PROJECT TITLE:
Plant Disease Classification

PROJECT SYNOPSIS

OF MAJOR PROJECT

BACHELOR OF TECHNOLOGY
Computer Science and Engineering

SUBMITTED BY:

Utkarsh Arora (2000290100177)

Vaibhav Singh (2000290100181)

Samaria Singh (2000290100130)

November 2023

KIET Group of Institutions, Delhi-NCR,
Ghaziabad (UP)
Department of Computer Science and Engineering

Plant Disease Classification

• Objective

To build a model, which can classify between healthy and diseased crop leaves and also if the crop has any disease, predict which disease is it.

• The Dataset Description:

This dataset consists of about 87K rgb images of healthy and diseased crop leaves which is categorized into 38 different classes. The total dataset is divided into 80/20 ratio of training and validation set preserving the directory structure. A new directory containing 33 test images is created later for prediction purpose.

We use the Plant Village dataset, which is an open access dataset available from Kaggle. The Plant Village dataset contains approximately 70,000 images of completely healthy and unhealthy crops, of which 38 categories have already been flagged. Our concern is to select input images of perfectly healthy apple leaves and input images of blight caused by the fungus. Domain experts, or botanists, assign each input image to an appropriate category: early healthy stage, intermediate healthy stage, healthy stage, or fully healthy. In the fully healthy stage, there are no spots on the leaves. In the early stage, healthy leaves have small round spots with a radius of about 2,5 mm. Middle stage, entire leaves with more spots that grow in a random or low shape. In the last stage, the healthy leaves are more polluted by the tree and cannot stay on the tree. All input images are screened by do- main experts and classified into appropriate disease. 180 images that experts found difficult to classify were removed from further processing.

• Problem Identification and Formulation

Plant diseases are known from times preceding the earliest writings. Fossil evidence indicates that plants were affected by disease 250 million years ago. Plant disease, an impairment of the normal state of a plant that interrupts or modifies its vital functions. All species of plants, wild and cultivated alike, are subject to disease. Although each species is susceptible to characteristic diseases, these are, in each case, relatively few in number. The occurrence and prevalence of plant diseases vary from season to season, depending on the presence of the pathogen, environmental conditions, and the crops and varieties grown.
Some plant varieties are particularly subject to outbreaks of diseases while others are more resistant to them.

Project Objective:

- The primary goal of this project is to develop a Deep Learning model to assist in the early and accurate identification of crop diseases. The model will classify crop leaves into healthy or diseased categories and, if diseased, identify the specific disease type.

- Real-world testing on a held-out dataset simulates practical usage, ensuring the model's utility. Interpretability techniques are applied to provide transparency in predictions.

- The trained model is deployed in a user-friendly application, allowing users to upload crop leaf images for health and disease predictions. Continuous monitoring and retraining are essential to adapt to evolving conditions and diseases.

Methodology/Planning of Work

This project employs a ResNet (Residual Network) methodology to address the task of classifying crop leaves as healthy or diseased and identifying specific diseases when present. It begins with the assembly of a well-annotated dataset of crop leaf images, encompassing both healthy and diseased samples. Image preprocessing ensures uniformity in size and pixel values, and data augmentation techniques boost dataset diversity.

The ResNet architecture is chosen for its proficiency in image classification, and we leverage transfer learning. The model's pre-trained weights are used, and the classification layer is adapted for our task. Fine-tuning focuses on optimizing the model's performance on our specific dataset.

Training involves appropriate optimizers and loss functions, with hyperparameters tuned using the validation set. Regularization techniques like dropout and batch normalization are employed to curb overfitting. Model interpretability is supported with visualization techniques such as class activation maps.

Model evaluation, using metrics like accuracy and precision, takes place on the validation set. The project's ultimate test comes with the assessment of real-world performance on a held-out test dataset. Deployment includes integrating the model into a user-friendly application.

Technologies Used:

RESNET-50: What Is the ResNet-50 Model?ResNet stands for Residual Network and is a specific type of convolutional neural network (CNN) introduced in the 2015 paper "Deep Residual Learning for Image Recognition" by He Kaiming, Zhang Xiangyu, Ren Shaoqing, and Sun Jian. CNNs are commonly used to power computer vision applications. ResNet-50 is a 50-layer convolutional neural network (48 convolutional layers, one MaxPool layer, and one average pool layer). Residual neural networks are a type of artificial neural network (ANN) that forms networks by stacking residual blocks.

