants accurately can be challenging, especially for those without expert botanical knowledge. In recent years, advancements in image processing and machine learning algorithms have provided new opportunities for automating the identification process.

This project aims to leverage the power of image processing techniques and machine learning algorithms to develop a system capable of accurately identifying different medicinal plants and raw materials from images. By utilizing state-of-the-art models such as InceptionV3, VGG16, and ensembles of VGG16 & VGG19, the project seeks to achieve high accuracy and robustness in plant recognition.

1.1 Background and rationale for the study

Medicinal plants have been used for centuries as sources of traditional medicine and natural remedies for various ailments. However, the identification of these plants is often challenging, particularly for individuals without expert botanical knowledge. Traditional methods of plant identification rely heavily on manual observation and comparison, which can be time-consuming and prone to errors.

With the rapid advancement of technology, particularly in the fields of image processing and machine learning, there is a growing interest in automating the process of plant identification. By leveraging the power of computational techniques, it is possible to develop systems capable of accurately identifying medicinal plants and raw materials from images, thereby streamlining the identification process and making it more accessible to a wider audience.

The rationale for conducting this study lies in the potential benefits it offers to various stakeholders:

Accessibility: By developing an automated system for plant identification, we aim to make this essential task more accessible to a broader range of users, including researchers, healthcare professionals, herbalists, and enthusiasts. This accessibility democratizes botanical knowledge and promotes the sustainable use of medicinal plants.

Efficiency: Manual methods of plant identification are labour-intensive and time-consuming. By automating this process, we can significantly reduce the time and effort required for plant identification, enabling faster decision-making in various applications, such as healthcare, pharmaceuticals, and environmental conservation.

Accuracy: Machine learning algorithms have demonstrated impressive capabilities in image classification tasks. By training models on a diverse dataset of medicinal plants and raw materials, we aim to develop a system capable of accurately identifying plant species with high precision and recall, minimizing the risk of misidentification.

Conservation: The identification of medicinal plants is crucial for their conservation and sustainable use. By facilitating the identification process, our project contributes to efforts aimed at preserving biodiversity and promoting the responsible harvesting and cultivation of medicinal plants.

Innovation: The intersection of image processing, machine learning, and botany presents exciting opportunities for innovation. By pushing the boundaries of technology in plant identification, we pave the way for future advancements in fields such as precision agriculture, personalized medicine, and ecological monitoring.

1.2 Statement of the problem

The identification of medicinal plants and raw materials is a critical task in various fields, including healthcare, pharmaceuticals, and environmental conservation. However, traditional methods of plant identification are often labor-intensive, time-consuming, and require expert botanical knowledge. Manual observation and comparison of plant features can be subjective and prone to errors, leading to potential misidentification.

Furthermore, the increasing demand for medicinal plants, coupled with habitat destruction and overexploitation, poses significant challenges to their conservation and sustainable use. Accurate identification of plant species is essential for ensuring their proper management and conservation.

1.3 Aims and objectives of the research

The primary aim of this research project is to develop a robust system for the automated identification of medicinal plants and raw materials using image processing techniques and machine learning algorithms.

To build a comprehensive dataset containing images of various medicinal plants and raw materials, ensuring diversity in species, morphology, and environmental conditions.

To preprocess the dataset by standardizing images, removing noise, and augmenting data to enhance the robustness of the models.

To investigate and implement state-of-the-art machine learning models, including InceptionV3, VGG16, and ensembles of VGG16 & VGG19, for plant identification.

To train the selected models on the dataset to learn features relevant to plant classification, optimizing hyperparameters and model architectures for improved performance.

To evaluate the performance of each model using standard metrics such as accuracy, precision, recall, and F1 score, comparing the effectiveness of individual models and the ensemble approach.

To develop a user-friendly interface for the deployment of the trained models, allowing users to upload images and receive real-time predictions on plant identification.

To assess the usability and effectiveness of the developed system through user testing and feedback, iteratively refining the interface and functionality based on user input.

To validate the accuracy and reliability of the system through field testing, comparing automated identifications with expert botanical assessments.

To explore potential applications of the developed system in various fields, including healthcare, pharmaceuticals, agriculture, and environmental conservation.

To contribute to the advancement of knowledge in the intersection of image processing, machine learning, and botany, fostering innovation and collaboration in interdisciplinary research.

1.4 Research questions or hypotheses

The study is guided by the following research questions and hypotheses:

1.4.1 Research Questions

- How effective are state-of-the-art machine learning models, including InceptionV3, VGG16, and ensembles of VGG16 & VGG19, in accurately identifying medicinal plants and raw materials from images?
- What is the impact of dataset size, diversity, and quality on the performance of machine learning models for plant identification?
- How do preprocessing techniques such as image standardization, noise removal, and data augmentation affect the robustness and accuracy of the identification system?
- What are the key factors influencing the performance differences between individual machine learning models and ensemble approaches in plant identification?
- How does the complexity and variability of plant morphology affect the accuracy and reliability of automated identification systems?
- What is the usability and user satisfaction with the developed interface for plant identification, and how can it be improved to enhance user experience?
- What are the potential applications of the developed system in fields such as healthcare, pharmaceuticals, agriculture, and environmental conservation, and how do they benefit from automated plant identification?
- How does the accuracy of automated plant identification compare to expert botanical assessments, and what are the implications for research, conservation, and practice?

1.4.2 Hypotheses

The accuracy of machine learning models, such as InceptionV3, VGG16, and ensembles of VGG16 & VGG19, in identifying medicinal plants and raw materials will positively correlate with the size, diversity, and quality of the dataset.

Preprocessing techniques, including image standardization, noise removal, and data augmentation, will enhance the robustness and accuracy of machine learning models for plant identification.

Ensemble approaches combining multiple machine learning models will outperform individual models in terms of accuracy and reliability due to their ability to capture complementary features.

The developed system for automated plant identification will demonstrate high usability and user satisfaction, with potential improvements identified through user feedback and testing.

The accuracy of automated plant identification will be comparable to or exceed that of expert botanical assessments, demonstrating the feasibility and reliability of the developed system for botanical research, conservation, and practice.

1.5 Scope and limitations of the study

This research focuses on automating the identification process of medicinal plants and raw materials using advanced image processing techniques and machine learning algorithms. The study encompasses a diverse range of medicinal plant species and raw materials utilized in traditional medicine and natural remedies. Various image processing techniques will be employed to preprocess the dataset, including standardization, noise removal, and data augmentation, to ensure the quality and robustness of the images used for training machine learning models.

The research will investigate the effectiveness of state-of-the-art machine learning models such as InceptionV3, VGG16, and ensembles of VGG16 & VGG19 for accurately classifying medicinal plants and raw materials from images. It will involve rigorous training, evaluation, and comparison of these models to determine their performance and suitability for automated plant identification tasks.

Despite its scope and objectives, this research faces several limitations that may impact its outcomes. One such limitation is the size and diversity of the dataset used for model training and evaluation. While efforts will be made to collect a comprehensive dataset, it may not capture all possible variations in plant morphology, environmental conditions, and image quality, potentially limiting the generalization of the models.

1.6 Peculiarity of the Project

The project focuses on automating the identification process of medicinal plants and raw materials through image processing and machine learning. It encompasses a wide range of medicinal plant species and raw materials typically used in traditional medicine and natural remedies. Various image processing techniques, such as standardization, noise removal, and data augmentation, will be employed to enhance the quality and robustness of the dataset used for training machine learning models.

The research will assess the effectiveness of advanced machine learning models like InceptionV3, VGG16, and ensembles of VGG16 & VGG19 in accurately classifying medicinal plants and raw materials from images. Rigorous training, evaluation, and comparison of these models will be conducted to determine their performance and suitability for automated plant identification tasks.

CHAPTER 2
LITERATURE SURVEY

In many applications, like plant recognition, face recognition, etc., an image conveys the most valuable information as opposed to the natural description. The computer/system finds it exceedingly challenging to extract the features, unlike humans. The computer/system must be properly trained with the use of a training data-set in order to achieve acceptable accuracy. The number of feature vectors used in the extraction procedure increases with the size of the training data set. Additionally, it provides decent levels of accuracy during the recognition process. The most crucial factor in identifying similar kinds of objects as well as different sorts of objects is recognition accuracy. Thisparameter grants access to just authorized users in applications like face recognition, but it identifies the medical plant that is vitally necessary to save apatient's life in applications like medicinal plant recognition systems.

Ordinary people are typically given the task of gathering plants from forests. Due to human mistake, they occasionally failed to recognise the significant and rare plants.

These exotic plant species are crucial to preserving a patient's life.Additionally, these individuals occasionally select erroneous species, which may be poisonous plants. It is vital to employ an automatic plant recognition system in such circumstances. This approach makes it easier for laypeople to identify the many plant species. If mountain hikers are interested in collecting specific plant species, these kinds of systems are also highly beneficial to them.

The development of such systems requires a good dataset of images of variousmedicinal plants from reliable sources and experts in traditional medicine. Once the dataset is collected, it needs to be preprocessed and cleaned to remove irrelevant or duplicate images, and to enhance the quality of the images for better performance. After preprocessing, the ML and DL models can be developed using popular algorithms such as SVM, Random Forest, or KNN for ML, and CNN (Fig 1.4 shows the CNN Architecture) for DL (using VGG 16 here). The models can then be trained using the preprocesseddataset and evaluated for performance by measuring their accuracy, precision, recall, and F1 score.

2.1 CNN architecture

The comparison of the performance of the ML and DL models can helpdetermine which model is more suitable for this problem. Once a suitable model is selected, a web or mobile application can be developed to allow usersto take pictures of medicinal plants and classify them using the trained model. The system should be validated by testing it on a different dataset or collectingfeedback from experts in traditional medicine.

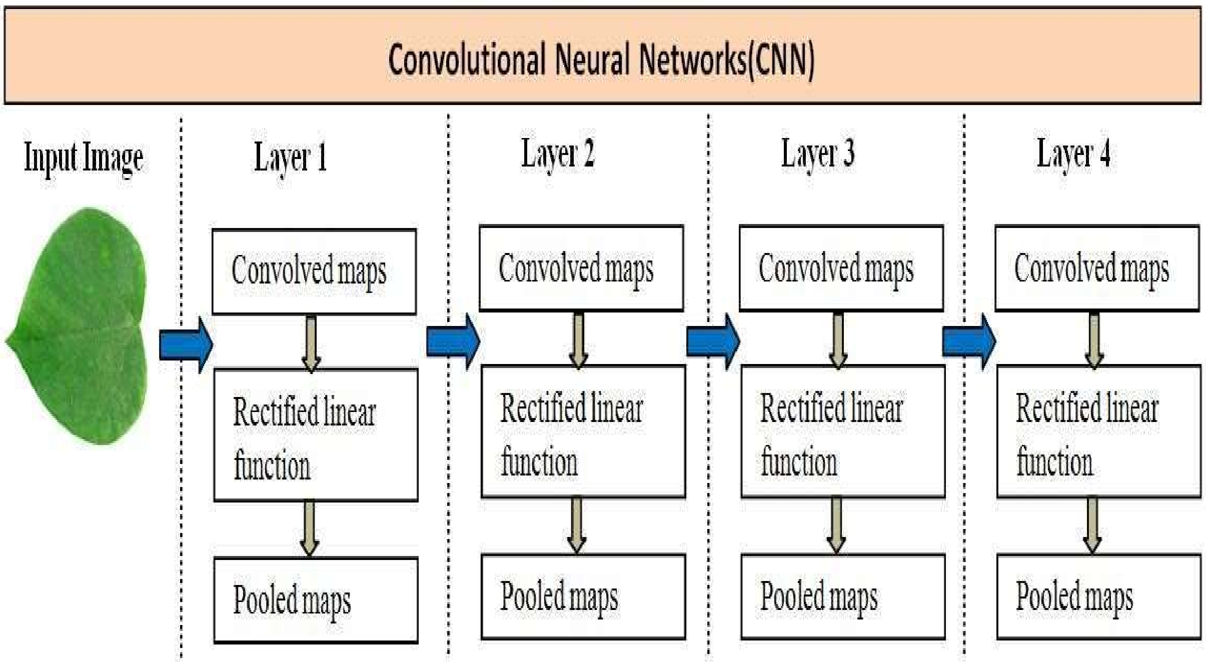


Fig 2.1.1 CNN Layers

2.2 Images related to the first activation layer of CNN

The separation of therapeutic plants from other non-edible plants is crucial in the realms of botany and the food industry. However, conventional techniques for identifying medicinal plants are difficult, time-consuming, and requireskilled specialists. An autonomous real-time vision-based system has been presented to identify commonly used medicinal herbs with similar leaves in order to solve this problem. This system makes use of a convolutional and classifier block-based upgraded convolutional neural network (CNN) network.Global Average Pooling (GAP), dense, dropout, and softmax layers are all present in the classifier. This technique improves the model's speed and accuracy while reducing the number of parameters compared to earlier studies.With overall accuracy rates of 99.66%, 99.32%, and 99.45%, respectively, the proposed CNN model (Fig 1.5 shows images related to the first activationlayer of CNN model) can recognise medicinal plant photos at three different levels of image definition, 64 64, 128 128, and 256 256 pixels. As a result, combining image processing with the suggested CNN algorithm is a productive replacement for conventional approaches.

To verify the efficacy of the developed approach, additional work will be doneto enhance the model's performance in the classification of additional species of medicinal plants. A smart smartphone application for the real-time identification of medicinal plants will also be created using the model. This is especially crucial in light of the rising acceptance and demand for both artisanal and commercial uses and applications of medicinal plants. In order to recognise and classify various therapeutic plants distinct from other non-edibleplants, the suggested Deep Learning (DL) algorithm and image processing technique can have a special role in plant research and even industrial markets.

