PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

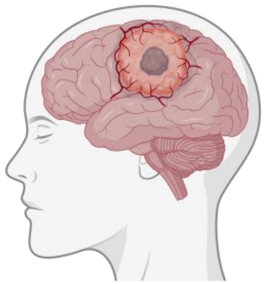
Radiogenomic modeling predicts survival-associated prognostic groups in glioblastoma

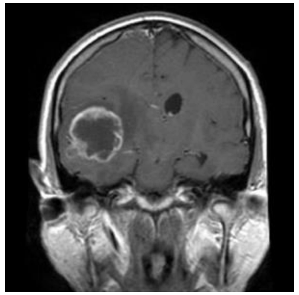
Nicholas Nuechterlein

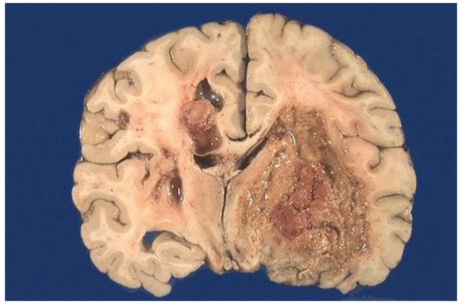
10/25/2021

The University of Washington
Paul G. Allen School of Computer Science & Engineering

What is Glioblastoma?

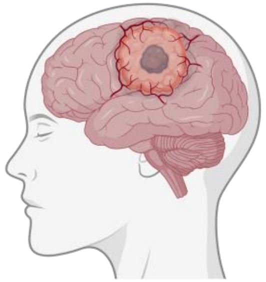


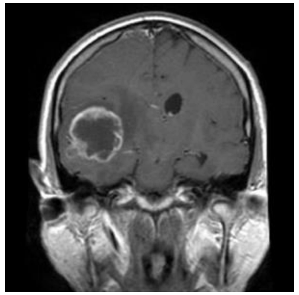


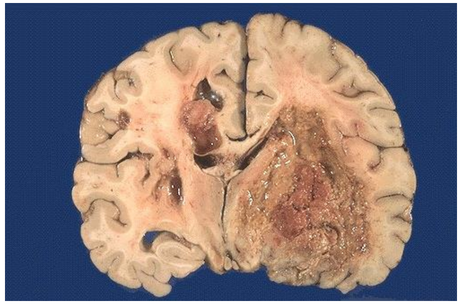


- Most common and aggressive primary adult malignant brain tumor
- Median survival of 15 months
- Incurable because
  - Extremely heterogeneous
  - Blood brain barrier
- Last approved therapeutic agent was in 2005

What is Glioblastoma?



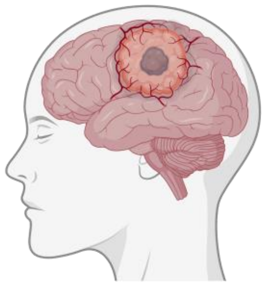


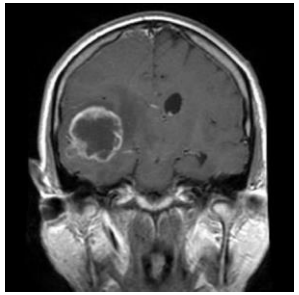


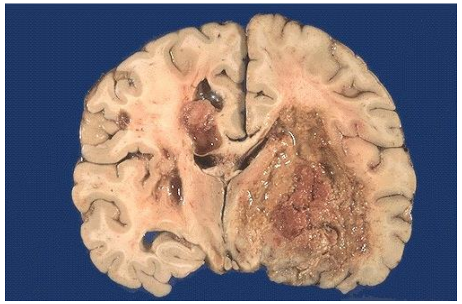
- Most common and aggressive primary adult malignant brain tumor
- Median survival of 15 months
- Incurable because
  - Extremely heterogeneous
  - Blood brain barrier
- Last approved therapeutic agent was in 2005

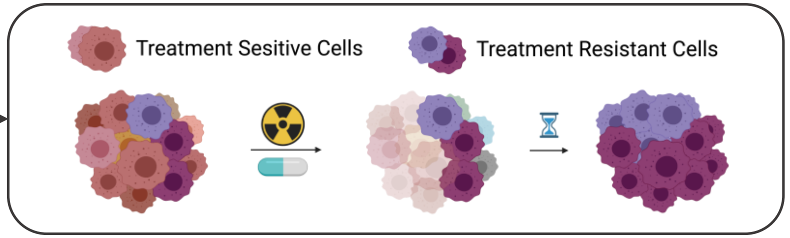
What is Glioblastoma?

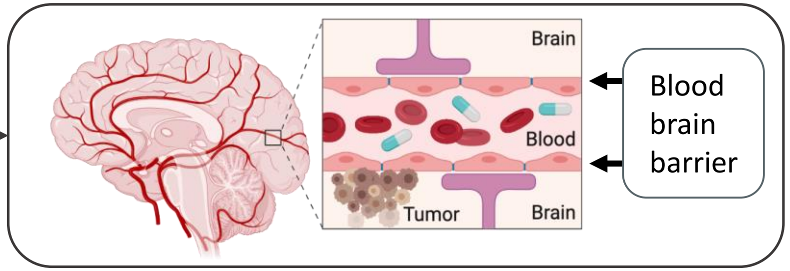
- Most common and aggressive primary adult malignant brain tumor
- Median survival of 15 months
- Incurable because
  - Extremely heterogeneous
  - Blood brain barrier
- Last approved therapeutic agent was in 2005



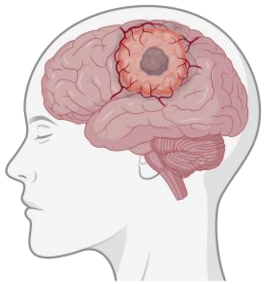


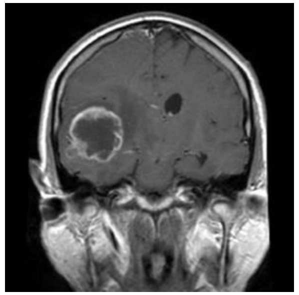


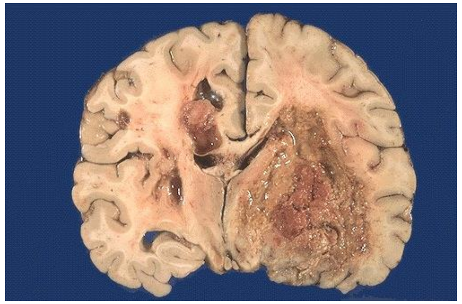




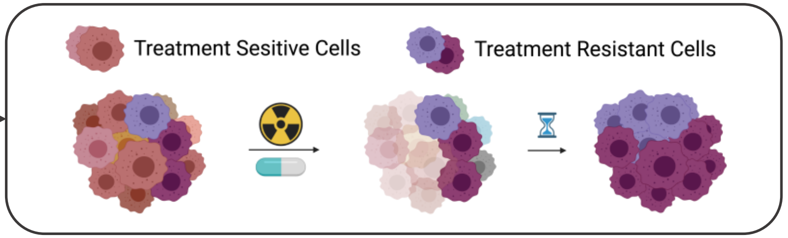
What is Glioblastoma?

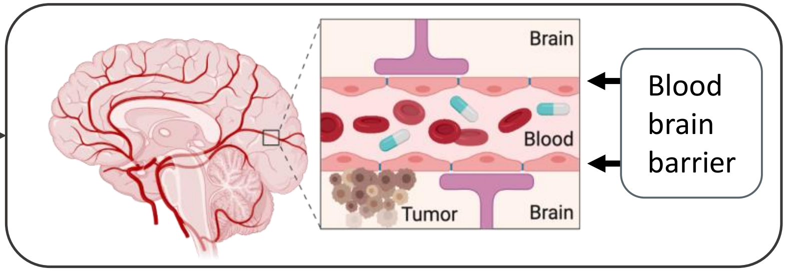


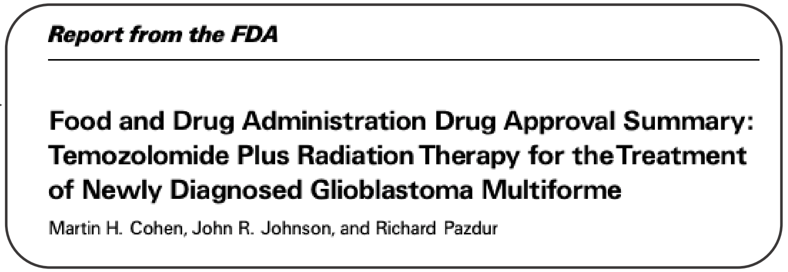




- Most common and aggressive primary adult malignant brain tumor
- Median survival of 15 months
- Incurable because
  - Extremely heterogeneous
  - Blood brain barrier
- Last approved therapeutic agent was in 2005



