



# Summer Internship 2024

<<Interpret-CXR >>



*Team members:*

*Faculty-in-charge:* Dr. Sowmya Kamath S.

Department of Information Technology  
National Institute of Technology Karnataka, Surathkal

**May 21, 2024**

- Original images were downloaded in JPG standard. Each image was converted to PNG using some post-processing.
- top/bottom 0.5% pixel values were clipped (to eliminate very dark or very bright pixel outliers)
- Pixel values scaled linearly to fit into 0–255 range
- Resized to 2048 on shorter side (to fit in Kaggle dataset limits)
- All these pre processing of images were done locally using a python script.

I have made two dataframes: **reports** and **projections**.

Projections provide **mapping between uid (unique id to each image) to image path and view of the scan (frontal or lateral)**.

Reports provide **mapping between uid and the captions**.

I have divided the UIDs into train and test set.

GP2 Used as Tokenizer

Llama Better but resource intensive

Fine tuning on of vision language model like Llava, MoonDream

I will first fine-tune gpt2 on the **corpus of findings and impressions**. This should improve GPT-2 alignment to medical report generation.

# Observations after Dataset analysis

- Top two most findings are '**no actue cardiopulmonary abnormality**' and '**no actue cardiopulmonary findings**' above 400 counts in feature.
- In impression feature, 50% data have less than 4 words per impression, it means only few impression have less words. 99% data have less than 39 words per impression. Only 1% data have legnth above 39.
- From wordcloud : **actue cardiopulmonary, cardiopulmonary abnormality, disease acute, heart size** are the highlighted words i.e. these are important words.
- Each patient have multiple x-rays associated with them. The maximum number of images associated with a report can be 5 while the minimum is 0. The highest frequency of being associated with a report are 2 images.

**THANK YOU**