# Automated Radiology Report Generation and Disease Classification for enhanced patient care

Aditya Sankhla - 12140060

## Introduction

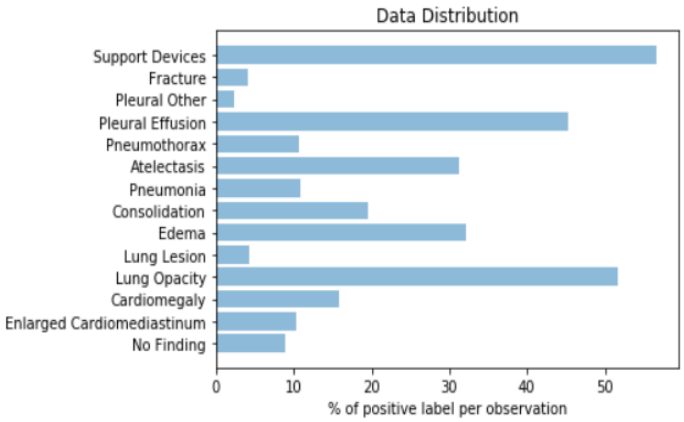
By the final evaluation, the project will deliver a functional portal where users can input symptoms and X-ray images for classification. This portal will provide an accessible diagnostic experience for users, marking the culmination of our project.
For the mid-evaluation, we successfully classified images using the CheXpert dataset and created a user-friendly portal powered by Hugging Face for answering basic queries.

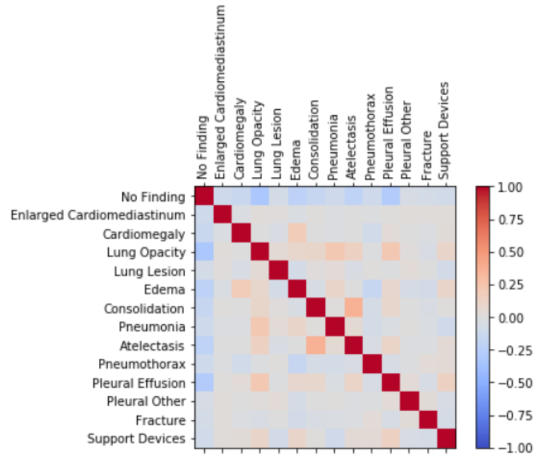
## Data Pre-Processing

The dataset used in this study is drawn from the CheXpert dataset for chest radiograph interpretation, comprising a vast collection of 224,316 chest radiographs belonging to 65,240 unique patients. It's important to mention that, for this classification, only frontal view chest X-ray images are taken into consideration. Additionally, the "Age," "Sex," and "AP/PA" features are excluded from the training process.

For data preprocessing, we employed various techniques, including image augmentation and exploratory data analysis (EDA). The EDA encompasses heatmaps, histograms, and pie charts to gain a comprehensive understanding of the dataset's characteristics and intricacies, which in turn aids in model development and evaluation.

To facilitate the training process, the dataset was divided into three subsets: a training set, a test set, and a validation set, with 10% of the data allocated to the validation set and another 10% to the test set. This strategic partitioning ensures a robust evaluation of our model's performance and generalization capabilities.




## Model Training and Evaluation

We trained fundamental models like ResNet, Inception, and DenseNet, which played a pivotal role in our study, shedding light on the performance of different architectures for chest X-ray image classification.

Our data pipeline was thoughtfully crafted, ensuring effective data preprocessing and augmentation, which contributed to the overall performance enhancement of our models.

The training phase was a crucial step, allowing us to fine-tune and optimize the models for multi-classification of chest X-ray images. This was closely followed by a rigorous validation process to assess their generalization capabilities.

Through these efforts, we established a robust data pipeline that encompassed model training, validation, and fine-tuning. The outcome of our endeavor includes critical performance metrics—accuracy, AUC, precision, recall, F1-score, and support—for all the classes within our dataset. These metrics offer valuable insights into the effectiveness of our models in classifying chest X-ray images across various medical observations, representing a significant achievement in our project.

## Model Performance and Selection

We trained several models, including ResNet, Inception, and DenseNet. Among these, the DenseNet model stood out as the top performer, achieving the highest accuracy and AUC values. We based our model architecture on pre-defined Keras models , adapting them to our specific task by modifying the final layers. These changes included adding a global spatial average pooling layer, a fully connected layer with the ReLU activation function, and a logistic layer with 14 outputs using sigmoid activation.

For training, we employed binary cross-entropy as the loss function, suitable for binary classification tasks. We optimized our models with the Adam optimizer, setting a learning rate (r) of 0.0001, along with hyperparameters $\beta1$ at 0.9 and $\beta2$ at 0.999. These configurations played a pivotal role in effective model training and enhancing performance throughout our study.

