Natural Language Processing
Final Project
DATS 6312

Medical Report Generation using Chest X Ray Images

Individual Project Report
(Team 6)

for the course

DATS 6312 ‘Natural Language Processing’

Fall 2024

Anand Raj

Introduction

Overview of the Project

The primary objective of this project is to develop an automated system capable of generating comprehensive medical reports from chest X-ray (CXR) images. Leveraging advanced deep learning architectures, the system inte grates image processing, multi-label classification, and natural language generation to produce accurate and clinically relevant reports. The project utilizes the MIMIC CXR dataset, which comprises 15,000 chest X-ray images paired with corresponding medical reports.

Outline of Shared Work

The project is a collaborative effort encompassing various components:

- **Label Extraction**: Utilization of ChexBERT to generate multi-label classifications from textual reports.
- **Model Architecture Changes and Improvements**: Developed architecture combining image encoders, clinical findings, alignment model, and language model. This work was making changes to the Base model architecture to fit ChexNet into the new architecture. Also changed the alignment model to get better results in terms of alignemnt.
- **Training and Evaluation**: Training of models using Parameter-Efficient Fine-Tuning (PEFT) techniques and evaluation based on metrics such as ROUGE-L.
- **Report Generation**: Modified report generation of BioGPT for generating natural language medical reports from aligned embeddings after concatenation of structural findings coming from chexnet.
- **Streamlit** Application **development:** I developed the streamlit app incorporating the best two models which were **BioVilt + ChexNet + Alignment + BioGPT** and **Blip2 + ChexNet + Alignment + BioGPT.** Also added a Chest X Ray classifier using simple image transforms.

My contributions to this project specifically involve the **development of scripts for multi-label extraction of structural findings** and the training of the **BioVilt + ChexNet + Alignment + BioGPT** architecture and **developing the streamlit application.**

2. Description of Individual Work

Multi-Label Extraction Script Development

Accurate extraction of structural findings from medical reports is crucial for training robust multi-label classification models like ChexNet. The reports in the MIMIC CXR dataset are initially in XML format, these are then converted to a csv containing image_id, image_path, findings, and impressions. To facilitate effective training, it is essential to convert these textual reports into structured multi-label annotations.

Script Development

The script I developed performs the following tasks:

1. **Data Parsing**: Reads the combined CSV file containing image_id, image_path, findings, and impressions.
2. **Label Extraction**: Utilizes ChexBERT, a pre-trained BERT model specialized for chest X-ray interpretation, to extract multi-label classifications from the impressions fields.
3. **Label Encoding**: Transforms the textual findings into binary vectors indicating the presence or absence of specific clinical conditions.
4. **CSV Generation**: Compiles the extracted labels alongside the corresponding image_id and image_path into a new CSV file, serving as the training dataset for ChexNet.

ChexBert:

ChexBERT is a specialized machine learning model designed for the classification of medical reports, particularly those related to chest X-rays. Building upon the foundational BERT (Bidirectional Encoder Representations from Transformers) architecture, ChexBERT is fine-tuned to understand and interpret the nuanced language found in medical documentation. Its primary function is to automatically categorize medical reports into multiple labels, facilitating tasks such as disease diagnosis, treatment planning, and research.

