NYU Langone Health

Impact of Non-Visual Data on Chest Radiography Classification

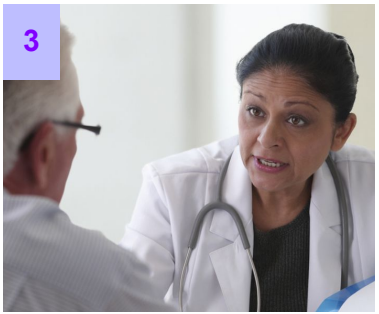
Deep Learning in Medicine
Spring 2024

Pratham Shah
Luke O'Donnell, MD

Introduction / Purpose

Project Hypothesis

Including non-visual data will improve image classification with the learning of additional features influencing the likelihood of a diagnosis

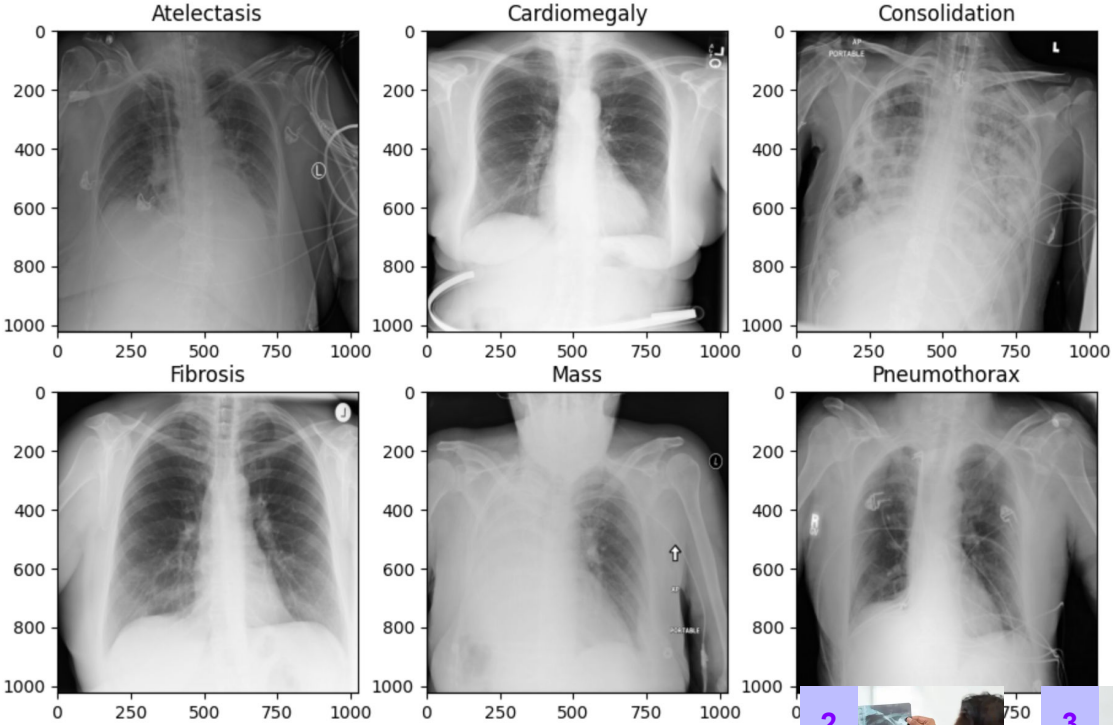


Diagnostic decisions do not happen in isolation

1. Physician evaluates a patient forming a pre-test probability for a condition
2. Diagnostic test performed
3. Physician forms post-test probability, the likelihood of condition based on pre-test probability and result of the diagnostic study

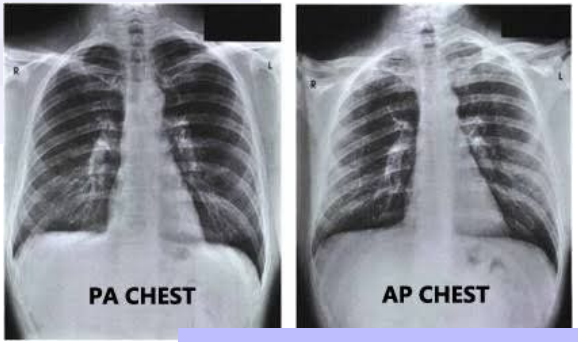
National Institutes of Health Chest Radiology Data

112,120 CXR Images in 1x1024x1024 PNG Format with 15 classifications that includes multi labels



