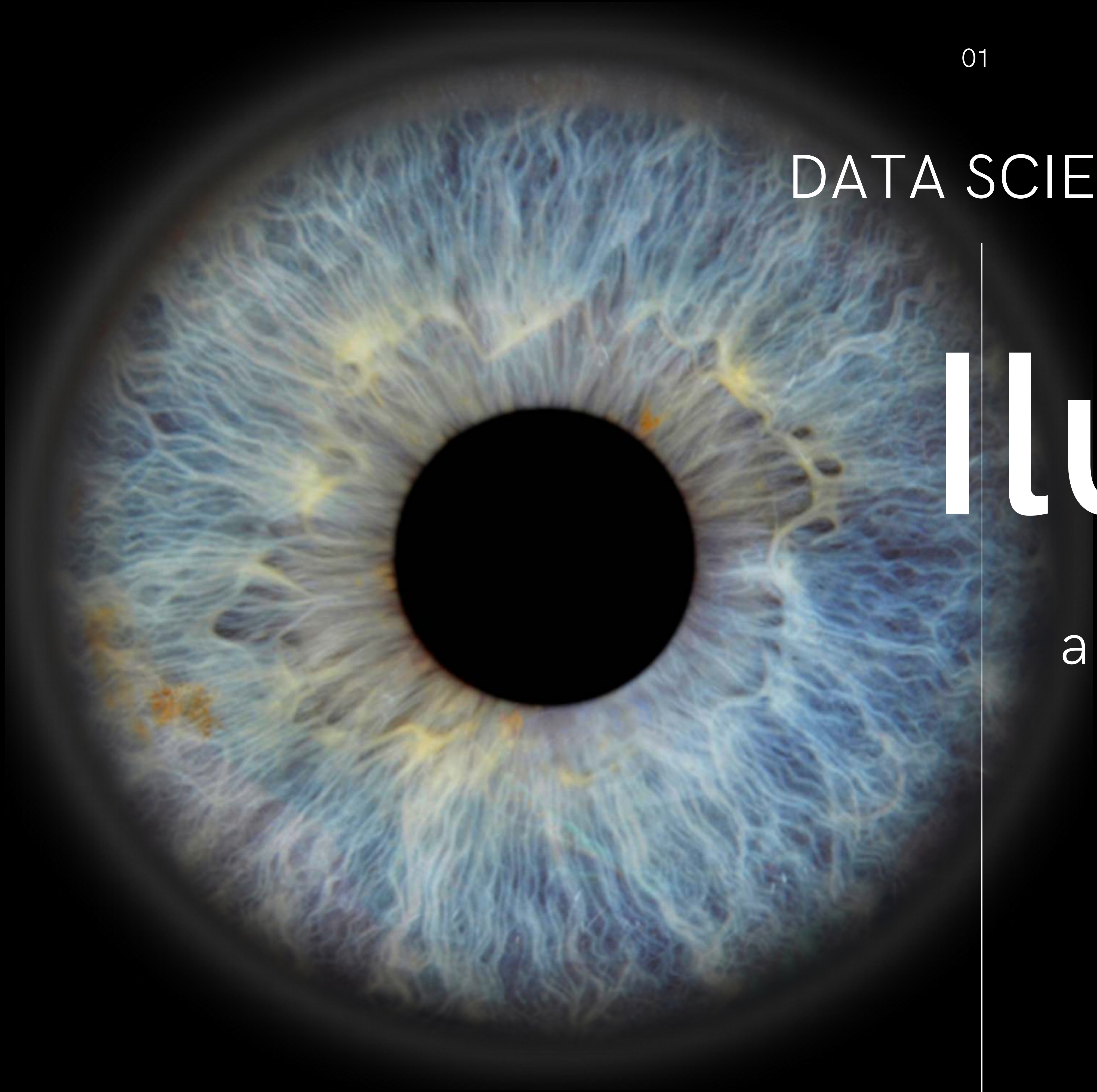


DATA SCIENCE CAPTSTONE PROJECT

Iluminado

a highly-accessible, low-cost initial eye disease risk diagnosis

By Cathy Kam

This was me
8 months ago

Problem Statement:

According to the World Health Organization:

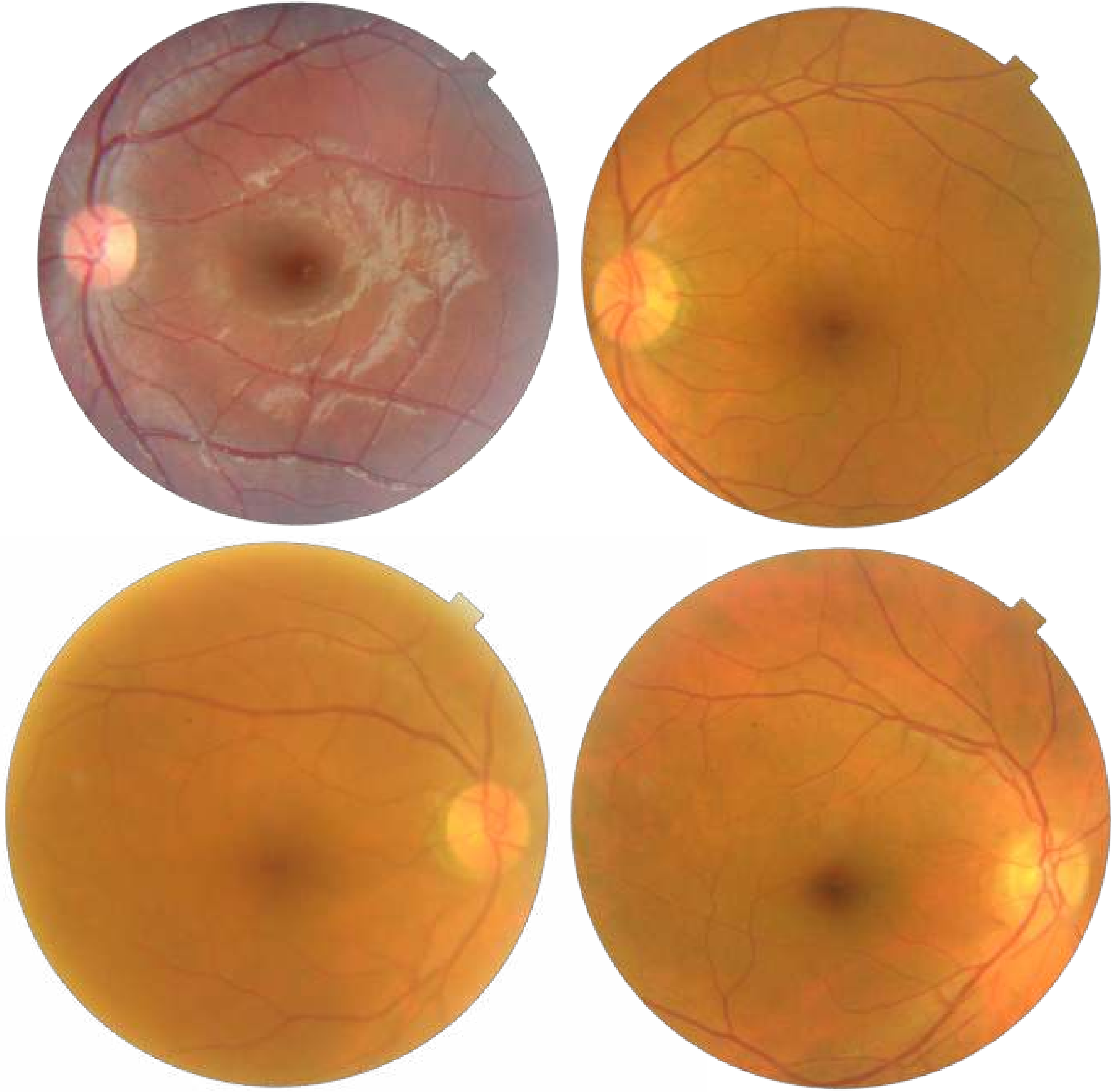
2.2 billion

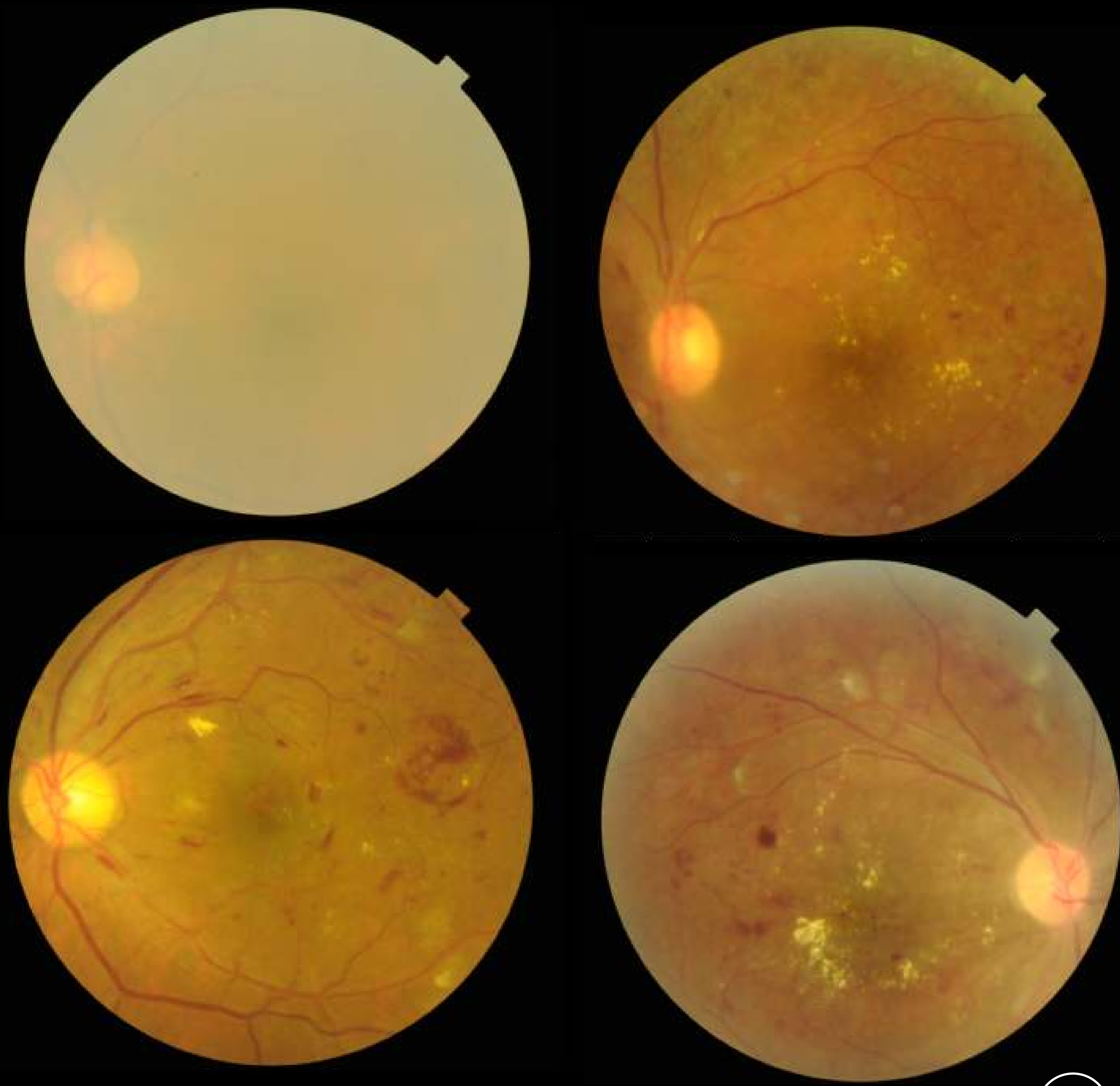
vision-impaired people worldwide

1 billion

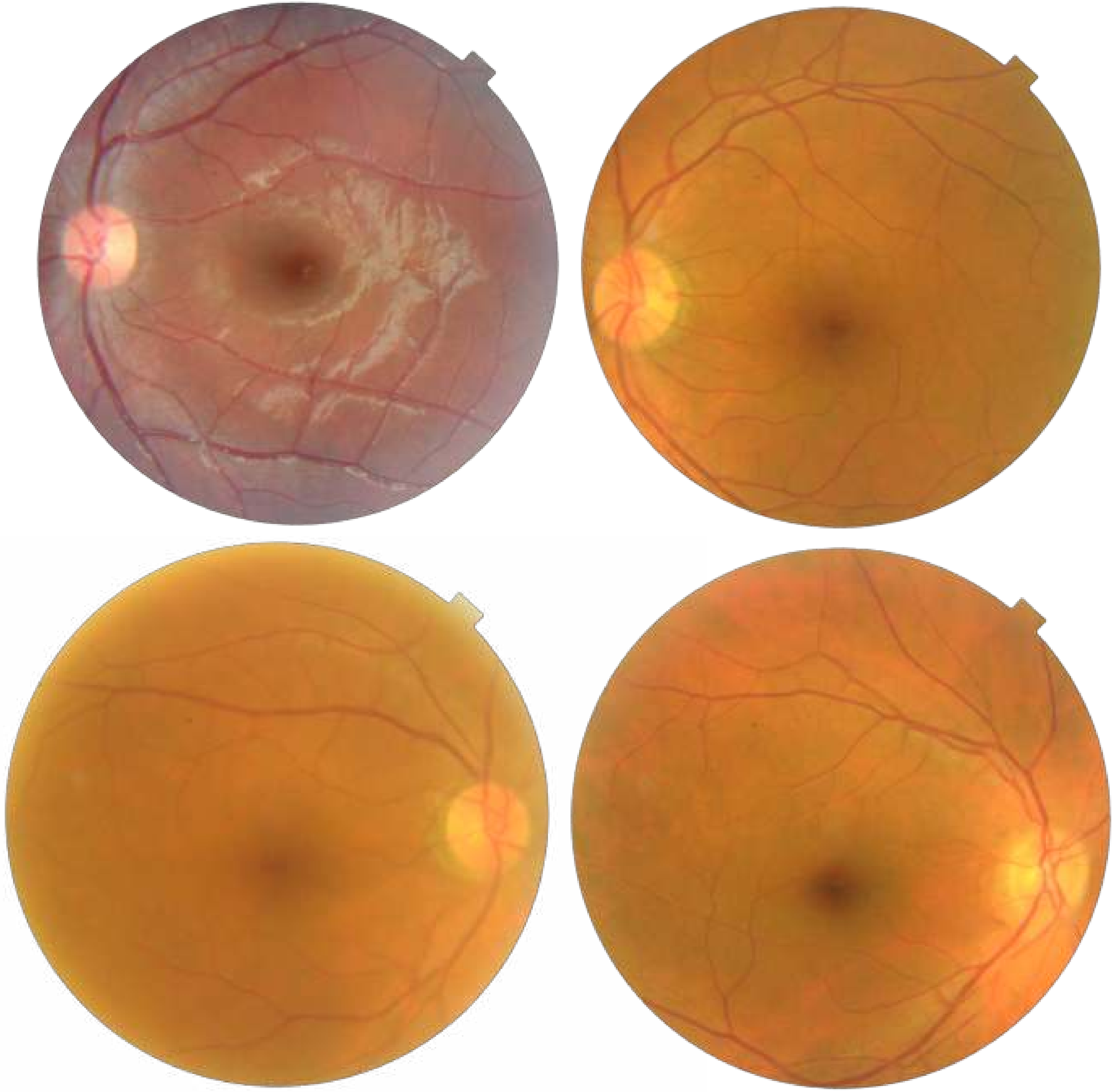
could have been prevented or are still untreated

Which side represents the healthy eyeballs?

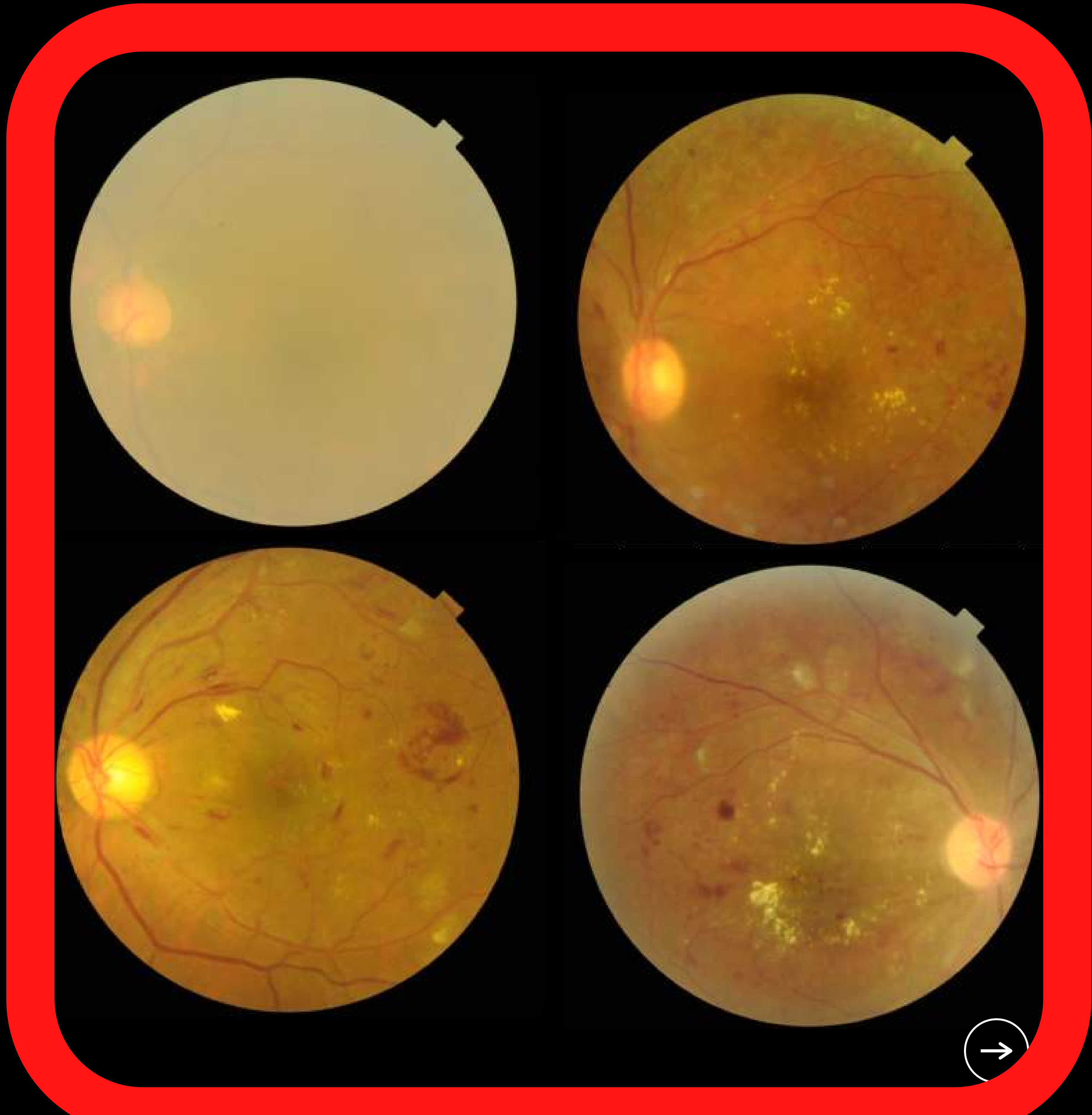




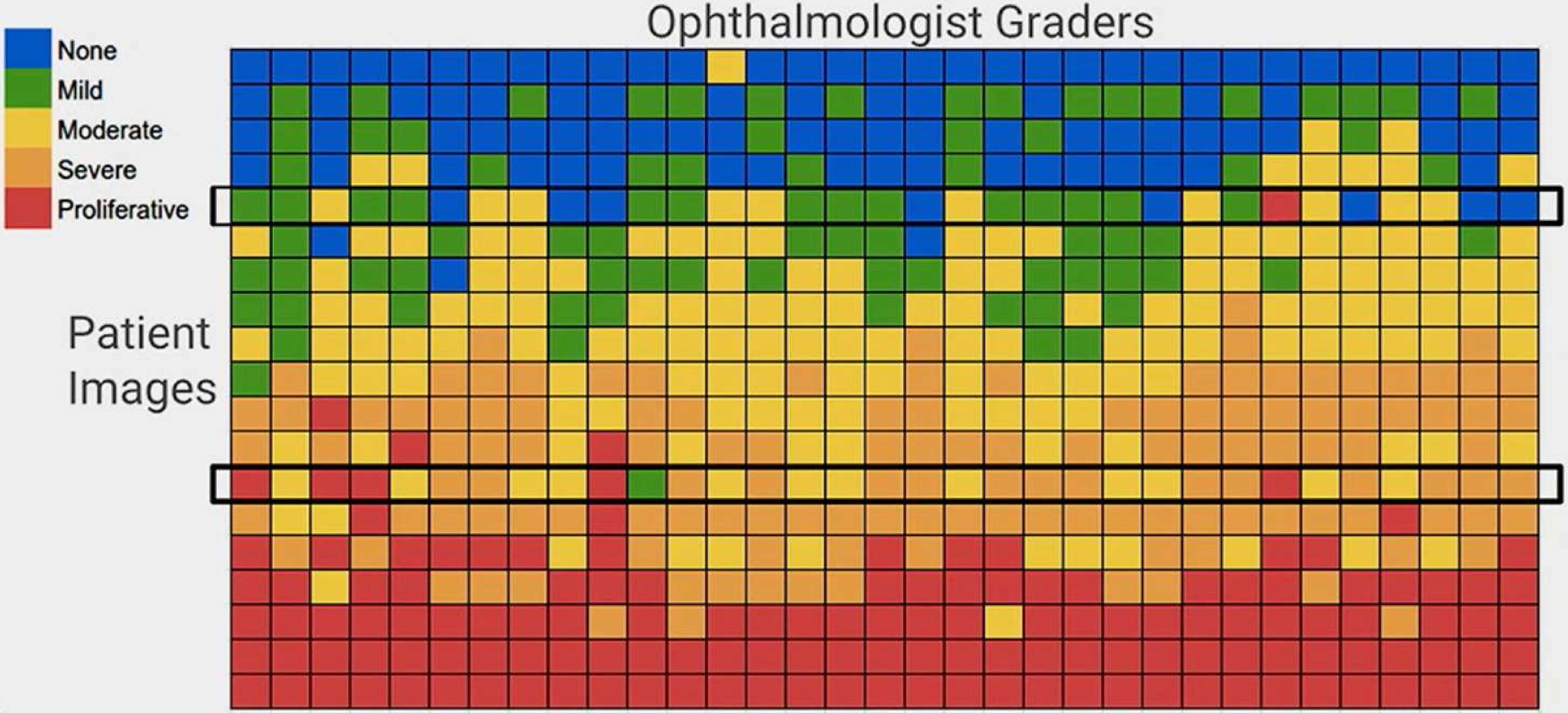
No_diesease



Yes_disease



Even when available, ophthalmologists are inconsistent



Consistency: intragrader ~65%, intergrader ~60%

The Aim of Iluminado:

To perform an initial eye disease risk diagnosis using convolutional network at a low-cost that is readily available online.

Results can then be sent to the nearest family Ophthalmologist to determine whether an intervention is needed immediately based on risk prediction through the model.

Where's the data?

Center for Precision Health, School of Biomedical Informatics, University of Texas Health Science Center

Retinal Fundus Multi-Disease Image Dataset (RFMiD): A Dataset for Multi-Disease Detection Research

The RFMiD is a new publicly available retinal images dataset consisting of 3200 images.

https://www.mdpi.com/2306-5729/6/2/14

Where does the image come from?

This image was captured using a tool known as the fundus camera — a specialized low power microscope attached to a flash-enabled camera.

Model	Hardware	FOV	Resolution (in Pixels)	Number of Images in Dataset
TOPCON 3D OCT-2000	Nikon D7000 digital camera	45°	2144 × 1424	2427
Kowa VX-10α	Nikon D70s digital camera	50°	4288 × 2848	467
TOPCON TRC-NW300	Integrated digital CCD camera	45°	2048 × 1536	306

2. Data Description

The RFMiD is a new publicly available retinal images dataset consisting of 3200 images along with the expert annotations divided into two categories, as follows:

- Screening of retinal images into normal and abnormal (comprising of 45 different types of diseases/pathologies) categories.
- Classification of retinal images into 45 different categories.

The dataset is split into 3 subsets: training 60% (1920 images), evaluation 20% (640 images), and test 20% (640 images) sets. The disease wise stratification on average in training, evaluation and test set is 60 ± 7 %, 20 ± 7%, and 20 ± 5%, respectively. The main motto of this dataset is to provide multiple diseases that appear in routine clinical practice. The labels are provided in three CSV files $RFMiD\_Training\_Labels.CSV$, $RFMiD\_Validation\_Labels.CSV$, and $RFMiD\_Testing\_Labels.CSV$. The information available in the CSV file is illustrated in **Figure 1**, with each column explanation given as follows:

	A	B	C	D	E	F	G	H	I	J	AU
1	ID	Disease_Risk	DR	ARMD	MH	DN	MYA	BRVO	TSLN	....	CL
2	1	1	1	1	0	0	1	0	0		0
3	2	1	0	0	0	0	0	1	0		0
4	3	1	1	0	1	1	0	0	1		0
5	4	1	0	0	0	0	0	0	0		1
6	5	0	0	0	0	0	0	0	0		0

Figure 1. Sample CSV files.

