CLASSIFICATION OF OPTICAL COHERENCE TOMOGRAPHY (OCT) IMAGES

Team 1:

Satya Sai Mandlem

Nikitha Guntaka

Hemanth Varma Dantuluri

DO YOU KNOW?

- According to National Library of Medicine, 2.2 Million people[4,5,6] globally are suffering from Vision Impairments, can grow to 5.4 million by 2050
- Most cases can be prevented when treated at early stages.

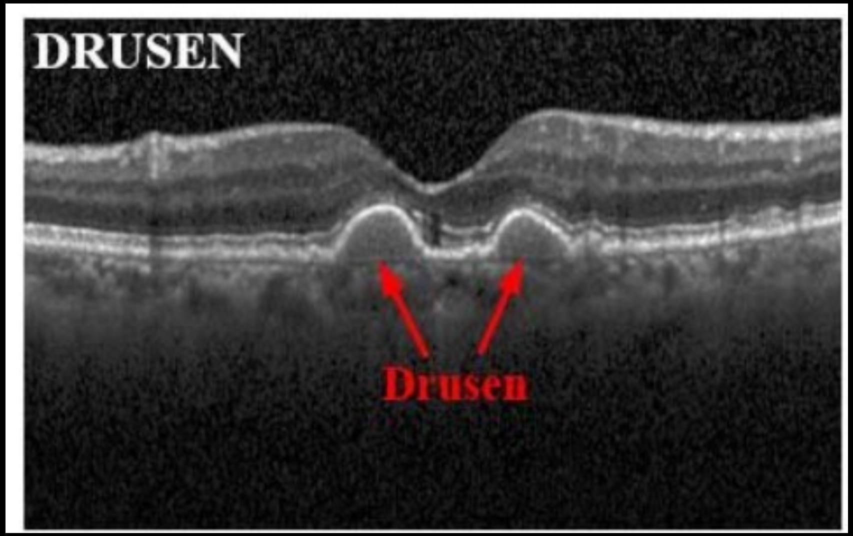
OPTICAL COHERENCE TOMOGRAPHY

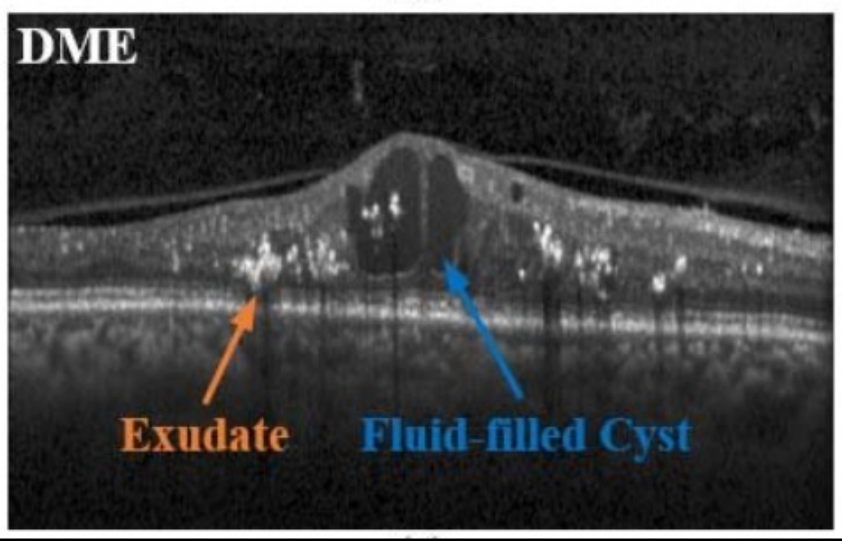
- Optical coherence tomography (OCT)[1] is a non-invasive imaging test. It uses light waves to take cross-section pictures of your retina.
- With OCT, your ophthalmologist can see each of the retina's distinctive layers. This allows your ophthalmologist to map and measure their thickness. These measurements help with diagnosis.

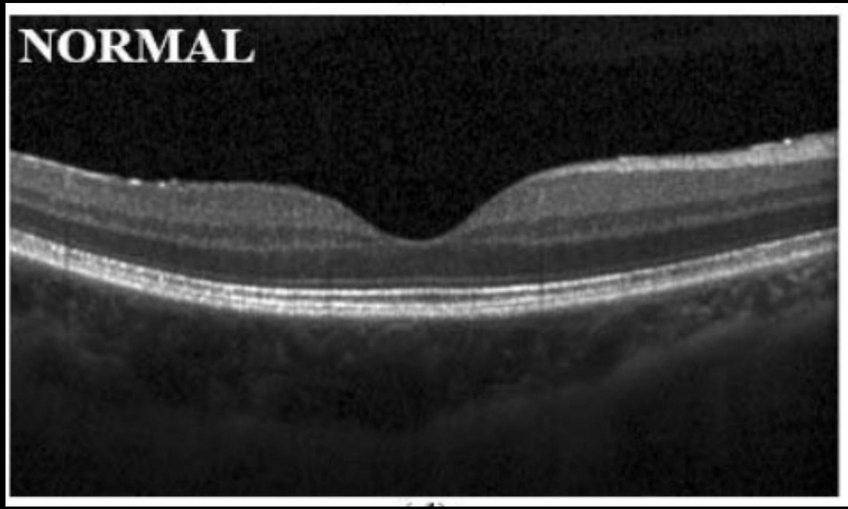
IMPORTANCE OF OCT IMAGES

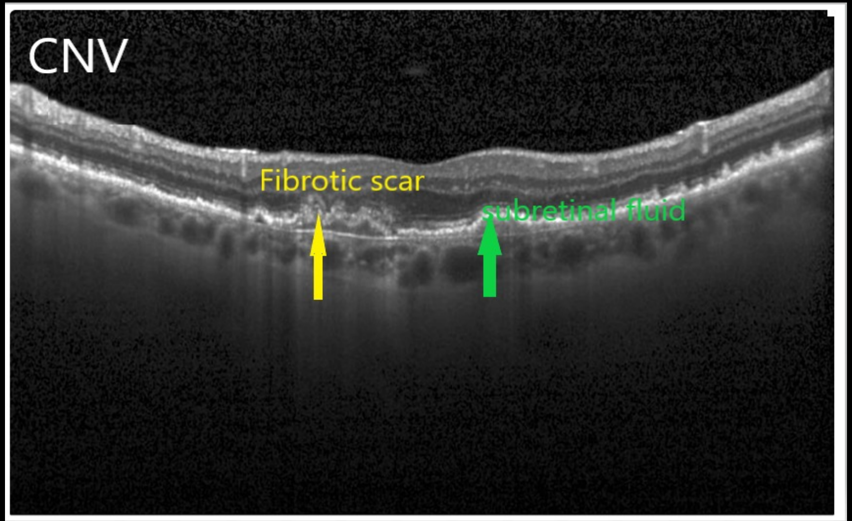
- According to National Library of Medicine[2], drusen, choroidal neovascularization (CNV), and diabetic macular edema (DME) are significant causes of visual impairment globally. Optical coherence tomography (OCT) imaging has emerged as a valuable diagnostic tool for these ocular conditions. However, subjective interpretation and inter-observer variability highlight the need for standardized diagnostic approaches.

IN-DEPTH CLASSIFICATION OF EACH CONDITION:









OBJECTIVES

- Develop deep learning models capable of accurately classifying retinal diseases such as CNV, DME, and Drusen based on OCT images.
- Evaluate the effectiveness of various deep learning architectures for OCT image classification.
- Investigate the impact of dataset size and diversity on model performance and generalization ability.
- Address class imbalance issues in the dataset, particularly for rare or underrepresented retinal diseases.
- Assess the influence of preprocessing techniques on model performance and interpretability.

OUR RESEARCH QUESTIONS

- How effective are deep learning models in accurately classifying retinal diseases based on OCT images?
- Which deep learning architectures prove most effective for OCT image classification?
- How does the size and diversity of the training dataset impact model robustness and generalization ability?
- What strategies can be employed to address class imbalance issues in the OCT image dataset?
- How do preprocessing techniques such as denoising, image registration, and normalization impact model performance and interpretability?

