Final Report  
Machine Learning: Detecting Brain tumors in Magnetic Resonance Imaging (MRI)

**By:**

**Zongzhu Li**

[Maliha Aziz](mailto:mlaziz@gwu.edu)

[**Project Proposal**](#_rt6uambuek8m) **2**

[**Introduction**](#_xgb53zwh1bmo) **4**

[**The Dataset**](#_vry8urwarh0j) **4**

[**Data preprocessing**](#_kjsbqjtfnvk8) **5**

[**Modeling, Prediction and Evaluation**](#_hsuuy47ldzoo) **6**

[**Deep Learning: Artificial neural networks**](#_ljb0gqu4o090) **9**

[**Discussion**](#_gyo8fyrj5zpe) **12**

[**References**](#_p2nyw9fsv1uo) **14**

[**Supplementary Material**](#_6xta7hs9innr) **16**

# 

# 

# Project Proposal

* What problem did you select, and why did you select it?

To sharpen our machine learning skills and understand real-world data , our group decided to select the topic of detecting brain tumors in Magnetic Resonance Imaging from a Kaggle competition: <https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri>

The objective of this project is trying to classify the tumor types after training all the MRI images.

The reason we select it is because automated classification techniques using Machine Learning(ML) and Artificial Intelligence(AI) has consistently shown higher accuracy than manual classification. Hence, proposing a system performing detection and classification by using Deep Learning Algorithms using Artificial Neural Network (ANN) would be helpful to doctors all around the world.

* What database will you use? Is it large enough to train a machine learning network or different algorithms?

We would use the dataset provided by Kaggle. This dataset is collected in the form of images through the scans which is the best technique to detect brain tumors. The dataset needs to be read out into matrix data. After that, we will have lots of data which is large enough to train a machine learning network.

* What neural network will you use? Will it be a standard form of the network, or will you have to customize it? What algorithms will you use?.

The ANN classification algorithm will be used in this project according to the objects. Some of the parameters, like layer number, learning rate, will be modified to make a comparison.

* What software will you use to implement the neural network or different algorithms?Why?

Python will be adopted here to implement the neural network because it is easy to use and many packages are plug-in-use.

* What reference materials will you use to obtain sufficient background on applying the chosen network or algorithm to the specific problem that you selected?

In this project, EDA, preprocessing, image reading out, model building and evaluation strategy, will be applied, so all the knowledge and reference materials related to the above topic will be referred.

* How will you judge the performance of the network? What metrics will you use?

After building a model with the 80% randomly selected data from the training dataset, we will apply the model to the 20% left dataset to evaluate the model performance and fitness. The evaluation metrics will be the accuracy score.

* Provide a rough schedule for completing the project.

June 15 - 16: proposal writing and data preparation;

June 17 - 20: code for model building and evaluation strategy;

June 21 - 23: presentation preparation and final report.

# **Introduction**

A typical brain tumor diagnosis requires an expert neuro-oncologist, a pathologist and series of scans and surgeries. The process is time consuming and causes distress to the patient while they wait for the results. On the other hand the oncologist has to wade through a series of MRI scans and make expert judgment on whether they identify a tumor or not. The recent groundbreaking research [1] utilizing stimulated Raman histology (SRH) - an optical imaging method in conjunction with artificial intelligence (AI) motivated our team to explore similar ideas with a different dataset i.e. MRI scans.

# **The Dataset**

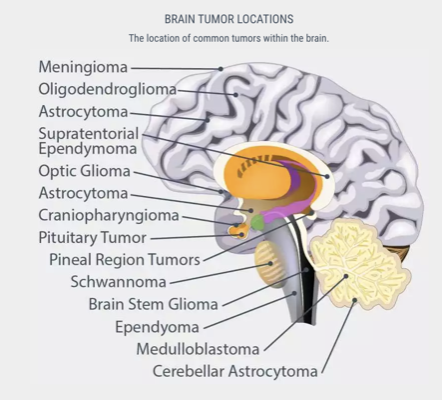


Figure 1. Types of Brain Tumors [9]