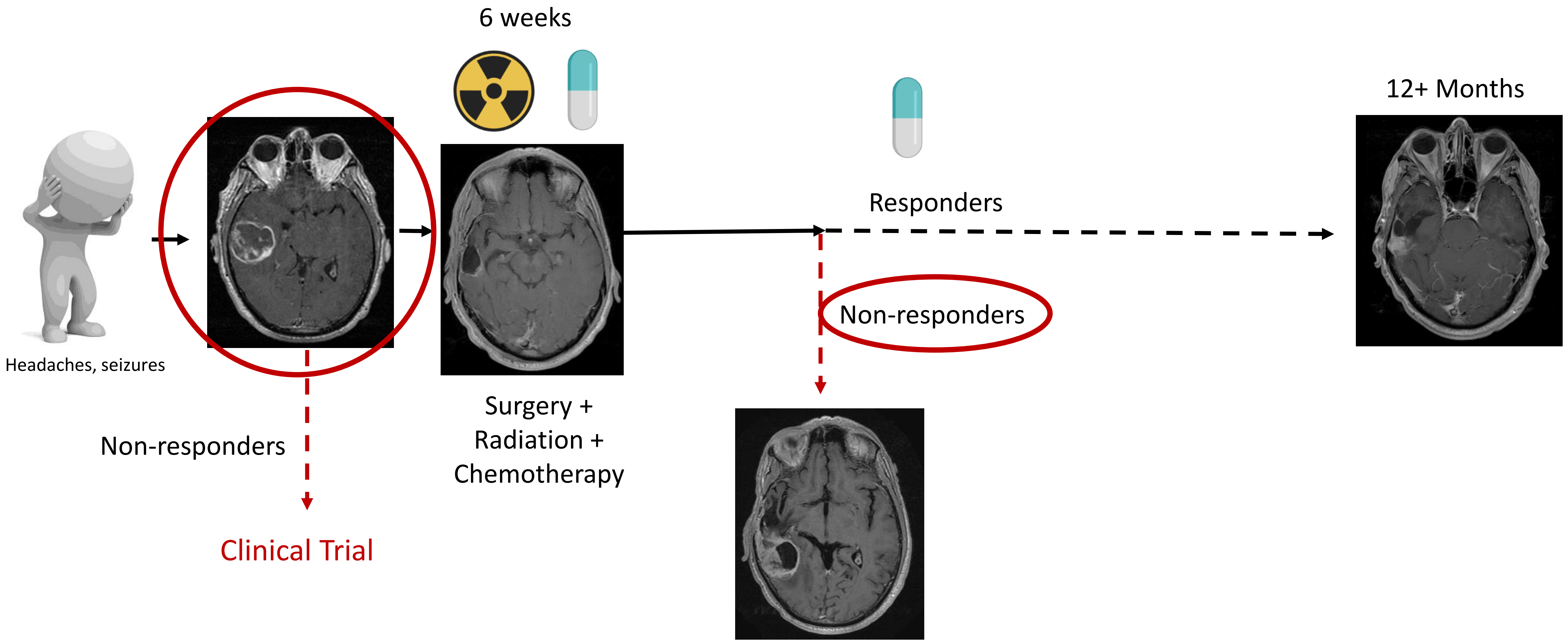



Report from the FDA

Food and Drug Administration Drug Approval Summary: Temozolomide Plus Radiation Therapy for the Treatment of Newly Diagnosed Glioblastoma Multiforme

Martin H. Cohen, John R. Johnson, and Richard Pazdur

Patient Clinical Course



Motivation for predicting short-term survivors

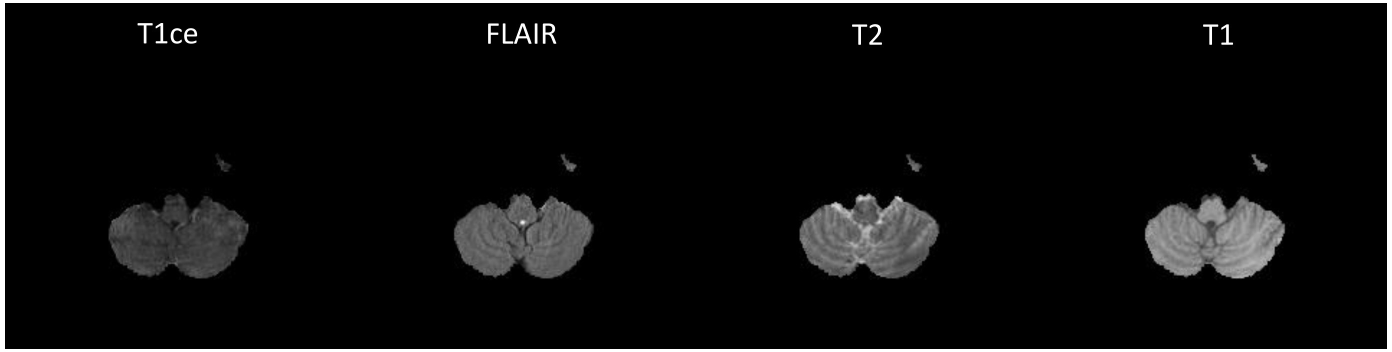
- Better for patients
  - Poor survivors have the most to gain from upfront trials
- Better for trials
  - Identifying poor survivors upfront can help balance clinical trial arms
  - Trials will run faster with poor survivors

- **But we need to know who the poor survivors are upfront**

Data

- MRI
  - Rich, global representation of tumor
  - Cheap, fast, non-invasive, repeatable
  - Volumetric
    - 255 x 255 x 155 x 4
    - (> 50 M voxels)
- Our data
  - 46 TCIA preoperative glioblastomas MRI with T1ce, FLAIR, T2, T1

Magnetic Resonance Imaging (MRI)

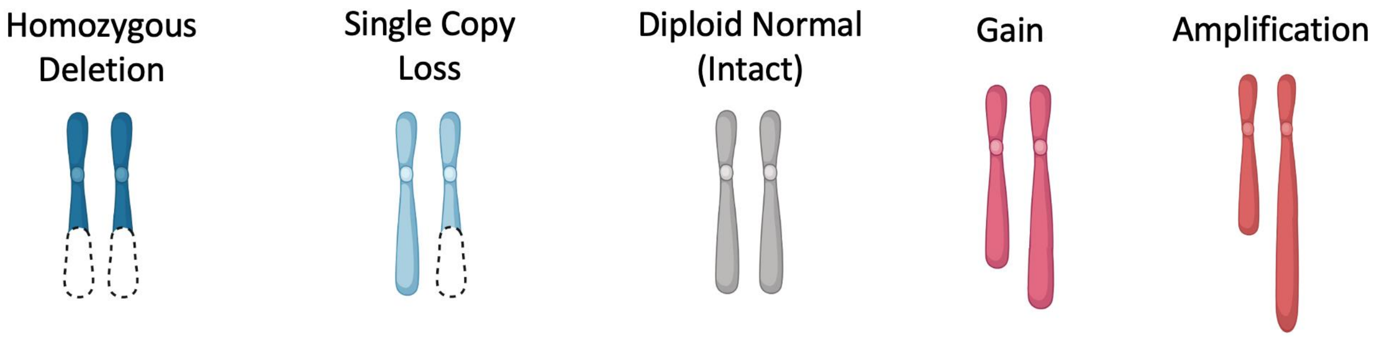


THE CANCER IMAGING ARCHIVE

NATIONAL CANCER INSTITUTE
THE CANCER GENOME ATLAS

Copy Number

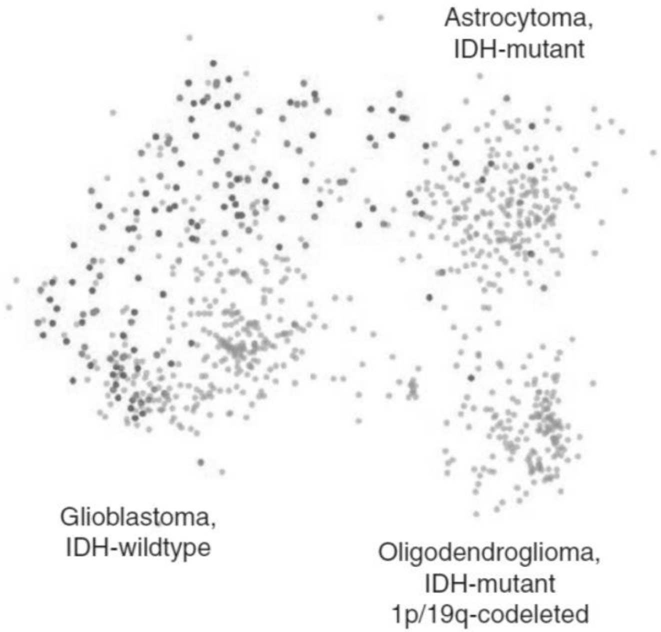
- Captures DNA structure
- Unlike MRI: invasive, expensive, not repeatable
- 23,000 x 1 (gene-level)
- Values in $\{-2, -1, 0, 1, 2\}$