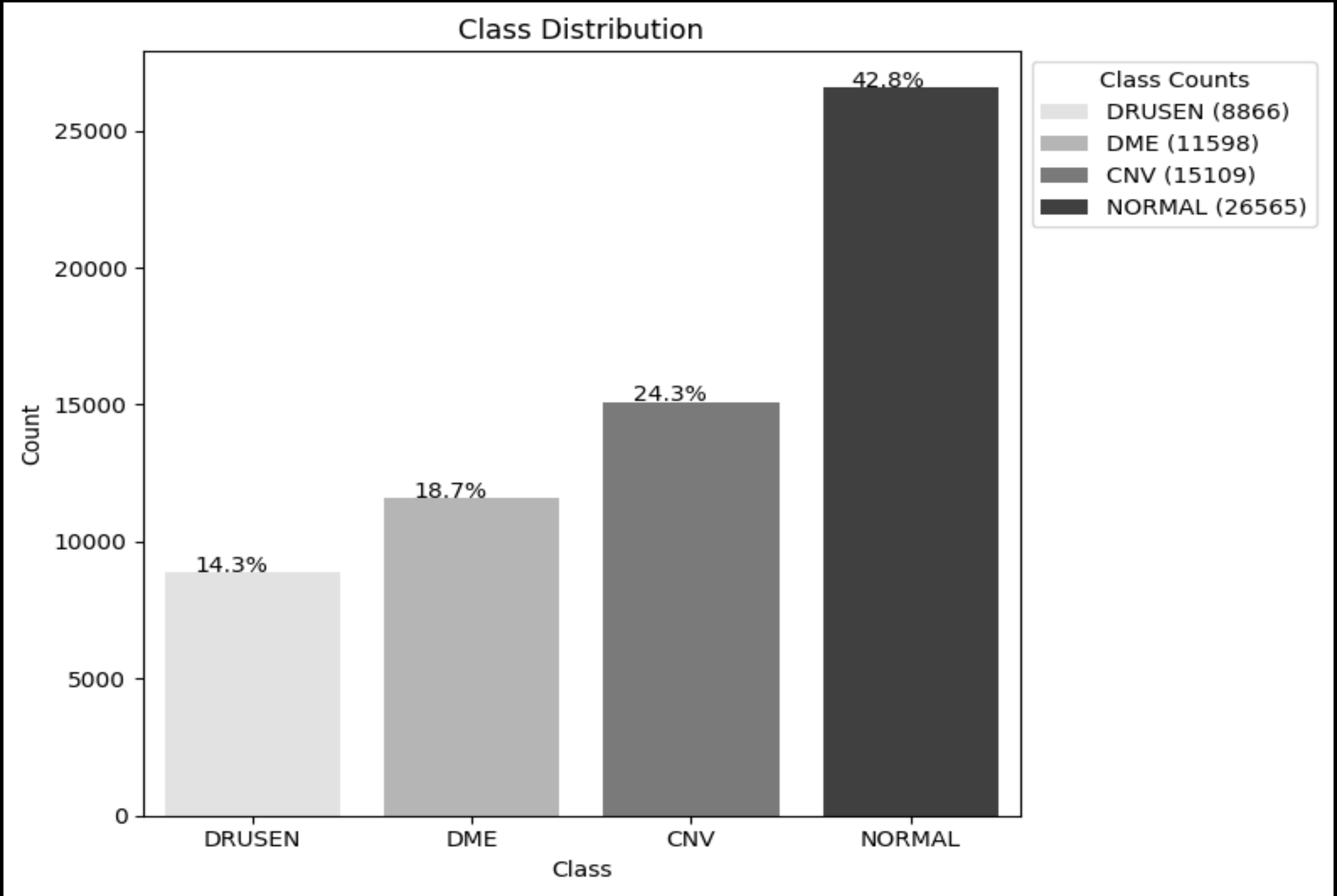
ABOUT THE DATASET

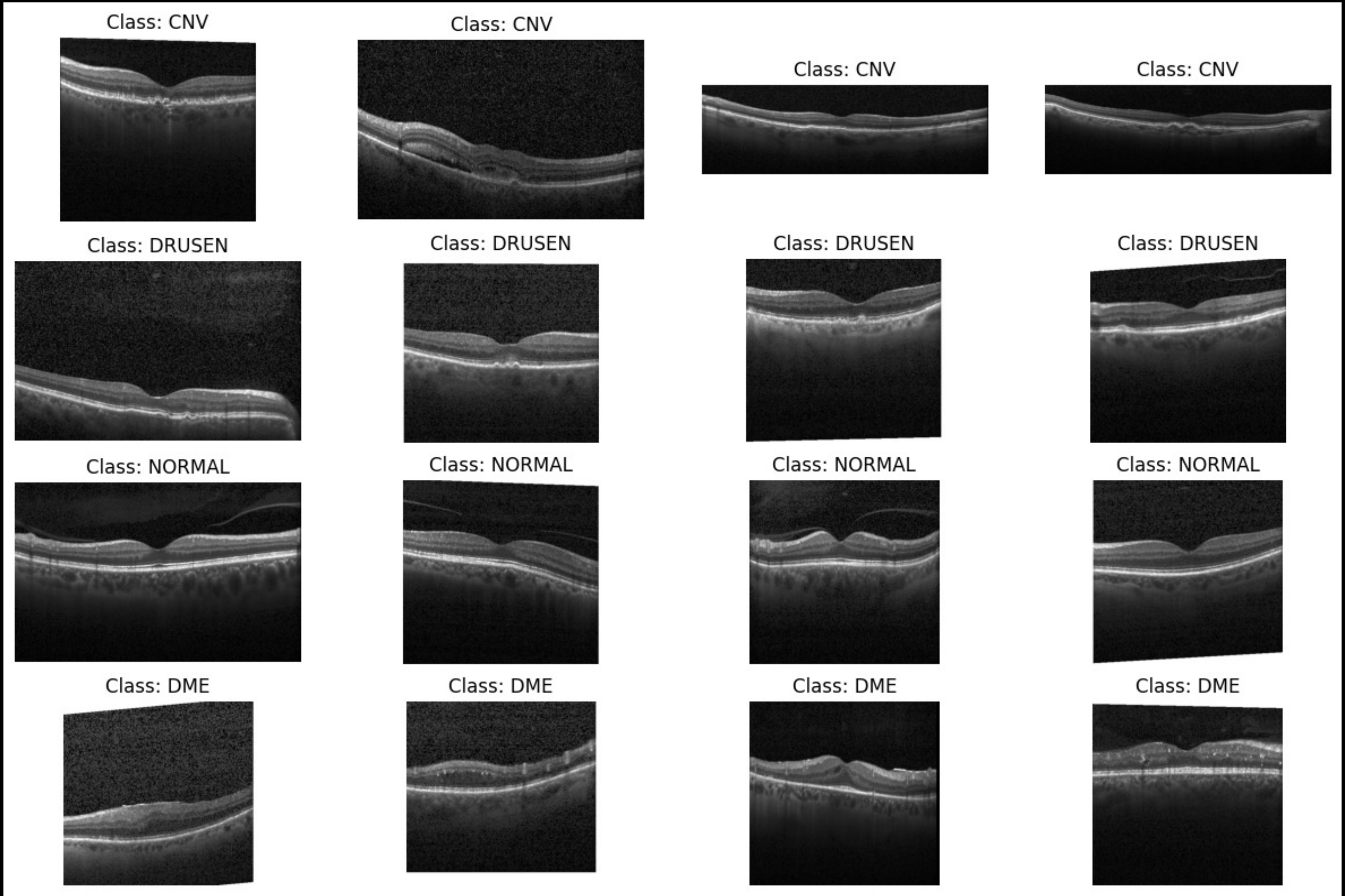
- Source:

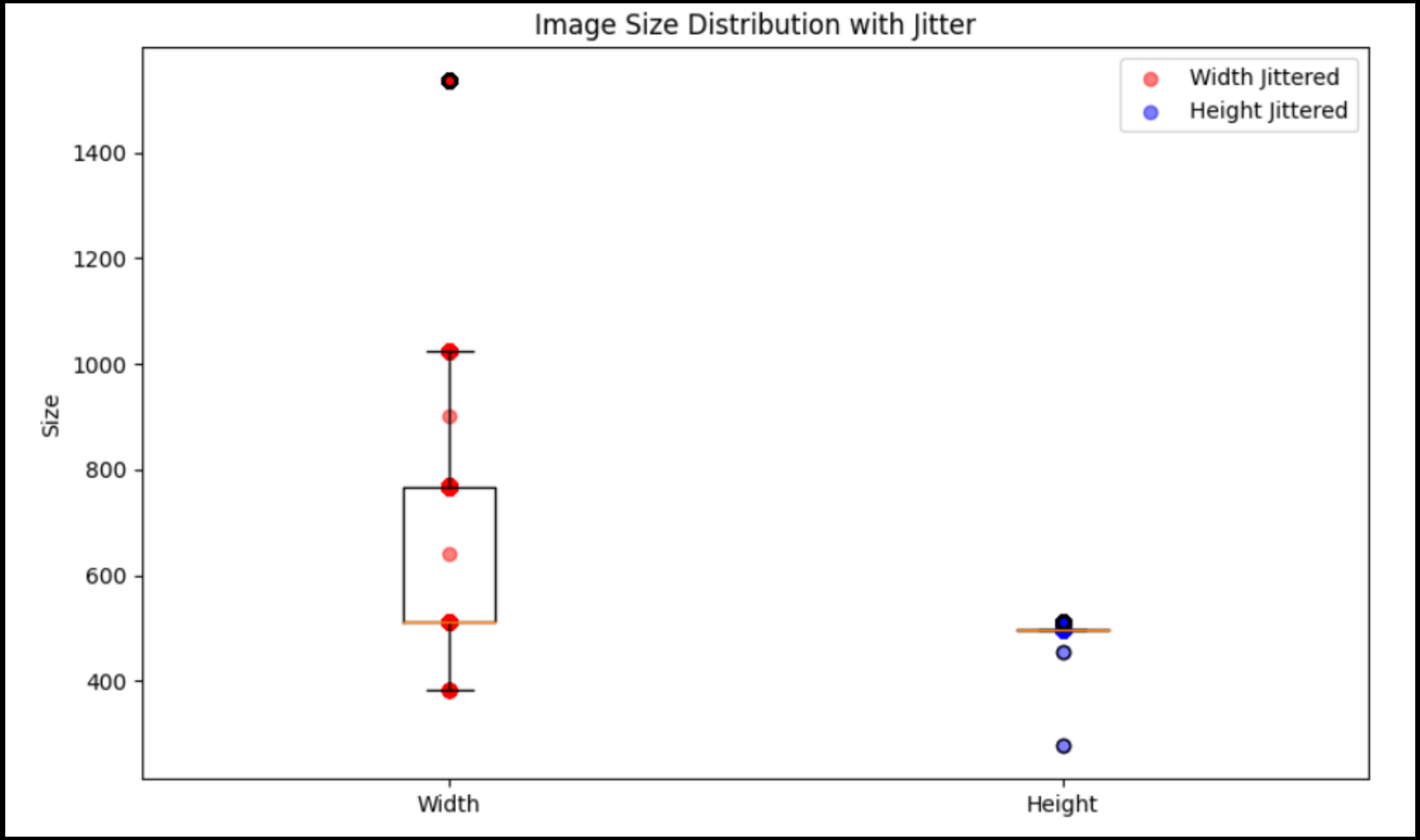
Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), “Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification”, Mendeley Data, V2, doi: 10.17632/rscbjbr9sj.2[3]

The dataset has 62,138 images of the following classes,

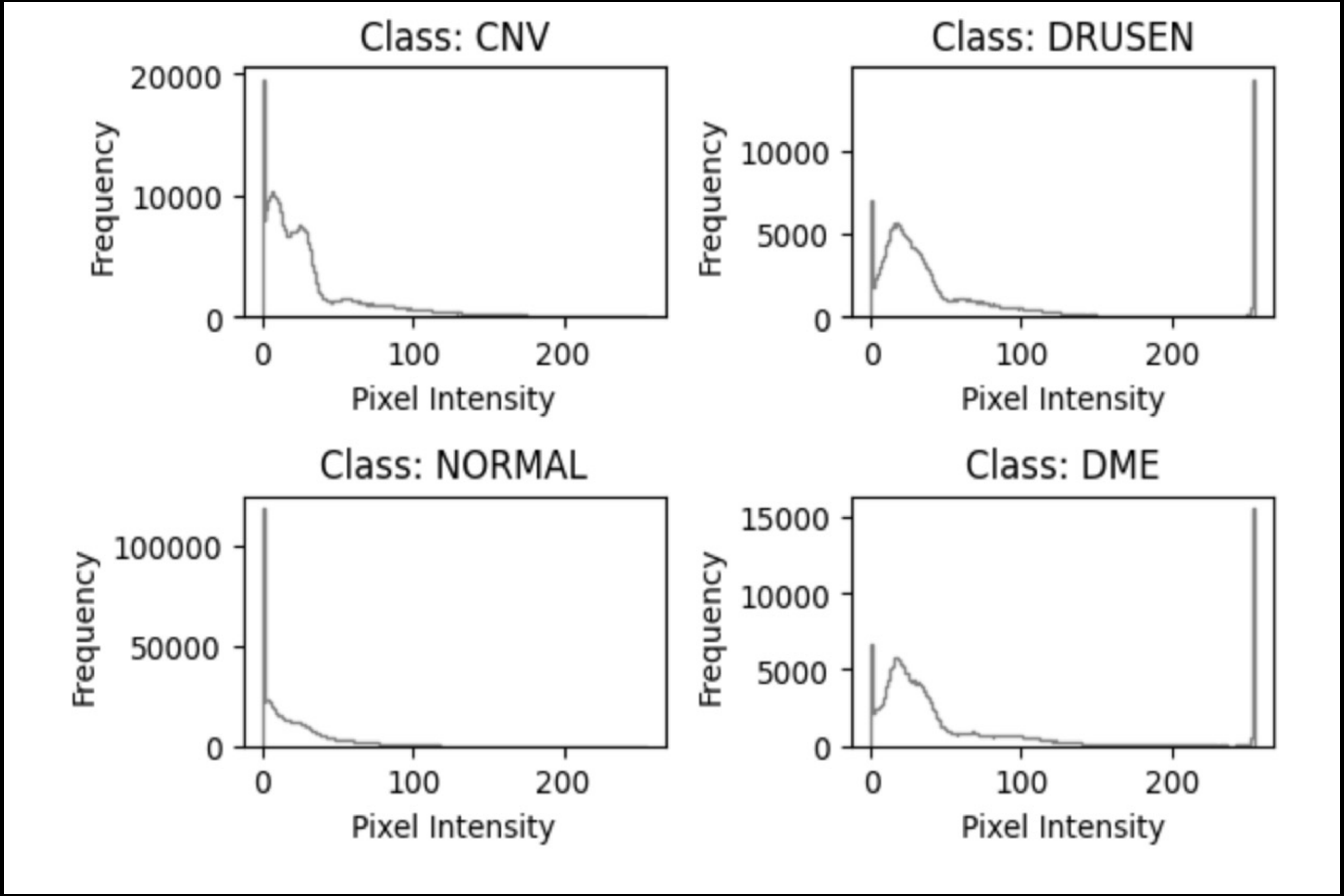
- CNV (Choroidal Neovascularization) – 15109 images
- DME (Diabetic Macular Edema) – 11598 images
- DRUSEN – 8866 images
- NORMAL – 26565 images


