Plant Leaf Disease Detection

Indian Institute of Information Technology, Guwahati
Advisor: Dr. Nilkanta Sahu

Ganesh Kavthekar
Roll no.: 2101099

This disseration is submitted for the degree of
Bachelors of Technology

IIIT GUWAHATI

April 2024

Table of Contents

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Ganesh Kavthekar
April 2024

Acknowledgement

I would like to thank the Department of CSE for going ahead and allowing me to pursue a research project aligning with the directions of my interest. I would like to thank my advisor, Dr. Nilkanta Sahu, for continuously guiding me throughout the project and motivating me to keep moving.
Finally, I'd like to thank that portion of the IIITG community who've supported me in keeping my morale high during difficult times.

Abstract

Crop production can be greatly reduced due to various diseases, which seriously endan- gers food security. Thus, detecting plant diseases accurately is necessary and urgent. Traditional classification methods, such as naked-eye observation and laboratory tests, have many limitations, such as being time consuming and subjective. Currently, deep learning (DL) methods, especially those based on convolutional neural network (CNN), have gained widespread application in plant disease classification.
They have solved or partially solved the problems of traditional classification methods and represent state-of-the-art technology in this field. In this work, we reviewed the latest CNN networks pertinent to plant leaf disease classification.
We summarized DL principles involved in plant disease classification. Additionally, we summarized the main problems and correspond- ing solutions of CNN used for plant disease classification. Furthermore, we discussed the future development direction in plant disease classification.

1 Introduction

In recent years, the field of computer vision has undergone significant advancements, particularly in the realm of plant disease detection leveraging deep learning methodologies. This study delves into the exploration of Convolutional Neural Network (CNN) architectures, such as VGG16, ResNet50, and EfficientNet, which have been extensively researched and applied for precise disease classification based on plant leaf imagery. Despite these efforts, further improvements in accuracy pose ongoing challenges.

The primary focus of this research is the integration of the Convolutional Block Attention Module (CBAM) into established CNN architectures for plant disease detection. CBAM is renowned for its prowess in capturing both spatial and channel-wise attention, demonstrating promising outcomes across various computer vision tasks. The overarching goal is to harness CBAM's attention mechanism to augment the discriminative power of baseline models, thereby enhancing disease classification accuracy.

Experimental evaluations were meticulously conducted on benchmark datasets, systematically comparing the performance of CBAM-enhanced models against their respective baseline counterparts. Unexpectedly, the results did not showcase a substantial increase in accuracy across all architectures. Several contributing factors were identified. Firstly, the inherent complexity and heterogeneity of plant diseases present formidable challenges in effectively capturing relevant features. Moreover, the limited size and diversity of datasets may impede the models' generalization capability, consequently constraining the potential benefits of CBAM integration.

2 Literature Review

Plant leaf disease classification using EfficientNet deep learning model[1]

Timely and accurate diagnosis of plant diseases is crucial for sustainable agriculture and resource management. While some diseases may not exhibit visible symptoms, the conventional approach involves expert plant pathologists visually inspecting infected leaves for characteristic signs. However, the increasing diversity of plants, variations in disease manifestation due to climate change, and rapid disease spread pose challenges even for experienced pathologists. The integration of expert systems and intelligent technologies for automated disease diagnosis offers significant advantages to agronomists and non-expert farmers alike.

Traditional machine learning models have been widely explored for plant disease detection, employing techniques such as spectral analysis, image segmentation, and feature extraction followed by classification using algorithms like Support Vector Machines (SVM) and Artificial Neural Networks (ANN). Recent years have witnessed a surge in research focusing on deep learning architectures for plant disease classification, leveraging the capacity of convolutional neural networks (CNNs) to process large-scale image data efficiently.

Studies such as those by Sladojevic et al. and Chen et al. have demonstrated the effectiveness of CNNs in classifying various plant diseases using datasets like PlantVillage. These studies have achieved high accuracy rates, indicating the potential of deep learning in disease diagnosis. However, many existing approaches focus on specific plants or diseases within limited datasets, leaving room for exploration of broader classification frameworks.

Efforts to address this gap have led to the emergence of novel deep learning architectures, including EfficientNet, designed to optimize computational efficiency while maintaining high performance. EfficientNet, characterized by compound scaling and inverted bottleneck building blocks, offers a promising avenue for achieving accurate and resource-efficient disease classification. Notably, its scalability across different model sizes presents opportunities for accommodating diverse hardware constraints and deployment scenarios.

