**Plant Species Classification Using Transfer Learning**

A Project Report Submitted in Partial Fulfilment of The Requirements for The Award of The Degree Of

**Bachelor of Technology**

**in**

**Computer Science and Engineering**

By

D Sunder Raj - 112015042

**Under the Supervision of: Ms. Anupama Arun**

**Semester: 7th Semester**



**Department of Computer Science and Engineering**

**Indian Institute of Information Technology, Pune**

**(An Institute of National Importance by an Act of Parliament)**

**NOVEMBER 2023**

# BONAFIDE CERTIFICATE

This is to certify that the project report entitled **“Plant Species Classification”** submitted by **D Sunder Raj** bearing the **MIS No: 112015042** in completion of his project work under the guidance of **Ms. Anupama Arun** is accepted for the project report submission in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in the **Department of Computer Science and Engineering**, Indian Institute of Information Technology, Pune (IIIT Pune),

|  |  |
| --- | --- |
| during the academic year **2023-24**. |  |
| **Ms. Anupama Arun** | **Dr. Sanjeev Sharma** |
| Project Guide | Head of the Department |
| Assistant Professor | Assistant Professor |
| Department of CSE | Department of CSE |
| IIIT Pune | IIIT Pune |

Project Viva-voce held on 07-11-2023.

# Undertaking for Plagiarism

I **D. Sunder Raj** solemnly declared that research work presented in the report titled **“Plant Species Classification”** is solely our research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete report/dissertation has been written by me. I understand the zero-tolerance policy of **Indian Institute of Information Technology Pune** towards plagiarism. Therefore, I declare that no portion of my report has been plagiarized and any material used as reference is properly referred/cited. I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of the degree, the Institute reserves the right to withdraw/revoke my B. Tech degree.

## Students’/ Student’s Name and Signature with Date

D. Sunder Raj - 112015042

# Conflict of Interest

**Project title:** Plant Species Classification

The authors whose names are listed immediately below certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers’ bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this project.

## Students’/ Student’s Name and Signature with Date

D. Sunder Raj - 112015042

# ACKNOWLEDGEMENT

This project would not have been possible without the help and cooperation of many. I would like to thank the people who helped me directly and indirectly in the completion of this project work. First and foremost, I would like to express my gratitude to our honourable Director, **Prof. O.G. Kakde**, for providing his kind support in various aspects. I would like to express my gratitude to my project guide **Ms. Anupama Arun**, **Department of CSE**, for providing excellent guidance, encouragement, inspiration, constant and timely support throughout this B.Tech Project. I would like to express my gratitude to **Dr. Sanjeev Sharma, Department of CSE**, for providing his kind support in various aspects. I would also like to thank all the faculty members in the Department of CSE and my classmates for their steadfast and strong support and engagement with this project.



# Abstract

In this project, we aim to develop an image classification system for plant species using deep learning techniques. We leveraged the Malayakew plant leaf dataset, which comprises a diverse set of plant species with 44 classes. The goal was to create a model capable of accurately identifying the plant species depicted in images, with class labels ranging from 1 to 44.

We utilized the EfficientNetB0 model, a state-of-the-art deep learning architecture, pre-trained on a large-scale image dataset, and fine-tuned it for our specific plant species classification task. The dataset was split into training and test sets, and data augmentation techniques were applied to enhance model generalization.

The model was trained using transfer learning, where the pre-trained layers of EfficientNetB0 were frozen to capture generic image features, and new classification layers were added for fine-tuning. The model was trained over several epochs, and early stopping techniques were employed to prevent overfitting.

After training, we used the model to make predictions on test data, assessing its ability to accurately classify plant species. We also implemented a method for making predictions on single images, allowing users to input their own images and receive predictions for plant species, starting from class index 1 to 44.

However, it's crucial to note that the model's performance may vary depending on the quality of the dataset and the specific plant species it was trained on. If the model does not accurately predict the class of a test image, it may indicate a need for further fine-tuning or dataset improvements.