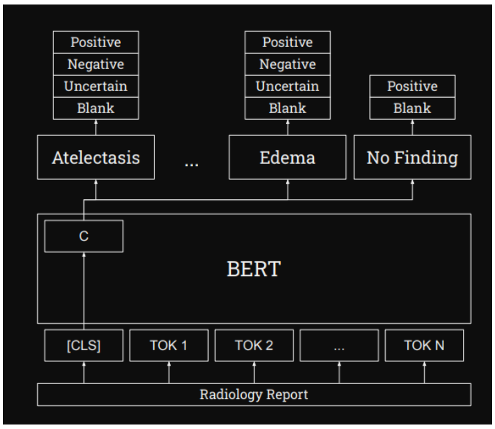


Figure 1: ChexBert Architecture

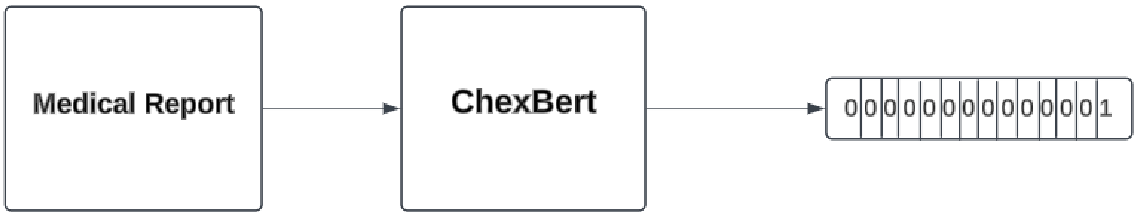


Figure 1: ChexBert for Multi-Label Extraction

Training of BioVilt + ChexNet + Alignment + BioGPT

Overview

The integration of BioVilt, ChexNet, an alignment model, and BioGPT forms a comprehensive architecture for generating medical reports from CXR images. This architecture combines high-level image feature extraction, structured findings classification, semantic alignment, and advanced language generation to produce coherent and clinically accurate reports.

Model Architecture

1. **BioVilt:** A ResNet-50 based image encoder specialized for biomedical image processing.
2. **ChexNet:** A DenseNet-121 based multi-label classifier identifying structural findings in CXR images.
3. **Alignment Model:** Bridges the gap between image embeddings and textual representations using linear projection layers. This was changed from what was used in the base architecture.
4. **BioGPT:** A pre-trained language model fine-tuned for biomedical text generation, employing Parameter-Efficient Fine-Tuning (PEFT) with LoRA.

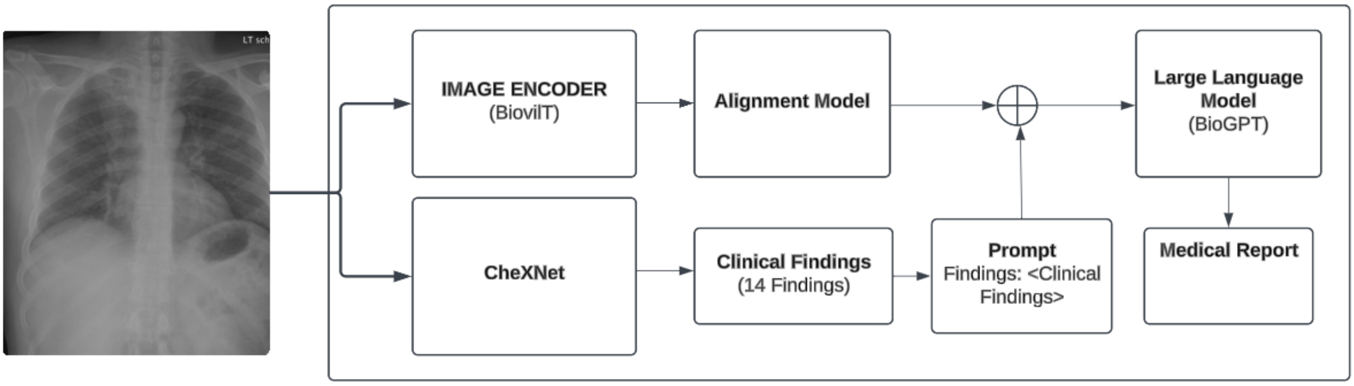


Figure 2: BioVilt + ChexNet + Alignment + BioGPT Architecture

Training Configuration

- **BioVilt**:
  - **Architecture**: ResNet-50 backbone.
  - **Output Dimensions**: Global embedding of 512 dimensions.
- **ChexNet**:
  - **Architecture**: DenseNet-121 backbone.
  - **Output Dimensions**: Multi-label predictions for 14 clinical findings.
- **Alignment Model**:
  - **Text Encoder**: BioGPT (For ground truth report).
  - **Projection Layers**: Linear layers to project image embeddings to 768 dimensions. A separate linear layer to project the text in the ground truth report as well.
  - **Loss Function**: Contrastive Loss.
- **BioGPT**:
  - **PEFT Configuration**: LoRA with rank 16, alpha 32, and dropout 0.1.
  - **Generation Parameters**: max_length=150, temperature=0.8, top_k=50, top_p=0.85.
  - **Generation Loss:** Cross Entropy Loss