Transfer learning, another prominent technique in deep learning, facilitates the adaptation of pre-trained models to new tasks with minimal data requirements. This approach has been instrumental in accelerating the development of disease classification models by leveraging knowledge acquired from large-scale image datasets like ImageNet.

The state-of-the-art CNN architectures, including AlexNet, VGG16, ResNet50, Inception V3, and EfficientNet, have been extensively studied for their efficacy in plant disease classification. Each architecture offers unique features and computational characteristics, contributing to the exploration of diverse model configurations for optimal performance.

Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification[3]

The excerpt provides an in-depth exploration of how convolutional neural network (CNN)-based deep learning (DL) techniques are being utilized to tackle the complex task of plant disease detection and classification. It starts by identifying key challenges that researchers encounter in this field. One major hurdle is the scarcity of diverse and extensive datasets. Without a sufficiently large and varied dataset, DL models may struggle to generalize well to new, unseen instances, leading to decreased performance when deployed in real-world settings. Additionally, the excerpt points out issues related to model robustness, highlighting how CNN models trained on ideal datasets may fail to perform adequately when faced with the diverse and unpredictable conditions encountered in agricultural environments.

Furthermore, the excerpt discusses the nuances of disease symptom variations and the influence of image backgrounds on classification accuracy. Plant diseases can manifest differently depending on various factors such as disease progression, environmental conditions, and plant characteristics. This variability poses a significant challenge for DL models, as they must be able to accurately classify diseases despite these fluctuations. Moreover, the background of images used for training and testing can affect model performance. Busy backgrounds or similar features between the background and the region of interest can confuse the model, leading to misclassification.

In response to these challenges, the excerpt outlines several solutions that researchers have proposed and implemented. Transfer learning, for example, allows DL models to leverage knowledge gained from pre-trained networks on large datasets, thus reducing the need for extensive amounts of labeled data. Data augmentation techniques, such as rotation, mirroring, and noise addition, help to diversify the training dataset, enabling models to learn robust representations of disease patterns. Additionally, robust model training methods, including multi-condition training and persistently enriching dataset diversity, aim to improve model generalization and adaptability to real-world conditions.

By providing detailed examples from relevant research studies, the excerpt illustrates how these solutions have been successfully applied in practice, demonstrating their effectiveness in enhancing the accuracy and reliability of plant disease classification systems. Overall, the excerpt offers a comprehensive overview of the advancements in CNN-based DL methods for plant disease detection and classification, shedding light on both the challenges faced and the innovative strategies developed to overcome them.

Efficientnet: Rethinking model scaling for convolutional neural networks[2]

EfficientNet is a convolutional neural network (CNN) architecture that has gained attention for its effectiveness in balancing model efficiency and performance. Developed by Tan and Le in 2019, EfficientNet introduces a novel scaling method that uniformly scales dimensions of depth, width, and resolution in a compound manner. This approach ensures that the network achieves optimal performance across various resource constraints, making it well-suited for deployment on different hardware platforms and in diverse computational environments.

The key innovation of EfficientNet lies in its compound scaling method, which involves systematically increasing the model's depth, width, and resolution simultaneously. By scaling these dimensions in a balanced and controlled manner, EfficientNet achieves superior performance compared to traditional scaling methods that only adjust one aspect of the network's architecture. This holistic approach enables EfficientNet to achieve state-of-the-art results on benchmark image classification tasks while maintaining high efficiency in terms of computational resources and model size.

EfficientNet consists of multiple blocks, including the compound scaling block, mobile inverted bottleneck convolution (MBConv) blocks, and squeeze-and-excitation (SE) blocks. These blocks are carefully designed to maximize computational efficiency without sacrificing performance. The compound scaling block determines the scaling coefficients for depth, width, and resolution based on a predefined compound coefficient, which ensures that the network achieves optimal trade-offs between accuracy and efficiency.

The MBConv blocks, inspired by MobileNetV2, serve as the primary building blocks of EfficientNet. They incorporate depthwise separable convolutions and inverted residuals to reduce computational complexity while maintaining expressive power. Additionally, the SE blocks enhance feature representation by adaptively recalibrating channel-wise feature responses, further improving the model's performance.

