Radiology Report Generation from Chest X-ray Image

Team Members:

Abilash Maharjan (THA077BCT004)

Anmol Kumar Gupta (THA077BCT011)

Shailendra Rawal (THA077BCT041)

Supervised By:

Er. Suramya Sharma Dahal

Department of Electronics and Computer Engineering

Institute of Engineering, Thapathali Campus

August 9, 2024

Presentation Outline

- Motivation
- Objectives
- Scope of Project
- Project Applications
- Methodology
- Instrumentation Tools
- Implementation Details
- Analysis and Discussion
- Remaining Tasks
- References

Motivation

- Increasing demand for radiological services
- Need to reduce diagnostic turnaround time
- Importance of enhancing consistency and accuracy in radiology reports
- Improve patient outcomes by enabling faster, reliable and better-informed treatment decisions
- AI-driven solutions to transform healthcare practices

Objectives

- To develop an AI-driven system that can accurately classify chest X-ray images into various disease categories and generate detailed findings and impressions.
- To create a user-friendly web portal that allows healthcare professionals to upload chest X-ray images, view AI-generated reports, edit report if necessary, and download the final report in PDF format.

Scope of Project [1]

- **Project Capabilities**
  - Advanced AI Integration for Radiology
  - User-Friendly Clinician Interface
  - Implement optimization techniques to enhance processing efficiency

Scope of Project [2]

- **Project Limitations**
  - Requires continuous maintenance and updates to adapt and learn the ever-evolving medical knowledge
  - Inconsistent Image Quality may lead to inaccurate results
  - Models trained on specific datasets may underperform with varied patient demographics or disease prevalence
  - Difficult to understand decision processes of trained model affecting trust and validation
  - Integration Challenges into EHR systems

Methodology

- ## System Block Diagram

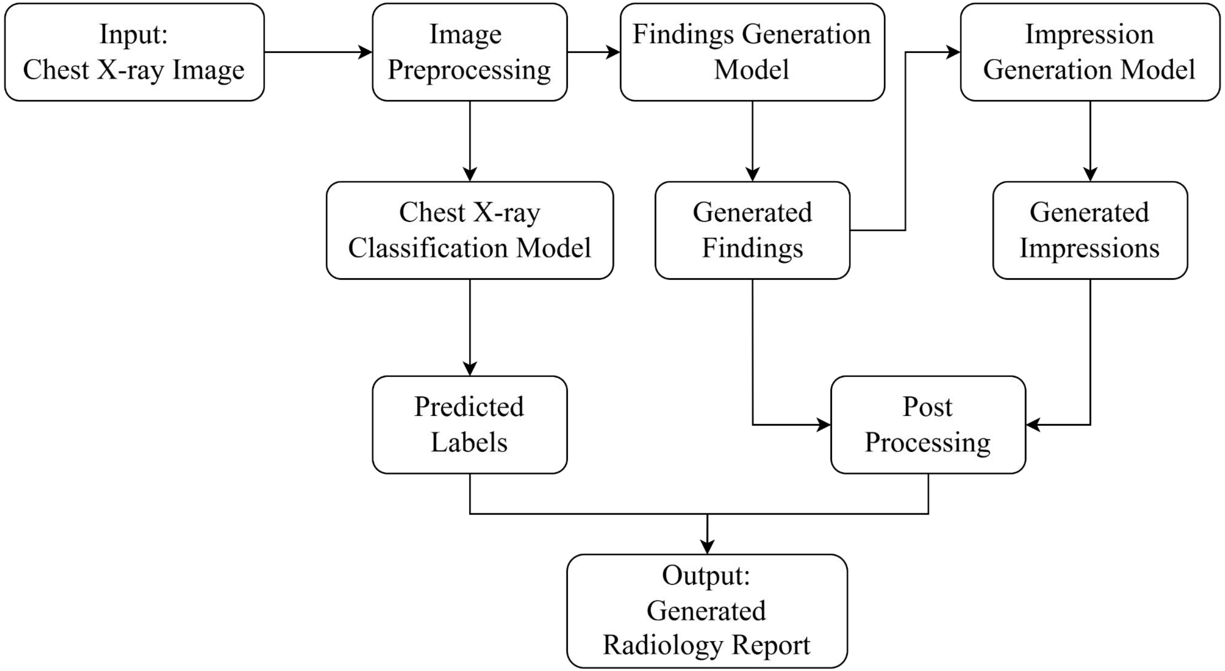


Fig: System Block Diagram

Methodology [Cont.]

- **Dataset Preparation**
  - Dataset Collection
  - Prepare Dataset for Chest X-ray Classification Model
  - Prepare Dataset for Findings Generation Model
  - Prepare Dataset for Impressions Generation Model
  - Train, Test, and Validation Split

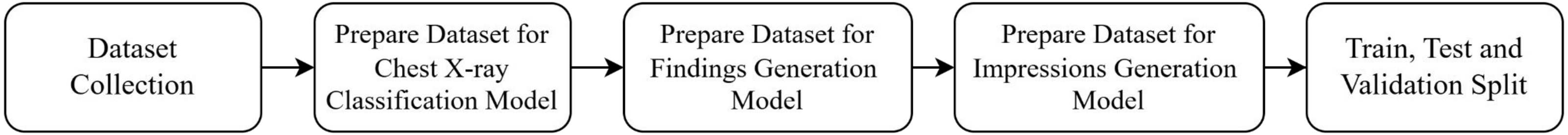
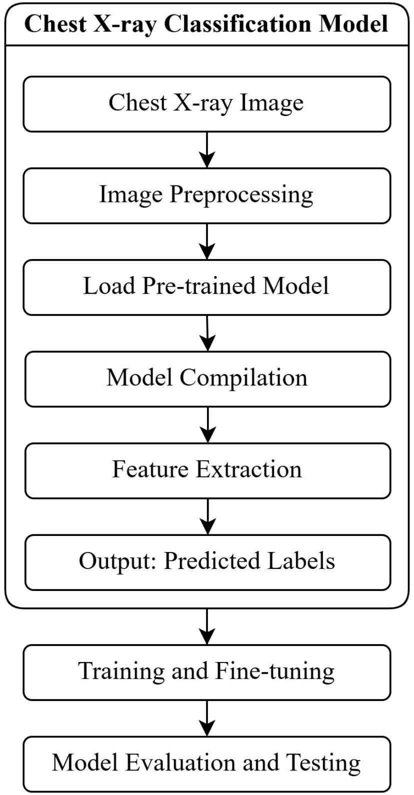


Fig: Dataset Preparation Workflow Diagram

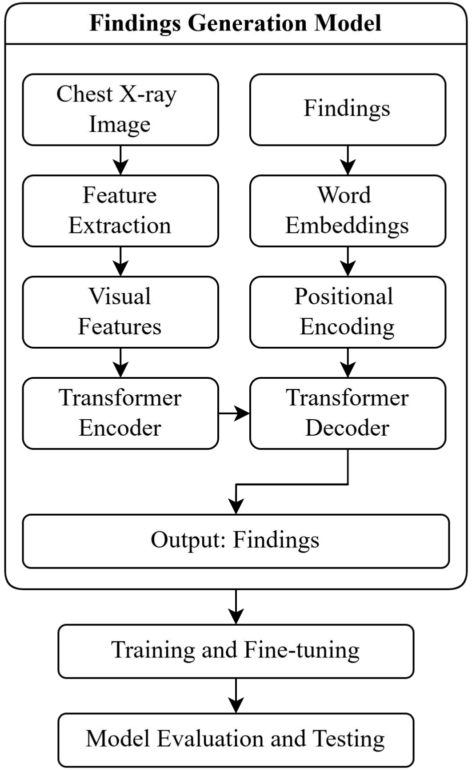
Methodology [Cont.]

