FINAL PROJECT – HW3 Extension

Team: 07 Brainy Brain

Member: 土木所碩一 r07521603 蔡松霖

Introduction

Gliomas are the most frequent primary brain tumors in adults. Since Gliomas is unfortunately an extreme fatal condition, providing an accurate survival estimate is really important for the patient. Brain Tumor Segmentation Challenge (BraTS) 2018 utilizes multi-institutional pre-operative MRI scans and focuses on the segmentation of intrinsically heterogeneous brain tumors along with the prediction of patient overall survival, via integrative analyses of radiomic features and machine learning algorithms. For this project, we will use the data provided from BraTS 2018, which includes multimodality MRI scans, annotations of tumors, and the age of the patients (Bakas et al., 2017; Menze et al., 2015). Our final goal is to advance the accuracy of the survival prediction.

Literature Review

The workflow of predicting survival of brain tumor patients can be divided into data preprocessing, tumor segmentation, feature extraction, feature selection and analysis for prediction. Deep learning methods and radiomics features play an important role in the whole run. Following are some related work done by others.

(Lundervold & Lundervold, 2019) gave an overview of deep learning in medical imaging focus on MRI. Deep learning has been applied to the entire MRI processing chain, from acquisition to image retrieval, from segmentation to disease prediction.

(Sharmila Agnal A, 2019) used VGG16 model to extract features from four modalities and got 25,088 features from MRI images (7*7*512). Further cleaning and feature selection enable a reduction to 4427 features. Then, the patient age feature is appended to the existing features to form a dataset with 4428 features. The proposed system uses an artificial neural network to classify the overall survival of the subjects as high risk (<10 months), moderate risk (>10 months and <15 months) or low risk (>15 months). However, they didn’t mention how they do for cleaning and feature selection.

(Shahzadi, Tang, Meriadeau, & Quyyum, 2018) proposed a cascaded CNN- LSTM model for volumetric classification of a brain tumor into High Grade and Low Grade glioma based on feature extracted from FLAIR sequence MRI.

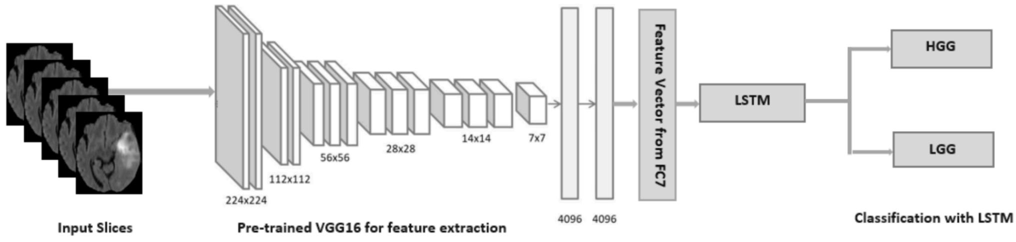


Fig. 4: Cascaded CNN-LSTM network architecture for brain tumour classification into Low Grade (LG) and High Grade (HG) glioma. Pre-trained VGG-16 is used for feature extraction by using transfer learning technique for training. Features from FC7 are then used as input to two layer LSTM for classification of volumetric 3D MRI brain tumour data.

(Kniep et al., 2019) investigated the feasibility of tumor type prediction with MRI radiomic image features of different brain metastases in a multiclass machine learning approach. Extracted features comprised 18 first-order features, 17 shape features, and 56 texture features. First-order and texture features were also calculated on the basis of eight wavelet decompositions (four decompositions for two-dimensional features), resulting in a total of 1423 image features. Algorithm-based feature selection was performed in their research.