3 Methodology

3.1 Architectures

VGG16:
VGG16, also known as the Visual Geometry Group 16, is a renowned convolutional neural network architecture devised by Simonyan and Zisserman in 2014. Comprising 16 layers, VGG16 stands out for its uniform design, employing 3x3 convolutional filters with a stride of 1 and same padding, along with 2x2 max-pooling filters with a stride of 2. With approximately 138 million parameters, it is celebrated for its simplicity and effectiveness in various image classification tasks.

AlexNet:
AlexNet, a groundbreaking convolutional neural network architecture, was developed by Krizhevsky et al. in 2012. This model secured victory in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012, marking a pivotal moment in deep learning for computer vision. Comprising 8 layers, including 5 convolutional and 3 fully connected layers, AlexNet introduced novel elements such as the Rectified Linear Unit (ReLU) activation function, local response normalization, and dropout regularization. With approximately 61 million parameters, AlexNet significantly contributed to the proliferation of deep learning in image recognition domains.

ResNet50:
ResNet50, proposed by He et al. in 2015, is a sophisticated convolutional neural network archi-

tecture that emerged victorious in the ILSVRC 2015 competition. Addressing the challenge of vanishing gradients in deep networks, ResNet50 incorporates residual connections, facilitating gradient flow during training. With its 50 layers predominantly composed of residual blocks, this architecture employs 3x3 convolutional filters. ResNet50 is acclaimed for its exceptional performance in image classification and feature extraction tasks, achieving remarkable accuracy while maintaining a parsimonious parameter count relative to preceding models.

Compound Scaling:
Compound scaling is a pioneering technique in neural network architecture design, meticulously crafted to optimize model efficiency and performance. Developed as a part of the EfficientNet framework, compound scaling systematically adjusts network depth, width, and resolution in a balanced manner, leveraging a user-defined coefficient to allocate computational resources judiciously. By uniformly scaling these dimensions, compound scaling ensures that the resulting models achieve superior accuracy while maintaining computational efficiency, making it a cornerstone methodology in modern deep learning research and applications.

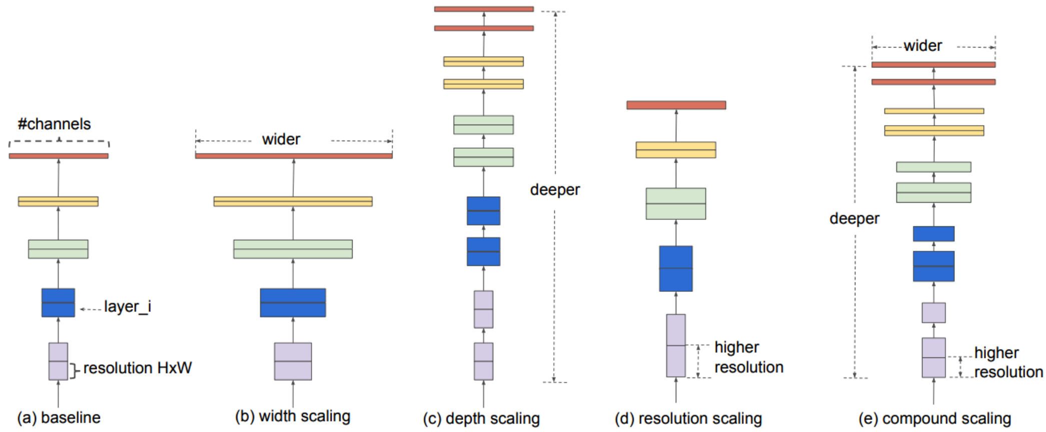

Figure 1: Compound Scaling

Efficient Net:
EfficientNet, a groundbreaking convolutional neural network architecture, offers a remarkable balance between model efficiency and performance. Introduced in 2019, EfficientNet achieves state-of-the-art accuracy on image classification tasks while requiring fewer computational resources compared to other models. By uniformly scaling depth, width, and resolution, EfficientNet optimizes model architecture across multiple dimensions, resulting in highly efficient and effective networks. This approach enables EfficientNet to achieve impressive accuracy even with smaller model sizes, making it well-suited for resource-constrained environments such as mobile devices or edge computing platforms.

