iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71619

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Plant Identification and Query System Using Deep Learning and RAG-Based Search

Abhishek Ganjigatti[1], Adarsh Gayadolli[2], Chaitali Joshi[3], Manoj Shekhar[4]
[1, 2, 3, 4]*Student, Artificial Intelligence and Data Science, Angadi Institute of Technology and Management, Belagavi, Karnataka, India*

Abstract: ***In this project, we present a dual-module system that utilizes a Convolutional Neural Network (CNN) for leaf-based plant identification and integrates a Retrieval-Augmented Generation (RAG) system for answering plant-related queries. The CNN was trained and fine-tuned using the LeafSnap dataset comprising 184 plant classes, achieving a final accuracy of 88%. Additionally, the RAG module leverages LangChain tools incorporating Wikipedia, Arxiv, and a custom PDF retriever for context-aware plant information search. This end-to-end solution is implemented using Flask and Streamlit, offering users a seamless interface to either upload plant leaf images or ask questions.***
Keywords: CNN, LangChain, RAG, Plant Identification, Flask, Streamlit, LeafSnap Dataset

I. INTRODUCTION

Accurate identification of plant species is crucial for botanists, researchers, and agricultural communities. This project aims to automate plant identification using a CNN trained on leaf images, while also offering a smart assistant to answer botanical queries using retrieval-augmented generation. The system combines deep learning for classification and transformer-based models for natural language question-answering.

II. LITERATURE REVIEW

The methodology in this project is inspired by existing work, including the IJCRT reference paper 'Plant Identification System Using Machine Learning' (IJCRT2304306), and LeafSnap, a computer vision system for automatic plant species recognition. These systems demonstrate that CNNs are effective in extracting relevant features from leaf images, while modern LLMs can enrich plant databases through intelligent querying.

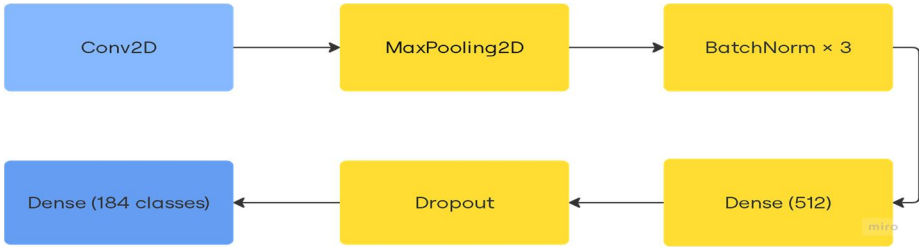
III. SYSTEM ARCHITECTURE

The proposed system is comprised of two primary components: (1) a Flask-based plant identification engine using a fine-tuned CNN model, and (2) a Streamlit-based document query tool powered by LangChain. The model predicts species from uploaded leaf images, and plant metadata is retrieved from a local dataset. The RAG system enhances functionality with external tools like Wikipedia and Arxiv to answer natural language questions.

IV. CNN MODEL AND TRAINING

The CNN model architecture consists of multiple convolutional and pooling layers followed by dense layers and batch normalization. It was trained for 20 epochs on the LeafSnap dataset and achieved the following metrics:
- Final Accuracy: 88%
- Final Validation Loss: ~0.9284
- Parameters: 13,036,152