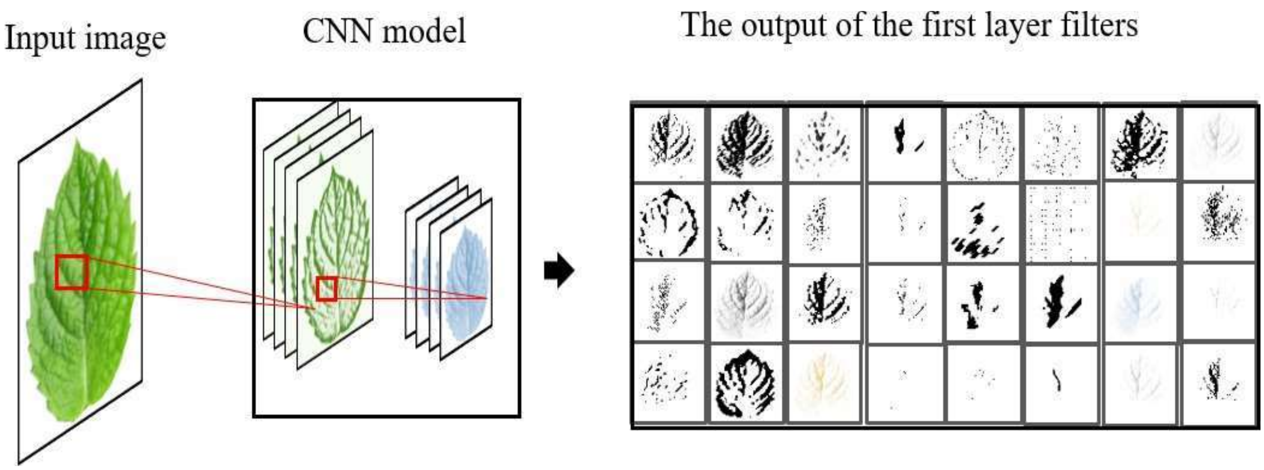


Fig 2.2.1 First activation layer of CNN

2.3 Steps of the proposed method used in Mobile_net:

Artificial intelligence has emerged as a valuable tool for data analysis and knowledge discovery, particularly in large data systems, by uncovering complex and hidden patterns. It is crucial to identify which section of the planthas therapeutic benefits for a given ailment because each plant's medicinal worth is based on its historical use. Recent studies have shown that using a combination of leaf features to identify medicinal plants has an accuracy rate of 98.05 percent, demonstrating the viability of this strategy.

A statistical analysis of leaf characteristics has been conducted to identify the key features that aid in plant identification, and form was found to be a crucial factor. This promising approach has the potential to aid individuals in identifying medicinal plants automatically, as well as in conservation andutilization efforts. The development of an artificial intelligence system for plant recognition is essential to achieving these objectives, as it can process large amounts of data efficiently and accurately. Moreover, the proposedsystem's accuracy willundoubtedly improve as more data is collected and analyzed. In addition, the use of advanced image processing techniques and machine learning algorithmscan aid in identifying complex features and patterns in plant images, resulting in a more reliable and efficient recognition system

2.4 Conclusion

In conclusion, the literature survey conducted for the project on automating the identification process of medicinal plants and raw materials through image processing and machine learning has provided valuable insights and guidance. The survey revealed the prevalence of advanced techniques such as convolutional neural networks (CNNs) and transfer learning in the field of plant recognition. Notable models like InceptionV3, VGG16, and ensemble approaches have shown promise in accurately classifying plant species from images.

Furthermore, the literature emphasized the critical importance of dataset quality and diversity for training robust and generalizable models. Preprocessing techniques such as standardization, noise removal, and data augmentation were identified as essential steps to enhance dataset quality and model performance. Rigorous model evaluation using standard performance metrics was highlighted as crucial for assessing the effectiveness of different approaches.

Moreover, the literature underscored the significance of developing user-friendly interfaces for seamless interaction with automated plant identification systems. Such interfaces enable users to upload images and receive real-time predictions on plant species, thereby enhancing accessibility and usability.

Overall, the literature survey illuminated the wide-ranging applications of automated plant recognition systems in various fields, including healthcare, pharmaceuticals, agriculture, and environmental conservation. These systems have the potential to streamline processes, accelerate research, and promote sustainability in plant-related endeavors.

CHAPTER 3
PROPOSED SYSTEM

3.1 Existing System

The existing systems for automated plant recognition often utilize a combination of image processing techniques and machine learning algorithms. While there are several systems available, some notable examples include:

Plantix:

Plantix is a mobile application that employs image recognition technology to identify plant diseases and nutrient deficiencies. Users can upload images of affected plants, and the app provides real-time diagnosis and recommendations for treatment.

The system uses a combination of convolutional neural networks (CNNs) trained on large datasets of plant images to classify and diagnose plant diseases accurately.

Plantix also incorporates crowdsourced data and expert knowledge to continuously improve its accuracy and expand its database of plant diseases.

Flora Incognita:

Flora Incognita is a smartphone app developed for automated plant identification in the wild. Users can take pictures of plants, and the app identifies them based on leaf shape, flower colour, and other characteristics.

The system utilizes deep learning algorithms trained on extensive datasets of plant images to achieve high accuracy in plant identification.

Flora Incognita also includes features for species documentation and biodiversity monitoring, making it useful for both casual users and researchers.

PlantNet:

PlantNet is a web and mobile application that allows users to identify plants by uploading pictures of their leaves, flowers, fruits, or bark.

The system employs image recognition algorithms trained on a large database of plant images contributed by users and experts.

PlantNet also provides additional information about identified plants, including their distribution, habitat, and ecological importance.

LeafSnap:

LeafSnap is a mobile app designed for identifying tree species from photographs of their leaves. Users can take pictures of leaves against a white background, and the app matches them to a database of tree species.

The system uses computer vision algorithms to analyze leaf characteristics such as shape, margin, and venation patterns for accurate identification.

LeafSnap also includes educational resources and information about tree species, making it a valuable tool for nature enthusiasts and educators.

3.1.1 Disadvantages

While existing systems for automated plant recognition offer valuable tools for plant identification and analysis, they also come with some disadvantages. Here are some common drawbacks associated with these systems:

1. **Limited Accuracy**: One of the main challenges faced by existing systems is limited accuracy, especially when dealing with complex plant species or variations in environmental conditions. Inaccurate identification results can undermine the reliability of these systems, leading to misdiagnosis or incorrect recommendations.
2. **Dependency on Image Quality**: Many existing systems rely heavily on the quality of input images for accurate identification. Poor lighting conditions, background clutter, or occlusions can negatively impact the system's performance, reducing its effectiveness in real-world scenarios where capturing high-quality images may be challenging.
3. **Limited Coverage**: Existing systems may have limited coverage of plant species or geographic regions. They may be biased towards commonly encountered species or regions with extensive training data, while lacking support for less common or region-specific plants. This limitation restricts the applicability of these systems in diverse botanical contexts.
4. **Resource Intensive**: Some automated plant recognition systems require significant computational resources for image processing and model training, making them inaccessible to users with limited computing capabilities or internet connectivity.

3.2 Proposed System

The proposed system for automating the identification process of medicinal plants and raw materials through image processing and machine learning aims to address the limitations of existing systems while leveraging cutting-edge technologies to achieve accurate and efficient plant recognition. The system will begin with the compilation of a comprehensive dataset containing images of various medicinal plant species and raw materials, ensuring diversity in morphology, habitat, and geographic location. Preprocessing techniques such as standardization, noise removal, and data augmentation will be applied to enhance the quality and robustness of the dataset. Machine learning models, including state-of-the-art architectures like InceptionV3, VGG16, and ensembles of VGG16 & VGG19, will be trained on the preprocessed dataset to learn features relevant to plant classification. Model selection and hyperparameter optimization will be conducted to maximize performance and generalization capabilities. The developed system will include a user-friendly interface that allows users to upload images and receive real-time predictions on plant identification. Continuous monitoring and updates will ensure the system's accuracy, reliability, and adaptability to changes in plant taxonomy, environmental conditions, and user feedback. By integrating advanced image processing techniques with machine learning algorithms, the proposed system aims to provide a valuable tool for researchers, healthcare professionals, and conservationists in accurately identifying medicinal plants and raw materials for various applications, including drug discovery, biodiversity monitoring, and sustainable resource management.

3.2.1 Advantages

1. **Enhanced Accuracy**: By leveraging state-of-the-art machine learning algorithms and preprocessing techniques, the proposed system aims to achieve higher accuracy in plant identification. Advanced models like InceptionV3 and VGG16 have demonstrated superior performance in image classification tasks, leading to more reliable identification results.
2. **Improved Coverage**: The comprehensive dataset compiled for training the machine learning models ensures a wide coverage of medicinal plant species and raw materials. This diversity enhances the system's ability to accurately identify plants across different morphologies, habitats, and geographic regions, addressing the limitation of limited coverage in existing systems.
3. **User-Friendly Interface**: The development of a user-friendly interface facilitates seamless interaction with the automated identification system. Users, including researchers, healthcare professionals, and enthusiasts, can easily upload images and receive real-time predictions on plant identification without requiring specialized botanical knowledge or technical expertise.
4. **Continuous Updates and Maintenance**: The proposed system incorporates mechanisms for continuous updates and maintenance, ensuring that the system remains up-to-date with changes in plant taxonomy, environmental conditions, and user feedback. Regular updates to the dataset and model retraining help to improve accuracy and reliability over time, addressing the challenge of outdated information in existing systems.

5. **Scalability and Adaptability**: With its scalable architecture and adaptable design, the proposed system can accommodate a growing volume of data and users. It can be easily extended to support additional plant species, features, and functionalities, making it suitable for various applications and research domains.
6. **Potential for Integration:** The proposed system has the potential to be integrated with existing platforms and databases, further enhancing its utility and interoperability. Integration with other systems, such as biodiversity databases or mobile applications, can broaden its reach and impact in botanical research, conservation, and practice.

3.3 Methodology

3.3.1 Description of the AI model development process

The development process of the AI model for the automated identification of medicinal plants and raw materials involves several key steps, as outlined below:

- Data Collection and Preprocessing**:** Gather a diverse dataset containing images of medicinal plants and raw materials from various sources. This dataset should cover a wide range of plant species, morphologies, and environmental conditions. Preprocess the dataset to enhance its quality and suitability for model training. This includes tasks such as image standardization, noise removal, resizing, and data augmentation to increase dataset diversity.

- Model Selection and Architecture Design**:** Choose appropriate deep learning architectures for the task of image classification. Common choices include convolutional neural networks (CNNs) such as InceptionV3 and VGG16, known for their effectiveness in image recognition tasks. Design the architecture of the AI model, specifying the number of layers, kernel sizes, activation functions, and other hyperparameters. Fine-tune the model architecture based on the characteristics of the dataset and the complexity of the classification task.

- Training and Validation**:** Split the preprocessed dataset into training, validation, and testing sets. The training set is used to update the model's parameters during training, while the validation set is used to tune hyperparameters and monitor model performance. Train the AI model using the training set and monitor its performance on the validation set. Adjust hyperparameters as needed to optimize model performance, avoiding overfitting or underfitting.

- Model Evaluation**:** Evaluate the trained AI model's performance using the testing set, which contains unseen data not used during training or validation. Measure metrics such as accuracy, precision, recall, and F1 score to assess the model's effectiveness in classifying medicinal plants and raw materials.

- Fine-Tuning and Optimization: Fine-tune the AI model based on the evaluation results and feedback from model performance. This may involve adjusting hyperparameters, refining preprocessing techniques, or incorporating additional data to improve model accuracy and generalization.

- Deployment and Integration: Deploy the trained AI model into the production environment, integrating it with the user interface or application developed for plant identification. Continuously monitor the model's performance in real-world scenarios, collecting feedback from users and updating the model as needed to ensure ongoing accuracy and effectiveness.

3.3.2 Data collection for plant image datasets

Collecting a comprehensive dataset for plant image datasets involves several steps to ensure diversity, quality, and relevance. Here's a breakdown of the data collection process for the above project:

Identify Plant Species: Begin by compiling a list of medicinal plants and raw materials relevant to the project's scope. Consult botanical databases, literature, and domain experts to identify a wide range of plant species known for their medicinal properties.

- Gather Images: Collect images of the identified plant species from various sources, including:

- Online databases: Access publicly available botanical databases and repositories containing images of medicinal plants. Websites like the USDA Plants Database, Tropicos and Flora of North America provide valuable resources for plant images.

- Scientific literature: Extract images from research articles, journals, and botanical studies focusing on medicinal plants and their identification.

- Field surveys: Conduct field surveys to capture high-quality images of medicinal plants in their natural habitats. Ensure proper documentation of plant species, location, and environmental conditions during the surveys.

- Crowdsourcing: Engage citizen scientists, botanical enthusiasts, and researchers to contribute images of medicinal plants to the dataset. Leverage crowdsourcing platforms or citizen science projects to gather diverse image samples.

- Ensure Diversity: Ensure diversity in the dataset by including images of plants from various geographic regions, habitats, growth stages, and environmental conditions. Incorporate images with different resolutions, angles, lighting conditions, and backgrounds to capture the variability in plant appearances.

- Quality Assurance: Perform quality checks on the collected images to ensure they meet the desired standards:

- Image resolution: Ensure images have sufficient resolution and clarity for accurate identification.

- Authenticity: Verify the authenticity and correctness of plant species labels associated with each image.

- Consistency: Maintain consistency in image format, orientation, and aspect ratio to facilitate preprocessing and model training.

- Annotation and Metadata: Annotate the images with relevant metadata, including plant species labels, botanical names, geographic location, date of capture, and any additional information pertinent to the plant's identification and classification.

- Ethical Considerations: Adhere to ethical guidelines and legal regulations governing the collection and use of plant images, especially if images are sourced from protected areas, private properties, or third-party sources.