Training of BioVilt + ChexNet + Alignment + BioGPT

Training Pipeline:

The training pipeline orchestrates the interaction between all components image preprocessing, image encoding, ChexNet-based multi-label classification, alignment modeling, and report generation to learn the mappings from images and structured findings to textual reports effectively.

Detailed Workflow (train.py)

1. **Environment Setup:**
    - **Device Selection:** Determines whether to use a GPU or CPU for training. (cuda used)
    - **Logging and Experiment Tracking:** Configured to provide real-time feedback on training progress, with optional integration with Weights & Biases (wandb) for experiment tracking.
2. **Model Initialization:**
    - **Image Encoder:** Instantiated using get_biovil_t_image_encoder(), which loads pre-trained BiovilT image encoder weights. Global embedding is extracted (<CLS>).
    - **Alignment Model:** ImageTextAlignmentModel maps image embeddings to text embeddings, facilitating semantic alignment.
    - **Report Generator:** MedicalReportGenerator utilizes BioGPT with LoRA for generating textual reports based on aligned embeddings.
3. **Data Loading:**
    - **Dataloaders:** Obtained via get_dataloaders, providing batched image, text findings, and findings lists for training and validation.
4. **Optimizer and Scheduler Configuration:**
    - **Alignment Optimizer:** Optimizes parameters of the alignment model using AdamW with a learning rate of 1e-4 and weight decay.
    - **Generator Optimizer:** Optimizes parameters of the report generator, including both the LoRA adapters and the image projection layer, using AdamW with separate learning rates.
    - **Schedulers:** Utilizes linear schedulers with warm-up steps (warmup_steps=1000) to adjust learning rates during training, enhancing convergence.
5. **Loss Function and Gradient Scaling:**
    - **Contrastive Loss: Contrastive Loss** is used to align image and text embeddings.
    - **Gradient Scaler:** Employs mixed precision training (torch.cuda.amp.GradScaler) to accelerate training while maintaining numerical stability.
6. **Training Loop:**
    - **Epoch Iterations:** Runs for a specified number of epochs (num_epochs=30), with the ability to resume from the last checkpoint if available.
    - **Training Phase:**
        - **Model Modes:** Sets the image encoder to evaluation mode to prevent gradient updates, while enabling training for the alignment model and report generator.

        - **Batch Processing:**
            - **Image Embeddings:** Extracted from BiovilT without gradient computation.
            - **ChexNet Labels:** Processed within the dataset and given to prompt as "Findings: <Clinical Findings>".
            - **Alignment Step:** Projects image and text (ground truth report) embeddings, computes alignment loss, and updates the alignment model.
            - **Concatenation of Image Embeddings and Prompt:** The Image embeddings and the prompt text embeddings are concatenated having a separator (<SEP>) token in between them and passed to BioGPT.
            - **Generation Step:** Generates reports based on concatenated embeddings, computes generation loss, and updates the report generator.
            - **Gradient Accumulation:** Accumulates gradients over multiple steps (gradient_accumulation_steps=4) to simulate larger batch sizes without increasing memory usage.
            - **Sample Outputs:** Periodically prints sample generated reports alongside their target findings for qualitative assessment.
    - **Validation Phase:**
        - **Model Modes:** Switches models to evaluation mode to assess performance on unseen data.
        - **Loss and Metrics Computation:**
            - **Losses:** Computes alignment (**Contrastive Loss**) and generation losses (**Cross Entropy Loss**) on the validation set.
            - **NLU Evaluation Metrics:** ROUGE-L scores to quantify report quality.
    - **Checkpointing and Best Model Saving:**
        - **Checkpoint Saving:** Saves the model state after each epoch to enable resumption.
        - **Best Model Tracking:** Updates and saves the best-performing model based on validation loss.