Class Distribution
Class Counts
DRUSEN (8866)
DME (11598)
CNV (15109)
NORMAL (26565)
42.8%
24.3%
18.7%
14.3%
25000
20000
15000
10000
5000
0
Count
DRUSEN
DME
CNV
NORMAL
Class


Class: CNV
Class: CNV
Class: CNV
Class: CNV
Class: DRUSEN
Class: DRUSEN
Class: DRUSEN
Class: DRUSEN
Class: NORMAL
Class: NORMAL
Class: NORMAL
Class: NORMAL
Class: DME
Class: DME
Class: DME
Class: DME




```
Image Size Statistics:
Minimum Width: 384
Maximum Width: 1536
Average Width: 695.7205536598084

Minimum Height: 277
Maximum Height: 512
Average Height: 497.2650742390837
```


Class: CNV
Class: DRUSEN
Class: NORMAL
Class: DME
Frequency
Pixel Intensity
20000
10000
0
100
200
15000
5000
100000
50000

PREPROCESSING STEPS

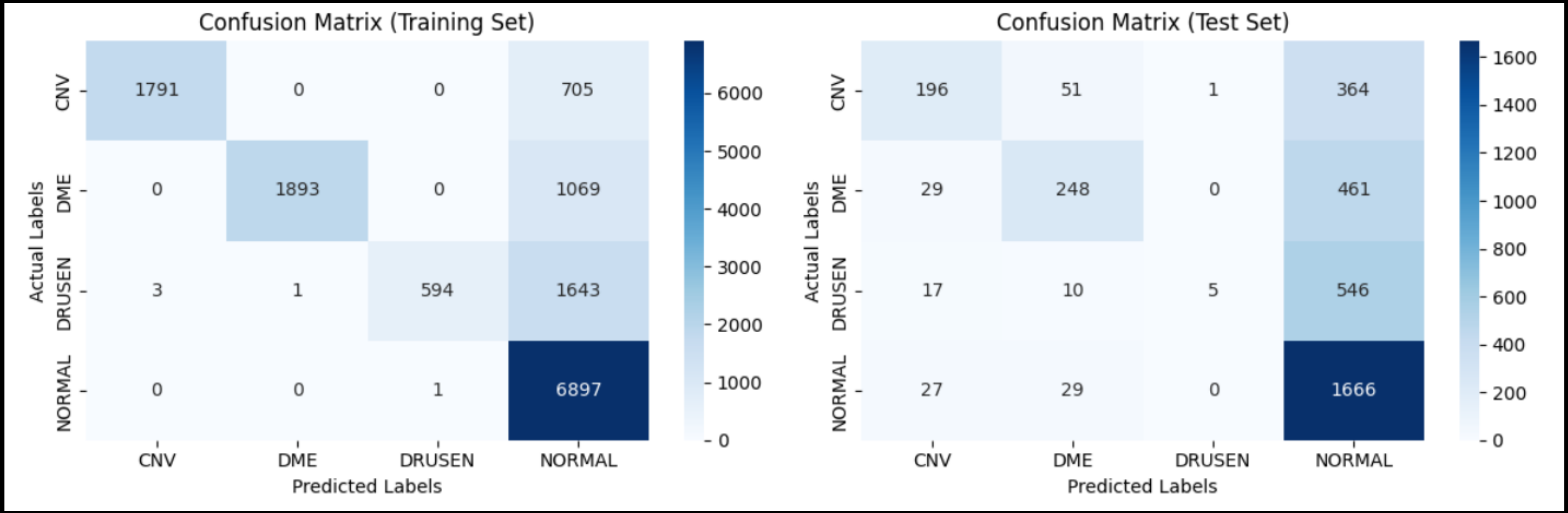
- We resized the images to 256x256 and normalized the pixels using OpenCV library.
- We then stored the images in NumPy arrays and performed augumentation.
- We used data augmentation setup with Keras' ImageDataGenerator, which will generate augmented versions of the original images on-the-fly during training. Each batch of images fed into the model for training will be augmented differently, introducing variability and helping the model generalize better.

MODELLING

RANDOM FOREST

- n_estimators=20
- max_depth = 10
- max_features= "sqrt“

```
Test Metrics:
Test Accuracy: 0.5794520547945206
Test Precision: 0.6612902812222957
Test Recall: 0.5794520547945206
Test F1-score: 0.5008363280543155

Train Metrics:
Train Accuracy: 0.7655682674522162
Train Precision: 0.8427901126668152
Train Recall: 0.7655682674522162
Train F1-score: 0.7440118708689517
```



K-NEAREST NEIGHBORS (KNN)

- algorithm='auto’
- n_neighbors= 7
- p= 1

Classification Report (Training Set - KNN):

	precision	recall	f1-score	support
CNV	0.63	0.49	0.55	791
DME	0.49	0.63	0.55	787
DRUSEN	0.53	0.53	0.53	806
NORMAL	0.57	0.54	0.56	816
accuracy			0.55	3200
macro avg	0.56	0.55	0.55	3200
weighted avg	0.56	0.55	0.55	3200

Classification Report (Testing Set - KNN):

	precision	recall	f1-score	support
CNV	0.31	0.21	0.25	209
DME	0.38	0.42	0.40	213
DRUSEN	0.34	0.36	0.35	194
NORMAL	0.30	0.35	0.33	184
accuracy			0.34	800
macro avg	0.33	0.34	0.33	800
weighted avg	0.33	0.34	0.33	800

SUPPORT VECTOR MACHINES (SVM)

- gamma='auto’
- C= 1
- kernel = 'rbf

Classification Report (Training Set):

	precision	recall	f1-score	support
CNV	0.56	0.48	0.51	791
DME	0.38	0.34	0.36	787
DRUSEN	0.39	0.33	0.36	806
NORMAL	0.42	0.58	0.49	816
accuracy			0.43	3200
macro avg	0.44	0.43	0.43	3200
weighted avg	0.44	0.43	0.43	3200

Classification Report (Testing Set):

	precision	recall	f1-score	support
CNV	0.52	0.41	0.46	209
DME	0.35	0.29	0.32	213
DRUSEN	0.32	0.30	0.31	194
NORMAL	0.35	0.52	0.42	184
accuracy			0.38	800
macro avg	0.38	0.38	0.37	800
weighted avg	0.39	0.38	0.37	800

ADABOOSTING

- n_estimators=200
- learning_rate=0.1

Train Metrics (200 Estimators):

Accuracy: 0.5746875
Precision: 0.5774133546676918
Recall: 0.5745430636476441
F1-score: 0.5751545273760863

Test Metrics (200 Estimators):

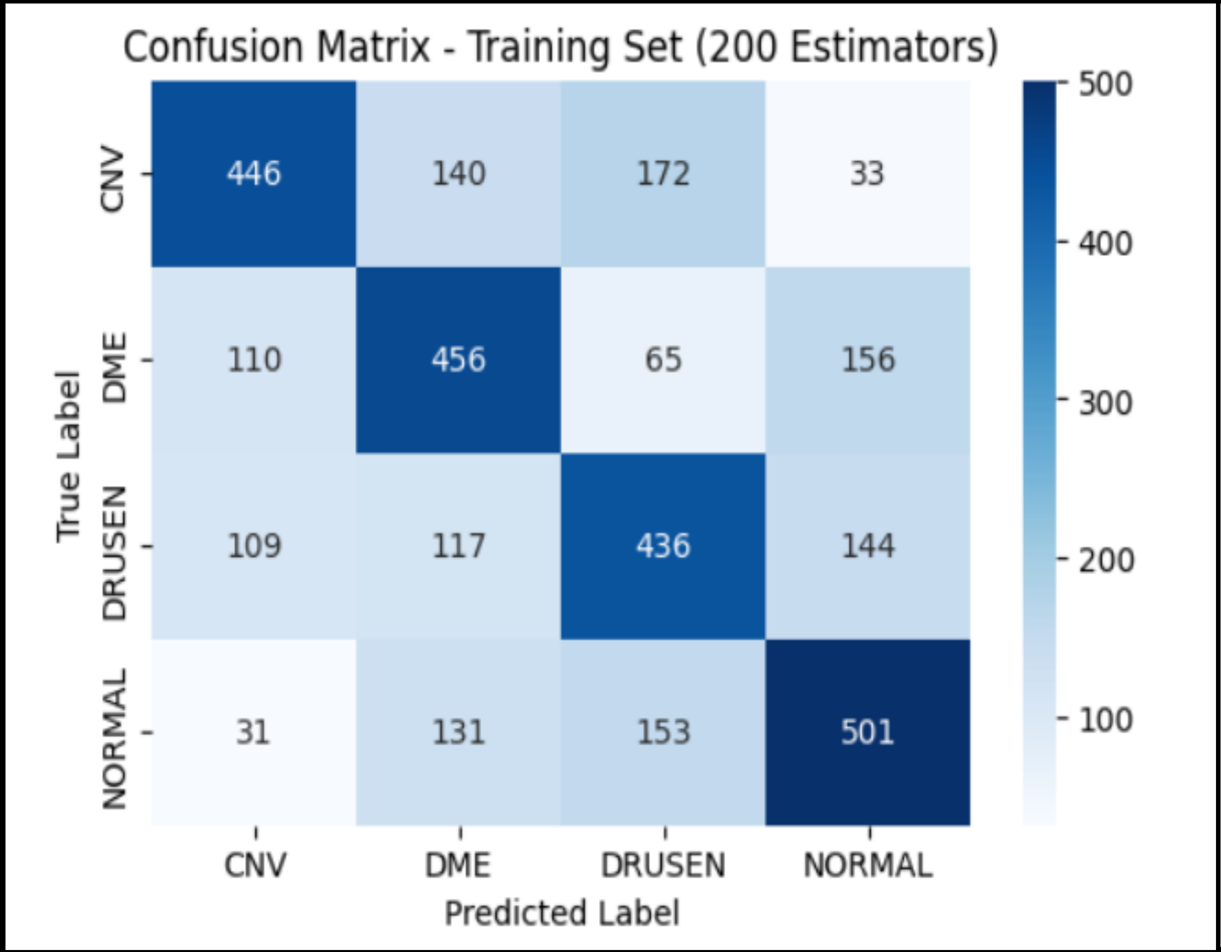
Accuracy: 0.44125
Precision: 0.4484732334907211
Recall: 0.4410732462642231
F1-score: 0.4424190224086838