# TABLE OF CONTENTS

**Abstract i**

**(i) List of Figures/Symbols/Nomenclature 6**

## (ii) List of Tables 3

### Introduction

1.1 Overview of work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .... 7

1.2 Motivation of work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . .. 8

1.3 Aim . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. 8

1.4 Objective. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

1.5 Literature Review . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. 9

1.6 Research Gap . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . . . . . . . . . .. 10

### 2 Problem Statement

2.1 Research Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . … 11

2.2 Methodology of work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

**3 Conclusion and Future Scope**

3.1 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

3.1 Future Scope . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ….19

# Chapter 1 Introduction

### 1.1 Overview of work

In the realm of plant biology and environmental science, the accurate identification and classification of plant species are essential for various applications, ranging from biodiversity conservation to agriculture and ecological research. Traditional methods of plant species recognition often involve time-consuming and manual processes, making it an ideal candidate for automation through advanced technologies. Deep learning, a subset of artificial intelligence, has emerged as a transformative tool for image analysis and classification tasks, offering new opportunities for efficient and precise plant species detection.

This project embarks on a journey into the intersection of deep learning and plant species identification, leveraging the Malayakew plant leaf dataset, which encompasses a diverse collection of 44 plant species. Deep learning, specifically through convolutional neural networks (CNNs), has showcased exceptional performance in image classification, outperforming conventional methods and human capabilities. Our primary objective is to harness the potential of deep learning to develop a robust model capable of accurately recognizing and categorizing plant species based on visual information extracted from images.

The fusion of deep learning and plant species identification signifies a promising and innovative approach to addressing real-world challenges. This project not only serves as a foundation for the continued development of AI-based solutions in botany and environmental science but also highlights the immense potential of deep learning in addressing critical global issues. It underscores the power of technology in advancing our understanding and preservation of the natural world.

### 1.2 Motive

The motive behind this project is to harness the capabilities of deep learning and artificial intelligence to streamline and enhance the process of plant species identification. Traditional methods of plant classification are often labour-intensive, time-consuming, and prone to errors. By employing cutting-edge technology, we aim to automate and accelerate the recognition of plant species from images, providing a valuable tool for various sectors, including biodiversity conservation, agriculture, and ecological research.

### 1.3 Aim

The aim of this project is to develop a robust and accurate deep learning model for plant species identification using the Malayakew plant leaf dataset. Our objective is to create a system capable of taking an image of a plant leaf as input and providing the correct plant species classification as output.

### 1.4 Objective

The primary objective of this project is to leverage the power of deep learning to revolutionize the process of plant species identification. Through the utilization of the Malayakew plant leaf dataset, our goal is to develop a sophisticated model that can accurately classify plant species from images.

.

Plant Species Identification



## 1.5 Literature Review

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.N o | Title of the Research Paper | Authors | Algorithms used | Learnings of the study |
| 1 | Pattern Recognition | S.H. Lee, C. S. Chan, S. Mayo and P. Remagnino | Deep learning (CNN)  algorithms | How Deep Learning Extracts and Learns Leaf Features for Plant Classification |
| 2 | Deep-Plant: Plant Identification with Convolutional Neural Networks | S.H. Lee, C. S. Chan, P. Wilkin and P. Remagnino | Deconvolutional  Network | Plant recognition Deep learning Feature visualisation |
| 3 | Automatic classification of plants based on their leaves | ] A. Aakif, M.F. Khan, | Artificial Neural Networks | The proposed algorithm identifies a plant in three distinct stages i) pre-processing ii) feature extraction iii) classification. |

## 1.6 Research Gap

The research gap in this project lies in the field of automated plant species identification using deep learning techniques. While there have been significant advancements in image classification and recognition across various domains, the specific application to plant species recognition is relatively underexplored. Existing methods for plant species identification often rely on manual expertise and are time-consuming, making large-scale identification and monitoring a daunting task.

The research gap also extends to the availability of comprehensive and diverse plant datasets, which are essential for training accurate deep learning models. The Malayakew plant leaf dataset provides a valuable resource, but there is still a need for more extensive and diverse datasets to encompass a broader range of plant species and variations in environmental conditions.