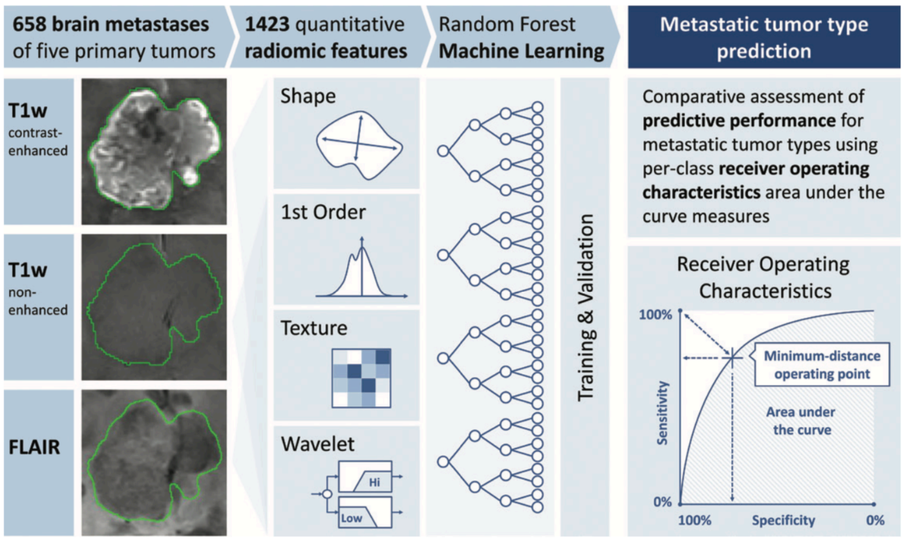


Figure 1: Conceptual overview of proposed metastatic tumor type prediction. *FLAIR* = fluid-attenuated inversion recovery, *T1w* = T1 weighted.

(Lao et al., 2017) proposed a deep feature-based radiomics model for prediction of OS in GBM patients. Both handcrafted features and deep features were extracted from multi-modality MR images. Deep features were extracted from the pre-trained CNN via transfer learning. After a four-step feature selection method, six most robust, nonredundant and predictive features were selected. Finally, a radiomics signature as well as a radiomics nomogram were constructed on a discovery cohort and validated on an independent validation cohort.

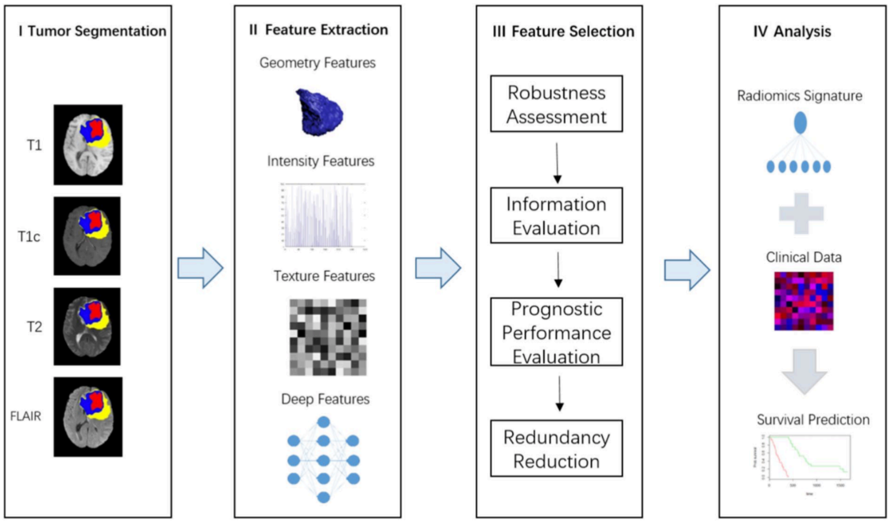
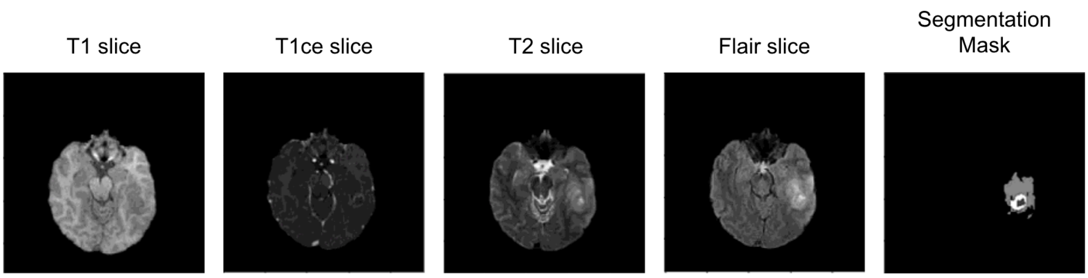


Figure 1. The workflow of radiomics analysis in this study.

In this project, we will focus on fine tuning the workflow of feature extraction, feature selection and deep learning model based on all the experiences from others' work. Combine both radiomics features and deep features and train a well performance predictive model.

Data

In this project, we use the data provided from BraTS 2018, which includes multimodality MRI scans, annotations of tumors, and the age of the patients. Four modalities including T1, T1ce, T2 and Flair are provided, and for the annotations of tumors 3 labels are provided (GD-enhancing tumor (ET — label 4), Peritumoral edema (ED — label 2), Necrotic and non-enhancing tumor core (NCR/NET — label 1). As for the size of the data, there are 39 training data (29 for training, and 10 for validation) and 20 testing data.



Task

The task of this project is to predict patient overall survival using Brain MRI Images It is a regression problem and the prediction will be evaluated by mean-square error (MSE).