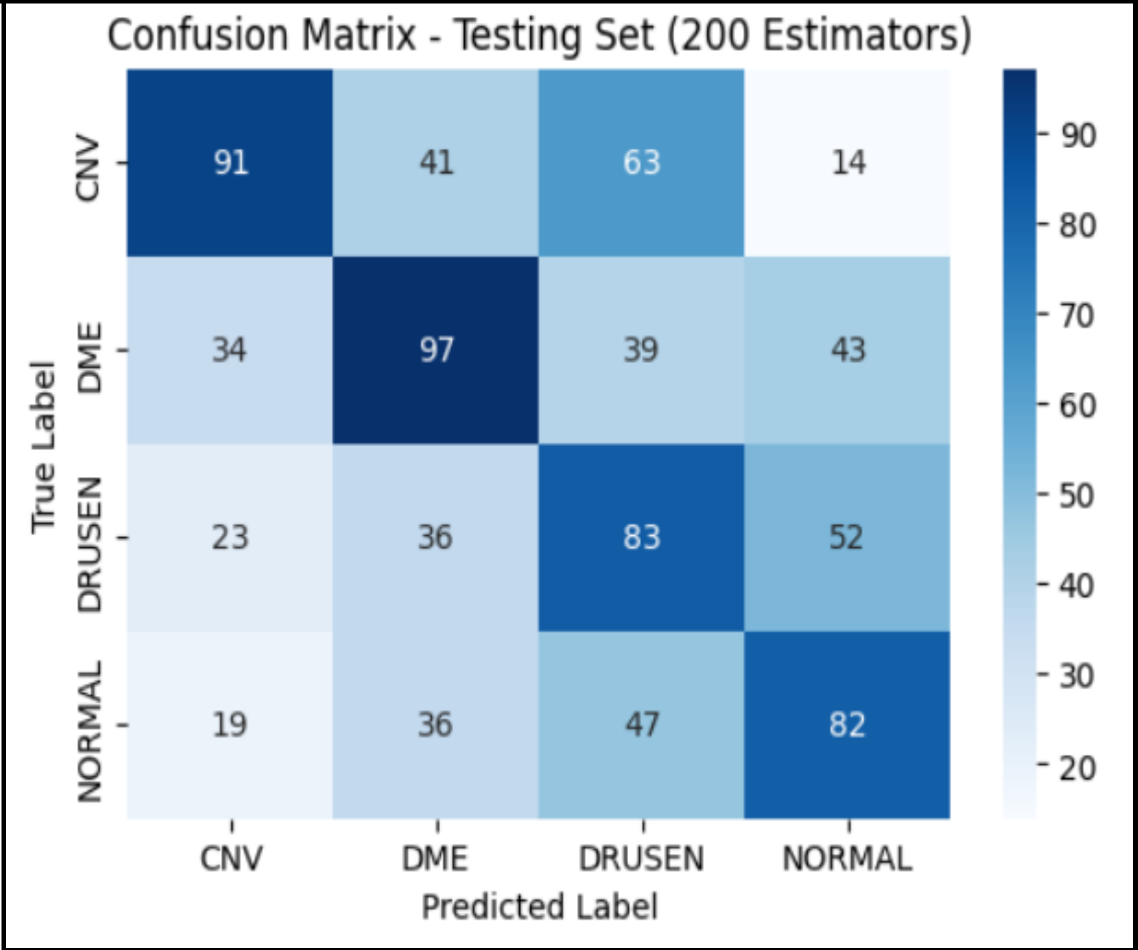
Classification Report (Training Set – 200 Estimators):

	precision	recall	f1–score	support
CNV	0.64	0.56	0.60	791
DME	0.54	0.58	0.56	787
DRUSEN	0.53	0.54	0.53	806
NORMAL	0.60	0.61	0.61	816
accuracy			0.57	3200
macro avg	0.58	0.57	0.58	3200
weighted avg	0.58	0.57	0.58	3200

Classification Report (Testing Set – 200 Estimators):

	precision	recall	f1–score	support
CNV	0.54	0.44	0.48	209
DME	0.46	0.46	0.46	213
DRUSEN	0.36	0.43	0.39	194
NORMAL	0.43	0.45	0.44	184
accuracy			0.44	800
macro avg	0.45	0.44	0.44	800
weighted avg	0.45	0.44	0.44	800


Confusion Matrix - Training Set (200 Estimators)
True Label
Predicted Label
CNV
DME
DRUSEN
NORMAL
446
140
172
33
110
456
65
156
109
117
436
144
31
131
153
501
500
400
300
200
100


Confusion Matrix - Testing Set (200 Estimators)
True Label
Predicted Label
CNV
DME
DRUSEN
NORMAL
91
41
63
14
34
97
39
43
23
36
83
52
19
36
47
82
90
80
70
60
50
40
30
20

GRADIENTBOOSTING

- n_estimators=50
- learning_rate=0.1

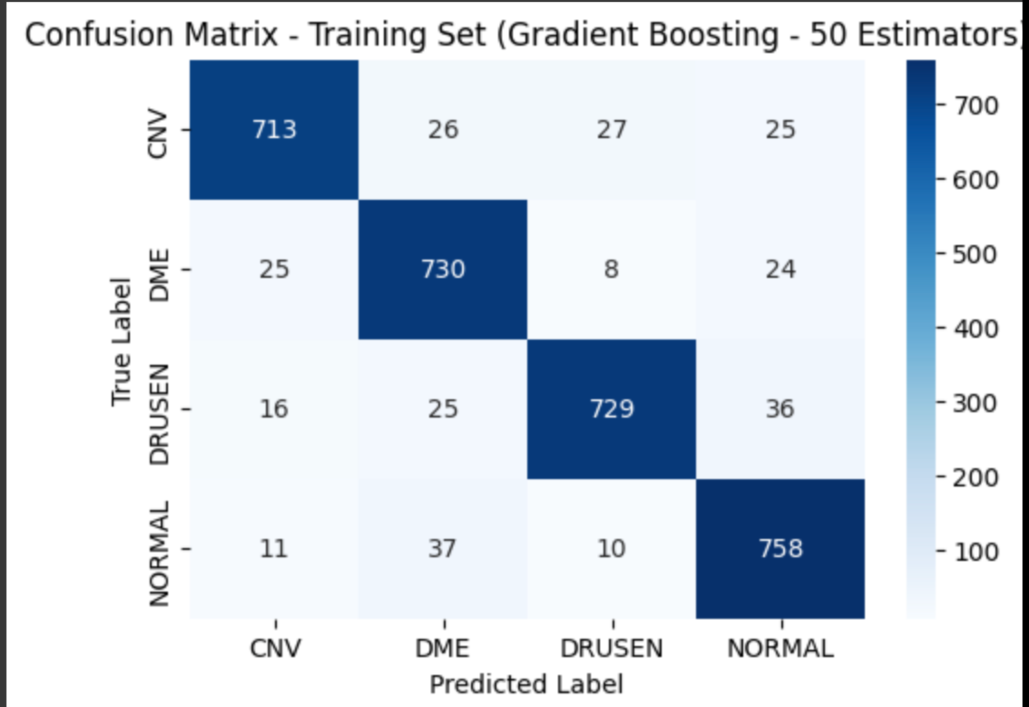
Classification Report (Training Set - Gradient Boosting - 50 Estimators):

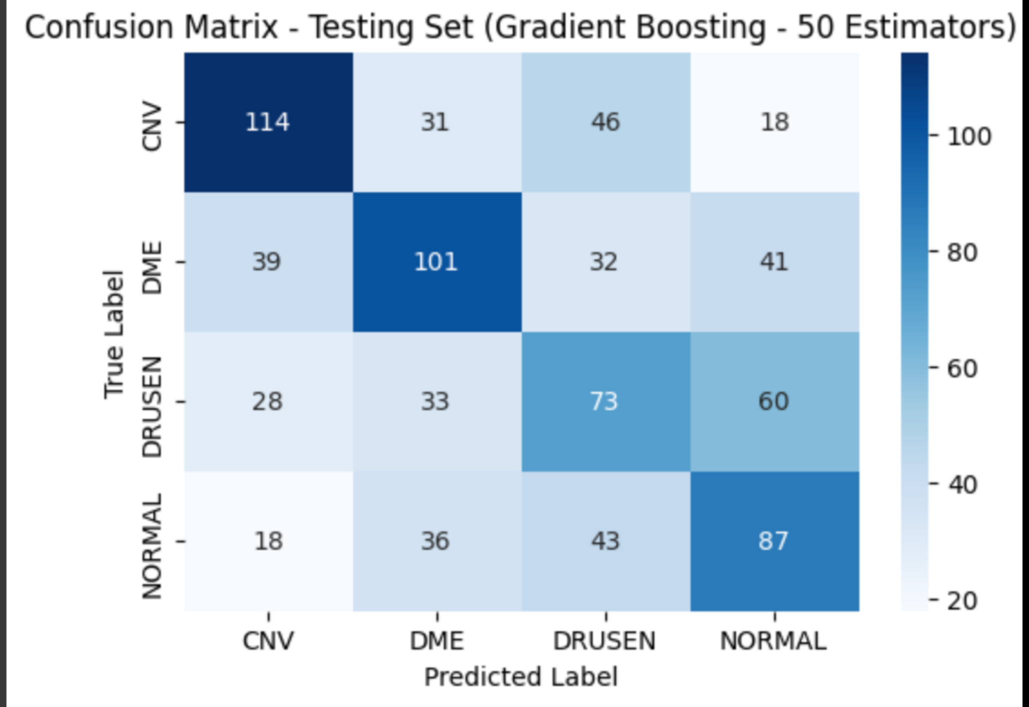
	precision	recall	f1-score	support
CNV	0.93	0.90	0.92	791
DME	0.89	0.93	0.91	787
DRUSEN	0.94	0.90	0.92	806
NORMAL	0.90	0.93	0.91	816
accuracy			0.92	3200
macro avg	0.92	0.92	0.92	3200
weighted avg	0.92	0.92	0.92	3200