7. **Metric Calculation (validate_epoch):**
    - **ROUGE-L:** Measures the longest common subsequence between generated and reference reports.
    - **Sample Generation Output:** Sample generated reports versus actual findings to monitor qualitative performance.
8. **Logging and Experiment Tracking:**
    - **Weights & Biases (wandb):** Optionally logs training and validation metrics for comprehensive experiment tracking and visualization.

3. Results

Performance of BioVilt + ChexNet + Alignment + BioGPT

The integration of BioVilt, ChexNet, the alignment Model, and BioGPT resulted in **significant improvements in report generation quality** when compared to the **base architecture**, as evidenced by the ROUGE-L scores over multiple epochs.

ROUGE-L scores for BioVilt + Alignment + BioGPT and BioVilt + ChexNet + Alignment + BioGPT

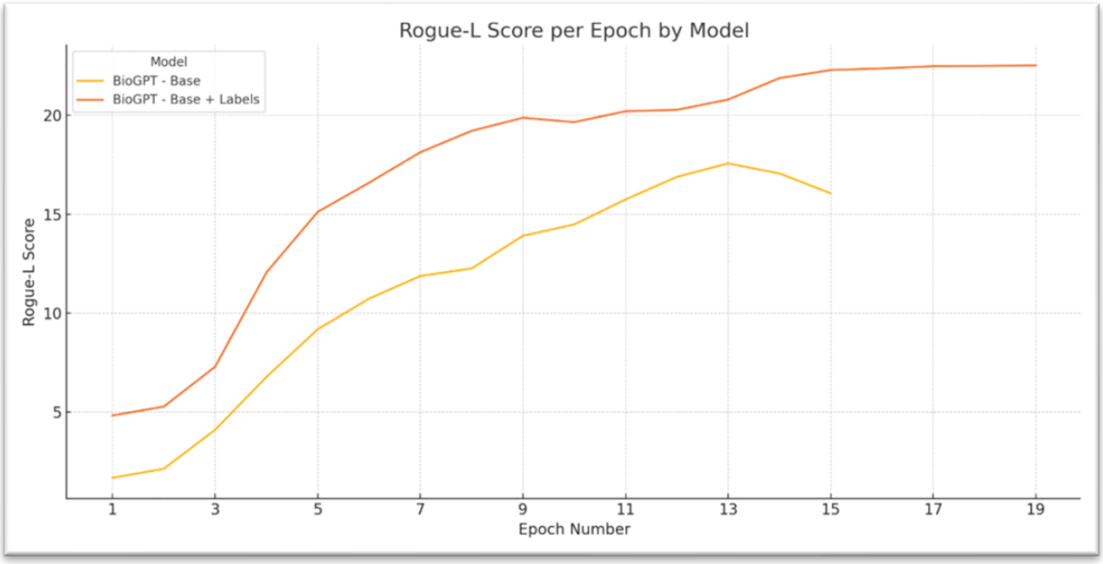


Figure 3: ROUGE-L Scores for Base Architecture vs BioVilt + ChexNet + Alignment + BioGPT Architecture

