



CSE 445 - Machine Learning

Report

Plant Disease Classification with ML & Android

Submitted to: Intisar Tahmid Naheen (ITN)

Section: 03

Group 03

Md-Raihanul Hoque Shanto - 1812746042

Md. Imanur Rahman Emon – 1813181642

Sourov Podder Surzo – 1731323042

Title: Plant Disease Classification with ML & Android

Abstract:

Agriculture is associated with us from the very beginning of our socialization and modernization. Our modern world is rooted from agriculture. People learned how to grow plants, how to farm animals, how to use the materials of the environment to aid themselves. Day by day they invented many things for the benefit of themselves. We humans invented more techniques and developed sophisticated machines for us. This development of science touched every aspect of our life, hence the agricultural field too. We studied, explored and invented ways to produce more with less resources to fulfill the demands of ours. Though we have gathered our knowledge, it is not always possible to apply the knowledge properly. There are many obstacles behind these situations. Maybe we have the right tool but not always on the right hand.

Computer science is one of the newest fields of study. With the help of computer science, we could solve many complex problems which were unthinkable once. This field of study is advancing so rapidly that every year we get to see more and more sophisticated technologies. Among these advancements, Artificial Intelligence or AI has shown us huge promises. Nowadays with the help of AI we are able to solve problems like detecting issues or predicting something with decent accuracy, in fact in some cases they surpassed the human.

This technology is being used in every sector and agriculture is also one of them. With the help of AI, we can track diseases on crops, predict production rates and many more. Scientists had been working to develop modern systems to aid the people associated with agriculture and farming to aid them in

various aspects. Most of the time the farmers are not aware of the disease they face in their crop and also it is not always possible to get help from the authorities in rural places. Now, with the advancement of technology and AI, we can build systems which can aid the farmers remotely without causing much trouble. So, in our project we intend to develop a system that can detect diseases in Potato with the help of AI and android.

Table of Contents:

- I. Abstract
- II. Introduction
 - A. Machine Learning
 - B. Android
- III. Related Works
 - A. Resource 1
 - B. Resource 2
 - C. Resource 3
- IV. Methodology
 - A. System Diagram
 - B. Dataset
 - C. Model
 - D. Work Plan
 - E. Software and Technology requirement
- V. Model Training and Result
- VI. Conclusion & Future Works
- VII. References

Introduction

Bangladesh is an agricultural country. Almost 70% to 80% of the population is somehow associated with agriculture. Among all, Potato is one of the most produced crop in Bangladesh. It is produced across the country and from cultivation to transportation, many people and industries are associated with it. But Potato can face several diseases which can disrupt the production. Among these, Early Blight and Late Blight are common. But our farmers are not qualified or privileged enough to detect these diseases by themselves. And it is not

always possible to contact proper authorities immediately in rural areas. Due to these issues, many farmers across the country face less production issues every year. And as we discussed above this can cause great impacts on all the persons and industries associated with it. Maybe a farmer can easily cure the disease but in the first place he has to be able to identify the disease. Also, in rural places across the country the internet can be an issue.

So, we intend to develop a system which can help the farmers to detect the diseases in Potato with the help of machine learning and android. Our application will be able to take pictures from the phone and detect common rice diseases without accessing the internet. The Potato diseases that we are willing to work on can be detected from the appearance of the rice leaves. Early Blight and Late Blight left various markings on the Potato leaves thus we can distinguish them from healthy Potato leaves. But to identify the specific disease someone must have that prior knowledge about the diseases. Our goal is to build such a system which will detect the specific disease from the leaf pictures.

A. Machine Learning:

Machine learning is a field of Artificial Intelligence where we let a machine learn something by itself based on the data provided. For example, in our project, we will provide some data on healthy and infected Potato leaves and with the help of some ML algorithms, our computer program will learn how to detect which Potato leaf is good and which has a disease.

The core concept of machine learning is to find features from provided data and establish a relation between them. The entire thing happens with the help of various

machine learning algorithms on mathematical models. First the machine assigns certain values to link the features of data and then with the help of loss functions it measures the accuracy. Then again depending on the result the machine sets new values for solving the problem.

B. Android:

Android is the most popular operating system in the world. There are more than 2.5 billion android users in the whole world. Because of android, almost all types of people got the opportunity to use smart phones. Nowadays, android powered devices are really good at performing basic to advanced tasks. With the development of faster chips and better cameras, any budget android phone can capture good pictures.

