Chest x-ray Report Generation and Chatbot

CMPE 258, Spring 2023

Presented by,

Team Ctrl Alt Del

Project Walkthrough

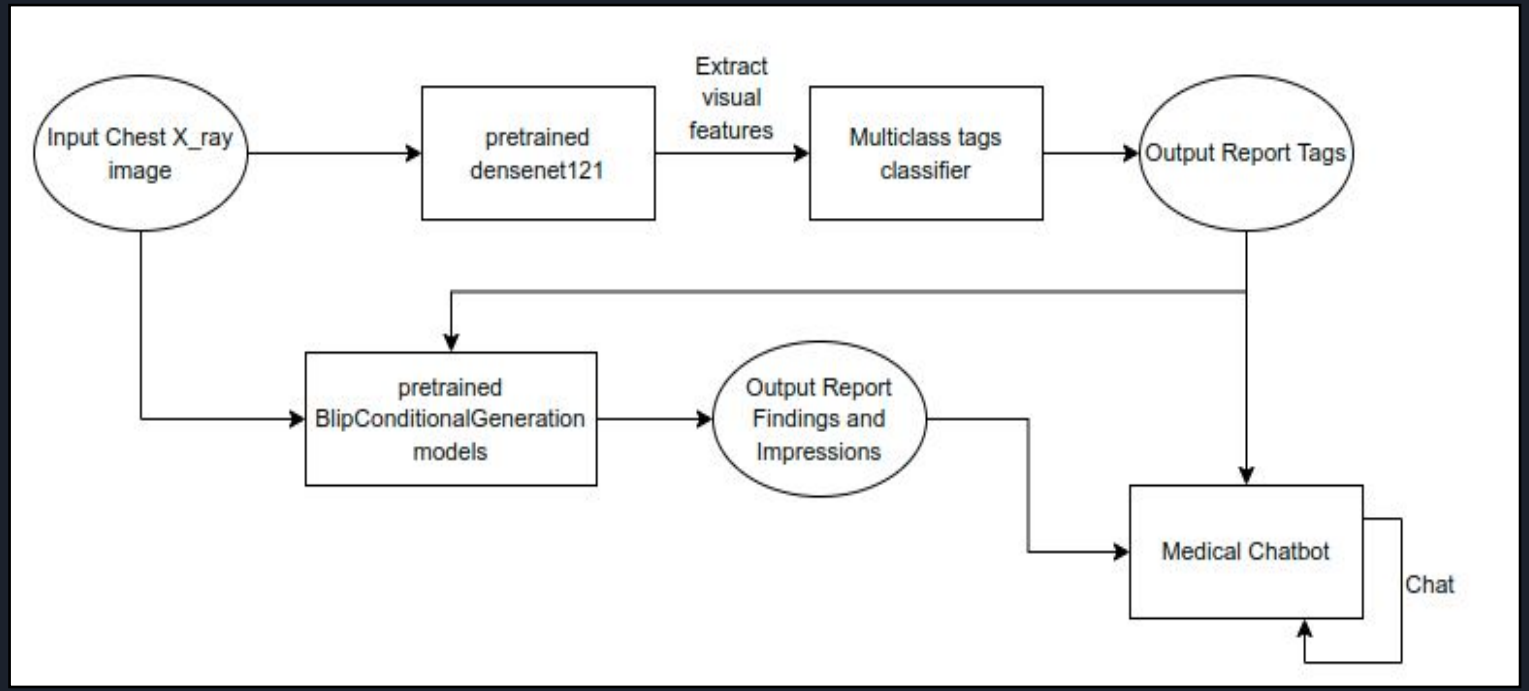
- Introduction
- Overall Process Flow
- Report Tags Generation
- Report Findings and Impressions Generator
- Evaluation Metrics
- Chatbot
- WebApp

Introduction

Chest X-rays are a common diagnostic tool used in detection and evaluation of various pulmonary diseases. However, generating reports based on these images can be time-consuming and prone to human errors, leading to delays in diagnosis and treatment.

To address this problem, we chose to build an AI tool that would read an image and generate tags from the image features and generate a report(impression and findings)for that image based on chest image features and tags. In this project, we use a dataset of chest X-ray images and their corresponding radiology reports. We then fine tune deep learning models which will extract relevant information from the images and generate corresponding reports.The generated reports (impressions and findings) are evaluated for the hamming loss and semantic similarity with the ground truth reports.

