Udacity Machine Learning Engineer Nanodegree
Capstone Proposal
Kareem Wahid
May 29th, 2017

# Predicting Glioblastoma Survival Outcomes from Radiomic Features

## Abstract

Supervised machine learning algorithms will be trained on radiomic features extracted from glioblastoma magnetic resonance images obtained from the 2017 BraTS Challenge to predict binary patient overall survival outcomes. Classifier performance will be evaluated using the area under the receiver operator curve metric and compared to previously published data.

## Domain Background

Radiomics is a budding new field of medical informatics that seeks to extract mathematically defined quantitative features such as statistics, shape, and texture from medical images [1]. Medical imaging up until recently could only provide visual qualitative information to a clinician. Radiomics allows hidden quantitative information in medical images to be deciphered and analyzed, with the potential to aid in prognosis of disease.

Cancer imaging in particular is a heavily researched area of radiomics. Tumors are often spatially and temporally heterogeneous. This frequently requires multiple tissue biopsies to be performed in order to capture the molecular heterogeneity of the tumor, which can be dangerous for the patient. Radiomics provides a non-invasive window into probing the heterogeneity of a tumor [2]. Gliomas are the most common variety of primary brain malignancies and have a high degree of intrinsic heterogeneity. This heterogeneity is apparent in their appearance and shape upon imaging, making prognosis difficult [3]. Radiomic analysis of glioma medical imaging can provide additional information about a patient's prognosis and likely survival outcomes [4-6].

Though significant research has been conducted on the application of machine learning algorithms to radiomic features for prognostic prediction [7-8], there is still much that is unknown about which models are best due to lack of standardization in the field.

## Problem Statement

The goal of this project is to determine which supervised machine learning classification models are the most suitable for predicting prognostic information from radiomic features of glioblastoma magnetic resonance imaging scans acquired from the 2017 Multimodal Brain Tumor Segmentation Challenge. A variety of supervised classification methods trained on radiomic feature data and known prognostic outcomes will be compared through quantifiable metrics to determine which method most accurately predicts the prognostic class for a set of new patients unseen radiomic feature inputs.

The Multimodal Brain Tumor Segmentation (BraTS) Challenge [9] is an annual competition that seeks to employ the brightest minds in computational radiology to develop the following:

- Accurate segmentation algorithms of gliomas in magnetic resonance images.
- *Prediction of patient overall survival through radiomics and machine learning algorithms.*

It is the second objective above that this project aims to address.

**Datasets and Inputs**

The BraTS Challenge provides a standardized magnetic resonance imaging (MRI) dataset of patients diagnosed with gliomas from the Cancer Imaging Archive; tumor volumes have been segmented by expert radiologists and are included in the dataset. In addition the BraTS Challenge provides overall survival data for patients corresponding to these MRI scans. The MRI dataset and patient survival data from the 2017 BraTS Challenge was obtained for this project by entering the competition.

Though the original 2017 BraTS Challenge dataset contains both high-grade glioblastoma (GBM) MRI images (n = 210) and low-grade glioma MRI images (n = 75), only the GBM MRI images will be utilized for this project. The inputs to the radiomics algorithm were FLAIR MRI scans as "Images" and corresponding tumor segmentations as "Masks". An example of a GBM MRI scan and corresponding tumor segmentation from the BraTS dataset is visualized in Fig 1.

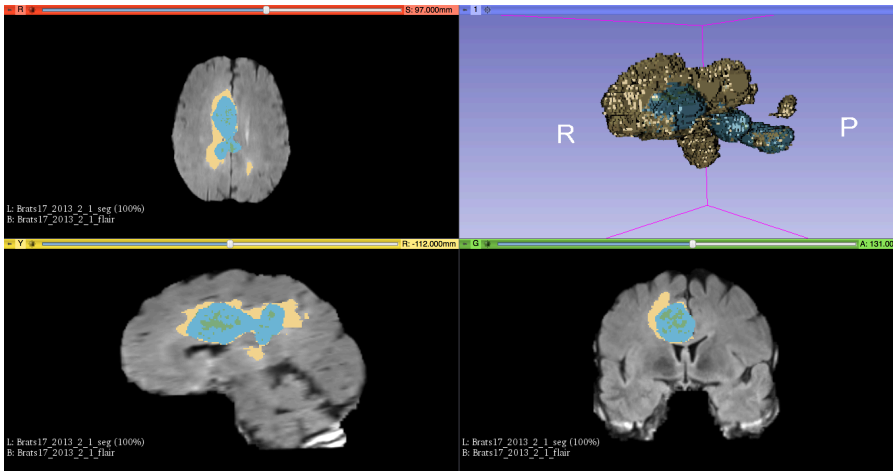


**Figure 1:** Visualization of GBM FLAIR MRI scan with corresponding tumor segmentation obtained from BraTS dataset. Tumor segmentation represented by colored areas. Color-coding represents subregions of tumor but will not be considered in this study. Tumor segmentation 3D volume is visualized in top right corner. Visualization performed in 3D Slicer [10].

