DATA ANALYTICS

PROJECT TITLE

LEAF DISEASE CLASSIFICATION



ABSTRACT

Disease detection in plants plays a very important role in the field of agriculture. This project aims to provide an image-based automatic inspection interface. Identification of the plant diseases is the key to preventing the losses in the yield and quantity of the agricultural product. It involves the use of self - designed image processing and deep learning techniques. It will categorize plant leaves as healthy or infected. The studies of the plant diseases mean the studies of visually observable patterns seen on the plant. Health monitoring and disease detection on plant is very critical for sustainable agriculture. It is very difficult to monitor the plant diseases manually. It requires tremendous amount of work, expertize in the plant diseases, and also require the excessive processing time. Detection of plant disease through some automatic technique is beneficial as it reduces a large work of monitoring in big farms of crops, and at very early stage itself it detects the symptoms of diseases i.e. when they appear on plant leaves. If automatic detection technique is used it will take less efforts, less time and become more accurate. In plants, some general diseases seen are brown and yellow spots, early and late scorch, and others are fungal, viral and bacterial diseases.

PROBLEM STATEMENT

To detect the leaf disease and classifying accordingly and to monitor the plant diseases manually it is very difficult. It requires tremendous amount of work, expertize in the plant diseases, and also require the excessive processing time.so we have proposed this.

DATASET:

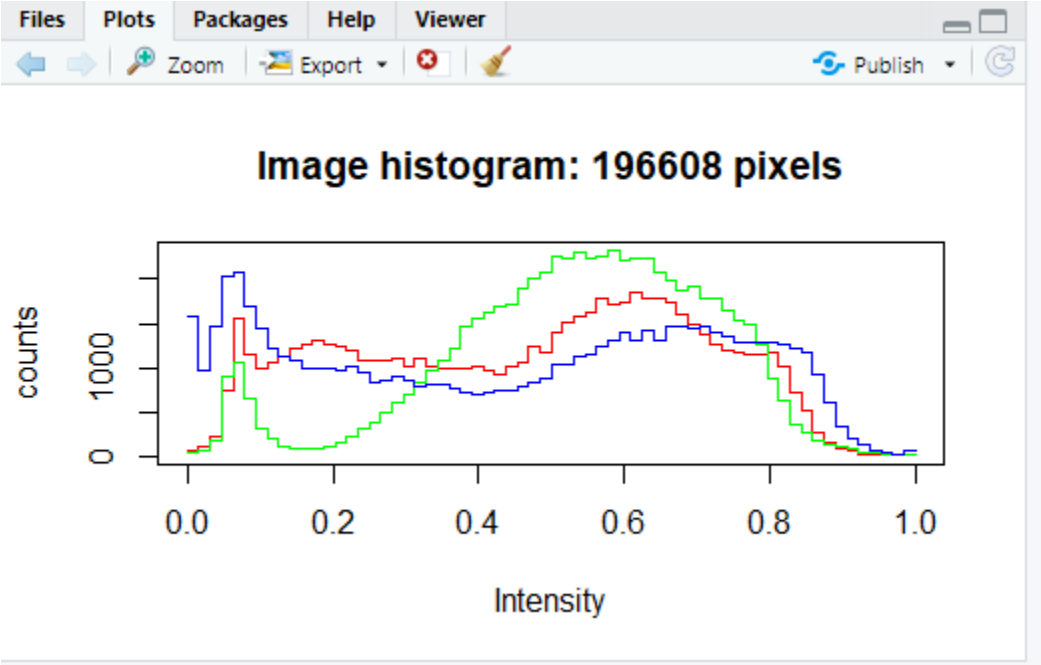
Here we use both infected and not infected images like fungi, bacteria, normal ,virus.

https://www.kaggle.com/sizlingdhairya1/leaf-disease

METHODOLOGY

The leaf disease dataset is taken of different types and some process have been done to train the images. Let's see, So for these we are using libraries keras which have deep learning, EBImage for image processing.

- The histogram of particular image. The intensity is between 0 and 1, if not we should make image to lie between 0 and 1.



