Meeting's report for ELLEN WANG

Honour's student

06-12-2021

1. Meeting Agenda

- This meeting was with Audrey, meeting minutes are with Pierre.

2. Meeting Minutes from last week.

- What files is the data stored in? NIFTI more supported, DICOM not so much.
- Neuroimaging in python - NIPY documentation read this.
- USE MRICRON TO visualise the data.
- Why are there different structure of random forests and algorithsm? What makes them different? Address this new data type.
- May need to modify the data to address different types of problems. Process thoughts that can lead to different versions of the same algorithm. Apply the same process of thought ot our own dataset and problem.
- Brain image data has a very specific structure and that's what we should be aiming for.
- WHAT IS RESEARCH:
  - Research is a process, so you need to work to very clearly define what your research question is. Why is your question important?
  - IMPORTANCE OF LITERATURE RESEARCH: What have all the researchers before you done? Dig a bit deeper into methods, techniques that they have used to address this question or similar questions in different context. Gives you a clue to improve what you initially thought you could be doing in that context.
  - Need to understand the limitations of random forests too.
  - Point of honours is to pick up research skills, read a lot of articles and understand what other people have done.
  - Try to improve that somehow based on what other people have done.
  - At least we need to see that process in your thesis somehow.
  - You need to have a good idea of all the different random forests which have been released, where they are made better to deal with medical images.
  - At the very minimum, you need to have a nice literature survey for random forests and brain images but doesn't seem like a big honours' thesis. Therefore you should try and build the random forest. Do go a bit faster because Melinda did a lot of work before you.
- THESIS STRUCTURE:
  - Big chapter on Random Forest algorithms (existing), explain what they are and how they originated, what are recent developments? Refer her work and compare the two approaches.

- USING SKLEARN CODE:
  - Give credit to whoever you are borrowing from, rewrite everything yourself but we are selling that the random forest is more advantageous to other ML techniques, and explain what these techniques are.
  - IE in terms of interpretation and feature importance it is much better, in terms of predictions it might also be better. Did it do better than Melinda's thesis in terms of predictions.
  - Melinda made a mistake in the way she split her train/test split. She modified the thesis somehow afterwards. Might also have been in the way she sampled her data, what exactly was this error so I don't make the same mistake?
- OTHER:
  - Common pitfalls: starting to work very hard near the end not at the beginning, bit of trouble because of lack of time at the end. Too good to be true results are just that!

4. Planned work after this meeting

- Email Melinda: where did she get her slices from, cc Pierre.
- Start scoping out how you want to structure your thesis.