A. Model Layers



©IJRASET: All Rights are Reserved

V. DATASET DESCRIPTION

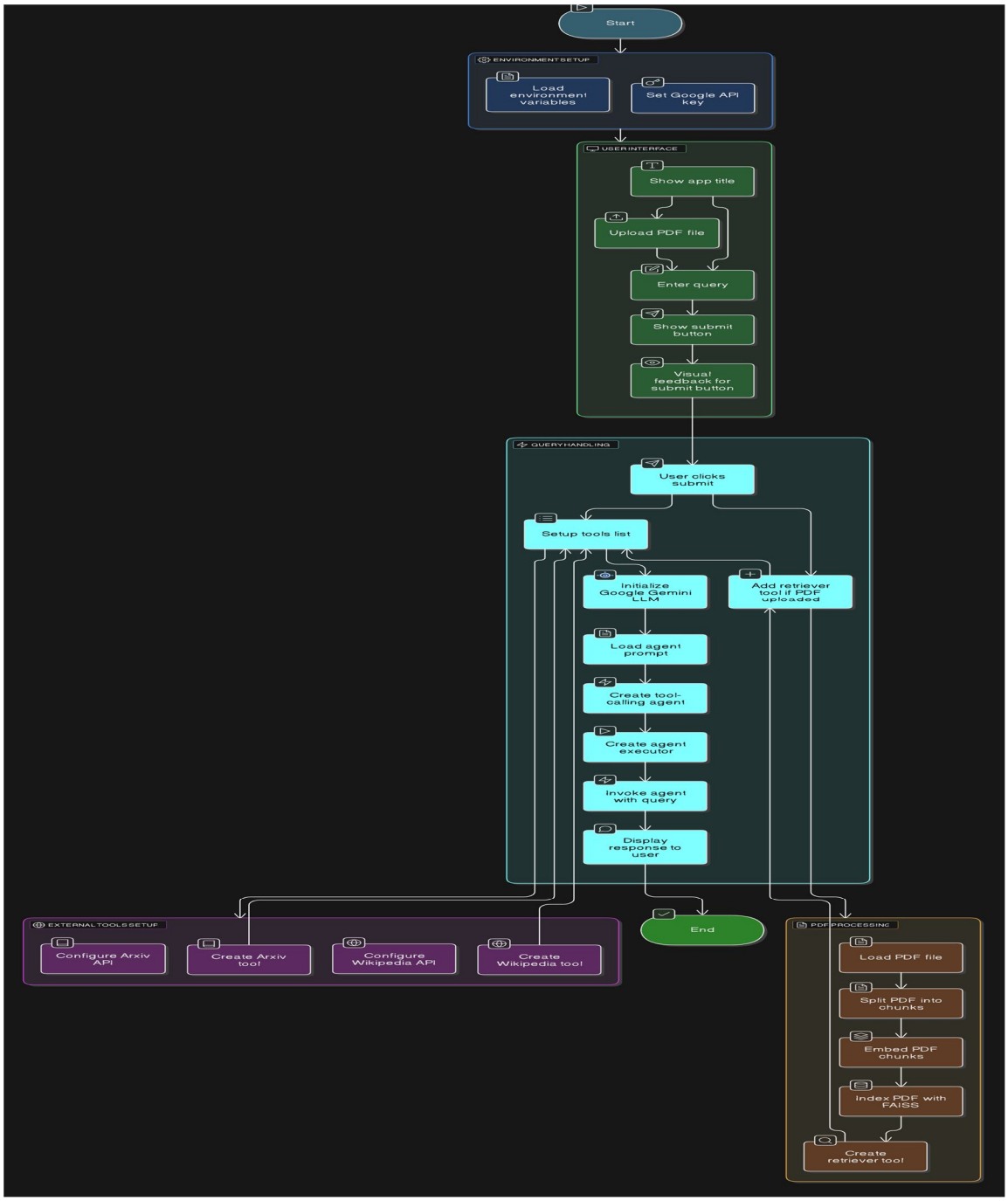
Sourced: Kaggle (https://www.kaggle.com/datasets/xhlulu/leafsnap-dataset). It includes labeled leaf images for 184 plant species. Images were normalized and resized to 224×224 pixels.

VI. FLASK WEB APPLICATION

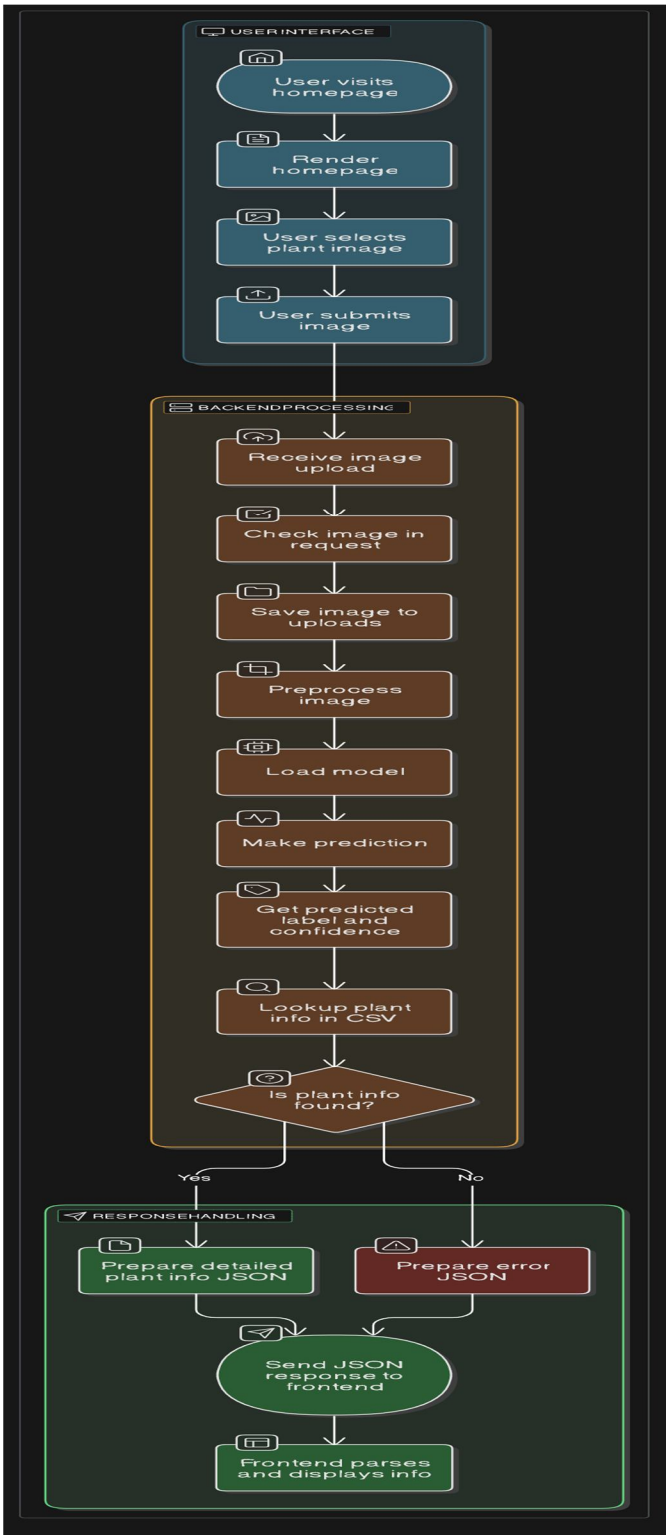
The Flask app handles image uploads, model inference, and returns plant details using a local CSV database of species metadata. It uses TensorFlow for prediction, Pillow for image processing, and pandas for data handling.

