ADS2002- Individual Portfolio

Week 2

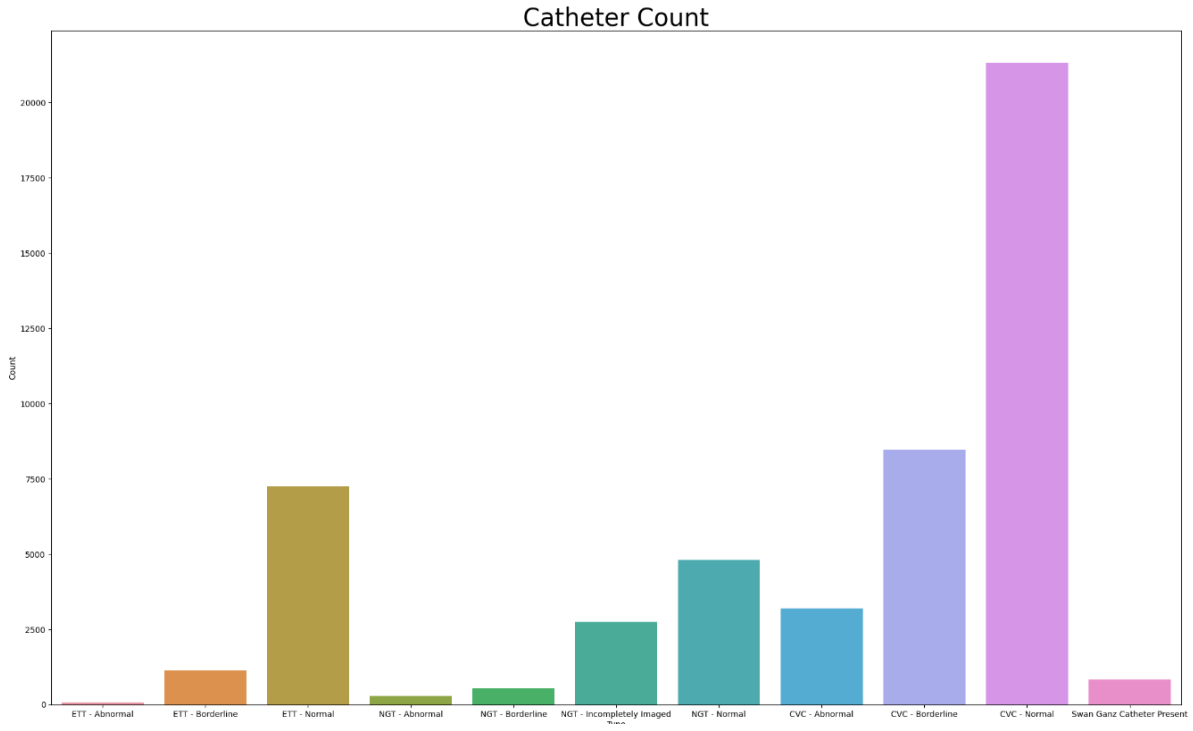
This week we first investigated google colab in which we will be conducting most of our coding. It is relatively similar to the Jupyter notebooks with which we have been working for the last few semesters. It seems slightly more difficult to actually access the dataframes as they need to be reuploaded into colab each time, however it is easier to integrate with github and so lends itself to more group collaboration which will be useful with the project. We also had our first introduction to our group members and fully explored the full project outline. We were tasked with coming up with meaningful research questions which we could ask our group mentor, Simon, in next week's class. We have explored the project introduction and basic overview of the task and were overall just brainstorming as a group possible approaches to the final task.

Week 3

This week we investigated the use of github, committing changes and pulling branches. Hopefully this will be useful further down the line with the group project work as in previous semesters projects collating all of the programming work at the end of the semester from a variety of jupyter notebooks has been difficult. Simon came into this week's class to clarify and give real details about the projects we will be undertaking. He explained the nature of the data as well as giving some recommendations for how to work with it. It will be difficult due to the size of the image data being so much larger than anything we have had to work with in the past but hopefully this leads to more learning and useful skills being developed. We decided to do some simple exploratory data analysis and form a basic understanding of the types of catheters involved as there are some technical terms which we were unfamiliar with. Following on from this, the only changes made to our group github were the creation of our own individual branches within our collaborative repository as well as the uploading of some trial programs in which we were just familiarising ourselves with the use of google colab and github.