94 radiomic features corresponding to statistics (19), shape (16), and texture (59) were extracted from each of the 210 GBM MRI images/segmentations by implementing a standardized open-source radiomics python library, PyRadiomics [11], with default parameters. Some image/mask combinations suffered from geometry mismatch, so these samples were discarded. A detailed list of the features extracted can be found in the PyRadiomics documentation. Statistical information for five of the radiomic features extracted is shown in Table 1. Dimensionality reduction will be applied to the full set of features to yield a subset of the most relevant features. The relevant features will then serve as inputs to the various machine learning models, described later in this document.

| | | | Textural Features | | |
|---|---|---|---|---|---|
| | **Variance (statistics)** | **Volume (shape)** | **Autocorrelation (GLCM)** | **Small Area Emphasis (GLSZM)** | **Short Run Emphasis (GLRLM)** |
| **count** | 150 | 150 | 150 | 150 | 150 |
| **mean** | 8643.94 | 11856.81 | 190.71 | 0.55 | 0.81 |
| **std** | 14328.38 | 14355.17 | 215.69 | 0.08 | 0.10 |
| **min** | 20.58 | 47.00 | 5.11 | 0.33 | 0.31 |
| **25%** | 2034.99 | 3274.75 | 67.45 | 0.50 | 0.75 |
| **50%** | 4273.07 | 7820.50 | 118.01 | 0.56 | 0.84 |
| **75%** | 8120.41 | 15919.75 | 240.79 | 0.61 | 0.88 |
| **max** | 98968.98 | 91299.00 | 1329.48 | 0.74 | 0.96 |

**Table 1**: Sample set of radiomic features derived via PyRadiomics for the 5 feature categories. GLCM = Gray Level Co-occurrence Matrix, Gray Level Size Zone Matrix = GLSZM, Gray Level Run Length Matrix = GLRLM are textural features.

The BraTS Challenge offers a .csv file with the number of days survived after diagnoses (overall survival) for 165 GBM patients. The patient overall survival data in this document are continuous numerical values. The data has been partitioned into two categorical values using a cutoff time of one year, which is based on median survival rates for glioblastoma [5]; patients who live longer than one year are assigned a 1, patients who live less than one year are assigned a 0. These new overall survival categories will serve as the output for our models.

The radiomic feature data and categorical overall survival data have been collated to yield the final dataset that will be used for this project. The dataset contains 150 rows, corresponding to the number of patients, and 95 columns corresponding to the radiomic features and the overall survival category.

**Solution Statement**

Four supervised machine learning algorithms will be compared to determine which is the most appropriate for classifying patients into their survival groups: Random Forest (RF), Naïve Bayes (NB), Decision Trees (DT), and Neural Networks (NN). These classifiers were chosen in accordance with Parmar et al., 2016 [7]; RF and NB showed high predictive performance while DT and NN showed low predictive performance. The resulting classifiers predicative performances will be measured using the area under the receiver operator characteristic curve (AUC) as a metric of comparison.

**Benchmark Model**

Parmar et al. published a landmark study comparing fourteen feature selection methods and twelve classification methods in terms of their performance and stability for predicting overall lung cancer patient survival [7]. This study contains AUC values for each of the feature selection and classification combinations, which can be directly compared with our projects AUC values.

Though the models in Parmar et al. were trained using lung cancer CT imaging and a different survival classification threshold (2 years instead of 1), it can be inferred that underlying radiomic principles are similar regardless of the tumor type studied or

imaging methodology. Therefore, it is logical to predict our classification methods will follow comparable predictive performance trends as observed in this study.

It can be seen clearly in Parmar et al. that the classifier type accounted for the vast majority of the variation of the AUC curve; therefore we will limit our study to compare only classifier methods and not feature selection methods.

**Evaluation Metrics**

The Receiver Operating Characteristics (ROC) curve is a commonly utilized metric to evaluate binary classifier output performance [12]. Often ROC curves display true positive rate on the Y-axis and false positive rate on the X-axis as discrimination threshold is varied. The area under the ROC curve, AUC, can be used to quantify the degree to which a model is able to accurately classify data. AUC values close to 0.5 are worse, signifying random guessing, while AUC values close to 1.0 are better, signifying accurate classification. Herein, we will use the AUC values to compare the predictive performance of our four classification methods, and compare them to the results in Parmar et al.

