RA-23

A Machine-Learning Based Classifier for Predicting a Multi-Parametric Probability Map of Active Tumor Extent within Glioblastoma Mulitforme

Sarah J. Fouke[1], Kilian Weinberger[2], Matthew Kelsey[2], Sharath Cholleti[3], David G. Politte[2], Daniel Marcus[2], Alicia Boyd[2], Bart Keogh[1], Tammie Benzinger[2], Mikhail Milchenko[2], Lauren Kim[4], Fred Prior[2]

[1]Swedish Neuroscience Institute, Seattle, WA 98122
[2]Washington University School of Medicine, St. Louis, MO 63110
[3]Emory University, Atlanta, GA
[4]Massachusetts General Hospital, Boston, MA 02114

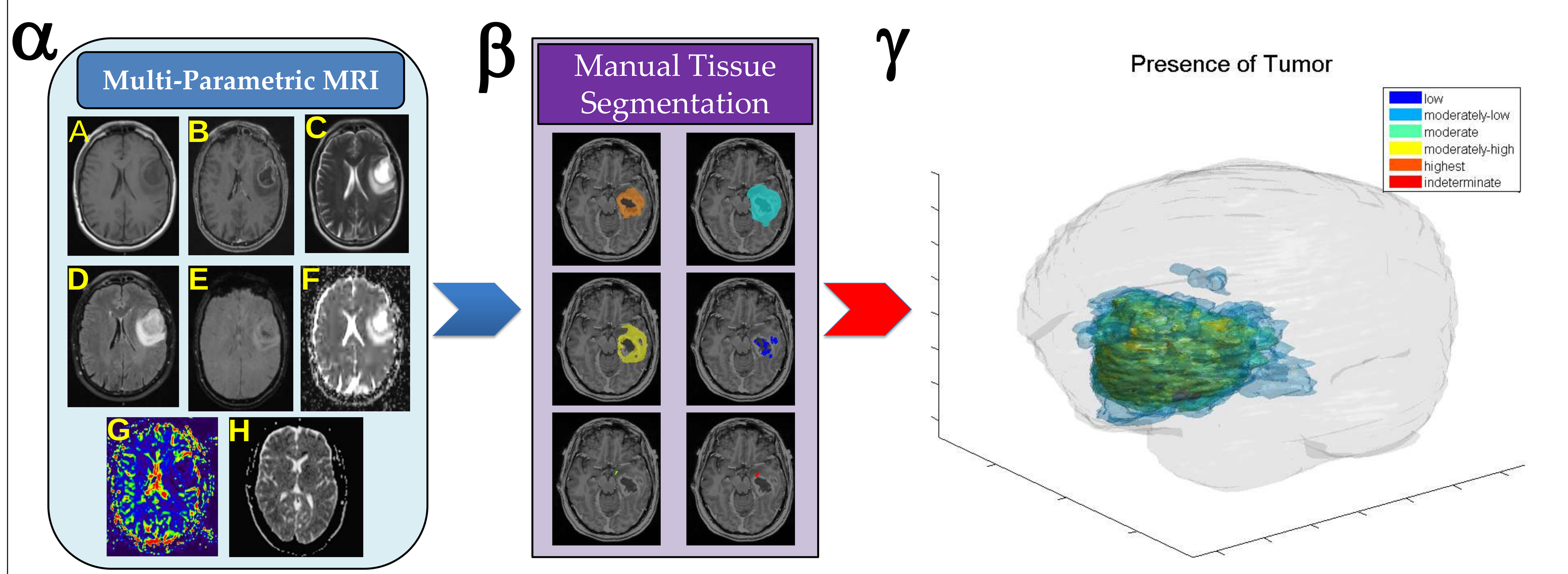
Introduction

- Glioblastoma Mulitforme (GBM) is highly infiltrative, making precise delineation of tumor margin difficult.
- Multi-parametric MRI has been shown to have advantages over contrast enhanced MRI as a method for determining the spatial extent of tumor involvement.
- Manual image segmentation and classification are time-consuming and prone to error. We propose a machine-learning based multi-parametric approach that uses radiologist generated labels to train a classifier that is able to classify tissue on a voxel-wise basis and automatically generate a tumor segmentation.