- **Machine Learning Models [1]**
  - **Chest X-ray Classification Model**
    - Image Preprocessing
    - Load Pre-trained Model (DenseNet121)
    - Model is compiled with appropriate loss functions, optimizers and evaluation metrics
    - Extract relevant features
    - Predicted Labels are given as output
    - Training and Fine-tuning
    - Model Evaluation and Testing



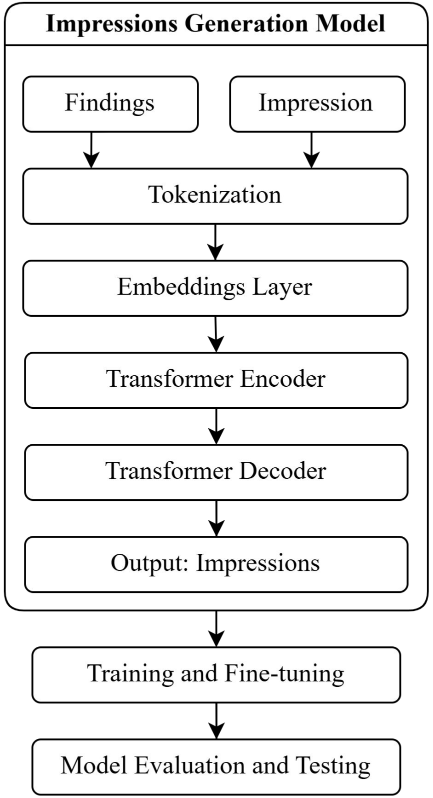
Methodology [Cont.]

- **Machine Learning Models [2]**
  - **Findings Generation Model**
    - CNN or ViT for visual features extraction
    - Feed into transformer encoder
    - Convert findings text into embeddings
    - Process the visual features and findings
    - Output detailed text findings
    - Training and Fine-tuning
    - Model Evaluation and Testing



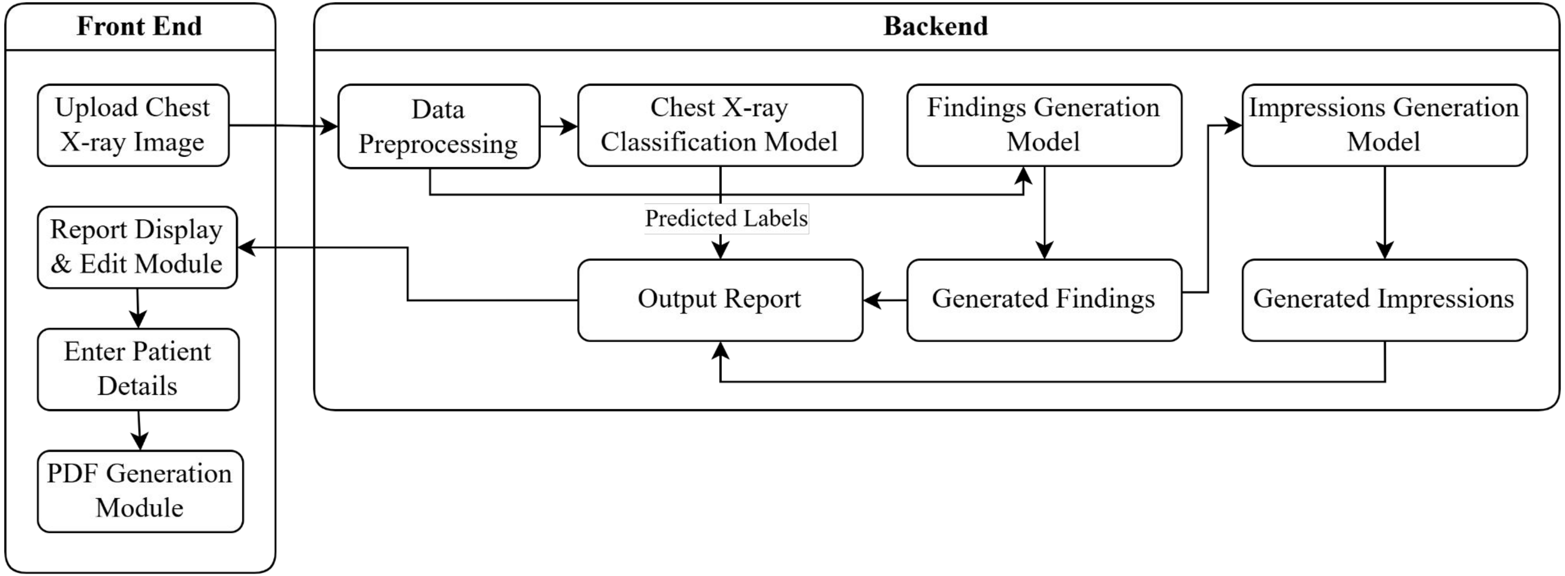
Methodology [Cont.]

- **Machine Learning Models [3]**
  - **Impressions Generation Model**
    - Tokenize findings
    - Convert into embeddings
    - Process embeddings
    - Generate high-level Impressions
    - Output summarized Impressions
    - Training and Fine-tuning
    - Model Evaluation and Testing



Methodology [Cont.]

- **Front End and Backend**



Methodology [Cont.]

- **Datasets**

For generating report from Chest X-ray image, the following set of datasets are used:

1. NIH Chest X-rays
2. Indiana University Chest X-ray Dataset
3. MIMIC-CXR Dataset

Methodology [Cont.]