Related Works

For our project we chose the following 3 research papers to understand how they did it.

Resource 1:

Rice Disease Identification and Classification by Integrating Support Vector Machine with Deep Convolutional Neural Network. [1]

Published by: Dept. of Electrical and Electronic Engineering.
Hajee Mohammad Danesh Science and Technology University.

In this research, they have used a deep convolutional neural network with support vector machine algorithm to detect 9 different classes of rice disease. The dataset contained high-resolution pictures taken with professional cameras. The dataset is prepared by themselves. They used Transfer learning and achieved 97.5% accuracy.

Resource 2:

Rice Crop Disease Identification and Classifier. [2]

Published by: Dept. of CSE, Rajiv Gandhi College of Engineering, India.

Here they also used a DCNN with SVM algorithm but they emphasized more on image processing. Here the images were taken with high-end mobile cameras as their intention was to deploy the system into smartphones. Their system uses a cloud-based solution where a photo is uploaded to the cloud and the cloud sends the prediction results back to the device

Resource 3:

Rice Disease Classification using Convolutional Neural Networks. [3]

Published by: Dept. of CSE, Rochester Institute of Technology.

This study is more focused on machine learning models. Here they trained the dataset on various small to complex pre-trained CNN models and showed the differences in training time, accuracy, and performance across various popular models like AlexNet, GoogLeNet, ResNet-50, Inception-v3, ShuffleNet and MobileNet-v2.

*None of the papers mentioned their dataset as public or provided any relevant resource.

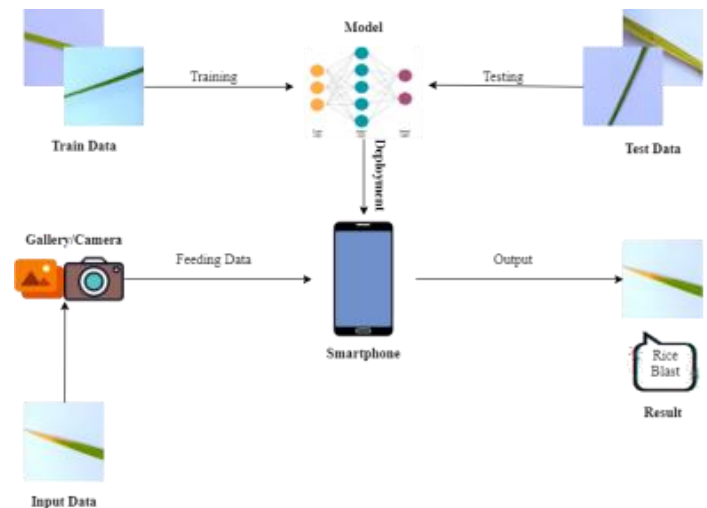
Methodology:

So, our goal is to build a system which can detect Potato diseases remotely from a smartphone. To reach our goal we have to go through various processes.

- First, we need to choose a dataset which we will use for our machine learning project. Then we need to study and organize the dataset.

- Then, we need to select a machine learning model to train with the dataset.
- Then, after testing and development we need to deploy the model into android as an android application.

System Diagram:



Dataset:

Dataset is one of the main components of the machine learning process. The Machine learning models are trained on datasets. A proper datasets can help to build an ML project with much more efficiency and lesser hassle. Due to the development of internet technologies, we have access to more data nowadays but it is not always easy to find a properly maintained dataset. Dataset exploration can be a tricky process but can't be avoided.

We know, Kaggle as our dataset resource as it is the most popular public dataset service platform. We worked with 2 different datasets.

Dataset: Plant Village Dataset by Abdallah Ali:



Plant Village Dataset contains 50,000 expertly curated images on healthy and infected leaves of various crop plants. These plants include Apple, Corn, Strawberry, Pepper, Potato, Tomato, etc. The dataset has 3 subfolders of 38 different classes, one contains a colorful pictures, one contains all greyscale pictures and one contains all colorful segmented pictures. We took only the potato leaves dataset from the entire dataset. There were 2152 pictures of 3 classes. [4]

Dataset Process and Augmentation:

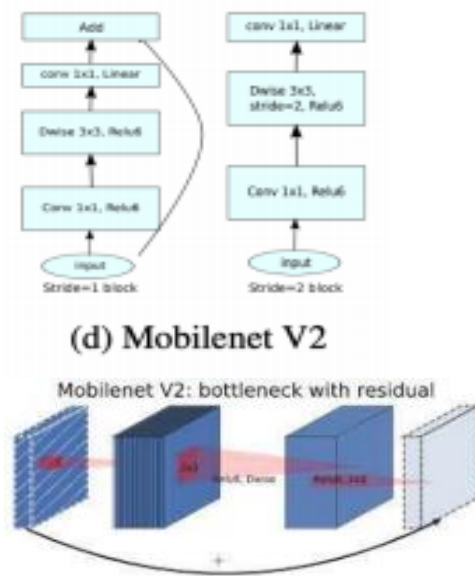
Dataset Process is an important task in training machine learning models. A dataset can have several issues like corrupt files, garbage values, impossible data combinations, and missing values. Dataset processing can take a short to a long time depending on the present condition of data and how we need data for the model.

At first we visualized the entire dataset and found do corrupt or irrelevant data .Then we renamed all the data with their parent folder names which gave them proper labels. In the case of the Potato dataset we did not need to downsize or resize the resolution as all the images had 256x256 pixel resolution. Hero labelling and resizing the data helped us to organize and visualize the data is a more efficient way when acceded. We split the dataset into 80:20 ratio for training and validation purposes.

Augmentation is generating more various of existing data by applying shifting, rotation, flipping, cropping, adding noise etc. We applied some basic data augmentation techniques following the Keras image augmentation documentation. [5]

Model:

For our project we chose to work with the MobileNetV2 model. MobileNetV2 is based on depth wise convolution and inverted residual blocks hence good for using in mobile devices. Convolution is a resource hungry task but in this particular model the main convolution is separated in multiple convolutions hence called the depth wise convolution. Also, it is based on inverted residual blocks where the blocks are connected in bottleneck layers, instead of expansion layers. What this does is it keeps the data flow among the model compressed but expands the data inside the block and runs convolutions. [6]



Model Training and Result:

Model training is the part where we finally train our model with the dataset we have chosen. During this process machine learning algorithms help the model to learn from the data provided. The model gets trained and tested simultaneously at the same time.

We used Categorical Cross Entropy to calculate the loss .It is used to measure loss in multiclass detection problems. As

optimizer for the model, we used Adam optimizer which can work as a replacement algorithm for SGD algorithm. The optimizer acts as the core algorithm which helps the model to learn. In the very last layer of our model we implemented a SoftMax function. This function is used for multiclass classification problems and it returns probabilities of each class.

We applied Transfer Learning in our training process. In transfer learning the model is already pre trained with similar kinds of data. Our model is pre-trained with imageNet dataset, hence it is good for detecting image. We used 224x 224 sized images and the number of images per batch is 64. The data is splitted into 80:20 ratio for trained and validation. We also kept source data for testing the application which were completely unseen by the model. The training Result is:

```
Epoch 1/10 [-----] 306s 15s/step - loss: 1.0853 - accuracy: 0.6647 - val_loss: 0.2324 - val_accuracy: 0.9273
Epoch 2/10 [-----] 100s 4s/step - loss: 0.1508 - accuracy: 0.9400 - val_loss: 0.1042 - val_accuracy: 0.9624
Epoch 3/10 [-----] 100s 4s/step - loss: 0.1002 - accuracy: 0.9633 - val_loss: 0.0818 - val_accuracy: 0.9694
Epoch 4/10 [-----] 99s 4s/step - loss: 0.0637 - accuracy: 0.9767 - val_loss: 0.0908 - val_accuracy: 0.9694
Epoch 5/10 [-----] 100s 4s/step - loss: 0.0700 - accuracy: 0.9767 - val_loss: 0.1163 - val_accuracy: 0.9624
Epoch 6/10 [-----] 100s 4s/step - loss: 0.0663 - accuracy: 0.9793 - val_loss: 0.0911 - val_accuracy: 0.9647
Epoch 7/10 [-----] 100s 4s/step - loss: 0.0673 - accuracy: 0.9633 - val_loss: 0.0813 - val_accuracy: 0.9741
Epoch 8/10 [-----] 99s 4s/step - loss: 0.0455 - accuracy: 0.9833 - val_loss: 0.0843 - val_accuracy: 0.9671
Epoch 9/10 [-----] 99s 4s/step - loss: 0.0508 - accuracy: 0.9828 - val_loss: 0.0908 - val_accuracy: 0.9718
Epoch 10/10 [-----] 99s 4s/step - loss: 0.0505 - accuracy: 0.9773 - val_loss: 0.0804 - val_accuracy: 0.9812
```

The training accuracy we found is 0.9773 and validation accuracy is 0.9812.