Brain tumors are named according to the region of the brain they occur in (Figure 1). We downloaded a training set of MRI scans for brain tumors from Kaggle [11]. The tumors were classified into categories detailed in Table 1.

|  |  |  |
| --- | --- | --- |
| **Tumor type** | **Dataset type** | **N** |
| No Tumor | Training | 395 |
| Glioma | Training | 826 |
| Meningioma | Training | 822 |
| Pituitary | Training | 827 |

Table 1. Classes of Brain Tumor in Kaggle training set

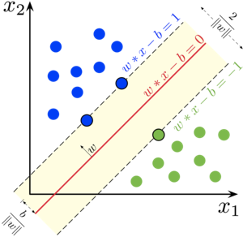
# **Data preprocessing**

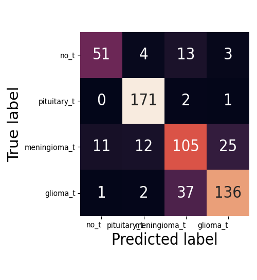
The MRI scans were put through a series of preprocessing steps before being fed into the deep learning layer. The images were of varying sizes therefore they were resized to a standard 128 by 128. Next we performed data augmentation ( rotation\_range=20,width\_shift\_range=0.2,height\_shift\_range=0.2,zoom\_range=0.2,horizontal\_flip=True) to include variations of the image to improve our prediction accuracy. The images were then transformed to an array, normalized by dividing the pixels by 255 and finally reshaped from a 3D to 2D.

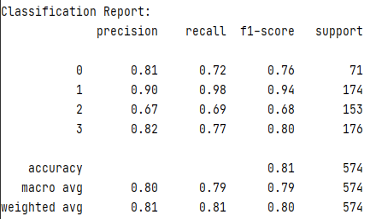
# **Modeling, Prediction and Evaluation**

SVM

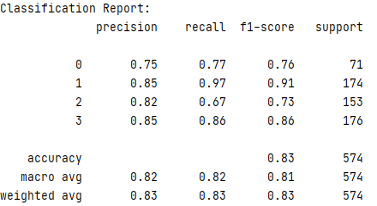
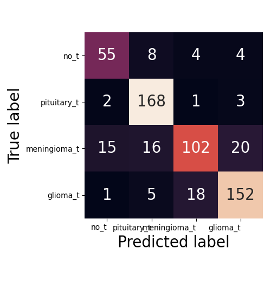
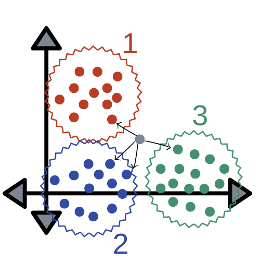
clf = SVC(kernel="poly")



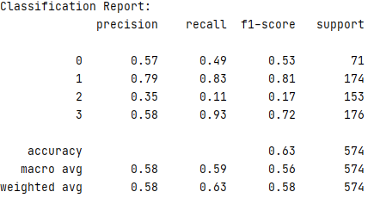
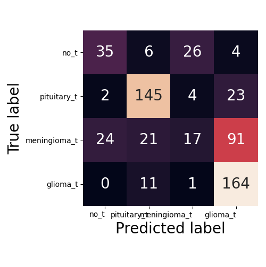
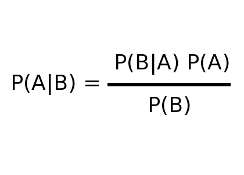




clf = KNeighborsClassifier(n\_neighbors=4)



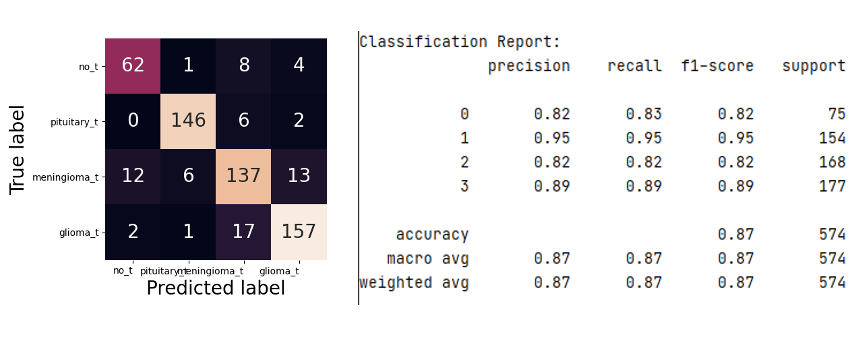
clf = GaussianNB()