- **Chest X-ray Images from NIH Chest X-ray Dataset**

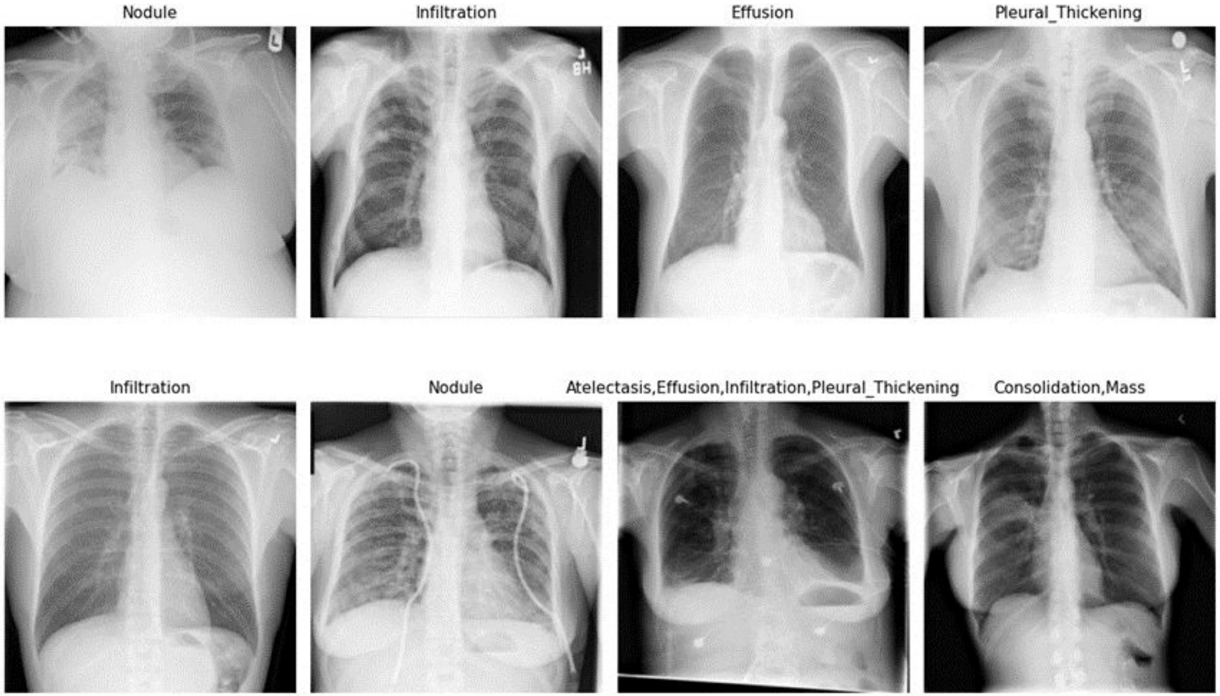


Fig: Chest X-ray Images with their Findings Labels

Methodology [Cont.]

- **Indiana University Chest X-ray Dataset**

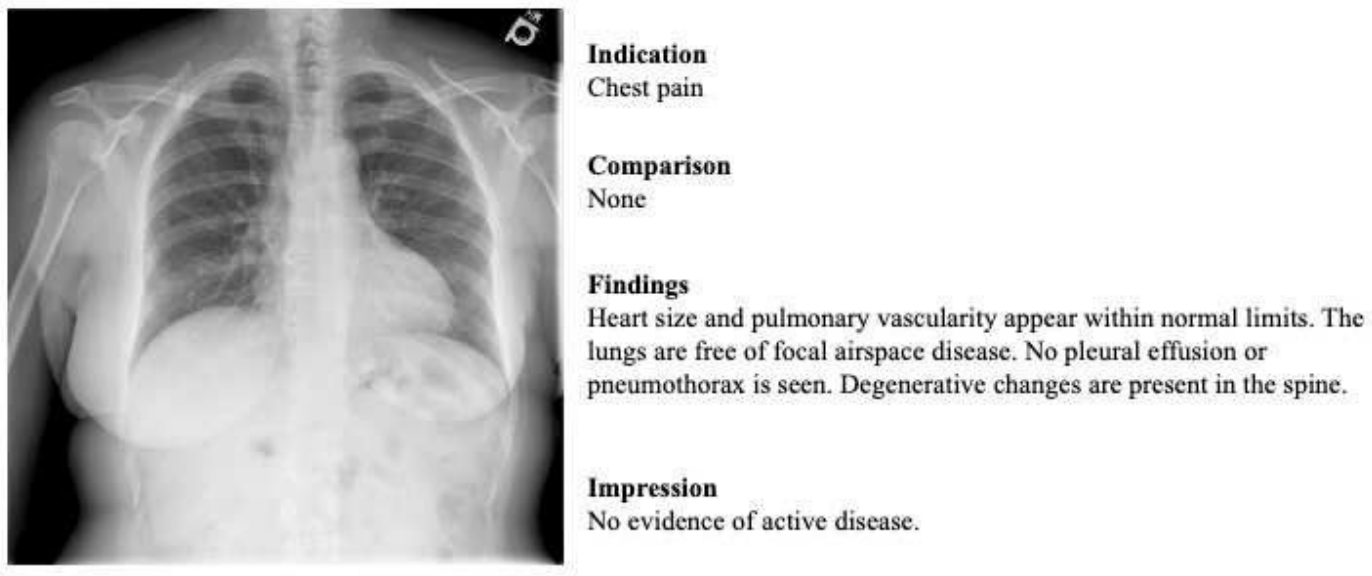
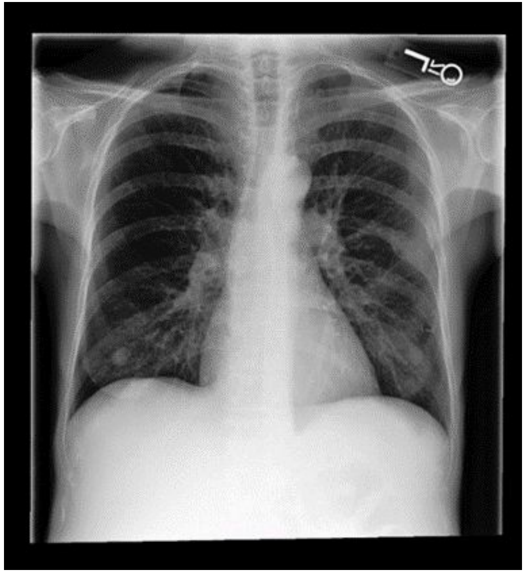


Fig: Chest X-ray Image with its Radiology Report from Indiana University Dataset

Methodology [Cont.]

- **MIMIC-CXR Dataset**



FINAL REPORT

EXAMINATION: CHEST (PA AND LAT)

INDICATION: ___F with new onset ascites // eval for infection

TECHNIQUE: Chest PA and lateral

COMPARISON: None.

FINDINGS:

There is no focal consolidation, pleural effusion or pneumothorax. Bilateral nodular opacities that most likely represent nipple shadows. The cardiomediastinal silhouette is normal. Clips project over the left lung, potentially within the breast. The imaged upper abdomen is unremarkable. Chronic deformity of the posterior left sixth and seventh ribs are noted.

IMPRESSION:

No acute cardiopulmonary process.

Fig: Chest X-ray Image with its Radiology Report from MIMIC-CXR Dataset

Methodology [Cont.]