Contributory Metadata

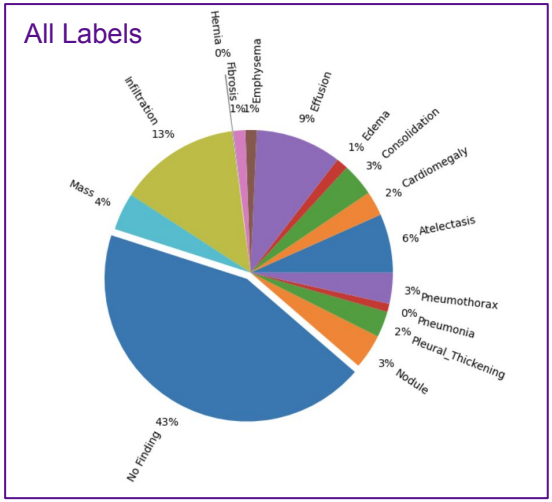
	image_name	age	age_normalized	sex	view	labels
0	00000001_000.png	58	0.368421	0	1	Cardiomegaly
23	00000008_000.png	69	0.440789	1	1	Cardiomegaly
34	00000011_006.png	75	0.480263	0	1	Atelectasis
49	00000013_011.png	60	0.381579	0	0	Pneumothorax
50	00000013_012.png	60	0.381579	0	0	Pneumothorax
...	...	...	...	...	...	...
112074	00030772_001.png	26	0.157895	1	0	Consolidation
112075	00030772_002.png	26	0.157895	1	0	Consolidation
112078	00030774_000.png	44				
112084	00030780_000.png	67				
112096	00030786_006.png	61				



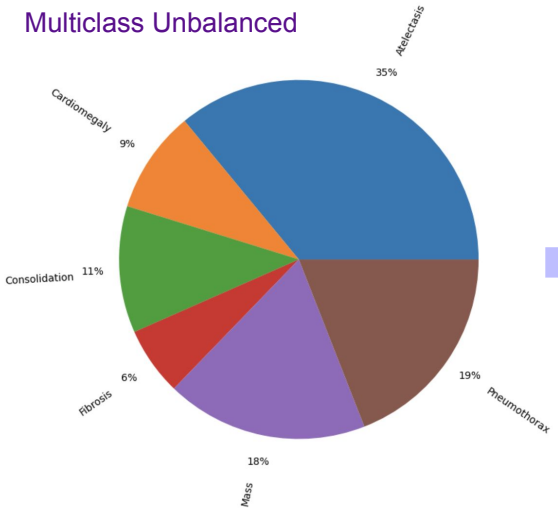
AP Images are usually performer is sicker hospitalized patients

Data Cleaning and Balancing

No Finding	60361
Infiltration	9547
Atelectasis	4215
Effusion	3955
Nodule	2705
Pneumothorax	2194
Mass	2139
Effusion|Infiltration	1603
Atelectasis|Infiltration	1350
Consolidation	1310
Atelectasis|Effusion	1165
Pleural_Thickening	1126
Cardiomegaly	1093
Emphysema	892
Infiltration|Nodule	829
Atelectasis|Effusion|Infiltration	737
Fibrosis	727
Edema	628
Cardiomegaly|Effusion	484
Consolidation|Infiltration	441

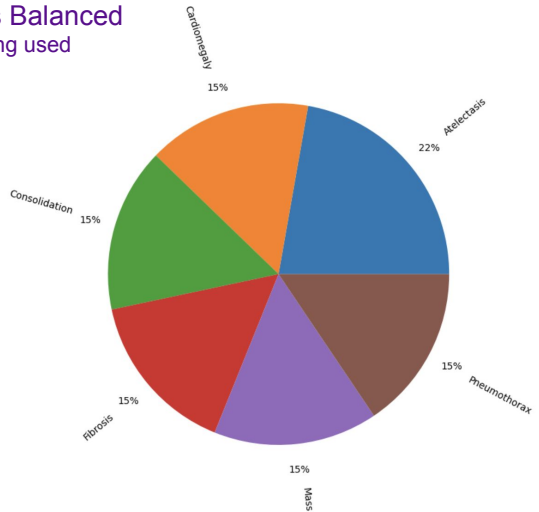


Multiclass Unbalanced



Multiclass Balanced

- Upsampling used



Materials and Methods

Two datasets

1. CXRs only
2. CXRs plus non-visual data

Models

1. Multilayer Perceptron
2. CNN
3. Combined CNN and Recurrent Neural Network (RNN)
4. Dynamic Affine Feature Map Transform (DAFT) CNN
5. InceptionResNetV2

