Medical OCR + Segmentation

*Abstract*—An application to develop an OCR which is capable of extracting information from Medical Reports accurately and effectively. The target being on inclusion of poor qualities

# Introduction

Segmenting of medical scans using medical reports as a context aims to improve Diagnostic capability and efficiency of doctors. A fined tuned ML model to segment alone itself is an option but having textual data helps the model to narrow down further for more accurate segmentation and detection. In this solution I propose a novel approach utilizing the best optimal problems. We are using ML Models and object detection models together to identify important tabular data and descriptions together and pass it to an LLM to extract important attributes and structures from the reports. Using this Attributes we predict top k possible diseases or diagnoses. This is done by RAG/ Fine tuned model. The images which are labeled with the probable diseases are selected. (Data preparation is required here. The image which is selected is suppose to have masked boundaries or boxes on the scan. )

Now the image is used as a prompt for the scan corresponding to the medical reports. Here for this prompt-based segmentation a fine tuned SegGPT is required. SegGPT is a model trained for variational of segmentation task. It takes an image and a prompt segmented image as an input and returns a segmented image as an output.

# OCR

## Image to Text

There are numerous models for text extraction from an image. However, for a medical record it is essential to have an awareness of the tabular data, descriptive data, etc. Because having only a high accuracy of OCR would not be able to interpret tabular data effectively. For now, I have used Pytesseract as the base ocr model. It is a pretrained model by google trained on a large dataset with proven high accuracy. The steps which we follow is :

1. Decide on the types of the report. (tabular/descriptive) or both
2. For tabular or mix we detect the regions and do the annotations
3. Preprocess the image (Not included) including contrast, grayscale and noise reduction
4. Use the YOLO v5 model to identify different sections in a report(table, description, name, etc)
5. Use tesseract to extract text information from the reports and store it in a csv sheet.

## Pdf to text

Here’s a bit different approach to my method. We use pdf2image library to convert a pdf document to an image and then follow the same procedure. The reason for the conversion is that

1. Many times, the reports in pdf are just collection of images. A pdf parser tends to fail to extract any information from that.
2. Content awareness. It is essential to identify the different fields or sections present in the document. Pdf parser does not provide that capability and might result in a non-segmented text data. Yolov5 does it for us in the image

Hence After conversion we follow the same procedure.

# extraction of attributes and relevant information

Now we have an unstructured text data extracted from the reports. These data cannot be directly provided to a segmentation model. A medical LLM is required. I have attached a sample dataset which can be used to train a Palm-2 model for diagnostics prediction. (A more relevant dataset would have been better in prediction).

## GPT-3.5-turbo-0613 function calling

We use this model to extract some important attributes from a scan. Labelling fields such as (lung function, CBC, etc) and providing description enables the gpt to search for the particular values just like what a doctor does. Function calling also enables the return of structured data in json format. Hence allowing a more persistent data flow each time.

## LLM(Palm-2/GPT/Llama2)

After getting the structured data from the GPT we pass the information to a fine-tuned model. Various dataset can be used- Pubmed QnA dataset, Palm’s sample dataset or any proprietary client dataset.

A model returning top k probable diseases is the target instead of a single deterministic disease as the clinical judiciary discourages to rely wholly on Computer diagnostics.

# Disease-image dictionary for image prompt

A tremendous manual labelling is required to create a key value pair of an image and a list of symptoms. The aim is to map a particular image which has a scan of a person with a particular disease or syndrome. For the top k symptoms which we obtained from the LLM, we assign an image according to the ranking system to each respectively. This image is the prompt for the patient’s actual scans. A mongo dB based dictionary is the preferred choice.

# STACKED ATTENTION MODEL + SEGGPT

1. SegGPT: model based on a research paper : [2304.03284.pdf (arxiv.org)](https://arxiv.org/pdf/2304.03284.pdf)

“*SegGPT is a generalist model that unifies various segmentation tasks into a single in-context learning framework. The architecture of SegGPT is designed to accommodate different kinds of segmentation data by transforming them into the same format of images1. The training of SegGPT is formulated as an in-context coloring problem with random color mapping for each data sample1. The objective is to accomplish diverse tasks according to the context, rather than relying on specific colors1.*

*After training, SegGPT can perform arbitrary segmentation tasks in images or videos via in-context inference, such as object instance, stuff, part, contour, and text1. SegGPT is evaluated on a broad range of tasks, including few-shot semantic segmentation, video object segmentation, semantic segmentation, and panoptic segmentation. The results show strong capabilities in segmenting in-domain and out-of-domain targets, either qualitatively or quantitatively.”*

If segmented images are available it is highly recommended to train the model to have a more precise segmentation.

1. Stacked Attention Network. Research paper : [Stacked Attention Networks for Image Question Answering (thecvf.com)](https://openaccess.thecvf.com/content_cvpr_2016/papers/Yang_Stacked_Attention_Networks_CVPR_2016_paper.pdf)

SAN is used for Image based Question Answering.

Combining these two gives a perfect Prompt and text contextual Medical image segmentation that takes the reports of a patient into consideration.

Our model gives output segmentation for the top k possible diseases. Hence giving the ultimate judgement in the hands of the doctor, a feature which although is a small thing but most necessary in a clinical project.