Android Application:

With the help of tflite converter we converted our model into a tflite format model. The converter provided us a model tflite and a labels txt file.

Then we designed and implemented our application with basic UI and functionally. We used Flutter as our android development framework and Android Studio as IDE. During the development process we tested everything on a virtual machine replication of the Pixel 2 device. Our application has two features, it can either capture a photo using the camera or it will access the file manager or gallery. Once the picture is provided the application returns the prediction result in a snap. The application automatically stores the prediction results in

its cache so that if anyone provides it the exact same image file again the application will directly return the result from cache, instead of running the prediction again.

Minimum Application Requirement:

Android 5, 3GB Ram, 80mh storage, camera and storage permission.

Software & Technology Requirement:

TensorFlow & Keras:

TensorFlow and Keras are python frameworks for developing machine learning and deep learning related tasks. These contain all the necessary algorithms and features we need to build, train, process machine learning and deep learning models.

Google Colab:

Google Collaboratory or Colab is a cloud-based platform for machine learning related works. Colab provides us all the necessary modules, software and hardware we need for training models. Though Colab can be slower compared to local machines, it provides all the tools we need.

Android Application: Flutter:

Flutter is a popular app development framework made by Google. Flutter is made for rapid UI based application development. Instead of using Java, Flutter uses a programming language called Dart. Flutter can be integrated into android studio easily to develop android applications.

Conclusion & Future Works

Though we tried to build a rice disease detection system, our main goal was to implement the idea that we can detect plant diseases with our phones without accessing

the internet. Throughout the development of this project, we faced various challenges and learned various workarounds.

We mainly struggled with our dataset. The development of the model and application was not that challenging but working with data was. We faced various performance issues due to our first dataset and we tried various techniques to fix that issue. But in the end, we had to move on with another dataset and we concluded that the previous dataset was not made properly. So in conclusion we can say that the dataset can be the most important thing in AI development. If the dataset has issues, no matter how many sophisticated algorithms we use we won't get proper results.

We have successfully demonstrated that we can implement advanced AI systems in our phones to solve problems like plant disease detection. With the help of modern mobile devices and AI technology we can solve many more problems in a much more efficient way. With the advancements of technology, the computers on our palms have now much more power than ever. We can use this power with AI to make our lives easier.

But we need properly maintained public datasets. We understand that making a dataset can be very costly but for the sake of research and development, researchers should try their best to make the dataset as much as public as possible.

With the help of AI, we can shape our future. We already can see and experience various benefits of artificial intelligence. We hope that people will develop more sophisticated AI systems for the betterment of us and everything around us, no one should use AI for the things which cause trouble.

References

- [S. M. a. S. A. M. J. Hasan, "Rice Disease 1 Identification and Classification by Integrating] Support Vector Machine With DeepConvolutional Neural Network," Academia.edu, [Online]. Available: https://www.academia.edu/43974637/Rice_Disease_Identification_and_Classification_by_Integrating_Support_Vector_Machine_With_Deep_Convolutional_Neural_Network. [Accessed 21 May 2021].
- [V. Nagrale, "Rice Crop Disease Identification 2 and Classifier," International Journal of] Computer Sciences and Engineering," [Online]. Available: https://www.academia.edu/42123981/Rice_Crop_Disease_Identification_and_Classifier. [Accessed 23 MAY 2021].
- [J. Pereira, "Rice Disease Classification using 3 Convolutional Neural Networks," Academia.edu, [Online]. Available: https://www.academia.edu/41821127/Rice_Disease_Classification_using_Convolutional_Neural_Networks. [Accessed 24 May 2021].
- [A. Ali, "PlantVillage Dataset," Kaggle, 01 4 September 2019. [Online]. Available:] <https://www.kaggle.com/abdallahalidev/plantvillage-dataset>.
- [K. Team, "Keras documentation: Image data 5 preprocessing," Keras, [Online]. Available:] <https://keras.io/api/preprocessing/image/>.
- [M. Sandler, "MobileNetV2: Inverted Residuals 6 and Linear Bottlenecks," 13 January 2018.] [Online]. Available: <https://arxiv.org/abs/1801.04381>.