**Garrett Wankel [Aug20]**

# Chest X-ray Abnormalities Detection

## Context

According to the Center for Disease Control (CDC), on average, 48 out of 100 people who will visit a hospital will receive an x-ray. Some of these will be in relation to a pre-identified condition, and others will be a diagnostic measure to confirm the presence or eradication of a condition. The latter of the two is an increasingly important measure to diagnose a possible ailment that could be life-threatening and it helps reduce total healthcare costs for individuals and the nation at large.

However, chest x-rays are infamously inconclusive and as we age, anatomical changes can cause even further ambiguity among results, as noted here. In cases like that, consensus matters, which is why improving the accuracy of identifying any given condition is important. Early and accurate detection reduces time spent in a hospital as well as visits, reduces the need for an invasive procedure, and it would also provide invaluable information regarding disease progression or regression.

From an x-ray, if a radiologist were able to identify an observed abnormality, and work in conjunction with a computer aided detection (CAD) algorithm that could tell with almost 100 percent certainty that the condition observed by the radiologist was indeed present or would need another opinion, this would greatly reduce hospital costs for patients and staffing issues for the hospital's managers. It would also be of great impact to developing countries where they may have one machine for every million people, because that one radiologist could tell with validated certainty what five radiologists would be needed to identify.

## Criteria For Success

1. For this task, finding the mean Average Precision (mAP) at IoU (Intersection over Union) > 0.4. That is to say, I will set the IoU threshold to be greater than 40 percent, and calculate the mean of the average precision of all observations that have been confirmed as a positive label matching.

## Scope of Solution

This analysis and modeling will be extremely valuable to physicians, biomedical manufacturers and hospital board members alike, because developing a program based around an incredibly accurate algorithm will drive cost-savings into the future for hospitals, provide an error-mitigated environment for physicians, and a program for manufacturers to build their technology around. Limiting a radiologist's hours and mental bandwidth could also help their effectiveness while working. In regards to developing nations, having new ground-breakinging technology and software could decrease mortality rates never seen before by expediting treatment and implementing more preventative care measures.

## Constraints

1. Dataset size. Often with Convolutional Neural Networks (CNNs), not having enough images to train on inhibits model performance.

## Key Stakeholders

1. VinBigData
    a. Hosts of the competition, they are a medical diagnostic imaging company based out of Vietnam.

## Key Data Sources

1. https://www.kaggle.com/c/vinbigdata-chest-xray-abnormalities-detection - link to the competition page and corresponding dataset.