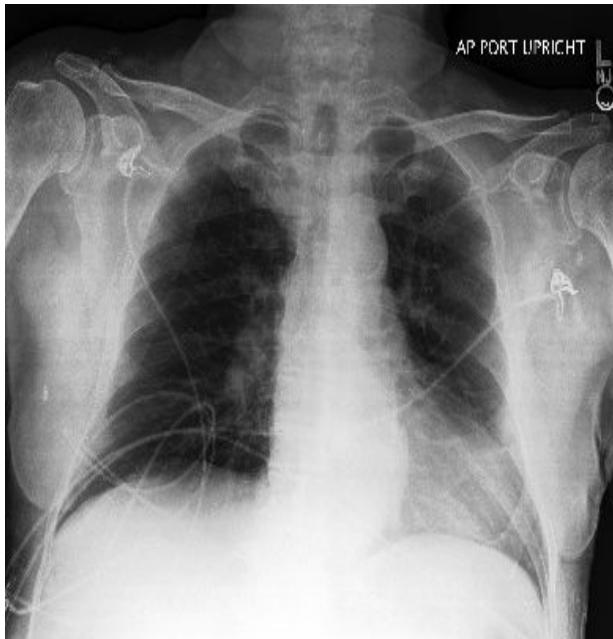


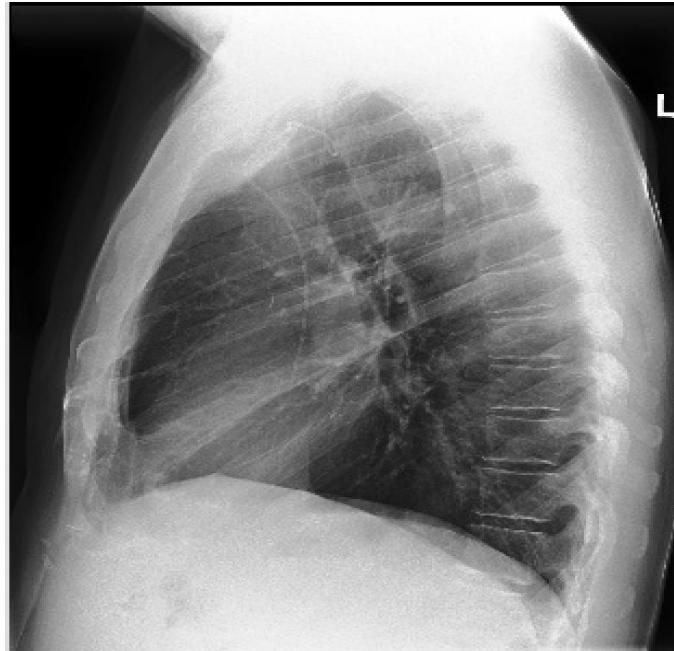
BD4H X Ray Classification Final Project Presentation