Changes From Initial Proposal

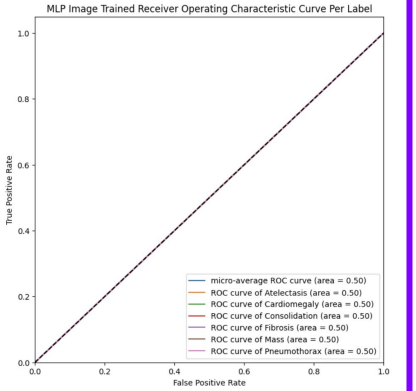
- Preformed multiclass, not multilabel classification

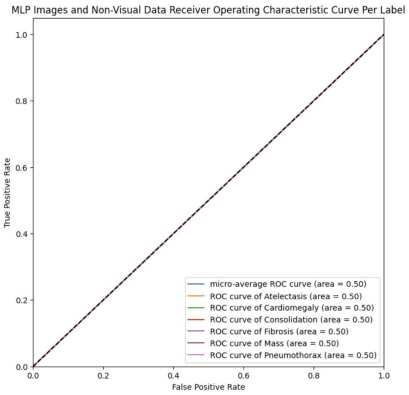
Libraries

```
import pandas as pd
import numpy as np
import os

from PIL import Image
import matplotlib.pyplot as plt

import time
from tqdm import tqdm
import random

import torch
from torch.utils.data import Dataset, DataLoader
import torch.nn as nn
import torch.nn.functional as F
import torch.optim as optim
import torchvision
import torchvision.transforms as transforms
from torchvision.io import read_image
from torchvision.utils import make_grid
from torchsummary import summary


from sklearn.preprocessing import MultiLabelBinarizer
from sklearn.metrics import f1_score
from sklearn.metrics import classification_report
from sklearn.metrics import roc_curve, auc

import timm
```

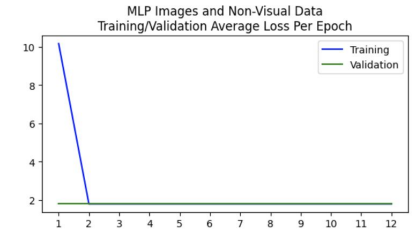
Multilayer Perceptron Model

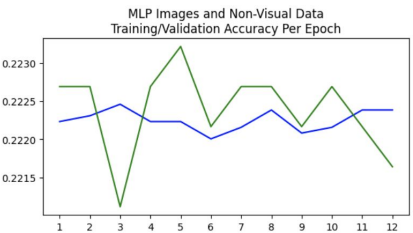
Images - Only

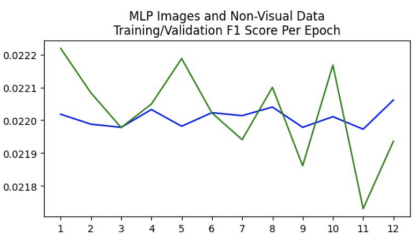




Images And Non Visual Data

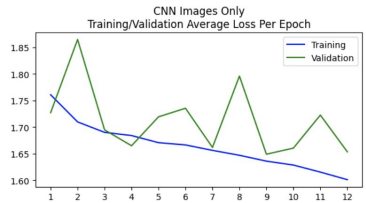


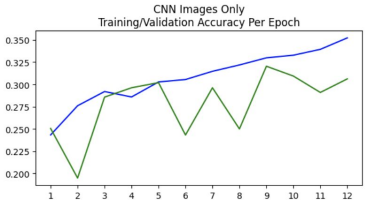


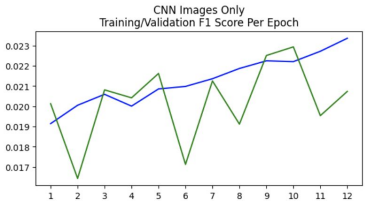


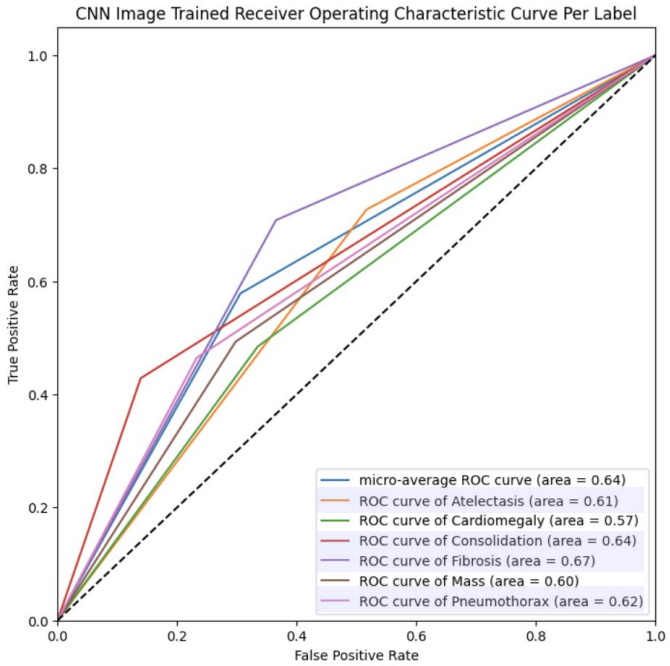
Convolutional Neural Network - Images Only