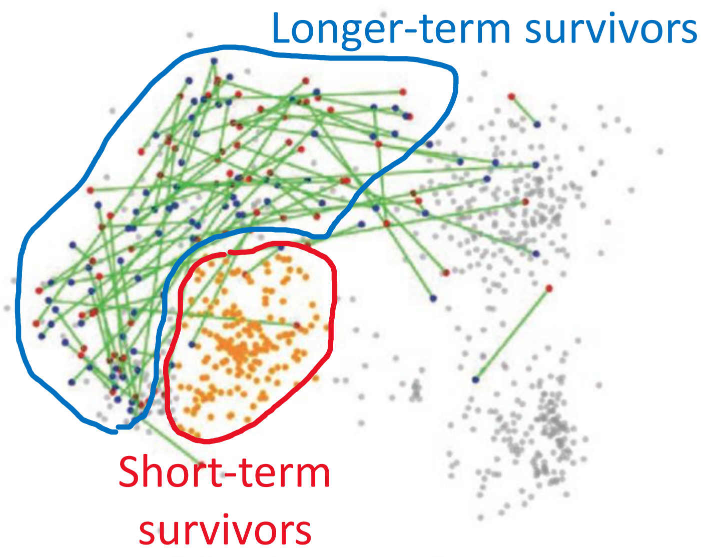


NATIONAL CANCER INSTITUTE
THE CANCER GENOME ATLAS

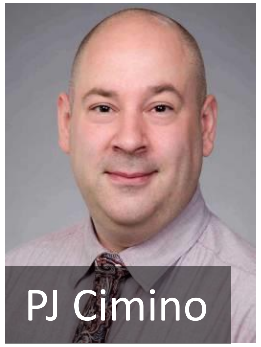
Poor survivor definition

Glioblastoma patients who undergo second resections live longer

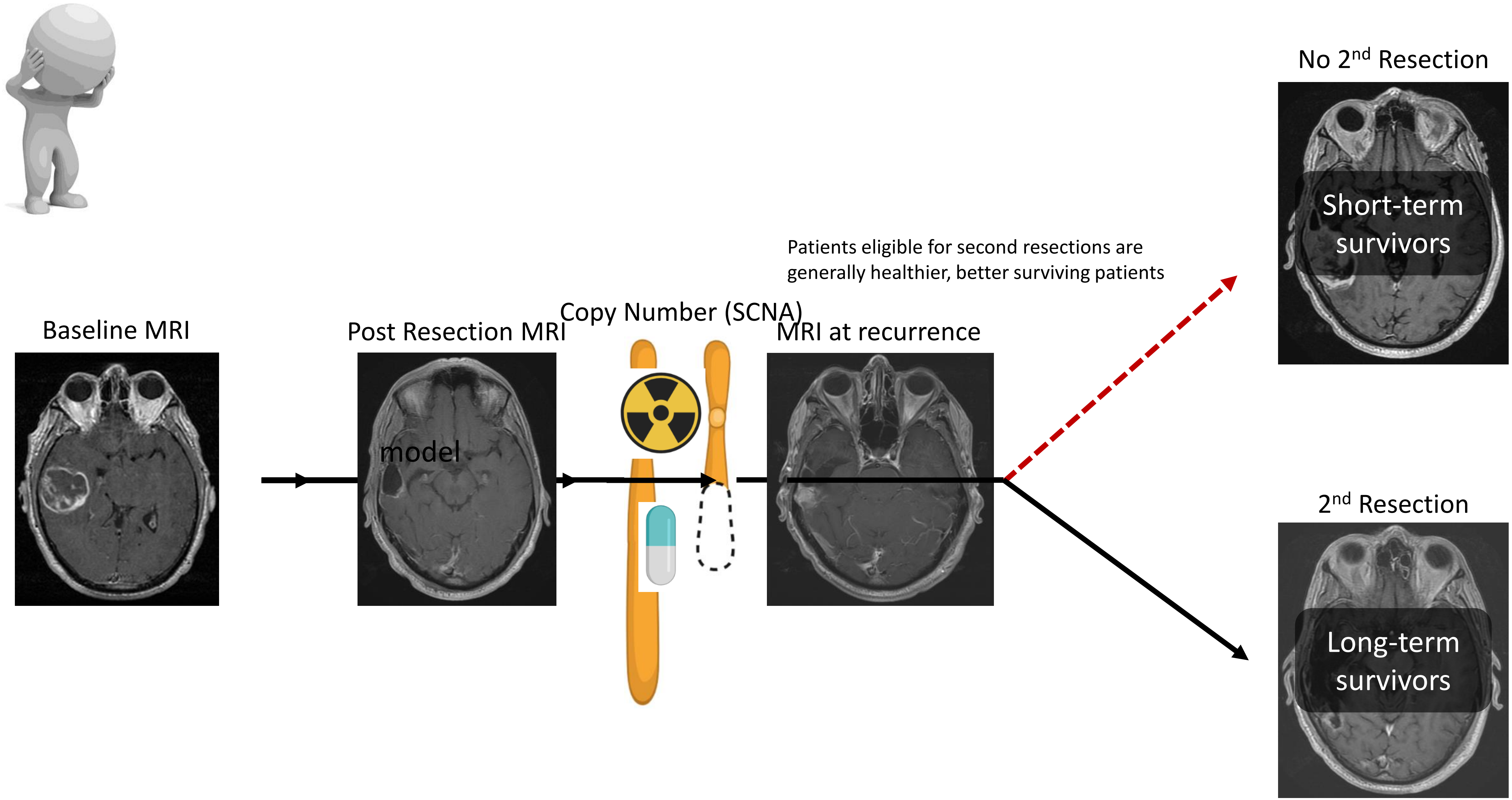




Distribution disparity of paired
initial/recurrent glioblastoma cohort



Cimino, et al. *Neuro-oncology* (2018)



No 2nd Resection
Short-term survivors
Patients eligible for second resections are generally healthier, better surviving patients
Copy Number (SCNA)
Baseline MRI
Post Resection MRI
MRI at recurrence
model
2nd Resection
Long-term survivors

Methods

- Radiogenomics/radiomics
  - An evolving field in medical imaging that strives to equate quantitative image features with the genomic profile of pictured tissues
- Pipelines
  - Image acquisition
  - Image normalization
  - Feature extraction
  - Feature selection
  - Prediction using ML models
  - (Or end-to-end deep learning models)
- Novelty
  - Feature selection method
  - Unique clinical application

Multicenter study demonstrates radiomic features derived from magnetic resonance perfusion images identify pseudoprogression in glioblastoma

Nabil Elshafeey[1], Aikaterini Kotrotsou[1,2], Ahmed Hassan[1], Nancy Elshafei[2,3], Islam Hassan[2], Sara Ahmed[2],

Neuro-Oncology

20(6), 848–857, 2018 | doi:10.1093/neuonc/nox188 |

Radiomic subtyping imp beyond key molecular, c characteristics in patien

Philipp Kickingereder, Ulf Neuberger, David Bonekamp, Paula L. Piechotta, Michael Götz,