Methods

- Preoperative MRI examinations of subjects with GBM were chosen from the COmprehensive Neuro-oncology Data Repository (CONDR) at Washington University in St. Louis and Swedish Neuroscience Institute (Seattle, WA).
- Eight MRI image types (primary and derived) were co-registered, transformed to a standard template space with 1mm isotropic voxels, and segmented (Fig 1α).
- A rule-based multi-parametric image analysis was conducted by a board certified neuroradiologist to define a standard for comparison (Fig 1β,γ, also see Poster RA-24 for details).
- Voxel Labels were generated by combining manual segmentations based on the radiologist's rule set and estimate of the probability of active tumor (Fig 1γ).
- A Random Forests classifier was trained using a leave-one-out experimental paradigm. A linear regression analysis was also implemented for comparison.
- A leave-one-out experiment with N labeled data sets uses N-1 data sets to train the classifier and then predicts the labels of the N^{th}. This process is repeated until all data sets have been predicted by the Classifier. Figure 2 illustrates the Radiologist's multi-parametric tumor map and the map predicted by the Random Forests Classifier for one case.
- Receiver Operating Characteristic (ROC) analysis was used to compare the predictions of the Random Forests classifier and a linear regression-based classifier relative to the manual segmentation standard.

Figure 1. Rule-Based Radiologist Analysis Establishes "Truth"



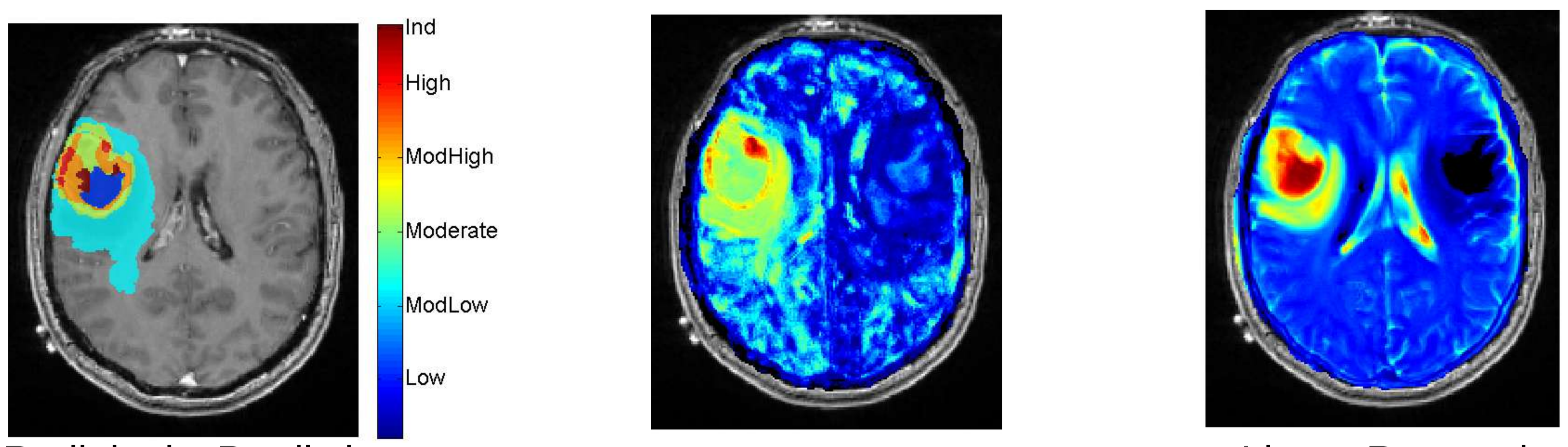
- Neuroradiologist use eight (8) image types to perform tissue segmentation:
  A. T1 pre-contrast
  B. T1 post-contrast
  C. T2
  D. Fluid Attenuated Inversion Recovery (FLAIR)
  E. Susceptibility Weighted Imaging (SWI)
  F. Apparent Diffusion Coefficient (ADC)
  G. relative Cerebral Blood Volume (rCBV)
  H. TRACEW