Layer (type)	Output Shape
Conv2d-1	[-1, 16, 511, 511]
BatchNorm2d-2	[-1, 16, 511, 511]
ReLU-3	[-1, 16, 511, 511]
Conv2d-4	[-1, 32, 255, 255]
BatchNorm2d-5	[-1, 32, 255, 255]
ReLU-6	[-1, 32, 255, 255]
Conv2d-7	[-1, 64, 85, 85]
BatchNorm2d-8	[-1, 64, 85, 85]
ReLU-9	[-1, 64, 85, 85]
Conv2d-10	[-1, 128, 29, 29]
BatchNorm2d-11	[-1, 128, 29, 29]
ReLU-12	[-1, 128, 29, 29]
Conv2d-13	[-1, 256, 27, 27]
BatchNorm2d-14	[-1, 256, 27, 27]
ReLU-15	[-1, 256, 27, 27]
AdaptiveAvgPool2d-16	[-1, 256, 1, 1]
Linear-17	[-1, 6]







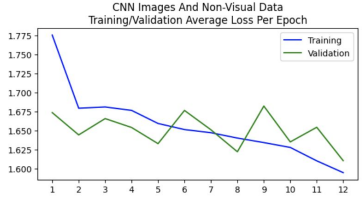


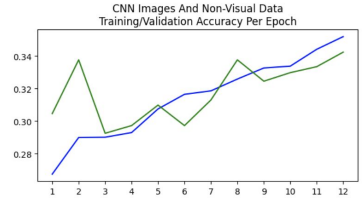
	precision	recall	f1-score	support
Atelectasis	0.45	0.73	0.55	850
Cardiomegaly	0.13	0.48	0.21	227
Consolidation	0.27	0.43	0.33	245
Fibrosis	0.11	0.71	0.19	144
Mass	0.28	0.49	0.36	446
Pneumothorax	0.31	0.46	0.37	424
micro avg	0.27	0.58	0.37	2336
macro avg	0.26	0.55	0.34	2336
weighted avg	0.32	0.58	0.40	2336
samples avg	0.28	0.58	0.37	2336

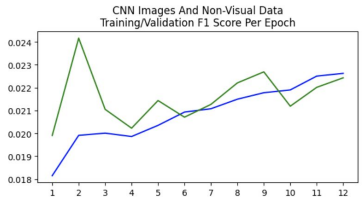
CNN - Images and Non-Visual Data

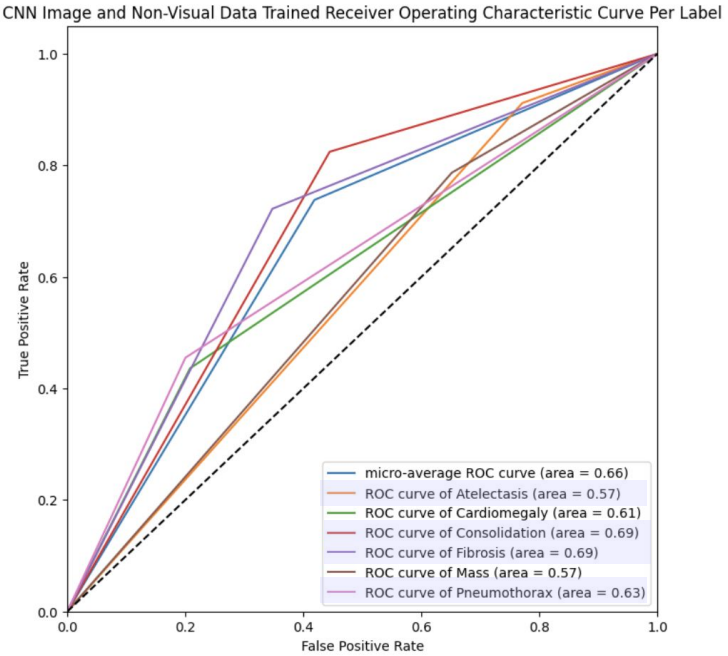
Architecture for Adding Non-Visual Data

```
def forward(self,x,non_visual_data):
    x = self.relu1(self.bn1(self.conv1(x)))
    x = self.relu2(self.bn2(self.conv2(x)))
    x = self.relu3(self.bn3(self.conv3(x)))
    x = self.relu4(self.bn4(self.conv4(x)))
    x = self.relu5(self.bn5(self.conv5(x)))
    x = self.avg(x)
    x = x.view(-1,256)

    z = non_visual_data.squeeze()
    z = z.view(x.shape[0],-1)
    x = torch.cat((x,z),1)

    x = self.bn6(x)
    x = self.linear(x)

    return x
```





