

Individual Portfolio

Week 2

This week, we all got into our groups for the project and met our mentor (Simon). We learnt about what kind of data we were dealing with, and the tools that would be required to process it. Because a lot of the tools and techniques Simon told us about were all quite novel to us, we decided to list down as many as we could, so that we could come back to it at another time. The main focus of the tools being given were using neural networks and image recognition in order to recognise the given images and classify them accordingly. After learning about the project and the different tools which would be needed to make models, we went back to our groups to finish off the administrative details. We very roughly assigned tasks to each other, and exchanged contact information to get everyone together. Furthermore, we briefly thought of the research questions and listed down all ideas, to get a basic understanding of the direction we wanted to go in. This was done during the rest of the class, as we did not yet have proper access to the data just yet. In addition, we were working on familiarising ourselves with google colab, as it was a new program we were required to use. During the week, we made a collaborative google doc to add any ideas to the project as we go along, I added some possible research questions and how to apply the tools we learnt during class. It was clear however, that during this week we were still quite unsure what we wanted to do, how we wanted to do it and what to do.

week 3

At the start of week three, unfortunately we still hadn't gotten access to the data, which meant that we could not properly determine what the data looked like up close. The data would become accessible later that week, therefore in the meantime we decided to continue adding to the collaborative doc of ideas. We created a basic plan of direction on how we wanted to approach this project, although quite basic, this would serve as the scaffold of our project, where we would refer back to the plan should we forget or become confused during the progress of our project. The plan was as follows:

- Identify normal placements of catheters from the csv file
- Link to corresponding images + coordinates
- Using tensorflow develop model based on coordinates of the placements
- Train on normal placements
- If the placement of the catheter is outside the "correct zone" (heatmap made from training of the normal placement data) the catheter is considered abnormal, if it is near the "correct zone" it would be considered borderline etc. we want the model to do this.

Although not detailed and simple, we all agreed on this plan and enabled all of us to be on the same page without any confusion. This week, we also started using github, to collaborate with the other group members with coding. We were all added to a project as collaborators to ensure that we could in fact share our code and combine it.

Week 4

After gaining access to the data we were finally able to get a better understanding of what the data was like. We had access to multiple datasets, each including different information. One problem we encountered was copying all the data to drive. Because there was so much data, it took quite a while to load where we didn't have the full range of data just yet. However, because we did have a dataset to work with for now, it was what we used and focused on. Some of the work we did with the data was some basic clean up work, in other words, getting rid of duplicate data. Another aspect of the data we looked into was the failure rate of each catheter to determine which was most prone to errors. We made a graph to visualise this and added it to our presentation. I added the code to my branch, requested to merge to the main branch so the others could see. Although a basic start, discovering this gave us a good idea of what each catheter was like in terms of how error prone they are, as well as determining the most common types of catheters people get. The rest of the week was mainly focused on familiarising ourselves with the given data, where we gained an understanding of the variables. Furthermore, we started looking into and investigating how to use tensorflow for our image recognition model. Getting ourselves familiar with tensorflow would be important in the upcoming weeks, as it would be the tool that'll allow us to reach our goal in making a model predicting the placement of the catheters.

Week 5

This week, our main focus was to figure out how to use tensorflow to make a classification model. Because it is quite new to all of us, it's taking time to fully understand how to utilise it for our project. In other areas of the project, we continued working on the slides for our presentation, and added a section. However, the area we put the most amount of work into this week was continuing investigating the nature of the data and discovering the non-obvious aspects. For example, we worked on trying to figure out what the coordinates actually meant and provided, as well as figuring out how different case studies of the same patient related to one another. Another discovery made was finding a patient with no catheter at all, where they did not have any catheter of any type. This raised the question of whether we include patients with no catheter or just exclude them, as they would have no contribution in training the classification model, yet they may also be useful in making the model more robust, where it is able to determine if a patient has no catheter.

Week 6

This week, although our classification model hasn't been completed yet, we decided that if we were starting to train it, we would start small with a couple of images then work our way up as we get more familiar with it. While I was looking at the multiple sets of data given, I realised it would be good if we merge the two training sets of data as each were equipped with different sets of information. For example, one dataset included the indication of whether each catheter existed in a patient, while another one included the coordinate position of each catheter of each patient corresponding to the image. It took some time to figure out how to do this, as the number of rows also differed by quite a large amount. After searching, a simple solution which I hadn't thought of got me through this task, where I was able to merge the two csv files into one, where it included all the relevant information. The

rest of the data which did not match up were dropped, leaving me with around 18 thousand rows. I pushed this code to git for the others to see.

Reflection:

The main thing I have learnt from working as a group was that everyone needs to be working together while being on the same page. I've learnt that collaboration can come in different forms, such as everyone doing their own tasks to produce an outcome or everyone working together to solve a problem. For us, the latter has been working very well, where if members were stuck on a particular task, other members would pitch in to give ideas on how to approach said task. The main challenge when working on this project was understanding and getting used to new tools and techniques, where all the group members had a hard time attempting to implement and apply them. To overcome this challenge our group had plenty of discussion and questions, in order to fully understand the concepts and to get everyone on the same page. My contribution to the project was manipulating and processing of the data, and making sure that the data we use for the model is in the optimal form. This included checking for any duplicate data, merging datasets, and making sure that there were no abnormalities. Furthermore, I had laid out a basic plan for us to have a rough destination. In the future, it would be good for me if I got more involved and engaged with some of the other group members to make sure everyone knows what is happening and where we are at. I would do this by frequently checking and updating other team members online or in person if I get the chance.