Overall process flow



Report Tags Generation

Model

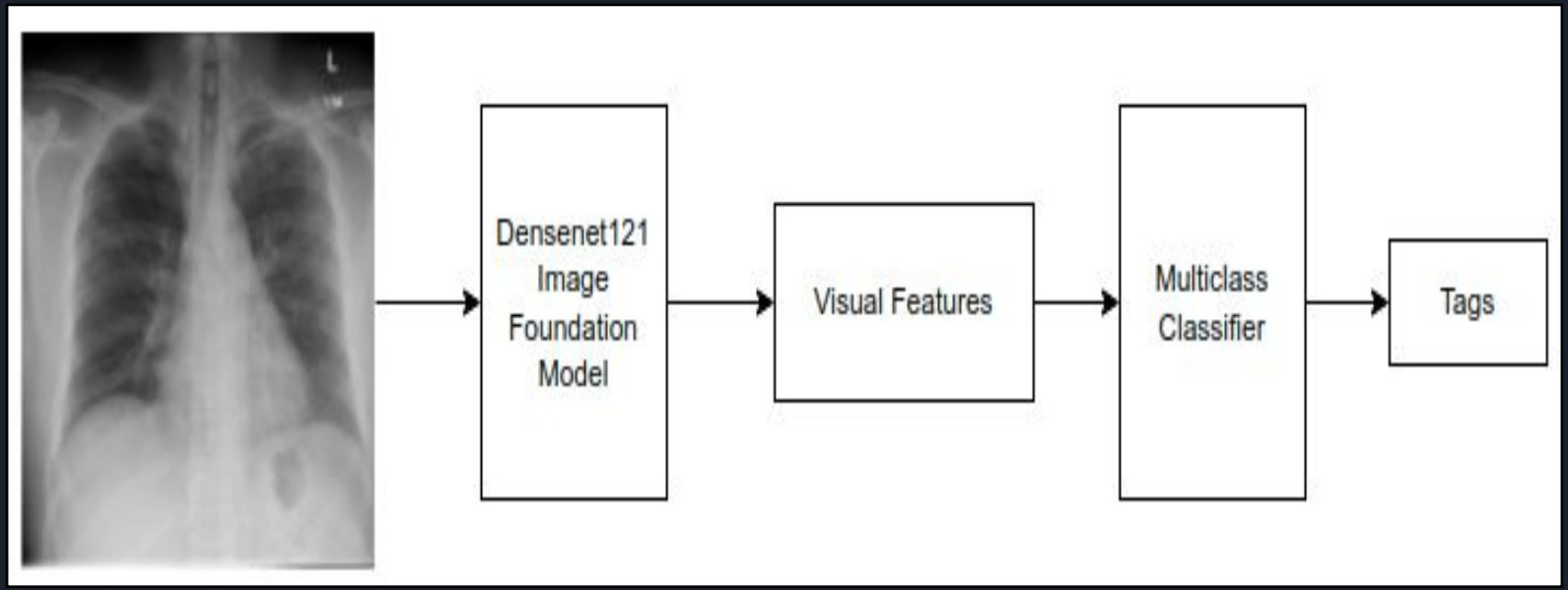
One of the parts of the x-ray report are the tags which list the critical keywords found in the findings section of the report. To generate these tags, we use the pretrained VGG-19 vision foundation model followed by a linear layer classifier for predicting the tags associated with the input x-ray image.

Methodology and Implementation

Tag generation involves two steps:

a. Extracting visual features from chest x-ray image
b. Using these visual features to predict the associated tags

Report Tags Generation



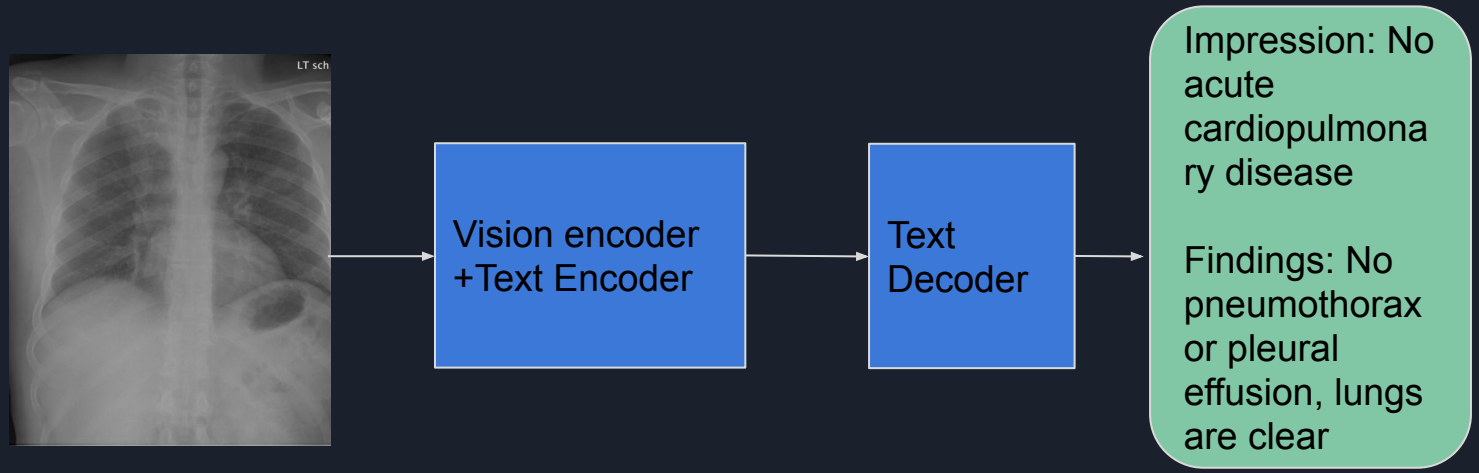
Report findings and impression generation