Classification Report (Testing Set - Gradient Boosting - 50 Estimators):

	precision	recall	f1-score	support
CNV	0.57	0.55	0.56	209
DME	0.50	0.47	0.49	213
DRUSEN	0.38	0.38	0.38	194
NORMAL	0.42	0.47	0.45	184
accuracy			0.47	800
macro avg	0.47	0.47	0.47	800
weighted avg	0.47	0.47	0.47	800

```
Training Metrics (Gradient Boosting - 50 Estimators):

Accuracy: 0.915625
Precision: 0.9163691947675392
Recall: 0.9155879442208441
F1-score: 0.9156745173100541

Testing Metrics (Gradient Boosting - 50 Estimators):

Accuracy: 0.46875
Precision: 0.46849266016957203
Recall: 0.4671869239901875
F1-score: 0.4672971851683588
```





XGBOOSTING:

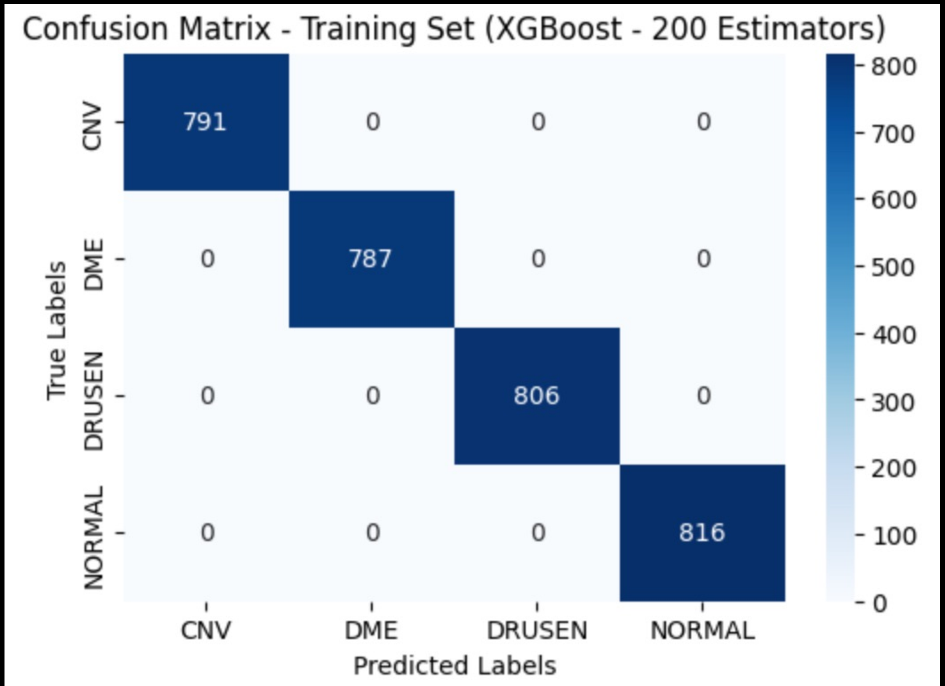
n_estimators=200, learning_rate=0.01, max_depth = 5)

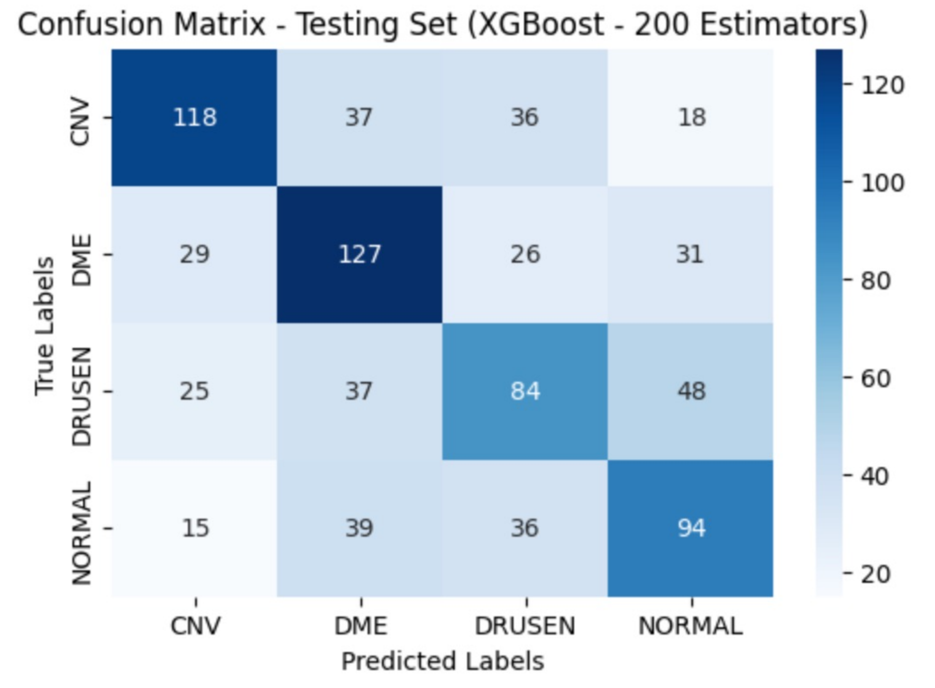
Classification Report (Training Set – XGBoost – 200 Estimators):

	precision	recall	f1-score	support
CNV	0.99	0.95	0.97	791
DME	0.95	0.94	0.95	787
DRUSEN	0.98	0.95	0.97	806
NORMAL	0.91	0.98	0.94	816
accuracy			0.96	3200
macro avg	0.96	0.96	0.96	3200
weighted avg	0.96	0.96	0.96	3200

Classification Report (Testing Set – XGBoost – 200 Estimators):

	precision	recall	f1-score	support
CNV	0.57	0.60	0.58	209
DME	0.55	0.50	0.52	213
DRUSEN	0.41	0.34	0.37	194
NORMAL	0.47	0.57	0.51	184
accuracy			0.50	800
macro avg	0.50	0.50	0.50	800
weighted avg	0.50	0.50	0.50	800





VGG-16

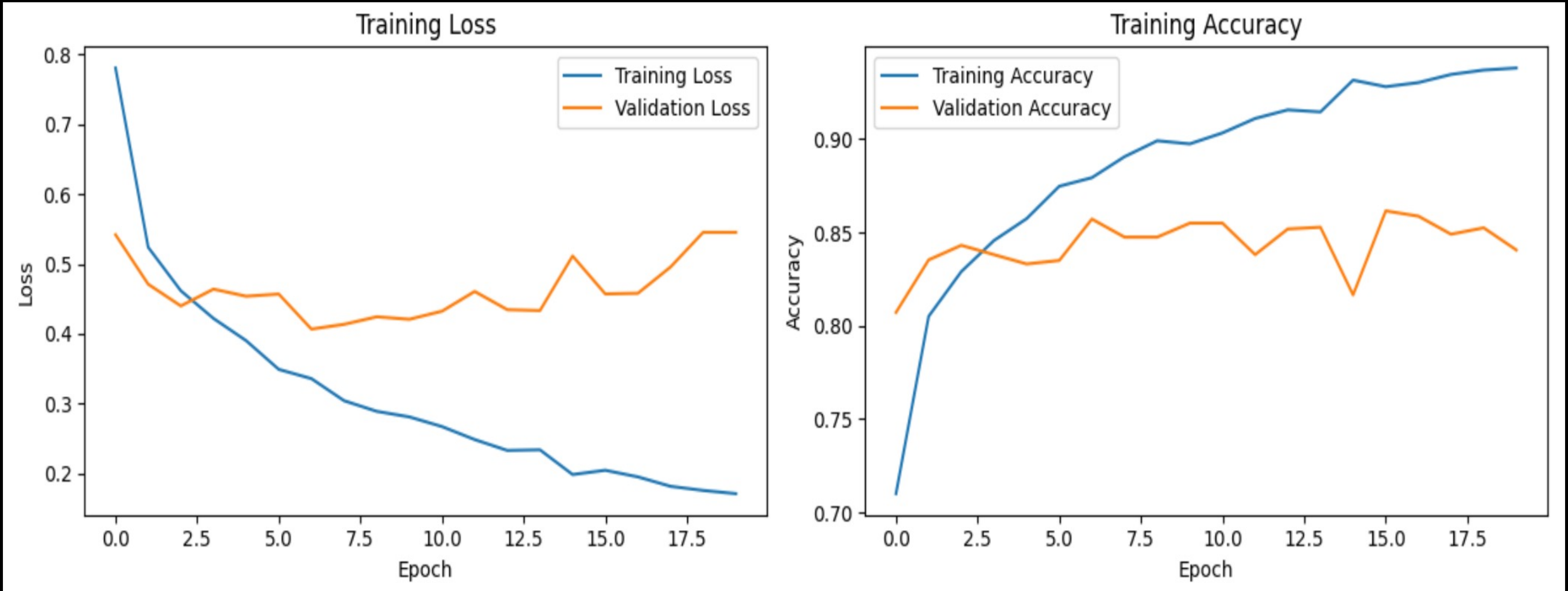
- optimizer=Adam(lr=0.001),
- loss='sparse_categorical_cross entropy'

Classification Report for Training Data:

	precision	recall	f1-score	support
0	0.87	0.99	0.92	3898
1	0.98	0.81	0.89	2942
2	0.86	0.82	0.84	2247
3	0.94	0.96	0.95	6912
accuracy			0.92	15999
macro avg	0.91	0.89	0.90	15999
weighted avg	0.92	0.92	0.92	15999

Classification Report:

	precision	recall	f1-score	support
0	0.77	0.94	0.85	962
1	0.91	0.70	0.79	758
2	0.70	0.62	0.66	572
3	0.89	0.90	0.90	1708
accuracy			0.83	4000
macro avg	0.82	0.79	0.80	4000
weighted avg	0.84	0.83	0.83	4000


Training Loss
Training Loss
Validation Loss
Loss
Epoch
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Training Accuracy
Training Accuracy
Validation Accuracy
Accuracy
Epoch
0.90
0.85
0.80
0.75
0.70
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

LE-NET MODEL

- optimizer='adam',
- loss='sparse_categorical_crossentropy'