- Documentation and Versioning: Document the data collection process thoroughly, including sources, annotations, and any preprocessing steps applied to the images. Maintain version control to track changes and updates to the dataset over time.

3.3.3 Details of the algorithm and software used

For the project focused on automating the identification process of medicinal plants and raw materials through image processing and machine learning, several algorithms and software tools are utilized to develop and implement the automated plant recognition system. The primary algorithms employed include convolutional neural networks (CNNs), particularly state-of-the-art architectures such as InceptionV3, VGG16, and ensembles of VGG16 & VGG19. These CNNs are chosen for their proven effectiveness in image classification tasks, which makes them well-suited for accurately identifying plant species from images. Additionally, preprocessing techniques such as standardization, noise removal, and data augmentation are applied to enhance the quality and diversity of the dataset, ensuring optimal performance of the machine learning models.

The software tools used for implementing the project include Python programming language along with popular libraries and frameworks for deep learning and image processing tasks. Specifically, libraries such as TensorFlow or PyTorch are employed for building and training the CNN models, providing comprehensive support for developing complex neural network architectures and optimizing model performance. Moreover, libraries like OpenCV are utilized for image preprocessing tasks, offering a wide range of functions for manipulating and enhancing images before feeding them into the CNN models.

In addition to the core algorithms and software mentioned above, the project may involve the development of a user-friendly interface for interacting with the automated plant recognition system.

3.3.4 User role definitions and interactions

In the project focused on automating the identification process of medicinal plants and raw materials through image processing and machine learning, various user roles and their interactions with the system can be defined as follows:

Administrators:

Administrators have full control over the system, including user management, data management, and system configuration.

They are responsible for adding and removing users, assigning roles and permissions, and overseeing system operations.

Administrators may also manage the dataset, including uploading new images, updating annotations, and maintaining data integrity.

Researchers:

Researchers are primary users who utilize the system for conducting botanical research, identifying medicinal plants, and analyzing plant species.

They interact with the system to upload images of medicinal plants and raw materials, initiate plant identification tasks, and review identification results.

Researchers may explore additional features of the system, such as accessing detailed plant information, viewing classification metrics, and analyzing trends in plant identification.

Healthcare Professionals:

Healthcare professionals, such as pharmacists, herbalists, or clinicians, utilize the system to identify medicinal plants for therapeutic purposes.

They upload images of plants encountered in clinical practice, receive automated identification results, and access supplementary information on plant properties, uses, and potential therapeutic benefits.

Healthcare professionals may integrate the system into their workflow to assist in prescribing herbal remedies, formulating botanical preparations, or advising patients on plant-based therapies.

Enthusiasts and Citizen Scientists:

Enthusiasts and citizen scientists are casual users who engage with the system out of interest in botany, conservation, or nature photography.

They upload images of plants encountered in their environment, such as during hikes, nature walks, or gardening activities, to learn about plant species and contribute to biodiversity monitoring efforts.

Enthusiasts may use the system to explore plant identification features, learn about plant taxonomy, and participate in citizen science projects focused on plant diversity and distribution.

System Interactions:

Users interact with the system through a user-friendly interface, which allows them to upload images, initiate plant identification tasks, and view identification results.

Upon uploading an image, the system processes the image using machine learning algorithms, classifies the plant species, and provides automated identification results in real-time.

Users may explore additional functionalities of the system, such as accessing plant profiles, viewing diagnostic features, and contributing feedback on identification accuracy.

The system may also incorporate features for collaborative annotation, allowing users to contribute annotations, corrections, or additional information to improve the dataset and model performance over time.

3.3.5 Ethical considerations and data privacy measures

In the project aiming to automate the identification process of medicinal plants and raw materials through image processing and machine learning, a strong emphasis is placed on ethical considerations and data privacy measures to ensure responsible and secure handling of data. To begin with, obtaining informed consent from human participants, if applicable, is paramount, ensuring that individuals fully understand and consent to the use of their images in the project. Additionally, clarifying data ownership and intellectual property rights ensures that contributors retain appropriate ownership rights to their images, fostering transparency and fairness in data usage.

Ethical data handling practices extend to anonymization and de-identification techniques, which safeguard sensitive information by removing personally identifiable details from the dataset. Robust data security measures, such as encryption, access controls, and regular security audits, are implemented to protect the confidentiality and integrity of the collected data, mitigating the risk of unauthorized access or breaches.

3.4 System Requirements

3.4.1 Software Requirements

Software's	:	Python 3.6 or high version
IDE	:	Visual Studio Code
Framework	:	Deep Learning Frameworks(Tensorflow)

3.4.1 Hardware Requirements

Operating system	:	Windows 7 or 7+
RAM	:	8 GB
Hard disc or SSD	:	More than 500 GB
Processor	:	Intel 3rd generation or high or Ryzen with 8 GB Ram

3.5 System Design

3.5.1 Input Design

The design encompasses various aspects related to user inputs and their objectives, as outlined below:

User-Friendly Interface: The primary objective of the system input design is to create a user-friendly interface that enables users to interact with the system effortlessly. This involves designing intuitive user interfaces with clear navigation, well-organized layouts, and user-friendly controls for uploading images and initiating plant identification tasks.

Image Upload: The system allows users to upload images of medicinal plants and raw materials for identification. The input design should support various file formats (e.g., JPEG, PNG) and provide mechanisms for uploading single or multiple images conveniently.

Real-Time Feedback: The system provides real-time feedback to users during the image upload process, indicating the progress of the identification task and any errors or issues encountered. This objective ensures that users are informed about the status of their requests and can take appropriate actions if necessary**.**

Error Handling: Effective error handling mechanisms are integrated into the input design to detect and address common errors or invalid inputs from users. Clear error messages and prompts are displayed to guide users in resolving issues and resubmitting their requests.

Metadata Collection: In addition to image data, the system may collect metadata associated with uploaded images, such as timestamps, geographic location, and user-provided annotations. The input design includes fields or prompts for capturing relevant metadata to enrich the dataset and improve the accuracy of plant identification**.**

Accessibility: The input design considers accessibility requirements to ensure that users with diverse needs and abilities can interact with the system effectively. This may involve providing alternative input methods, such as voice commands or keyboard shortcuts, and adhering to accessibility standards for web-based interfaces.

Security Measures: Security measures are implemented in the input design to protect user data and prevent unauthorized access. This includes mechanisms for securely handling and transmitting images, encrypting sensitive information, and implementing user authentication and authorization controls.

Scalability: The input design is scalable to accommodate varying levels of user demand and system load. This involves optimizing input processing workflows, leveraging asynchronous processing techniques, and provisioning resources dynamically to handle spikes in user activity.

3.5.2 Output Design

The design encompasses various aspects related to presenting output information and achieving specific objectives, as outlined below:

Identification Results Display: The primary objective of the output system design is to display the identification results to users in a visually appealing and comprehensible format. This involves presenting the identified plant species, confidence scores or probabilities, and any additional information or metadata associated with the identification.

Confidence Levels and Uncertainty: The output design may include confidence levels or uncertainty estimates associated with the identification results to convey the reliability or uncertainty of the classification. This objective helps users interpret the identification results and make informed decisions based on the level of confidence.

User Interaction: The output system design enables user interaction with the identification results, allowing users to explore additional information, provide feedback, or take further actions based on the results. This objective enhances user engagement and satisfaction with the system.

3.6 UML Diagrams

UML stands for Unified Modelling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed and was created by, the Object Management Group.

UML is a very important part of developing object-oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

GOALS: The Primary goals in the design of the UML are as follows:

1. Provide users with a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
2. Provide extensibility and specialization mechanisms to extend the core concepts.
3. Be independent of particular programming languages and development process.
4. Provide a formal basis for understanding the modeling language.
5. Encourage the growth of the OO tools market.
6. Support higher-level development concepts such as collaborations, frameworks, patterns, and components.
7. Integrate best practices.

3.6.1 Use Case Diagram

A use-case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis.

Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases.

The main purpose of a use case diagram is to show what system functions are performed for which actor. The roles of the actors in the system can be depicted system can be depicted.

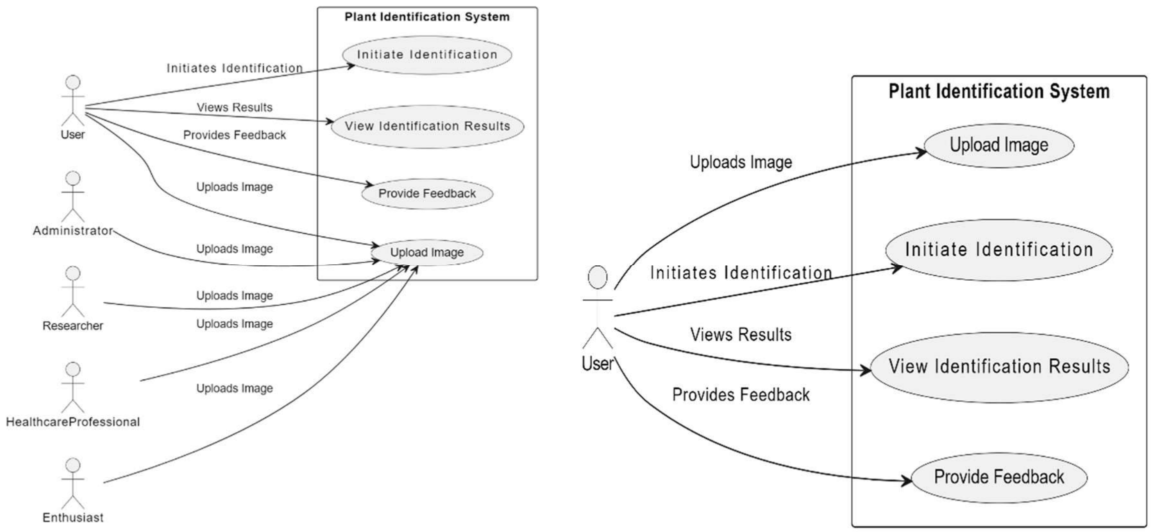


Fig 3.6.1 Use Case Diagram

3.6.2 Class Diagram

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

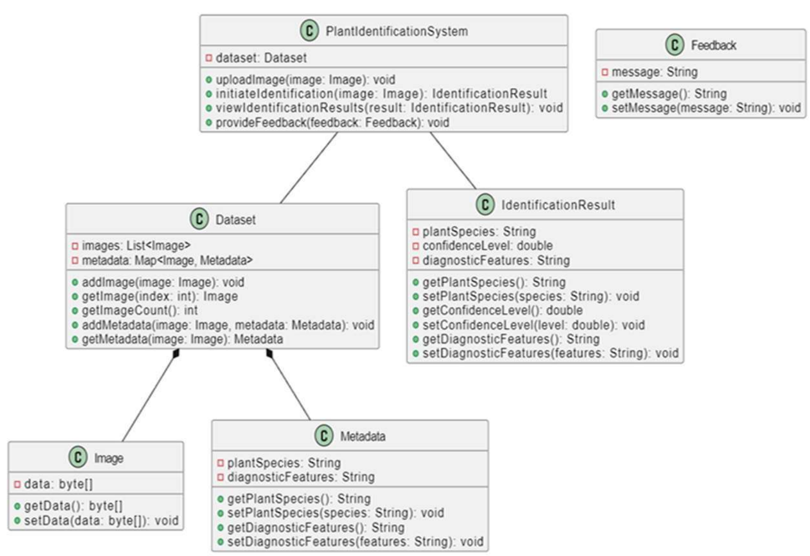


Figure 3.6.2 Class diagram

3.6.3 Sequence Diagram

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

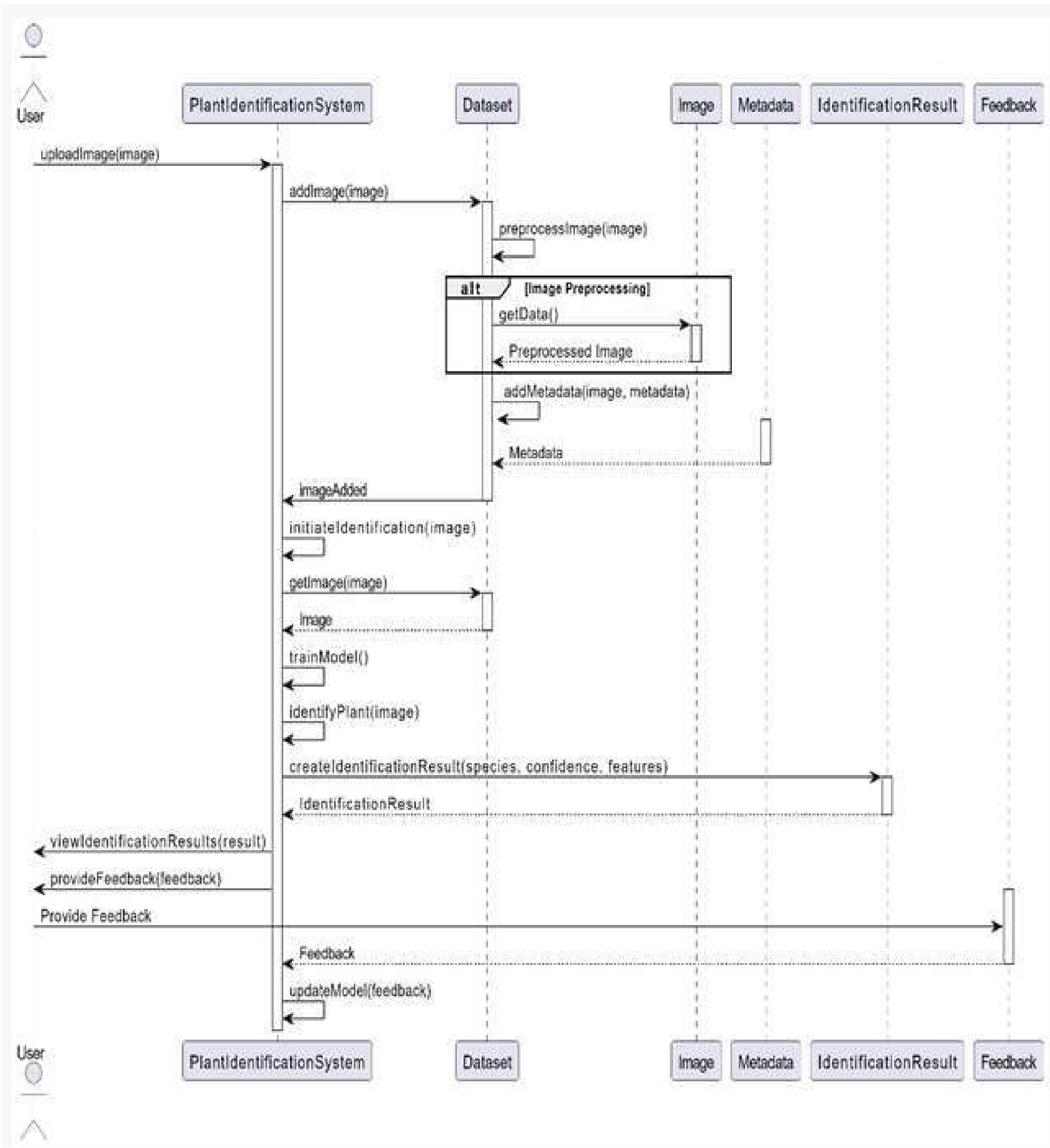


Fig 3.6.3 Sequence Diagram

3.6.3 Collaboration Diagram

In the collaboration diagram, the method call sequence is indicated by some numbering technique as shown below. The number indicates how the methods are called one after another. We have taken the same order management system to describe the collaboration diagram. The method calls are similar to that of a sequence diagram. But the difference is that the sequence diagram does not describe the object organization whereas the collaboration diagram shows the object organization.