The DenseNet model exhibited a training accuracy of 91.67% (0.9167), underlining its robust learning capability. In the validation phase, it maintained an impressive accuracy of 90.67% (0.90674), demonstrating its capacity to generalize well. The final test accuracy also remained strong at 87.7% (0.877), reinforcing the reliability and effectiveness of the DenseNet model in classifying chest X-ray images across various medical observations.

**Table 3: Test Accuracy and Test Scores for different models**

| Model | Train Accuracy | Test Accuracy |
|---|---|---|
| **DenseNet -121** | 0.375 | 0.270 |
| **DenseNet with lr scheduler and fine tuning** | 0.188 | 0.241 |
| **DenseNet with Pre-trained weights (Final)** | 0.916 | 0.877 |

**Table 1 : F1-score on Test Data with DenseNet 121 model**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| No Finding | 0.90 | 0.89 | 0.89 | 6903 |
| Enlarged Cardiomediastinum | 0.68 | 0.69 | 0.69 | 6456 |
| Cardiomegaly | 0.84 | 0.73 | 0.78 | 8741 |
| Lung Opacity | 0.86 | 0.80 | 0.83 | 24517 |
| Lung Lesion | 0.93 | 0.97 | 0.95 | 8806 |
| Edema | 0.91 | 0.59 | 0.71 | 15229 |
| Consolidation | 0.66 | 0.86 | 0.75 | 12202 |
| Pneumonia | 0.89 | 0.55 | 0.68 | 6855 |
| Atelectasis | 0.74 | 0.91 | 0.81 | 17793 |
| Pneumothorax | 0.79 | 0.76 | 0.78 | 5654 |
| Pleural Effusion | 0.86 | 0.89 | 0.87 | 20018 |
| Pleural Other | 0.94 | 0.99 | 0.96 | 5451 |
| Fracture | 0.85 | 0.99 | 0.92 | 7729 |
| Support Devices | 0.91 | 0.89 | 0.90 | 24988 |
| micro avg | 0.84 | 0.83 | 0.83 | 171342 |
| macro avg | 0.84 | 0.82 | 0.82 | 171342 |
| weighted avg | 0.84 | 0.83 | 0.83 | 171342 |
| samples avg | 0.83 | 0.82 | 0.81 | 171342 |

**Table 2: F1-score on test data with DenseNet model with lr Scheduler and Fine Tuning**

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| No Finding | 0.60 | 0.47 | 0.53 | 6808 |
| Enlarged Cardiomediastinum | 0.71 | 0.01 | 0.02 | 6368 |
| Cardiomegaly | 0.75 | 0.21 | 0.33 | 8831 |
| Lung Opacity | 0.63 | 0.82 | 0.71 | 24254 |
| Lung Lesion | 0.78 | 0.25 | 0.38 | 8935 |
| Edema | 0.62 | 0.80 | 0.70 | 15267 |
| Consolidation | 0.53 | 0.10 | 0.17 | 12168 |
| Pneumonia | 0.54 | 0.13 | 0.21 | 6714 |
| Atelectasis | 0.60 | 0.52 | 0.56 | 17469 |
| Pneumothorax | 0.56 | 0.36 | 0.44 | 5658 |
| Pleural Effusion | 0.81 | 0.58 | 0.67 | 19738 |
| Pleural Other | 0.68 | 0.14 | 0.23 | 5461 |
| Fracture | 0.80 | 0.15 | 0.25 | 7974 |
| Support Devices | 0.84 | 0.62 | 0.71 | 24838 |
| micro avg | 0.68 | 0.48 | 0.56 | 170483 |
| macro avg | 0.67 | 0.37 | 0.42 | 170483 |
| weighted avg | 0.69 | 0.48 | 0.52 | 170483 |
| samples avg | 0.60 | 0.46 | 0.49 | 170483 |