Part 1

Data cleaning + Explanatory Data Analysis

Data Cleaning and Sorting

```
train = pd.read_csv('RFMiD_Training_Labels.csv')
train
```

	ID	Disease_Risk	DR	ARMD	MH	DN	MYA	BRVO	TSLN	ERM	...	CME	PTCR	CF	VH	MCA	VS	BRAO	PLQ	HPED	CL
0	1	1	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	2	1	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	3	1	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	4	1	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	5	1	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1915	1916	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1916	1917	1	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1917	1918	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1918	1919	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1919	1920	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

1920 rows × 47 columns

```
#sort out the ID that has disease risk == 1 and convert them into a list
```

```
yes_disease = list(train.loc[train['Disease_Risk'] == 1]['ID'].astype(str))
yes_disease[4]
```

'5'

```
no_disease = list(train.loc[train['Disease_Risk'] == 0]['ID'].astype(str))
no_disease[4]
```

'28'

Data Cleaning and Sorting

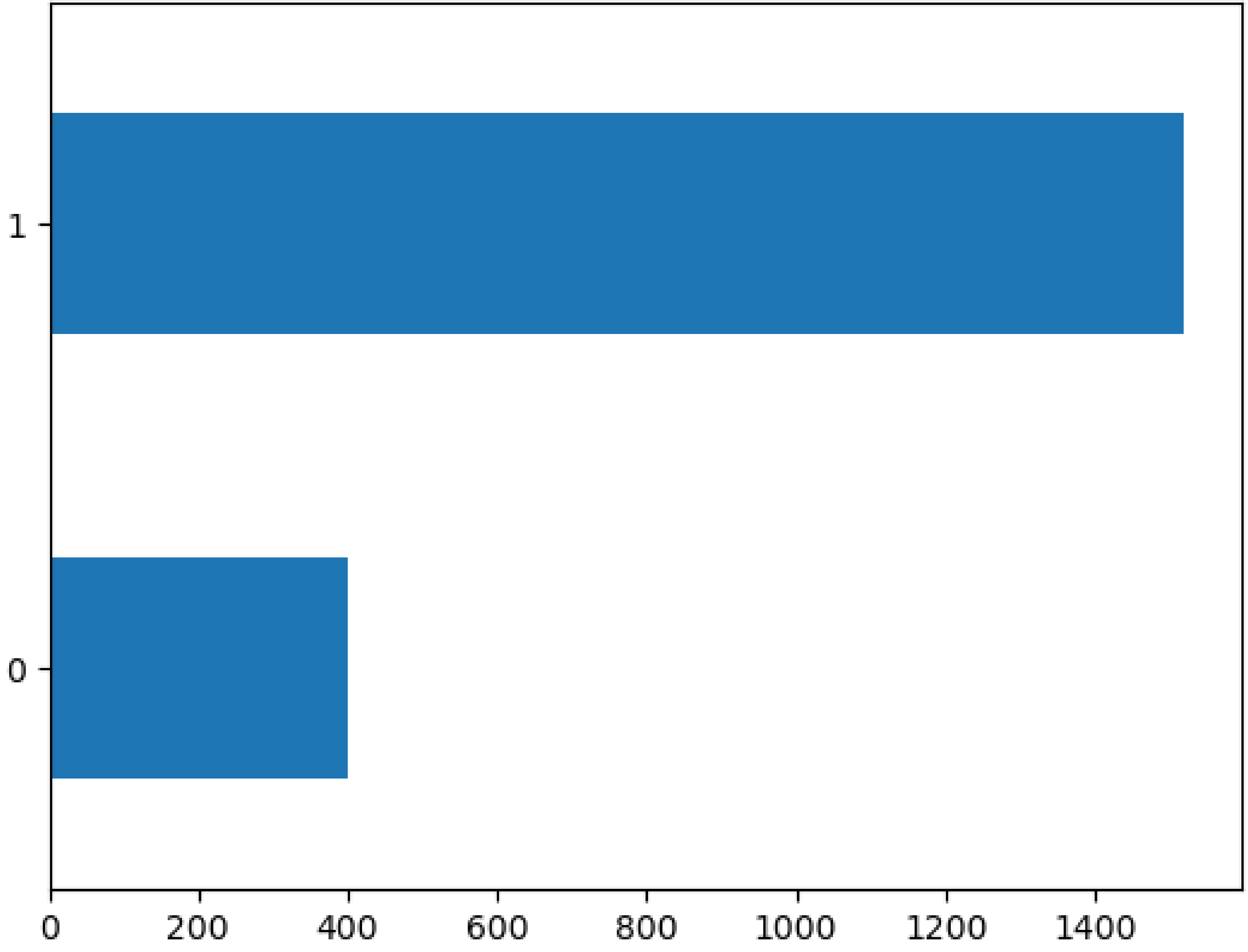
```python
#sorting the images into the Yes_disease pile

ims = os.listdir()

for im in ims:
    if im.replace('.png', '') in yes_disease:
        os.rename(im, f"Yes_disease/{im}")
```

```python
#sorting the images into the No_disease pile

for im in ims:
    if im.replace('.png', '') in no_disease:
        os.rename(im, f"No_disease/{im}")


# os.rename("1920.png", "Yes_disease/1920.png")
```

Pathologies and Their Distributions Within the Training and Evaluation Set of the RFMiD Dataset

Pathology Label	Pathology Description	Count in Training Set	Count in Evaluation Set
DR	Diabetic retinopathy	376	132
MH	Media haze	317	102
ODC	Optic disc cupping	282	72
TSLN	Tessellation	186	65
DN	Drusen	138	46
MYA	Myopia	101	34
ARMD	Age-related macular degeneration	100	38
BRVO	Branch retinal vein occlusion	73	23
ODP	Optic disc pallor	65	26
ODE	Optic disc edema	58	21
LS	Laser scars	47	17
RS	Retinitis	43	14
CSR	Central serous retinopathy	37	11
Other	Other	34	21
CRS	Chorioretinitis	32	11
CRVO	Central retinal vein occlusion	28	8
RPEC	Retinal pigment epithelium changes	22	6
AION	Anterior ischemic optic neuropathy	17	5
AH	Asteroid hyalosis	16	4
MS	Macular scars	15	5
EDN	Exudation	15	5
ERM	Epiretinal membrane	14	7
RT	Retinal traction detachment	14	6
PT	Parafoveal telangiectasia	11	2
MHL	Macular hole	11	3
TV	Tortuous vessels	6	2
RP	Retinitis pigmentosa	6	2
ST	Optociliary shunt	5	4

The overall distribution of yes_disease [1] and no_disease [0] images in the training set



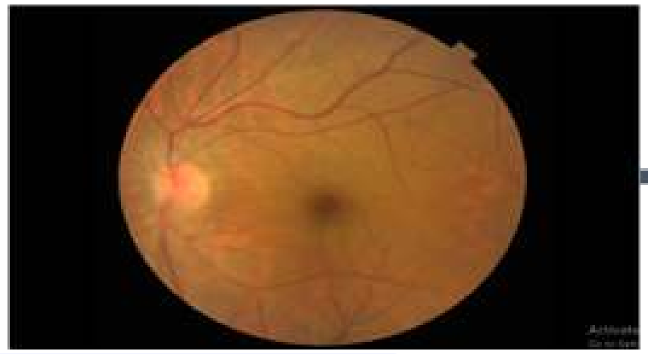
1: 1519

0: 401

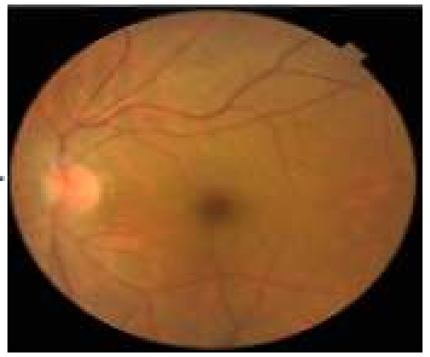
Part 2

Modeling for Binary Classification

Data Augmentation + Binary Classification



Pre-processing



```
ImageDataGenerator(rescale = 1/255., shear_range= 0.1, zoom_range= 0.2, horizontal_flip = True, vertical_flip = True)
```

```
top_layer = Sequential()
top_layer.add(Dense(100, activation = 'relu'))
top_layer.add(Dropout(0.2))
top_layer.add(Flatten())
top_layer.add(Dense(512,activation="relu"))
top_layer.add(Dense(1, activation = 'sigmoid'))
```

Disease Classes (1 = Yes_disease, 0 = No_disease)

Disease risk	0.5%
Disease risk	0.87%
Disease risk	0.95%
⋮	⋮

Success Metrics - Model Performance from 5 Different Base Convolution

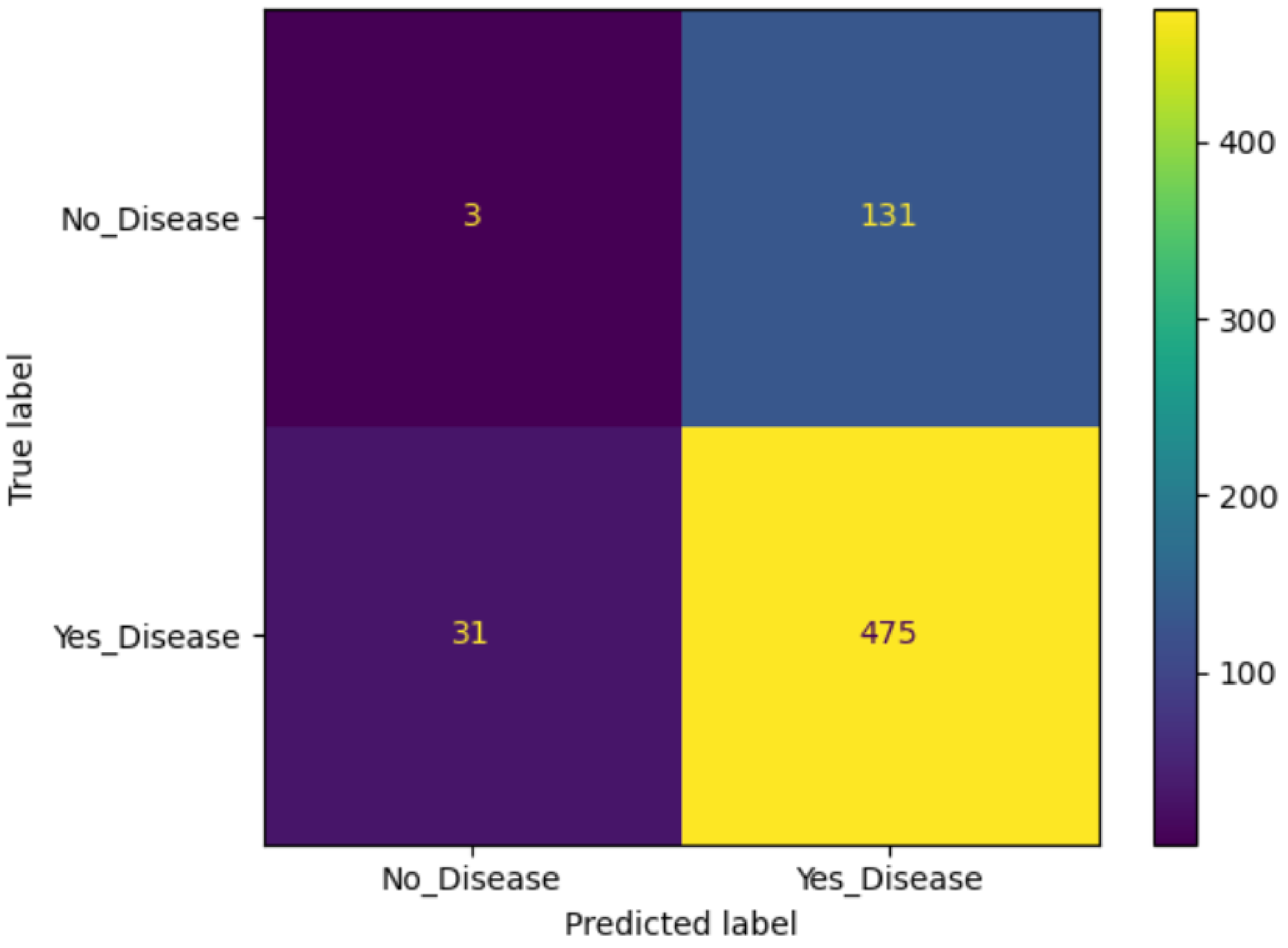
Architectures	Epochs	Training loss	Testing loss	Training accuracy	Testing accuracy	Training precision	Testing precision
InceptionV3	70	0.0123	2.1166	0.9969	0.7469	0.9980	0.7838
Xception	70	0.0025	2.2295	0.9995	0.7391	0.9993	0.7927
VGG16	30	0.1782	0.9857	0.9276	0.7375	0.9378	0.7975
MobileNetV2	30	0.0290	1.7801	0.9896	0.7437		
EfficientNetB5	20	0.5123	0.5205	0.7932	0.7844	0.7940	0.7911
SE-ResNeXt	20	0.5173	0.51	0.7911	0.7906	0.7911	0.7906