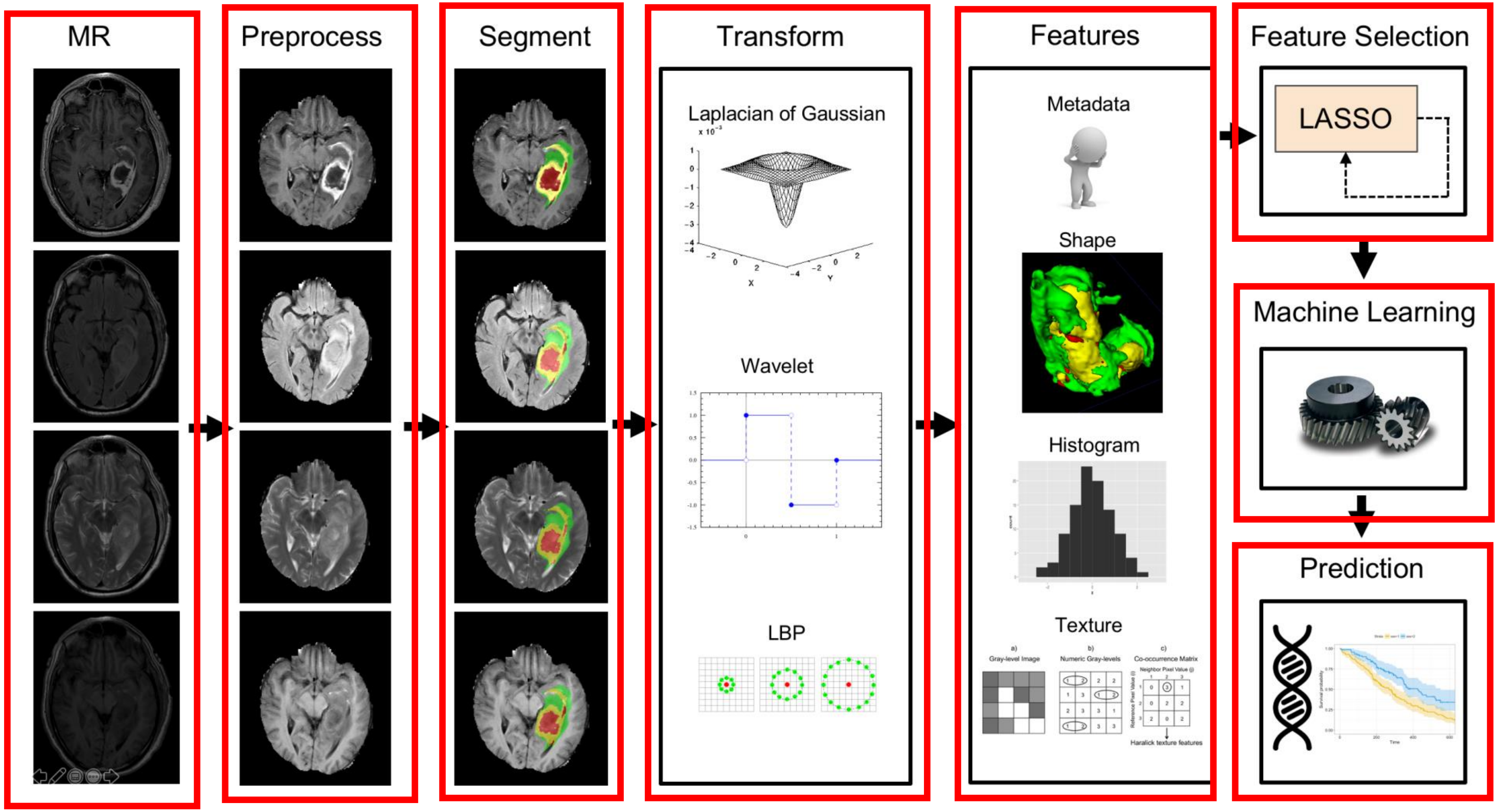
Radiomics: Images Are More than Pictures, They Are Data[1]

Prediction of IDH and TERT promoter mutations in low-grade glioma from magnetic resonance images using a convolutional neural network

Ryohei Fukuma[1,2], Takufumi Yanagisawa[1,2,3*], Manabu Kinoshita[1*], Takashi Shinozaki[4,22],

Li, Qihua, et al., *Scientific reports* (2017)
Arita, Hideyuki, et al., *Scientific reports* (2018)
Fukuma, Ryohei, et al., *Scientific reports* (2019)
Matsui, Yutaka, et al., *Journal of neuro-oncology* (2020)

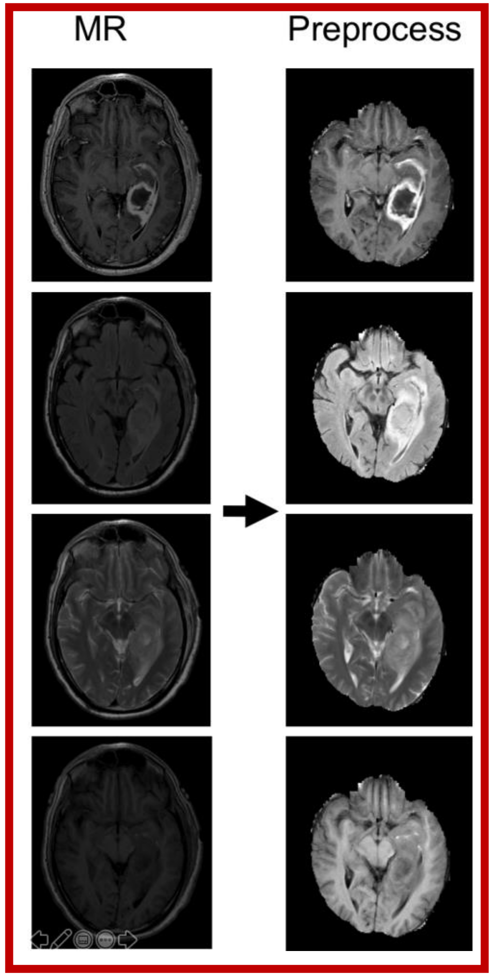
Method



Nuechterlein, et al. *Neuro-oncology advances* (2021)

Preprocessing

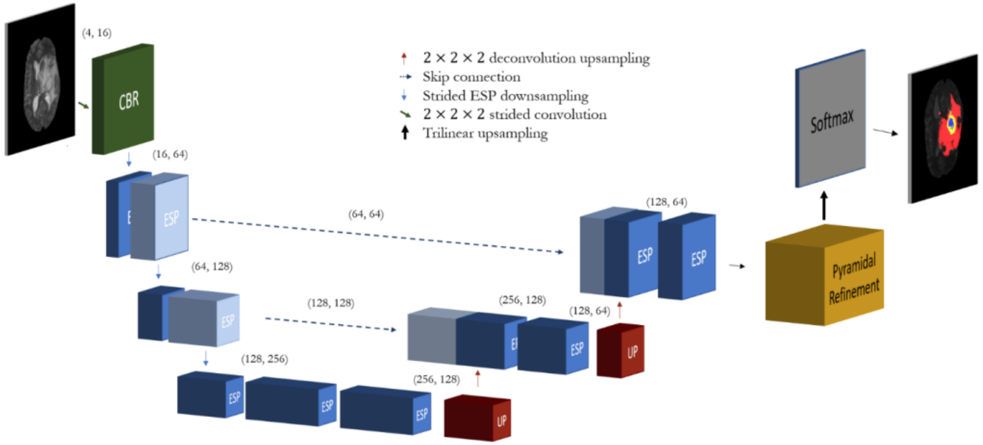
1. DICOM -> NIfTI
   - dcm2niix
2. Skull-strip
   - The Brain Extraction Tool (BET)
3. Co-register same-subject MRI sequences
   - FMRIB's Linear Image Registration Tool (FLIRT) from the FMRIB Software Library (FSL)
4. Normalize/bias correct
   - N4 Bias Field Correction

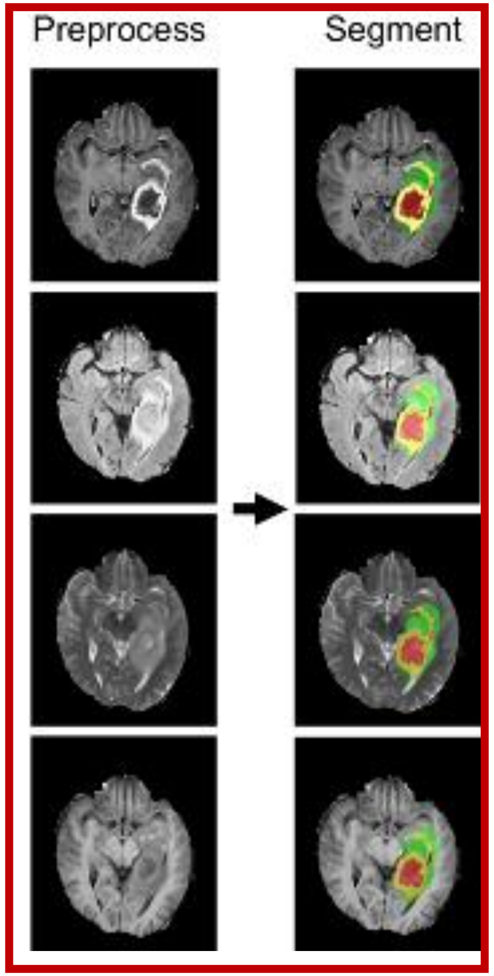


Li, et al., J Neurosci Methods (2016)
Jenkinson et al. Med Image Anal (2001)
Jenkinson et al. Neuroimage (2002)
Tustison NJ, Avants BB, et al. IEEE Trans Med Imaging (2010)

Segmentation

- U-net based architecture
- Used ESP blocks

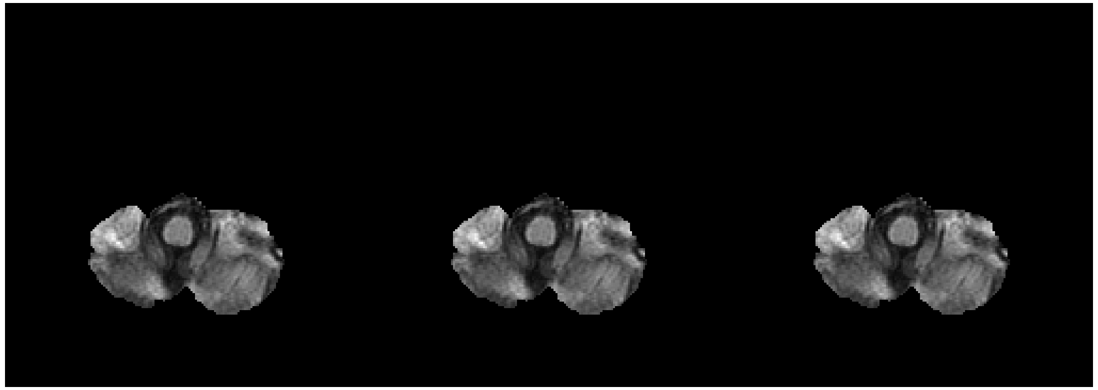




Nuechterlein, et al., *International MICCAI Brainlesion Workshop* (2018)
Mehta et al., *ECCV* (2018)
Ronneberger et al., *MICCAI* (2015)

