Vision Transformers on Chest X-Ray Dataset

The availability of datasets in healthcare is limited. We are interested in exploring how vision-transformers pre-trained on generic, non-healthcare data perform on highly specialized datasets like chest X-Ray images. We plan to compare the performance of vision transformers with supervised pre-training, self-supervised pre-training and no pre-training with our baseline model.

Our team comprises of three members:

1. Aishwarya Raman (ar6381)
2. Nidhi Ranjan (nr2387)
3. Shreya Gupta (sg6606)

Dataset

We will be using the National Institutes of Health Chest X-Ray Dataset (link) which comprises of 112,120 X-ray images of size 1024 x 1024 with disease labels from 30,805 unique patients.There are 15 classes (14 diseases, and one for "No findings"). Images can be classified as "No findings" or one or more disease classes. The labels were created using Natural Language Processing to text-mine disease classifications from the associated radiological reports. The original dataset size is ~45 GB. However, considering the project time and resource availability, we will be using a smaller sample to the data set.

Implementation

We will be comparing the accuracy of the following models on our dataset:

A. *Baseline*: We will be implementing a Convolution neural network as the baseline model similar to the one mentioned in ChestX-ray8

B. *Vision Transformers(without pretraining)*: We will be implementing vision transformers trained only on the Chest X-Ray dataset [https://arxiv.org/pdf/2010.11929.pdf]

C. *Vision Transformers (with supervised Pre Training)*: We will be implementing vision transformers pre trained on ImageNet dataset [https://arxiv.org/pdf/2010.11929.pdf] and fine tuned on ChestX-Ray dataset.

D. *Vision Transformers (with Self Supervised Pre training)*: We will be implementing vision transformers pre trained using self supervision on ImageNet dataset [https://arxiv.org/abs/2104.14294] and fine tuned on ChestX-Ray dataset.