# **Deep Learning: Artificial neural networks**

The Winning architecture:

Hidden layers and Nodes: (500,500,20)/ Iteration: 500/ Activation: “sigmoid”/ Solver: RMSpro (lr=0.001)



Other Models that were tried:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANN** | **Layers & Nodes** | **Iteration** | **Activation** | **Early-stopping** | **Solver** | **Accuracy**  **(%)** |
| 01 | (20, 20) | 500 | “sigmoid” | False | Adam | 84.5 |
| 02 | (20, 20) | 500 | “relu” | False | Adam | 79.4 |
| 03 | (20, 20) | 500 | “sigmoid”, ”relu” | False | Adam | 82.1 |
| 04 | (500, 500) | 500 | “sigmoid” | True | Adam | 85.2 |
| 05 | (20, 20, 20,20) | 500 | “sigmoid” | False | Adam | 80.8 |
| 06 | (20, 20) | 2000 | “sigmoid” | False | Adam | 82.2 |
| 07 | (20,20) | 500 | “sigmoid” | False | SGD | 81.2 |
| 08 | (20, 20) | 500 | “sigmoid” | False | RMSprop | 84.8 |
| 09 | (20, 20) | 500 | “sigmoid” | False | Adagrad | 81.0 |
| 10 | (20, 20) | 500 | “sigmoid” | False | Adadelta | 78.6 |
| 11 | (20,20) | 500 | “sigmoid” | False | Adamax | 73.3 |
| ANN | (500, 500, 20) | 500 | “sigmoid” | False | RMSprop | 87.5 |

# **Discussion**

It feels like technology has come a full circle. A method that was adapted from the functioning of the brain is being used to detect its aberrations and help heal it. We applied various Machine learning and artificial neural net methods on the MRI scans of the brain tumor dataset. The successful network gave us a f1-score of 87% where the precision and recall were 87%. We observed that activation with “sigmoid” shows better performance and higher accuracy than activation with “relu” and “tanh”. Furthermore, the RMSprop solver showed higher prediction accuracy with the brain cancer dataset.

Our next step will be to apply more advanced methods such as convolution and maxpool to obtain higher accuracy.

# References

1. Hollon TC, Pandian B, Adapa AR, et al. Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat Med*. 2020;26(1):52-58. doi:10.1038/s41591-019-0715-9
2. Aldape, K., Brindle, K.M., Chesler, L. et al. Challenges to curing primary brain tumours. Nat Rev Clin Oncol 16, 509–520 (2019).<https://doi.org/10.1038/s41571-019-0177-5>
3. Cao, H., Erson-Omay, E.Z., Li, X. et al. A quantitative model based on clinically relevant MRI features differentiates lower grade gliomas and glioblastoma. Eur Radiol 30, 3073–3082 (2020).<https://doi.org/10.1007/s00330-019-06632-8>
4. Nadeem, M. W., Ghamdi, M., Hussain, M., Khan, M. A., Khan, K. M., Almotiri, S. H., & Butt, S. A. (2020). Brain Tumor Analysis Empowered with Deep Learning: A Review, Taxonomy, and Future Challenges. Brain sciences, 10(2), 118.<https://doi.org/10.3390/brainsci10020118>
5. Hollon, T.C., Pandian, B., Adapa, A.R. *et al.* Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat Med* **26,** 52–58 (2020).<https://doi.org/10.1038/s41591-019-0715-9>
6. Kingma, D.P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980.
7. https://ml-cheatsheet.readthedocs.io/en/latest/index.html
8. https://ruder.io/optimizing-gradient-descent/index.html#adamax
9. <https://miamineurosciencecenter.com/en/conditions/brain-tumors/types/>
10. <https://mlfromscratch.com/optimizers-explained/#/>
11. <https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri>

# Supplementary Material