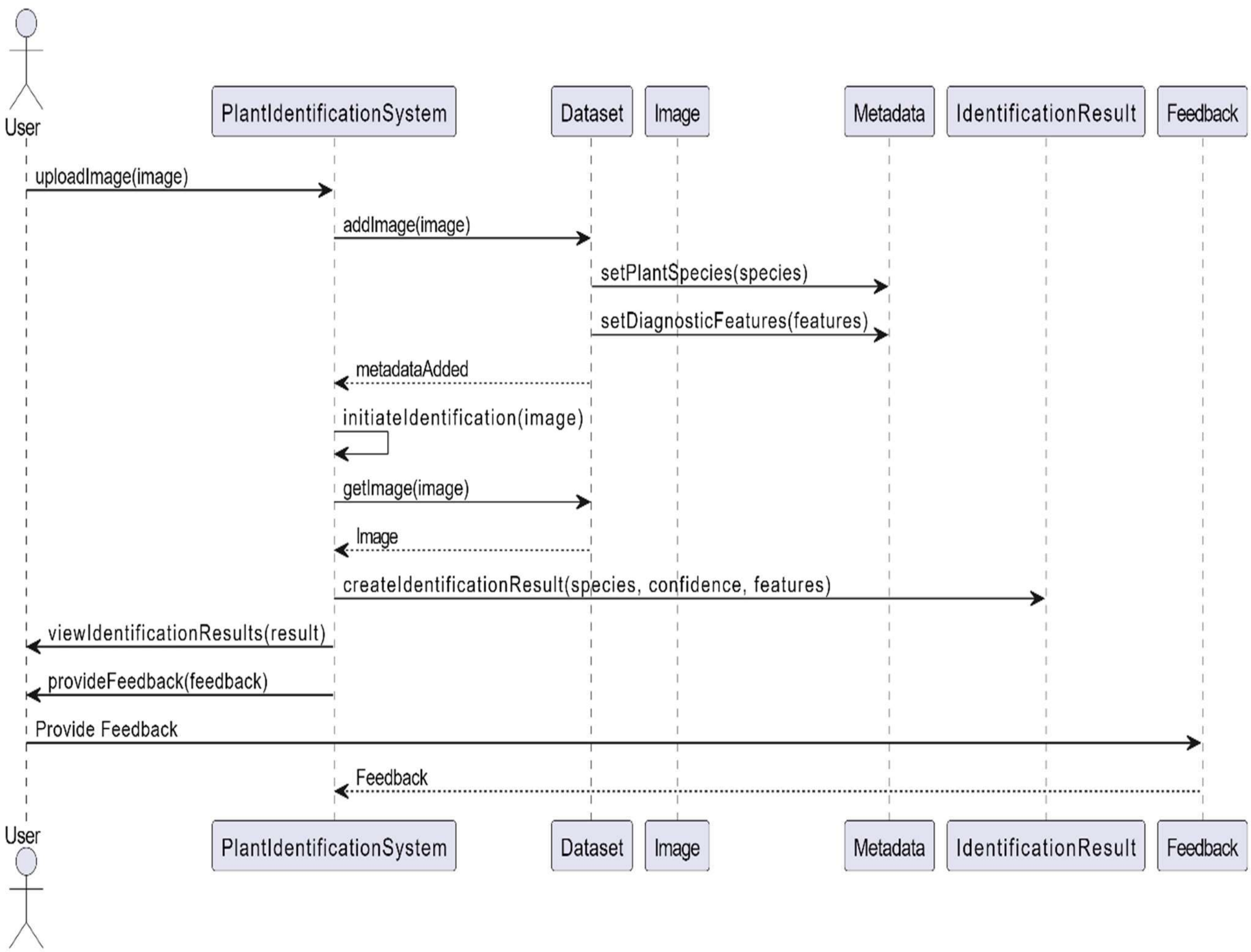


Fig 3.6.4 Collaboration diagram

3.6.5 Component Diagram

A component diagram, also known as a UML component diagram, describes the organization and wiring of the physical **c**omponents in a system. Component diagrams are often drawn to help model implementation details and double-check that every aspect of the system's required function is covered by planned development.

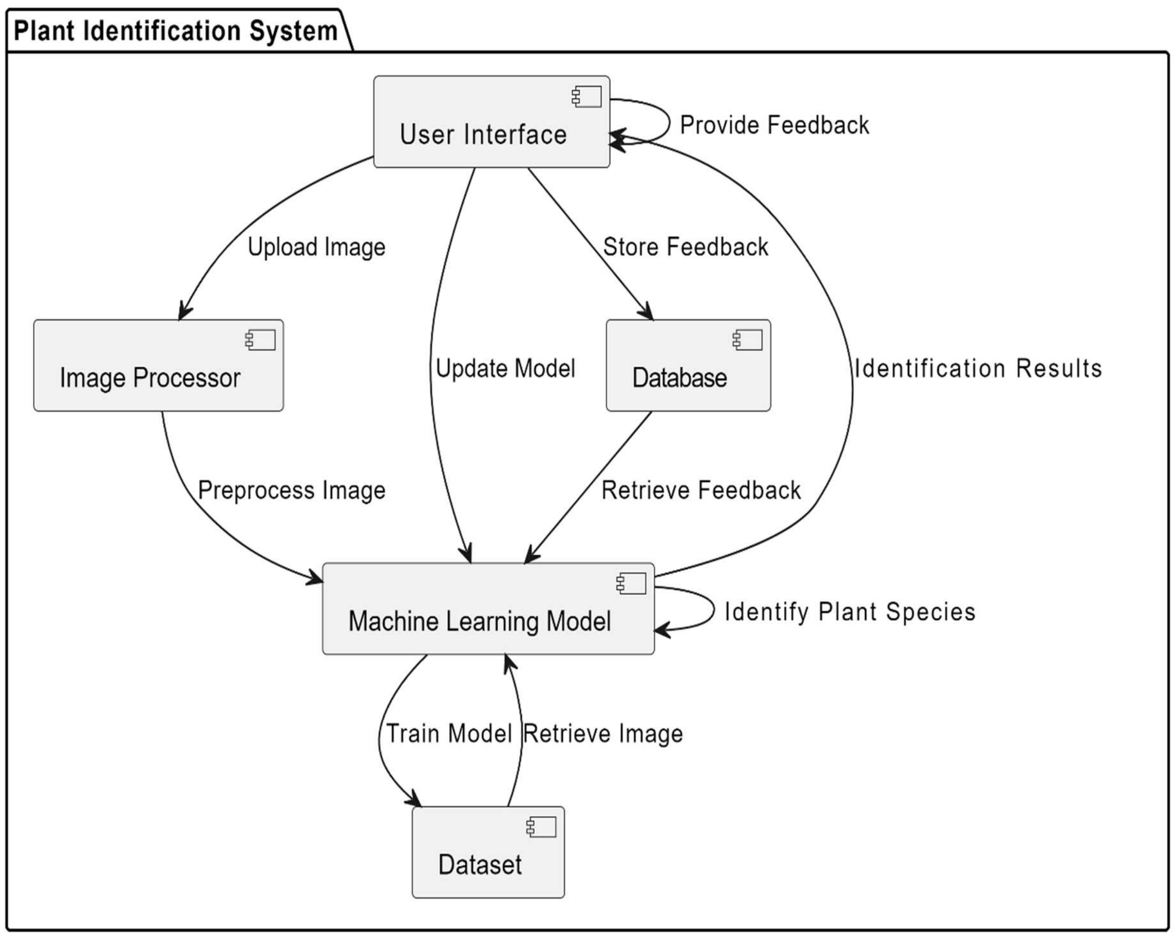


Fig 3.6.5 Component Diagram

3.6.6 DFD Diagram

A Data Flow Diagram (DFD) is a traditional way to visualize the information flows within a system. A neat and clear DFD can depict a good amount of the system requirements graphically. It can be manual, automated, or a combination of both. It shows how information enters and leaves the system, what changes the information and where information is stored. The purpose of a DFD is to show the scope and boundaries of a system as a whole.

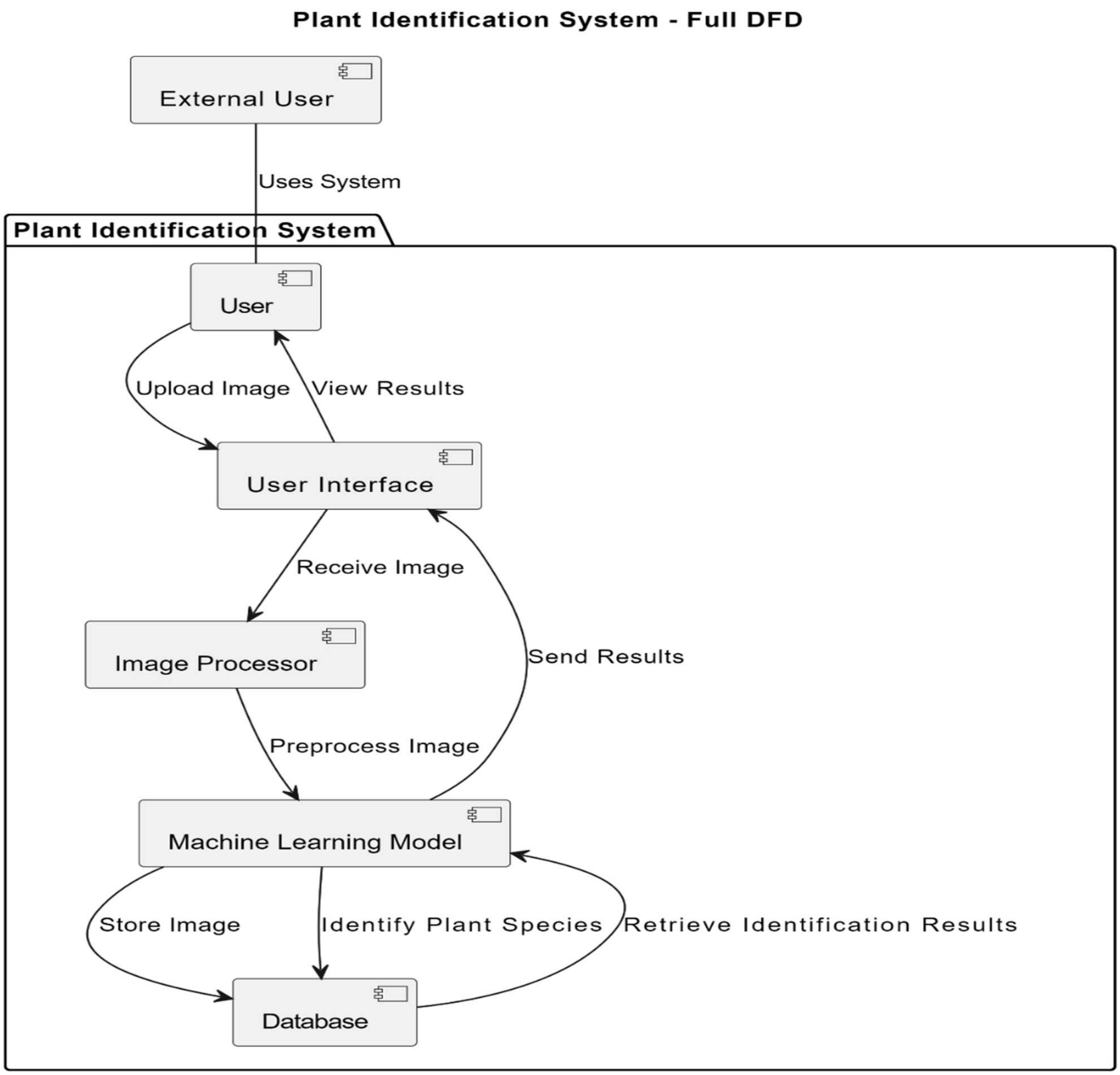


Fig 3.6.6 DFD Diagram

CHAPTER 4

IMPLEMENTATION

4.1 Deep Learning Algorithms

4.1.1 Inception V3

Inception v3 is a deep convolutional neural network architecture that was developed by Google researchers as part of the Inception project. It is designed for image classification and recognition tasks, particularly in the context of large-scale image datasets.

