Deep-Learning Based Predictive Models for Chest X-ray Diagnosis

Hoseinali (Ali) Borhan
hborhan@stanford.edu
Recorded presentation: https://www.youtube.com/watch?v=MS67Bbuo2RA

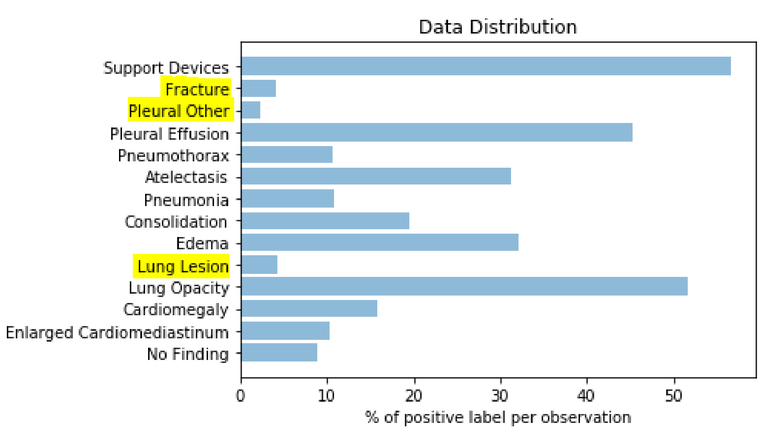
ABSTRACT AND OBJECTIVES

Predictive multi-class models are trained for **chest x-ray diagnosis of 14 observations** using **different deep learning architectures** and a large dataset of x-ray images (**CheXpert**).

- VGG-16, ResNet-50, and **DenseNet-121** are trained on an Amazon AWS EC2 GPU instance with Keras.
- For DenseNet-121, both **transfer learning** and **full training** are applied.
- Error analysis of the data indicating **unbalance dataset**
- **Up-sampling** data to fix unbalance issue leading to significant improvement of F1 scores over test data
- Gradient weighted **Class Activation Map** application

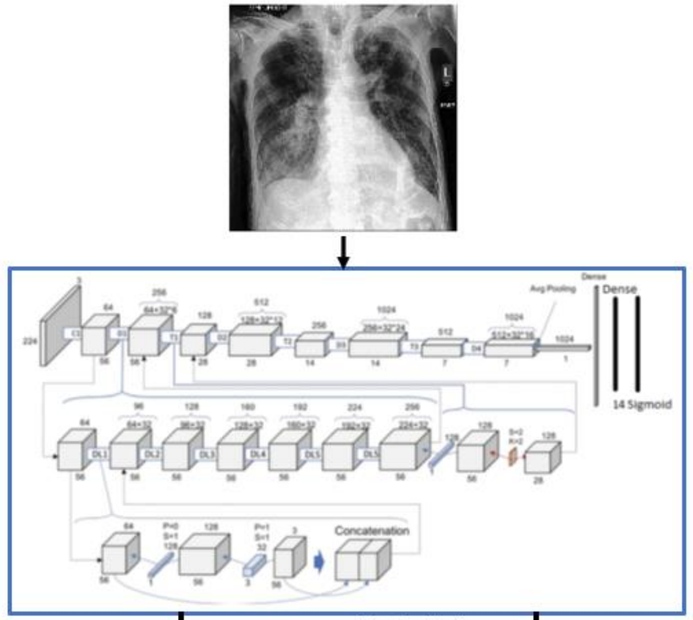
DATA AND FEATURES

- CheXpert, a large public dataset for chest radiograph interpretation, consisting of 224,316 chest radiographs of 65,240 patients.
- Features/Model Inputs: Chest X-Ray Data
- Model Outputs: 14 class observations prediction
- Model Category: Computer Vision for Multi Classification
- Data split: train/development (10%)/test(10%)
- Data distribution:

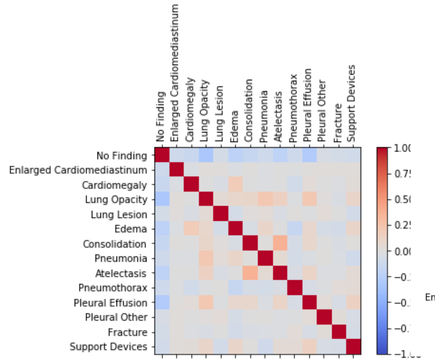


METHODS

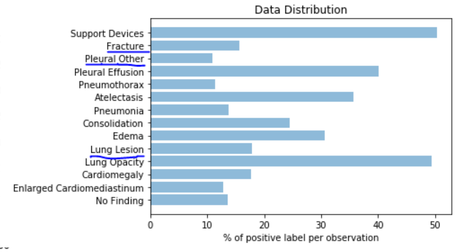
- Deep Learning Models for Multi-Class Prediction
  - **VGG-16**
  - **ResNet-50**
  - **DenseNet-121**
- Weighted Gradient Class Activation – Image Localization



RESULTS



Up-sampling Improved data Balance



	f1-score Baseline DenseNet-121	f1-score DenseNet-121 balanced Data
No Finding	0.27	0.77
Enlarged Cardiomediastinum	0.00	0.42
Cardiomegaly	0.25	0.64
Lung Opacity	0.47	0.73
Lung Lesion	0.00	0.81
Edema	0.22	0.71
Consolidation	0.00	0.55
Pneumonia	0.00	0.52
Atelectasis	0.00	0.66
Pneumothorax	0.00	0.57
Pleural Effusion	0.53	0.77
Pleural Other	0.00	0.73
Fracture	0.00	0.68
Support Devices	0.74	0.84
micro avg	0.42	0.70
macro avg	0.18	0.67
weighted avg	0.34	0.70
samples avg	0.34	0.65

CONCLUSIONS

- Best model: DenseNet-121 after balancing training data
- The importance of other indexes like f1-score besides accuracy
- The importance of splitting data into train, development and test
- Localization insights with weighted gradient class activation
- **Future steps**: combine CheXpert and MIMIC-CXR to further improve unbalance data, apply image segmentation