Baseline:

```
y_test.mean()
```

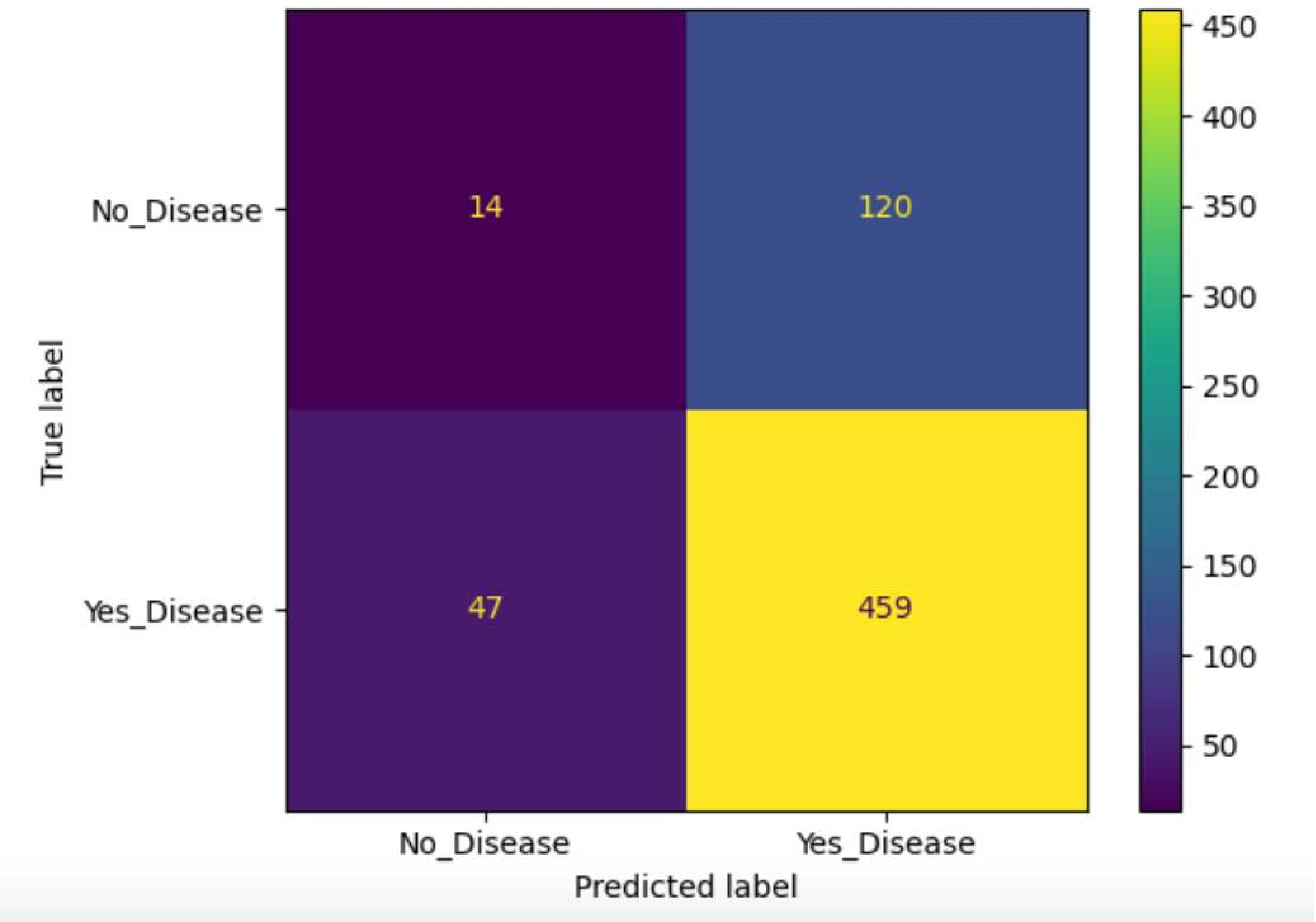
```
0.790625
```

InceptionV3



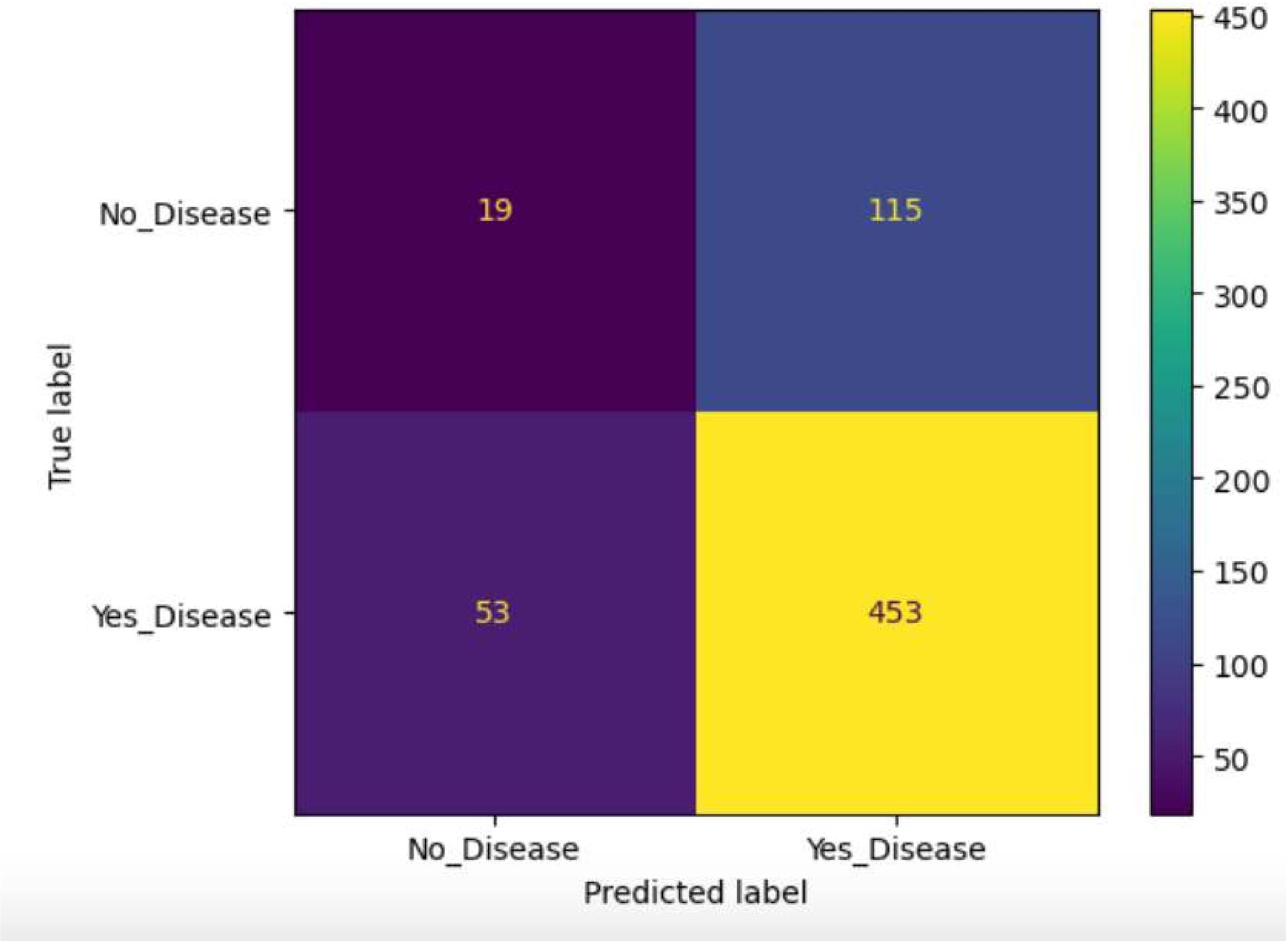
val_recall: 0.9387 - val_auc: 0.4908 - val_prc: 0.7873