Sample Result with BioVilt + ChexNet + Alignment + BioGPT Architecture

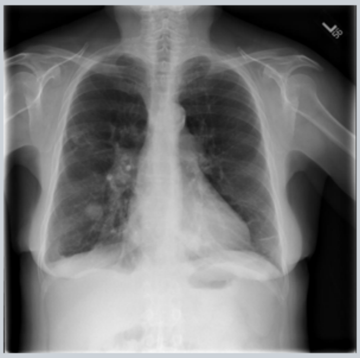
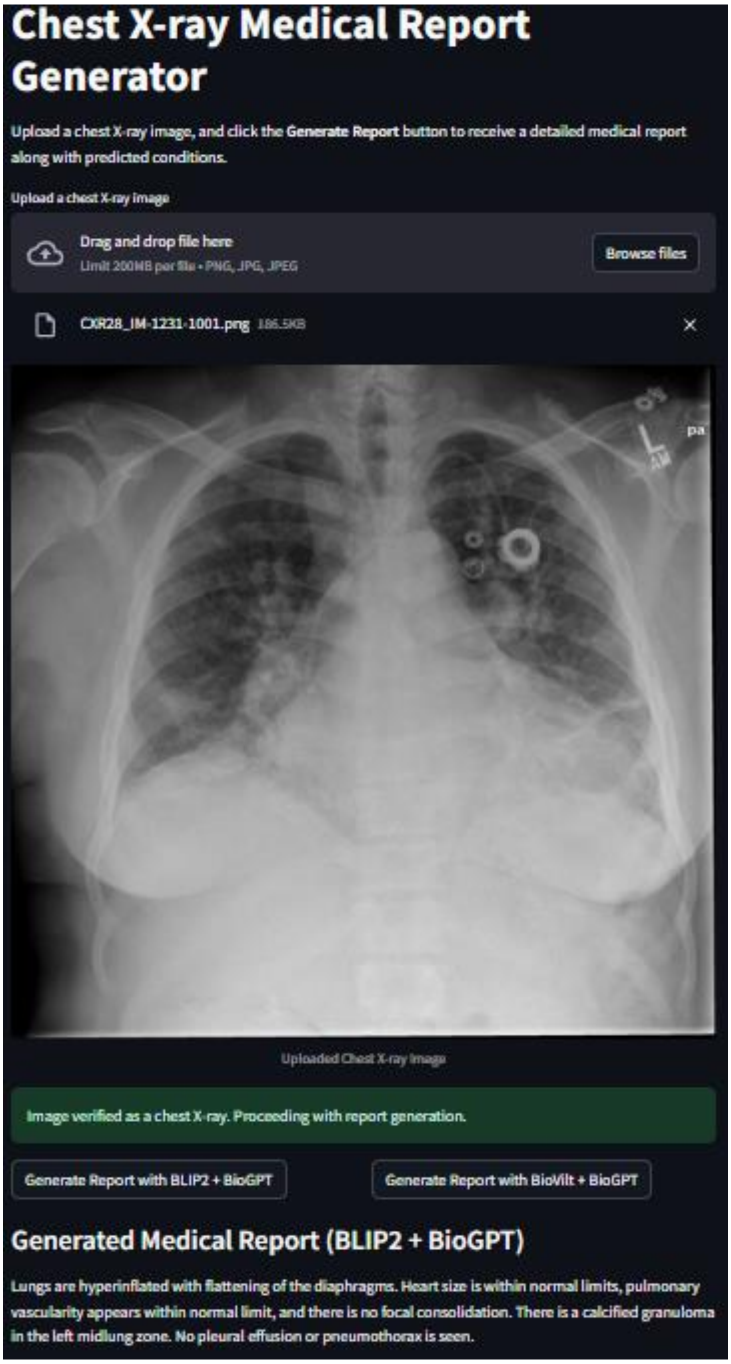
Chest X Ray Image	Ground Truth	Prediction
	Clinical Findings: Opacity Report: there is a large right hilar opacity, xxxx in the posterior segment of the lung. the heart size and mediastinal contour are normal. no pneumothorax or pleural effusions. this appears to be hyperinflated no focal airspace consolidations.	Clinical Findings: Opacity Report: there is a rounded opacity in the right lower zone measuring 2.0 cm which is xxxx to be in the posterobasal segment. there is of uncertain etiology but would benefit from followup at xxxx some concern for neoplasm. a xxxx is recommended. no airspace disease, effusion or cavitary nodule. normal heart size and mediastinum. visualized xxxx of the chest xxxx are within normal limits.

Figure 4: Sample report for BioVilt + ChexNet + Alignment + BioGPT Architecture

StreamLit Application:



Chest X-ray Medical Report Generator

Upload a chest X-ray image, and click the **Generate Report** button to receive a detailed medical report along with predicted conditions.

