Blindness Detection

The data was downloaded from Kaggle. The data when extracted gave a warning that the file is corrupt. There were images missing in the folder. I made a new dataframe when contained only the images available and only for those images the labels were available.

Diabetic eye disease comprises a group of eye conditions that affect people with diabetes. These conditions include diabetic retinopathy, diabetic macular edema (DME), cataract, and glaucoma. All forms of diabetic eye disease have the potential to cause severe vision loss and blindness. Diabetic retinopathy involves changes to retinal blood vessels that can cause them to bleed or leak fluid, distorting vision. Diabetic retinopathy is the most common cause of vision loss among people with diabetes and a leading cause of blindness among working-age adults. DME is a consequence of diabetic retinopathy that causes swelling in the area of the retina called the macula. Because diabetic retinopathy often goes unnoticed until vision loss occurs.

Diabetic retinopathy affects blood vessels in the light-sensitive tissue called the retina that lines the back of the eye. It is the most common cause of vision loss among people with diabetes and the leading cause of vision impairment and blindness among working-age adults. For building a machine learning model, I have downloaded the data from an open source website Kaggle.com.

The model building would help speed up disease detection and this would help to prevent lifelong blindness.

There are two types of files, one being the CSV file for storing the labels of the images and another folder containing the images. The data size is of 929 images and the images are labeled into 5 categories.

1) Class 0 : No Diabetic retinopathy

2) Class 1 : Mild

3) Class 2 : Moderate

4) Class 3 : Moderate

5) Class 4 : Severe

In [2]:
```
# importing necessary packages
import numpy as np
import pandas as pd
```

In [7]:
```
#link for the image file
image_file="/Users/rohitbohra/Documents/aptos2019-blindness-detection/
```

In [8]:
```
#assigning a variable to the label of the image
file_name=pd.read_csv("/Users/rohitbohra/Documents/aptos2019-blindness
```

In [9]:
```
# count of the labeled variable and the first 5 rows of data
print(file_name.count())
print(file_name.head())
```

```
id_code      3662
diagnosis    3662
dtype: int64
        id_code  diagnosis
0  000c1434d8d7          2
1  001639a390f0          4
2  0024cdab0c1e          1
3  002c21358ce6          0
4  005b95c28852          0
```

In [10]:
```
# adding .png to every image name in the csv file
file_name["id_code"]=file_name["id_code"].apply(lambda x:x+".png")
```

In [11]:
```
# first 5 rows of data. The image name contains .png to every image na
file_name.head()
```

Out[11]:

	id_code	diagnosis
0	000c1434d8d7.png	2
1	001639a390f0.png	4
2	0024cdab0c1e.png	1
3	002c21358ce6.png	0
4	005b95c28852.png	0

In [12]:
```
# checking for the count of images in the image folder
import os

path, dirs, files = next(os.walk("/Users/rohitbohra/Documents/aptos201
file_count = len(files)
print(file_count)
```

```
929
```

In [13]:
```
# the name of the files in the image folder
for i in files:
    print(i)
```

```
4a5a6efc0bef.png
*
a8582e346df0.png
*
b09101adb478.png
*
3b58b02c89ed.png
*
d74ccc796517.png
*
a821b6ecef33.png
*
90a786abe58e.png
*
35d6c4c50072.png
*
3710ff45299c.png
*
67f5d89da548.png
```

In [56]:
```
# The file name is stored into a dataframe
image_name_df = pd.DataFrame({'id_code':files})
```

In [58]:
```
# merging the dataframe of image_name_df and file_name with id_code co
# assigning it to a vriable
for i in files:
    for j in file_name['id_code']:
        df = file_name.merge(image_name_df, on=['id_code'])
```

In [59]:
```
# dataframe
df
```

Out[59]:

	id_code	diagnosis
0	000c1434d8d7.png	2
1	00a8624548a9.png	2
2	00cb6555d108.png	1
3	0104b032c141.png	3
4	0124dffecf29.png	1
5	014508ccb9cb.png	0
6	01b3aed3ed4c.png	1
7	01c7808d901d.png	2
8	0212dd31f623.png	0
9	02dda30d3acf.png	4
10	0304bedad8fe.png	0

In [60]:
```
# the dataframe is now reduced from 3662 to 929.
df.count()
```

Out[60]:
```
id_code      929
diagnosis    929
dtype: int64
```

In [61]:
```
# this dataframe is now saved into label.csv which was created in adva
df.to_csv(r'/Users/rohitbohra/Documents/aptos2019-blindness-detection/
```

In [62]:
```
# checking if the dataframe saving is correctly.
label=pd.read_csv("/Users/rohitbohra/Documents/aptos2019-blindness-det
```

In [74]:
```
# first 5 rows in the dataframe.
# the dataframe is now correctly saved.
label.head()
```

Out[74]:

	Unnamed: 0	id_code	diagnosis
0	0	000c1434d8d7.png	2
1	1	00a8624548a9.png	2
2	2	00cb6555d108.png	1
3	3	0104b032c141.png	3
4	4	0124dffecf29.png	1