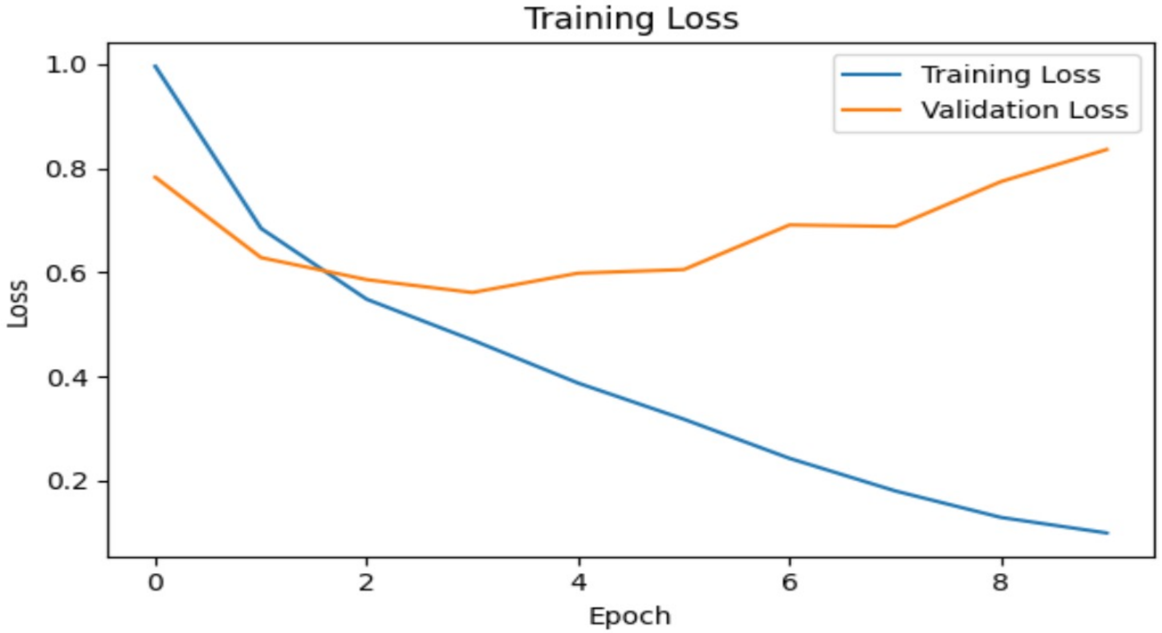
```
Model: "sequential_8"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d_16 (Conv2D)          (None, 124, 124, 6)       156

 max_pooling2d_16 (MaxPooli  (None, 62, 62, 6)         0
 ng2D)

 conv2d_17 (Conv2D)          (None, 58, 58, 16)        2416

 max_pooling2d_17 (MaxPooli  (None, 29, 29, 16)        0
 ng2D)

 flatten_9 (Flatten)         (None, 13456)             0

 dense_26 (Dense)            (None, 120)               1614840

 dense_27 (Dense)            (None, 84)                10164

 dense_28 (Dense)            (None, 4)                 340

=================================================================
Total params: 1627916 (6.21 MB)
Trainable params: 1627916 (6.21 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

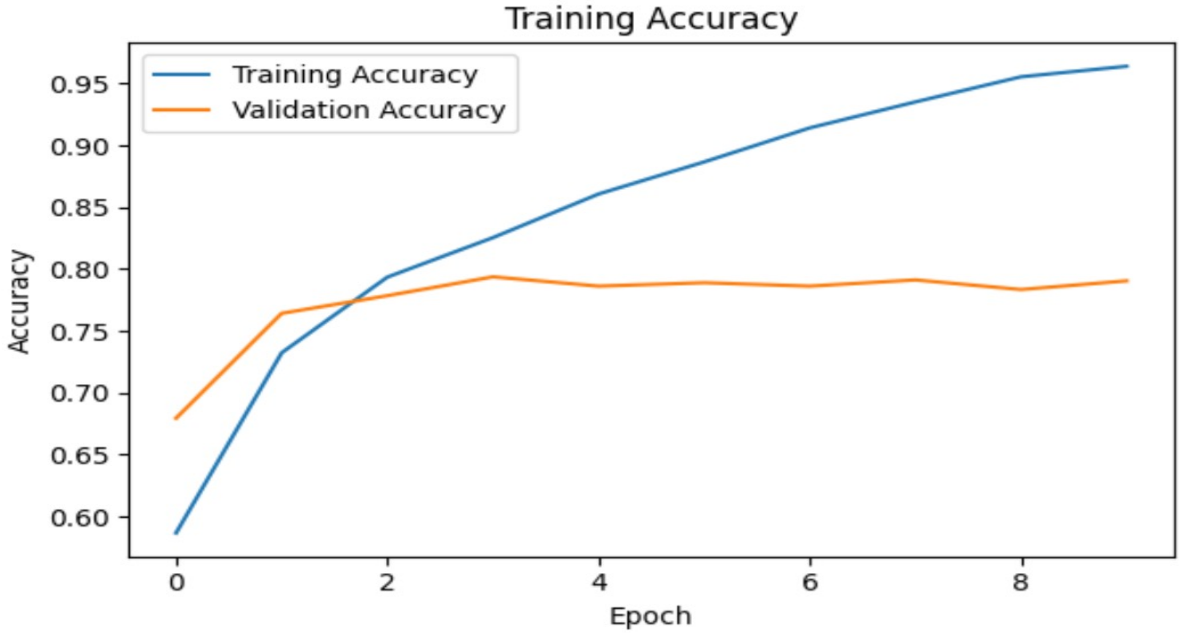
Classification Report for Train Set:

	precision	recall	f1-score	support
CNV	1.00	0.98	0.99	3898
DME	0.99	0.98	0.98	2942
DRUSEN	0.97	0.95	0.96	2247
NORMAL	0.98	1.00	0.99	6912
accuracy			0.98	15999
macro avg	0.98	0.98	0.98	15999
weighted avg	0.98	0.98	0.98	15999

Classification Report for Test Set:

	precision	recall	f1-score	support
CNV	0.85	0.82	0.84	962
DME	0.76	0.69	0.73	758
DRUSEN	0.58	0.58	0.58	572
NORMAL	0.84	0.89	0.86	1708
accuracy			0.79	4000
macro avg	0.76	0.75	0.75	4000
weighted avg	0.79	0.79	0.79	4000





CONVOLUTION NEURAL NETWORKS

- 3 Convolutional layers (conv2d)
- 3 Pooling layers (max_pooling2d)
- 1 Flatten layer
- 3 Dense layers (dense)

Model: "sequential_2"

Layer (type)	Output Shape	Param #
conv2d_6 (Conv2D)	(None, 254, 254, 32)	320
max_pooling2d_6 (MaxPooling2D)	(None, 127, 127, 32)	0
conv2d_7 (Conv2D)	(None, 125, 125, 64)	18496
max_pooling2d_7 (MaxPooling2D)	(None, 62, 62, 64)	0
conv2d_8 (Conv2D)	(None, 60, 60, 128)	73856
max_pooling2d_8 (MaxPooling2D)	(None, 30, 30, 128)	0
flatten_2 (Flatten)	(None, 115200)	0
dense_6 (Dense)	(None, 128)	14745728
dense_7 (Dense)	(None, 64)	8256
dense_8 (Dense)	(None, 4)	260

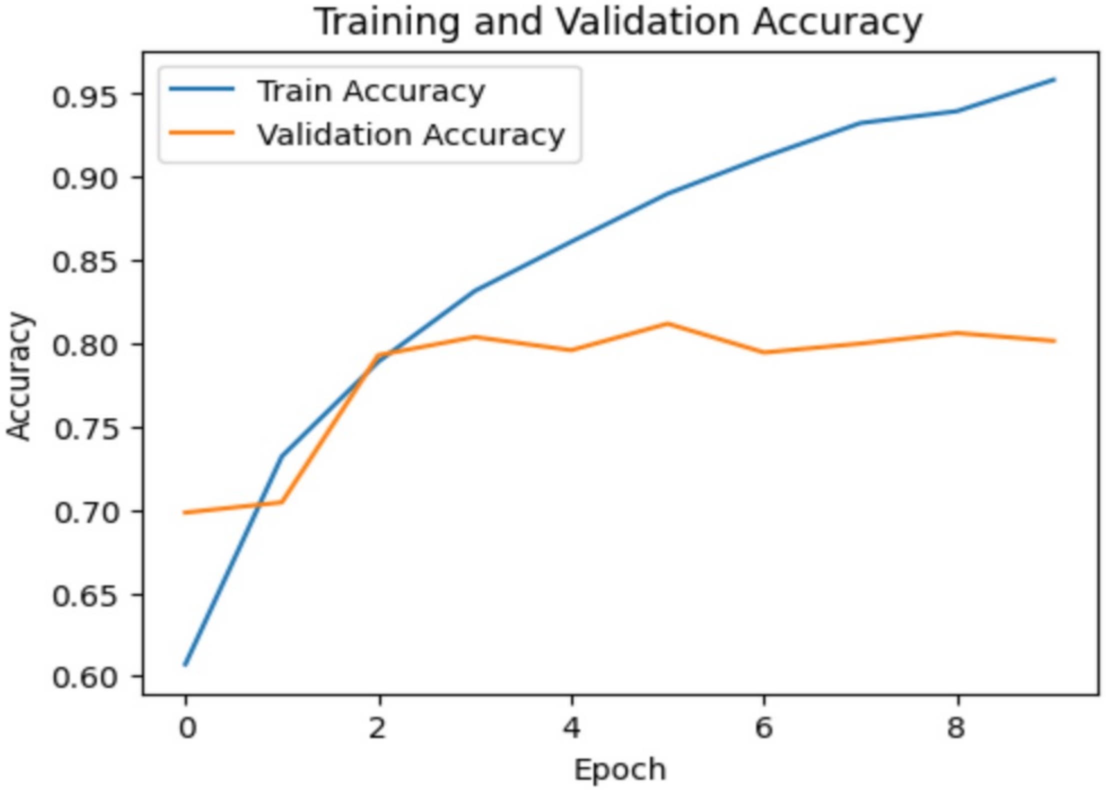
Total params: 14846916 (56.64 MB)
Trainable params: 14846916 (56.64 MB)
Non-trainable params: 0 (0.00 Byte)

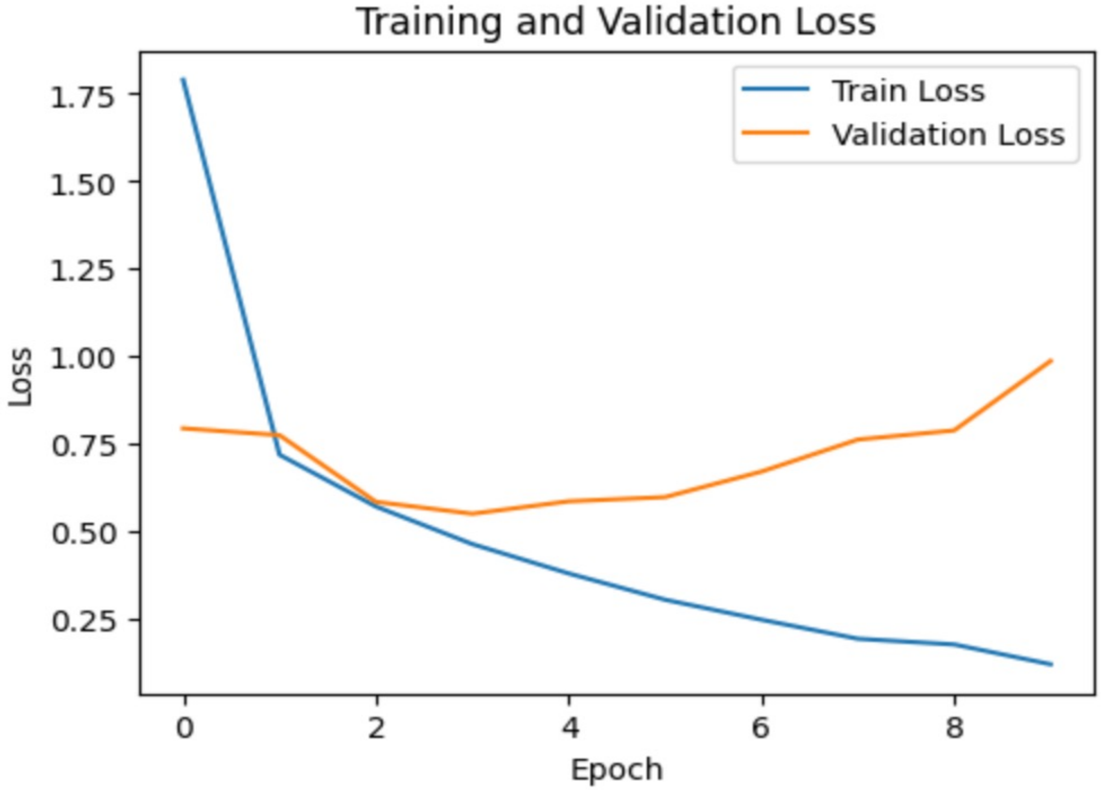
Train Classification Report:

	precision	recall	f1-score	support
CNV	0.98	0.98	0.98	10380
DME	0.98	0.95	0.96	7443
DRUSEN	0.90	0.95	0.92	6031
NORMAL	0.98	0.97	0.97	18420
accuracy			0.97	42274
macro avg	0.96	0.96	0.96	42274
weighted avg	0.97	0.97	0.97	42274

Test Classification Report:

	precision	recall	f1-score	support
CNV	0.81	0.83	0.82	4449
DME	0.81	0.73	0.77	3189
DRUSEN	0.57	0.59	0.58	2585
NORMAL	0.87	0.88	0.87	7895
accuracy			0.80	18118
macro avg	0.77	0.76	0.76	18118
weighted avg	0.80	0.80	0.80	18118





CONVOLUTION NEURAL NETWORKS

- 4 Convolutional layers (conv2d)
- 4 Pooling layers (max_pooling2d)
- 1 Flatten layer
- 3 Dense layers (dense)

- Shape of X_train: (42274, 256, 256)
- Shape of y_train: (42274,)
- Shape of X_test: (18118, 256, 256)
- Shape of y_test: (18118,)

Model: "sequential_19"

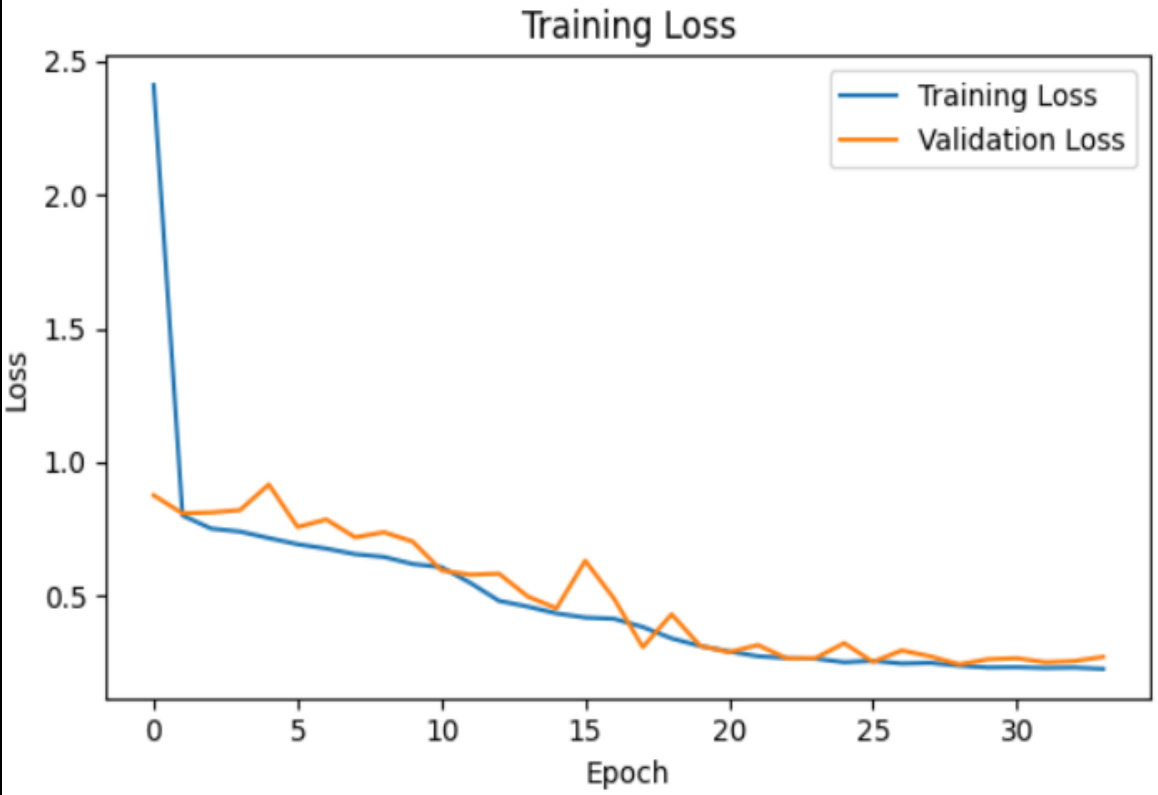
Layer (type)	Output Shape	Param #
conv2d_91 (Conv2D)	(None, 256, 256, 32)	320
