Can we utilize data to build an ML model that can reduce Brain Cancer mortalities?

BrainStation Capstone Report

Author: Bilal Munir

Due: June, 2023

Problem Statement: Can we leverage machine learning to provide timely diagnoses of brain cancer by detecting tumors prematurely?

Magnetic Resonance Imaging (MRI) is a medical imaging technique that uses magnetic fields to generate images of the internal structures of the body. It is useful in providing medical insights and **diagnostic information** to specialists so they can determine the appropriate course of action in treating their patients.

The survival rate for cancer in the past ten years has been around 30%. Unfortunately, you cannot prevent brain tumors by simply changing your lifestyle (diet, exercise etc). According to the American Cancer Society, you can reduce the risk by limiting radiation exposure to the head.

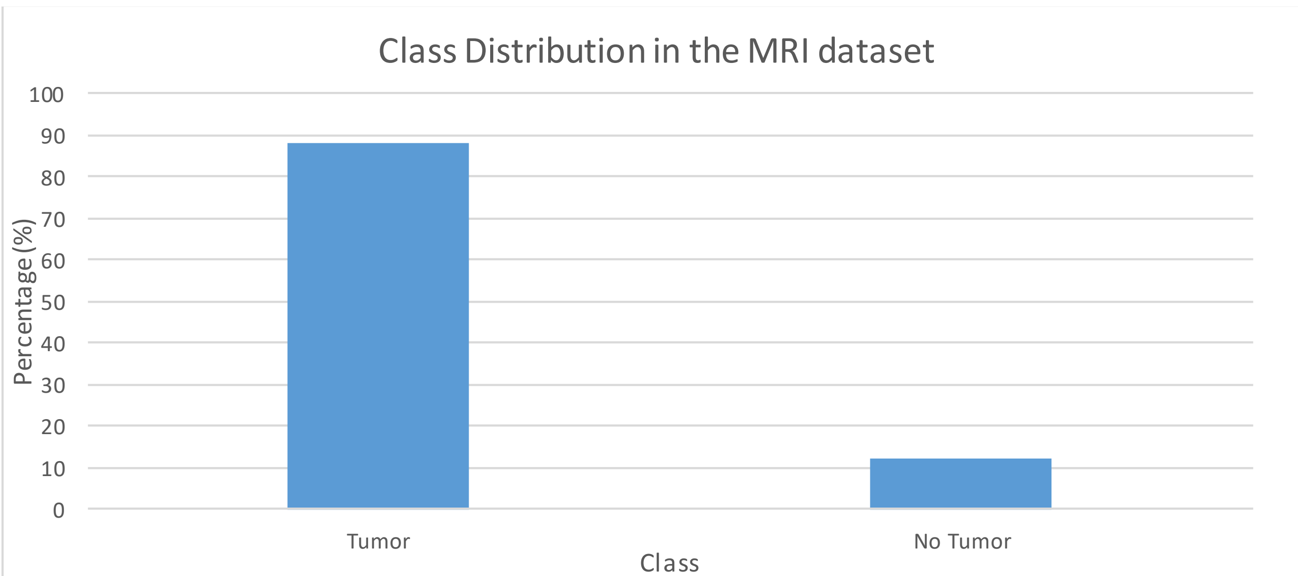
Since there is no way to prevent it from happening, the best solution is to identify the anomalies early and begin treatment as soon as possible. This project will hopefully serve as a small step towards accurately detecting the presence of tumors and in the future, be able to classify them, thereby speeding up the diagnosis.

Dataset:

This notebook explores a dataset that contains almost 4500 Brain MRIs. It was obtained online from a diagnostic imaging clinic in Brazil.