**Project Design**

Having already performed radiomic feature extraction via PyRadiomics, the immediate next step is to perform any necessary data preprocessing, such as normalization of features, removal of outliers, etc. The data will then be split into a training and testing set, with a testing size = 0.2. Since the dataset currently contains 96 features for 150 patients, the feature space will need to be reduced significantly to avoid overfitting. Feature selection will be performed to determine and select the most relevant radiomic features. Dimensionality reduction will be performed using principle component analysis or a related algorithm. The four previously discussed machine learning algorithms (RF, NV, DT, NN) will then be trained using the relevant radiomic features as input and survival classification as output via cross validation. Grid search will also be implemented to find the optimal parameters for each model. Finally, the models will each be evaluated on the test set. Performance between classification methods will be quantified with AUC values and compared with the results in Parmar et al.

## References


1. Gillies, R. J.; Kinahan, P. E.; Hricak, H., Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016,** *278* (2), 563-577.
2. Aerts, H. J. W. L.; Velazquez, E. R.; Leijenaar, R. T. H.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; Hoebers, F.; Rietbergen, M. M.; Leemans, C. R.; Dekker, A.; Quackenbush, J.; Gillies, R. J.; Lambin, P., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications* **2014,** *5*, 4006.
3. Network, T. C. G. A. R., Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *New England Journal of Medicine* **2015,** *372* (26), 2481-2498.
4. Kotrotsou, A.; Zinn, P. O.; Colen, R. R., Radiomics in Brain Tumors: An Emerging Technique for Characterization of Tumor Environment. *Magnetic Resonance Imaging Clinics of North America* **2016,** *24* (4), 719-729.
5. Narang, S.; Lehrer, M.; Yang, D.; Lee, J.; Rao, A., Radiomics in glioblastoma: current status, challenges and potential opportunities. *Translational Cancer Research* **2016,** *5* (4), 383-397.
6. Chaddad, A.; Zinn, P. O.; Colen, R. R. In *Radiomics texture feature extraction for characterizing GBM phenotypes using GLCM*, 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), 16-19 April 2015; 2015; pp 84-87.
7. Parmar, C.; Grossmann, P.; Bussink, J.; Lambin, P.; Aerts, H. J. W. L., Machine Learning methods for Quantitative Radiomic Biomarkers. *Scientific Reports* **2015,** *5*, 13087.
8. Parmar, C.; Grossmann, P.; Rietveld, D.; Rietbergen, M. M.; Lambin, P.; Aerts, H. J. W. L., Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer. *Frontiers in Oncology* **2015,** *5*, 272.
9. Menze, B. H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; Lanczi, L.; Gerstner, E.; Weber, M. A.; Arbel, T.; Avants, B. B.; Ayache, N.; Buendia, P.; Collins, D. L.; Cordier, N.; Corso, J. J.; Criminisi, A.; Das, T.; Delingette, H.; Ç, D.; Durst, C. R.; Dojat, M.; Doyle, S.; Festa, J.; Forbes, F.; Geremia, E.; Glocker, B.; Golland, P.; Guo, X.; Hamamci, A.; Iftekharuddin, K. M.; Jena, R.; John, N. M.; Konukoglu, E.; Lashkari, D.; Mariz, J. A.; Meier, R.; Pereira, S.; Precup, D.; Price, S. J.; Raviv, T. R.; Reza, S. M. S.; Ryan, M.; Sarikaya, D.; Schwartz, L.; Shin, H. C.; Shotton, J.; Silva, C. A.; Sousa, N.; Subbanna, N. K.; Szekely, G.; Taylor, T. J.; Thomas, O. M.; Tustison, N. J.; Unal, G.; Vasseur, F.; Wintermark, M.; Ye, D. H.; Zhao, L.; Zhao, B.; Zikic, D.; Prastawa, M.; Reyes, M.; Leemput, K. V., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging* **2015,** *34* (10), 1993-2024.
10. Fedorov, A.; Beichel, R.; Kalpathy-Cramer, J.; Finet, J.; Fillion-Robin, J.-C.; Pujol, S.; Bauer, C.; Jennings, D.; Fennessy, F.; Sonka, M.; Buatti, J.; Aylward, S.; Miller, J. V.; Pieper, S.; Kikinis, R., 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging* **2012,** *30* (9), 1323-1341.
11. Joost JM van Griethuysen, A. F., Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, Hugo JWL Aerts, Computational Radiomics System to Decode the Radiographic Phenotype. *Submitted* **2017**.
12. Fawcett, T., An introduction to ROC analysis. *Pattern recognition letters* **2006,** *27* (8), 861-874.