Additionally, the incorporation of efficient deep learning models, such as EfficientNet, in the context of plant species recognition is a relatively novel approach. The optimization of such models for plant species identification presents a promising research area that can significantly enhance the accuracy and efficiency of the classification process.

Furthermore, the development of user-friendly interfaces and tools for non-experts in the field of botany is an underexplored aspect. Bridging this gap will enable a wider audience to access and utilize plant species identification technology for various applications.

Overall, the research gap in this project encompasses the need for advanced deep learning models, extensive datasets, and user-friendly tools to streamline the process of plant species identification, ultimately addressing the challenges in biodiversity conservation, agriculture, and ecological research.

# Chapter 2 Problem Statement

## Problem Statement:

## the problem statement involves streamlining and automating the process of plant species identification through advanced deep learning techniques.

## 2.1. Research Objectives

The objectives of the project can be summarized as follows:

1.Develop and implement a robust deep learning model for automated plant species identification using the Malayakew plant leaf dataset, with a particular focus on optimizing the EfficientNet architecture.

2.Create a user-friendly interface that allows non-experts in botany to easily input images and receive accurate plant species identification results, thereby democratizing access to this technology.

## 2.2. Methodology of the Work

The methodology for this project involves several key steps to achieve the automated identification of plant species using deep learning techniques. Here is a detailed explanation of the methodology:

**Data Acquisition and Preprocessing:**

* The project starts by acquiring the Malayakew plant leaf dataset, which includes images of various plant species. This dataset is divided into training and testing sets, each containing 44 species of plants.
* Data preprocessing is a crucial step that includes resizing the images to a consistent resolution (e.g., 256x256 pixels) and normalizing pixel values to the range [0, 1]. This ensures uniformity in the dataset and enhances model training.
* The Malayakew dataset is a collection of images of plant leaves, designed for plant species classification and recognition.
* It is commonly used for research and experimentation in the field of computer vision, deep learning, and botany.
* The dataset is categorized into 44 different plant species, with each species having a variable number of images.
* **D1 (Directory 1):**
* Total number of classes: 44
* Total number of images in the **training dataset**: 44 classes \* 52 images/class = 2,288 images
* Total number of images in the **test dataset**: 44 classes \* 12 images/class = 528 images
* **D2(Directory 2):**
* Total number of classes: 44
* Total number of images in the **training dataset**: 44 classes \* 788 images/class = 34,672 images
* Total number of images in the **test dataset**: 44 classes \* 200 images/class = 8,800 images