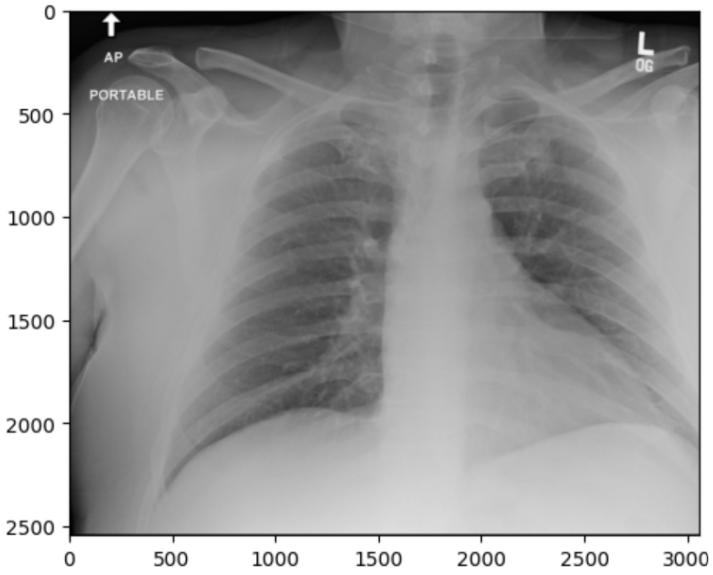
Week 4

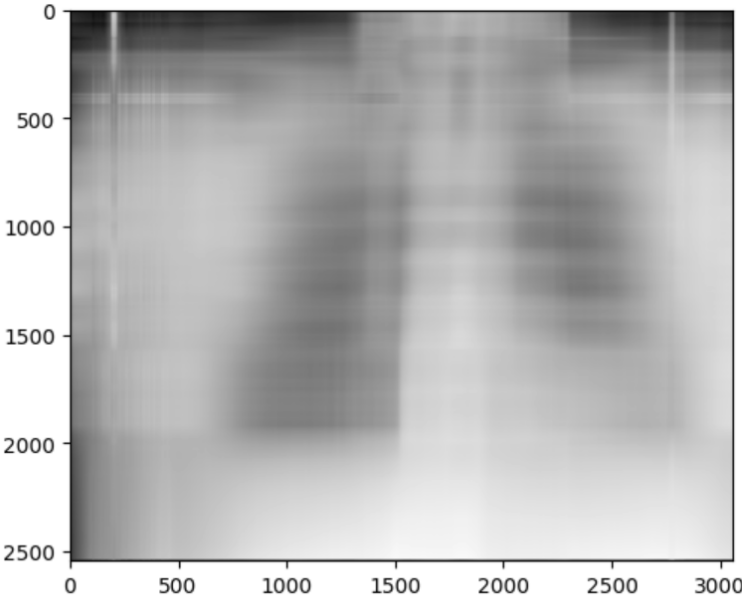
This week we were introduced to the concept of gradient descent. We viewed a video which was pretty informative and it seems to be how most of the regression models we have been using function. Following this, we finally managed to import the proper train dataset into colab to begin having a basic browse over the data and continue with further exploratory data analysis. There doesn't seem to be too much information within this dataset, just the type of catheter and whether or not it is present. However, the real challenge seems to lie within accessing the full array of image data within colab which is something we haven't had to deal with previously. We further analysed the train dataset and it became evident that most of the catheters were in the CVC - Normal definition. As a group we have also been trying to gain a better understanding of programs such as tensorflow and neural networks which may be useful in our image analysis later on but at this phase they seem a little out of scope of our current ability. We are also having difficulty accessing the full image dataset but we have communicated with Simon and this hopefully be fixed soon.