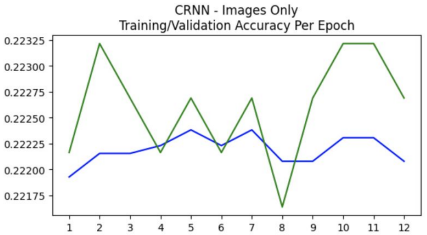


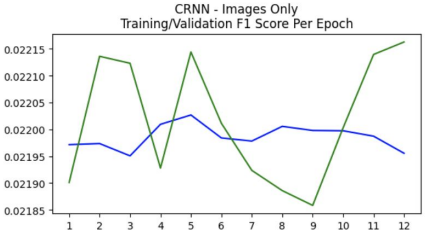


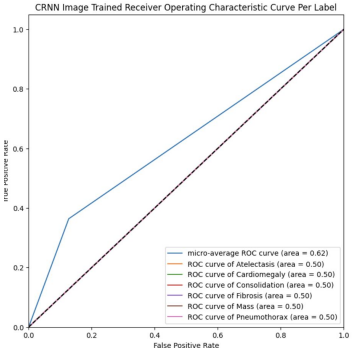
	precision	recall	f1-score	support
Atelectasis	0.40	0.91	0.56	850
Cardiomegaly	0.18	0.44	0.26	227
Consolidation	0.18	0.82	0.29	245
Fibrosis	0.12	0.72	0.21	144
Mass	0.22	0.79	0.35	446
Pneumothorax	0.34	0.46	0.39	424
micro avg	0.26	0.74	0.39	2336
macro avg	0.24	0.69	0.34	2336
weighted avg	0.29	0.74	0.41	2336
samples avg	0.27	0.74	0.39	2336

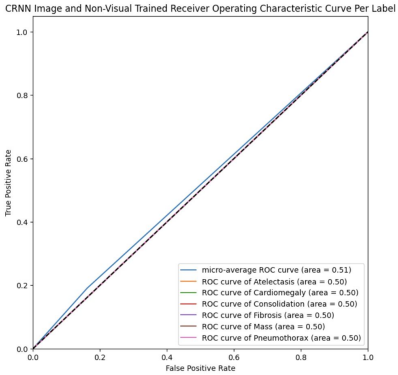
Combined CNN and Recurrent Neural Network (RNN)