- **Architecture of DenseNet121**

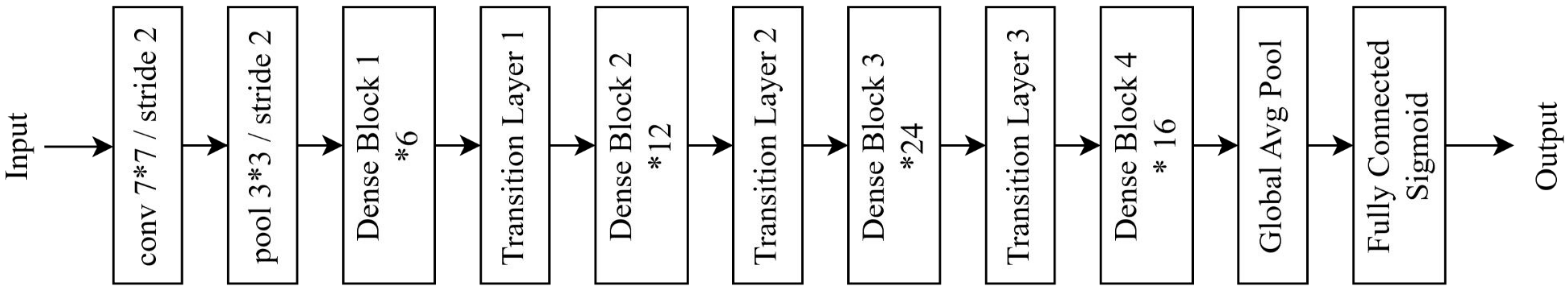


Fig: Architecture of DenseNet121

- **Input:** RGB image of size 224x224x3
- **Initial Convolution:** convolution layer with 64 filters, k.s. of 7x7, stride 2, padding 3 which is followed by Batch Normalization layer and ReLU activation
- **Initial Pooling:** max pooling with pool size of 3X3 and stride 2

Methodology [Cont.]

- **Dense Block and Transition Layer of DenseNet121**

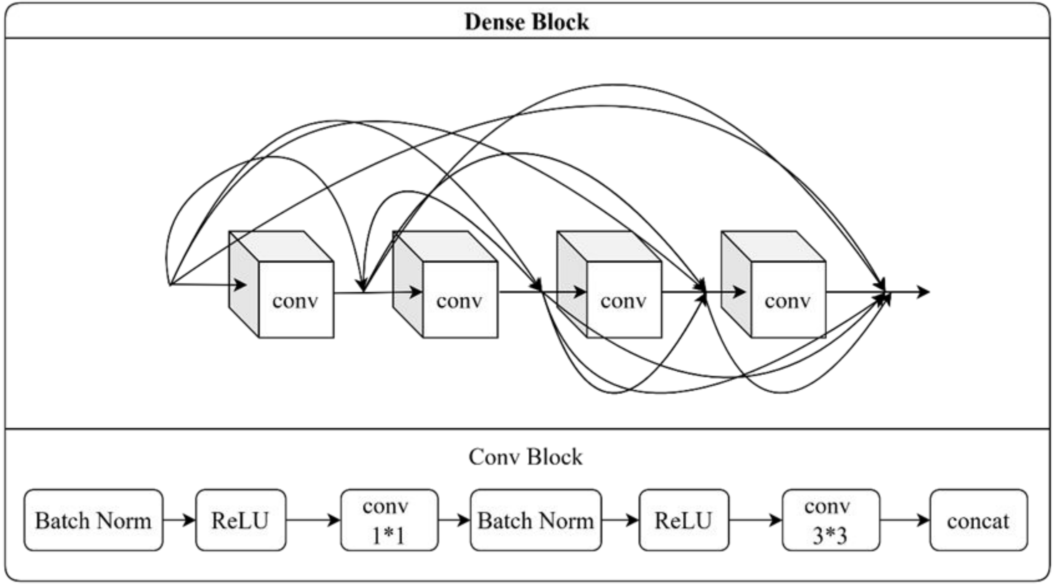


Fig: Dense Block of DenseNet121

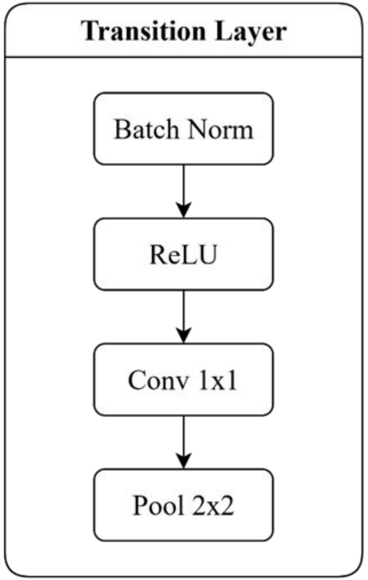


Fig: Transition Layer of DenseNet121

Methodology [Cont.]

- **Dense Block**
  1. **Batch normalization:** It scales down the value to bring stability to the training process.
  2. **ReLU function:** It is the activation function.
  3. **1*1 convolution:** It is 1*1 matrix size filter.
  4. **3*3 convolution:** It is 3*3 matrix size filter.
  5. **Concatenation:** The outputs form previous layer are merged and passed on to next layer.

Methodology [Cont.]

- **Transition Layer**

  1. **Batch normalization:** It scales down the value to bring stability to the training process
  2. **ReLU function:** It is the activation function.
  3. **1*1 convolution:** It reduces no. of feature maps.
  4. **Average Pooling:** It is a 2*2 size filter which reduces the dimension of feature maps.

Methodology [Cont.]

- **Classification Layer**

  1. **Global Average Pooling (GAP):** It reduces the spatial dimensions of the feature maps to a single value per map. Thus summarizing the image.
  2. **Dropout:** It is a regularization technique used to prevent overfitting. It is applied to GAP and dense layer (prediction).

Methodology [Cont.]

- **Classification Layer**

  3. Dense Layer (Prediction): It is a fully connected layer consisting of 256 neuron.

  4. Output Dense Layer: The final layer has 12 neurons, each providing the probability for a findings labels.

Methodology [Cont.]

- **Activation Functions**

1. Sigmoid Function

- It maps the input value in the range between 0 and 1.
- We have applied sigmoid as activation function in dense layer (prediction) and output layer.

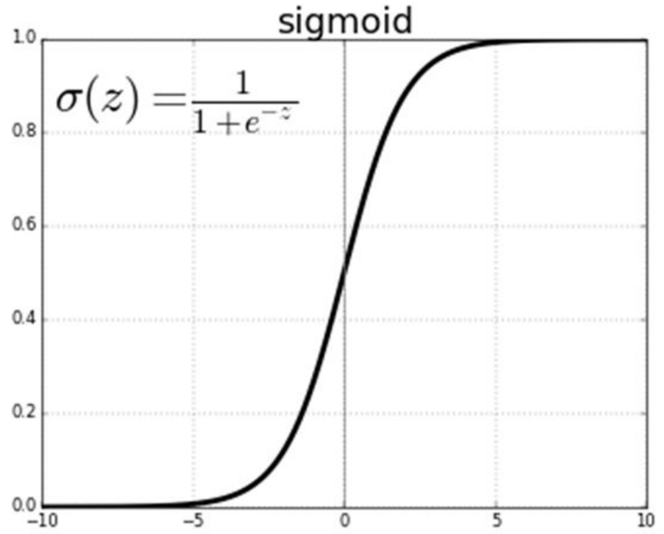


Fig: Graph of Sigmoid Activation Function

Methodology [Cont.]

