

Project 1

Zhaowen Zhou BUID: U20617932

Problem Statement

According to the statistics given by International Agency for Research on Cancer, WHO, in 2020, the new cases of brain and nervous system cancer has ranked among top 20. The number of deaths caused by the brain and central nervous system ranked the 12th. According to other statistics given by IARC, the incidence and mortality caused by brain cancer will increase continuously from 2020 to 2040 for both males and females. The current method for analyzing cancer requires multiple weeks to do surgery and determine the characteristics of the tumor. This may extend the treatment time. Through predicting the genetic subtype of glioblastoma, which is a kind of aggressive cancer begins with brain, using MRI scans of brain, the model can detect the presence of MGMT promoter methylation, which is a mechanism of MGMT gene silencing and can predict good outcome for glioblastoma patients ^[1]. This program includes extracting the MRI picture as data, learning the training set and setting up a model to detect for the presence of MGMT promoter methylation and using the test set to test the model.

Applications

MGMT promoter is the only confirmed molecular predictive factor in glioblastoma ^[2], thus, with the success of this topic, the hurting surgeries and treatments can be reduced, the treatment period can be shortened. Due to the severe lethality of glioblastoma ^[3] and the estimated increasing number of patients, this program can reduce the number of deaths caused by glioblastoma and increase the life span of elder people.

Area of Focus

In this project, the most attractive part for me is how to choose features from data describing the glioblastoma to predict the genetic subtype and how to find out the relationship between the genetic subtype and the presence of MGMT promoter methylation.

Literature Review

The CNN architecture can capture different features with 2-dimensional special image and RNN can capture the temporal context. ^[4] Due to the 3-dimensional space of the MRI scan picture, a method combines the architecture of CNN and RNN can reveal new features which is related to the MGMT methylation.

Methylation data from different GBM patients are provided, the methylation data are from the regions which have the minimal promoter and enhancer regions show to have

maximal methylation activity. 3 sites are extracted. A threshold of 0.2 is set up to decide whether the methylation site is positive or not. Any of these three sites are positive is the criterion for deciding whether the patient is methylation positive.

For the data of MRI scan of the brain with the glioblastoma, each scan is a 3-dimensional constructure of the brain which contains a set of image frames according to the slice thickness. 5235 MRI scans contain 458951 image frames. The researchers chose MRI scans with 3 label “T1/T2/Flair axial”. For these image frames, the researchers normalize all the features (slice thickness, sex, age, weight, etc.) and remove the noise. Then the researchers limit the MRI scan to the slices that contain the tumor to the nearest 10th size. All MRI scan are rotated every 4 degrees from -90 to 90 so that the dataset is augmented to combat overfitting.

First, each image frame of the MRI scan is put in to CNN to extract essential features. Then the image frame goes through 2 fully connected neural networks to create a length 512 vector. These vectors represent each frame go through the RNN to analyze the sequence of MRI image frames and output a binary classification of methylation status per scan.

The gene expression data of glioblastoma, MRI data, genomic data are included, MGMT methylation status was determined by level 3 values.^[5] After extracting the quantitative and qualitative features describe the size, location of the tumor and the volume of three abnormal regions segmented. The data is reduced using variable selection or reduction. The classification model is validated through 10-fold cross validation. The predictions are made through four algorithms: random forest, KNN, Gaussian naïve Bayes and the J48 tree.

In conclusive, the whole progress contains data collection, tumor segmentation, feature extraction, selection, reduction, and model training.

Open Resource Research

The open source I found came from the Kaggle code submitted by Yaroslav Isaienkov, he pydicom library, which is a library used for analyzing DICOM files. MRI are DICOM files. Another library is cv2, which is also called as open cv. This library provides a real-time optimized computer vision library. There are also some modules for operating files and data such as glob and json.

The pydicom library is used to analyze the data of MRI scan.

This project uses the ROC curve, this can be imported from sklearn.

For learning model, PyTorch is used.

In the several other open resources, these libraries, pydicom, cv2, nn from torch are also used.

In conclusion, this project can be used by doctors to detect the presence of MGMT so that less surgeries are needed. This can lead to a shorter treatment time and a higher survival rate.

References

- [1] Weller M, Stupp R, Reifenberger G, et al. MGMT promoter methylation in malignant gliomas: ready for personalized medicine[J]. Nature Reviews Neurology, 2010, 6(1): 39-51.
- [2] Bleeker F E, Molenaar R J, Leenstra S. Recent advances in the molecular understanding of glioblastoma[J]. Journal of neuro-oncology, 2012, 108(1): 11-27.
- [3] Gallego O. Nonsurgical treatment of recurrent glioblastoma[J]. Current oncology, 2015, 22(4): 273-281.
- [4] Han L, Kamdar M R. MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks[C]//PACIFIC SYMPOSIUM ON BIOCOMPUTING 2018: Proceedings of the Pacific Symposium. 2018: 331-342.
- [5] Kanas V G, Zacharaki E I, Thomas G A, et al. Learning MRI-based classification models for MGMT methylation status prediction in glioblastoma[J]. Computer methods and programs in biomedicine, 2017, 140: 249-257.