Segmentation Results

Raw MR Prediction Ground Truth



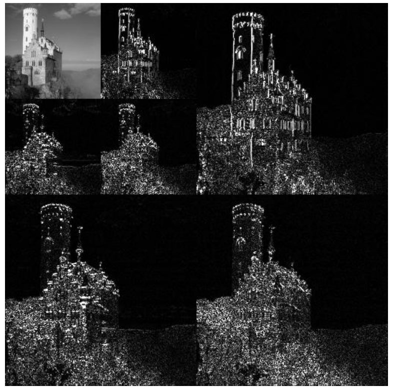
Raw MR Prediction Ground Truth

Image filters / transformations

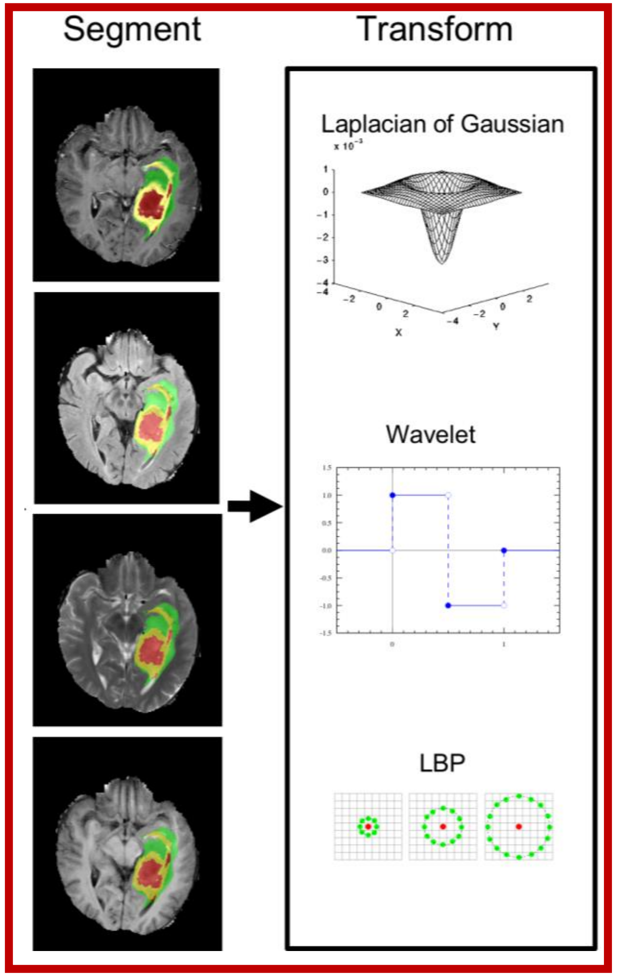
- Identity
- Laplacian of Gaussian (LoG)
- Wavelet
- Local binary patterns (LBP)
- Exponential, logarithm, square , square root

LoG

Wavelet



LBP



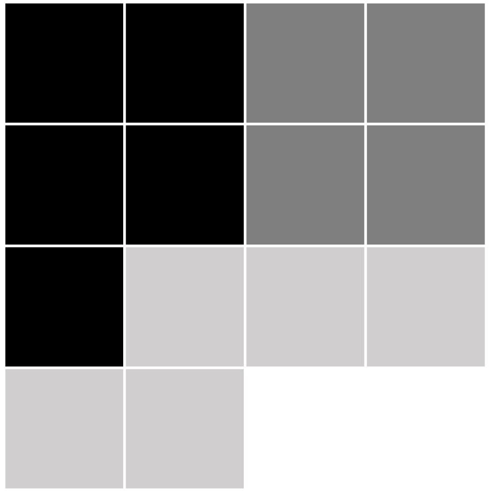
Zhang, et al. *Math. Biosci. Eng* (2020)

Texture analysis

- Image texture gives us information about the spatial arrangement of color or *intensities* in an image
- Example: Grey-level co-occurrence matrix (GLCM)

GLCM

Image



Values are image grey levels (GLs)

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Prepare GLCM matrix: values are descriptions of GLCM values

i/j	0	1	2	3
0	(0,0)	(0,1)	(0,2)	(0,3)
1	(1,0)	(1,1)	(1,2)	(1,3)
2	(2,0)	(2,1)	(2,2)	(2,3)
3	(3,0)	(3,1)	(3,2)	(3,3)

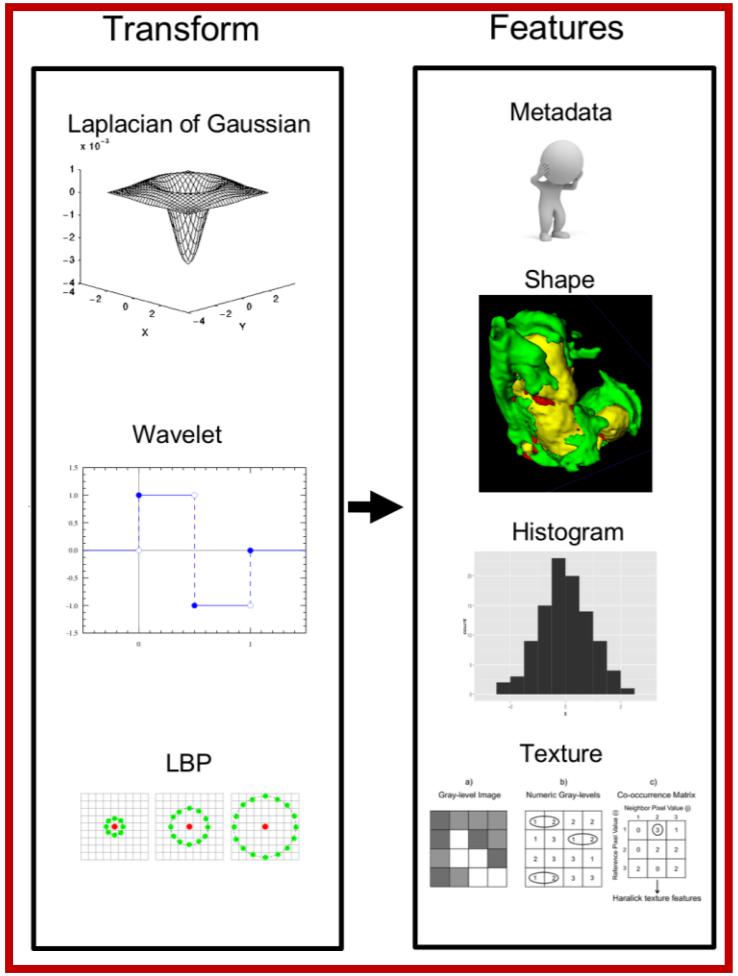
Values are counts of frequencies of the neighboring pairs of image pixel values