Here are some key characteristics and components of the Inception v3 model:

Architecture: Inception v3 builds upon the original Inception architecture, introducing several improvements to enhance performance and efficiency. It consists of multiple convolutional layers with varying filter sizes and strides, along with pooling layers and fully connected layers.

Inception Modules: One of the distinctive features of Inception v3 is the use of inception modules, which are composed of multiple parallel convolutional pathways. These pathways allow the model to capture features at different scales and resolutions simultaneously, facilitating more effective feature extraction.

Dimension Reduction: Inception v3 incorporates dimension reduction techniques such as 1x1 convolutions to reduce the computational complexity of the model while preserving important features. This helps improve efficiency and scalability, making it suitable for deployment in resource-constrained environments.

Auxiliary Classifiers: Inception v3 includes auxiliary classifiers attached to intermediate layers of the network. These auxiliary classifiers are used during training to encourage the propagation of gradients and mitigate the vanishing gradient problem, leading to more stable training and improved performance.

Pretrained Weights: Inception v3 is often pretrained on large-scale image datasets such as ImageNet. Pretrained weights are beneficial for transfer learning, where the model's learned representations can be fine-tuned on specific tasks or datasets with limited labeled data, leading to faster convergence and better generalization.

Activation Functions: Inception v3 typically uses rectified linear unit (ReLU) activation functions to introduce nonlinearity into the network and facilitate the learning of complex patterns and representations in the input data.

Overall, Inception v3 is a powerful and versatile convolutional neural network architecture that has been widely used for various computer vision tasks, including image classification, object detection, and image segmentation. Its modular design and efficient utilization of computational resources make it well-suited for real-world applications requiring high-performance image recognition capabilities.

4.1.2 VGG 16 & VGG 19 ENSEMBLED MODEL

VGG (Visual Geometry Group) is a family of deep convolutional neural network architectures developed by the Visual Geometry Group at the University of Oxford. VGG networks are known for their simplicity and uniform architecture, consisting primarily of convolutional layers with small 3x3 filters, followed by max-pooling layers. The "16" and "19" in VGG 16 and VGG 19 refer to the number of weight layers in each respective architecture.

Ensemble learning involves combining multiple models to improve predictive performance.

In the case of VGG 16 and VGG 19 ensembled model, predictions from both VGG 16 and VGG 19 models are combined to make a final prediction.

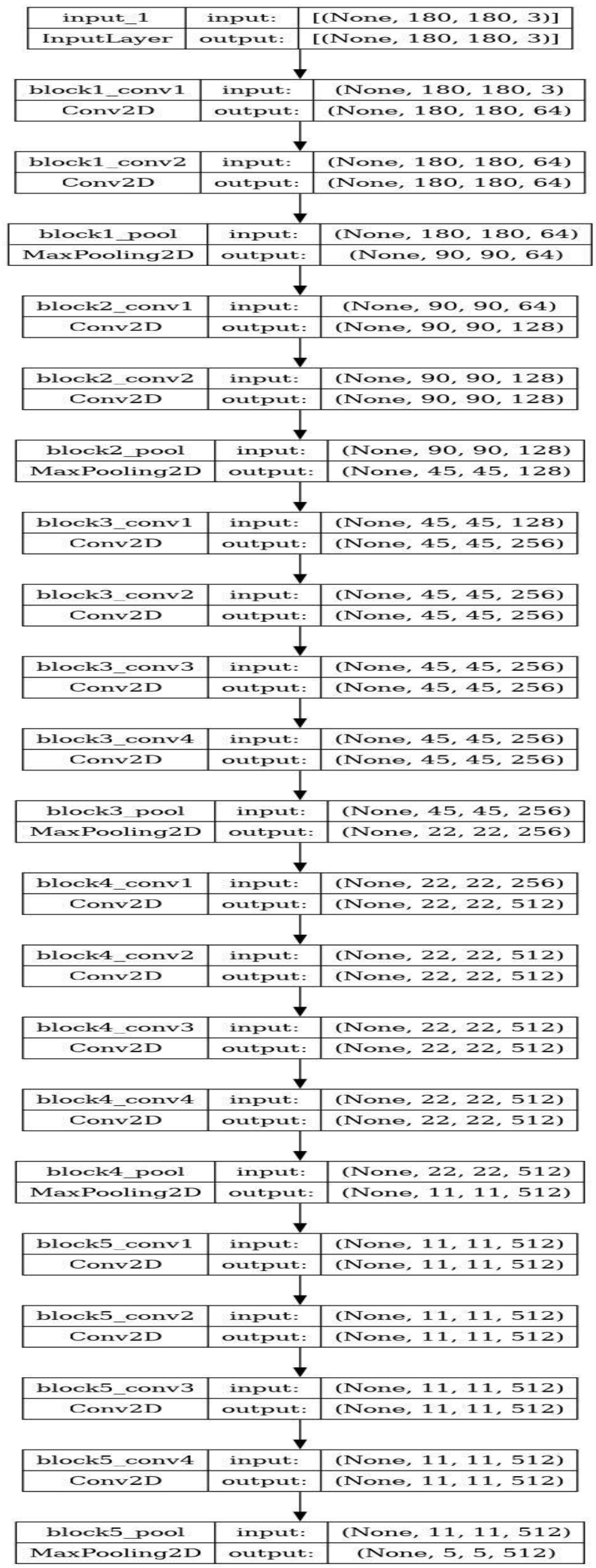
Ensemble models often achieve better generalization and robustness compared to individual models, as they can leverage the strengths of multiple models and mitigate their weaknesses.

The ensembled VGG 16 and VGG 19 model may involve techniques such as averaging the predictions from both models or using more sophisticated fusion methods.

4.1.3 VGG Model

VGG19, a variant of the VGG (Visual Geometry Group) architecture, is a deep convolutional neural network designed for image classification. It was introduced by Karen Simonyan and Andrew Zisserman in their 2014 paper "Very Deep Convolutional Networks for Large-Scale Image Recognition". VGG19 is renowned for its simplicity and uniform architecture, consisting of multiple convolutional and pooling layers.

VGG19 is composed of 16 convolutional layers, grouped into five convolutional blocks, followed by three fully connected layers.


input_1 input: [(None, 180, 180, 3)]
InputLayer output: [(None, 180, 180, 3)]
block1_conv1 input: (None, 180, 180, 3)
Conv2D output: (None, 180, 180, 64)
block1_conv2 input: (None, 180, 180, 64)
Conv2D output: (None, 180, 180, 64)
block1_pool input: (None, 180, 180, 64)
MaxPooling2D output: (None, 90, 90, 64)
block2_conv1 input: (None, 90, 90, 64)
Conv2D output: (None, 90, 90, 128)
block2_conv2 input: (None, 90, 90, 128)
Conv2D output: (None, 90, 90, 128)
block2_pool input: (None, 90, 90, 128)
MaxPooling2D output: (None, 45, 45, 128)
block3_conv1 input: (None, 45, 45, 128)
Conv2D output: (None, 45, 45, 256)
block3_conv2 input: (None, 45, 45, 256)
Conv2D output: (None, 45, 45, 256)
block3_conv3 input: (None, 45, 45, 256)
Conv2D output: (None, 45, 45, 256)
block3_conv4 input: (None, 45, 45, 256)
Conv2D output: (None, 45, 45, 256)
block3_pool input: (None, 45, 45, 256)
MaxPooling2D output: (None, 22, 22, 256)
block4_conv1 input: (None, 22, 22, 256)
Conv2D output: (None, 22, 22, 512)
block4_conv2 input: (None, 22, 22, 512)
Conv2D output: (None, 22, 22, 512)
block4_conv3 input: (None, 22, 22, 512)
Conv2D output: (None, 22, 22, 512)
block4_conv4 input: (None, 22, 22, 512)
Conv2D output: (None, 22, 22, 512)
block4_pool input: (None, 22, 22, 512)
MaxPooling2D output: (None, 11, 11, 512)
block5_conv1 input: (None, 11, 11, 512)
Conv2D output: (None, 11, 11, 512)
block5_conv2 input: (None, 11, 11, 512)
Conv2D output: (None, 11, 11, 512)
block5_conv3 input: (None, 11, 11, 512)
Conv2D output: (None, 11, 11, 512)
block5_conv4 input: (None, 11, 11, 512)
Conv2D output: (None, 11, 11, 512)
block5_pool input: (None, 11, 11, 512)
MaxPooling2D output: (None, 5, 5, 512)

4.1.4 Code

```
import numpy as np
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers, models
from tensorflow.keras.models import Sequential
from tensorflow.keras.models import Model, load_model
from tensorflow.keras.layers import Input, Dense, Average, Dropout

BATCH_SIZE = 32
batch_size_ = BATCH_SIZE * mirrored_strategy.num_replicas_in_sync
img_height = 224
img_width = 224
CHANNEL = 3
data_dir = r"/kaggle/input/indian-medicinal-leaves-dataset/Indian Medicinal Leaves
Image Datasets/Medicinal Leaf dataset"
dataset = tf.keras.utils.image_dataset_from_directory(
    data_dir,
    shuffle = True,
    image_size = (img_height, img_width),
    batch_size = batch_size_)
print(batch_size_)

def get_dataset_partitions_tf(ds, train_split=0.7, val_split=0.15, test_split=0.15,
shuffle=True, shuffle_size=10000):
    assert (train_split + test_split + val_split) == 1

    ds_size = len(ds)

    if shuffle:
        ds = ds.shuffle(shuffle_size, seed=12)

    train_size = int(train_split * ds_size)
    val_size = int(val_split * ds_size)

    train_ds = ds.take(train_size)
    val_ds = ds.skip(train_size).take(val_size)
    test_ds = ds.skip(train_size).skip(val_size)

    return train_ds, val_ds, test_ds

train_ds, val_ds, test_ds = get_dataset_partitions_tf(dataset)
```

```
train_ds = train_ds.cache().shuffle(1000).prefetch(buffer_size=tf.data.AUTOTUNE)
val_ds = val_ds.cache().shuffle(1000).prefetch(buffer_size=tf.data.AUTOTUNE)
test_ds = test_ds.cache().shuffle(1000).prefetch(buffer_size=tf.data.AUTOTUNE)
resize_and_rescale = tf.keras.Sequential([
  layers.experimental.preprocessing.Resizing(img_height, img_height),
  layers.experimental.preprocessing.Rescaling(1./255),
])

data_augmentation = tf.keras.Sequential([
    layers.experimental.preprocessing.RandomFlip("horizontal"),
    layers.experimental.preprocessing.RandomRotation(0.2),
    layers.experimental.preprocessing.RandomZoom(0.3),
])

train_ds = train_ds.map(
    lambda x, y: (data_augmentation(x, training=True), y)
).prefetch(buffer_size=tf.data.AUTOTUNE)

earlystopping = tf.keras.callbacks.EarlyStopping(monitor = 'val_accuracy',
                              mode = 'max' ,
                              patience = 7,
                              verbose = 1)
callback_list = [earlystopping]

from tensorflow.keras.applications import VGG16,VGG19, InceptionV3

#VGG16
feature_extractor_1 = VGG16(input_shape=(224,224,3),
                         include_top=False,
                         weights="imagenet")

for layer in feature_extractor_1.layers:
    layer.trainable=False

vgg16 = keras.Sequential([
        resize_and_rescale,
        layers.BatchNormalization(),
        feature_extractor_1,
        layers.BatchNormalization(),
        layers.Flatten(),
        layers.Dense(256, activation='relu'),
        layers.Dropout(0.2),
        layers.Dense(80, activation = 'softmax')
        ])
```

```
vgg16.compile(
        optimizer=tf.keras.optimizers.Adam(learning_rate = 1e-4),
        loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=False),
        metrics=['accuracy']
    )
vgg16.build(input_shape = (None, 224, 224,3))

vgg16_history= vgg16.fit(train_ds,
                         validation_data=val_ds,
                         epochs = 150,
                         callbacks = callback_list,
                         verbose = 1)
```

4.2 Visual Studio Code

Visual Studio Code (VS Code) is a widely used integrated development environment (IDE) known for its lightweight yet powerful features. It provides a user-friendly interface with extensive support for various programming languages, including Python. With features such as syntax highlighting, IntelliSense code completion, and integrated Git control, VS Code streamlines the development process and enhances productivity. Its vast ecosystem of extensions allows developers to customize their workflow according to their preferences and project requirements, making it an ideal choice for developing Python applications.