Week 5

This week through the notebook we were investigating the feature selection and how it can be implemented. While this was interesting to know it doesn't seem as relevant to our project which is more focused towards image analysis for the different types of catheters. Moving on, we can now fully access all of the images and I figured out how to load them into google colab and plot them as well. I attempted to implement some PCA on the images which we have been provided at the recommendation of my group members. Ideally this process would reduce the dimension of each of the images making them easier to process as there is a very large amount of data we need to work with. It proved difficult however to load an entire folder of images into the notebook and I am still researching how this can be done the way we need. Ultimately, I managed to get some reduction working on a single x-ray image but it also raises concerns of how accurate our model will end up being with this reduced detail as the actual catheters are already pretty difficult to distinguish in the x-rays.





Week 6

This week we spend the first half of the class time working on the imputation activity for the abalone dataset. In terms of our group project we don't have any values in the dataset that need imputation however it may be useful further in the future. In regards to the group project I began attempting to implement a k-means classification model on some of the images from the dataset. This is proving to be difficult as when researching online the majority of examples appear to be classifying sets of images with only two options, such as dog or a cat. In the data we have been provided there are eleven different types of catheter which can be found, oftentimes multiple of them being within the same image. In order to simplify the starting of this process I think it would be appropriate to create a subset of the data to work on, only trying to decide whether a catheter is CVC - Normal or not. Additionally, all of the x-ray images are not of the same dimension which means they don't work with the kmeans model so this is another issue that needs to be corrected.

1st Reflection

So far, working on this catheter group project has presented quite a challenge both in terms of the physical work as well as the group dynamics. It has been challenging managing people's expectations of one another as well the timeframe for the entire project. While it was stressful to begin with, I believe that we have developed more of an understanding of one another to produce a less stressful environment more conducive to success. Additionally, we faced challenges when two of our original group members left the class altering the team interactions and workload required. Overall, I have mostly been focused on some of the basic image analysis, like kmeans, which will hopefully provide a solid foundation upon which we will be able to develop our project and more complex image analysis models. I believe that I could have better implemented github into more work at the start of the project so that all group members are working from the same initial point, preventing issues later on. Overall, I believe we are on track to understanding and developing successful outcomes for this project.

Week 7

This week we first investigated the use of perceptrons. It was interesting to finally be working with something we haven’t encountered before as most of the previous concepts I was at least somewhat familiar with, although not completely fresh in my memory. Continuing on from this, I was further attempting to implement a kmeans model on the subset of images that i chose last week. I finally had some success with importing all of these images from their folder on google drive into colab. The command I was using also had the added feature of being able to resize the images which is useful. As well as this we also needed to make sure that the images were brought in grayscale so that there was not a third dimension of RGB value which was negatively affecting the model. It was still proving difficult to actually get the kmeans model up and running however but there is still time to fix these issues. It will be a difficult week to work further on the group project as there are multiple mid semester exams as well as the essay due this week so I don't expect too much group work to be conducted.

Week 8

This week we started taking a look at multilayer perceptrons, building upon last week's notebook. These were pretty confusing to try and understand however they are supposedly good for image analysis so may be useful in future project work. Following this Simon joined the class to discuss MASSIVE with us all and how to access and use it. I was a little unsure about some of the process but there are resources on moodle which should be able to assist me. It also seems that it will be critical to our project work as analysing even a small sample of the images on colab is very difficult. We then moved on to delivering our mock presentations. We didn’t have an overwhelming amount of data to present as it has been quite difficult to get models up and running however it was still beneficial to get feedback on our general outline and recommendations of what else we should include in our final presentation. However this did cause some stress within members of the group when comparing ourselves to other groups but I think it is important to focus on our own progress as we still have plenty of time.

Week 9

This week Simon led the class as Zach was away on leave. We initially went through the notebook on neural networks which was a little confusing to understand however it helped to have Simon explaining it directly. Following this we did a small amount of work on the notebook exercises before continuing on to further group project work. It was mentioned in the previous week's class that multilayer perceptrons could be well utilised for image analysis so I spent some time researching the possibility of this implementation on our dataset. The most prevalent issue continues to be the format of the data however as once imported in its feature array form my dataframe always seems to be missized or just unable to be used in a model. This frustration is added to as all of the online examples are initially set up perfectly with the correctly formatted data so it is difficult to find solutions. Other members of the group have been exploring balancing the data as it is predominately CVC - Normal. Simon recommended that we next explore convolutional neural networks and using resnet-50 so that will likely be where we next put our effort to hopefully get a model working on a smaller sample.