2	2	1	0
0	2	0	0
0	0	3	1
0	0	0	1

The diagonal elements all represent pixel pairs with no grey level difference

Feature Extraction

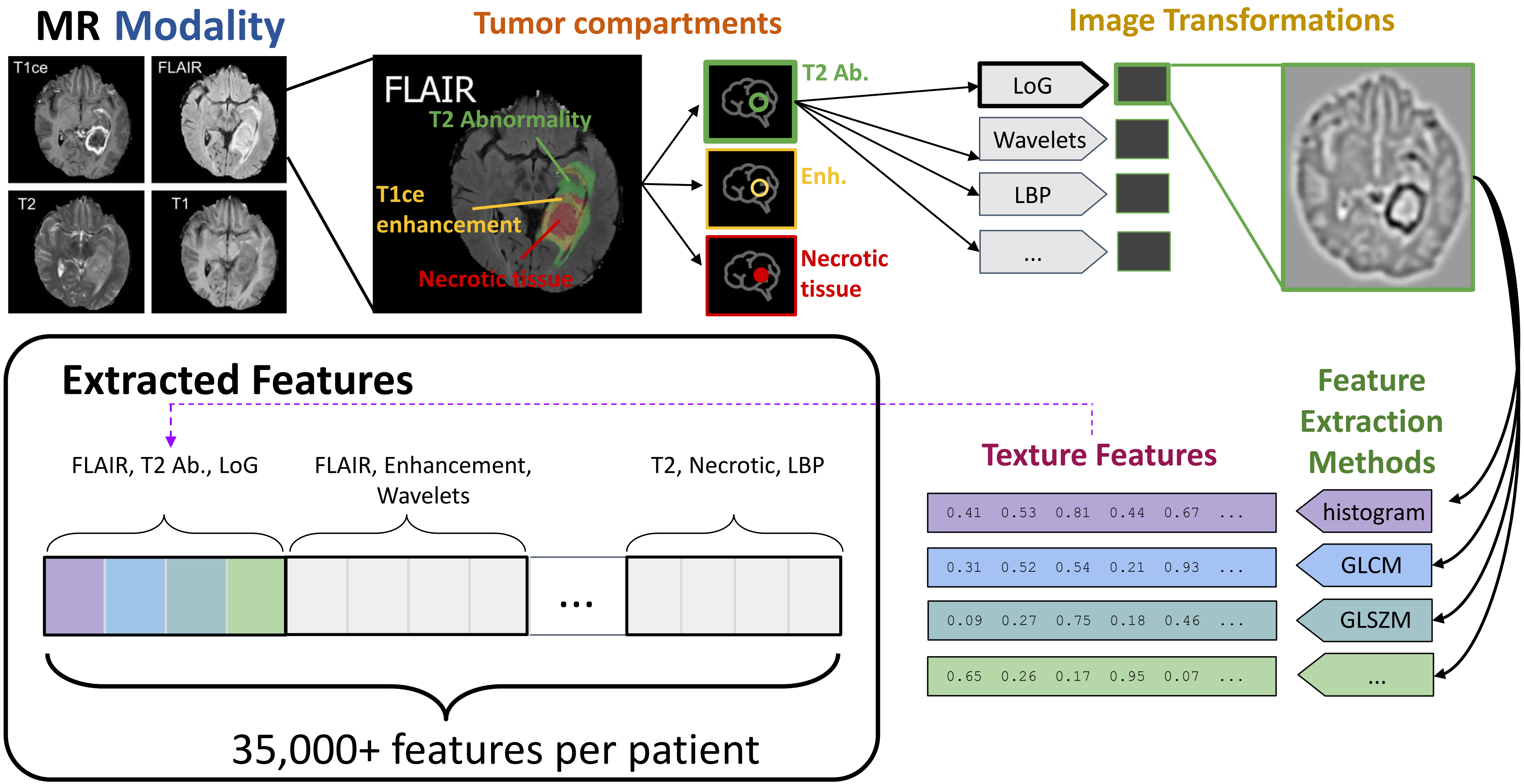
- Histogram
  - Percentile, energy, entropy, kurtosis, skewness, uniformity, etc.
- Texture
  - GLCM (Gray Level Co-occurrence Matrix)
    - Contrast, correlation, etc.
  - GLRLM (Grey-Level Run Length Matrix)
  - GLSZM (Gray Level Size Zone Matrix)
  - GLDM (Gray Level Dependence Matrix)
  - NGTDM (Neighboring Gray Tone Difference Matrix)
- Implementation
  - *pyradiomics*



Griethuysen, et al. *Cancer research* (2017)

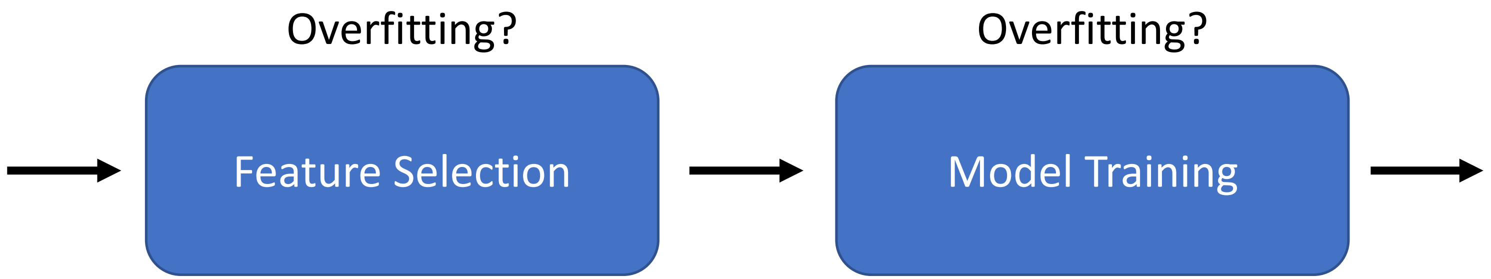
Putting it all together

Feature = $\{m, c, t, e, f\}$

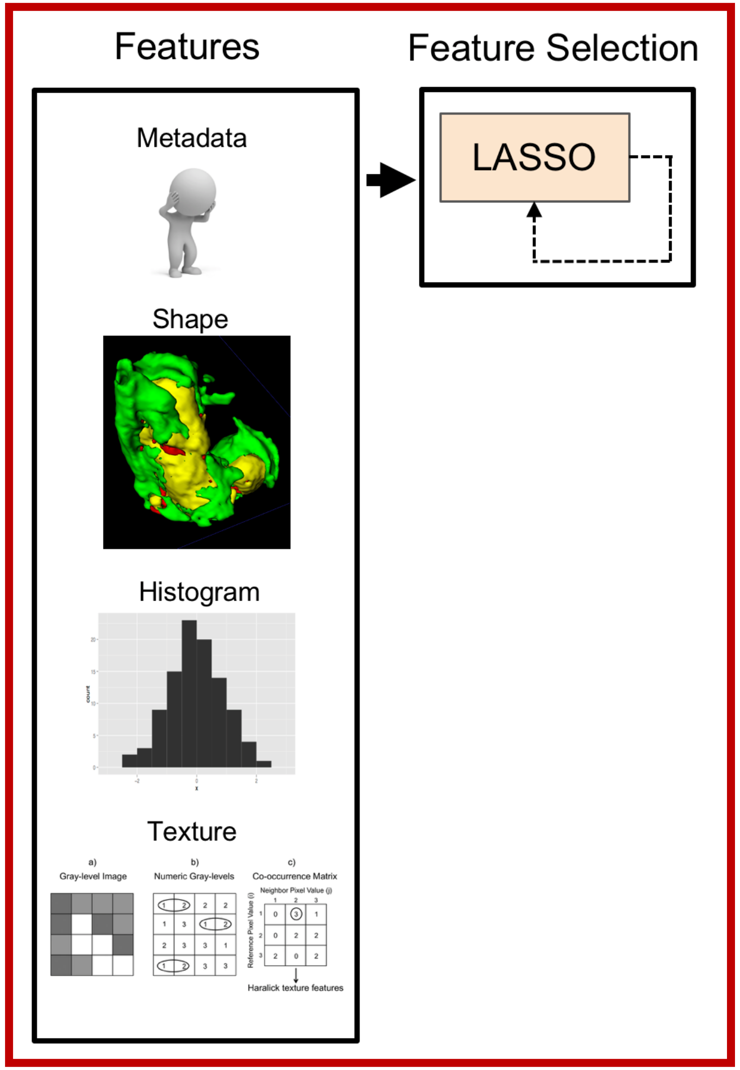


Feature selection

- Feature set is far too large for modeling a few number of samples
- Feature selection overfits
  - Recursive feature elimination
  - Variance thresholding
  - LASSO feature selection