Images - Only

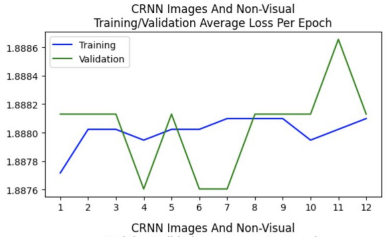


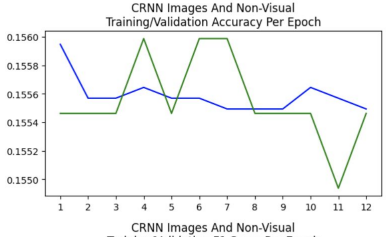


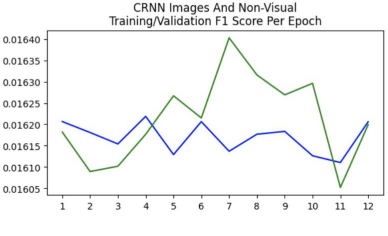




Images And Non Visual Data

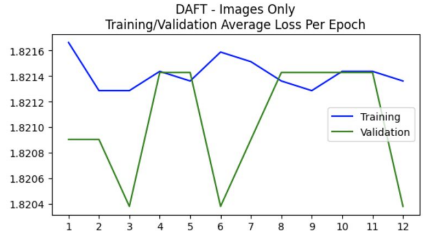


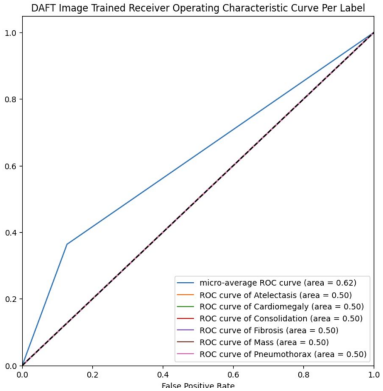


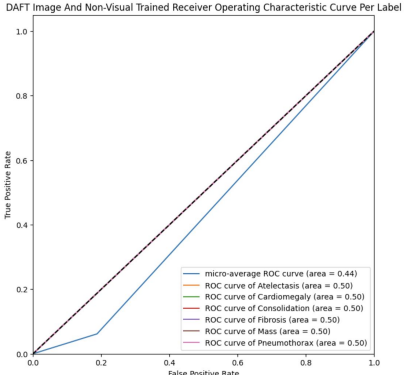


Dynamic Affine Feature Map Transform (DAFT) CNN

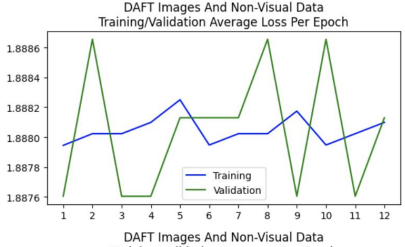
Images - Only

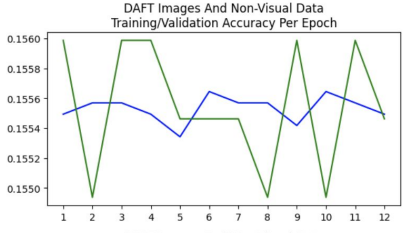


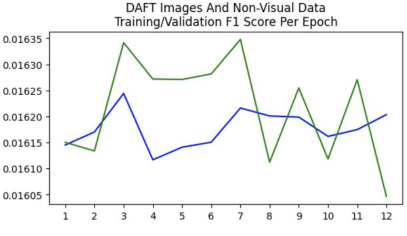




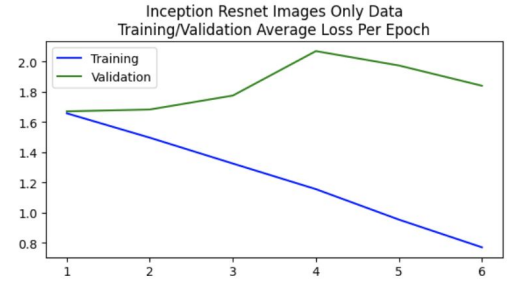
Images And Non Visual Data

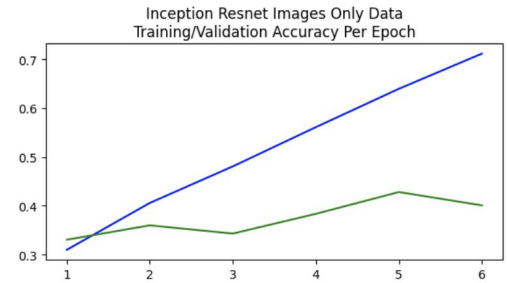


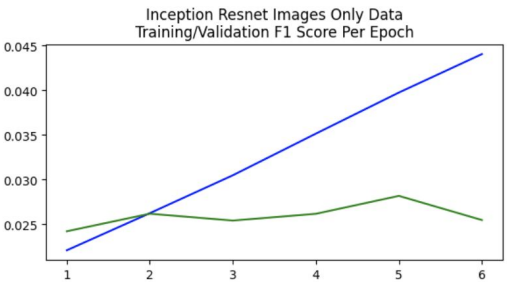


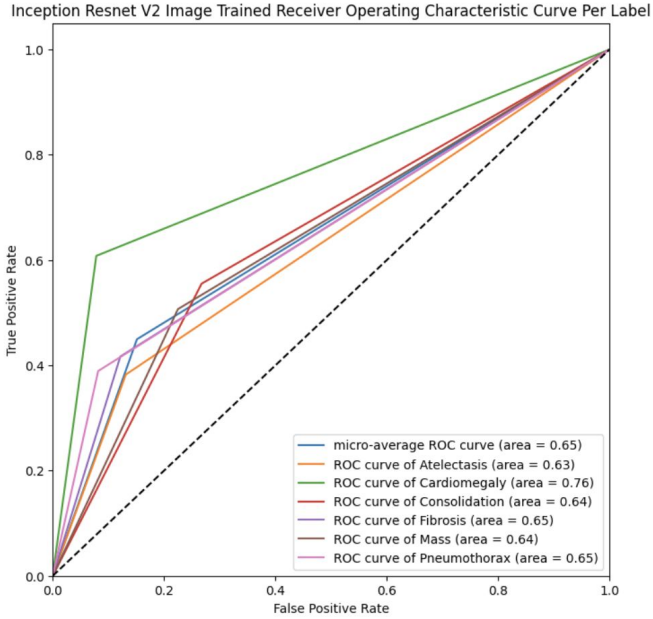


Inception Resnet v2 - Images Only









	precision	recall	f1-score	support
Atelectasis	0.62	0.38	0.47	850
Cardiomegaly	0.46	0.61	0.52	227
Consolidation	0.20	0.56	0.29	245
Fibrosis	0.18	0.42	0.26	144
Mass	0.35	0.51	0.41	446
Pneumothorax	0.51	0.39	0.44	424
micro avg	0.37	0.45	0.41	2336
macro avg	0.39	0.48	0.40	2336
weighted avg	0.46	0.45	0.43	2336
samples avg	0.32	0.45	0.36	2336

Inception Resnet v2

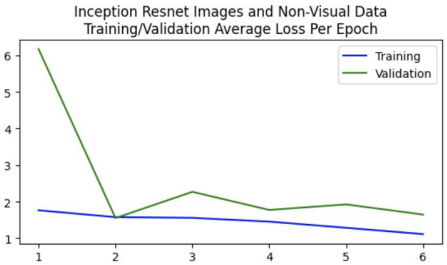
Images Only

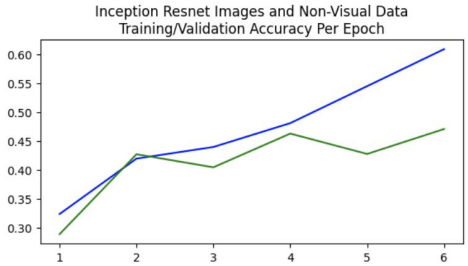
```
**********Epoch: 1**********
Training Phase:
100%|██████████| 826/826 [1:07:05<00:00,  4.87s/it]
Train set:
        Average Accuracy For Epoch: 0.3097003631961259
        Average Cross Entropy Loss for Epoch: 1.6575194590028204