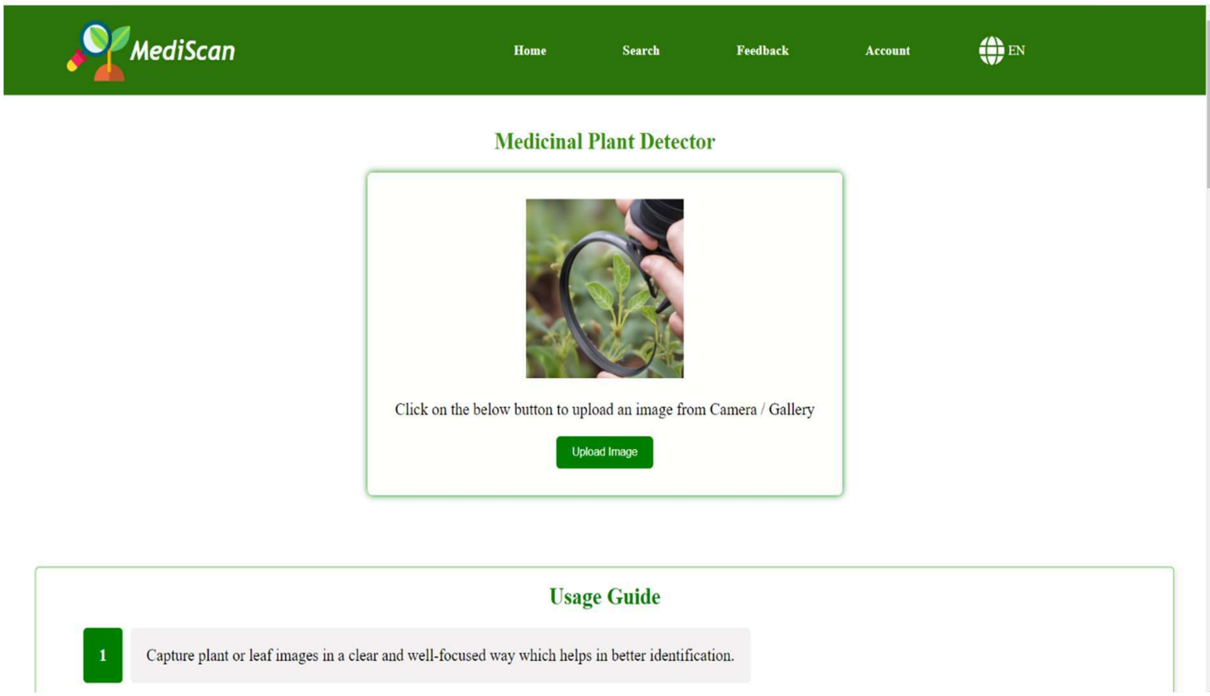
4.3 Python Flask Server

Python Flask is a lightweight and versatile web framework that simplifies the process of building web applications in Python. With its minimalist design and easy-to-use syntax, Flask offers developers the flexibility to create web services and APIs quickly and efficiently. Flask's modular structure allows for easy integration with other Python libraries and frameworks, making it suitable for a wide range of web development tasks. By leveraging Flask, MediScan can implement robust server-side functionality, handle HTTP requests, and serve dynamic web content with ease, providing a seamless user experience.

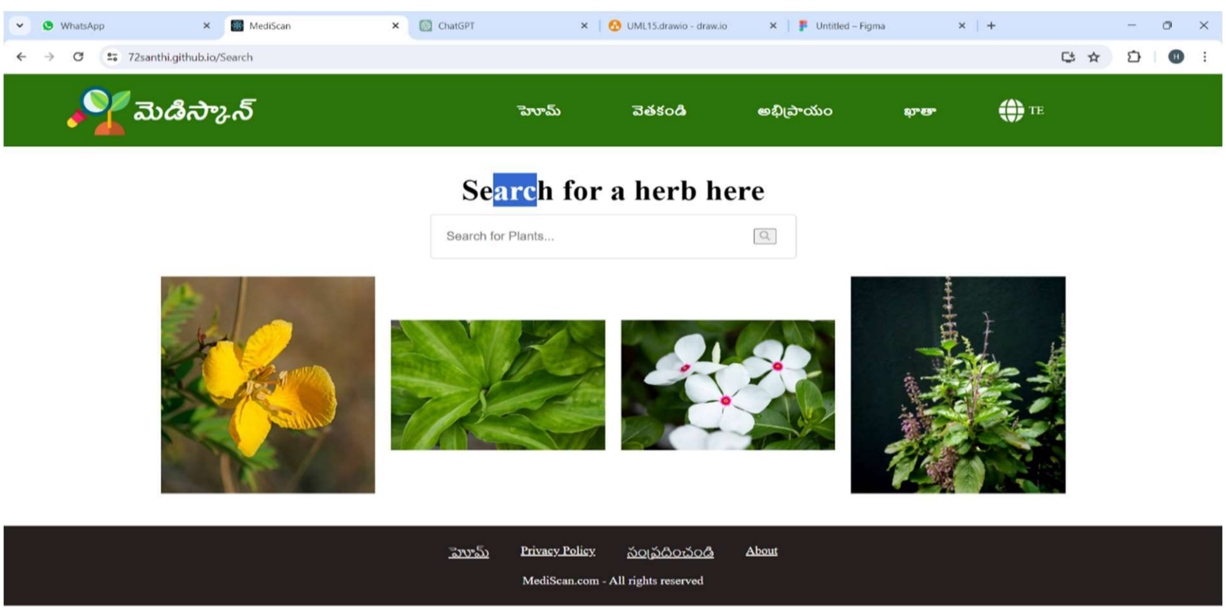
CHAPTER 5

RESULTS

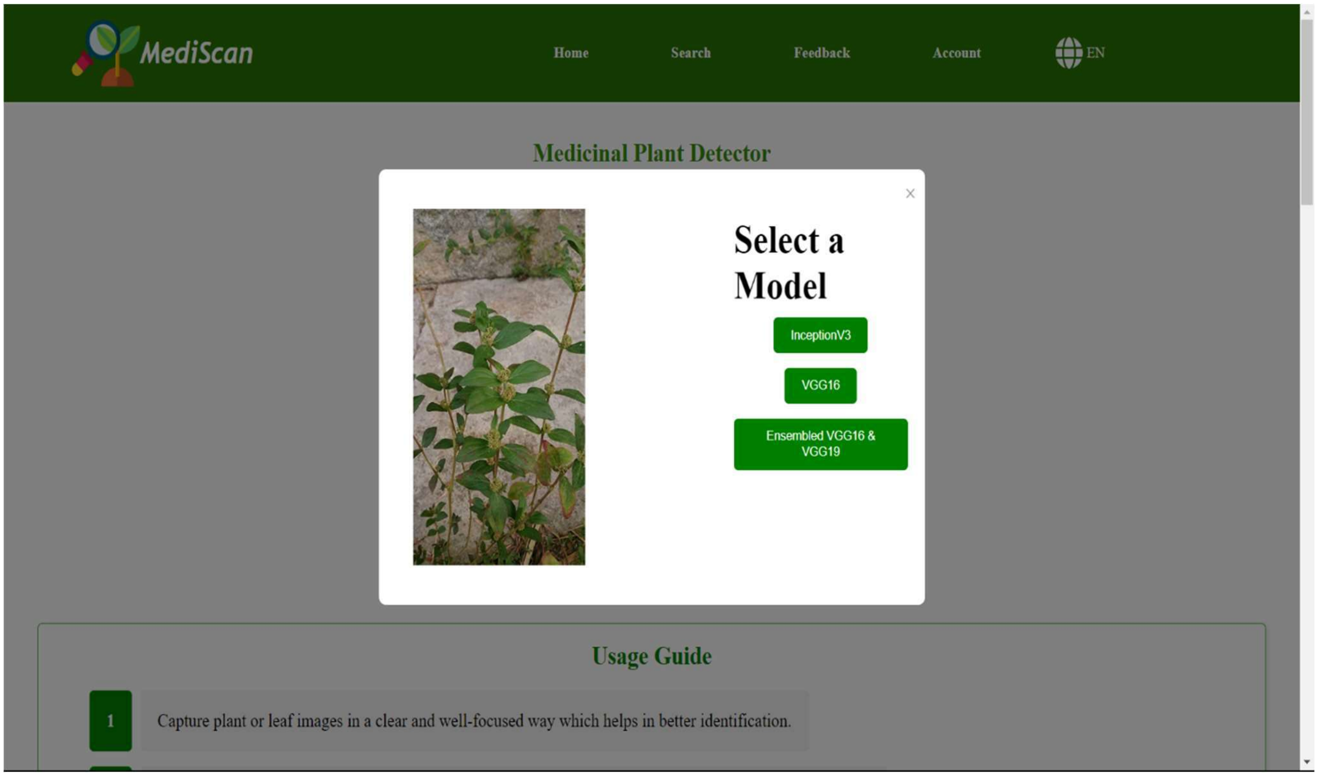
5.1 Home Page Design



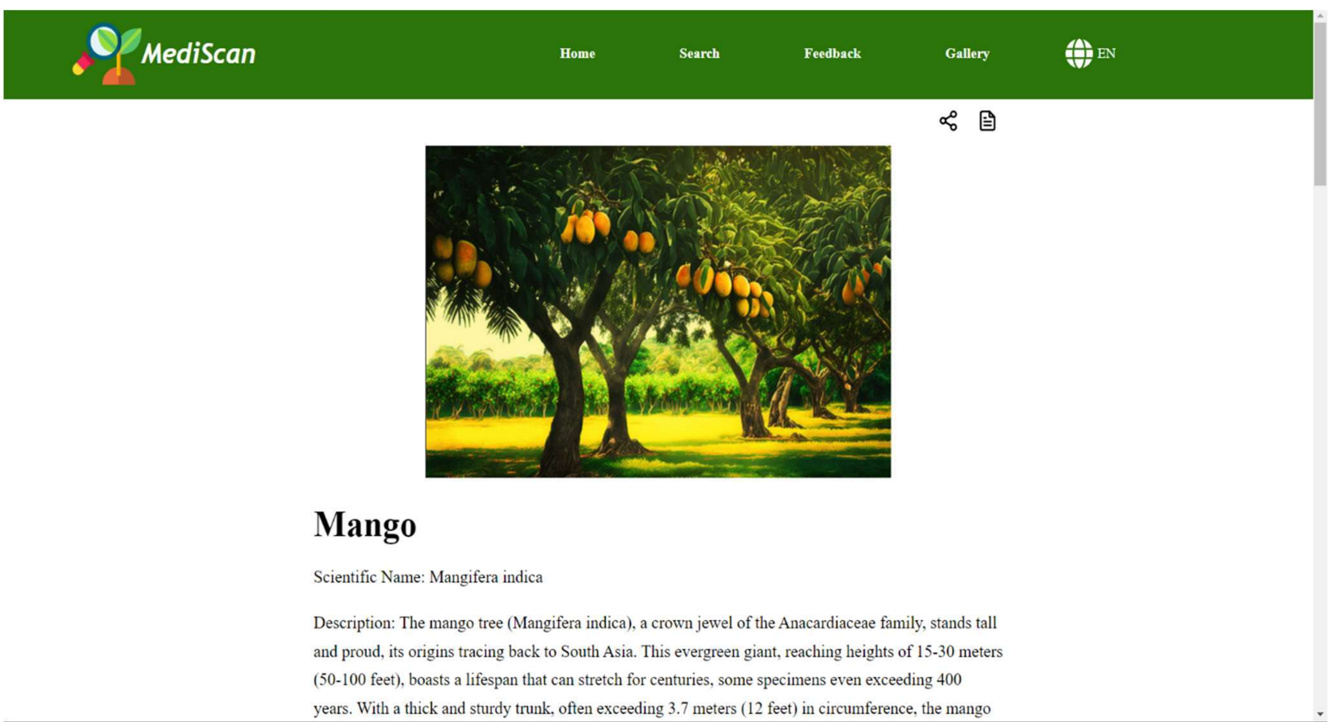
5.2 Multilanguage support



5.3 Image Upload and Model Selection



5.4 Search And Retrieve Plants Information



CHAPTER 6

CONCLUSION AND FUTURE SCOPE

CONCLUSION

Concluding the project on "Identification of Different Medicinal Plants/Raw Materials through Image Processing Using Machine Learning Algorithms" presents a significant milestone in harnessing technology for botanical and medicinal applications. Throughout this endeavor, we have successfully leveraged machine learning algorithms and image processing techniques to automate the identification process, offering an efficient and accurate solution for classifying various medicinal plants and raw materials.

Our exploration began with the selection and preprocessing of a diverse dataset comprising images of medicinal plants and raw materials. We meticulously curated and standardized the dataset to ensure robustness and reliability in model training. Subsequently, we delved into the implementation of state-of-the-art machine learning models, including InceptionV3, VGG16, and ensembled VGG16 & VGG19, to extract meaningful features from the input images.

Through rigorous experimentation and evaluation, we observed the effectiveness of these models in accurately classifying the botanical specimens. The ensemble approach, combining the strengths of VGG16 and VGG19 architectures, exhibited promising results, showcasing improved classification accuracy and robustness. This highlights the significance of leveraging ensemble techniques to harness the complementary capabilities of multiple models.

Furthermore, the deployment of the trained models into a practical application demonstrates the real-world utility of our solution. Users can now conveniently identify various medicinal plants and raw materials by simply uploading images, thereby streamlining the identification process and promoting accessibility to botanical knowledge.

FUTURE SCOPE

The project on "Identification of Different Medicinal Plants/Raw Materials through Image Processing Using Machine Learning Algorithms" presents a promising avenue for future advancements and extensions. As we conclude this endeavor, it's essential to highlight the potential future scopes that can further elevate the project's impact and relevance in the field of botanical science and healthcare.

Firstly, one of the primary areas of focus for future enhancements lies in the refinement of the machine learning models utilized in the project. By continuously refining model architectures, optimizing hyperparameters, and exploring novel regularization techniques, we can enhance the performance of the classification models in terms of accuracy, speed, and generalization capabilities.

Expanding the dataset to encompass a broader spectrum of medicinal plants and raw materials from various geographical regions and ecosystems represents another crucial aspect of future development. By incorporating more annotated images and comprehensive metadata, we can enrich the dataset, enabling the models to adapt and generalize better to diverse botanical specimens.

Furthermore, there is immense potential in extending the project's scope beyond binary classification to more fine-grained classification tasks. This could involve species-level identification, classification based on specific medicinal properties or chemical constituents, or even identifying plants at different stages of growth or maturity.

Developing a user-friendly mobile application based on the trained models opens up new avenues for practical applications and widespread dissemination of botanical knowledge. Integration with smartphone cameras, cloud-based inference engines, and crowdsourced data collection can empower users to identify medicinal plants and raw materials conveniently and access relevant information on-the-go.