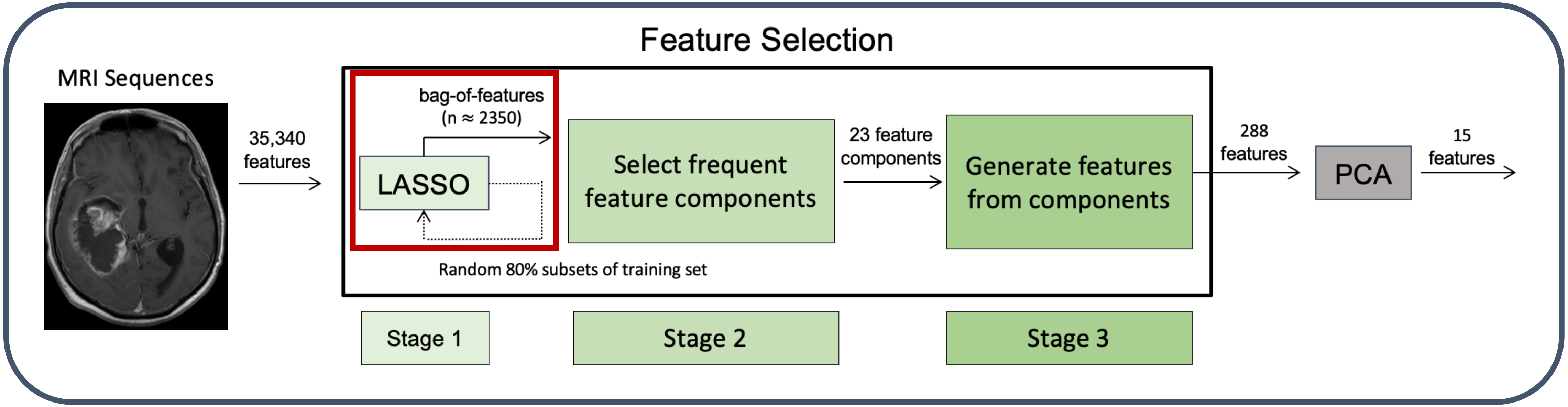


- We want to leverage the structure of our features

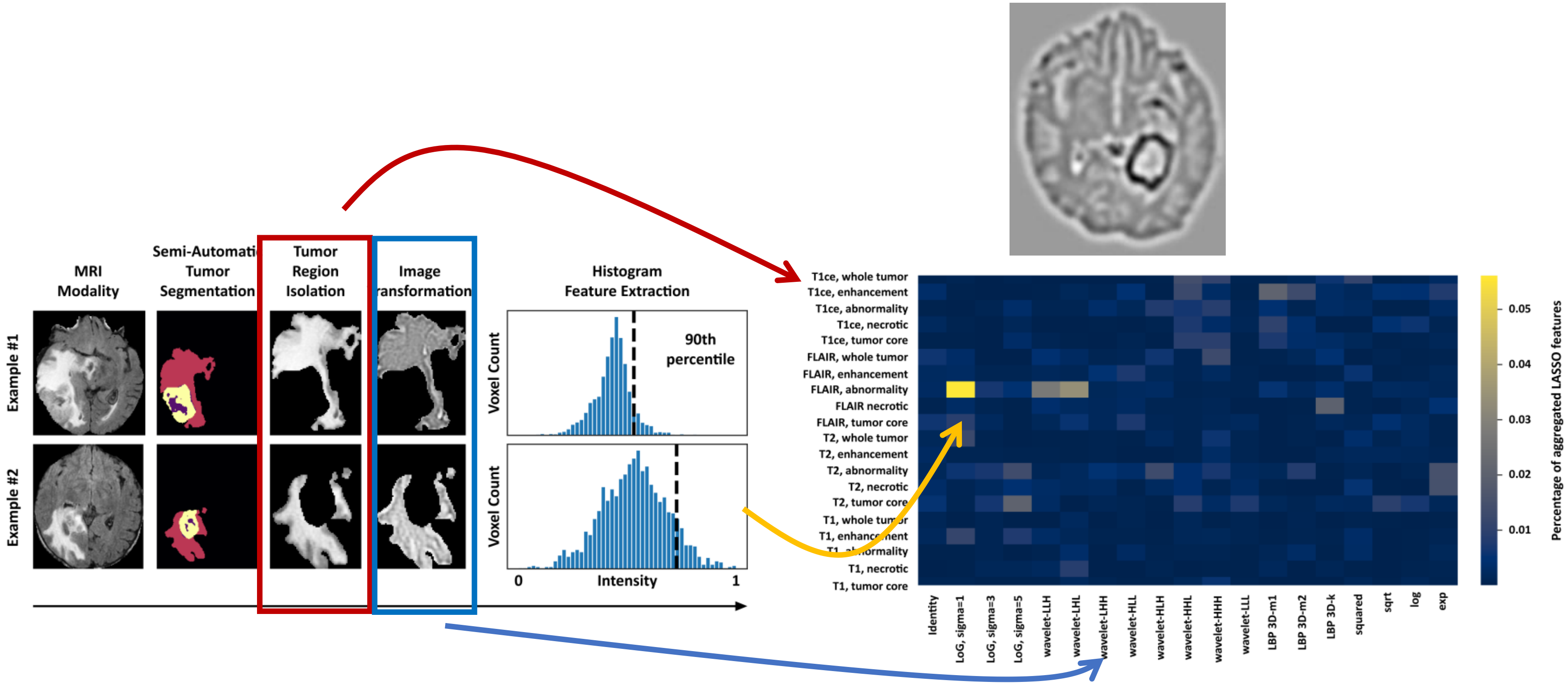


Feature Selection

- Stage 1
  - Aggregate a bag *B* of LASSO-selected features, including duplicates, by training LASSO models on random subsets of the training data

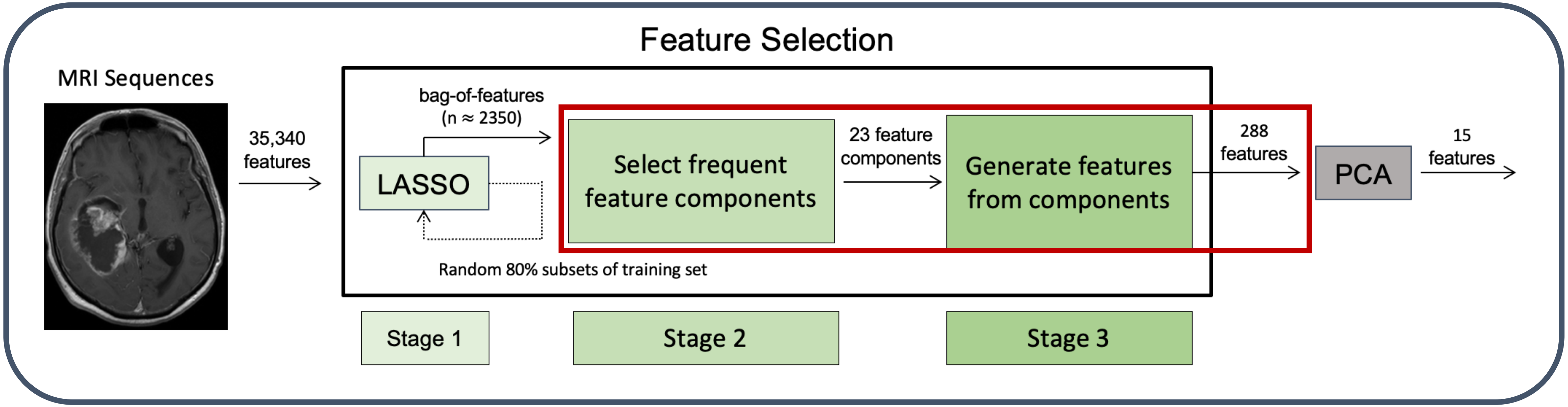


Feature Selection



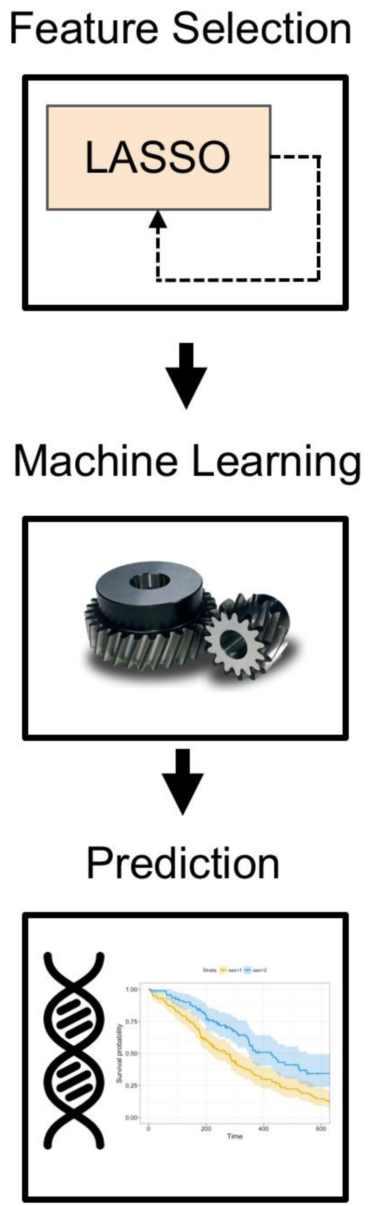
Feature Selection

- Stage 2 & 3
  - Use *B* to determine which *feature components (C)* are most relevant to the classification task
  - Generate the set of 288 features whose components were determined from the set *C*
  - Use PCA to further reduce the dimensionality of our feature set to 15