        Average F1 for Epoch: 0.02208084347647203
Validation Phase:
100%|██████████| 119/119 [02:51<00:00,  1.44s/it]
Validation set:
        Average Accuracy: 0.33035714285714285
        Average Cross Entropy Loss: 1.6704451867512293
        Average F1 for Epoch: 0.02420807310392764
Time elapse: 01:09:56
**********Epoch: 2**********
Training Phase:
100%|██████████| 826/826 [1:06:11<00:00,  4.81s/it]
Train set:
        Average Accuracy For Epoch: 0.4057203389830508
        Average Cross Entropy Loss for Epoch: 1.4960555999244385
        Average F1 for Epoch: 0.026236713113930273
Validation Phase:
100%|██████████| 119/119 [02:48<00:00,  1.42s/it]
Validation set:
        Average Accuracy: 0.3597689075630252
        Average Cross Entropy Loss: 1.682669072591958
        Average F1 for Epoch: 0.02617780690814821
Time elapse: 02:18:55
**********Epoch: 3**********
Training Phase:
100%|██████████| 826/826 [1:05:20<00:00,  4.75s/it]
Train set:
        Average Accuracy For Epoch: 0.480705205811138
        Average Cross Entropy Loss for Epoch: 1.3250154181052063
        Average F1 for Epoch: 0.030482550128404718
Validation Phase:
100%|██████████| 119/119 [02:53<00:00,  1.46s/it]
Validation set:
        Average Accuracy: 0.3429621848739496
        Average Cross Entropy Loss: 1.7748976780586885
        Average F1 for Epoch: 0.025410558110147596
Time elapse: 03:27:09
```

