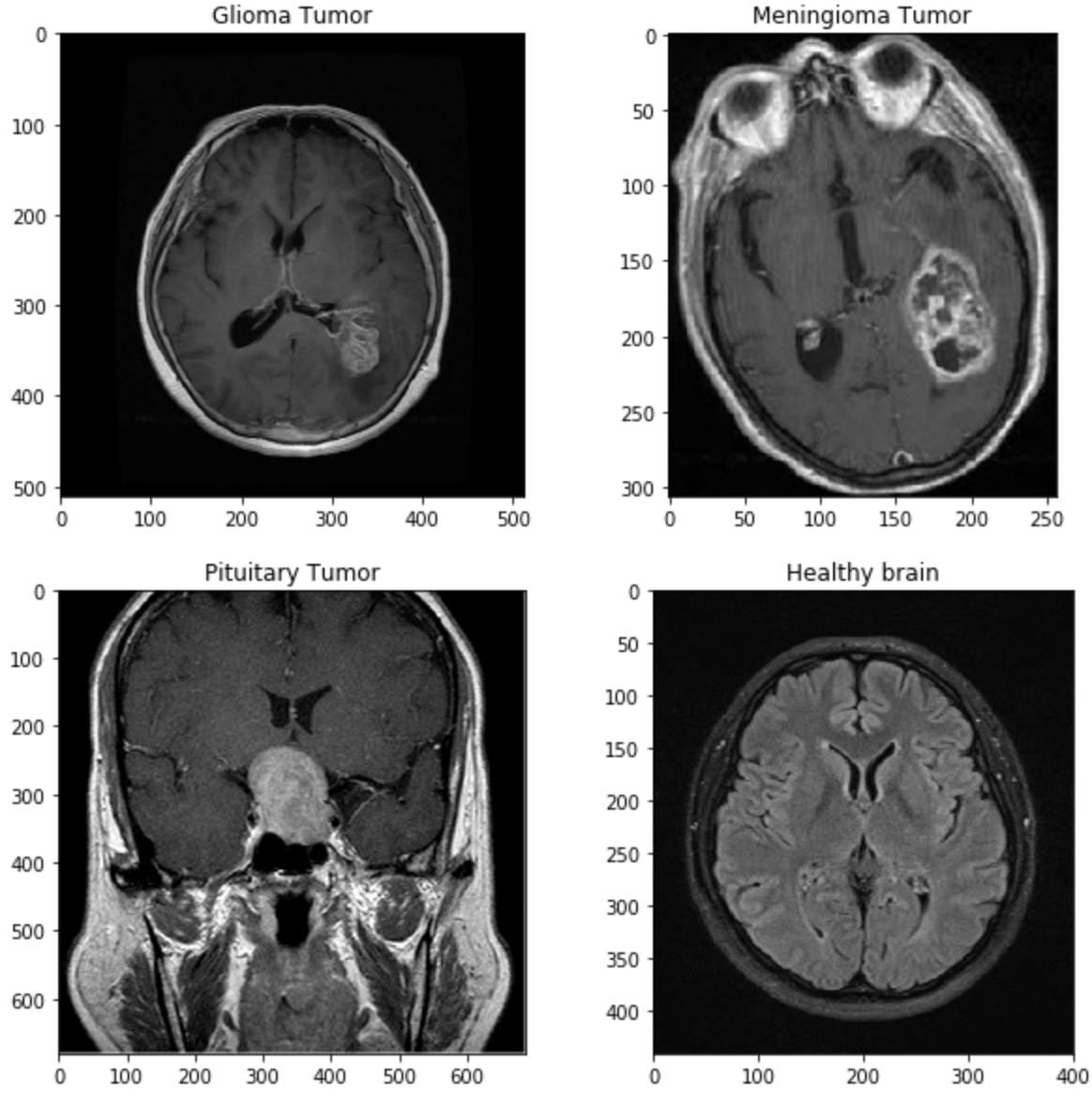

Brain Tumor Detection

CS156

—

Alma Gashi

1. Problem Definition

Brain tumors are a mass of cells that grow abnormally in the brain tissue, and serve no (good) purpose. Benign tumors grow slowly, and usually do not invade the other parts of the brain, unlike malignant tumors, which spread fast, and overgrow cancerous cells and harm normally-helpful cells (Harvard Health Publishing, 2019; Ambardekar, 2019). While these tumors make up for only 1.8% of the overall incidence rate of cancers, they prove to be more mortal than the rest of the cancers (Ferlay et. al, n.d.).