Xception

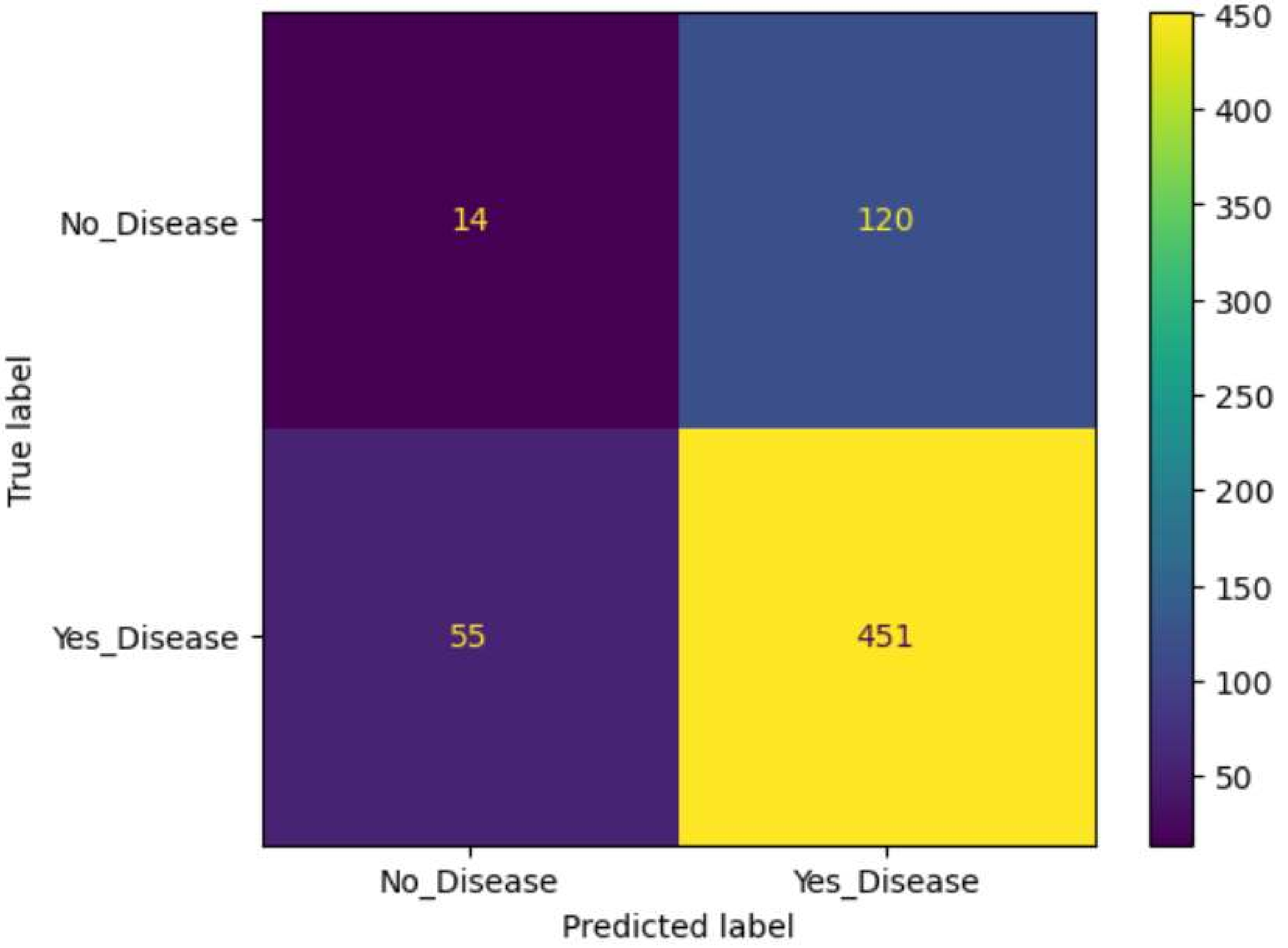


val_recall: 0.9071 - val_auc: 0.5234 - val_prc: 0.8003

VGG16

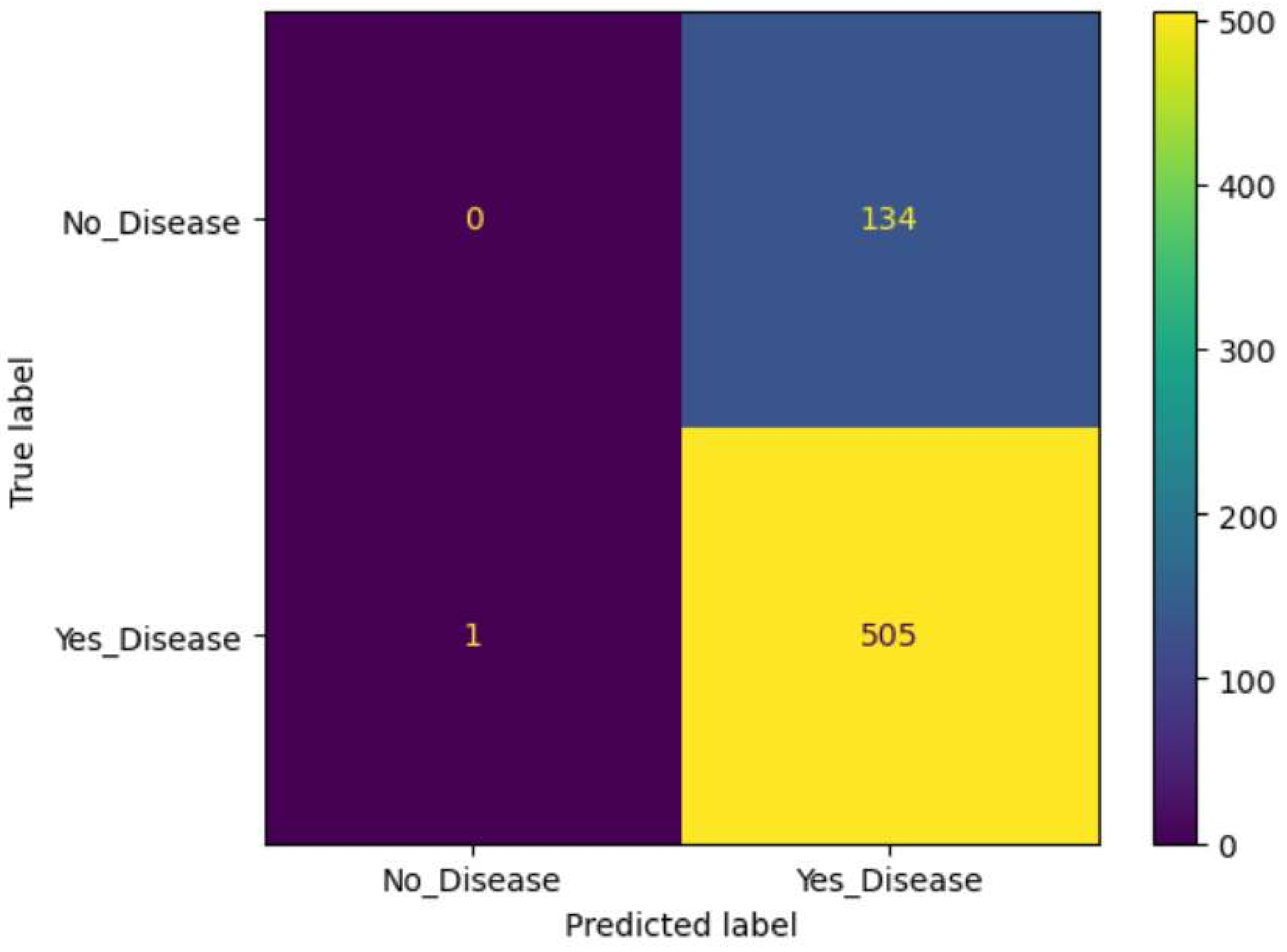


MobileNetV2



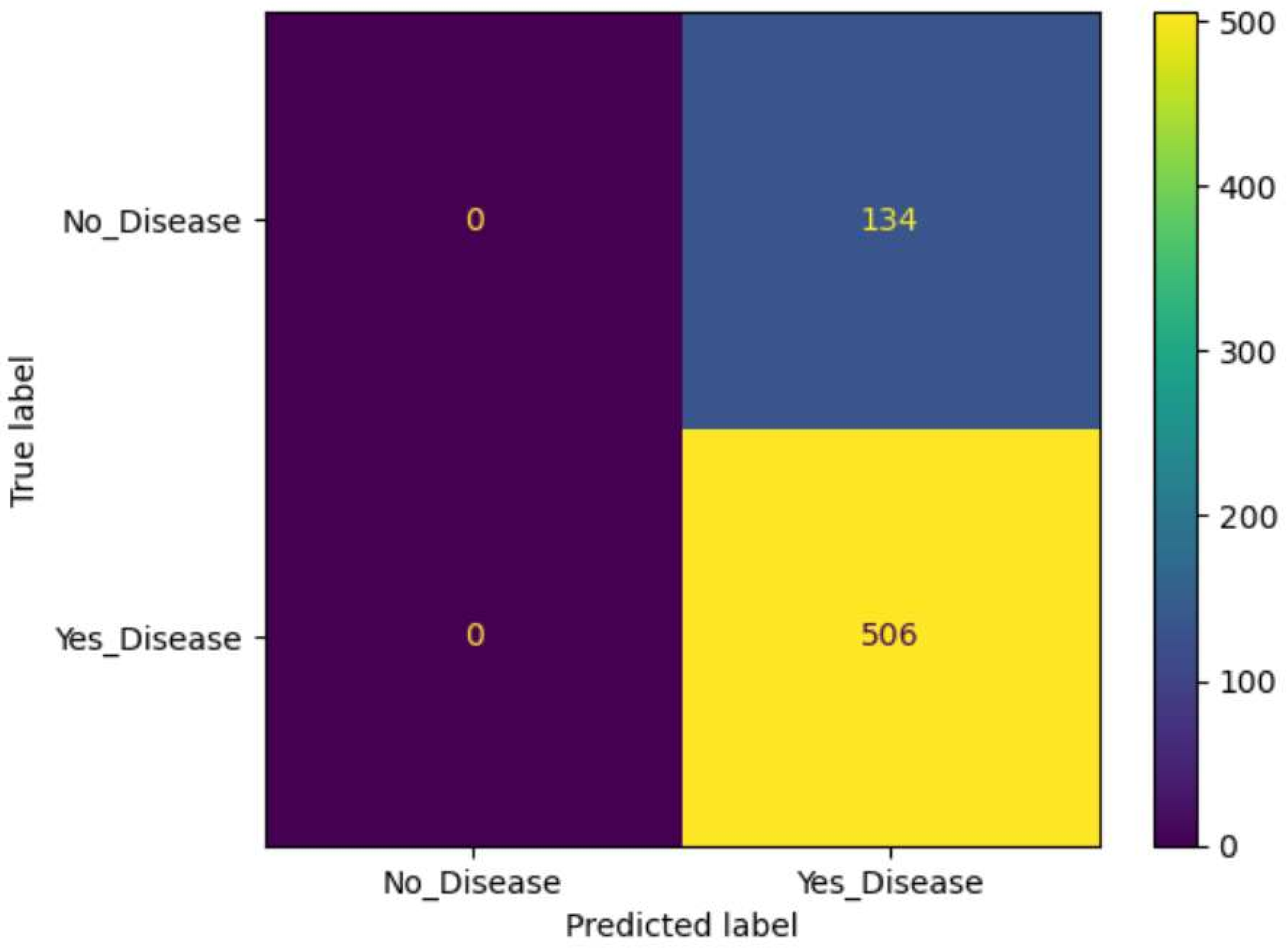
val_recall: 0.8953 - val_auc: 0.5339 - val_prc: 0.7990

EfficientNetB5



val_recall: 0.9980 - val_auc: 0.4556 - val_prc: 0.7687

SE-ResNeXt



val_recall: 1.0000 - val_auc: 0.5289 - val_prc: 0.8103

Part 3

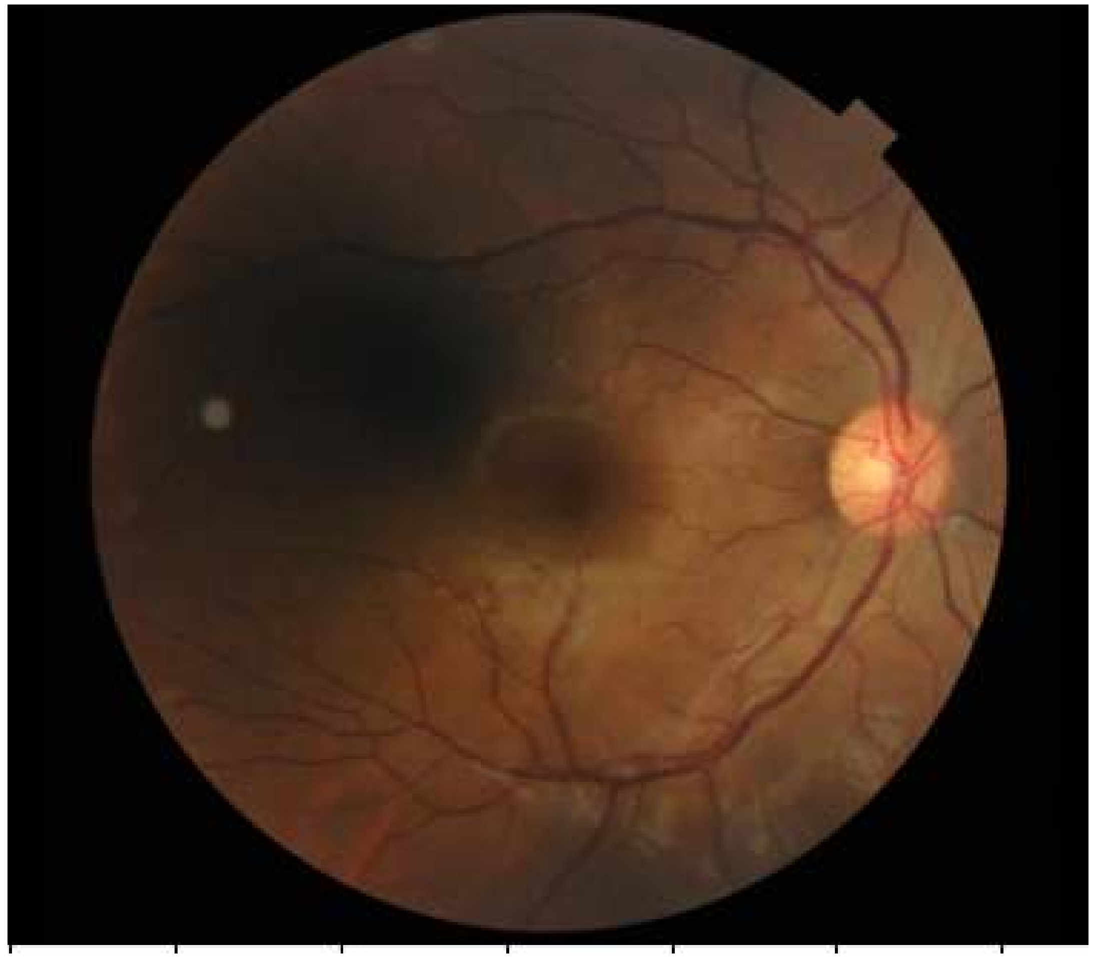
Modeling for Multi-Classification

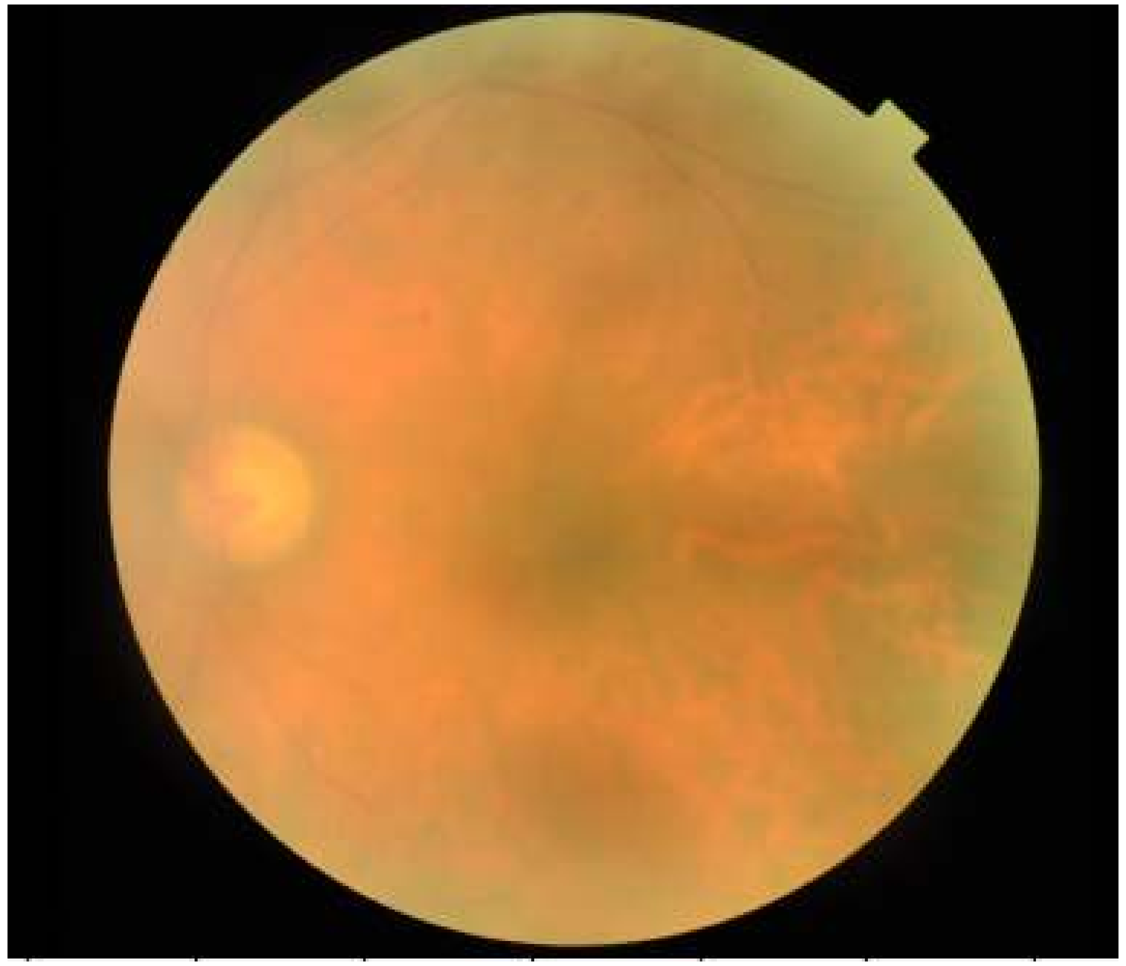
Diabetic retinopathy (DR)