VII. WORKFLOW DIAGRAM

A. RAG-Based Search



B. Plant-Detection and Data-Mapping

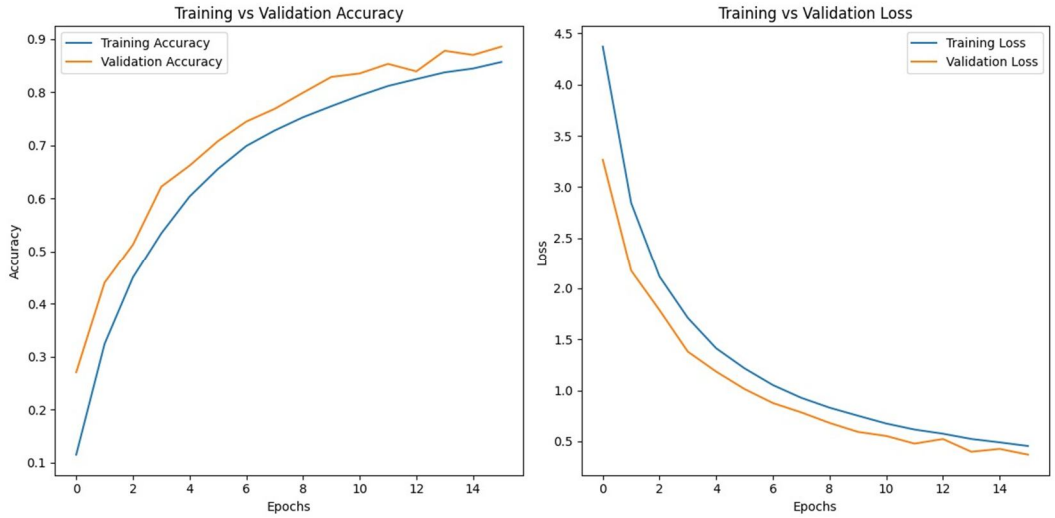


VIII. LANGCHAIN RAG SYSTEM

The RAG-based query system is implemented in Streamlit. Users can upload their own plant-related PDFs or query directly using Wikipedia and Arxiv tools via LangChain. The backend uses sentence-transformers and FAISS for vector similarity retrieval.

IX. RESULTS AND EVALUATION

The model showed excellent convergence and performance over epochs. Below is the training/validation accuracy and loss graph:



The frontend of the web application is simple and user-friendly:

X. CONCLUSION AND FUTURE WORK

This project successfully demonstrates a unified plant identification and query system using CNNs and transformer-based retrieval. Future work includes developing a mobile application and expanding the training dataset to include more plant features such as flowers or stems.

