

Project 1

First Name: Shuyi Last Name: Fan BU ID: U14145401

Email: shuyifan@bu.edu

Github Username: shuyifan1024

Topic: RSNA-MICCAI Brain Tumor Radiogenomic Classification

Problem Statement

Q:What does the topic cover ?

A:Data Science and Intelligent System. Data science and intelligence systems have been a hot technology in recent years. It is widely used in medical treatment, communication, education, computer and other fields. This topic covers many basic sciences, such as Statistics, Probability, Software, Hardware, etc. It is a comprehensive and large concept, which is hard to be generalized by several words.

Q:Why it is important ?

A: Data analysis is a new technology based on mathematical statistics and big data science. This technology is to extract undiscovered rules or information from existing data, and then use these information to assist the work or the judgment. The biggest role of data analysis is to use existing data to mine new treasures. For example, just like we invented the excavator, and we also know that there is abundant treasure in the ground. Now, what we are going to do is digging the ground with the excavator and finding the treasure.

Applications

Q:What are applications of the topic ?

A:The application of data analysis is very extensive. At present, the mainstream application fields include financial analysis (risk analysis, futures analysis, etc.), unmanned driving field (data analysis of obstacles on the road, or pedestrians, etc.), and medical data analysis that our group is interested in. Medical imaging is very detailed and complex. It is extremely difficult to identify by human alone. Data analysis of medical images, finding cancerous cells and areas of injury, are all hot applications right now.


Q:What is the social significance of the research ?

A: Data analysis techniques are of great importance to society. This is the engine that drives social and technological innovation. Data analysis can upgrade the traditional industry experience analysis to data quantitative analysis. The advantage of this is that

it can refine the vague explanation or judgment and reduce the error or error caused by the empirical judgment. Let's take a concrete example. Now the vast majority of cars are artificially driven, on the highway, there are many accidents caused by drivers fatigue or inexperience. Now unmanned driving, through the analysis of road data, can quickly identify obstacles or pedestrians, vehicles, high precision prediction in advance, and make vehicle adjustment. It is well known that computer calculations are faster and more accurate than judgments made by human minds. Artificial driving can liberate the driver labor force, and it is also of great significance for social development. The driverless cars will be able to operate 24 hours a day, while human workers will only be able to work for eight hours a day. In logistics alone, efficiency will be at least tripled. In this way, data analysis can greatly improve the operation efficiency of the logistics industry. Introducing data analysis to other industries will also improve accuracy and efficiency.

Pick an area of focus that interests you in the topic

Our group is interested in the diagnosis of brain tumors in the field of medical imaging. Sum up our subject in one sentence. Data analysis techniques are used to analyze complex medical images of the brain to determine and predict potential cancer sites. The greatest advantage of this method is that it can reduce the pain of surgical examination for patients.


RSNA Radiological Society of North America · 1,128 teams · a month to go (20 days to go until merger deadline)

RSNA-MICCAI Brain Tumor Radiogenomic Classification

Predict the status of a genetic biomarker important for brain cancer treatment

\$30,000

Prize Money

[Overview](#)
[Data](#)
[Code](#)
[Discussion](#)
[Leaderboard](#)
[Rules](#)

Join Competition

Overview

Description

Evaluation

Timeline

Prizes

Code Requirements

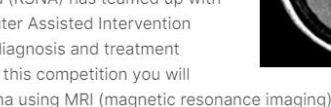
Acknowledgments

A malignant tumor in the brain is a life-threatening condition. Known as glioblastoma, it's both the most common form of brain cancer in adults and the one with the worst prognosis, with median survival being less than a year. The presence of a specific genetic sequence in the tumor known as MGMT promoter methylation has been shown to be a favorable prognostic factor and a strong predictor of responsiveness to chemotherapy.

Currently, genetic analysis of cancer requires surgery to extract a tissue sample. Then it can take several weeks to determine the genetic characterization of the tumor. Depending upon the results and type of initial therapy chosen, a subsequent surgery may be necessary. If an accurate method to predict the genetics of the cancer through imaging (i.e., radiogenomics) alone could be developed, this would potentially minimize the number of surgeries and refine the type of therapy required.