Diabetic retinopathy (DR) is a microvascular complication of diabetes mellitus and is a leading cause of vision loss in the elderly and working population. The image is labeled as DR if it shows any of the following clinical findings: microaneurysms, retinal dot and blot hemorrhage, hard exudates or cotton wool spots

Media Haze (MH)

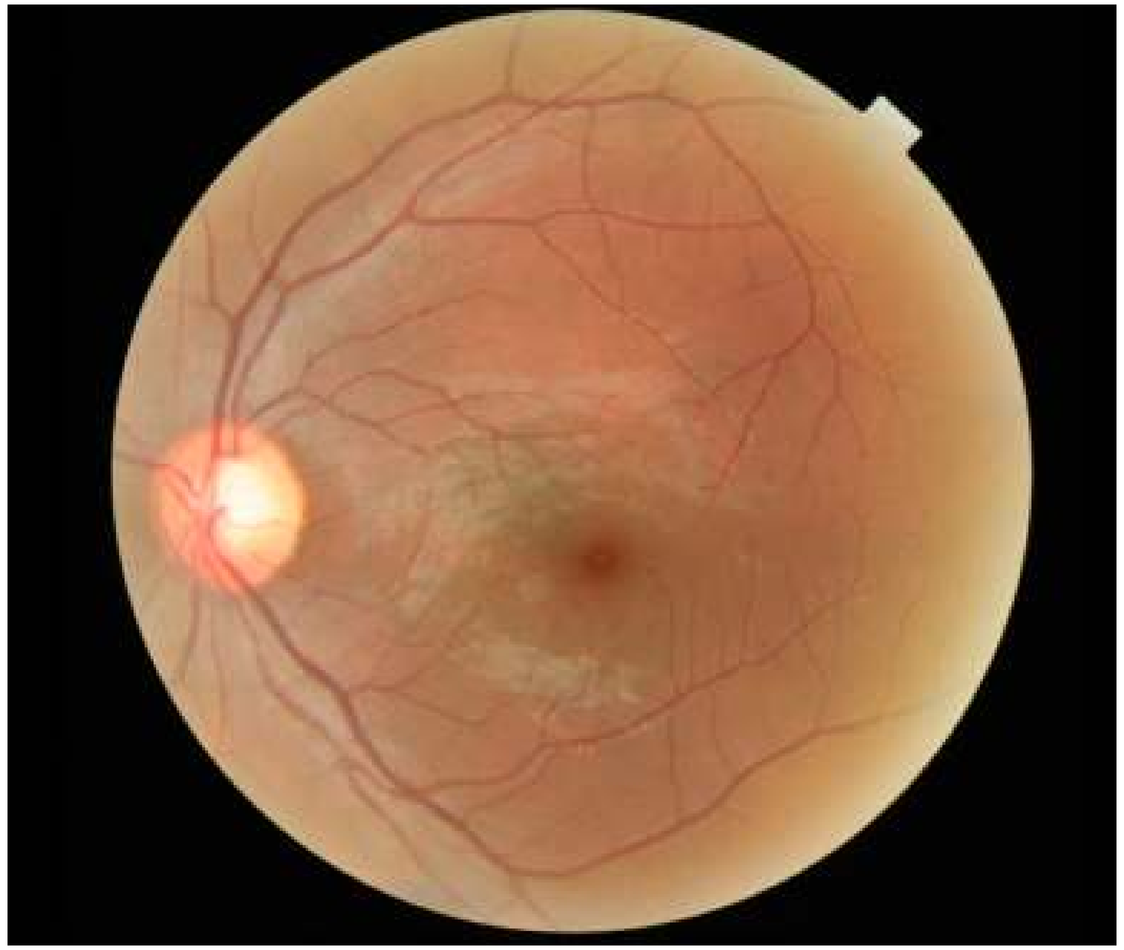
Media Haze (MH): The opacity of media can be a hallmark for the presence of cataracts, vitreous opacities, corneal edema or small pupils. Moreover, some other artifacts may be introduced as a result of acquisition procedures, such as eyelash artifacts and artifacts introduced by the instrument. All these are labeled under the category MH

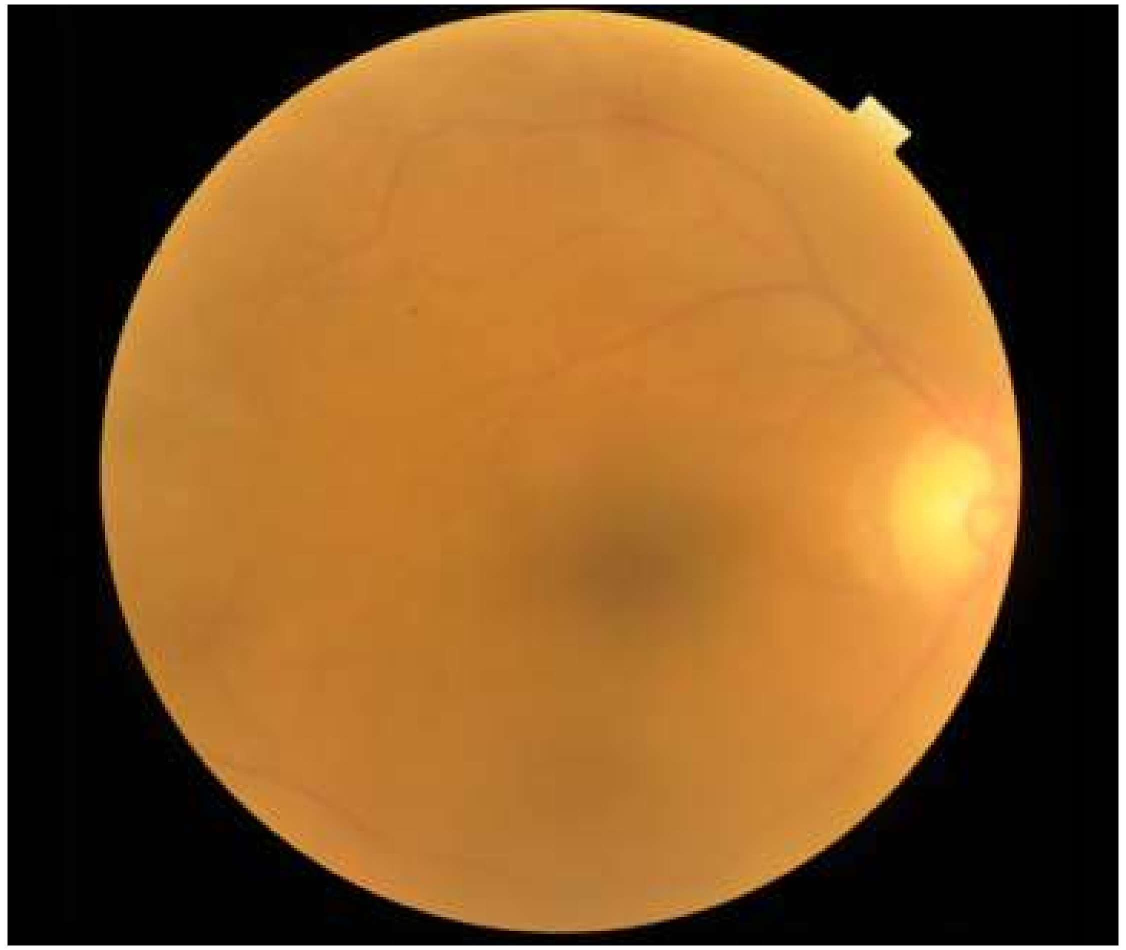




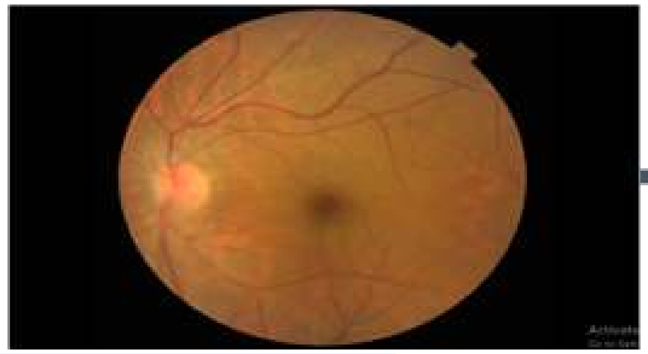
Optic disc cupping (ODC)

Optic disc cupping (ODC) is the thinning of neuroretinal rim such that optic disc appears excavated as shown in Figure 5 c. Pathological ODC is generally referred to as glaucoma. However, several other non-glaucomatous diseases, such as arteritic anterior ischemic optic neuropathy and central retinal vein occlusion, also result in ODC. Thus, it is very important to separately evaluate ODC.

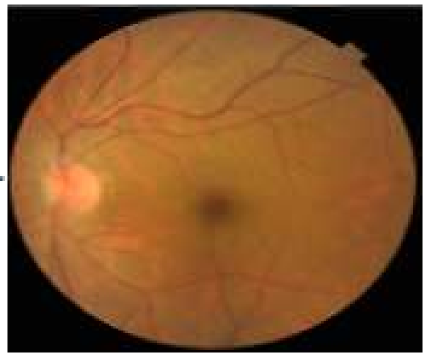




Data Augmentation + Multi-Classification



Pre-processing



ImageDataGenerator(rescale = 1/255., shear_range= 0.1, zoom_range= 0.2, horizontal_flip = True, vertical_flip = True)

top_layer = Sequential()
top_layer.add(Dense(100, activation = 'relu'))
top_layer.add(Dropout(0.2))
top_layer.add(Flatten())
top_layer.add(Dense(512,activation="relu"))
top_layer.add(Dense(1, activation = 'sigmoid'))

Disease Classes (0 =DR/ 1= MH/ 2= ODC)

Disease type	[1]
Disease type	[0]
Disease type	[2]
⋮	⋮

Success Metrics - Model Performance from 5 Different Base Convolution

Architectures	Epochs	Training loss	Testing loss	Training accuracy	Testing accuracy
InceptionV3	50	0.5472	3.8138	0.8494	0.4182
Xception	50	1.0094	1.0212	0.4779	0.5091
VGG16	20	0.0379	2.6346	0.9862	0.4727
MobileNetV2	20	0.0699	2.9862	0.9779	0.4591
EfficientNetB5	30	0.1221	4.4094	0.9572	0.4727
SE-ResNeXt	20	0.0729	3.9207	0.9696	0.4136