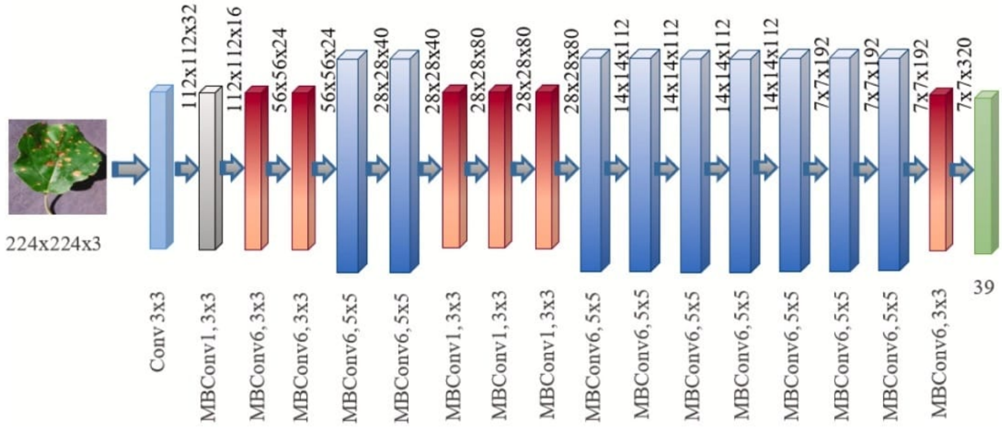

Figure 2: Efficient Net

CBAM:
CBAM, or Convolutional Block Attention Module, represents a breakthrough in convolutional neural network (CNN) architectures, emphasizing the importance of attention mechanisms in feature extraction. Introduced as an enhancement to existing architectures like ResNet50, CBAM integrates spatial and channel-wise attention mechanisms into CNN blocks, enabling the network to focus on salient features while suppressing irrelevant information. This adaptive attention mechanism significantly enhances the network's ability to capture contextually relevant features, leading to improved performance in various computer vision tasks, including image classification and object detection. With its innovative attention-based approach, CBAM stands at the forefront of advancements in deep learning architectures, paving the way for more efficient and accurate models in the field of computer vision.

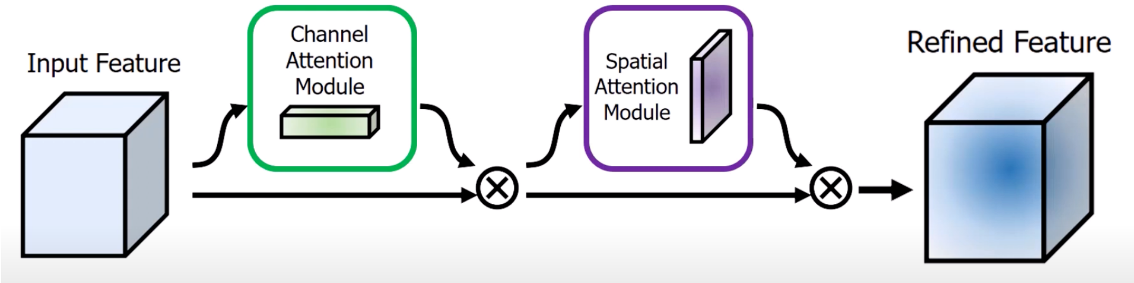

Figure 3: CBAM

The **channel attention** mechanism in CBAM enhances feature representations by highlighting informative channels and suppressing less relevant ones. It computes attention maps to capture the importance of each feature channel, allowing the network to focus on discriminative channels while attenuating noise. This adaptive recalibration of feature responses improves the network's discriminative power, leading to enhanced performance in tasks like image classification and object detection.

