Brain Tumor Dataset

|  |  |
| --- | --- |
| **Size** | 7MB |
| **Dataset Characteristics:** | Image |
| **Attribute Characteristics:** | N/A |
| **Associated Tasks:** | Classification |
| **Number of Instances:** | 231 |
| **Number of Attributes:** | Variable sized images |
| **Missing Values?** | No |
| **Number of Classes (for classification)** | 2 |
| **Area:** | Other |
| **Additional Details:** | N/A |

# Dataset Information:

In this dataset, we provide MRI images of two types of brain tumors: malignant and benign. in each folder, the images are saved as JPG files. Most of the images are RGB images, a few are grayscale images. You can remove the grayscale images from your project if you want. But if you can keep them and handle the task, then you get additional credit for it.

# Dataset Organization:

data/

* benign/
  + 1.jpg
  + 2.jpg

- …

* malignant/
  + 1.jpg
  + 2.jpg

- …

Class Information:

|  |  |
| --- | --- |
| **Name** | **Number of instances** |
| Malignant | 154 |
| Benign | 77 |
| Total | 231 |

Task Expectation:

1. You should use pretraining schemes for neural network models as the number of samples is very limited.
2. You should look at smart feature extraction techniques from the provided images so that your model can classify properly in spite of having such few number of samples during training.
3. You are allowed to augment the number of original samples during training; but the augmentation needs to be performed only on the training data and not on the test data after train-test splitting.
4. You are allowed and encouraged to use external image samples for brain MRI, but make sure that you are using them just to do unsupervised or semi-supervised pretraining of your models (no malignant or benign labels used)