Baseline:

```
(y_test == 0).mean()
```

```
0.4727272727272727
```

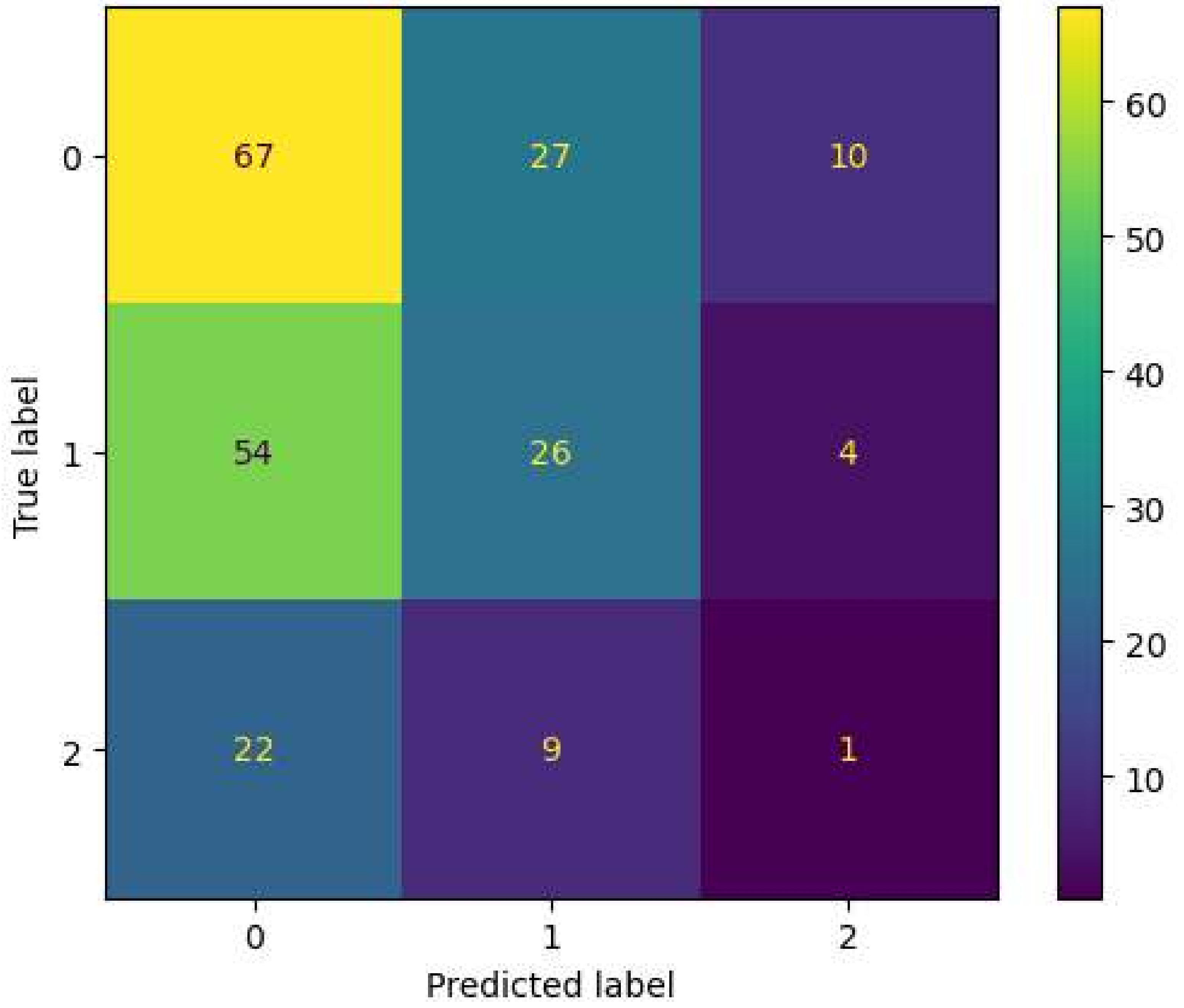
```
(y_test == 1).mean()
```

```
0.38181818181818183
```

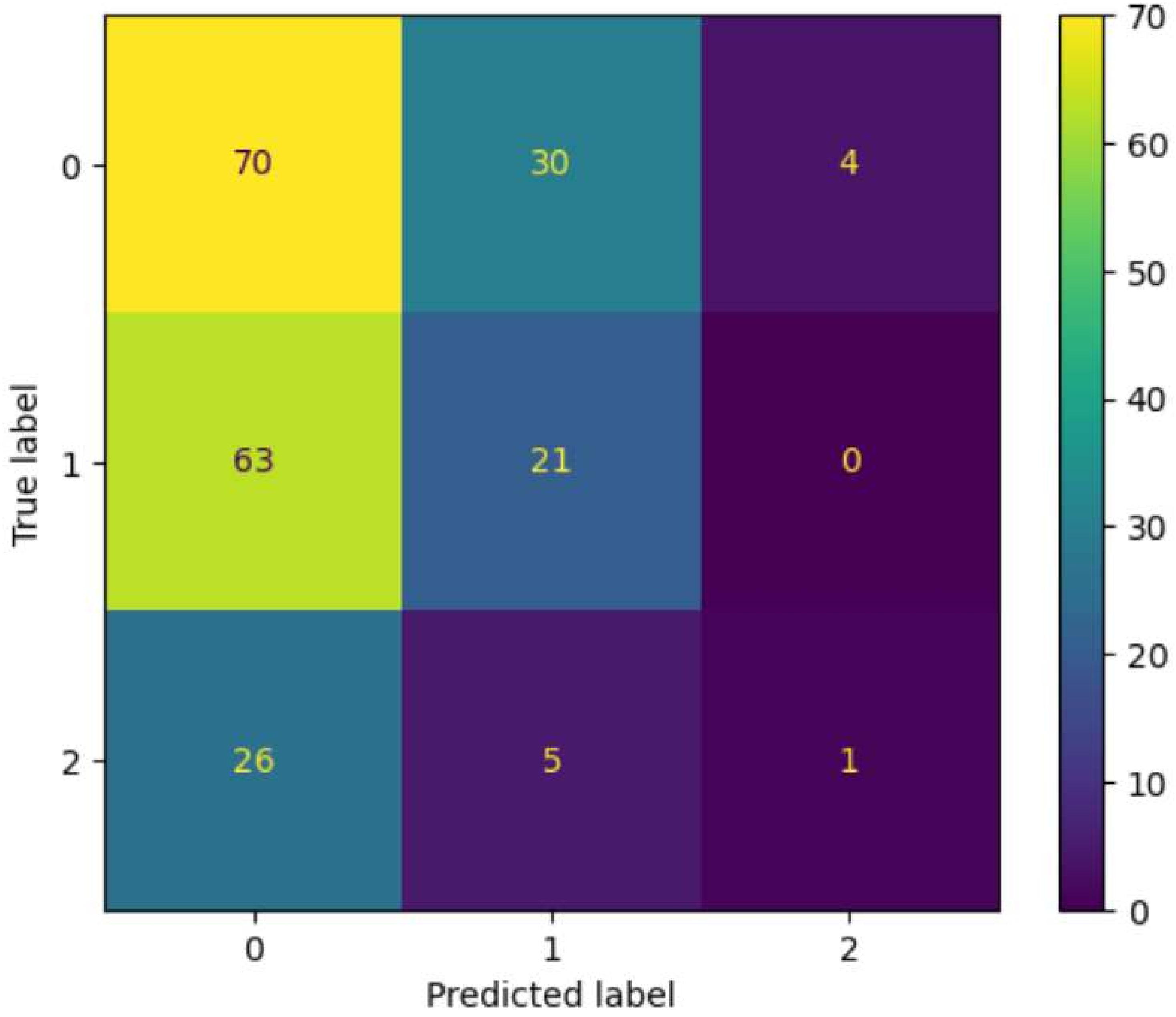
```
(y_test == 2).mean()
```

```
0.14545454545454545
```

