Interpretable Brain Tumor Classification

Nicholas Nuechterlein, Xingfan Huang, Erin Wilson
Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle

Background: treating patients with Glioblastoma

Glioblastoma (GBM) is the most common and most aggressive primary brain tumor. Patients with this disease live on average only 12-15 months, however there exists a subgroup of extremely short-term survivors who usually live less than 6 months. These short-term survival patients are critical to identify so they can be treated with experimental therapies right away.

Recent work has shown that the genomes in the short-survivor tumors exhibit extreme changes in their DNA code but it requires dangerous and costly brain surgery to get a tumor sample to confirm. Instead, we aim to use non-invasive magnetic resonance (MR) scans to predict these extreme genomic changes in tumors using computer vision techniques.

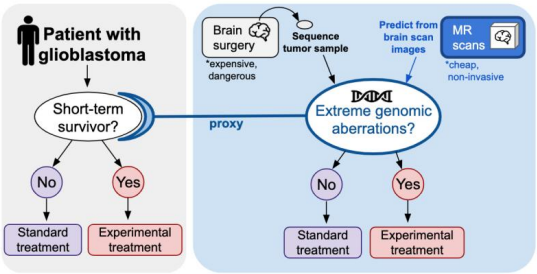


Figure 1: Overview of glioblastoma patient treatment decision flow.

Problem Statement

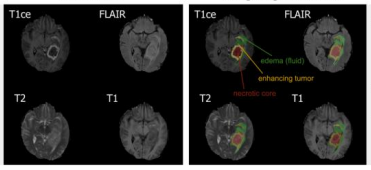
To better recommend appropriate treatment plans to GBM patients while avoiding dangerous surgery options, we aim to use computer vision techniques to train a model to predict the presence of extreme genomic changes in tumors from non-invasive MR image scans.

Primary dataset: GBM MR Scans

The Cancer Imaging Atlas (TCIA)

- 4 types of MR scan modalities (T1ce, FLAIR, T2, T1)
- Images segmented into 3 tumor compartments (necrotic tissue, enhancing tumor, edema)

Raw Scans | **Image segmentations**

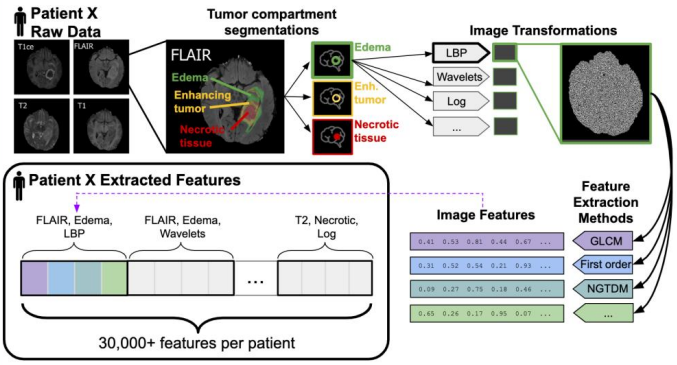


Data Challenges

- Extremely high-dimensional data (50,000,000 voxels per patient)
- Small example set (46 GBM patients)
- High variation due to noise

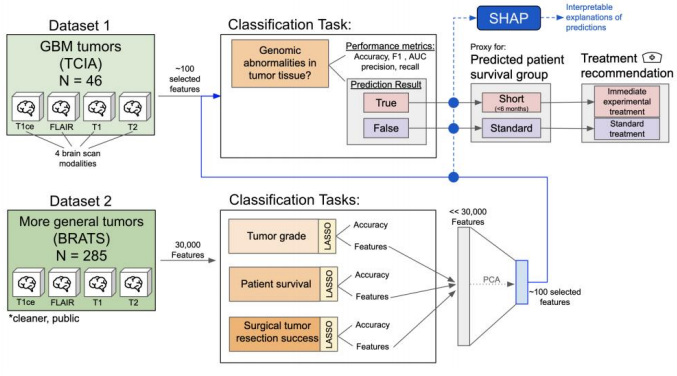
Image feature generation

Groups of features are generated combinatorially by choosing: 1) a modality (e.g., FLAIR), 2) a tumor compartment (e.g., edema), 3) an image transformation method (e.g., LBP), and then applying a number of feature extraction methods (e.g., GLCM) to get numerical features. All combinations result in about 30,000 features per patient.