- All image types are co-registered and interpolated to 1mm isotropic voxels.
- Neuroradiologist uses rules for combining the segmentation maps and assigns a tissue classification and probability estimate (likelihood of being active tumor).
- A MatLab routine takes this information and computes a probability estimate for each voxel in the image volume to be used as labels for machine learning.

Figure 2. Machine Learning Classifiers Generate Segmentations and Tumor Multi-Parametric Probability Maps

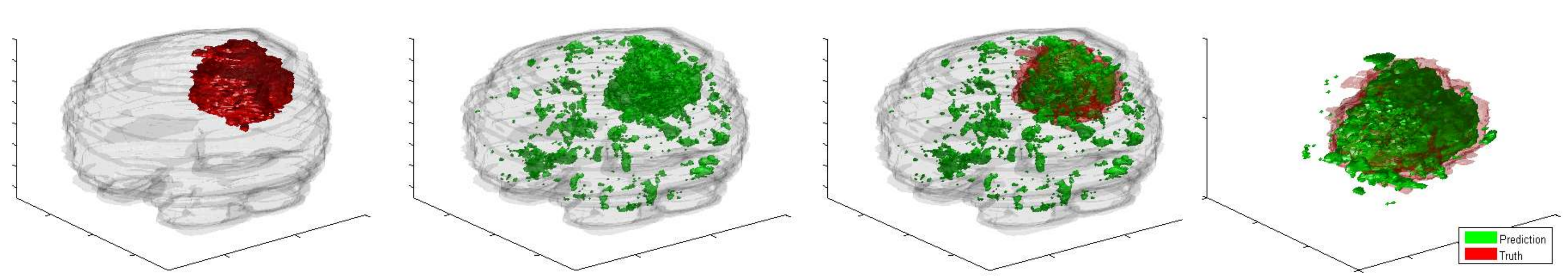
- The Random Forests (RF) algorithm constructs an ensemble of decision trees. Each decision tree is constructed by selecting a random subset of features and training examples, creating a variety of *experts*. The leaves of the trees are associated with constant predictions. RF combines the votes of all the trees for overall prediction.
- To construct a tumor probability estimate and tissue segmentations, feature vectors are constructed for each voxel in the brain, using the value of that voxel in each of the 8 MR data types.
- Once trained, the classifier is applied to every voxel (every feature vector) in the test set and classifies the tissue as normal or malignant.