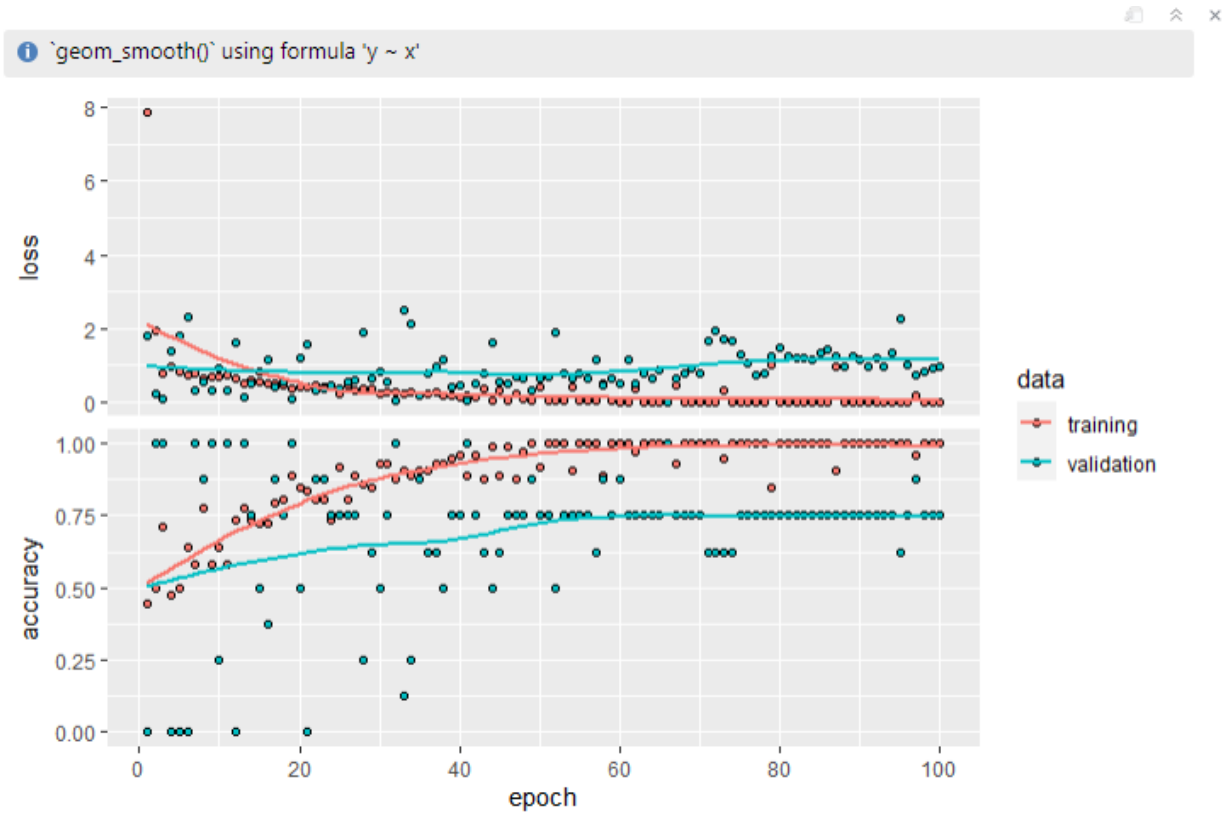
- Resize: Sometimes there will be unequal sizes, so we are going to resize the image. Now when we see the structure , after resize

```
List of 100
 $ :Formal class 'Image' [package "EBImage"] with 2 slots
  .. ..@ .Data    : num [1:28, 1:28, 1:3] 0.694 0.69 0.662 0.712 0.591 ...
  .. ..@ colormode: int 2
```

- Reshape: So now we have to change the single dimension for training purpose

```
List of 100
 $ : num [1:28, 1:28, 1:3] 0.694 0.69 0.662 0.712 0.591 ...
```

- So now we are splitting the images into training and testing and also test for only 1 image ,whether it can correctly classify or not.
- We do one hot encoding to change into binary vectors.
- We built a multilayer perceptron model which had input layer, hidden layer, output layer.
- So after building the model ,compile and fit the training set to model, first we gave 50 epochs and 16 batch size ,which didn't give satisfactory results it gave around 79per accuracy, after changing into 100 epochs and 6 batch size it gave 97per accuracy.



Plot gives Loss - prediction error and accuracy how well the model is able to predict.

- Evaluate and prediction on training the model,0-fungi,1-normal/healthy leaf.

```
##      loss  accuracy
## 0.1237612 0.9750000
```

```
##          Actual
## predicted  0  1
##         0 40  2
##         1  0 38
```

- The accuracy and prediction results. model is able to predict/classify well that diagonal 40 images true negative (fungi) and38 images as true positive (healthy/normal) and remaining is misclassified 2img as false negative(fungi) .The accuracy is also good that means 97% model is able to predict well..

EVALUATING ON TESTING DATA IMAGES

- Now we are testing the model with 20 images, 0-fungi,1-normal/healthy leaf.

```
##      loss  accuracy
## 0.7579941 0.8500000
```

```
##          actual
## predicted 0 1
##         0 8 1
##         1 2 9
```

So here diagonal images has been correctly predicted as 8imgs as true negative(fungi) and 9imgs as true positive(healthy) so model is able to predict well on testing images also.

- To test for only 1 image that is I have selected fungi image(binary(0)).

```
##     loss accuracy
## 4.136773 1.000000
```

```
##          actual
## predicted 0
##         0 1
```

So the image has correctly predicted as true negative(fungi).

- so now we took the dataset contains c-(fungi),n-(normal/healthy),p -(bacteria),v-(virus) images which contains overall 200 images.
- Now I'm testing the model with 40 images that is 10 images from each #0-fungi,1-normal/healthy,2-bacteria,3-virus leaf.

```
##      loss  accuracy
## 0.6111615 0.8750000
```

```
##          actual
## predicted  0  1  2  3
##         0  8  0  2  0
##         1  2 10  0  1
##         2  0  0  8  0
##         3  0  0  0  9
```

So here 35 images has been correctly predicted so model is able to predict well on testing images also diagonal 8imgs are true fungi, 10imgs true healthy, 8imgs true bacteria, 9imgs true virus, rest are all misclassified as 2imgs false fungi, 3imgs false healthy.

- To test for only 1 image from each

```
##         loss     accuracy
## 2.765522e-05 1.000000e+00
```

```
##          actual
## predicted 0 1 2 3
##         0 1 0 0 0
##         1 0 1 0 0
##         2 0 0 1 0
##         3 0 0 0 1
```

So final the image has correctly predicted as true fungi, healthy, bacteria, virus.

So the model is able to predict and classify the images well, so accordingly to the results they can analyze and take measures to prevent and correct the disease.

REFERENCES

[1] https://tensorflow.rstudio.com/tutorials/beginners/basicml/tutorial_basic_text_classification/

[2] https://blogs.rstudio.com/ai/posts/2017-12-14-image-classification-on-small-datasets/

[3] https://tensorflow.rstudio.com/tutorials/beginners/basicml/tutorial_basic_classification/