Methodolgy

Referring to others work on similar cases, in this project we use a well-known convolutional neural networks (CNN) called VGG16 to extract features from MRI images, and pyradiomics package to extract radiomics features from tumor segmentation. Since there are too many radiomics features, feature selection method based on random forest regressor (use random search CV to find best parameters) and SelectFromModel (select features that are higher than the threshold) is applied to select radiomics features with higher importance in order to get better results. The details of the features we use in this project are shown in the figure below.

- **Age**
- **MRI: 4 modalities** (T1, T1ce, T2, Flair)
  - Image (155 x 240 x 240)
  - VGG extraction (155 x 7 x 7 x 512)
- **Radiomics features** (4 modalities + segmentation label)
  - 1130 x 4 modalities x 3 labels (1, 2, 4) = 13,560

Filter
- Original
- Wavelet (8)
- Laplacian of Gaussian (sigma = 1,2,3)

Feature Class
- First order
- GLCM
- GLDM
- GLRLM
- GLSZM
- NGTDM
- shape

In this project, we try three ways to train model for prediction:

1. Use only with radiomics features to train deep neural networks (DNN) .
2. Use MRI image features and age information to train DNN.
3. Combine radiomics features, MRI image features and age all together to train DNN.

Results

1. Use only with radiomics features to train deep neural networks (DNN)

In this part, we try different settings of the threshold of SelectFromModel, and see improvement when fewer features are selected.

The 12 features selected by threshold 5e-3:

```
t1 2 original_firstorder_Minimum
t1 2 log-sigma-2-mm-3D_ngtdm_Complexity
t1 4 log-sigma-2-mm-3D_gldm_DependenceEntropy
t1ce 1 wavelet-LHL_glcm_Imc1
t1ce 1 wavelet-HLL_firstorder_Variance
t1ce 1 wavelet-HHL_firstorder_RobustMeanAbsoluteDeviation
t1ce 4 wavelet-HHL_firstorder_Skewness
t2 1 log-sigma-2-mm-3D_glcm_ClusterShade
t2 4 wavelet-HLL_glcm_ClusterShade
t2 4 log-sigma-2-mm-3D_glcm_Idn
flair 4 wavelet-HHH_glcm_SumSquares
flair 4 log-sigma-3-mm-3D_ngtdm_Busyness
```

Mean-square error of validation data:

- All 13560 features (before selection): 10256321061951898
- 1565 features (threshold: 1e-4): 3755067596800
- 207 features (threshold: 1e-3): 9266382438.4000
- 12 features (threshold: 5e-3): 384280.2438

However, the testing score using 12 features on Kaggle is 258098.10637, 77159.56196 on private and public leaderboard, which is somehow a poor performance. So, I also try random forest regressor to replace the original DNN model to fit the training data, and get a better result:

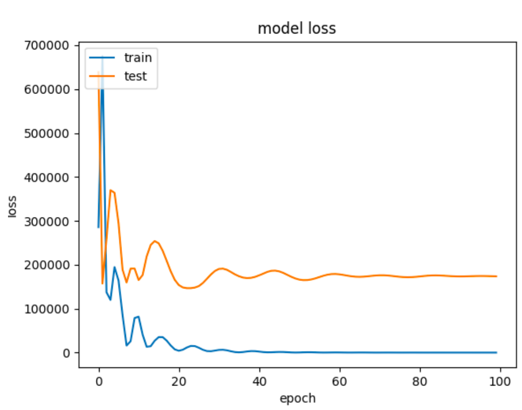
- All 13,560 features: 60575.03819, 93456.37464
- 207 selected features: 66556.21716, 102876.76061
- 12 selected features: 32025.95643, 100518.22383

2. Use MRI image features and age information to train DNN.

In this part, we try VGG16 to extract features from different modalities.
First, we use VGG16 to extract features from flair only and take all features concatenated with age information into a NN model with only one hidden layer with 16 units.

NN settings:

```
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         (None, 3888641)           0
_________________________________________________________________
dense_1 (Dense)              (None, 16)                62218272
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 17
=================================================================
Total params: 62,218,289
Trainable params: 62,218,289
Non-trainable params: 0
_________________________________________________________________
```

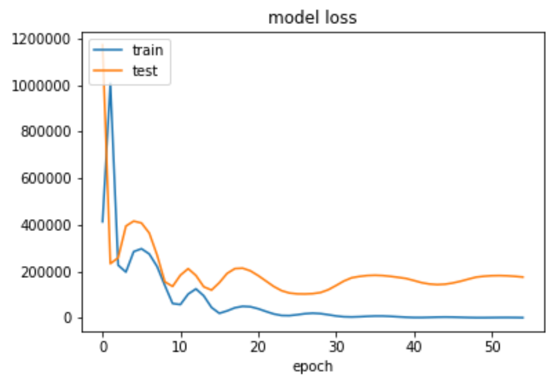


The result on Kaggle for this model is: 33651.91340, 28428.36611.

