

Dataset Description

Since both of our datasets are well over 5 MB, below are brief descriptions of each dataset that was utilized in implementing our final project.

Our first dataset consisted of four different classes of MRI scans – gliomas, meningiomas, pituitary tumors, and a control group with no tumors. It was pre-separated into training and testing sets, which we preserved. The training set was mostly evenly distributed with sample sizes of 800, except for the no tumor group which was about only 400 samples large. The testset was also mostly evenly distributed with sample sizes of 100 across each class, with only the pituitary tumor group having about 75 samples.

Our second dataset consisted of classes that were somewhat similar but with greater number and granularity of brain tumor classes. These classes include: gliomas (T1, T1C+, T2), meningiomas (T1, T1C+, T2), neurocitomas (T1, T1C+, T2), schwannomas (T1, T1C+, T2), other (T1, T1C+, T2), and a no tumor control group (T1, T2). The dataset - with 4449 total images - was not separated into train and test sets, but rather by class label. The number of images in each class differed greatly, but we left the sample sizes as is to preserve the rarity of each tumor type across the population.