ECS 170 Project Check-in Team 21

Group members: Issac Craig, Chelsea Huffman, Jack Thomson, Parinita Gupta, Maggie Chen, Arvind Tawker, Kirin Kapoor

Project overview

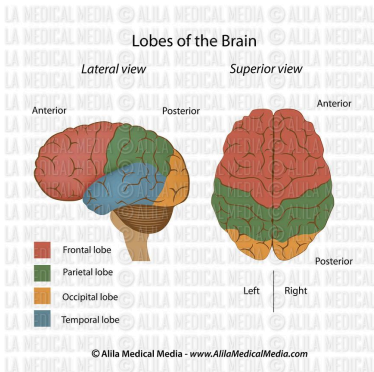
Our project has developed our idea in that instead of identifying the presence of a tumor, we want to identify which lobe a tumor is located in. We are also using a dataset that divides different types of tumor into glioma, meningioma, and pituitary tumors. Our project now aims to identify the presence of a tumor, determine if it is a glioma, meningioma, or pituitary tumor, and then locate which lobe of the brain it is located in and return potential effects. So far, we have gained clarity on the aim of our project and the datasets required to achieve our goal. We realized in order to expand our project beyond the scope of what has been done before, we need to find datasets including different views of the brain from MRI scans. We have now found an open source dataset that contains different scans of patients diagnosed with different types of tumors. These also include different views of the brain, from which we can attempt to identify the location of the tumor and the type of tissue it originates from (meningioma, pituitary, or glioma).

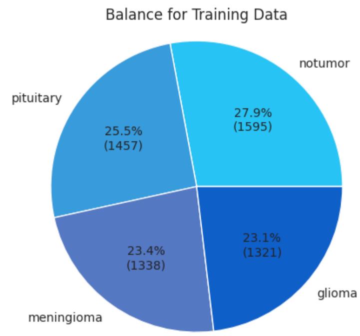
The feedback we received addressed potential biases in using datasets, especially from Kaggle. We have chosen different datasets aside from Kaggle to avoid any biases. This includes datasets that don't have super obvious tumors present, so that we are not only providing late stage tumors for our system to classify. We also hope to further explore other architectures mentioned from our graded project proposal, like DenseNets or ResNets, which are indicated to improve accuracy with limited datasets. We are also considering using more intelligent/powerful CNNs for edge detection to improve input quality.

Roadblocks and challenges

One of our biggest challenges has been finding a dataset that provides multiple views of the brain, which we will need in order to find the location of the tumor if we chose to identify the location and possible effects of the tumor. We addressed this by reading through many research articles, and trying to find the datasets that they used. We also are limited in the different types of tumors that we can identify, since visually, it can be hard for even humans to identify what type of tissue the tumor originated from. We are addressing this by doing more research into if these different types of tumors generally appear in certain regions of the brain, or if they can be identified based on location or other features. We anticipate a roadblock facing low contrast in images provided by datasets. We hope to overcome this by increasing the contrast in the images (using black point or white point) in order to more easily distinguish between the tumors and the healthy brain tissue. <u>There is also an issue with using different views of MRI scans as this can lead to complications when identifying portions of the lobes as the brain is not as easily distinguishable with images of other portions of the head.</u> *We would like feedback on how to address this!*

Figures





Left: These are the parts of the brain that we will train our model to identify, so it can tell us exactly where the tumor is located. **Right:** This figure is a pie chart that quantifies the categories of tumors in one of our datasets to determine if the dataset used for training is balanced.

Reflection

Many of us come from backgrounds that are not super familiar with brain anatomy, neuroscience in general, or cancer physiology. Through understanding how to implement AI that can identify cancer in brain scans, we have learned a lot about the mechanisms of cancer, and the different types of tumors that can form from different tissues. Additionally, some of us are less technically inclined, and have learned more about different approaches at this stage. Specifically, we have all contributed to the data pre-processing stage, which has in turn primed some of us of the work that goes into preparing a dataset for a machine learning application. Throughout this experience, we are learning a lot about the design process as well, in terms of determining what methods would address our problems in the best manner. Despite our different backgrounds, we have all learned new topics through this collaborative process as we share our areas of expertise.

Contribution & Work Division

So far, we all worked on preprocessing different datasets in order to gather ideas on what is important to consider. Our future work involves more preprocessing with our newly chosen dataset, applying a neural network, forming KNN clusters, and analyzing results. This division of work is yet to be determined.

Jack: Searched Kaggle and conducted pre-processing of the image data to create different sets and labels. Researched available datasets outside of Kaggle.

Kirin: Used a library preprocessor and created and tuned a Feed-Forward Neural Network. Raised accuracy on preliminary data from ~0.45 to ~0.60. Set up a shared virtual environment for OS, package conflict management.

Maggie: Searched Kaggle datasets, wrote preprocessing code for a certain dataset for quantitative analysis of the given dataset's glioma, meningioma, pituitary, or no tumors in the form of a pie chart.

Isaac: Searched Kaggle for datasets, researched and implemented preliminary preprocessing.

Arvind: Researched (hours), found image dataset (various angles) - classify effects of tumor.

Chelsea: Preprocessed image data from a Kaggle dataset so that all the images were stored in a uniform format. Also added labels for the images provided so data would be ready for training and testing.

Pari: Brought extensive domain knowledge on neurobiology to identify limitations of datasets and explore avenues of AI classification in terms of actual medical impact.