The Radiological Society of North America (RSNA) has teamed up with the Medical Image Computing and Computer Assisted Intervention Society (the MICCAI Society) to improve diagnosis and treatment planning for patients with glioblastoma. In this competition you will predict the genetic subtype of glioblastoma using MRI (magnetic resonance imaging) scans to train and test your model to detect for the presence of MGMT promoter methylation.

If successful, you'll help brain cancer patients receive less invasive diagnoses and treatments. The introduction of new and customized treatment strategies before surgery has the potential to improve the management, survival, and prospects of patients with brain cancer.



Literature Review

Because of the particularity of selecting the project-selection, our team's project-selection was a world-class medical problem. Previously, it was only possible to diagnose this particular type of brain cancer through extraction and analysis of traditional surgical brain slices. Data analysis technology has only been introduced in the last year or two, so there are not many solutions to this topic. We will only give an overview of traditional surgical methods and try to describe how the most advanced data analysis techniques are applied to this field.

Problem Introduction

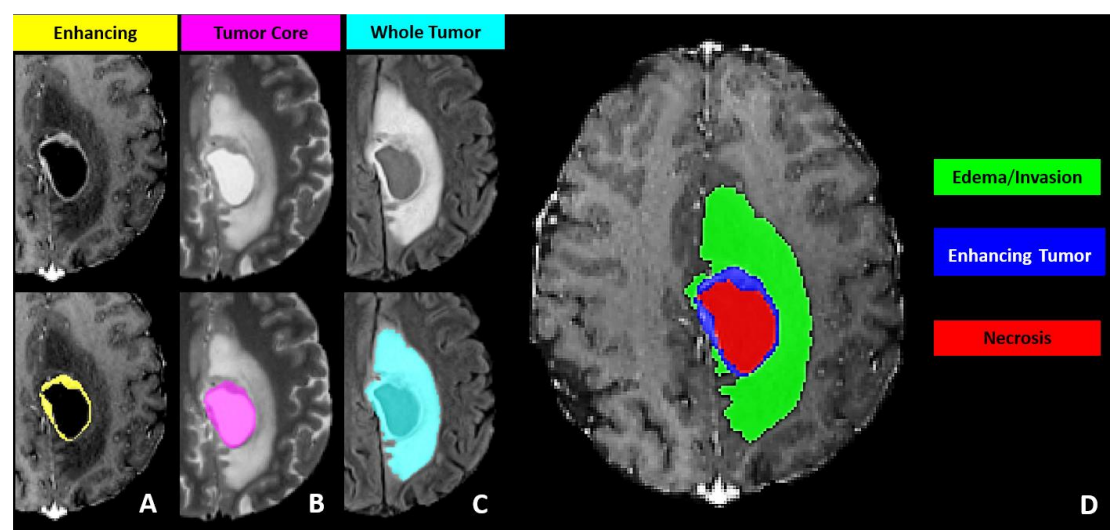
manual detection and tracing of tumor sub-regions is tedious, time-consuming, and subjective. In a clinical setup, this manual process is carried out by radiologists in a qualitative visual manner, and hence becomes impractical when dealing with numerous patients. This highlights the unmet need for automated deterministic segmentation solutions that could contribute in expediting this process.

Traditional Method:

The method used by the medical community is craniotomy, then the brain tissue cells are extracted and analyzed. This process can take several weeks before a diagnosis can be made using cell analysis.

Advanced Method:

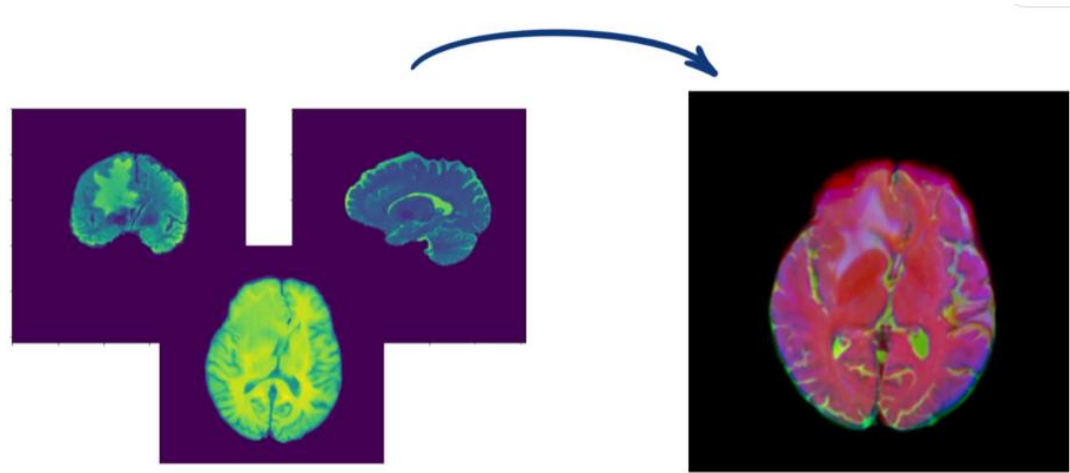
The most advanced treatment is to take images of existing brain tumors and have multiple doctors annotate the data. The existing data are then learned using machine learning techniques. Finally, our algorithm can diagnose the unlabeled brain image.



CITE FROM arXiv:2107.02314v2

Fig. 1: Glioma sub-regions considered in the RSNA-ASNR-MICCAI BraTS 2021 challenge. Image panels with the tumor sub-regions annotated in the different mpMRI scans. The image panels A-C denote the regions considered for the performance evaluation of the participating algorithms and specifically highlight (from left to right): the enhancing tumor (ET - yellow) visible in a T1Gd scan, surrounding the cystic/necrotic components of the core (panel A), the tumor core (TC - magenta) and the whole tumor (WT - cyan) visible in the corresponding T2 (panel B) and T2-FLAIR (panel C) scans, respectively. Panel D, depicts the combined segmentations generating the final tumor sub-region labels, as provided to the BraTS 2021 participants: enhancing core (yellow), necrotic/cystic core (red), and edema/invasion (green).

Open Source Research



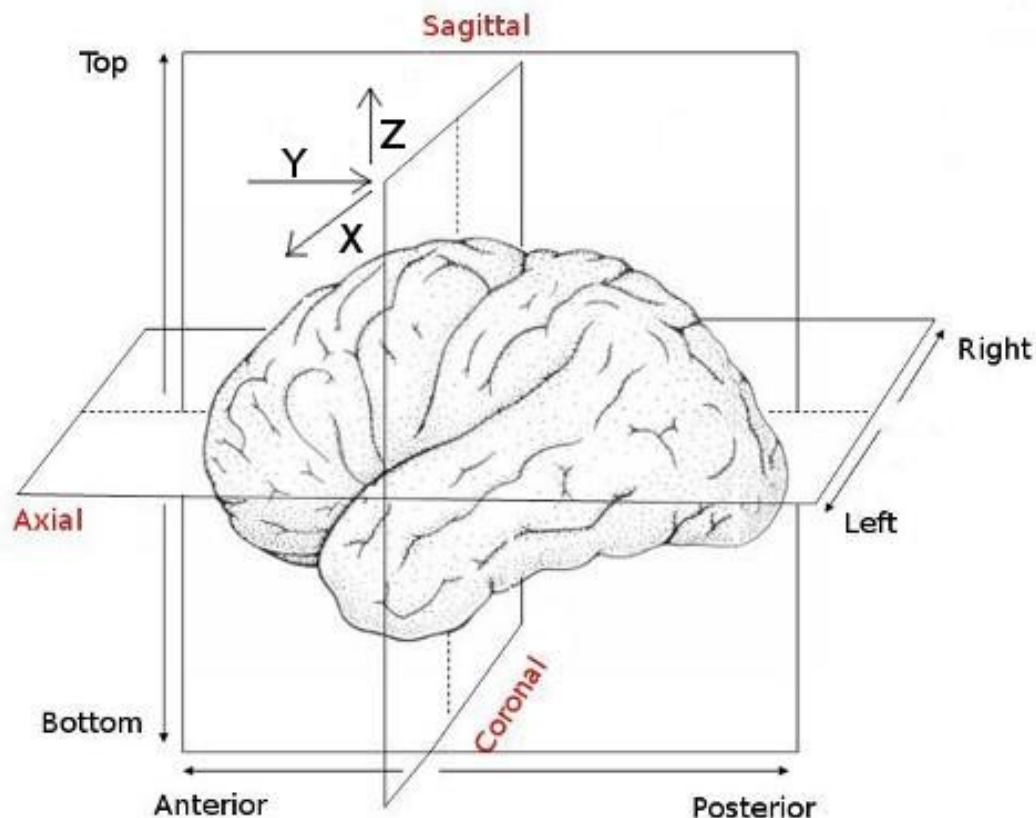
Data:

	BraTS21ID	MGMT_value	imfolder	path	FLAIR_count	T1w_count	T1wCE_count	T2w_count	allsame
0	0	1	00000	../input/rsna-miccai-brain-tumor-radiogenomic-...	400	33	129	408	False
1	2	1	00002	../input/rsna-miccai-brain-tumor-radiogenomic-...	129	31	129	384	False
2	3	0	00003	../input/rsna-miccai-brain-tumor-radiogenomic-...	129	33	129	408	False
3	5	1	00005	../input/rsna-miccai-brain-tumor-radiogenomic-...	400	28	129	424	False
4	6	1	00006	../input/rsna-miccai-brain-tumor-radiogenomic-...	129	32	129	408	False
...
580	1005	1	01005	../input/rsna-miccai-brain-tumor-radiogenomic-...	23	23	23	23	True
581	1007	1	01007	../input/rsna-miccai-brain-tumor-radiogenomic-...	124	124	124	124	True
582	1008	1	01008	../input/rsna-miccai-brain-tumor-radiogenomic-...	60	192	192	64	False
583	1009	0	01009	../input/rsna-miccai-brain-tumor-radiogenomic-...	23	23	23	23	True

BraTS21ID contains the patient's Id and MGMT_value contains information on whether the MGMT promoter has methylation.

Each case consists of four structural multi-parametric MRI (mpMRI) scans. These scans are -

- FLAIR = Coronal
- T1w = Axial
- T1wCE = Coronal
- T2w = Sagittal



Training Data:

The one with MGMT_value = 1

Visualize the person with gene methylation = people with favorable prognostic factors and strong predictors of responsiveness to chemotherapy