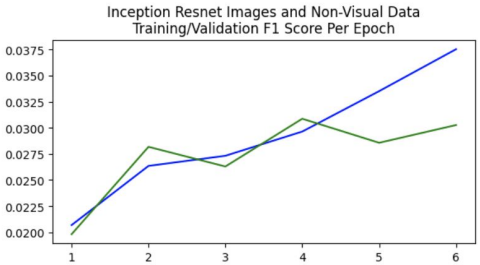
Images And Non-Visual Data

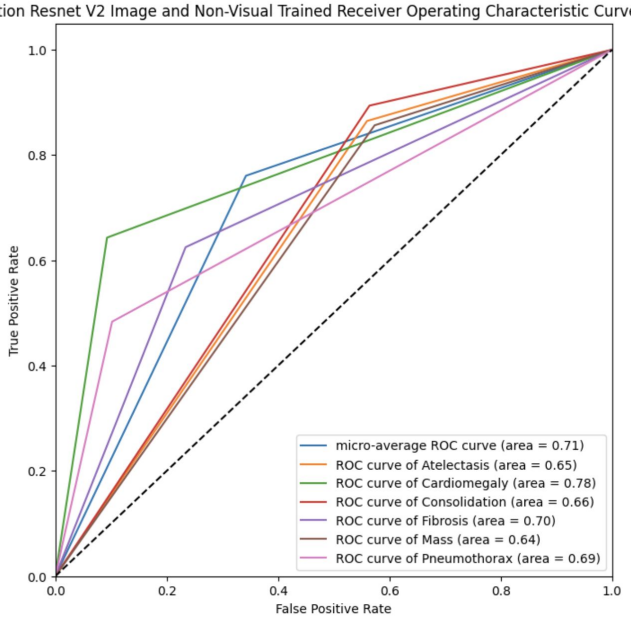
```
**********Epoch: 1**********
Training Phase:
100%|██████████| 826/826 [1:08:39<00:00,  4.99s/it]
Train set:
        Average Accuracy For Epoch: 0.32369854721549635
        Average Cross Entropy Loss for Epoch: 1.7645719561536433
        Average F1 for Epoch: 0.020688274431634197
Validation Phase:
100%|██████████| 119/119 [02:59<00:00,  1.51s/it]
Validation set:
        Average Accuracy: 0.28886554621848737
        Average Cross Entropy Loss: 6.166285808346853
        Average F1 for Epoch: 0.01981405756247034
Time elapse: 01:11:39
**********Epoch: 2**********
Training Phase:
100%|██████████| 826/826 [1:04:48<00:00,  4.71s/it]
Train set:
        Average Accuracy For Epoch: 0.4194915254237288
        Average Cross Entropy Loss for Epoch: 1.5784103531237088
        Average F1 for Epoch: 0.026353357420340617
Validation Phase:
100%|██████████| 119/119 [02:00<00:00,  1.01s/it]
Validation set:
        Average Accuracy: 0.4269957983193277
        Average Cross Entropy Loss: 1.5526511453780807
        Average F1 for Epoch: 0.02818086751746153
Time elapse: 02:18:28
**********Epoch: 3**********
Training Phase:
100%|██████████| 826/826 [1:03:24<00:00,  4.61s/it]
Train set:
        Average Accuracy For Epoch: 0.43954297820823246
        Average Cross Entropy Loss for Epoch: 1.5585870971546911
        Average F1 for Epoch: 0.027322922256147176
Validation Phase:
100%|██████████| 119/119 [02:06<00:00,  1.06s/it]
Validation set:
        Average Accuracy: 0.40441176470588236
        Average Cross Entropy Loss: 2.269154444462111
        Average F1 for Epoch: 0.026296803959783233
Time elapse: 03:23:59
```

Inception Resnet v2 - Images and Non-Visual Data









	precision	recall	f1-score	support
Atelectasis	0.47	0.86	0.61	850
Cardiomegaly	0.43	0.64	0.51	227
Consolidation	0.16	0.89	0.27	245
Fibrosis	0.15	0.62	0.24	144
Mass	0.26	0.86	0.40	446
Pneumothorax	0.51	0.48	0.50	424
micro avg	0.31	0.76	0.44	2336
macro avg	0.33	0.73	0.42	2336
weighted avg	0.38	0.76	0.48	2336
samples avg	0.34	0.76	0.45	2336

Discussion

- Results on training resized sample set images did not translate into full-sized full dataset training
- Significant difficulty with multilabel classification, opted for multiclass and selected discordant pathologies with similar levels of samples
- Imbalance datasets with classification
    - Accuracy vs F1 score
- CNN with non-visual data added and batch normalized for last day did perform better
- RNN and DAFT performed worse, and worsened when non-visual data add
    - Global understanding unneeded for CXR classification?
- NIH CXR dataset maybe best for normal/abnormal classification given roughly 50/50 breakdown of dataset
- Recognized similar challenges to those mentioned in Dr. Krysztof Geras lecture
    - Natural images vs medical images
    - Sufficient data size for multiclass/multilabel
    - Need for large images, very slow processing time

NYU Langone Health

Questions?