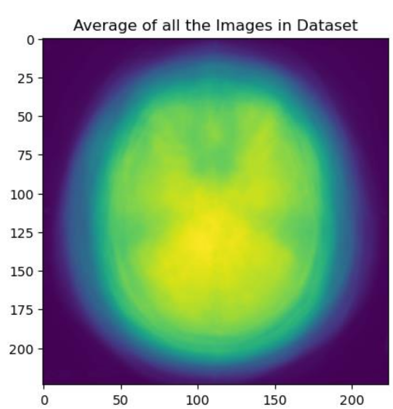
Dataset Link: https://www.kaggle.com/datasets/fernando2rad/brain-tumor-mri-images-44c

The dataset contained brain images of 15 different categories. 14 of which were brains with tumors and the one other with normal brains.



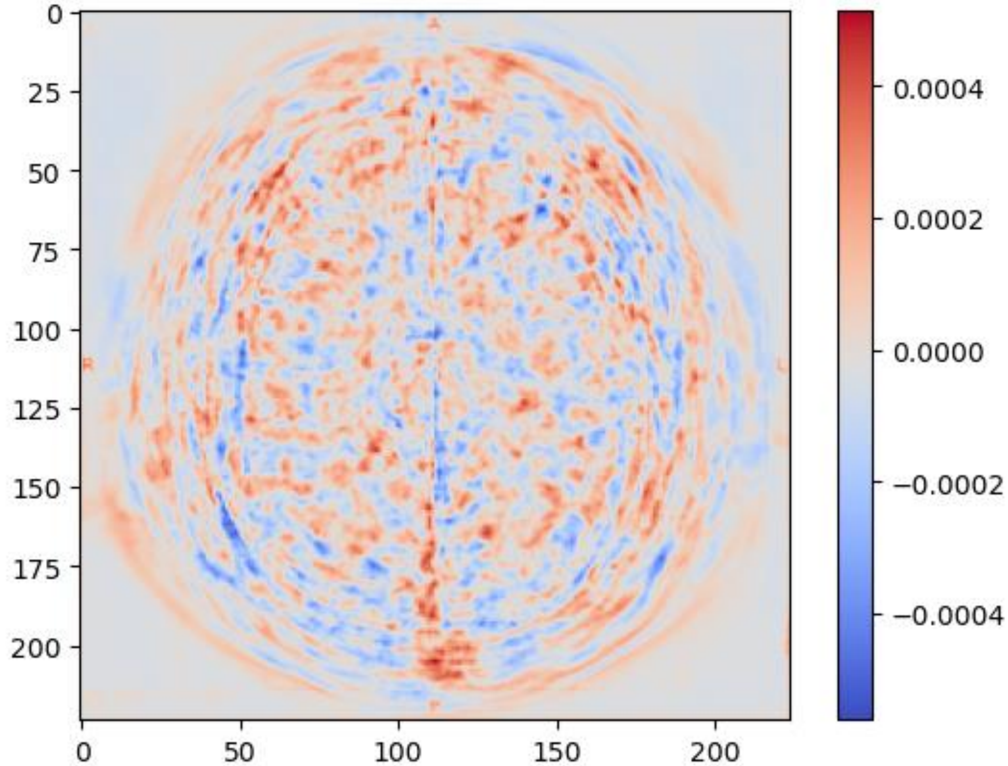
As you can see, the two classes have a noticeable difference in size. The Tumor class is a clear majority.

To ensure that the models have to deal with as little complexity as possible, I resized all the images to 224x224. This resulted in varying brain sizes as you can see in this mean of all the MRIs.



As you can see there is considerable variance in the shape of the brain when the images are resized to perfect squares.

During this project I learned that plotting the **coefficients of a fitted Logistic Regression Model** allow you to see which areas of an image have a high likelihood of tumors present.



The red areas symbolize all the points where there is a high probability of a tumor present.

Next Steps:

The next obvious challenge is to figure out how to properly optimize the models such that they can only focus on the Brain itself and not the black background in the MRI.

One way this might be achieved is that instead of resizing, we crop them. It might be a time consuming task but I think it would be worth it if we can get an ideal model with the least amount of misclassifications.

Another way to streamline my process is to learn how to use functions in .py files. I can refrain from having to repeat code all the time.

Finally, in order to fix the class imbalance, we up sample the normal brains in order to provide a better model for classification.