- **Activation Functions (cont.)**

2. ReLU Function

- It outputs the input value directly if is positive, otherwise it outputs zero.

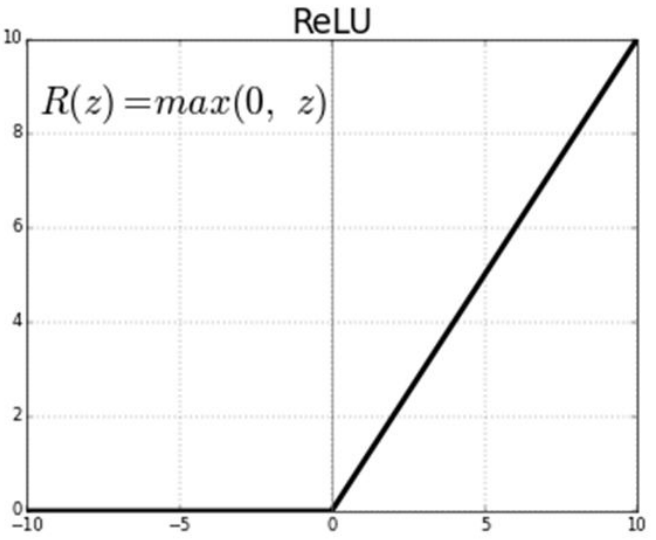


Fig: Graph of ReLU Activation Function

Methodology [Cont.]

- **Loss Function**

  1. **Binary Cross Entropy Loss**

  - It is well-suited for multilabel classification loss calculation.
  - It calculates loss independently for each label and then averages over all labels.

  Mathematically:

$$L = -\frac{1}{N * C}\sum_{n=1}^{N}\sum_{c=1}^{C}[y_{nc}\log(p_{nc}) + (1 - y_{nc})\log(1 - p_{nc})]$$

  Where,
  N is the number of samples C is the number of labels
  y_{nc} is the true label true label for sample n and class c.
  p_{nc} is predicted probability for sample n and class c.

Methodology [Cont.]

- **Optimizer**

  1. **Adam Optimizer**

  - It is used to update the weights and learning rate of the neural network.
  - The formula for update rule is

$$\theta_t = \theta_{t-1} - \frac{\alpha \hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}$$

Methodology [Cont.]

- **Evaluation Metrics**

1. **Binary Accuracy:**

   For each instance i and each label j:

$$Accuracy_j = \frac{1}{N}\sum_{i=1}^{N} I(y_{ij} = \hat{y}_{ij})$$

$$Binary\ Accuracy = \frac{1}{L}\sum_{j=1}^{L} Accuracy_j$$

Methodology [Cont.]

- **Evaluation Metrics (cont.)**

2. Precision:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

3. Recall:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

4. F1 Score:

$$F1\ Score = 2 \times \frac{Precision\ \times Recall}{Precision + Recall}$$

5. ROC Curve

$$TPR = \frac{TP}{TP + FN} \qquad FPR = \frac{FP}{FP + TN}$$

Instrumentation Tools [1]

1. Software Requirements

- Python
- Tensorflow
- PyTorch
- Jupyter Notebook
- NumPy
- Pandas
- Scikit-learn
- Matplotlib
- Seaborn
- OpenCV
- Pillow
- PyPDF2
- VS Code
- Streamlit

Instrumentation Tools [2]

2. Hardware Requirements

- **CPU**: Intel Octa-core i5 1.60GHz or higher CPU
- **GPU**: NVIDIA K80, T4, P4 or P100 (Google Colab)

  NVIDIA Tesla P100 (Kaggle)

- **RAM**: Minimum 8 GB
- **Storage**: 512 GB SSD or higher

Implementation Details

- **Image Preprocessing**

1. Resizing

- Chest X-ray Image resized to 224 by 224

2. Normalization

- It adjust the range of pixel intensity values in an image to standard scale
- Two types of normalization were employed:

Symmetric Normalization:

$$normalized_{sym} = \frac{image - image_{min}}{\frac{image_{max} - image_{min}}{2}} - 1$$

Symmetric-positive Normalization:

$$normalized_{pos} = \frac{image - image_{min}}{image_{max} - image_{min}}$$

Implementation Details [Cont.]

- **Image Preprocessing (cont.)**

  3. CLAHE (Contract Limited Adaptive Histogram Equalization)

  - Advanced image processing technique designed to enhance the contrast of an image
  - Key steps includes:
  - **Division**: dividing image into grid of small tiles to apply the equalization process locally.
  - **Histogram Equalization**: The contrast of each tile is enhanced by redistributing pixel intensities to achieve a uniform histogram.

Implementation Details [Cont.]

3. CLAHE (cont.)

$$New\ Value = round\big(CDF(c) * (L-1)\big) \qquad CDF(x) = \frac{\sum_{i=0}^{x} PDF(i)}{N}$$

- **Contract Limiting:** Contract limit is applied to avoid over-amplification of noise by clipping at a predefined threshold.

$$Clipped\ PDF(x) = \min(PDF(x), Clip\ Limit)$$

$$Redistributed\ PDF(x) = Clipped\ PDF(x) + \frac{Excess\ Pixels}{Number\ of\ Bins}$$

- **Interpolation:** The local tiles are smoothly combined using interpolation to remove boundaries and ensure a seamless final image

Implementation Details [Cont.]

3. CLAHE (cont.)

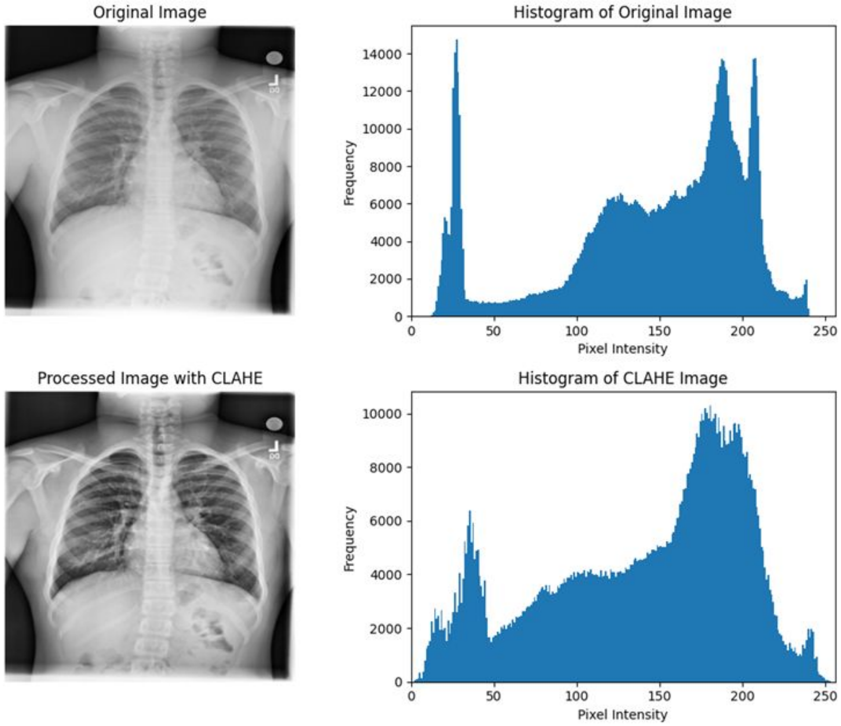


Fig: Comparison Original Image and Image after applying CLAHE Interpolation along with their respective Histogram