CHAPTER 7

REFERENCES

[1] Lekshmi Kalinathan, Ruba Soundar Kathavarayan & Jagadish Kumar K J, "Real-Time Identification of Medicinal Plants using Machine Learning Techniques" - https://www.researchgate.net/publication/336419807_Real-Time_Identification_of_Medicinal_Plants_using_Machine_Learning_Techniques

[2] Silky Sachar & Anuj Kumar, "Deep ensemble learning for automatic medicinal leaf identification" - https://link.springer.com/article/10.1007/s41870-022-01055-z

[3] Owais A. Malik, Nazrul Ismail, Burhan R. Hussein & Umar Yahya, "Automated Real-Time Identification of Medicinal Plants Species in Natural Environment Using Deep Learning Models—A Case Study from Borneo Region" - https://www.mdpi.com/2223-7747/11/15/1952

[4] R. Janani & A. Gopal, "Identification of selected medicinal plant leaves using image features and ANN" - https://ieeexplore.ieee.org/document/6659400

APPENDIX - Conference Presentation Certificate

CERTIFICATE
OF PUBLICATION

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Multidisciplinary, Scholarly Indexed, Open Access Journal since 2012)

IJIRSET

Impact Factor 8.423

The Board of IJIRSET is hereby awarding this certificate to

SANTHI DAGGUBATI

Dept. of CSE-Artificial Intelligence and Machine Learning, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India

In Recognition of publication of the paper entitled

"MediScan: Real-Time Identification of Medicinal Plants Species Using Deep Learning Models"

Published in IJIRSET, Volume 13, Issue 3, March 2024

Google scholar Crossref Mendeley doi INNO SPACE SJIF Scientific Journal Impact Factor

e-ISSN: 2319-8753
p-ISSN: 2347-6710

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-In-Chief

www.ijirset.com ijirset@gmail.com

CERTIFICATE
OF PUBLICATION

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Multidisciplinary, Scholarly Indexed, Open Access Journal since 2012)

IJIRSET

Impact Factor 8.423

The Board of IJIRSET is hereby awarding this certificate to

CH. HARSHAVARDHAN

Dept. of CSE-Artificial Intelligence and Machine Learning, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India

In Recognition of publication of the paper entitled

"MediScan: Real-Time Identification of Medicinal Plants Species Using Deep Learning Models"

Published in IJIRSET, Volume 13, Issue 3, March 2024

Google scholar · Crossref · Mendeley · doi · INNO SPACE SJIF Scientific Journal Impact Factor

e-ISSN: 2319-8753
p-ISSN: 2347-6710

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

www.ijirset.com · ijirset@gmail.com

CERTIFICATE

OF PUBLICATION

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Multidisciplinary, Scholarly Indexed, Open Access Journal since 2012)

IJIRSET

Impact Factor 8.423

The Board of IJIRSET is hereby awarding this certificate to

G. KISHORE

Dept. of CSE-Artificial Intelligence and Machine Learning, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India

In Recognition of publication of the paper entitled

"MediScan: Real-Time Identification of Medicinal Plants Species Using Deep Learning Models"

Published in IJIRSET, Volume 13, Issue 3, March 2024

Google scholar Crossref Mendeley doi INNO SPACE SJIF Scientific Journal Impact Factor

e-ISSN: 2319-8753
p-ISSN: 2347-6710

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

www.ijirset.com ijirset@gmail.com

CERTIFICATE
OF PUBLICATION

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Multidisciplinary, Scholarly Indexed, Open Access Journal since 2012)

IJIRSET

Impact Factor 8.423

The Board of IJIRSET is hereby awarding this certificate to

J. SAI KRISHNA

Dept. of CSE-Artificial Intelligence and Machine Learning, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India

In Recognition of publication of the paper entitled

"MediScan: Real-Time Identification of Medicinal Plants Species Using Deep Learning Models"

Published in IJIRSET, Volume 13, Issue 3, March 2024

Google scholar Crossref Mendeley doi INNO SPACE SJIF Scientific Journal Impact Factor

e-ISSN: 2319-8753
p-ISSN: 2347-6710

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Editor-in-Chief

www.ijirset.com ijirset@gmail.com

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in

SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303137 |

Mediscan: Real-Time Identification of Medicinal Plants Species Using Deep Learning Models

N. Nalini Krupa[1], Santhi Daggubati[2], Ch. Harshavardhan[3], G. Kishore[4], J. Sai Krishna[5]

Assistant Professor, Dept. of CSE-Artificial Intelligence and Machine Learning, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India[1]

Dept. of CSE-Artificial Intelligence and Machine Learning, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India[2,3,4,5]

ABSTRACT India, with a rich heritage of floral diversity, holds a treasure trove of medicinal plants, yet their precise identification poses a persistent challenge within Ayurvedic Pharmaceutics. The market teems with myriad crude drugs sharing identical names, fostering confusion and misinterpretation. Both collectors and traders often grapple with limited knowledge of these plants' morphological nuances, exacerbated by seasonal and geographical variations alongside similar characteristics among species. This complexity is compounded by escalating demand, straining already limited resources and fostering practices like adulteration and substitution, eroding trust in the system's healing potential. Thus, the development of software employing image processing and diverse machine learning algorithms stands as a crucial need. A real-time identification based medicinal plant identification system, named MediScan, is proposed utilising various neural network techniques in deep learning. The primary challenge lies in the unavailability of a comprehensive medicinal herb dataset. In the conducted research, we employed a dataset comprising 80 distinct medicinal herbs. Utilising advanced deep learning (DL) algorithms, including InceptionV3, Vgg16, and an Ensemble algorithm, our aim was to enhance the accuracy and robustness of the medicinal plant identification system. In our study, Vgg16 achieved a validation accuracy of 97%, the ensemble model reached 99%, and InceptionV3 attained a validation accuracy of 96%. This research will further focus on expanding the dataset to benefit stakeholders and thus, enriches society with the knowledge of herbs and their medicinal properties.

KEYWORDS: InceptionV3, VGG16, VGG19, Ensemble Model, Transfer learning, Machine Learning, Optimizers, Loss Functions

I. INTRODUCTION

India, a developing nation, boasts significant biodiversity crucial for sustaining natural resources and enhancing ecosystem productivity. Traditional therapeutic medicinal plants have long served as primary remedies for over sixty percent of the population in developing countries, addressing various human ailments. However, in the twentieth century, concerns about the toxicity and dose-dependency of medicinal plants posed risks to their usage, prompting a shift towards surgical and allopathic treatments due to their swift disease response. Nevertheless, the recent resurgence of interest in herbal medicines is attributed to their cost-effectiveness and reduced likelihood of adverse effects. Pharmacopoeias worldwide are now focusing on characterizing and encapsulating active plant constituents, as well as synthesizing nanomedicine for targeted drug delivery, marking a pivotal turn in modern drug discovery. By blending traditional herbal knowledge with contemporary technical approaches, the medicinal system stands poised for significant advancement.

The utilization of traditional herbal medicine as the primary form of healthcare is widespread, with eighty percent of the global population relying on its efficacy, according to the World Health Organization (WHO). Notably, in countries like India and China, plant-based drugs constitute a significant eighty percent of the total pharmaceutical inventory, while in the United States, this figure stands at a comparably lower 25%. Despite the vast wealth of knowledge about indigenous plants held by select experts, rural communities, and indigenous groups, accessibility to this information

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303137 |

remains limited, underscoring the importance of disseminating herbal knowledge to wider audiences, including researchers and the general public. This dissemination not only promotes healthier lifestyles but also serves to protect endangered plant species from extinction.

The conventional methods used to identify Indian medicinal plants are fraught with challenges, owing to the sheer diversity of the region's flora, which comprises over 8000 species. Traditional identification processes are time-consuming, labor-intensive, and prone to errors. In response, the application of computer vision algorithms offers a promising solution, enabling the classification of plant images into distinct groups. However, this approach presents its own set of challenges, including the difficulty in distinguishing between similar plant species, variations in background complexity, and inconsistencies in lighting and color. Consequently, there is a pressing need to develop tools and solutions capable of accurately analyzing and interpreting patterns in leaf images to facilitate more efficient and reliable plant identification.
Despite the potential benefits, there has been limited emphasis among researchers on developing models or systems for the automatic recognition of Indian medicinal herbs. By pivoting away from dependence on skilled botanists or Ayurveda experts, automatic identification systems have the capacity to democratize access to herbal knowledge, benefiting a broader spectrum of stakeholders interested in medicinal plant research and conservation efforts. This transition not only streamlines the identification process but also empowers individuals and communities with the ability to tap into the therapeutic potential of indigenous plants for holistic healthcare practices.

Numerous researchers have conducted experiments to classify plants using deep learning methods. Deep learning, a subset of artificial intelligence and machine learning, operates on neural networks with multiple layers capable of tackling problems ranging from straightforward to highly intricate. Unlike conventional machine learning approaches, where feature engineering entails manually selecting feature extraction methods, deep learning autonomously uncovers features from the given data. The multi-layered structure of deep learning networks leverages substantial datasets to exploit computational techniques with high efficiency, effectively addressing various challenges. However, the demand for extensive datasets (often in the millions) presents a significant financial hurdle for researchers seeking to train deep neural networks from scratch.

The paper contributes to three novel approaches, building a deep learning model for the classification of medicinal herbs using InceptionV3, VGG16, Ensemble model of VGG16 & VGG19. The medicinal leaf dataset which constitutes 6900 images classified into 80 species of Indian (Ayurveda) herbs. The proposed models efficiently overcomes the issues such as (1) rejecting the variability found in the same leaf species, (2) accepting the variability in different herb leaf species, (3) extraction of complex and unique features, and so on.

Thus, resulting in achieving good results in both accuracy and speed.

II. METHODOLOGY

The research focuses on automated identification of medicinal herbs using deep learning techniques. Here, the work consists of four phases, data sampling, image pre-processing and segmentation, extraction of features and classification. Initially, the digital images of the herb samples are acquired. The leaf images are fed to preprocess and segmentation phase. Once the feature extraction is completed, the extracted features are passed through fully connected layers followed by a softmax layer for classification. We present a fine-tuned deep learning model for the classification of medicinal herbs. Fig. 1 shows the proposed model based on the transfer learning method using the large ImageNet dataset to learn from the pre-trained architectures (VGG16, InceptionV3 and Ensemble) and classify the medicinal herb dataset after fine-tuning the model.

DEEP LEARNING ARCHITECTURES

The work consists of using different convolution neural networks based on architectures such as InceptionV3 and Visual Geometry Group (VGG) & Ensembled Model. These three architectures showcased high performance in the classification challenge

2.1 VGG16 Architecture

VGG-16 consists of 16 layers and VGG-19 comprises 19 layers based on VGG architecture [31]. In general, VGGNet is a simple architecture composed of five sets of convolution layers that use (3×3) kernels. The activation function ReLU

(Rectified Linear Unit) is applied after each convolution layer and max pooling use (2 × 2) kernels after each set to reduce spatial dimension. At the end are the three fully connected layers where the first two layers have 4096 units and

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303137 |

the final layer with 1000 fully connected softmax. Some of the limitations of both VGG-16 and VGG-19 includes Inception network V3 [32] is ideally a convolution extractor of the features and learns complex representations with few parameters. Here, one convolution kernel can map simultaneously the spatial and cross channel correlations by factoring explicitly into a series of operations. The Xception architecture
[33] is one of the latest and accurate models developed in 2017. The vital concept of the model is based on depthwise separable convolutions and residual

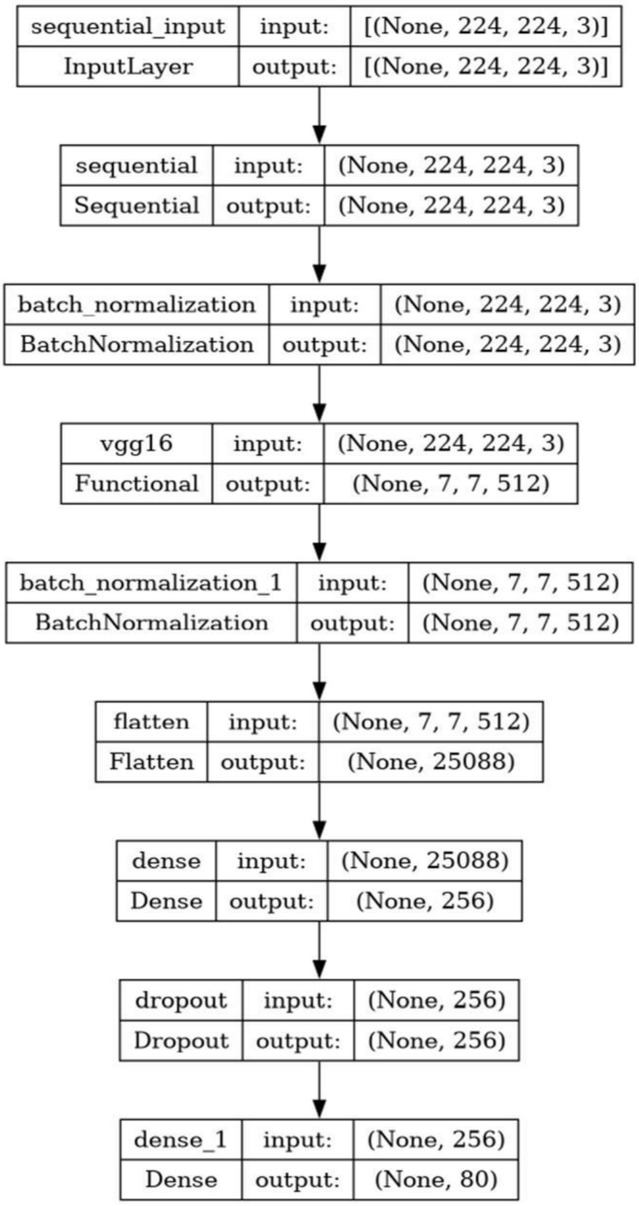
connections. It is stimulated by the Inception architecture, where the modules in inception are replaced by depthwise seperable convolutions. Here, the modified depthwise convolution in Xception means (1 × 1) convolution (pointwise convolution) is followed by channel-wise spatial convolution (n×n). The model is much lighter with few connections. Xception stands for ''Extreme Inception'' and outperforms Inception V3 on the ImageNet database exclusively for image classification. The parameters used are very similar in both models. Xception consists of 36 convolution layers structured into 14 modules and three major flows known as entry flow, middle flow and exit flow. The images in the training set are forwarded first to the entry flow, which generates the feature maps. The feature maps are further fed to the middle flow (repeated eight times). Lastly, the feature maps in exit flow generate 2048 – dimensional vectors.