Special characteristics of ResNet-50

ResNet-50 has an architecture based on the model depicted above, but

with one important difference. The 50-layer ResNet uses a bottleneck design for the building block. A bottleneck residual block uses 1x1 convolutions, known as a "bottleneck", which reduces the number of parameters and matrix multiplications. This enables much faster training of each layer. It uses a stack of three layers rather than two layers.

PyTorch: This tutorial introduces the fundamental concepts of PyTorch through self-contained examples. At its core, PyTorch provides two main features:

- An n-dimensional Tensor, like numpy but can run on GPUs.
- Automatic differentiation for building and training neural networks we will use problem of fitting y = sin(x) with a third order polynomial as our running example. The network will have four parameters and will be trained with gradient descent to fit random data by minimizing the Euclidean distance between the network output and the true output.

Literature Review:

Research Paper 1:

Title: Plant diseases detection with low resolution data using nested skip connections

Author: **Hilman F. Pardede1* , Endang Suryawati1 , Vicky Zilvan1 , Ade Ramdan1 , R. Budiarianto S. Kusumo1 , Ana Heryana1 , R. Sandra Yuwana1 , Dikdik Krisnandi1 , Agus Subekti2 , Fani Fauziah3 and Vitria P. Rahad**

Now, there are increasing trends of using deep learning for plant diseases detection. However, their implementations may be difficult in developing countries due to several reasons. First, existing deep learning models are usually trained with images with adequate resolutions. In developing countries however, with limited internet connection, models that would perform well even when data with low resolution are used are needed. Secondly, the generated models are large. Hence, most deep learning-based applications are available on-line. Unfortunately, the trend for new deep learning architectures is either have larger models or require a heavy memory usage. So, models with smaller size would be preferred. In this paper, we evaluate various existing deep learning models for plant diseases detection when low resolution data are used. They are: VGGNet, AlexNet, Resnet, Xception, and MobileNet. Our focus is deep convolutional neural network (DCNN) which is commonly applied for image data. We also propose a new DCNN architecture with two branches of concatenated residual networks. It is well known that the deeper the networks the better performance of DCNN. However, DCNN with very deep networks and large number of training parameters is prone to vanishing gradient problems. One solution for that is to apply residual networks as branches to DCNN. While it is found that increasing the branch of the networks beneft the performance, larger memory is required to train the networks. So, we apply two concatenated residual networks only. We called it Compact Networks (ComNet). We compare our method other with six popular CNN architectures. We evaluate the performance on the PlantVillage dataset and our own dataset. We

collected images of tea leaves which consist of 6 classes: 5 classes of diseases that are commonly found in Indonesia and a healthy class. Our experiments show that our method is generally better than referenced DCNN networks. Keywords: Deep Learning, Convolutional neural network, Light CNN, Plant diseases detection, Tea diseases detection.

Research paper 2:

Title: Deep Learning-Based Leaf Disease Detection in Crops Using Images for Agricultural Applications

Author: **Andrew J. , Jennifer Eunice , Daniela Elena Popescu , M. Kalpana Chowdary, and Jude Hemanth**

The agricultural sector plays a key role in supplying quality food and makes the greatest contribution to growing economies and populations. Plant disease may cause significant losses in food production and eradicate diversity in species. Early diagnosis of plant diseases using accurate or automatic detection techniques can enhance the quality of food production and minimize economic losses. In recent years, deep learning has brought tremendous improvements in the recognition accuracy of image classification and object detection systems. Hence, in this paper, we utilized convolutional neural network (CNN)-based pre-trained models for efficient plant disease identification. We focused on fine tuning the hyperparameters of popular pre-trained models, such as DenseNet-121, ResNet-50, VGG-16, and Inception V4. The experiments were carried out using the popular PlantVillage dataset, which has 54,305 image samples of different plant disease species in 38 classes. The performance of the model was evaluated through classification accuracy, sensitivity, specificity, and F1 score. A comparative analysis was also performed with similar state-of-the-art studies. The experiments proved that DenseNet-121 achieved 99.81% higher classification accuracy, which was superior to state-of-the-art models.