Week 10

This week we explored neural networks for image classification specifically for the MNIST dataset. I think that this notebook will require some later reviewing as it seems very useful however was confusing upon initial viewing. Moving on from this, we went into even more project work. Things are becoming slightly more tense due to the incoming due date for the finalised project however I still believe that it will come together quite well. I have been attempting to further develop my KMeans model which was utilised earlier in the semester to get it functioning over more images using MASSIVE however I have had some difficulties in doing this as some of the packages used in my colab version of the modelling aren't implemented in MASSIVE. However, I think in addition to this that it is just as important to get what we have already produced into the report and presentation as it all helps to show our process of development.

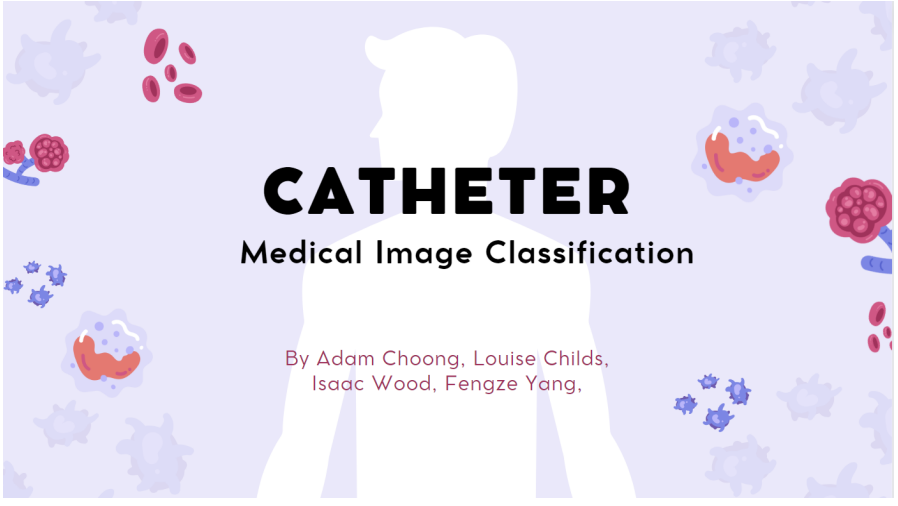
Week 11

Starting this week's class we analysed the notebook which was investigating model interpretability specifically using LIME and SHAP values. These seem like they could be quite useful to seeing what features are most critical in a model and how they affect it, however I'm not as sure how it can relate to our project as we only really analyse image feature arrays. Continuing on we moved on to further group project work. This week we finally managed to get some progress with MASSIVE and some of the image analysis models. I have got KMEans working with 1000 of the train images set and ideally would like to investigate more. We also did a lot of work on finalising the presentation as that is due relatively soon next week. It is looking quite nice in comparison to past projects I have presented, however we additionally need to ensure that the content is all well communicated and effective.

Week 12

This was the final week of class and we finally presented our findings! Overall I think our presentation went quite well. There was some tension in the group beforehand in relation to the content but the presentation was very aesthetically pleasing while also containing quite a wide variety of knowledge and exploration of the data. We will now have to wait for the marks to be released but I believe it went well. Following this, we now need to finalise our final report. There is still some content that needs to be worked upon before we can add it to the report as there were some suggestions in questions from the mentors after our presentation but as long as we all describe what we talked about during the presentation in the report it should be very effective at communicating our research findings.

Final Presentation:



2nd Reflection

This second half of the semester presented vastly more challenges within the dynamics of the group and the attitudes towards one another. Likely due to the closer due date of the entire project to be finalised there was quite a bit of stress and some displeasure with work of others within the group but I learnt how in these difficult situations it is really important to remain respectful of one another and understand that we are still all working towards the same final goal. I was quite proud of my contribution to the presentation with a variety of KMeans modelling as background research however I believe that I could have improved my organisation and time management in order to further explore more models using MASSIVE utilising that improved computing power.

Overall, this project keenly demonstrated the importance of communication within a group, effectively managing our time, as well as how critical it is to be understanding of each other in the group and doing your best to meet deadlines in order to keep stresses minimised and achieve a favourable outcome.