Ahmad Khan

Goal: Correctly predict the existence of 14 pathologies in X Ray Images



Front View X Ray



Lateral View X Ray

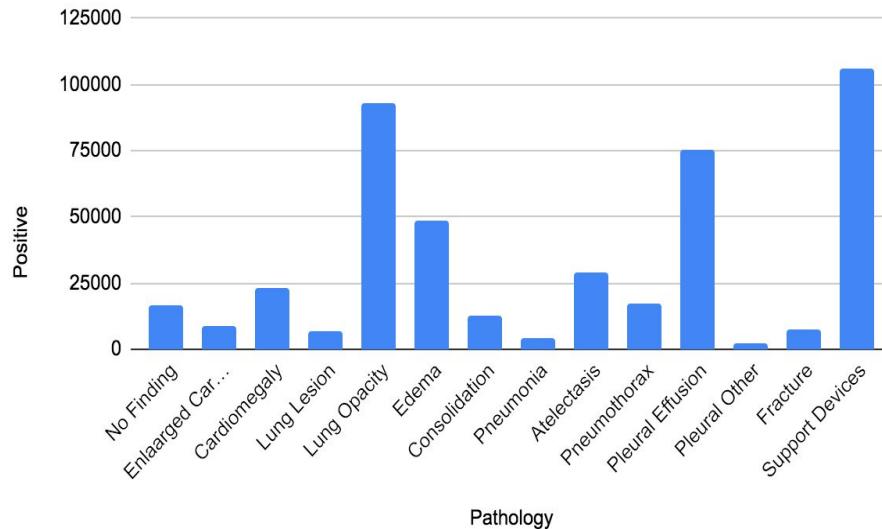
Data

224K images
from ChexPert
data with front
and lateral views
of patients

224 images with
radiologist
ground truth for
validation

14 Disease Pathologies with 3 labels (Positive, Negative, Uncertain)

Positive Occurrences vs Pathology



2 Classification Approaches utilized

- a) **Binary 'Ones' Classification** - Treat Uncertain labels as Positive
- b) **3 Class Multi-Classification** - Treat Uncertain as its own label. Predict for all class types and use Positive scores as final prediction

Binary Cross Entropy used as Loss Function for both with Adam Optimizer

4 Modeling Approaches Investigated.

First used pretrained general architectures:

- 1) ResNet 152 (Kaiming et al)
- 2) DenseNet 121 (baseline model) (Irvin et al, Huang et al)

Next used Attention Learning architectures:

- 3) Attention Guided Convolutional Neural Network (Guan et al 2018)
- 4) Category Wide Residual Attention Learning (Guan et al)
 - a) Attention Based Learning
 - b) Attention Based Learning using Residual Attention Network (Wang et al)

Baseline Model: Pretrained General Architectures

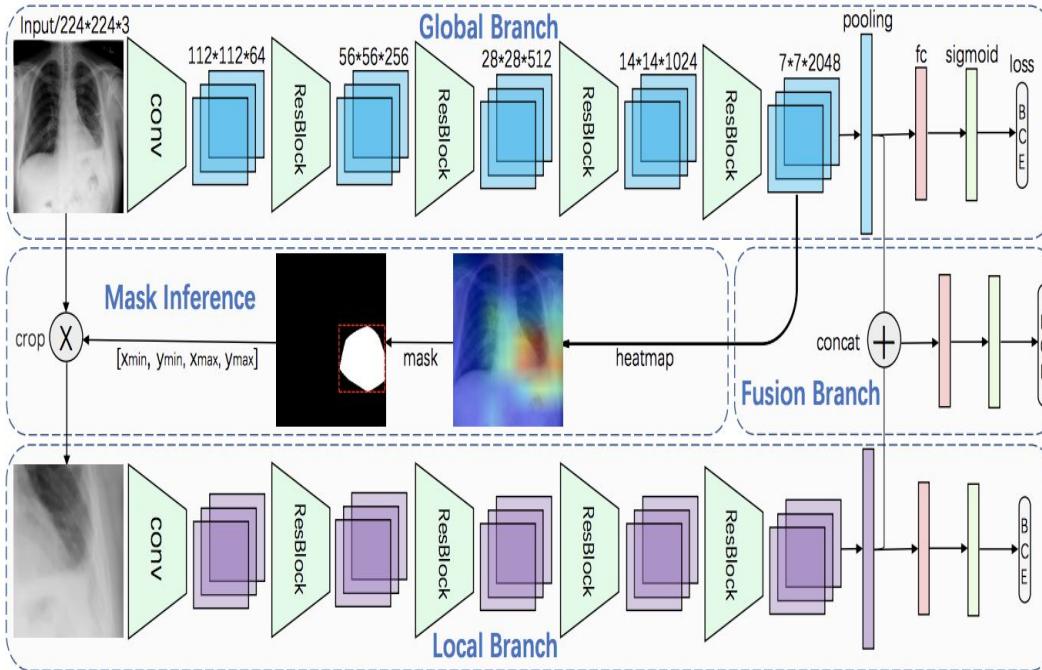
Mean Class AUC Results for Each Pretrained Architecture Model (DenseNet121 retained as Baseline since it performed better)

