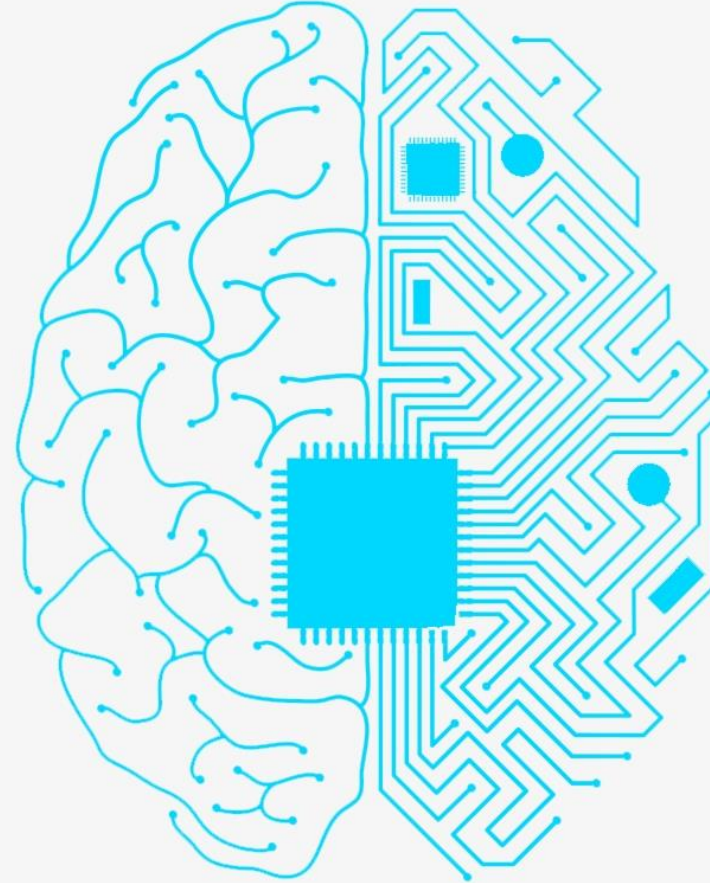


Predicting Survival of Glioblastoma Patients

Jamie Hobbs

New College of Florida

May 2020



What is a Glioblastoma?

- Fast-growing glioma(tumor arising from glial cells)
 - Astrocytes and oligodendrocytes
- "de novo" or secondary evolution
- Low prognosis: median survival of 15 months
- Risk factors:
 - Age: median of 64 y/o
 - Exposure to radiation
 - Ionizing vs electromagnetic
 - No current link to family history

Data Description

Regression & Classification Techniques

- 504 Variables
 - 4 Basic Info: Gender, Race, Age at diagnosis & Days of survival
 - 500 Genes with Somatic Mutations
 - 300 patients who have passed due to a glioblastoma(GBM)
 - NCI data portal (Kaggle), 2017
-