Upload a chest X-ray image

CXR28_IM-1231-1001.png 186.5KB

Uploaded Chest X-ray Image

Image verified as a chest X-ray. Proceeding with report generation.

Generated Medical Report (BLIP2 + BioGPT)

Lungs are hyperinflated with flattening of the diaphragms. Heart size is within normal limits, pulmonary vascularity appears within normal limit, and there is no focal consolidation. There is a calcified granuloma in the left midlung zone. No pleural effusion or pneumothorax is seen.

Drag and drop file here
Limit 200MB per file • PNG, JPG, JPEG
Browse files
pa
L
Generate Report with BLIP2 + BioGPT
Generate Report with BioVilt + BioGPT

Figure 5: Stremlit Application

Explanation of Results

- **Initial Epochs (1-3)**: Demonstrated a gradual increase in ROUGE-L scores, indicating the model's initial learning phase.
- **Mid Training (4-10)**: Marked a significant improvement in performance, with ROUGE-L scores exceeding 15 by epoch 5 and reaching above 19 by epoch 8.
- **Later Epochs (11-19)**: Achieved peak performance at 17th epoch with ROUGE-L scores stabilizing around 22.5, training was stopped here.

Explanation: The plot illustrates the steady increase in ROUGE-L scores, highlighting the effectiveness of integrating ChexNet with BioVilt and BioGPT. The plateau observed in later epochs indicates the model's convergence.

Comparative Analysis

- **Baseline (BioVilt + Alignment + BioGPT)**: ROUGE-L scores peaked at 18.06, indicating moderate performance.
- **With ChexNet Integration**: Substantial improvement to ROUGE-L scores of up to 22.51, demonstrating the value of structured findings in enhancing report generation.

Explanation of Sample Reports

Explanation: The generated reports exhibit improvement and accurate descriptions of the CXR images, closely aligning with the target impressions. It can understand the exact disease depicted in the CXR image. This qualitative assessment complements the quantitative ROUGE-L scores, validating the model's effectiveness.

4. Summary and Conclusions

Summary of Results

My contributions to the project involved developing a script for **extracting multi-label structural findings** from medical reports, training the **BioVilt + ChexNet + Alignment + BioGPT** architecture and the development of **Streamlit application**. The integration of ChexNet significantly enhanced the model's performance, as evidenced by the improvement in ROUGE-L scores from 18.06 to 22.51 over 19 epochs. This improvement underscores the importance of incorporating structured findings into the report generation pipeline.

Lessons Learned

- **Importance of Structured Data**: Integrating multi-label classifications provides valuable context, enhancing the quality and accuracy of generated reports.
- **Model Integration**: Seamlessly combining different models (BioVilt, ChexNet, BioGPT) requires careful alignment of embeddings and thoughtful architecture design.
- **PEFT Techniques**: Utilizing Parameter-Efficient Fine-Tuning (LoRA) effectively reduces computational overhead while maintaining high performance.

Suggested Improvements

- **Multi Image Input:** The initial plan was to use multiple images as input but due to less data availability we only used single image as input.
- **Hyperparameter Optimization**: Further tuning of learning rates, dropout rates, and LoRA configurations could yield even better performance.
- **Incorporation of Additional Models**: Exploring more advanced image encoders or language models may enhance report generation quality.
- **Enhanced Data Augmentation**: Implementing more diverse augmentation techniques could improve model robustness and generalization.
- **Evaluation Metrics**: Incorporating additional evaluation metrics, such as BLUE and METEOR scores, would provide a more comprehensive assessment of report quality.

5. Calculating the Percentage of Copied Code

Lines Copied: 715

Lines Modified: 0

Lines of Code written: 1684

Percentage of Copied Code: 42.45%

6. References

[1]. ChexBERT

[2]. ChexNet

[3]. BioVilt

[4]. BioGPT

[5]. RaDialog-RG

[6]. METransformer