max_pooling2d_41 (MaxPooling2D)	(None, 128, 128, 32)	0
conv2d_92 (Conv2D)	(None, 128, 128, 64)	18496
max_pooling2d_42 (MaxPooling2D)	(None, 64, 64, 64)	0
conv2d_93 (Conv2D)	(None, 64, 64, 128)	73856
max_pooling2d_43 (MaxPooling2D)	(None, 32, 32, 128)	0
conv2d_94 (Conv2D)	(None, 32, 32, 256)	295168
max_pooling2d_44 (MaxPooling2D)	(None, 16, 16, 256)	0
flatten_19 (Flatten)	(None, 65536)	0
dense_55 (Dense)	(None, 512)	33554944
dropout_1 (Dropout)	(None, 512)	0
dense_56 (Dense)	(None, 4)	2052

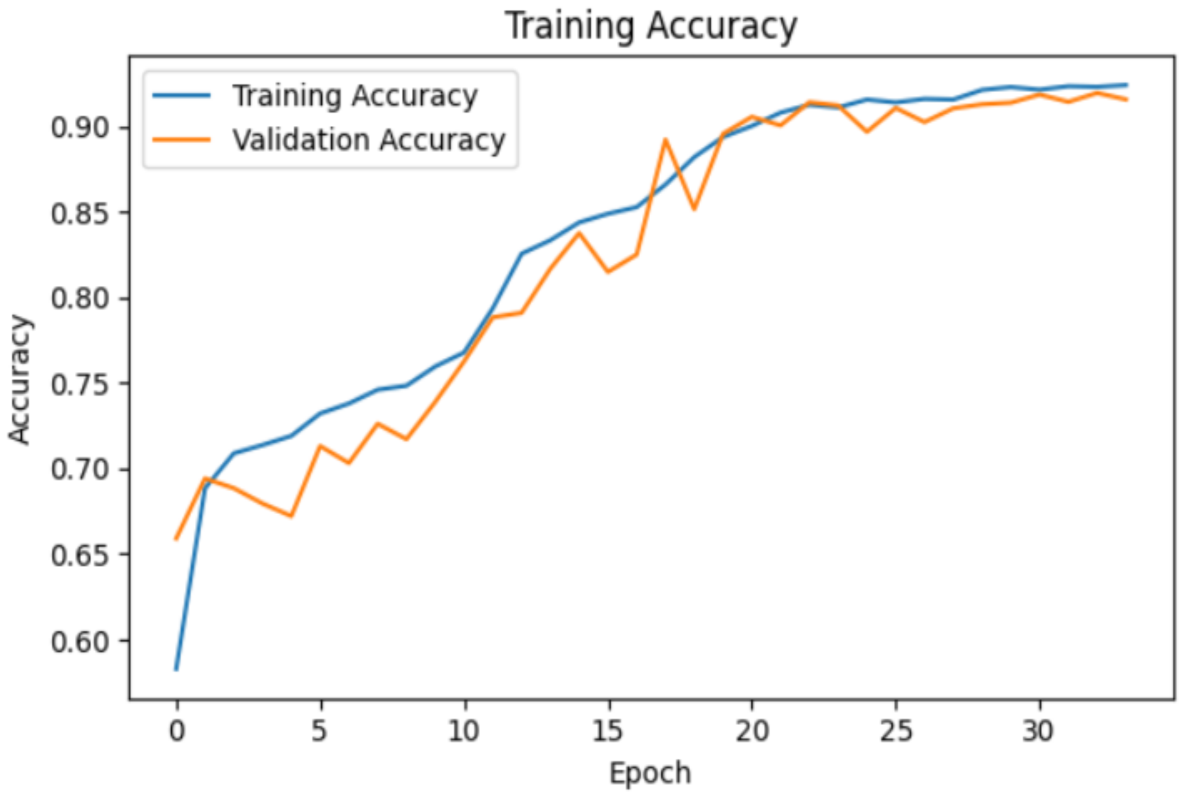
Train Classification Report:

	precision	recall	f1-score	support
0	0.88	0.96	0.92	9341
1	0.98	0.79	0.87	6680
2	0.93	0.78	0.84	5427
3	0.92	0.99	0.95	16598
accuracy			0.92	38046
macro avg	0.92	0.88	0.90	38046
weighted avg	0.92	0.92	0.91	38046

Test Classification Report:

	precision	recall	f1-score	support
0	0.87	0.96	0.91	4449
1	0.97	0.79	0.87	3189
2	0.93	0.77	0.84	2585
3	0.92	0.99	0.95	7895
accuracy			0.91	18118
macro avg	0.92	0.88	0.89	18118
weighted avg	0.92	0.91	0.91	18118





BEST PERFORMING MODEL- CNN

- 5 Convolutional layers (conv2d).
- 5 Pooling layers (max_pooling2d).
- 1 Flatten layer.
- 2 Dense layers (dense, dense_1)
- 1Dropout layer

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 256, 256, 32)	320
max_pooling2d (MaxPooling2D)	(None, 128, 128, 32)	0
conv2d_1 (Conv2D)	(None, 128, 128, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 64)	0
conv2d_2 (Conv2D)	(None, 64, 64, 128)	73856
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 128)	0
conv2d_3 (Conv2D)	(None, 32, 32, 256)	295168
max_pooling2d_3 (MaxPooling2D)	(None, 16, 16, 256)	0
conv2d_4 (Conv2D)	(None, 16, 16, 512)	1180160
max_pooling2d_4 (MaxPooling2D)	(None, 8, 8, 512)	0
flatten (Flatten)	(None, 32768)	0
dense (Dense)	(None, 512)	16777728
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 4)	2052

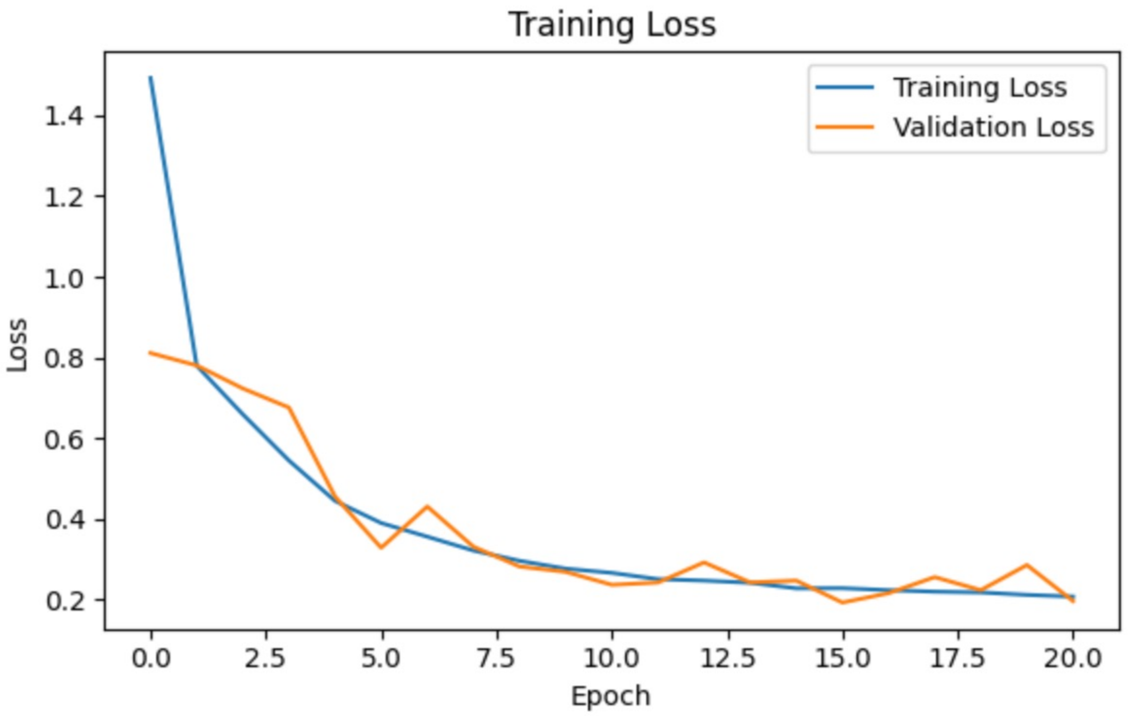
Total params: 18347780 (69.99 MB)
Trainable params: 18347780 (69.99 MB)
Non-trainable params: 0 (0.00 Byte)

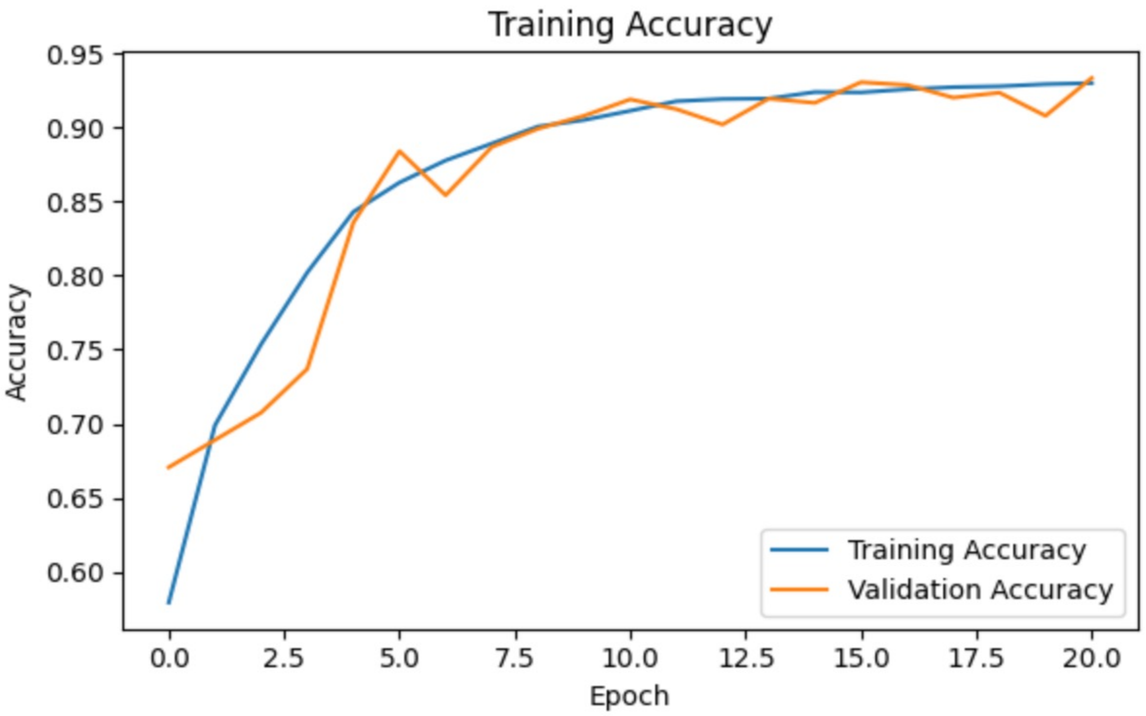
Train Classification Report:

	precision	recall	f1-score	support
0	0.93	0.93	0.93	9341
1	0.95	0.88	0.91	6680
2	0.89	0.86	0.88	5427
3	0.94	0.98	0.96	16598
accuracy			0.93	38046
macro avg	0.93	0.91	0.92	38046
weighted avg	0.93	0.93	0.93	38046

Test Classification Report:

	precision	recall	f1-score	support
0	0.93	0.92	0.93	4449
1	0.94	0.87	0.91	3189
2	0.88	0.86	0.87	2585
3	0.94	0.98	0.96	7895
accuracy			0.93	18118
macro avg	0.92	0.91	0.92	18118
weighted avg	0.93	0.93	0.93	18118





	CNV	DME	DRUSEN	NORMAL	Overall F-1 Score	
XGBoosting (n_estimators=200, learning_rate=0.01, max_depth = 5)						
Train	0.97	0.95	0.97	0.94	0.95	
Test	0.58	0.52	0.37	0.51	0.5	
Gradient Boosting (n_estimators=50, learning_rate=0.1)						
Train	0.92	0.91	0.92	0.91	0.91	
Test	0.56	0.49	0.38	0.45	0.46	
AdaBoosting (n_estimators=200,learning_rate=0.1)						
Train	0.6	0.56	0.53	0.61	0.57	Number of images = 20,000
Test	0.48	0.46	0.39	0.44	0.44	
K-Nearest Neighbors (n_neigbors=7, p=1)						
Train	0.55	0.55	0.53	0.56	0.54	
Test	0.25	0.4	0.35	0.33	0.33	
VGG16 (optimizer=Adam(lr =0.001), loss='sparse categorical cross-entropy')						