- Multiple Linear Regression
- Random Forest

Variable Selection

High Survivor: ≥ 377 Days

Gene	N
NF1	18
USH2A	17
RB1	16
COL6A3	16
MUC17	15
LAMA1	15
TRRAP	15
GRIN2A	15
RIMS2	15
DNAH17	15

Low Survivor: < 377 Days

Gene	N
PIK3CA	22
SPTA1	21
SYNE1	19
PIK3R1	19
OBSCN	18
RB1	18
LRP2	18
DNAH2	18
FLG2	17
PCLO	17

Multiple Linear Regression

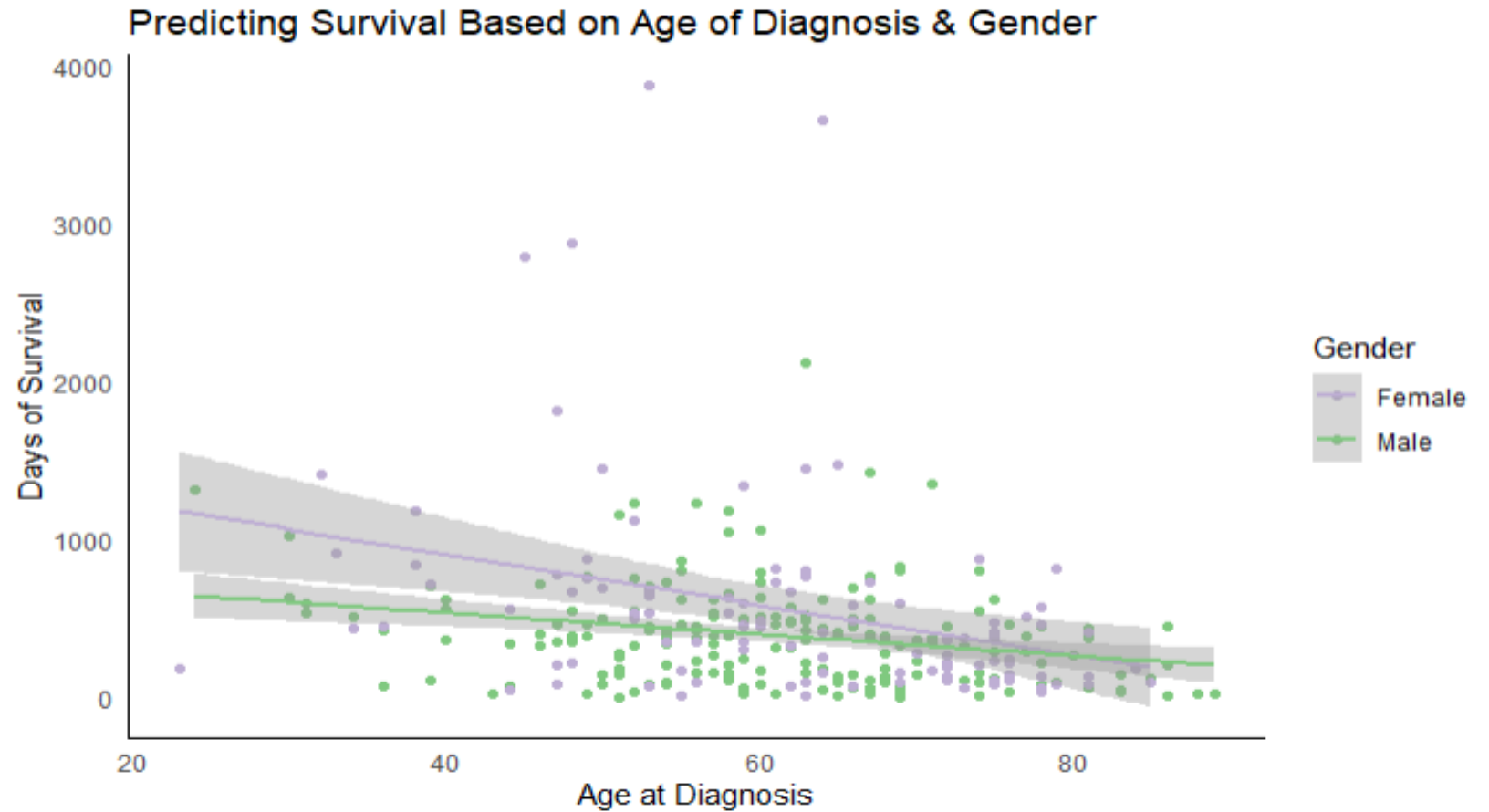
No collinearity diagnosed

Backwards AIC dropped Race from the original fit

Females average 1,201.5 days of survival

Males average 176 days less than females

As age of diagnosis increases, days of survival decrease



$$\hat{days} = 1201.525 - 10.349age - 175.99gender_{male}$$

Multiple Linear Regression Comparison

Top 10 Genes w/o Race

```
lm(formula = Days.Till.Death ~ Age.at.Diagnosis + Gender + TTN +  
PTEN + TP53 + EGFR + FLG + MUC16 + RYR2 + SPTA1 + OBSCN +  
PKHD1 + NF1, data = gbm)
```

Residuals:

Min	1Q	Median	3Q	Max
-788.85	-252.72	-64.74	142.70	3031.92

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1103.633	139.677	7.901	6.00e-14 ***
Age.at.Diagnosis	-9.726	2.080	-4.676	4.52e-06 ***
GenderMale	-177.387	55.129	-3.218	0.00144 **
TTN1	118.325	56.251	2.104	0.03629 *
PTEN1	-16.718	56.761	-0.295	0.76857
TP531	142.598	58.663	2.431	0.01568 *
EGFR1	68.321	59.291	1.152	0.25016
FLG1	-60.925	67.972	-0.896	0.37084
MUC161	-116.270	67.968	-1.711	0.08823 .
RYR21	16.862	80.287	0.210	0.83380
SPTA11	-120.943	81.833	-1.478	0.14053
OBSCN1	120.023	84.061	1.428	0.15444
PKHD11	-37.833	82.754	-0.457	0.64789
NF11	19.115	76.318	0.250	0.80240