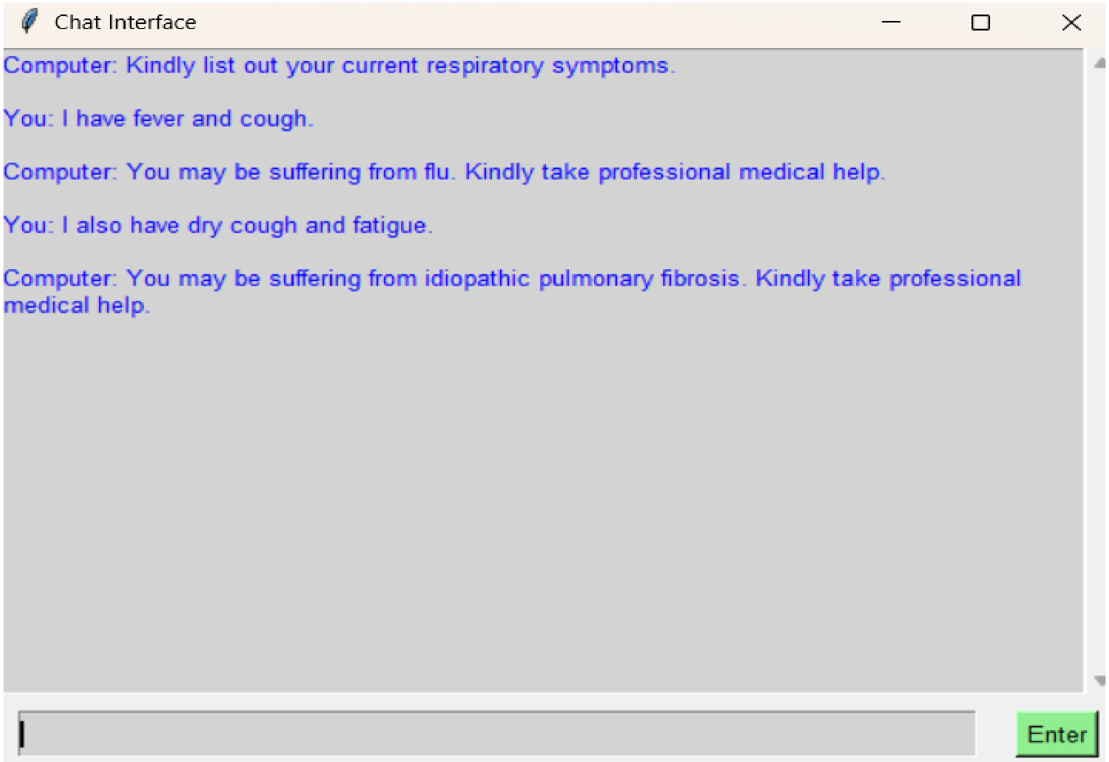



**Figure: GUI developed so far working on Hugging Face Model**

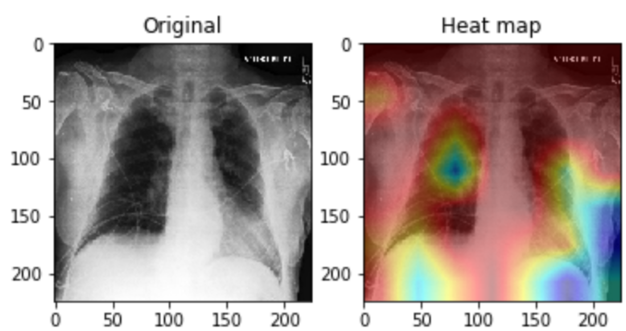



**Figure : A sample of Class Activation Map application**

## Challenges Encountered:

1. The dataset is very large which causes even a single epoch training to be very time consuming. We tried using smaller batch sizes and training over more compute

resources yet it was taking a lot of time. We plan to solve this using pipeline training i.e. instead of training the entire model from scratch, we will train in stages. We shall start with a subset of the data and a simpler model, and then progressively increase the complexity as we go.

2. Another challenge that we are facing is that we need to hardcode the context currently. This limits the amount of data we are able to provide to the model. We are planning on overcoming this hurdle using either RAG (yet to be finalised) which will index the medical documents and provide us with better query results at runtime.

## Future Plan:

- Channeling the multiclass prediction obtained from the model through a text generation model to obtain detailed medical reports which can potentially be used to assist physicians.
- Using the results obtained from both our classifier and the results from evaluation of the symptoms listed down by the user and finding the intersection of potential diseases for further enhancing the credibility of reports generated.
- Making the portal user-end ready by incorporating a mechanism to input their symptoms and X-rays and getting the required report generated directly.
- If time permits, the goal would be to work further on increasing the accuracy of the model.

## Individual Contributions (Aditya Sankhla) :

- **Data Preprocessing:** I played a crucial role in data preprocessing, which included data augmentation to increase the diversity of our dataset and improve model generalization. Additionally, I was involved in data visualization to better understand the data distribution and detect any potential imbalances.
- **Created and Evaluated ResNet Model:** I developed and tested the ResNet model, which proved to be highly accurate in identifying diseases from X-ray images.
- **GUI Creation:** I developed the graphical interface (GUI), simplifying data input and making it more user-friendly. This portal shall further be developed to accommodate the user text as well as the X-Ray images too.