Model

In addition to tags, a chest x-ray report also consists of findings and impressions. The impression section provides diagnosis and the findings section lists the observations regarding each body part examined in the imaging study. For the purpose of generating impressions and findings, we use the BlipForConditionalGeneration model from Hugging Face transformer architecture based on the Generative Pre-Trained Transformer (GPT) architecture. For our purpose of generating text from the image of the chest x-ray, we use the generate-cxr model which is modified for taking into account both image and text inputs to generate output text. Internally, it uses a Blip Processor to achieve this.

Report findings and impression generation

A new transformer architecture particularly suited for this type of multi-modal problem was just released by SalesForce (Li et al. 2022). BLIP has a dual text and vision encoder paired with a text decoder. This allows it to continue generating new text for a radiology report from a given prompt's starting point.



Evaluation Metrics

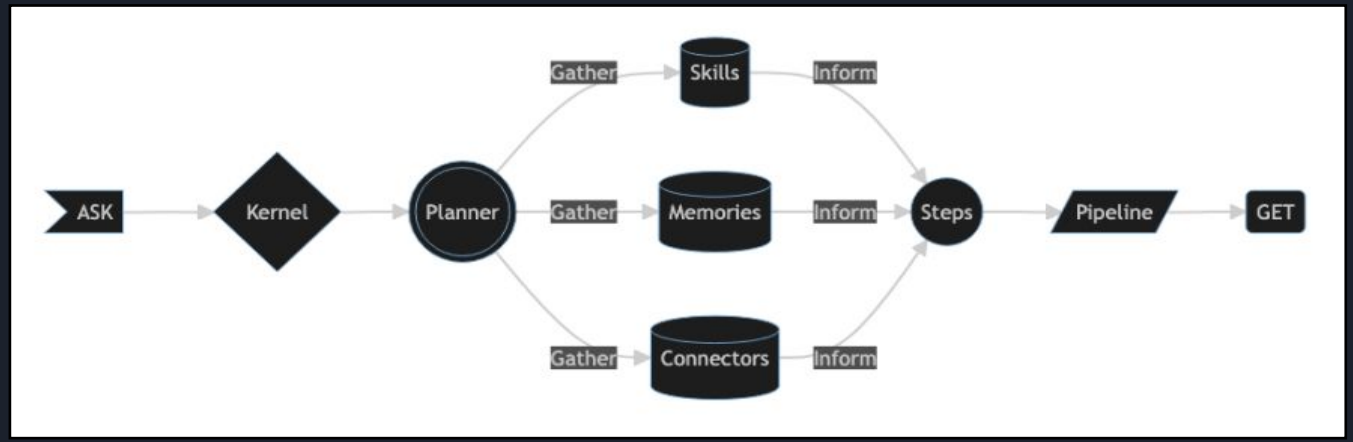
To measure the semantic similarity between the generated impressions and the ground truth impressions, as well as the generated findings and ground truth findings, we embed them using the medium sized corpora from spacy and then perform semantic similarity.

A higher similarity score suggests that the predicted and actual report components match closely.

We also used hamming loss that calculates the fraction of misclassified labels/tags.
i.e., the fraction of tags that are predicted incorrectly, compared to the true tags.

A lower hamming loss suggests that the tags generated are closer to the ground truth tags

ChatBot using semantic kernel



ChatBot using semantic kernel

- A user’s query is sent to a semantic kernel.
- The kernel orchestrates this query, the planner in the above diagram basically breaks down the query and gathers available resources that are in skills(typical questions - prompts), memories(context) and connects them. Basically it leverages the available resources from the populated memory.
- It then executes the steps in the pipeline to get the response for the query and sends it to the user.

Demo the streamlit app

Conclusion

The goal of the project was to generate reports from chest x-rays where the only available information one has is the chest x-ray.

By leveraging the developments in large language models for both text and vision, we aim to reduce the burden on physicians of creating these reports manually.

Additionally, using a semantic kernel for building a chatbot that can help a patient in answering questions about their report can be a big boost especially in scenarios where medical professions are not immediately available for translating the report to the common man.

Thank you