Tumor Segmentations



Radiologist Prediction | Random Forests | Linear Regression

Tumor Multi-Parametric Probability Maps

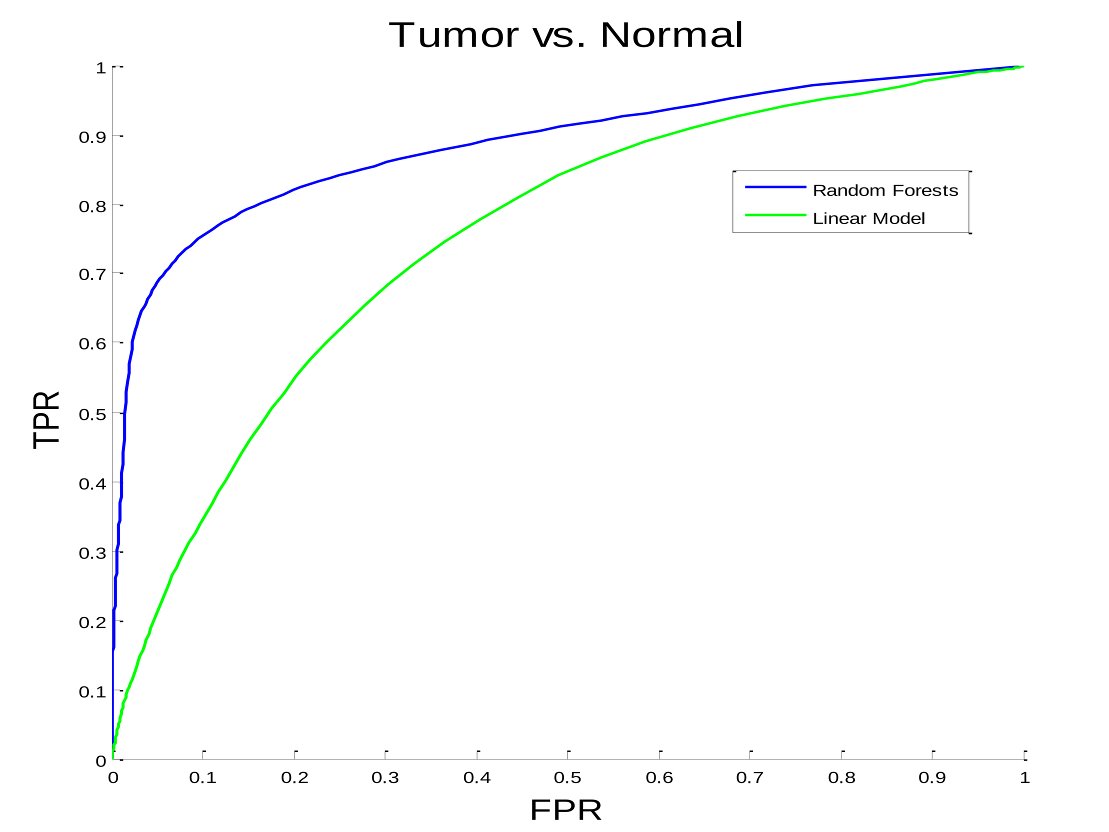


Radiologist Generated | Random Forests Predicted | Overlap of Total Map, Tumor Region

- The Random Forests algorithm classifies other brain tissue as having a high probability of being cancer. Separating false positives from true positives in these areas is an area of active research.

Results

- The Random Forests classifier generated a multi-parametric probability map that more accurately predicted radiologist-generated segmentations and tumor extent than did the linear classifier.



ROC Analysis (AUC)		
Subject	Random Forests	Linear Classifier
W001	0.918	0.702
W010	0.905	0.707
W015	0.83	0.725
W019	0.939	0.786
W025	0.927	0.804
Overall	**0.885**	**0.75**

Area Under Curve (AUC) for Random Forests and Linear classifiers partitioning normal from diseased brain.

Data were normally distributed with a Shapiro Wilk W test ($P \geq 0.64$). Paired t tests indicated that random forests resulted in higher AUC values for tumor vs. normal and moderate tumor vs. normal ($P = 0.002$ and 0.005).

Conclusions

- The infiltrative nature of gliomas makes assessment of tumor burden a challenge, and multi-parametric imaging markers may offer a method to improve our measures of tumor invasion and, ultimately, extent of resection.
- By enhancing our multi-parametric approach with Machine Learning we eliminate manual segmentation and generate a probability map that incorporates contrast enhancement with additional MRI markers to produce a composite image that predicts the probability of viable tumor and tissue type.

References

1. McMillan, K.M., et al., *An objective method for combining multi-parametric MRI datasets to characterize malignant tumors.* Medical Physics, 2007. **34**(3): p. 1053-1061.
2. Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32.
3. Corso, Jason J., et al. *Efficient multilevel brain tumor segmentation with integrated bayesian model classification.* Medical Imaging, IEEE Transactions on 27.5 (2008): 629-640.

Acknowledgments

- This work was partially supported by NIH grant 1R01NS066905-01. Data were provided by clinical trial NCT01124461.
- The authors wish to thank Stephen Tyree for his assistance running the machine learning algorithms.