Abstract:

The global demand for diagnostic imaging has continued to increase significantly and the demand is exceeding the supply of radiologists. Can a deep learning model be trained to detect pathologies in X-ray images at an expert level? Can this be used to help alleviate some demand and improve healthcare access to parts of the world where it is limited? By using data provided by Stanford for their MURA competition a model can be built and evaluated.

Intro:

The demand for medical imaging has only increased over the years with advancements in medical technology. I am hoping to make a deep learning model with a convolutional neural network (CNN) to detect abnormalities in an X-ray image at the level of expert. Applications of this model can be applied towards:

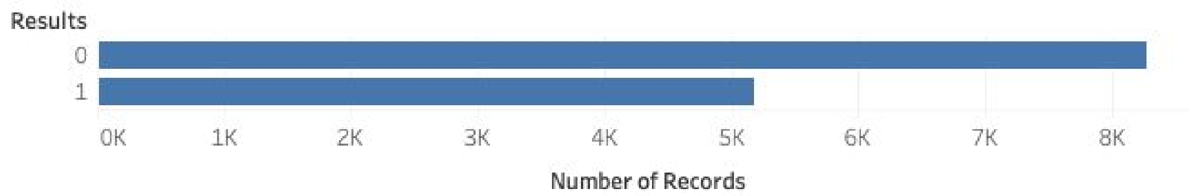
- Improving healthcare access to parts of the world where access to skilled radiologists is limited.
- Improving patient care/experience by decreasing wait times.
- Alleviating workload for radiologists
- Aiding radiologists and other physicians in diagnosing patients.
- Aid in training new radiologists

Methodology:

The dataset that I will be using is from the Stanford ML Group and it is called MURA (musculoskeletal radiographs). The dataset contains 40,561 multi-view X-ray images from 14,863 studies. The images are of the upper extremities and are categorized into seven different categories: finger, hand, wrist, forearm, elbow, humerus, and shoulder. Each image is further labeled to be normal or abnormal.

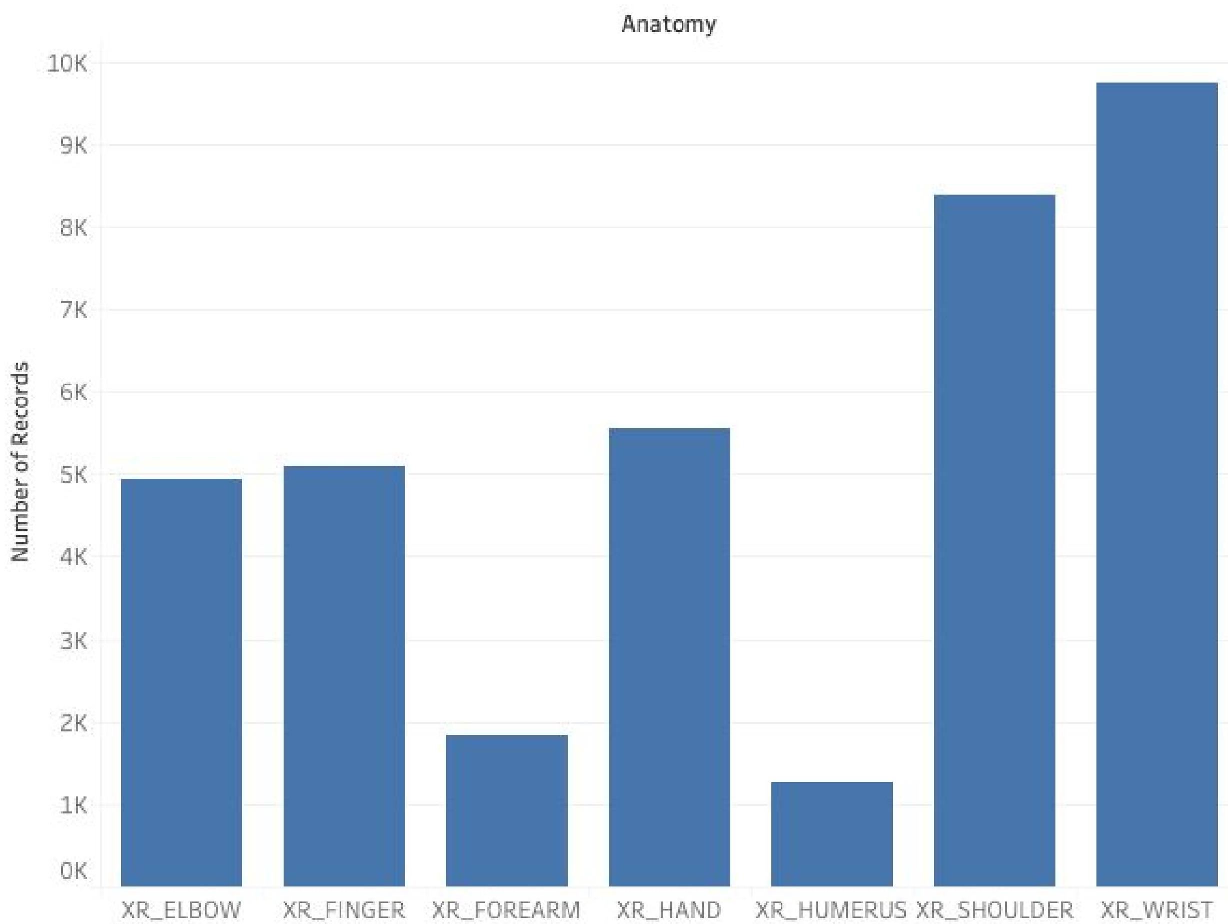
The goal of this project is to build a CNN deep learning model that can perform as well as medical professionals. This can be measured by comparing my model's performance to the scores on the MURA website (https://stanfordmlgroup.github.io/competitions/mura/) where not only other competitors are scored but where the radiologist score is displayed as well.

Challenges that this project face are large images that are not uniform in dimensions. The anatomy is not always orientated the same way and a transformation may need to be applied to the images. There is also a slight imbalance in the dataset where roughly 60% of the cases are normal with only 40% being abnormal.



Sum of Number of Records for each Results.

Fig 1



Sum of Number of Records for each Anatomy.

Fig 2

Within the dataset the images in the seven different categories is imbalanced as well with only 1272 cases for humerus where as wrist cases numbered at 9752 (Fig 2). Other challenges include training the model which will take a lot of processing power and time. The dataset has seven different categories of anatomy so the CNN ensemble can be considered but this will take even more training and processing.

Building and training the model properly I hypothesise that I can achieve a score close to or even above what the Stanford radiologists were able to score, which was a cohen kappa score of 0.778.

Conclusion:

For the final project a total of 8 different models were trained and deployed on a website to classify upper extremity X-rays and to detect any abnormalities in the X-rays. The anatomy classification model was 91% accurate and could classify the different parts of the arm. The 7 other models were made to detect abnormalities within each separate body part. Their accuracies ranged from 46%-76% which is quite a large range and below what the Stanford radiologists were able to score. The models that had the lowest accuracies were the forearm and wrist models. With more time and adjustments in the models trained I am sure that these scores can be improved. Overall I believe that I have achieved my goal in creating models that can classify different body parts as well as detect anomalies in the X-rays as well.