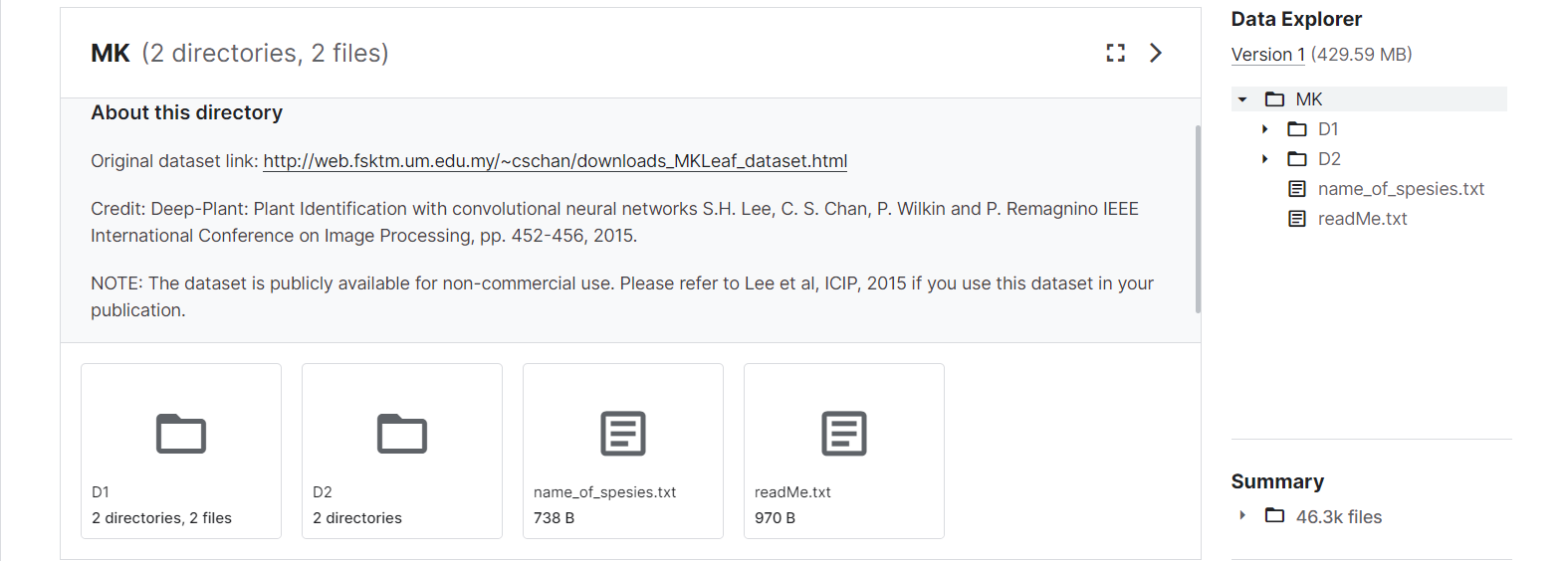


Fig 1: Malayakew Dataset

**Model Selection and Fine-tuning:**

EfficientnetB0, a state-of-the-art deep learning architecture And Xception Model is chosen as the base model for this project. The pre-trained weights of an EfficientNet and Xception variant are loaded to leverage knowledge learned from a broad range of images.

Transfer learning is employed, where the pre-trained model's weights are fine-tuned on the plant species dataset. The final layers of the model are customized to match the number of plant species in the dataset.

**EfficientnetB0:**

* **EfficientnetB0 is a pretrained model (Transfer learning)**
* **It achieves competitive performance with much fewer parameters compared to other models like ResNet or Inception.**
* **EfficientNetB0 is used in a variety of computer vision tasks, including image classification, object detection, and image segmentation.**

