Project Title: Medical Image Captioning for Chest X-Rays

Team members:

Full name	ASU ID
Amey Bhilegaonkar	1225368924
Disha Agarwal	1225441776
Ninad Nale	1225710226
Janaki Venkata Ramachandra Sai Nayani	1225418207

Problem statement:

ML-based interpretation of medical X-Ray images for purposes such as generating medical reports, and answering diagnoses-related questions are gaining popularity. This involves processing image data to not just detect objects in an image or classify an image but also identify relationships between different objects in an image and generate a natural language response, hence is a challenging problem.

Most research has explored caption generation or Visual Question Answering(VQA) as separate tasks and from the perspective of use by medical practitioners. This project proposes to build a single model for both tasks and convert the task of captioning also to a prompt-based input so as to take input from the patient and practitioner as to what specific/detailed information they want to know.

Further, the aim is to explore state-of-the-art models like VisualBERT for this task in addition to the existing methods. The extension of the project aims to explore methods to mark the region in the image which influences the answer the most.

Related work:

Existing methods have explored Retrieval based, Template based, and Generative models for the task of generating captions. There have been comprehensive review papers on the existing literature also. Below are findings from a few recent papers:

1) **Automatic captioning for medical imaging (MIC): a rapid review of literature[1]:**
Is an all-encompassing survey paper on biomedical image captioning. They have used the web of science, Scopus, Medline, and ImageCLEF to identify papers filtering them based on an inclusion criteria. Most papers have used the IU chest X-ray dataset and the Imageclef dataset. The work explored papers consisting of several techniques grouped as template-based, retrieval-based, generative-based, and hybrid models. It concludes that medical image captioning is a challenging task due to the nature of medical images and reports which are different from natural images and generic captions. However, hybrid models that combine generative and retrieval-based models, perform better.

2)**A Sequence-to-Sequence Model Approach for ImageCLEF 2018 Medical Domain Visual Question Answering[2]:**

There exist fewer papers on Medical VQA as compared to the general domain, this paper aspires to reduce this gap and hence nominated a new approach for the medical-image-to-question for ImageCLEF 2018 VQA-med dataset. It uses VGG net for image feature generation and 3 GRU layers for encoding the question and 3 GRU layers(signifying captioning, translation, and the last layer used for the final answer) for answer generation. Their model had higher BLEU and WBSS scores than other teams participating in the ImageCLEF challenge.

Initial hypothesis:

The main objective of this project is to come up with a multitask model for generating captions for medical X-ray images as well as answering diagnostic questions based on a given image. Methods we want to explore:
1) Image features using Densenet121 and Xception
2) Generating captions/answers using simple encode-decoder framework
3) Generating captions/answers using attention-based framework
4) Generating captions/answers using VisualBERT

It is expected that using data instances from the caption and VQA datasets should increase performance on both tasks.

Datasets:

We will be using 2 datasets(as of now) for this project.

1) Visual Question Answering in Radiology is a VQA dataset wherein questions and answers are collected by various clinical trainees through web UI. The Dataset has 2248 images and different types of questions(size, presence, abnormality.etc.) and answers(open ended/closed) related to them. This Data set is collected by Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda. In this dataset, each image represents a different person.

2) Indiana University Chest X-ray Collection is a dataset that has images and captions given to the images by health professionals. It has 7470 X-ray images and 3955 Radiology reports associated with them. The Images are in DICOM format(Digital imaging and communications on medicine), preprocessing might be needed to change the orientation. The dataset was created by Indiana University(IU) for research purposes but not clinical purposes. This dataset is the most widely used dataset in research related to medical captioning.

Management plan:

Task Name	Subtask Name	Assignee	Due Date
Identify the problem statement, define project scope, objectives, and deliverables	Explore problem statement and project objective.	Disha	28/02/2023
	Explore datasets that exist for the problem statement.	Ram	
	Define project communication and collaboration.	Ninad	
	Define project plan and timeline.	Amey	
Literature Survey	Explore arxiv, ACL, Springer, IEEE, and other conference papers.	Everyone	20/03/2023
EDA and Preprocessing	EDA Data preprocessing on IU chest X-ray collection	Amey & Ninad	03/04/2023
	EDA and Data Preprocessing on IU VQA-RAD	Disha & Ram	
Model Development	Implementing VisualBERT	Disha	15/04/2023
	Implementing simple encode-decoder framework	Ram	
	Implementing Densenet121 and Xception for image feature generation	Ninad	
	Implementing attention-based framework.	Amey	
Develop and testing a user interface for the application - Stretch Goal	Implement the prompts and user input	Disha & Ram	20/04/2023
	Integrate the trained models with the user interface and deployment	Amey & Ninad	
	End-to-End Testing- finding bugs, performance issues and resolving them.	Everyone	
Project Presentation	Prepare PPT slide for the presentation	Everyone	23/04/2023

References:

[1] Beddiar, DR., Oussalah, M. & Seppänen, T. Automatic captioning for medical imaging (MIC): a rapid review of literature. *Artif Intell Rev* (2022). https://doi.org/10.1007/s10462-022-10270-w

[2] R. Ambati and C. Reddy Dudyala, "A Sequence-to-Sequence Model Approach for ImageCLEF 2018 Medical Domain Visual Question Answering," *2018 15th IEEE India Council International Conference (INDICON)*, Coimbatore, India, 2018, pp. 1-6, doi: 10.1109/INDICON45594.2018.8987108.