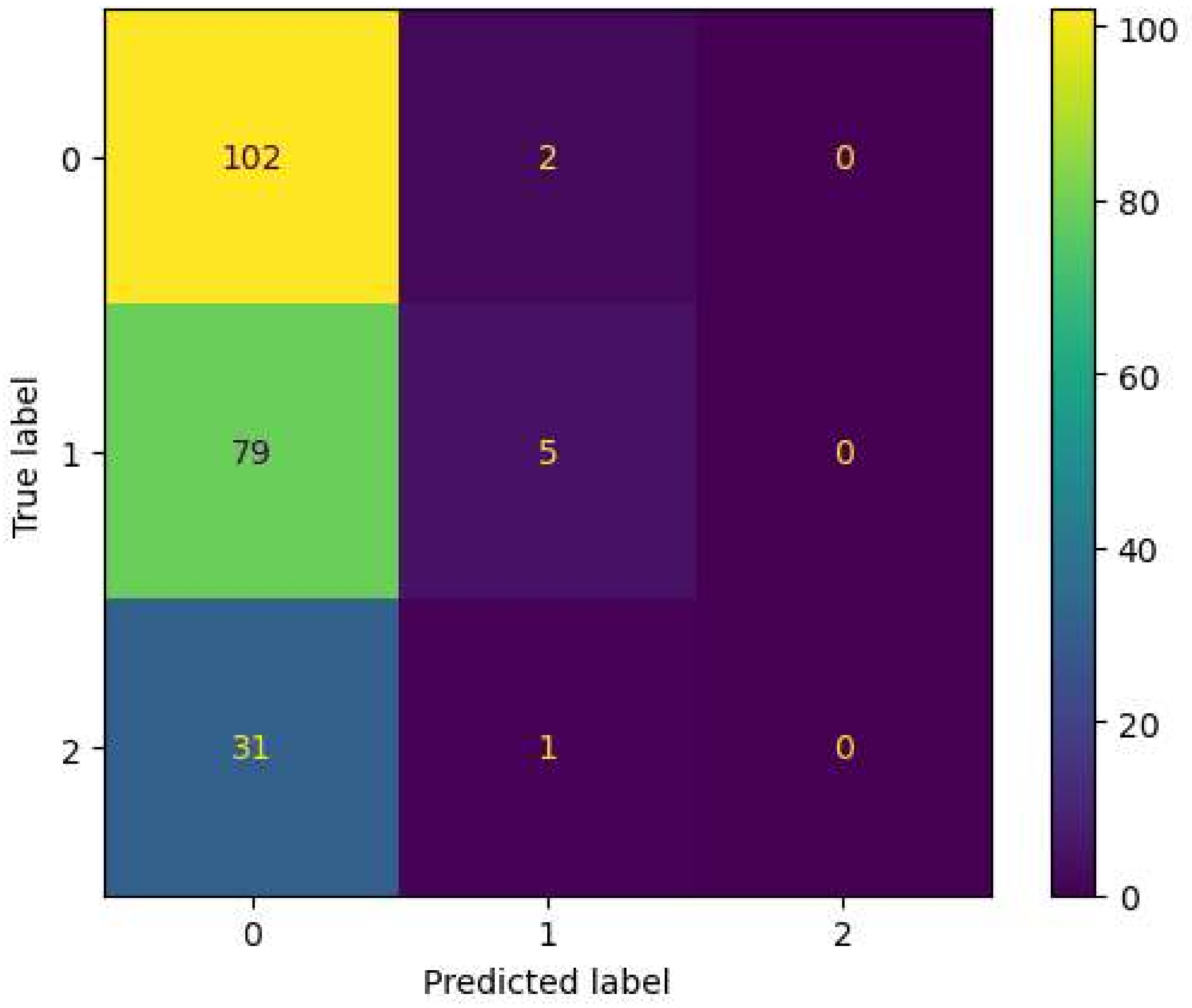
InceptionV3



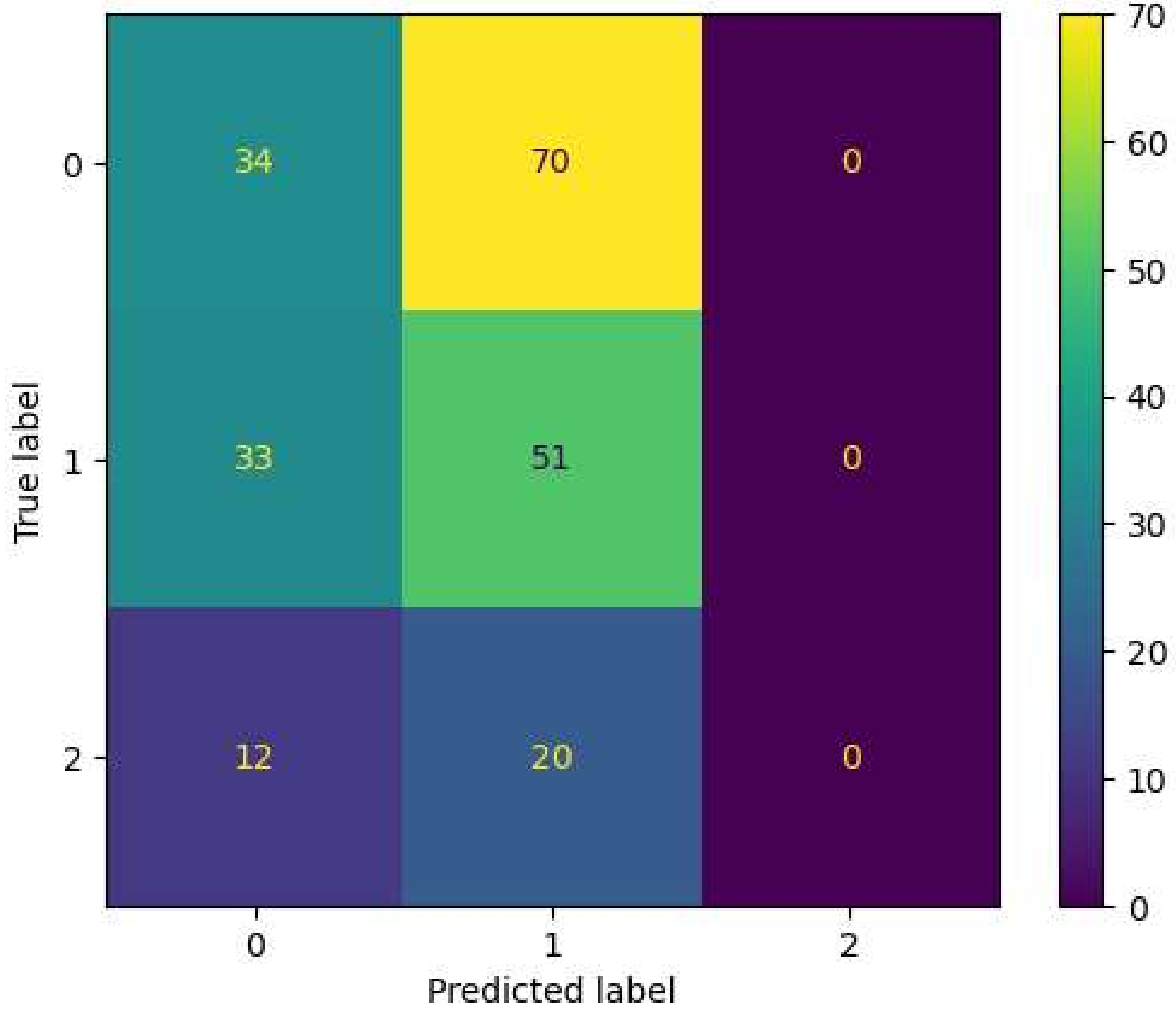
Xception



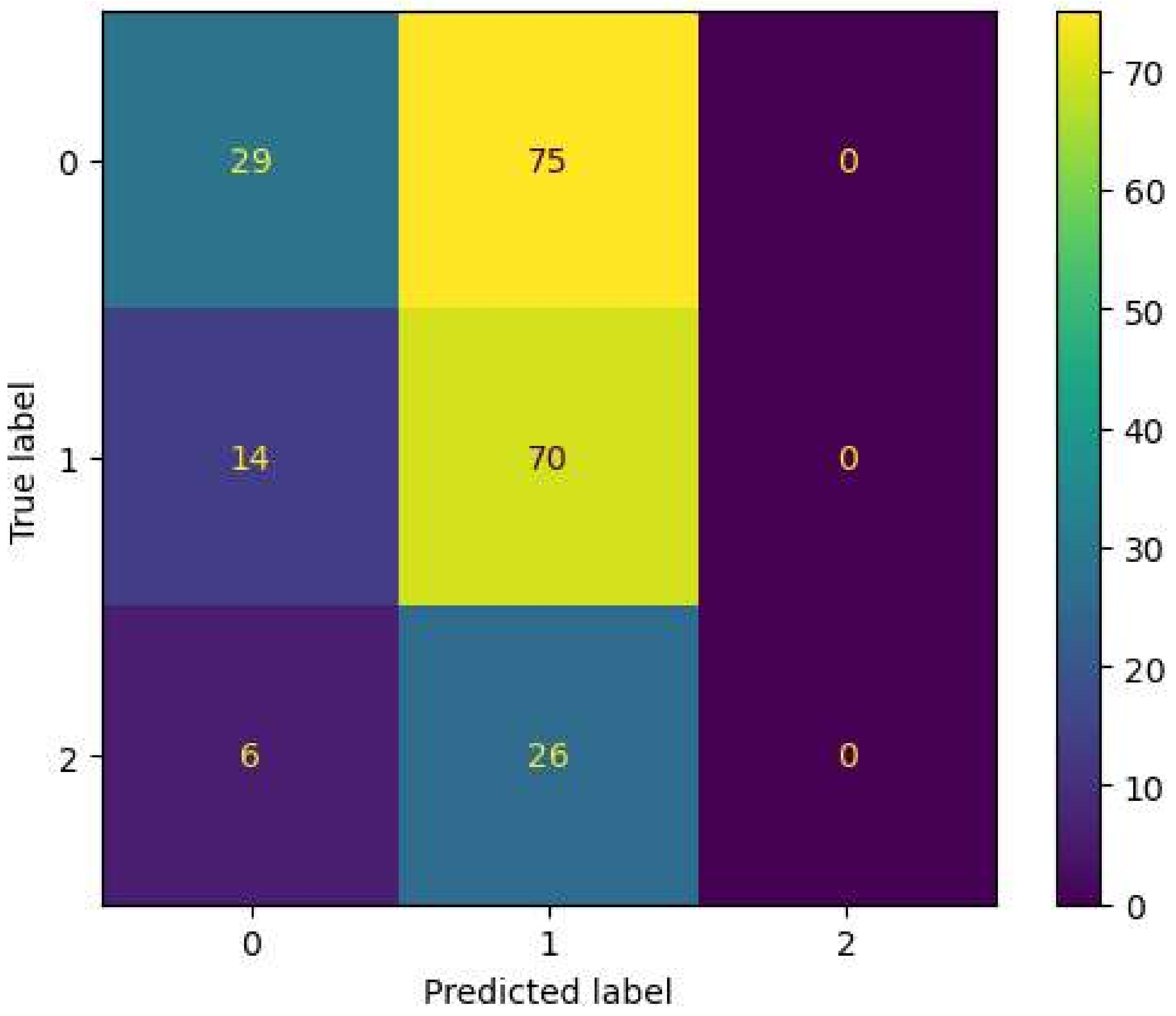
VGG16



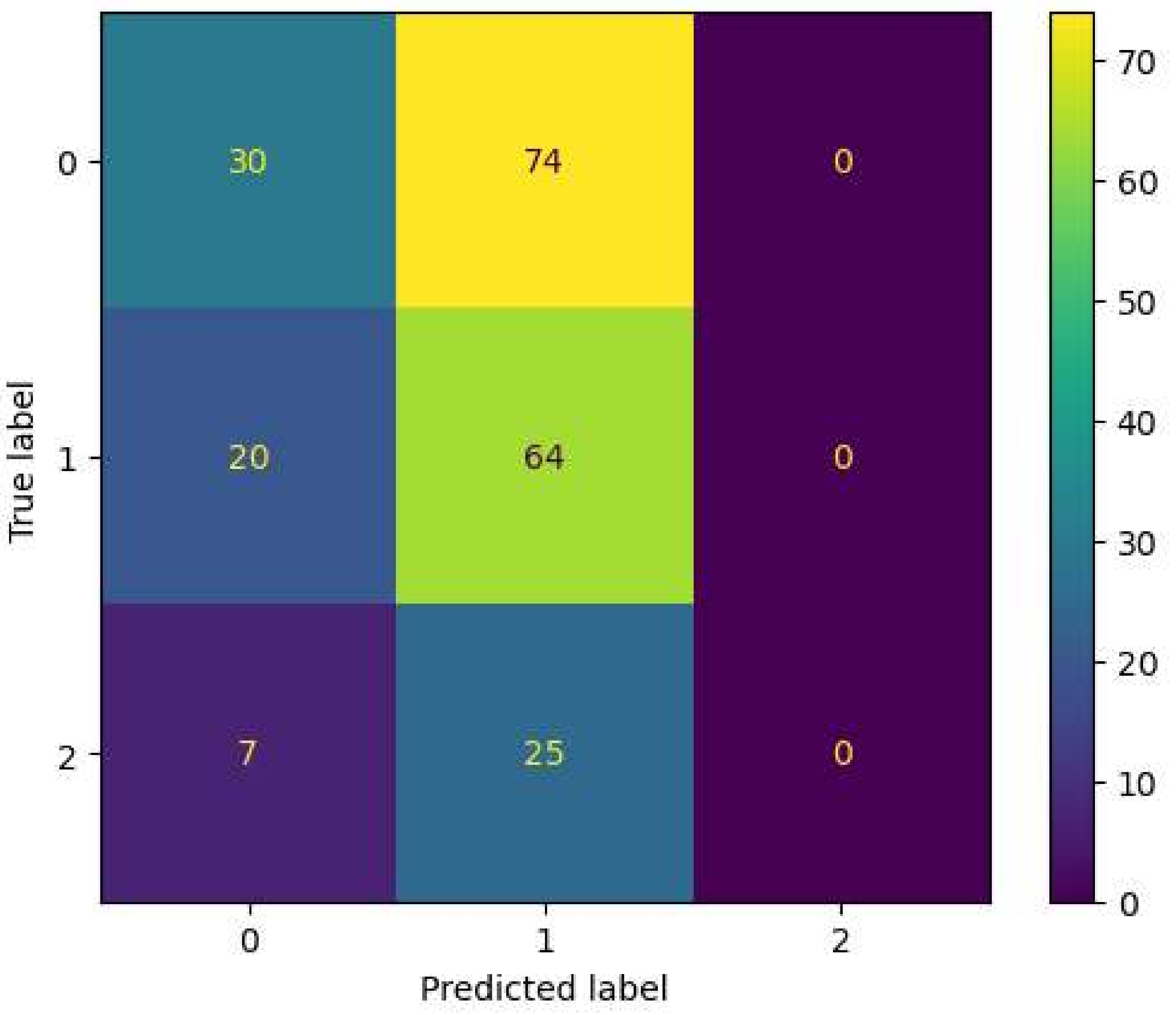
MobileNetV2



EfficientNetB5



SE-ResNeXt



Part 4

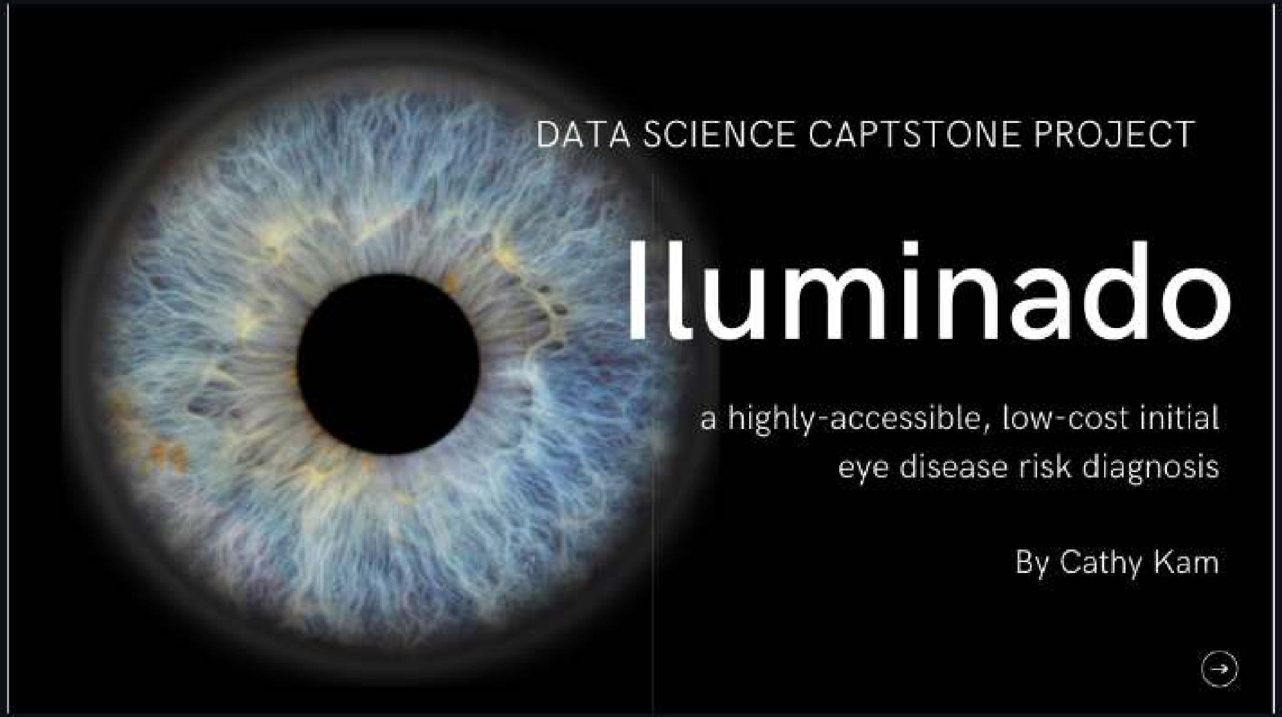
Deployment on Streamlit

Iluminado

Initial Eye Disease Risk Diagnosis

This convolutional net aims to assist ophthalmologist to determine whether an intervention is needed immediately based on retinal fundus photography.

You may upload your retinal image down below taken the fundus camera to see your result of an initial diagnosis.



Project Cover Picture

☑ Consent form: Your participation in this initial diagnosis is voluntary. After you tick the consent form, the image you submitted will be kept as part of data collection for advancing the current model.

```
I have read and I understand the provided information. I understand that my particip
```

Drag and drop your retinal image here

Drag and drop file here
Limit 200MB per file • PNG, JPEG, JPG

Browse files

Please upload an image.


```
  ...
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]]

 [[0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  ...
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]]

 [[0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  ...
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]]

 ...

 [[0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  ...
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]]

 [[0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  ...
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]]

 [[0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]
  ...
  [0. 0. 0.]
  [0. 0. 0.]
  [0. 0. 0.]]]]
```


```
If the result is or close to 1, it means there is a high-risk of eye disease.

If the result is or close to 0, it means your eye is healthy and there is a low-risk
```

The predicted image is [[0. 0. 0. ... 5.7967043 0. 0.] [0. 0. 0. ... 0. 0. 0.] [0. 0. 0. ... 0. 0. 0.]]

Future Adaptation / Improvement:

- Display the type(s) of eye disease that the image is classified to in the app
- Several images could be processed at the same time with a list of patients' name and disease risk prediction generated
- Online machine learning should be incorporated, where data is acquired sequentially and is utilized to update the best predictor for future data at each step
- Patient can choose to allow the diagnosis to be automatically sent to the patient’s preferred ophthalmologist as a medical reference and stored in a stable storage system

Do you have any questions?