Brain tumors are difficult to detect before they progress. Any misdiagnosis could be detrimental; not diagnosing a real tumor means that the tumor will progress, and misdiagnosing cell growth for a tumor, could mean that the patient will take treatment that would kill their healthy cells.

Magnetic Resonance Imaging shows to have one of the best performances in tumor detection. Using magnetic fields and radio waves, MRI is able to create images of brain areas that cannot be seen with CT-scans or X-rays.

A pitfall of this method is the human error - doctors read the scans. While there are well-established guidelines on reading neuroimages (i.e. obtaining a second opinion), this human error could be mitigated. With the machine learning classification, we could use previously correctly-labelled MRI neuroimages, and train the machine to output the diagnosis in either the form of binary classification: Tumor / No tumor; or multi-class classification to tell the type of the tumor.

To create these models, I use this dataset, which contains pictures of healthy brains and 3 types of tumors:

1. Glioma
2. Meningioma
3. Pituitary

The dataset consisted of class imbalance as seen in Fig 1. To address this issue, I used undersampling

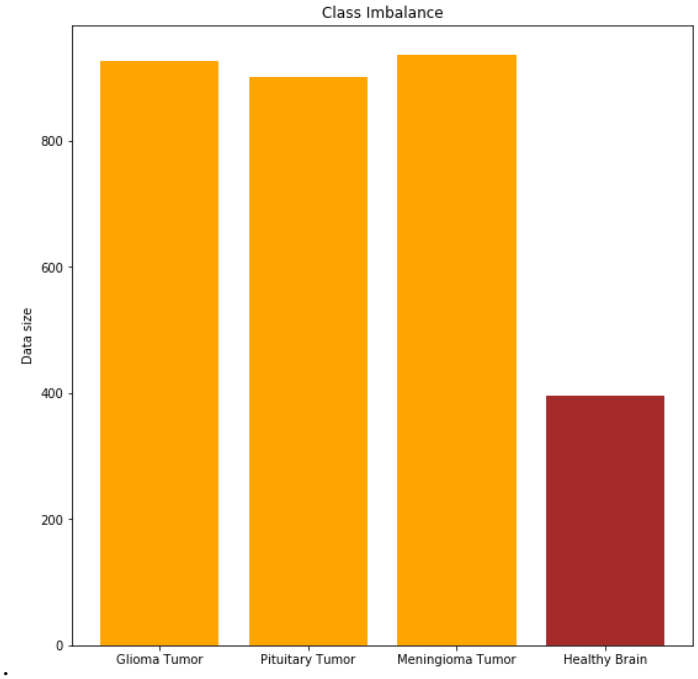

Figure 1. Class Imbalance in the dataset

2. Solution Specification

I. Binary Classification

A. Logistic Regression
   1. Glioma or Healthy
   2. Meningioma or Healthy
   3. Pituitary or Healthy

B. Logistic Regression & Principal Component Analysis
   1. Glioma or Healthy
   2. Meningioma or Healthy
   3. Pituitary or Healthy

C. Support Vector Machines
   1. Glioma or Healthy
   2. Meningioma or Healthy
   3. Pituitary or Healthy

II. Multi-class Classification

A. Neural Networks
   1. Glioma | Meningioma | Pituitary | Healthy

3. Testing and Analysis

I use different metrics to analyze the models; accuracy, precision, f1 score.

For different tumors, we have different accuracy results. Across all binary classification models, Support Vector Machines tend to do best at classifying, achieving highest accuracy score.

Multiclass classification on the other hand only achieves ~ 62% accuracy.

Thus, I would choose Support Vector Machines for these types of images, or add more neurons on the Neural Network and observe the accuracy.

4. References

Ambardekar, N. (2019). Benign Tumors. \textit{WebMD}. Retrieved from https://www.webmd.com/a-to-z-guides/benign-tumors-causes-treatments

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. (2012). Cancer Incidence and Mortality Worldwide.\textit {IARC CancerBase No. 11} [Internet].

Harvard Health Publishing. (2019). Brain Tumor Overview. \textit {Harvard Medical School.}

Retrieved from https://www.health.harvard.edu/a_to_z/brain-tumor-overview-a-to-z\

5. Appendices.

Binary Classification Glioma Tumor

Binary Classification Meningioma Tumor

Binary Classification Pituitary Tumor

Multiclass Classification