Machine Learning Engineer Nanodegree
Capstone Proposal
Kareem Wahid
May 28th, 2017

**Predicting Survival Outcomes for Glioblastoma Patients via Radiomic Features**

**Domain Background**
Radiomics is a budding new field of medical informatics that seeks to extract
mathematically defined quantitative features such as shape, statistics, 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 in 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 the best to use due to lack of standardization in
the field. As a current medical student and aspiring radiologist, exploring the new and
exciting field of radiomics through this project gives me a chance to help develop a
modality that could one day impact my future patients.

**Problem Statement**
The goal of this project is to determine which 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; 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.
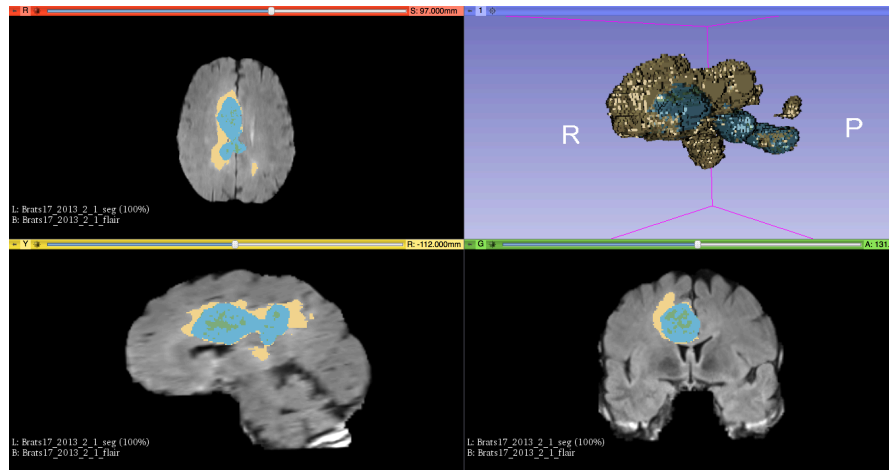


Figure 1: Visualization of GBM FLAIR MRI scan with corresponding tumor segmentation obtained from BraTS dataset. Tumor segmentation represented by colored areas. Color-coding represent subregions of tumor but will not be considered in this study. Tumor segmentation 3D volume is visualized in top right 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 GBM MRI images/segmentations by implementing a standardized open-source radiomics python library, PyRadiomcs [11], with default parameters. A sample list of radiomic features are shown in Table 1. A detailed list of the features extracted can be found at the following link: http://pyradiomics.readthedocs.io/en/latest/features.html. These features are continuous numerical values. Dimensionality reduction will be applied to the 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.

| Statistical | Shape | GLCM | GLSZM | GLRLM |
|---|---|---|---|---|
| Variance | Volume | Autocorrelation | Small Area Emphasis | Short Run Emphasis |
| Uniformity | Surface Area | Joint Average | Large Area | Long Run |

| | | | Emphasis | Emphasis |
|---|---|---|---|---|
| Kurtosis | Sphericity | Cluster Prominence | Gray Level Non-Uniformity | Gray Level Non-Uniformity |

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 each patient survived after diagnoses (overall survival). The patient overall survival is a continuous numerical value. It will be 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 will be assigned as 1, patients who live less than one year will be assigned as 0. These new overall survival categories will serve as the output for our models.

**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 prediction 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 similar 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 (sensitivity/recall) on the Y-axis and false positive rate (fall-out) 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 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 obtained the GBM MRI data with survival outcome data from the BraTS challenge and performed radiomic feature extraction via PyRadiomics, the immediate next steps would be the following:

- Transform patient survival outcome data from continuous numerical values and remove any patients who do not have survival data from the dataset.
- Perform any necessary data preprocessing (normalization, removal of outliers, etc.).

Once the dataset has been finalized, the data will be split into a training and testing set, with a testing size = 0.2. Since the dataset currently contains 96 features for about 200 patients, the feature space will need to be reduced significantly. Feature selection will be performed to determine and select the most relevant radiomic features. A metric such as variance will be used to select the top features to use for the machine learning 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.