XI. ACKNOWLEDGEMENTS

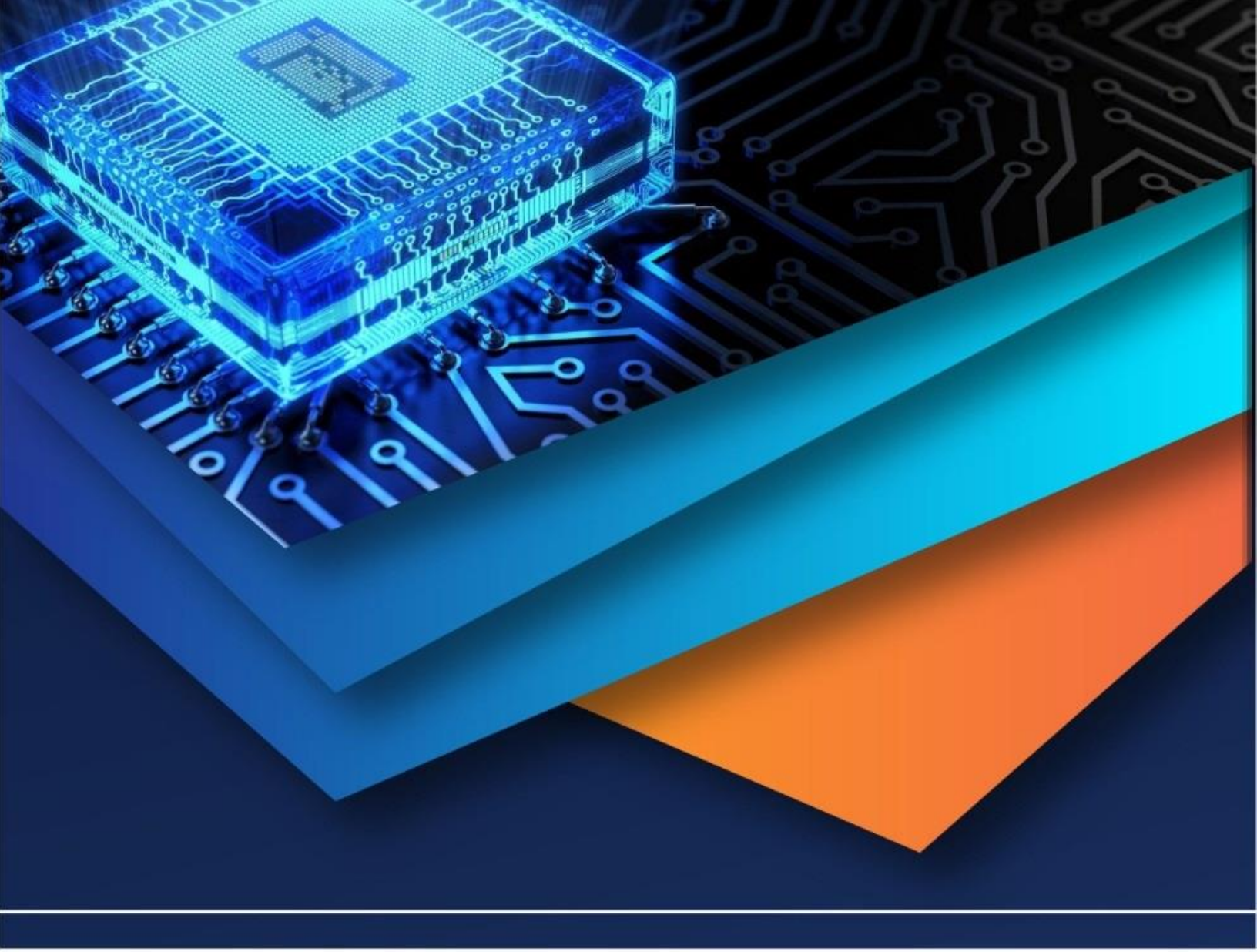
The authors sincerely thank **Angadi Institute of Technology and Management, Belagavi, Karnataka, India**, for providing the resources and support required to carry out this project. We extend our heartfelt gratitude to our guide and Head of Department, **Dr. Sagar Birje** (Professor, Dept. of AI&DS), for his invaluable mentorship, constant support, and expert guidance throughout the project. We also acknowledge the support of **Prof. Dattatreya Choudhari** (Mini-Project Coordinator, Assistant Professor, Dept. of AI&DS) and **Dr. Anand Deshpande** (Principal and Director) for their encouragement and institutional support.

REFERENCES

[1] N. Kumar et al., 'Leafsnap: A Computer Vision System for Automatic Plant Species Identification', ECCV 2012.
[2] IJCRT2304306, 'Plant Identification System Using Machine Learning', IJCRT Vol. 11, Issue 4, April 2023.
[3] https://www.kaggle.com/datasets/xhlulu/leafsnap-dataset
[4] LangChain Documentation: https://docs.langchain.com/

Link

https://github.com/abhishek-ganjigatti/Leaf_image_detection
https://github.com/abhishek-ganjigatti/PLD_RAG-TT-VT



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)