Implementation Details[Cont.]

- **Image Preprocessing (cont.)**

4. Gaussian Blur

- Smooths an image by reducing its noise by applying Gaussian filter.
- **Gaussian Kernel:** a matrix used to average each pixel's value with its neighbors, weighted by their distance, to achieve smooth, blurred effect that preserves edges
- **Convolution Operation:**

$$I_{blurred}(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x+i, y+j).K(i, j)$$

Implementation Details [Cont.]

- **Model Initialization**
  - Load pre-trained weights of ImageNet dataset
  - Exclude top classification layers
  - Set model configuration to trainable
- **Model Compilation**
  - Loss Function: Binary Cross Entropy
  - Optimizer: Adam
  - Metrics: Binary Accuracy, F1-Score and AUC Value

Implementation Details [Cont.]

- **Model Training**
  - **Setting Hyperparameters**
    - Epochs: 20
    - Initial Learning Rate: 0.001
    - Batch Size: 32
    - Input Shape: 224,224,3
    - Number of classes: 12

Implementation Details [Cont.]

- **Model Training (cont.)**
  - **Assigning Class Weights**
    - Address class imbalance in the dataset
    - Balance influence of each class to improve performance across all classes

$$class\_positive\_counts[i] = \sum_{j=1}^{N} x_{j,i}$$

$$class\_weights[i] = \frac{N}{num\_classes \cdot class\_positive\_counts[i]}$$

Implementation Details [Cont.]

- **Model Training (cont.)**
  - **Learning Rate Scheduler**
    - Adjusts the learning rate according to predefined schedule or strategy during training process.
    - Improves convergence by gradually reducing the learning rate

$$new\_lr = lr * \exp(-0.1)$$

    - Reduces the learning rate by about 9.5% each time the epoch count crosses the threshold.

Implementation Details [Cont.]

- **Model Training (cont.)**
  - **Reduce Learning Rate on Plateau**
    - Adjusts the learning when performance metrics on validation set plateaus
    - Learning rate is reduced by a specified factor which is typically less than 1.

$$(\eta \leftarrow \eta * factor)$$

Implementation Details [Cont.]

- **Model Training (cont.)**
  - **Early Stopping**
    - Regularization technique that monitors the performance of the model on the validation set during training.
    - Halts training when the model's performance on the validation set does not improve for a specified no. of consecutive epochs.
    - Helps in preventing overfitting and unnecessary training.

Results [1]

- **Image Pre-processing**

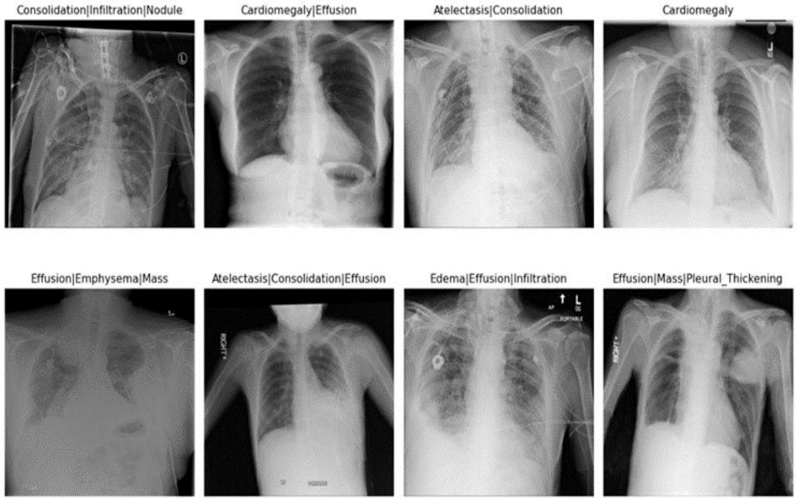


Fig: Image Before Pre-processing

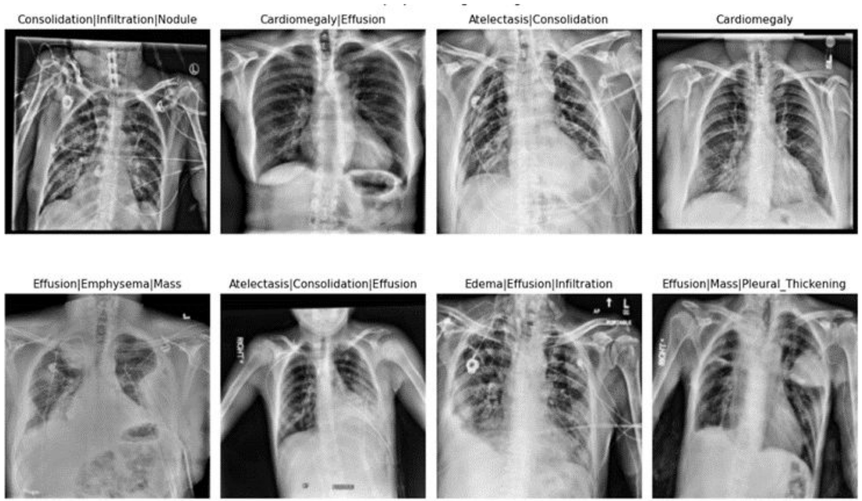


Fig: Image After Pre-processing

Results [2]

