4510 Project Proposal

*By: The ‘Learning Machine’*

###### Introduction to our problem

Brain tumors are recognized as one of the most vicious diseases that mankind has faced. Primary brain tumors occur in around 250,000 people a year globally. According to the National Cancer Institute of USA, the average 5-year survival rate for domestic malignant brain cancers is 33%. The disease has posed great threats to all age groups, from children to elderly citizens. The treatment of brain tumors differs for different sub-categories, and the best diagnostic technique is through Magnetic Resonance Imaging (MRI).

However, manual classification by reading MRI scans could be error-prone owing to the complex characteristics of variant types of brain tumors. The abnormalities in sizes and lesions of brain tumors complicate the understanding of their nature. Furthermore, in less developed countries, the lack of professional neurosurgeons with knowledge and diagnostic experience adds challenges to interpret MRI reports and identify tumor categories.

The past decade has witnessed a rise in utilizing Machine Learning (ML) methods to classify brain tumors due to its priority in accuracy and time. Hence, in this project, we focus on constructing a reliable (ML) model for the classification of different types of brain tumors.

###### Description of dataset

Our choice of data is the public dataset Brain Tumor Classification (MRI), found in the public domain of Kaggle. Authored by Sartaj et.al, the dataset comprises 394 pieces of testing data and 2870 pieces of training data. Each piece of data is an MRI scan with a scale of 495\*629 pixels. There are four classes of interest: *No Tumor*, *Meningioma Tumor*, *Glioma Tumor*,and *Pituitary Tumor*.

Link: <https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri>

###### Objectives

We aim to build ML classification models that could use MRI images as input and determine whether the patient has a brain tumor and identify the tumor type.

We would compare the efficiency and accuracy among our different models, and suggest a robust model that could potentially be adopted by the medical society.

###### Proposed method

We have decided to use Python to write the code (with Pytorch framework), and collaborating using Github

The models of choice are as follows:

1. KNN
2. Support Vector Machine
3. Random Forest
4. Convolutional Neural Networks (CNN) (ResNet)
5. Vision Transformer (ViT)