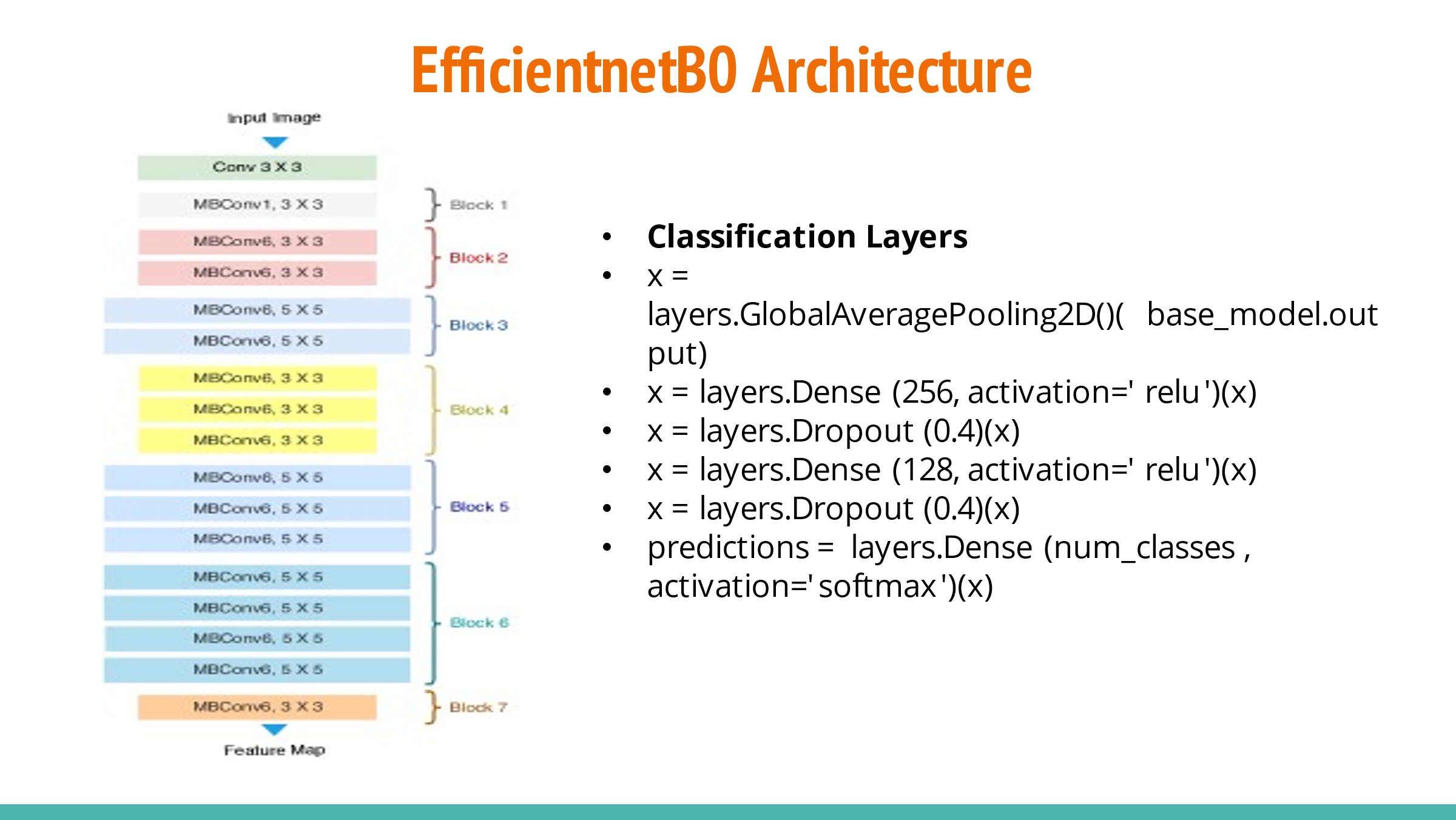


Fig 2: EfficientnetB0 Architecture

**Xception:**

* **Transfer learning** on pre-trained model (ImageNet).
* Already learnt complex **feature extraction** from large dataset.
* **Learning rate scheduling** is used to prevent overshooting the minima.
* **Image augmentation** is used to virtually increase training set size.
* **Test time augmentation** is used to get multiple outputs for voting.

**Training the Model:**

The model is trained on the training dataset, with a focus on optimizing for plant species identification. During training, data augmentation techniques may be applied, such as random rotations, flips, and brightness adjustments, to increase the model's robustness.

Training metrics, including loss and accuracy, are monitored to evaluate the model's performance. The training process continues for a specified number of epochs.

**Training the Model using EfficientnetB0:**

A screenshot of a computer program

Description automatically generated**A screenshot of a computer program

Description automatically generated**

Fig 3: Training the Model using EfficientnetB0.

**Training the Model using Xception:**

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

Fig 4: Training the Model using Xception

**Model Evaluation:**

The trained model is evaluated on the test dataset to assess its generalization capability. Evaluation metrics such as Training Loss And validation Loss and Training accuracy and Validation accuracy are calculated to measure the model's performance in identifying plant species.