- **Chest X-ray Classification Model Performance and Evaluation**

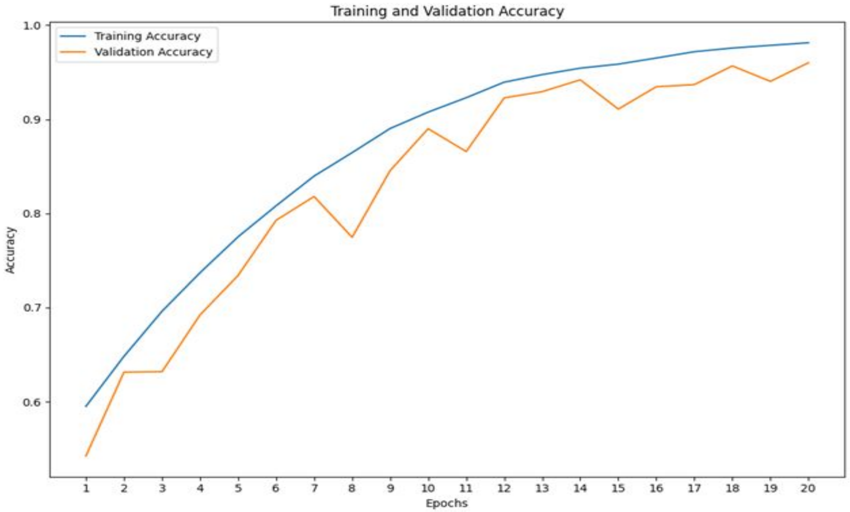


Fig: Training and Validation Accuracy Curve

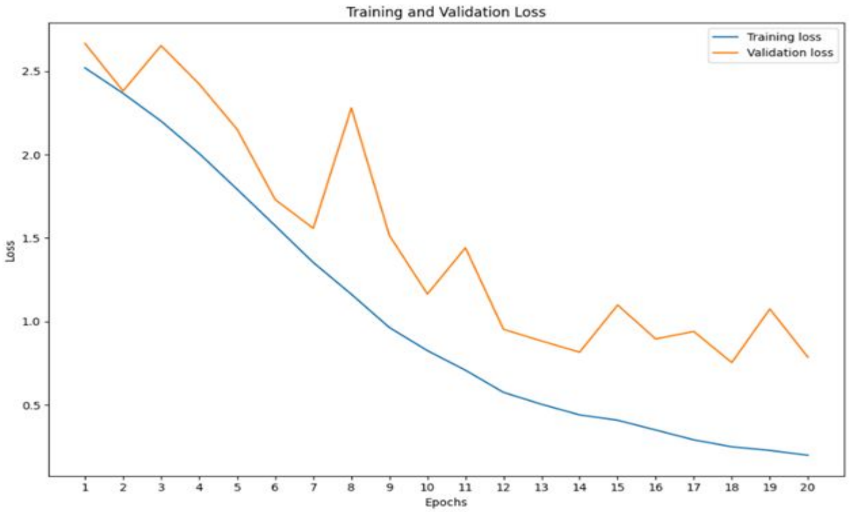


Fig: Training and Validation Loss Curve

Results [3]

- **Chest X-ray Classification Model Performance and Evaluation (cont.)**

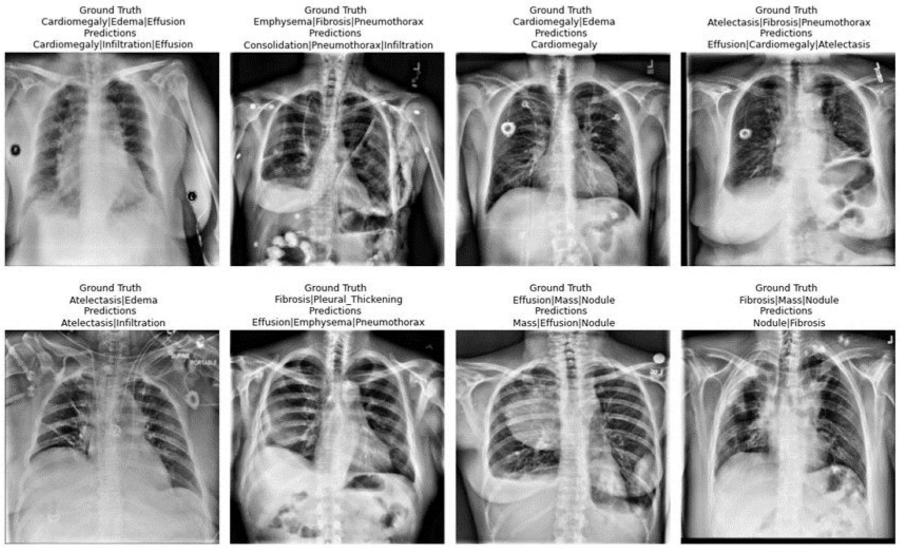


Fig: Ground Truth vs Predictions

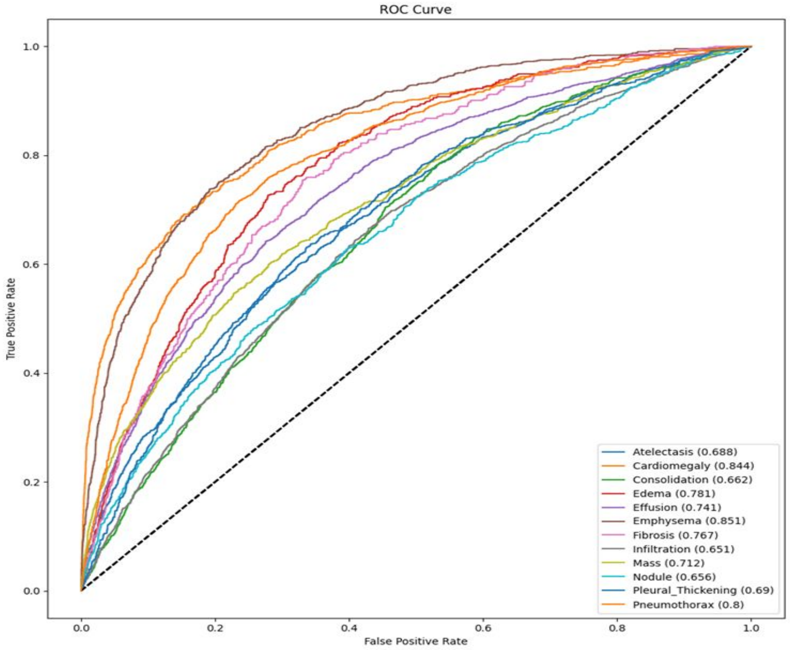


Fig: ROC curve

Results [4]

- **Chest X-ray Classification Model Performance and Evaluation (cont.)**

Findings Labels	Precision	Recall	F1 Score
Atelectasis	0.569157	0.137112	0.220988
Cardiomegaly	0.740841	0.361111	0.485549
Consolidation	0.285714	0.002920	0.005780
Edema	0.304928	0.282857	0.293478
Effusion	0.576792	0.501890	0.536740
Emphysema	0.618231	0.419730	0.500000
Fibrosis	0.242236	0.061321	0.097867
Infiltration	0.502848	0.259072	0.341962
Mass	0.423329	0.337513	0.375581
Nodule	0.379747	0.093400	0.149925
Pleural_Thickening	0.326724	0.226810	0.267750
Pneumothorax	0.642643	0.309920	0.418173
Micro avg	0.514662	0.272760	0.356554
Macro avg	0.467766	0.249471	0.307816
Weighted avg	0.499527	0.272760	0.334976