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 448.3 on 286 degrees of freedom
Multiple R-squared: 0.1609, Adjusted R-squared: 0.1227
F-statistic: 4.217 on 13 and 286 DF, p-value: 1.926e-06

Top 11-20 Genes w/o Race

```
lm(formula = Days.Till.Death ~ Age.at.Diagnosis + Gender + NF1 +  
USH2A + RB1 + COL6A3 + MUC17 + LAMA1 + TRRAP + GRIN2A + RIMS2 +  
DNAH17 + PIK3CA + SPTA1 + SYNE1 + PIK3R1 + OBSCN + LRP2 +  
DNAH2 + FLG2 + PCLO, data = gbm)
```

Residuals:

Min	1Q	Median	3Q	Max
-1279.33	-228.00	-56.58	119.58	3040.01

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1139.976	142.777	7.984	3.75e-14 ***
Age.at.Diagnosis	-9.604	2.166	-4.434	1.33e-05 ***
GenderMale	-166.389	56.800	-2.929	0.00368 **
NF11	-2.882	80.559	-0.036	0.97149
USH2A1	-30.023	94.599	-0.317	0.75120
RB11	73.036	91.722	0.796	0.42655
COL6A31	32.017	98.816	0.324	0.74618
MUC171	-7.247	91.901	-0.079	0.93721
LAMA11	-9.878	98.200	-0.101	0.91995
TRRAP1	178.564	105.015	1.700	0.09018 .
GRIN2A1	-2.998	105.899	-0.028	0.97744
RIMS21	136.748	108.352	1.262	0.20798
DNAH171	255.562	124.846	2.047	0.04160 *
PIK3CA1	-93.915	88.819	-1.057	0.29126
SPTA11	-118.248	85.488	-1.383	0.16771
SYNE11	-20.218	96.723	-0.209	0.83458
PIK3R11	-77.011	91.809	-0.839	0.40229
OBSCN1	72.944	90.118	0.809	0.41896
LRP21	-25.583	100.237	-0.255	0.79874
DNAH21	28.750	99.380	0.289	0.77257
FLG21	6.137	100.122	0.061	0.95117
PCLO1	-107.370	98.423	-1.091	0.27626

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 455.4 on 278 degrees of freedom
Multiple R-squared: 0.1586, Adjusted R-squared: 0.09503
F-statistic: 2.495 on 21 and 278 DF, p-value: 0.0004132