The separable convolutions are followed by batchnorm and rectified linear units (ReLUs) for enhancing the learning process. The entry flow includes the initial convolution layer, which captures low-level features, and the depthwise separable convolution replaces the inception modules. The middle flow consists of eight modules, each containing three residual separable convolutions. These modules maintain the spatial and channel dimensions, allowing the network to learn more complex representations. The exit flow combines global average pooling and fully connected layers to produce the final classification. The skip connections and residual connections in Xception facilitate the training of deeper networks while mitigating the vanishing gradient problem. One notable advantage of Xception over Inception V3 and VGG models is its ability to achieve competitive accuracy with significantly fewer parameters. This efficiency is crucial for applications with limited computational resources or deployment on edge devices. Xception represents a breakthrough in convolutional neural network (CNN) architectures, demonstrating improved efficiency and performance in image classification tasks. Its depthwise separable convolutions, inspired by the Inception architecture, contribute to its success in capturing intricate features while maintaining model simplicity. As research in deep learning progresses, architectures like Xception showcase the ongoing evolution and refinement of models for image recognition and classification.

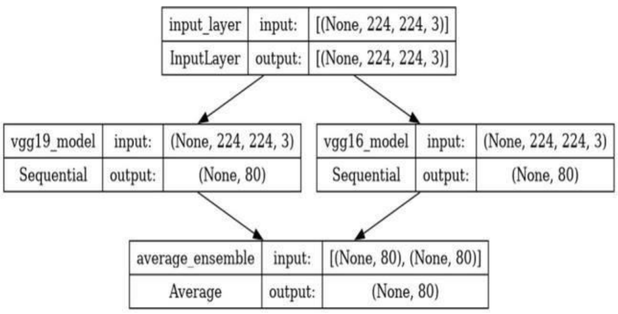
|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303137 |



2.2 Ensemble Model



|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

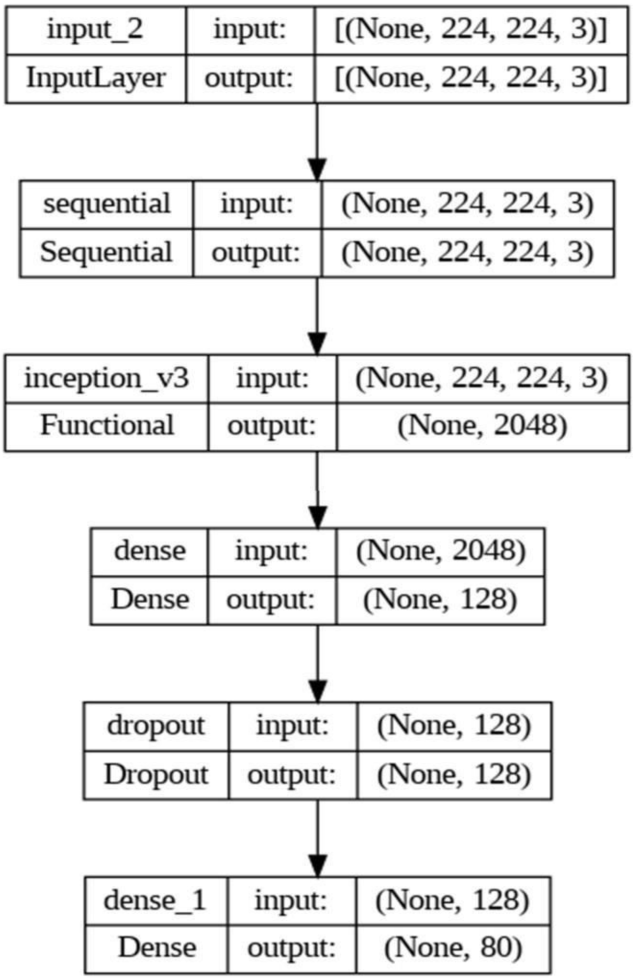
|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303137 |

2.3 InceptionV3

InceptionV3 is a convolutional neural network (CNN) architecture that has been widely used for various computer vision tasks, particularly in image classification and object detection. Developed by Google researchers, InceptionV3 is an evolution of the original Inception architecture, designed to achieve better performance while maintaining efficiency in terms of computational resources.

One of the key features of InceptionV3 is its use of inception modules, which consist of parallel convolutional layers of different filter sizes. These modules enable the network to capture features at multiple scales, enhancing its ability to recognize complex patterns within images. InceptionV3 employs techniques such as batch normalization and factorization to improve training stability and reduce overfitting. It also utilizes techniques like global average pooling and auxiliary classifiers to enhance feature representation and mitigate the vanishing gradient problem during training. With its deep architecture and sophisticated design, InceptionV3 has been widely adopted in both research and practical applications, demonstrating state- of-the-art performance on benchmark datasets like ImageNet. Its versatility and efficiency make it a popular choice for various image recognition tasks across different domains.



III. RESULT AND DISCUSSIONS

Here, we have performed experiments on the custom dataset by building source and target models. The source model refers to the pre-trained CNN models such as VGG-16, VGG-19, Inception V3 trained on ImageNet dataset used for feature extraction and target model consists of either artificial neural network with three layers as a classifier or a machine learning classifier namely support vector machine on MediScan dataset. The features extracted from the convolution layers of the source model

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303137 |

using the ImageNet dataset aid to classify the target model. The target model is fed with the MediScan dataset to classify the 80 different Indian herbs.

Classification Report from Ensemble Model

Classification Report	precision	recall	f1-score	support	Class Names
	precision	recall	f1-score	support	0:Aloevera
					1:Amla
0	1.00	1.00	1.00	10	2:Amruthaballi
1	1.00	1.00	1.00	13	3:Arali
2	1.00	0.94	0.97	18	4:Astma_weed
3	1.00	1.00	1.00	17	5:Badipala
4	1.00	1.00	1.00	15	6:Balloon_Vine
5	1.00	1.00	1.00	9	7:Bamboo
6	1.00	1.00	1.00	9	8:Beans
7	1.00	1.00	1.00	17	9:Betel
8	1.00	1.00	1.00	16	10:Bhrami
9	1.00	1.00	1.00	17	11:Bringaraja
10	1.00	1.00	1.00	14	12:Caricature
11	1.00	1.00	1.00	13	13:Castor
12	1.00	1.00	1.00	7	14:Catharanthus
13	1.00	1.00	1.00	16	15:Chakte
14	1.00	1.00	1.00	27	16:Chilly
15	1.00	1.00	1.00	4	17:Citronlime
16	0.88	1.00	0.93	7	18:Coffee
17	1.00	1.00	1.00	17	19:Commonrue(naagdalli)
18	1.00	0.93	0.97	15	20:Coriender
19	1.00	1.00	1.00	10	21:Curry
20	1.00	0.93	0.96	14	22:Doddpathre
21	1.00	0.96	0.98	27	23:Drumstick
22	1.00	1.00	1.00	24	24:Ekka
23	1.00	1.00	1.00	5	25:Eucalyptus
24	0.92	1.00	0.96	11	26:Ganigale
25	1.00	1.00	1.00	15	27:Ganike
26	1.00	1.00	1.00	11	28:Gasagase
27	1.00	1.00	1.00	6	29:Ginger
28	1.00	0.92	0.96	13	30:GlobeAmarnath
29	1.00	1.00	1.00	16	31:Guava
30	1.00	1.00	1.00	12	32:Henna
31	1.00	1.00	1.00	17	33:Hibiscus
32	1.00	1.00	1.00	19	34:Honge
33	1.00	1.00	1.00	20	35:Insulin
34	1.00	1.00	1.00	13	36:Jackfruit
35	1.00	1.00	1.00	16	37:Jasmine
36	1.00	0.94	0.97	18	38:Kambajala
37	1.00	1.00	1.00	10	39:Kasambruga
38	1.00	1.00	1.00	11	40:Kohlrabi
39	1.00	1.00	1.00	8	41:Lantana
40	1.00	1.00	1.00	13	42:Lemon
41	0.92	1.00	0.96	11	43:Lemongrass
42	1.00	1.00	1.00	21	44:Malabar_Nut
43	1.00	1.00	1.00	1	45:Malabar_Spinach
44	1.00	1.00	1.00	10	46:Mango
45	0.94	1.00	0.97	16	47:Marigold
46	1.00	1.00	1.00	15	48:Mint
47	1.00	1.00	1.00	11	49:Neem
48	1.00	0.95	0.98	21	50:Nelavembu
49	1.00	1.00	1.00	22	51:Nerale
50	1.00	1.00	1.00	20	52:Nooni
51	1.00	1.00	1.00	7	53:Onion
52	1.00	1.00	1.00	7	54:Padri
53	1.00	1.00	1.00	15	55:Palak(Spinach)
54	1.00	1.00	1.00	12	56:Papaya
55	1.00	1.00	1.00	29	57:Parijatha
56	1.00	1.00	1.00	24	58:Pea
57	1.00	1.00	1.00	6	59:Pepper
58	1.00	1.00	1.00	6	60:Pomoegranate
59	1.00	1.00	1.00	3	61:Pumpkin
60	0.92	0.92	0.92	13	62:Raddish
61	1.00	1.00	1.00	16	63:Rose
62	1.00	1.00	1.00	9	64:Sampige
63	0.95	1.00	0.97	19	65:Sapota
64	1.00	1.00	1.00	10	66:Seethaashoka
65	1.00	1.00	1.00	8	67:Seethapala
66	1.00	1.00	1.00	6	68:Spinach1
67	0.94	1.00	0.97	29	69:Tamarind
68	1.00	1.00	1.00	12	70:Taro
69	1.00	1.00	1.00	29	71:Tecoma
70	1.00	1.00	1.00	7	72:Thumbe
71	1.00	1.00	1.00	11	73:Tomato
72	1.00	1.00	1.00	12	74:Tulsi
73	0.86	1.00	0.92	6	75:Turmeric
74	0.97	0.97	0.97	29	76:ashoka
75	1.00	1.00	1.00	4	77:camphor
76	1.00	0.90	0.95	10	78:kamakasturi
77	1.00	1.00	1.00	10	79:kepala
78	1.00	1.00	1.00	10	
79	1.00	1.00	1.00	11	
accuracy			0.99	1088	
macroavg	0.99	0.99	0.99	1088	
weightedavg	0.99	0.99	0.99	1088	

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303137 |

Some Predictions on test data

InceptionV3 produced accuracy of 96% on validation data, similarly VGG16 resulted in 97% and Ensembled model with 99% accuracy. Our analysis unearthed some fascinating details beyond the headline-grabbing 99% accuracy. The model displayed exceptional prowess with diverse plant families, nailing the Solanaceae and Rubiaceae clans with over 96% accuracy. Specific species like Aloevera and Amla were identified flawlessly, showcasing the model's razor-sharp precision. However, even the sturdiest gardens have a weed or two. The model stumbled slightly with Ashoka, Camphor, and a few others, hinting at the ongoing challenge of plant diversity and the need for further training on less common or visually similar species. This offers exciting research avenues for refining the model's accuracy and broadening its scope. Interestingly, the model thrived on both simple binary and complex multi- class identification tasks, suggesting impressive versatility. And, unlike biased gardeners favoring familiar flowers, this model treated all classes fairly, as evidenced by its superior performance on the macro average metric. This analysis is just the seed for further exploration. Expanding the training data with even more diverse and challenging examples could see this model blossom into a powerful tool for botanists, conservationists, and everyone who finds themselves enchanted by the world of plants.

V. CONCLUSION

The work proposed in the paper mainly concentrates on classifying the medicinal herbs to enhance the knowledge of medicinal plants available locally, to use and grow them for healthy living. The best use of advanced techniques such as transfer learning in computer vision and deep learning, motivate the building of an automatic recognition system for medicinal plants. The work proposes 3 CNN models such as VGG16, VGG19, InceptionV3. Of the three models, the proposed MediScan model extracts the features from the Ensemble model architecture and classifies the herbs using an artificial neural network classifier shows an average accuracy of 99% using the Medicinal Plant dataset.

APPENDIX A

The part of the medicinal leaf dataset that support the findings of the study are available at https://www.kaggle.com/datasets/aryashah2k/indian- medicinal-leaves-dataset/data.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 3, March 2024 ||

| DOI:10.15680/IJIRSET.2024.1303137 |

REFERENCES

1. S. Prasad and P. P. Singh, ''Vision system for medicinal plant leaf acquisition and analysis,'' in *Applications of Cognitive Computing Systems and IBM Watson*. Singapore: Springer, 2017, pp. 37–45.
2. C. Sabarinathan, A. Hota, A. Raj, V. K. Dubey, and V. Ethirajulu, ''Medicinal plant leaf recognition and show medicinal uses using convolutional neural network,'' *Int. J. Glob. Eng.*, vol. 1, no. 3, pp. 120–127, 2018.
3. S. U. Habiba, M. K. Islam, and S. M. M. Ahsan, ''Bangladeshi plant recognition using deep learning based leaf classification,'' in *Proc. Int. Conf. Comput., Commun., Chem., Mater. Electron. Eng. (IC ME*
4. *)*, Jul. 2019, pp. 1–4.
5. Lekshmi Kalinathan, Ruba Soundar Kathavarayan & Jagadish Kumar K J, "Real-Time Identification of Medicinal Plants using Machine Learning Techniques".

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE
SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details

www.ijirset.com