Train	0.92	0.89	0.84	0.95	0.9	
Test	0.85	0.79	0.66	0.9	0.8	
LeNet (optimizer='adam', loss='sparse categorical cross-entropy')						
Train	0.99	0.98	0.96	0.99	0.98	
Test	0.82	0.69	0.58	0.86	0.73	
CNN (3 Convolutional layers (conv2d), 3 Pooling layers (max_pooling2d), 1 Flatten layer, 3 Dense layers (dense))						
Train	0.98	0.96	0.92	0.97	0.95	Number of images = 62,138 (Entire Dataset)
Test	0.82	0.77	0.58	0.87	0.76	
CNN (5 Convolutional layers (conv2d), 5 Pooling layers (max_pooling2d), 1 Flatten layer, 2 Dense layers (dense, dense_1), 1Dropout layer)						
Train	**0.93**	**0.91**	**0.88**	**0.96**	**0.92**	
Test	**0.93**	**0.91**	**0.87**	**0.96**	**0.92**	

ADDRESSING OUR RESEARCH QUESTIONS

- How effective are deep learning models in accurately classifying retinal diseases based on OCT images?
- Which deep learning architectures prove most effective for OCT image classification?
- How does the size and diversity of the training dataset impact model robustness and generalization ability?
- What strategies can be employed to address class imbalance issues in the OCT image dataset?
- How do preprocessing techniques such as denoising, image registration, and normalization impact model performance and interpretability?

REFERENCES

1. Turbert, D. (2023, April 27). *What is optical coherence tomography?*. American Academy of Ophthalmology. https://www.aao.org/eye-health/treatments/what-is-optical-coherence-tomography
2. Riazi Esfahani, P., Reddy, A. J., Nawathey, N., Ghauri, M. S., Min, M., Wagh, H., Tak, N., & Patel, R. (2023, July 9). *Deep Learning Classification of drusen, choroidal neovascularization, and diabetic macular edema in Optical Coherence Tomography (OCT) images*. Cureus. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10411652/#:~:text=Age%2Drelated%20macular%20degeneration%20(AMD,tool%20for%20these%20ocular%20conditions
3. Kermany, D. (2018, January 6). *Labeled optical coherence tomography (OCT) and chest X-ray images for classification*. Mendeley Data. https://data.mendeley.com/datasets/rscbjbr9sj/2
4. Sharif, Z., & Sharif, W. (2019). *Corneal neovascularization: Updates on pathophysiology, Investigations* & *Management*. Romanian journal of ophthalmology. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6531773/
5. Yau, J. W. Y., Rogers, S. L., Kawasaki, R., Lamoureux, E. L., Kowalski, J. W., Bek, T., Chen, S.-J., Dekker, J. M., Fletcher, A., Grauslund, J., Haffner, S., Hamman, R. F., Ikram, M. K., Kayama, T., Klein, B. E. K., Klein, R., Krishnaiah, S., Mayurasakorn, K., O’Hare, J. P., … Wong, T. Y. (2012, February 10). *Global prevalence and major risk factors of diabetic retinopathy*. American Diabetes Association. https://diabetesjournals.org/care/article/35/3/556/28568/Global-Prevalence-and-Major-Risk-Factors-of
6. Johnson, P. T., Lewis, G. P., Talaga, K. C., Brown, M. N., Kappel, P. J., Fisher, S. K., Anderson, D. H., & Johnson, L. V. (2003, October 1). *Drusen-associated degeneration in the retina*. Investigative Ophthalmology & Visual Science. https://iovs.arvojournals.org/article.aspx?articleid=2182253