Model	Mean Validation AUC	Mean Validation Accuracy	Mean Validation Loss
Densenet121 with no fixed weights (Baseline)	0.776	85.13%	0.357
ResNet152 with no fixed weights	0.769	84.76%	0.347

1. Using pre-trained architectures on ImageNet the DenseNet 121 and ResNet 152 models were again retrained on ChexPert data and performance noted on validation set of 224 images
2. Weights were not fixed as fixing weights caused a decline in performance
3. DenseNet 121 performed slightly better on average 14 class AUC using Ones Classification approach and was retained as the **baseline model**

Attention Learning with Attention Guided CNN (AG-CNN)

AG-CNN Architecture (Credit: AG-CNN Paper)

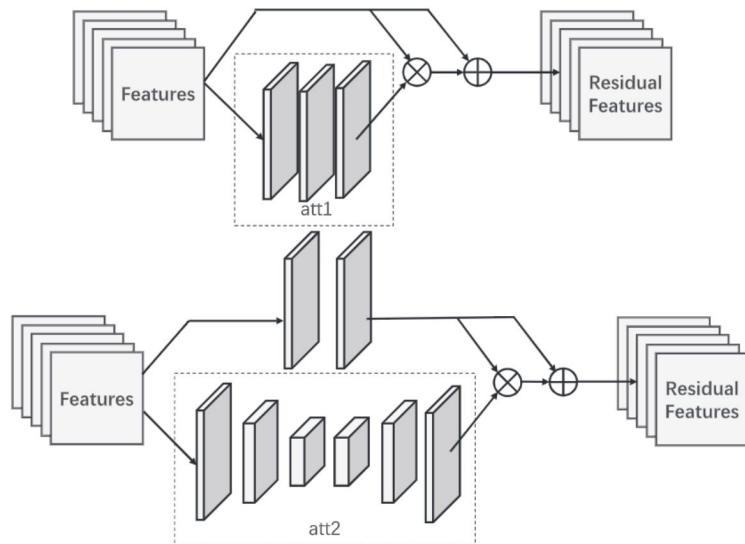


AG-CNN Architecture

1. First global branch learns global features using DenseNet 121 on entire image
2. Thresholding masked operation on Global feature map done to extract smaller cropped high activated heatmap
3. DenseNet 121 again trained locally on only this image subset
4. Predictions from both Global and Local branch Pooled and then combined in Fusion branch for final prediction

Category Wise Residual Attention Learning (CRAL)

Attention Module architectures used (Credit: CRAL Paper)



Att1 attention module : 2 3x3 conv layers each with
Relu activation followed by 1x1 convolution and a
sigmoid activation

CRAL Architecture

1. First global branch learns global features using DenseNet 121 on entire image
2. Then global feature maps from last convolutional layer passed into an attention module (2 types of attention modules both shown below) to emphasize good features for each class
3. Attention module acts as feature selector enhancing good features while suppressing bad ones
4. Attention Maps then combined with Global Branch feature map using hadamard product and addition

Model Results: One Classification Approach

ClassName	DenseNet-121 Ones (Baseline)	CRAL-ATT1 Ones	CRAL-ATT2 Ones	AG-CNN Ones	Irvin et al Comparison
No_Finding	0.863	0.861	0.868	0.879	
Enlarged Cardio	0.528	0.548	0.431	0.443	
Cardiomegaly	0.817	0.797	0.744	0.822	0.832
Lung_Opacity	0.900	0.896	0.883	0.898	
Lung_Lesion	0.262	0.270	0.296	0.562	
Edema	0.931	0.934	0.926	0.933	0.941
Consolidation	0.899	0.896	0.823	0.879	0.899
Pneumonia	0.864	0.767	0.700	0.800	
Atelectasis	0.810	0.810	0.762	0.848	0.858
Pneumothorax	0.843	0.851	0.920	0.742	
Pleural_Effusion	0.921	0.919	0.918	0.926	0.934
Pleural_Other	0.790	0.824	0.742	0.888	
Support_Devices	0.936	0.924	0.932	0.939	
Average AUC	0.797	0.792	0.765	0.812	
5 Class AUC	0.875	0.871	0.834	0.882	0.893