Second, we use the same way to extract features from t1, t1ce and flair and only take the features from 40 slices in the middle due to memory issue on Google Colabatory.

NN settings:

```
Layer (type)                 Output Shape              Param #
=================================================================
input_5 (InputLayer)         (None, 1003521)           0
_________________________________________________________________
dense_7 (Dense)              (None, 16)                16056352
_________________________________________________________________
dense_8 (Dense)              (None, 1)                 17
=================================================================
Total params: 16,056,369
Trainable params: 16,056,369
Non-trainable params: 0
_________________________________________________________________
```



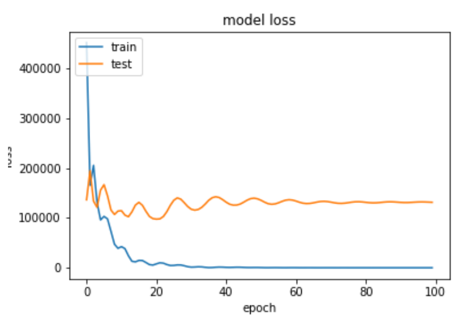
The result on Kaggle for this model is: 32290.46856, 48287.07711. The private score is lower than previous model while the public score is higher which is somehow a good try to train a model with fewer parameters and get similar result. By taking more time to try some settings on the NN model, I think there is possibility to get better score.

3. Combine radiomics features, MRI image features and age all together to train DNN.

Due to the time issue, I simply give a quick try to combine all features to train a model. MRI image features are extracted using VGG16 and only take the middle 40 slices as mentioned above. Along with the image features, 12 selected radiomics features and the age information are used.

- T1, T1ce, Flair -> VGG
  - ✧ 40 * 25,088 = 1,003,520
- Radiomics
  - ✧ Selected 12 features
- Age
  - ✧ 1 feature

```
Layer (type)                 Output Shape              Param #
=================================================================
input_2 (InputLayer)         (None, 1003533)           0
_________________________________________________________________
dense_1 (Dense)              (None, 16)                16056544
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 17
=================================================================
Total params: 16,056,561
Trainable params: 16,056,561
Non-trainable params: 0
_________________________________________________________________
```



The result on Kaggle for this model is: 39360.35041, 60306.04520. Though not yet an improvement, maybe with a deeper or wider NN model can improve the performance.

Conclusion

In this project, we use pyradiomics to extract radiomics features and further select 12 features using SelectFromModel. As for image features, we use VGG16 to extract features. With these features mentioned above along with the age information, several models are trained to predict the patient survival. Currently, the model using features extracted from only flair images and age has the best performance on Kaggle leaderboard. NN models is considered too small and shallow by TA and leave lots of rooms for improvement.

Future Work

- Train models with more data
- Train deeper and larger NN models
- Try on LSTM models
- Try on classification task
- Apply cross validation techniques
- Ensemble models for better performance
- Try on Multimodal Brain Tumor Segmentation Challenge 2019

Reference

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., . . . Davatzikos, C. (2017). Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data, 4*, 170117. doi:10.1038/sdata.2017.117

Kniep, H. C., Madesta, F., Schneider, T., Hanning, U., Schonfeld, M. H., Schon, G., . . . Gellissen, S. (2019). Radiomics of Brain MRI: Utility in Prediction of Metastatic Tumor Type. *Radiology, 290*(2), 479-487. doi:10.1148/radiol.2018180946

Lao, J., Chen, Y., Li, Z. C., Li, Q., Zhang, J., Liu, J., & Zhai, G. (2017). A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. *Sci Rep, 7*(1), 10353. doi:10.1038/s41598-017-10649-8

Lundervold, A. S., & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Z Med Phys, 29*(2), 102-127. doi:10.1016/j.zemedi.2018.11.002

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., . . . Van Leemput, K. (2015). The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans Med Imaging, 34*(10), 1993-2024. doi:10.1109/TMI.2014.2377694

Shahzadi, I., Tang, T. B., Meriadeau, F., & Quyyum, A. (2018). *CNN-LSTM: Cascaded Framework For Brain Tumour Classification*. Paper presented at the 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES).

Sharmila Agnal A, A. D. C., Venkatesh J, Sudarshan S, Pranav A (2019). Predicting Survival of Brain Tumor Patients using Deep Learning. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8*(6).