**Image Classification Brain Tumor**

# **Bhushan Suryawanshi** [**DSC680-T302 Applied Data Science (2215-1)**](https://cyberactive.bellevue.edu/webapps/blackboard/execute/courseMain?course_id=_500232_1) **https://github.com/BhushanGitHub/bhushanGitHub.github.io**

# **Domain and Background:**

Images are used in various fields to make the problem easier to understand. Image processing techniques are most widely used in medical imaging to identify the affected area through an X-ray, computed tomography scan(CT scan), MRI scan(Magnetic resonance images). These images used to detect, identify, and locate the infections, abnormal growths from the human body. Heart diseases, Cancer, Brain tumor, Blood clotting, these are some of the abnormalities that can be found by medical imaging techniques. We can use different machine learning techniques to classify different types of brain tumors by using MRI. The convolutional neural network (CNN) is a class of deep learning neural networks that are highly effective with image classifications.

References -

[1] <https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.median_filter.html>

[2] <https://www.pyimagesearch.com/2016/04/11/finding-extreme-points-in-contours-with-opencv/>

[3] <https://machinelearningmastery.com/batch-normalization-for-training-of-deep-neural-networks/>

[4] [https://www.analyticsvidhya.com/blog/2020/01/fundamentals-deep-learning-activation-functions-when-to-use-them/#:~:text=The%20ReLU%20function%20is%20another,neurons%20at%20the%20same%20time](https://www.analyticsvidhya.com/blog/2020/01/fundamentals-deep-learning-activation-functions-when-to-use-them/" \l ":~:text=The ReLU function is another,neurons at the same time).

[5] Brain Tumor Detection Using K-Means Clustering Algorithm: V.Surudhi, K.Sanjana, R.Saravanan, G.Santhosh, S.Kirubha <https://pdfs.semanticscholar.org/b70e/4a5455a4531ca650272474d49d29e5e1da5c.pdf>

[6] The Complete Beginner’s Guide to Deep Learning: Convolutional Neural Networks and Image Classification: Anne Bonner: <https://towardsdatascience.com/wtf-is-image-classification-8e78a8235acb>

[7] QUICK BRAIN TUMOR FACTS: National Brain Tumor Society: <https://braintumor.org/brain-tumor-information/brain-tumor-facts/>

[8] Pei, L., Vidyaratne, L., Rahman, M.M. *et al*  Context aware deep learning for brain tumor segmentation, subtype classification, and survival prediction using radiology images. *Sci Rep* 10, 19726 (2020). https://doi.org/10.1038/s41598-020-74419-9

[9] Seetha J, Raja S. S. Brain Tumor Classification Using Convolutional Neural Networks. Biomed Pharmacol J 2018;11(3).

[10] An Hoai Truong, Viktoriia Sharmanska, Clara Limbӓck-Stanic, Matthew Grech-Sollars, Optimization of deep learning methods for visualization of tumor heterogeneity and brain tumor grading through digital pathology, Neuro-Oncology Advances, Volume 2, Issue 1, January-December 2020, vdaa110, https://doi.org/10.1093/noajnl/vdaa110

# **Data Understanding:**

Dataset URL: https://figshare.com/articles/brain\_tumor\_dataset/1512427.

The brain tumor dataset contains 3064 T1-weighted contrast-enhanced images from 233 patients. The dataset includes the above discussed three kinds of tumor images. The data set comprises the following.

* 708 Meningioma images,
* 1426 Glioma images,
* 930 pituitary tumor images.

The dataset contains files in Matlab data format(.mat files). Every file contains a structure with the following fields.

1. cjdata.label: 1 for meningioma, 2 for glioma, 3 for pituitary tumor
2. cjdata.PID: patient ID
3. cjdata.image: image data
4. cjdata.tumorBorder: a vector storing the coordinates of discrete points on the tumor border.
5. For example, [x1, y1, x2, y2,...] in which x1, y1 are planar coordinates on tumor border. It was generated by manually delineating the tumor border. We can use it to create a binary image of the tumor mask.
6. cjdata.tumorMask: a binary image with 1s indicating tumor region

# Research Question:

With every year, the number of patients with a brain tumor is increasing. There are two classes of brain tumors, primary and secondary tumors. Primary tumors have several types; one of the frequently found is a meningioma type. It is very challenging to locate, detect, and select the infected tumor portion in the brain from the MRI (Magnetic resonance images). This tedious and time-consuming job requires radiologists and medical field experts. The accuracy of this task is mainly subject to the experience and expertise of the person performing this task. So, if we use a machine learning model to perform this task, it will help to overcome the shortcomings of the person involved in performing this task. So, I think if we can automate this process of classifying the tumors by using machine learning algorithms, it will improve the accuracy of the results and cost due to the expertise required.

# **Method:**

Using Keras sequential model for CNN model will be easier. The CNN model will take the Input as image data in batches with each image shape as 65536. Since I am using grayscale images, the size of each image will be smaller compared to using RGB. A similar picture with RGB will need an array of size 196, 608. The first layer used in the network will be the Batch normalization layer. The batch normalization is a technique for training deep neural networks that normalize the input to a layer for every mini-batch. The batch normalization has the effect of stabilizing the learning process and dramatically reducing the number of training epochs required to train deep networks. The second and fourth layers of the model will be the activation layers with activation type as ReLU. ReLU stands for the Rectified Linear Unit. The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time. That means that the neurons will only be deactivated if the output of the linear transformation is less than 0. The third layer is a densely connected layer with size as 100 with activation as ReLU. For the final layer, we will use the Dense layer with activation type as “softmax” to obtain the probabilities of all brain tumor classes.

# Potential Issues:

There are multiple risks involved in getting the data science project completing successfully. In the larger organization setup you encounter risks such as human resources, technology (hardware/software), poor direction from management, poor requirements, bad project planning etc. However for this course work I already eliminated people risk by working individual. However, following are some of the risks still expected and respective mitigation plan.

1. Low Data Quality

Since I am using data from online data source where I don’t have control on its availability, I have kept a copy downloaded.

2. Technological Risk

Since the work is done in a local environment with windows laptop, some of the libraries may give problem due to the setup environment. I do have setup and verified Anaconda environment with image dataset to work with image classification using mnist dataset.

3. Wrong Model and Libraries

While working I may end up choosing wrong model. I have to wait till the time model building start but will keep researching on using right algorithm for image classification. However CNN is widely used for image processing and expecting to work for this problem as well.

# Concluding Remarks:

Image processing with deep learning has wide variety of usage. Working with brain tumor classification not only help to solve current problem in hand but also has opportunity for future work in other areas where images are used. The dataset has huge training data which will help to train the model for greater performance and accuracy. If we use a machine learning model to perform this task, it will help to overcome the shortcomings of the person involved in performing manual analysis by only expert Neurologists. This way it will be cheaper as well as less time consuming.