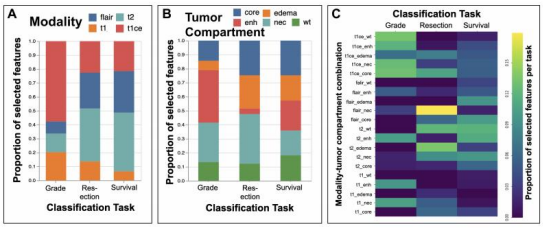
Computational Methods Overview

- Extract 30,000 image features from each patient
- Group features by MR modality, tumor compartment, and feature extraction method
- Use lasso to select 50 groups of features and use PCA to project each group to two dimensions, resulting in 100 features
- Use selected features found in the larger dataset as features for the smaller GBM dataset
- Train ML models with transferred features; record accuracy; use SHAP to extract interpretable explanations of final predictions



Trends in lasso selected features

- A: T1ce-based features are most important for tumor grade classification and T2-based features are important for survival prediction
- B: The enhancing tumor compartment is most important for tumor grade classification and nearly irrelevant to resection prediction
- C: Features from the necrotic tissue compartment on the FLAIR modality are most important for resection prediction, but inconsequential for survival prediction



BraTS and TCIA Classification Accuracy

Our method (1*) outperforms vanilla ML models applied to the entire dataset

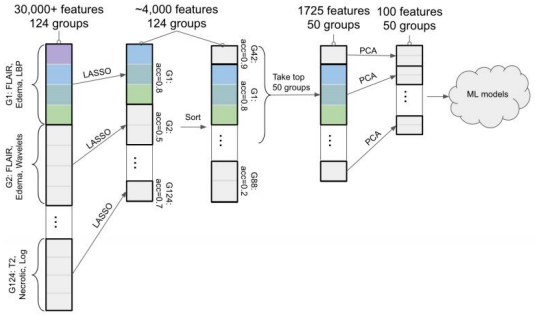
Table 1: BraTS Experimental Results (Accuracy)

Task	Method	lasso	SVM	MLP	XGBoost	DT	RF	LR
Survival	1*	**0.700**	0.700	0.675	**0.681**	**0.626**	**0.650**	0.700
	2	0.650	0.589	0.540	0.577	0.571	0.583	0.589
	3	0.681	**0.736**	**0.687**	0.656	0.564	0.607	**0.736**
	4	0.663	0.613	0.601	0.620	0.577	0.564	0.613
	All Features	0.607	0.632	0.650	0.620	0.583	0.546	0.632
Resection	1*	**0.843**	**0.855**	**0.795**	**0.759**	**0.711**	**0.723**	**0.855**
	2	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-
	All Features	0.639	0.711	0.687	0.711	0.687	0.639	0.711
Tumor Grade	1*	**0.933**	**0.926**	**0.930**	**0.944**	**0.891**	**0.930**	**0.926**
	2	0.923	0.923	0.923	0.930	0.853	**0.930**	0.921
	3	0.930	0.919	0.895	0.919	0.884	0.926	0.919
	4	0.910	0.919	0.905	0.940	**0.891**	0.905	0.919
	All Features	0.919	0.905	0.909	0.930	0.877	0.902	0.905

Table 2: TCIA Experimental Results (Accuracy)

Task	Training Features	lasso	SVM	MLP	XGBoost	DT	RF	LR
Copy-Number	Survival Features	0.696	0.783	0.739	0.739	0.652	0.739	0.783
	Resection Features	0.717	**0.804**	**0.783**	**0.804**	**0.761**	**0.761**	**0.804**
	Tumor Grade Features	**0.913**	**0.804**	0.587	**0.804**	**0.761**	0.696	**0.804**
	All Features	0.761	0.761	0.587	0.761	**0.761**	0.739	0.761

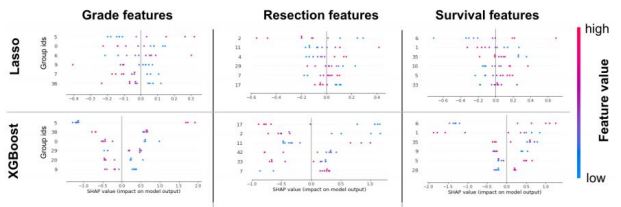
Feature selection method 1



Model Interpretability with SHAP

SHapley Additive exPlanation (SHAP) quantifies the contribution of each input group of features in machine learning model predictions with SHAP values.

- Running SHAP on various models for the TCIA classification tasks identifies groups of features that are important for prediction
- Groups of features with the top SHAP scores are consistent across models, but not when using different sets of input features



Top SHAP feature groups common to all 3 models

Group 0: t1ce, wt, glcm
Group 5: shape
Group 9: t1ce, wt, glrlm

Group 2: t2, core, ngtdm
Group 11: t2, edema, first order

Group 1: t1, nec, ngtdm
Group 5: t1ce, enh, gldm
Group 6: t2, nec, glszm
Group 35: t2, wt, ngtdm

Conclusions

- Our grouped feature selection method outperforms vanilla machine learning models without compromising interpretability
- Transferring selected features from tasks on the BraTS dataset to the smaller TCIA dramatically improves performance on the TCIA dataset far beyond human proficiency
- Our feature selection results confirm prior domain knowledge and suggest new conjectures about the role of the necrotic area on the FLAIR modality in neurosurgery