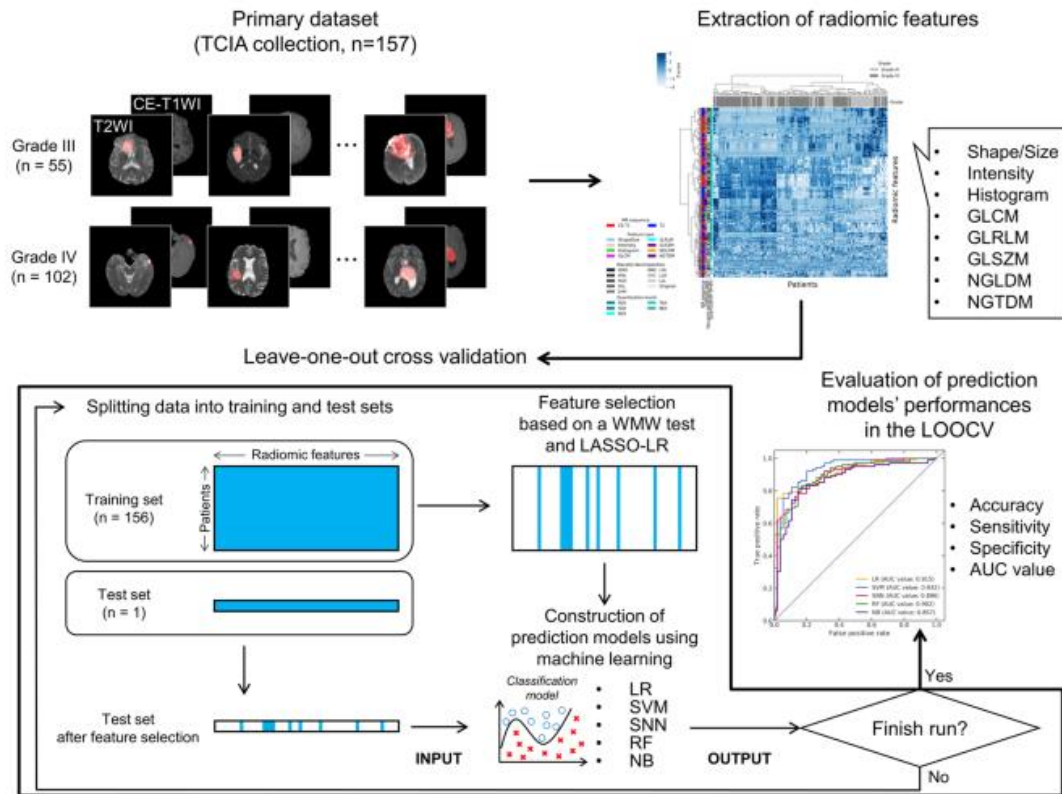
The one with MGMT_value = 0

Visualize the person who does not have gene methylation = those who do not have favorable prognostic and strong predictors of responsiveness to chemotherapy

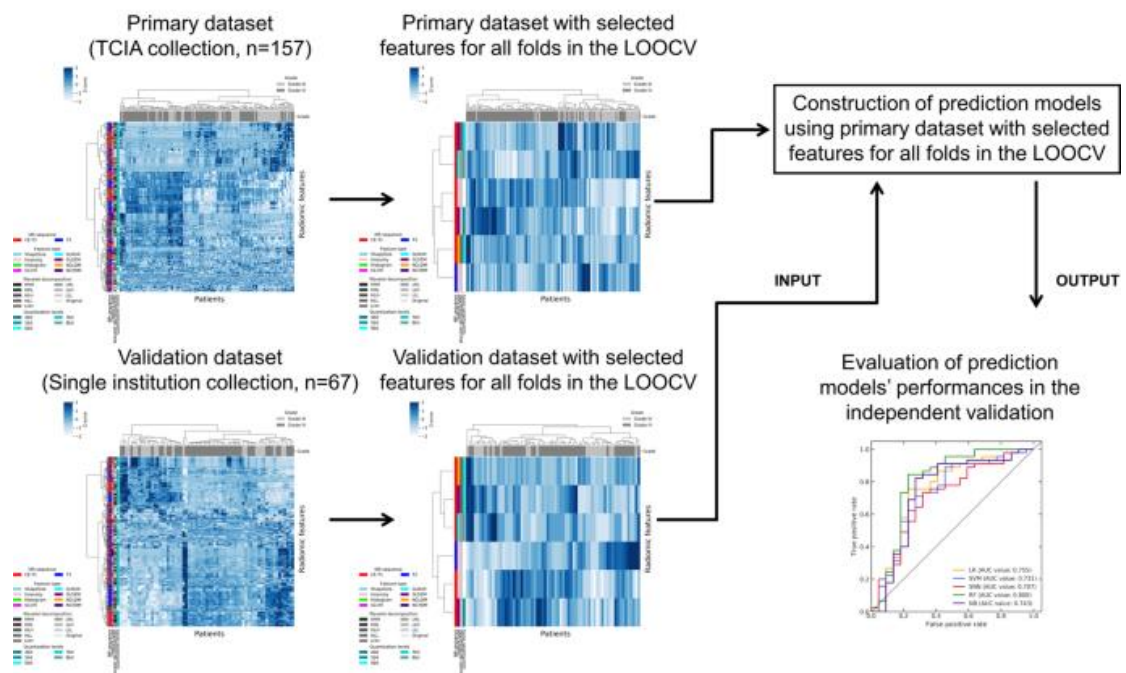
Training Model

The training model is very complex. After consulting a large number of papers, I can only understand the algorithm process in a short time.

Leave-one-out cross validation in the primary dataset



Independent validation



CITE FROM Sci Rep 9, 19411 (2019)

In this algorithm, the first is to enter the image, and then modify the image size. After the size is modified, the image data is classified into training set and test set. WMW and Lasso-LR algorithms are used to extract features from the images, and then the images are classified. Finally, the results are output.

Conclusion

That was my work in project1. In this project, the difficulty of this topic exceeds my knowledge range, and my understanding of algorithm is not in place. The open source algorithm in Kaggle does not introduce the code in detail, so there is no detailed analysis of the complex algorithm model in the limited time. In this project, what I do is to understand this topic and learn how to apply technology to practical problems according to the lecture content.