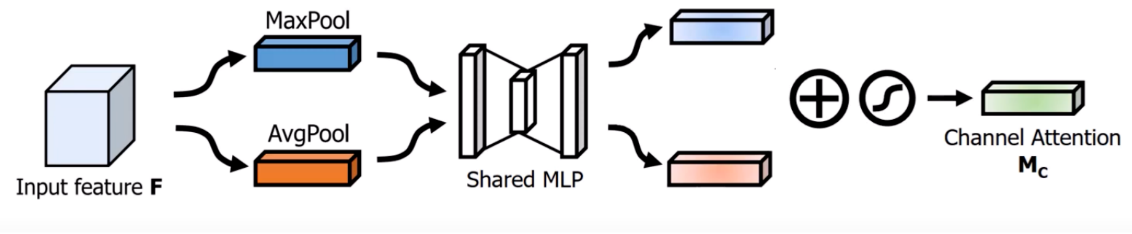

Figure 4: Channel Attention

The **spatial attention** mechanism in CBAM selectively emphasizes relevant spatial regions within feature maps while suppressing irrelevant ones. By generating attention maps that highlight informative regions and suppress distracting ones, spatial attention enables the network to focus on salient image regions, enhancing its ability to capture spatial context and fine-grained details. This mechanism improves the network's spatial awareness and enables more precise localization of objects or features, contributing to enhanced performance in tasks such as object recognition and semantic segmentation.

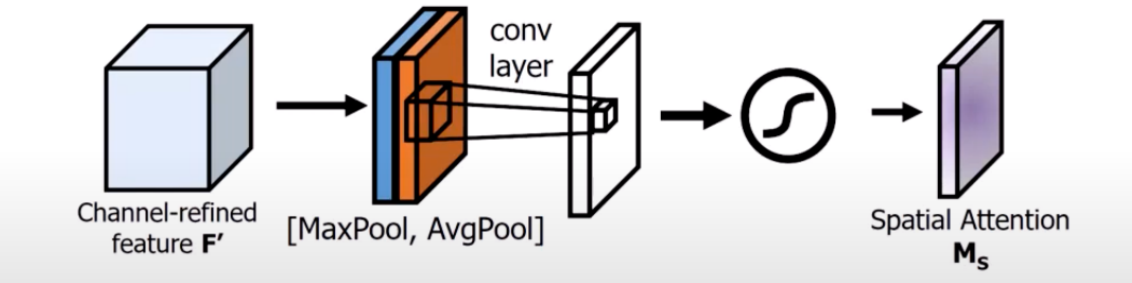

Figure 5: Spatial Attention

3.2 Details

The several methodologies used in the comparison of state-of-the-art CNN architectures such as AlexNet, VGG16, ResNet50, Inception V3, and EfficientNet. Here are the methodologies mentioned:
Model Architectures: Each CNN architecture (AlexNet, VGG16, ResNet50, EfficientNet) has its unique design, including the number of layers, types of layers (convolutional, pooling, fully connected), activation functions (ReLU, Softmax, Swish), and other architectural features.
Parameter Comparison: The comparison includes the number of parameters each model has. For example, AlexNet has approximately 61 million parameters, VGG16 has around 138 million parameters, ResNet50 has a specific number of parameters, and EfficientNet models vary between B0 and B7, with the number of parameters increasing slightly with each model number.
Input Image Size: Each model has a specific input image size. For example, AlexNet takes images of size 227x227 pixels, VGG16 takes images of size 224x224 pixels, ResNet50 takes images of size 224x224 pixels, Inception V3 takes images of size 299x299 pixels, and EfficientNet models have varying input image sizes depending on the model.
Activation Functions: Different activation functions are used in different architectures. For example, AlexNet and VGG16 use ReLU activation functions, Inception V3 uses Softmax activation function, and EfficientNet uses Swish activation function instead of ReLU.
Efficiency Metrics: The efficiency of the architectures is evaluated based on factors such as accuracy, computational load and resource utilization. EfficientNet, in particular, is highlighted for its efficiency in terms of accuracy and computational load compared to other state-of-the-art models.
Scaling Methodology: EfficientNet introduces a novel scaling methodology called compound scaling, which uniformly scales depth, width, and resolution of the network to achieve better performance without significantly increasing computational load. This methodology involves grid search to find the best scaling coefficients and then scaling up the baseline network to create EfficientNet models of different sizes.

3.3 Datasets

Soynet:
Indian Soybean Image dataset with quality images captured from the agriculture field (healthy and diseased Images).
Raw dataset and preprocessed dataset with a resolution of 256x256 pixels present in the dataset. This dataset consists of 9000+ high-quality images of soybeans (healthy and Disease quality).
Plant Village Dataset:
In this data-set, 39 different classes of plant leaf and background images are available. The data-set containing 61,486 images. We used six different augmentation techniques for increasing the data-set size. The techniques are image flipping, Gamma correction, noise injection, PCA color augmentation, rotation, and Scaling.