1. AG-CNN outperforms DenseNet 121 and CRAL on Average AUC and 5 Class AUC and most diseases
2. CRAL does best for certain diseases like Edema (ATT1) and Pneumothorax (ATT2)
3. No Model however outperforms for each individual disease pathology though

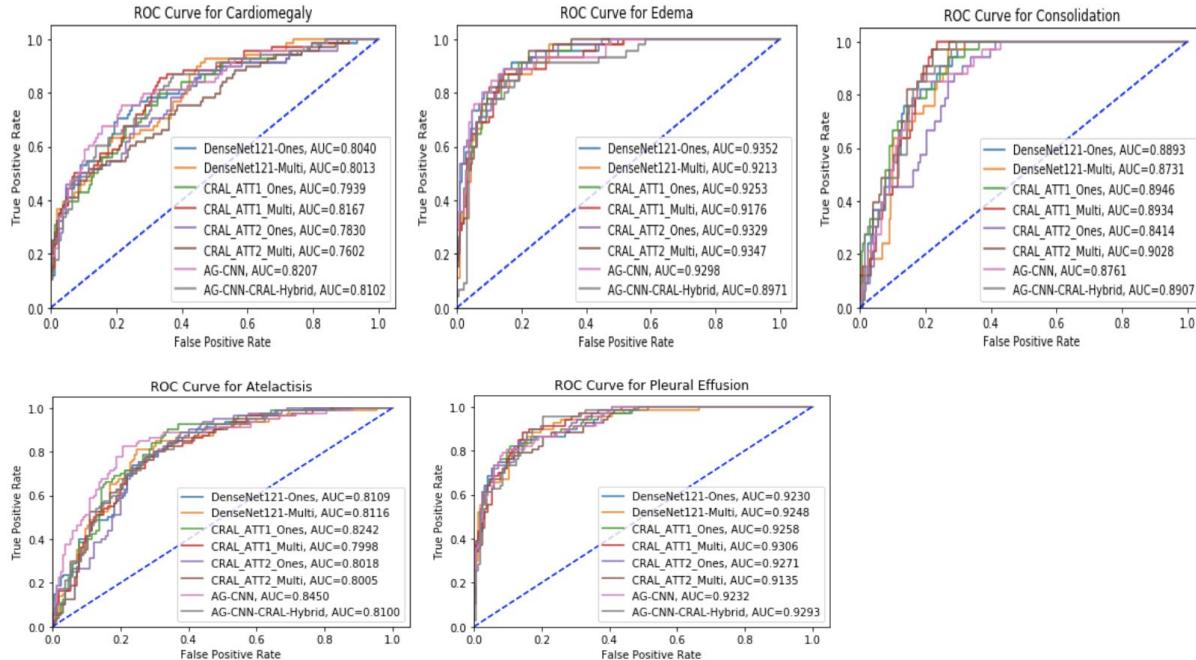
Model Results: 3 Class Classification Approach

ClassName	DenseNet 121 Multi	CRAL-ATT1 Multi	CRAL-ATT2 Multi	AG CNN (Hybrid) Multi	Irvin et al Comparison
No_Finding	0.8471	0.900	0.867	0.909	
Enlarged Cardio	0.4868	0.522	0.569	0.492	
Cardiomegaly	0.8013	0.817	0.760	0.810	0.854
Lung_Opacity	0.9079	0.913	0.911	0.903	
Lung_Lesion	0.3605	0.202	0.202	0.446	
Edema	0.9213	0.918	0.935	0.897	0.928
Consolidation	0.8731	0.893	0.903	0.891	0.937
Pneumonia	0.6831	0.748	0.752	0.775	
Atelectasis	0.8116	0.800	0.800	0.810	0.821
Pneumothorax	0.8451	0.858	0.888	0.799	
Pleural_Effusion	0.9248	0.931	0.913	0.929	0.936
Pleural_Other	0.9485	0.893	0.811	0.871	
Support_Devices	0.9073	0.919	0.919	0.928	
Average AUC	0.794	0.793	0.787	0.805	
5 Class AUC	0.866	0.872	0.862	0.867	0.895