Table: Classification Report

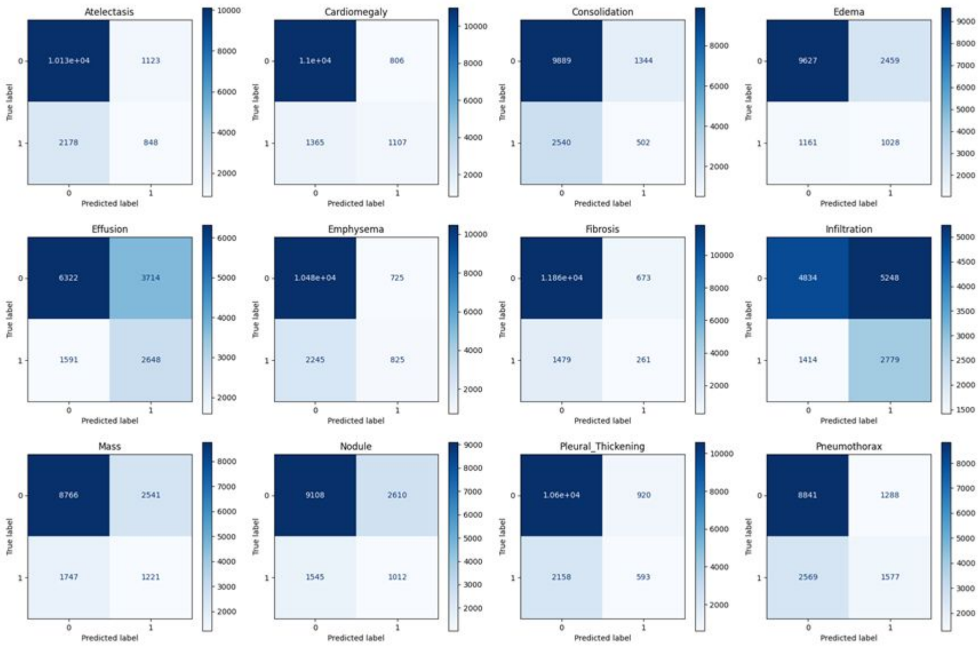


Fig: Confusion Matrix

Results [5]

- **User Interface for Chest X-ray Classification**

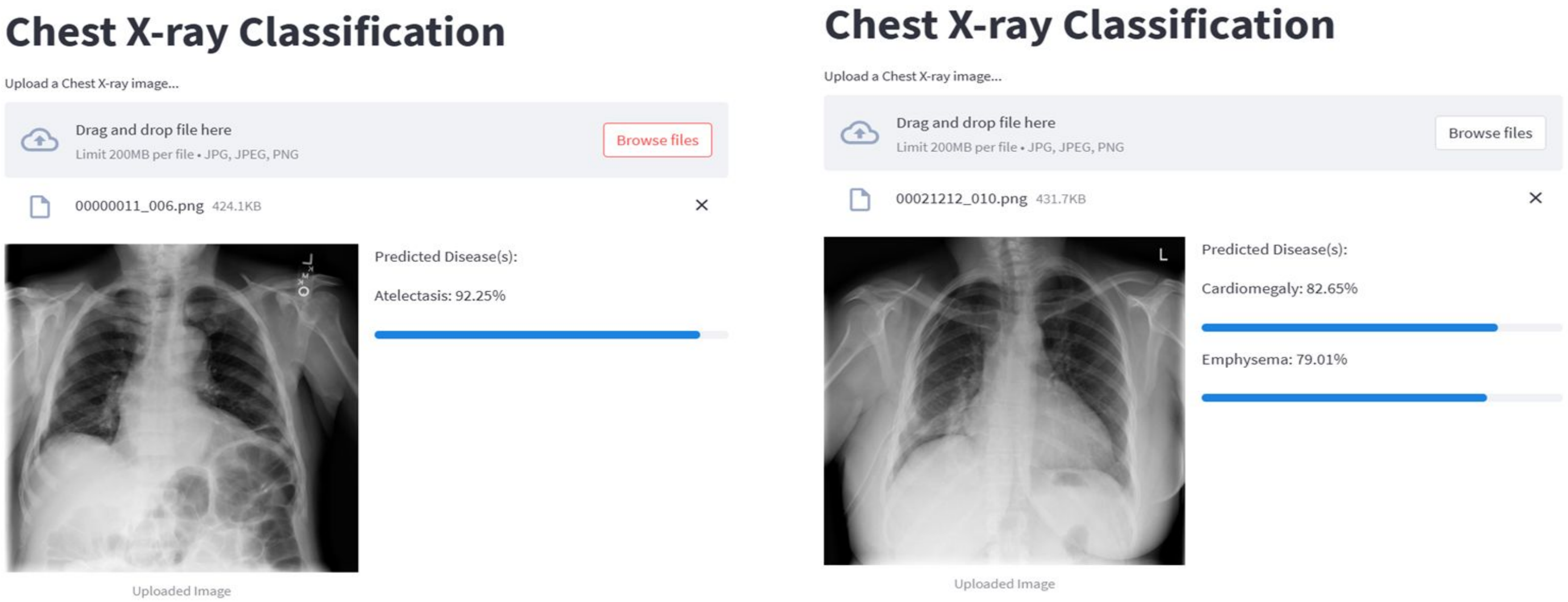


Fig: UI showing Prediction for different diseases

Analysis and Discussion

- Overfitting Problem encountered in the Chest X-ray Classification model
- Low Precision, Recall and F1-Score obtained for most of the classes
- Model performance isn’t satisfactory on test data
- Challenge for obtaining high AUC value for each findings labels

Remaining Tasks

- Fine-Tuning Chest X-ray Classification Model
- Preparing Dataset for Findings and Impressions Generation Models
- Development, Training and Fine-Tuning of Findings and Impressions Generation Models
- Web Interface Development

References

[1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Jul. 2017, pp. 3462–3471. doi: 10.1109/CVPR.2017.369.

[2] P. Rajpurkar et al., “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” Nov. 2017.

[3] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, “Learning to diagnose from scratch by exploiting dependencies among labels,” Oct. 2017.

[4] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.

[5] Y. Wang, “Reading Radiology Imaging Like The Radiologist,” Jun. 2023.

References

[6] M. Nishio et al., “Fully automatic summarization of radiology reports using natural language processing with large language models,” in Informatics in Medicine Unlocked, 2024, p. 101465. doi: 10.1016/J.IMU.2024.101465.

[7] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Jul. 2017, pp. 3462–3471. doi: 10.1109/CVPR.2017.369.

[8] V. Kougia, J. Pavlopoulos, and I. Androutsopoulos, “A Survey on Biomedical Image Captioning,” May 2019.

[9] A. E. W. Johnson et al., “MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports,” Sci Data, vol. 6, no. 1, p. 317, Dec. 2019, doi: 10.1038/s41597-019-0322-0.

[10] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” Aug. 2016.