4 Results and Discussion

4.1 Results

Architecture	No. Of parameters	test accuracy	% validation accuracy	loss
$VGG16$	134,420,327	0.9314	0.9567	0.7122
$Alexnet$	58,441,127	0.8995	0.9363	1.0201
$Resnet50$	25,636,712	0.9539	0.9778	0.8712
$EfficientNet$	5,330,571	0.9526	0.9843	1.1230

Table 1: Results for differenct architectures without CBAM

Architecture	No. Of parameters	test accuracy	% validation accuracy	loss
$VGG16$	134,573,398	0.8823	0.9185	0.8832
$Alexnet$	58,476,496	0.8401	0.8779	1.0448
$Resnet50$	25,650,588	0.9138	0.9457	1.5673
$EfficientNet$	5,343,579	0.9137	0.9567	0.7862

Table 2: Results for differenct architectures with CBAM

4.2 Discussion

Complexity Overhead: The Convolutional Block Attention Module (CBAM) introduces additional complexity to the model architecture, involving computations for generating attention maps. This added complexity may not always translate to a proportional improvement in model performance, especially in scenarios where the task can be effectively handled by simpler architectures.
Overfitting Risk: The inclusion of CBAM increases the model's capacity to capture intricate patterns and features in the data. However, in cases where the dataset is limited or lacks diversity, the use of complex attention mechanisms like CBAM may lead to overfitting. The model may become overly specialized in capturing nuances present in the training data but fail to generalize well to unseen data.
Training Data Characteristics: The effectiveness of attention mechanisms like CBAM heavily relies on the presence of salient features and patterns in the training data that warrant selective attention. If the dataset lacks sufficient complexity or diversity, CBAM may not have significant features to attend to, resulting in minimal performance improvement compared to simpler architectures.
Computational Overhead: CBAM requires additional computational resources for computing attention maps, which can increase both training and inference times. In scenarios where computational resources are limited or time constraints are stringent, opting for simpler architectures without attention mechanisms may be preferred to achieve faster model training and deployment.
Hyperparameter Sensitivity: Architectural modifications like integrating CBAM require careful tuning of hyperparameters to ensure optimal performance. The effectiveness of CBAM may be sensitive to hyperparameters such as learning rate, regularization strength, and batch size. Inadequate hyperparameter tuning could lead to suboptimal performance of models with CBAM compared to simpler architectures.
Task Complexity and Dataset Size: The impact of CBAM on model performance may vary depending on the complexity of the task and the size of the dataset. For tasks with

inherently complex structures or large-scale datasets, CBAM may offer significant performance gains by enabling the model to focus on relevant features. However, for simpler tasks or smaller datasets, the benefits of CBAM may be marginal compared to the computational costs it incurs.
Model Interpretability: The presence of CBAM may enhance the model's ability to selectively attend to informative features, but it can also compromise model interpretability. Understanding how attention mechanisms influence model decisions and interpreting the attention maps generated by CBAM may pose challenges, potentially hindering the model's explainability and transparency.

5 Conclusion

The study encompasses a range of prominent CNN architectures, including VGG16, AlexNet, ResNet50, and EfficientNet, each offering unique design principles and trade-offs in terms of model complexity, parameter count, and computational efficiency.
Comparative analysis reveals the strengths and weaknesses of each architecture concerning tasks such as image classification and feature extraction. ResNet50, with its innovative residual connections, stands out for its exceptional performance and robustness, while EfficientNet demonstrates impressive accuracy with reduced computational overhead.
The project delves into advanced methodologies such as compound scaling and attention mechanisms like CBAM, shedding light on their contributions to improving model efficiency and performance in deep learning tasks.
The insights gained from this study hold significant implications for real-world applications in computer vision, ranging from medical image analysis and autonomous driving to industrial automation and augmented reality.

References

[1] Atila, Ümit, Murat Uçar, Kemal Akyol, and Emine Uçar. "Plant leaf disease classification using EfficientNet deep learning model." Ecological Informatics 61 (2021): 101182.

[2] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." In International conference on machine learning, pp. 6105-6114. PMLR, 2019.

[3] Lu, Jinzhu, Lijuan Tan, and Huanyu Jiang. "Review on convolutional neural network (CNN) applied to plant leaf disease classification." Agriculture 11.8 (2021): 707.