Stanford CheXpert Chest X-Ray Data Classification

Dataset contains:

* 14 training labels
* Labels on the training images are generated from previous radiology reports using NLP classification methods.
* Validation dataset consist of ~200 images which are hand-labeled by expert radiologists.

Labels used ( 0 = no presence, 1 = presence, -1 = uncertain):

'No Finding'

'Enlarged Cardiomediastinum'

'Cardiomegaly'

'Lung Opacity'

'Lung Lesion'

'Edema'

'Consolidation'

'Pneumonia'

'Atelectasis'

'Pneumothorax'

'Pleural Effusion'

'Pleural Other'

'Fracture'

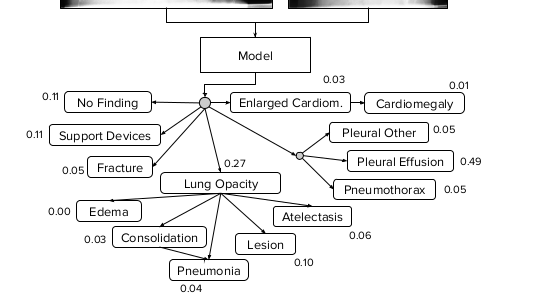
'Support Devices'

Looking at cross correlation between different disease labels:



Observations from cross-correlation values between labels:

* Some of the validation dataset labels (labeled by human experts) are well correlated with each other, indicating that certain diseases can be labeled together into larger groups (ie. group 1 = “Enlarged Cardiom. + Cardiomegaly”, group 2 = all sub-groups under “Lung Opacity” etc ). Similar to the graph below.
* The actual training dataset (labeled by NLP method) doesn’t appear to correlate similar to the validation dataset or the graph below, indicating there could be inaccurate or missing labels.



**Possibilities for Capstone project scope:**

* **Re-label Training dataset by grouping labels into following four groups:**
  + **No-Finding: No problems present**
  + **Lung-opacity group: includes edema, consolidation, pneumonia, lesion, atelectasis.**
  + **Cardiology problems group: includes Enlarged Cardiomediastinum and Cardiomegaly.**
  + **Pleural group: includes pleural effusion and pleural other.**
* **Ignore all other original labels like “Support devices” and “fractures” etc.**
* Main Stanford study using the same dataset and the 14 training labels above: <https://arxiv.org/pdf/1901.07031.pdf>
  + Trained 5 separate models based on criteria for utilizing the uncertain label (-1) .
  + Some models performed better for certain diseases.
  + The paper displays a probability tree showing the relationship between different diseases.

Other studies:

1. <https://lukeoakdenrayner.wordpress.com/2019/02/25/half-a-million-x-rays-first-impressions-of-the-stanford-and-mit-chest-x-ray-datasets/>
   * Someone who has some domain knowledge in radiology took a look at the CheXpert datasets and made some observations:
     + The labeling quality on the training dataset (labeled using NLP methods using past radiology reports) is sometimes poor and inaccurate.
     + 70% of the images come from 31% of the patients.
     + The “support devices” and “fracture” labels are quite inaccurate. Should not be used.
2. <https://www.seas.upenn.edu/~chanana/Chexpert.html>
   * Used the CheXpert dataset and generated ResNet and DenseNet models, one for each type of X-ray (frontal and lateral)
   * Validation tests showed it was unable to distinguish between 'Enlarged Cardiomediastinum' and ‘Cardiomegaly'.
     + [These two labels](https://arxiv.org/pdf/1711.05225.pdf) have high correlation since 'Enlarged Cardiomediastinum' is cause by ‘Cardiomegaly’ → These are related heart illnesses.
3. <https://arxiv.org/pdf/1711.05225.pdf> : Previous Stanford attempt on NIH Chest X-Ray dataset
   * Diseases labeled:
     + - Atelectasis
       - Cardiomegaly
       - Consolidation
       - Edema
       - Effusion
       - Emphysema
       - Fibrosis
       - Hernia
       - Infiltration
       - Mass
       - Nodule
       - Pleural Thickening
       - Pneumonia
       - Pneumothorax
   * First trained a model to detect ‘pneumonia’ only. Then extended it to detect all 14 diseases with high success rate.
4. <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0204155&type=printable>
   * Used a different dataset. Images only had 4 labels to indicate problems:
     + “Enlarged cardiac silhouette” : similar to 'Enlarged Cardiomediastinum' and 'Cardiomegaly' from the Stanford dataset.
     + Pleural effusion : similar to 'Pleural Effusion' and 'Pleural Other'
     + Pulmonary opacity : equivalent to 'Lung Opacity', ‘atelectasis’, ‘pneumonia’, ‘consolidation’,
     + Hilar prominence : Doesn’t seem to have any equivalence to Stanford labels.
5. 3. <https://arxiv.org/pdf/1711.06373.pdf>
   * Used the NIH Chest X-Ray dataset:
     + Diseases labeled:
       - Atelectasis
       - Cardiomegaly
       - Consolidation
       - Edema
       - Effusion
       - Emphysema
       - Fibrosis
       - Hernia
       - Infiltration
       - Mass
       - Nodule
       - Pleural Thickening
       - Pneumonia
       - Pneumothorax
     + Used labeled ‘bounding boxes’ to identify regions where diseases existed to help guide the learning. It helped to increase accuracy but was a manual effort to generate these additional bounding boxes and needed the expertise of radiologists.