1. AG-CNN outperforms DenseNet 121 and CRAL on 5 Class AUC but not by as much in the Ones case
2. No Model outperforms for each individual disease pathology though
3. 3 Class models generally don't do as well as the Ones Classification model

Model ROC Curves

AUC curve for the 5 major disease curves with each model results shown (Cardiomegaly, Edema, Consolidation, Atelectasis, Pleural Effusion)



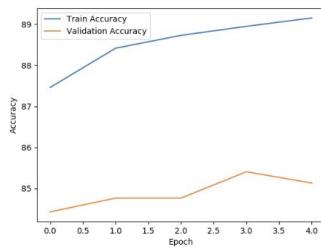
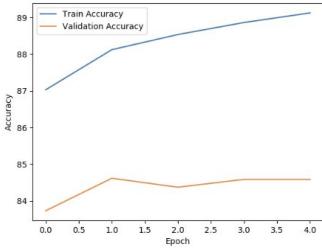
1. AG-CNN outperforms DenseNet 121 and CRAL on Average AUC and 5 Class AUC but not by as much in the Ones case
2. No Model outperforms for each individual disease pathology though
3. 3 Class models generally don't do as well as the Ones Classification model

3 Class Classification: Overfitting Observed

Figure 7: Three Class Classification Accuracy by Epoch (shows much greater divergence between training and validation accuracy)

DenseNet 121

Cral ATT1 Model



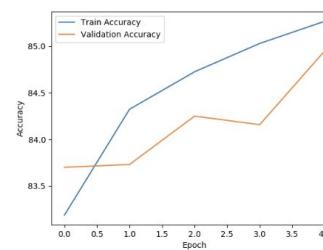
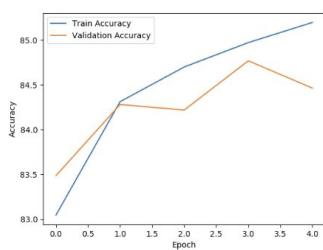
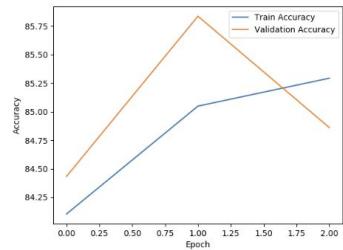
However, for the Ones classification approach models this overfitting was not as severe as the difference between training and validation accuracy was much smaller as shown in the learning curves below in Figure 8 for each model and each epoch.

Figure 8: Ones Classification Accuracy by Epoch

AGCNN Model

DenseNet 121 Model

CRAL ATT1 Model



Greater divergence in training and validation accuracy for 3 class approach which can suggest why 3 Class approach doesn't perform better than Ones for many disease types

Challenges/Areas of Improvement

1. Slow training time and memory bottleneck. training had to be stopped early after only 5 epochs or so and a parameter grid search for each model was not done as it would have taken too much time.
2. Averaging model performance over multiple runs could also lead to less variance in predictions and more stable results.
3. AG-CNN and CRAL papers did not explicitly mention some network architecture parameters that could be critical in model performance
 - a. In AG-CNN the cutoff threshold was mentioned to be 0.7 but it was not clear if this implied 70% percentile value or absolute value
 - b. CRAL paper it was not specified how many filters would be in the convolutional layers of the attention module and had to be guessed

Conclusion

1. Attention Learning can be utilized to improve model performance as seen by the superior performance of AG-CNN model (and also CRAL for certain diseases)
2. However, no model outperforms every model on every disease pathology. Some diseases are better suited than others for attention architectures especially those that occur locally
3. 3 Class classification approach better for some diseases vs others but not for all disease types