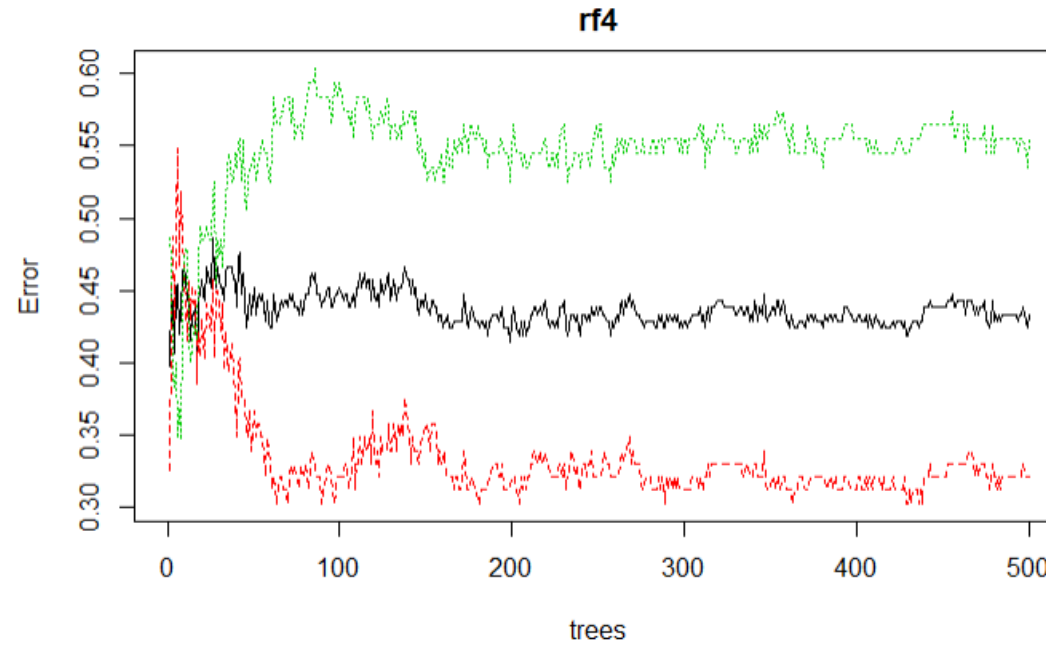
Random Forest

Initial Fit

Data: Top 11-20 Genes, Age
& Gender

```
Call:
  randomForest(formula = Days.Till.Death ~ ., data = brain.new_train,      nodesize = 1)
      Type of random forest: classification
      Number of trees: 500
No. of variables tried at each split: 4

      OOB estimate of  error rate: 40.95%
Confusion matrix:
      Above Below class.error
Above    75    34   0.3119266
Below    52    49   0.5148515
```



Random Forest

Training

Data: Top 11-20 Genes,
Age & Gender

Random Forest

210 samples
21 predictor
2 classes: 'Above', 'Below'

No pre-processing

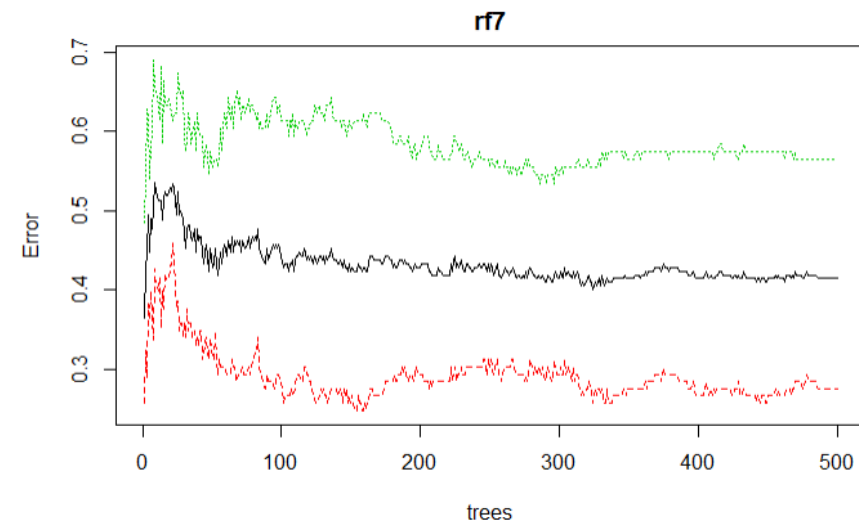
Resampling: Cross-validated (10 fold)

Summary of sample sizes: 189, 188, 189, 189, 189, 190, ...

Resampling results across tuning parameters:

mtry	Accuracy	Kappa
2	0.5627922	0.1190516
11	0.5582900	0.1128603
21	0.5439827	0.0838158

Accuracy was used to select the optimal model using the largest value.
The final value used for the model was
mtry = 2.



Call:

```
randomForest(formula = Days.Till.Death ~ ., data = brain.new_train, nodesize = 1, mtry = 2)  
Type of random forest: classification  
Number of trees: 500
```

No. of variables tried at each split: 2

OOB estimate of error rate: 41.43%

Confusion matrix:

	Above	Below	class.error
Above	79	30	0.2752294
Below	57	44	0.5643564

10-fold CV

Confusion Matrix and Statistics

	Reference	
Prediction	Above	Below
Above	40	47
Below	1	2

Accuracy : 0.4667

95% CI : (0.3607, 0.5749)

No Information Rate : 0.5444

P-Value [Acc > NIR] : 0.9435

Kappa : 0.015

Mcnemar's Test P-Value : 8.293e-11

Sensitivity : 0.97561

Specificity : 0.04082

Pos Pred Value : 0.45977

Neg Pred Value : 0.66667

Prevalence : 0.45556

Detection Rate : 0.44444

Detection Prevalence : 0.96667

Balanced Accuracy : 0.50821

'Positive' class : Above

Considerations
not taken into
account within
dataset

Location of tumor

Treatment resistance

Brain's ability to repair

Blood brain barrier

Neurotoxicity from treatments

Sources

- Dataset originally pulled from NCI data portal, Aug 2017
 - <https://www.kaggle.com/palashio/glioblastomamutations>
- Glioblastoma, ch.8 Epidemiology and Outcome of Glioblastoma
 - <https://www.ncbi.nlm.nih.gov/books/NBK470003/>

Resources

-
- Duke University Cancer Institute
 - <https://dukecancerinstitute.org/topics/glioblastoma>
 - National Cancer Institute
 - <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/studied-cancers/glioblastoma>

Questions?

