**PROBLEM STATEMENT**

In medicine, the next frontier for AI is anomaly localization in medical imaging. Localization of anomalies refers to both predicting anomalies and their boundaries. Automatic detection algorithms to locate the position of inflammation in an image can help the physicians in making the better clinical decisions. In this project we analyze data with the knowledge of EDA, we build a detection model and we present our findings based on the evaluations with the RSNA Pneumonia Detection Challenge dataset.

**DATA AND FINDINGS**

In 2018, RSNA organized an AI challenge to detect pneumonia, one of the leading causes of mortality worldwide, as part of its efforts to help improve artificial intelligence (AI) instruments for radiology. RSNA Pneumonia dataset consists of 29684 thousand images. All the images are in dicom format. There are 26684 images for training and 3000 images for testing.

**Dicom images:** In a special format called DICOM files (\*. dcm), medical images are stored. They contain a mix of header metadata as well as pixel data underlying raw image arrays.

There are three classes featured in the dataset Normal, Not normal/No opacity and Lung opacity. Normal class indicates there is no anomaly in the lungs. Not normal/No opacity indicates while it was decided that pneumonia was not present, there was still some sort of picture abnormality and sometimes this finding could mimic the appearance of true pneumonia. Lung opacity class indicates there is definite pneumonia in the lungs. These three classes are divided as two target variables 0 and 1 the images with lung opacity comes under target 1 and 0 is assigned to other two classes.

Along with the images two csv files are provided. Detailed class info file consists of image name and the class it belonged to. Train labels file consists of the bounding box coordinates belonging to each image. Bounding box coordinates are given in the following format as follows,

* x -- the upper-left x coordinate of the bounding box.
* y -- the upper-left y coordinate of the bounding box.
* width -- the width of the bounding box.
* height -- the height of the bounding box.

With these bounding box coordinates target column is provided which discriminates classes into categories of 0 and 1. All the images in the dataset belongs to stage 2 pneumonia.