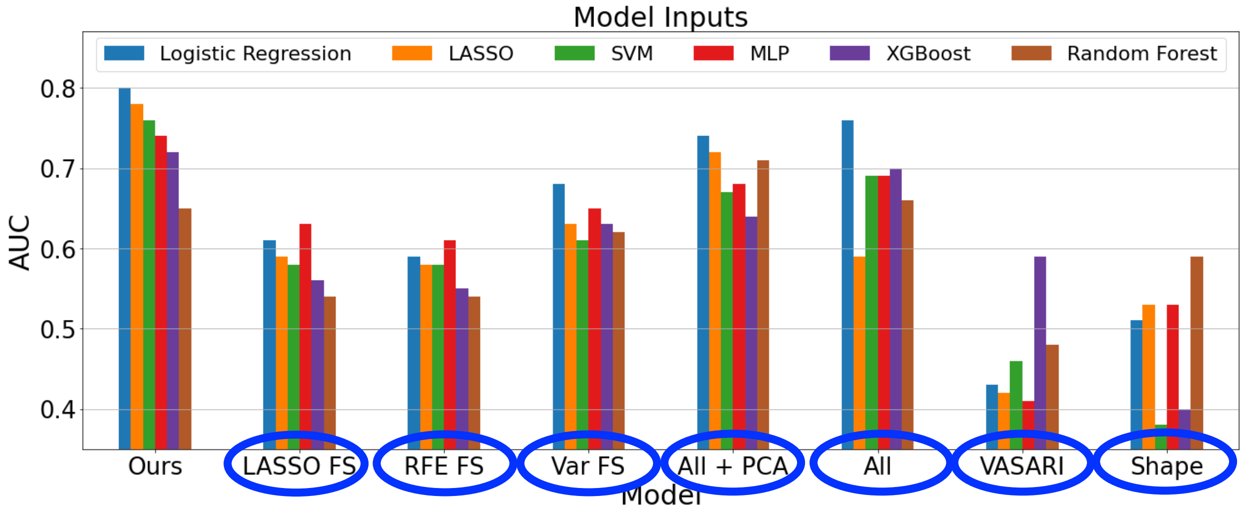


Modeling

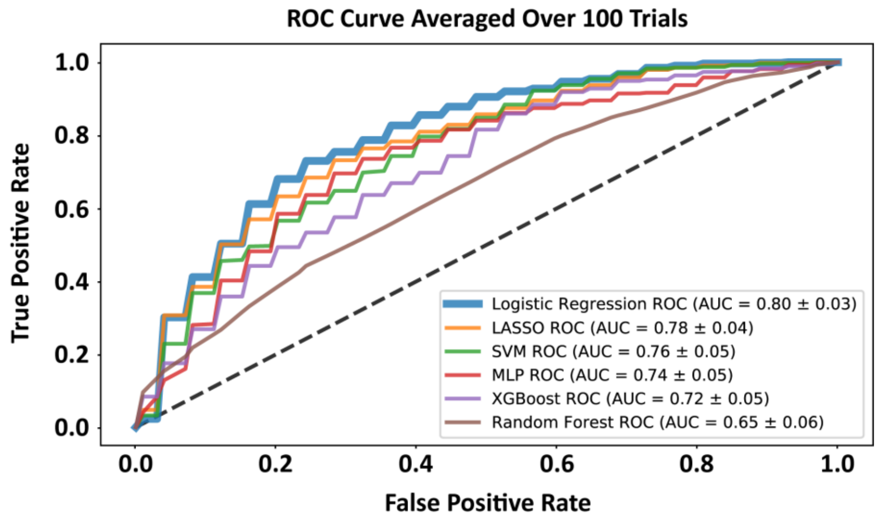
- 15 PCA Features
- Collection of small machine learning models
- Cross validation



Results

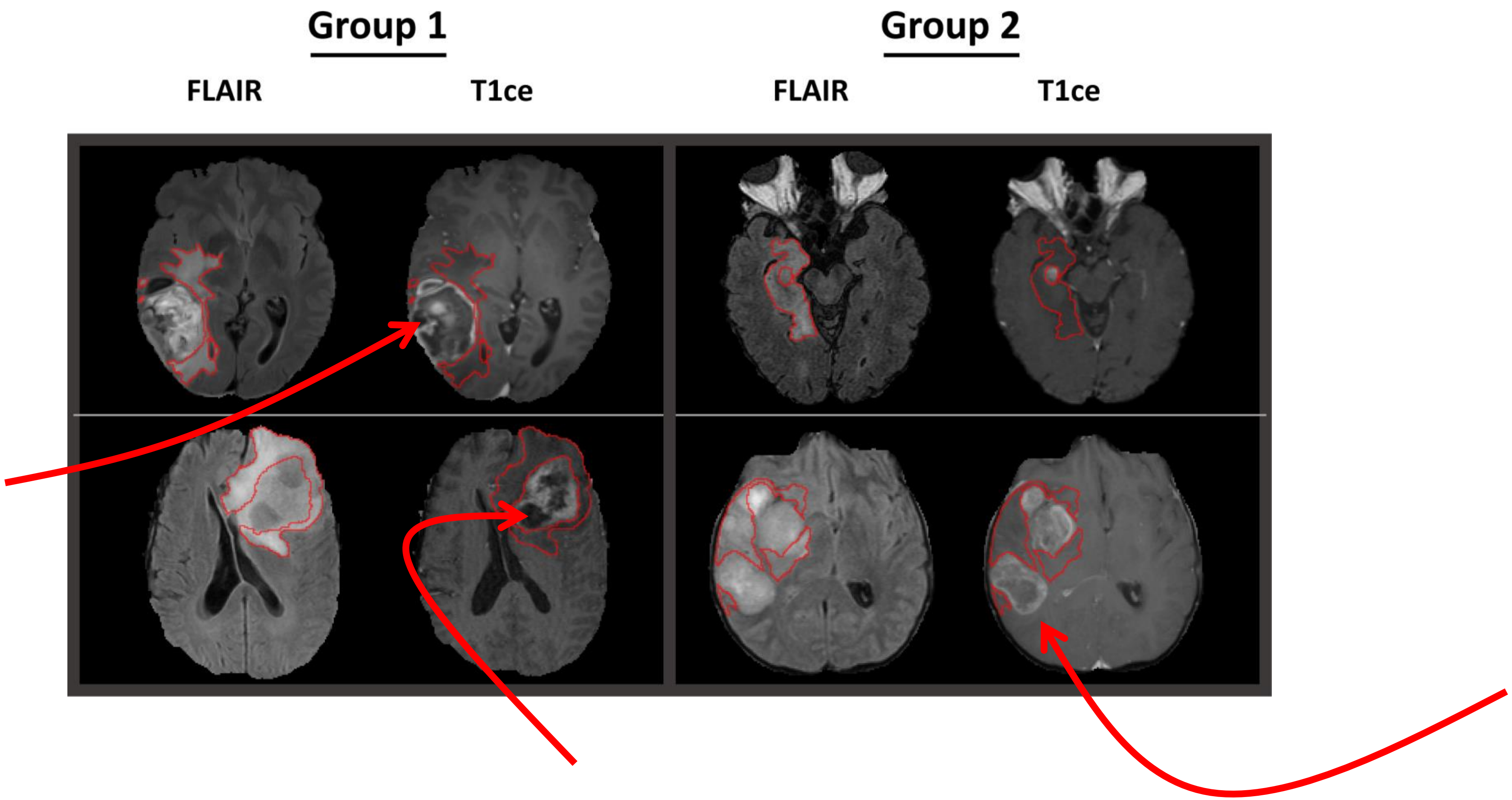


Model Inputs
Logistic Regression
LASSO
SVM
MLP
XGBoost
Random Forest
AUC
0.8
0.7
0.6
0.5
0.4
Ours
LASSO FS
RFE FS
Var FS
All + PCA
All
VASARI
Shape
Model



ROC Curve Averaged Over 100 Trials
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression ROC (AUC = 0.80 ± 0.03)
LASSO ROC (AUC = 0.78 ± 0.04)
SVM ROC (AUC = 0.76 ± 0.05)
MLP ROC (AUC = 0.74 ± 0.05)
XGBoost ROC (AUC = 0.72 ± 0.05)
Random Forest ROC (AUC = 0.65 ± 0.06)

Results



Discussion

- AUC > 0.80
- Attributes Not Selected
  - Enhancing tumor!
  - Identity transformation
- Attributes Selected
  - Laplacian of Gaussian transform (edge detector)
  - T2 Abnormality on FLAIR

Imaging Summary

- Developed a custom feature selection method that allows for the prediction of poor surviving glioblastoma patients, but leaves room for improvement
- Imaging limitations
  - Until scanner protocol is standardized, noise will interfere with model reliability
  - Low sample counts
  - Patients almost always get first resections, thus the fact that MRI is cheap and non-invasive is not necessarily an advantage in the upfront setting

Acknowledgments

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

University of Washington, Seattle, WA

- Linda Shapiro, PhD
- Beibin Li
- Mehmet Saygin Seyfioğlu
- Sachin Mehta, PhD
- Shima Nofallah

FRED HUTCH CURES START HERE™

Fred Hutch Cancer Research Center, Seattle, WA

- Eric Holland, MD, PhD
- Sonali Arora

UW Medicine LABORATORY MEDICINE & PATHOLOGY

DEPARTMENT OF RADIOLOGY UNIVERSITY of WASHINGTON

University of Washington, Seattle, WA

- PJ Cimino, MD, PhD
- Michelle Casad, PhD
- James Fink, MD
- David Haynor, MD, PhD

Funding Sources

NSF

IGERT Integrative Graduate Education and Research Traineeship

ARCS ADVANCING SCIENCE IN AMERICA