p9120 Final Project Proposal

Zongchao Liu | zl2860

11/17/2020

Introdcution

This project is a real data analysis on 2D imaging data and patients' clinical outcomes.

Data

Here, we are using real world data. The patients in our data have enrolled in the FOWARC (NCT01211210) clinical trial.

The data we use for this project has two parts:

- patients' imaging data(original 2D CT and MRI images)
- patients' raw clincial outcomes

In this project, we will firstly process the raw imaging data by extracting 2D imaging features from each CT/MRI image, and then select the most predictive features for our prediction in pCR. To make the prediction more interpretable, we are going to combine the imaging features with the clincial outcomes to build and compare the machine learning models.

Project Description

The data is from a research project. The main goal is to use machine learning methods to predict the pathological complete response (pCR) in patients with rectal cancer after neoadjuvant treatment. We will compare different machine learning models and select the best one for the prediction.

Methods

- For predicting pCR, we plan to use and compare logistics regression, SVM, random forest and gradient boosted machine
- If we have time, We also want to explore the capability of a simple structured neural network to predict pCR by using only the imaging features. We are not going to apply networks with convolutional layers due to the limit of computational memory. Instead, we would only implement a simple multilayer perceptron in R. We expect the performance of this neural network on the imaging features will be bad due to the overfit problem. The resaon is that we have only less than 200 subjects' data.