

EEP 596: AI and Health Care || Conceptual 1

Dr. Karthik Mohan

Univ. of Washington, Seattle

May 12, 2022

Read Me

This assignment gets you thinking more freely on the concepts we have discussed so far in class. But it also does have couple of coding questions that you are free to use any method/library for. Include code with your submission and also include insights where applicable. For each problem, clearly state your assumptions, data set you are using (if any) and metrics you computed (if any).

Problem 1: Radiation Therapy (15 pts)

In radiation therapy, isolated cluster of cancer cells are usually targeted. Given a pet scan image of the body with dense white patches indicating cancer cells - what would be your algorithm/method to help detect cancer cells suitable for radiation therapy? Clearly mention the steps in your method and the reasoning behind it.

Problem 2: Prioritizing patients (15 pts)

A clinic would like to prioritize patients for pre-emptive screening of cancer based on an automated assessment of the patients health. Doctors agree that preliminary indicators for further assessment come from the complete blood count history (wbc, platelets and hb counts among others) and any other abnormalities noticed in the patient history. Given that the clinic has data on patients detected for early stage cancer (of various types) and their past history - How would you go about building an automated assessment method for the clinic to help with the pre-emptive screening? Discuss data, ML methods and metrics that you would use to build the assessment method.

Problem 3: Generating Images (35 pts)

Pick any set of medical images of the same category (e.g. mri scans or ct scans or biopsy slides) that are publicly available and in a sufficient quantity. Train a CNN autoencoder-decoder architecture on these images. Take any 5 random images and pare them through the neural network to then produce 5 new images. Plot the two sets of images side by side. Also use a suitable image similarity metric to measure the similarity between the two images. Does your method for data augmentation seem to work for the data set you picked? How could you improve the quality of the augmented data, if not? Include your code along with the submission.

Problem 4: Summarization (35 pts)

One of the important aspects of automating an Electronic Health Record (EHR) using the digital scribe is to summarize the medical notes into a structured EHR. Assume for now that we are doing **extractive summarization** - i.e. picking specific sentences from a document to include in the summary (without modifying it or adding words to it). Pick any publicly available summarization data set - i.e. a data set that has a document that is paired with its corresponding summary, as target. Train a model, which when given a document can generate a summary from it. Clearly mention the data set you will be using for training. What model architecture are you using for extractive summarization? (you can pick any of the available libraries to play with as well) What metrics would be suitable to compare your generated summary against the ground truth summary for a document? Include the code and also the metrics along with your submission. How good is your extractive summarizer? How would this extrapolate to a medical notes data set?