**For EfficientnetB0:**

**A graph with orange lines and blue lines

Description automatically generated**

Fig 5: EfficientnetB0 Model Accuracy

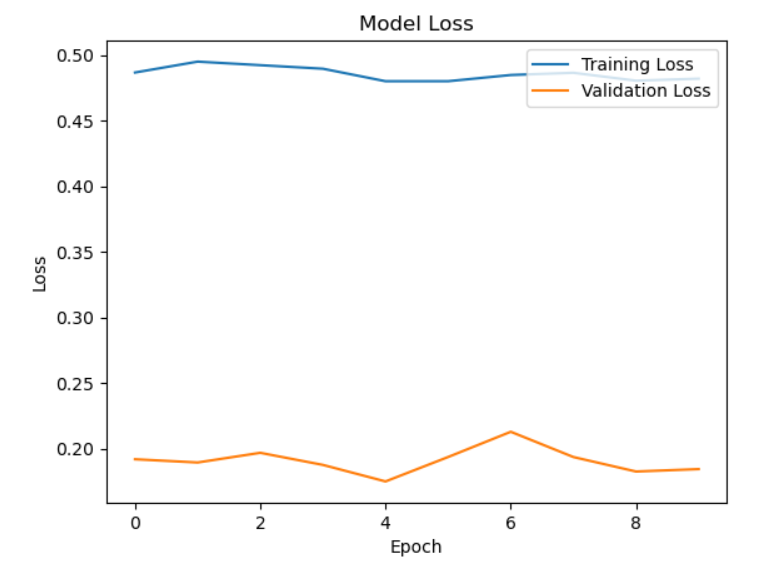
****

Fig 6: EfficientnetB0 Model Loss

**For Xception:**

**A graph of a graph showing the loss of a model

Description automatically generated with medium confidence**

Fig 7: Xception Model Accuracy

**A graph of a graph

Description automatically generated with medium confidence**

Fig 8: Xception Model Accuracy

**User Interface Development:**

To make the technology accessible to a wider audience, a user-friendly interface is developed. This interface allows users to upload images of plant leaves and receive real-time predictions of the plant species.

**Dataset Contribution:**

As part of addressing the research gap, efforts are made to contribute to the availability of plant datasets. This involves curating and expanding the Malayakew dataset to include a more diverse set of plant species and images captured under varying environmental conditions.

**Result Reporting:**

The final step involves reporting the results of the model's predictions. Users can input an image through the user interface, and the model will provide the predicted plant species name along with a probability score.

**Continuous Improvement:**

The project is an iterative process, and the model can be fine-tuned further based on feedback and new data. Continuous improvement and expansion of the dataset are essential to enhance the accuracy and scope of plant species identification.

# Chapter 3

**Conclusion and Future Scope**

## 3.1 Conclusion

## This project successfully developed an efficient plant species identification system using deep learning. By harnessing the potential of the Malayakew plant leaf dataset and the state-of-the-art EfficientNet model, the system accurately classifies plant species based on leaf images. This innovation simplifies the task of identifying plant species and has significant implications for botanists, researchers, and environmental enthusiasts. The project also contributes to expanding the dataset, filling a research gap, and demonstrates the remarkable capabilities of deep learning in addressing real-world challenges, particularly in the context of plant biodiversity. It is a significant step forward in the application of artificial intelligence for plant species recognition, with broad potential across various domains, including conservation, agriculture, and ecological research.

## 3.2 Future Scope

The future scope of this project is promising and multifaceted. Firstly, the model's accuracy and efficiency can be further improved by collecting and incorporating more diverse and extensive datasets, allowing it to recognize a broader range of plant species and variations. Additionally, the system can be extended to include additional features such as disease detection and growth stage prediction, making it a comprehensive tool for plant health monitoring. Integration with mobile applications for real-time plant identification in the field is another exciting avenue to explore, making the technology accessible to a broader audience. Collaboration with botanists and environmental organizations can further refine and validate the system. Moreover, exploring the potential for automating data collection and analysis for ecological studies is a direction that aligns with emerging trends in conservation and biodiversity research. This project lays a strong foundation for future developments in the domain of plant species identification and has the potential to make a substantial impact on the field of botany and environmental science.

19

# References

1. S.H. Lee, C. S. Chan, S. Mayo and P. Remagnino  
   Pattern Recognition, vol. 71, pp. 1-13, 2017 **(PR 2017)**

https://web.fsktm.um.edu.my/~cschan/doc/PR2017.pdf

1. Deep-Plant: Plant Identification with convolutional neural networks  
   S.H. Lee, C. S. Chan, P. Wilkin and P. Remagnino  
   IEEE International Conference on Image Processing, pp. 452-456, 2015 **(ICIP *2015)***

https://web.fsktm.um.edu.my/~cschan/doc/150608425v1.pdf

1. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems (NIPS).
2. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).
3. Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).

20