



CLASSIFICATION OF HIGH GRADE VS LOW GRADE GBM TUMORS



{ VINCENT-PIERRE BERGES, VICTOR STORCHAN, KEVIN LUO }

PROBLEM

Glioblastoma (GBM) disease is a very frequent brain cancer in adults. Classifying these tumors into high grade vs low grade tumors by noninvasive methods to improve the accuracy of targeted therapy. Challenging issues:

1. Unbalanced data set
2. Features extraction
3. Validation and clinical consistency of the results

DATA SET AND FEATURES

Provided by Professor O. Gevaert *et al*, the data set is highly unbalanced with 274 patients consisting of:

- 220 high grade tumors
- 54 low grade tumors

Tumor: 4 regions, imaged with different modalities. Here, we used 24 features.

- means of histograms
- standard deviation of histograms

MODELS USED

- Logistic Regression (LR)
- Support Vector Machine (SVM)

A FUTURE DIRECTION

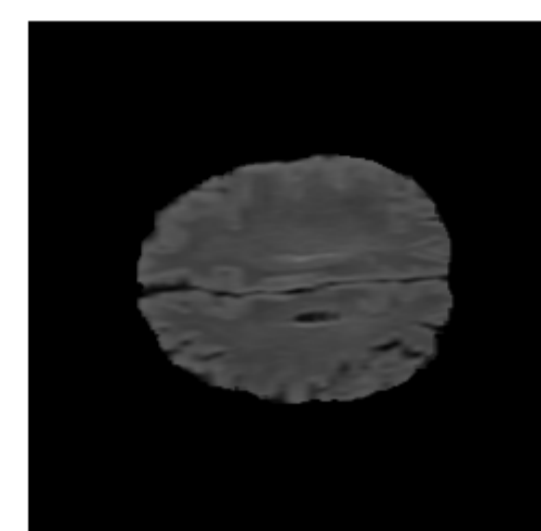
Point of view adopted in this work: Level of phenotypic information. Another interesting approach: Radiogenomics tried to establish links between image features and gene expression.

SOURCE CODE

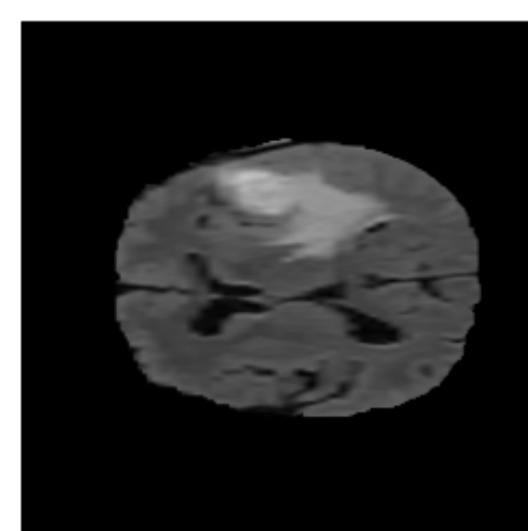
The source code is available at:

<https://github.com/cs229classif>

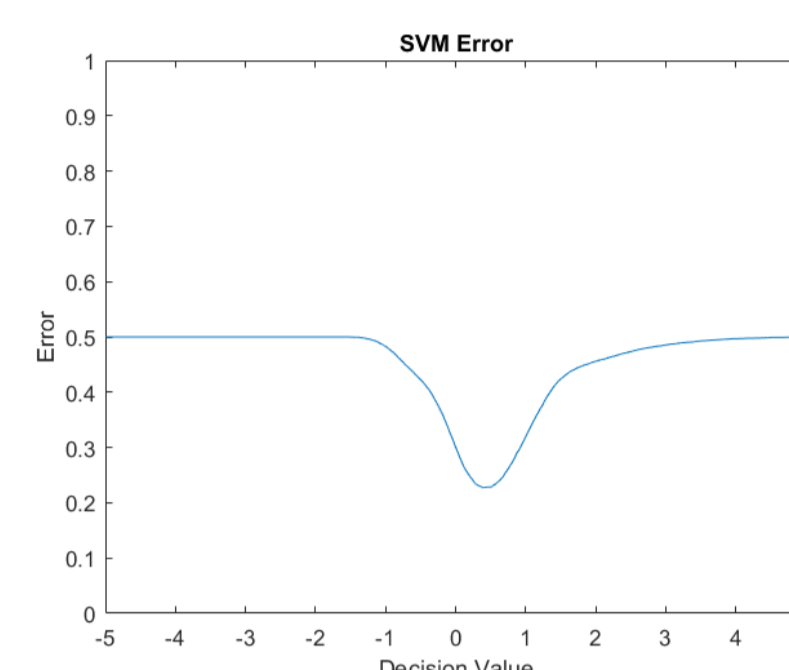
RESULTS



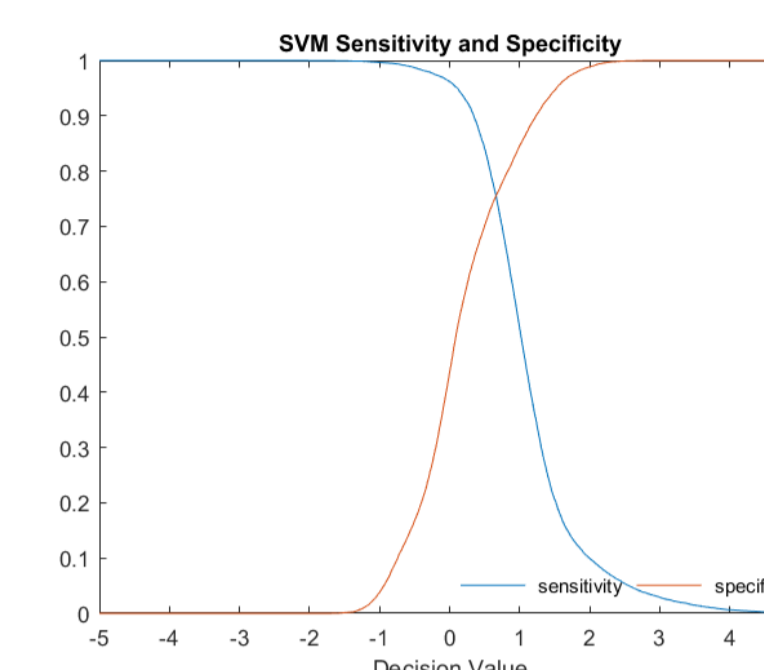
Slice no Tumor



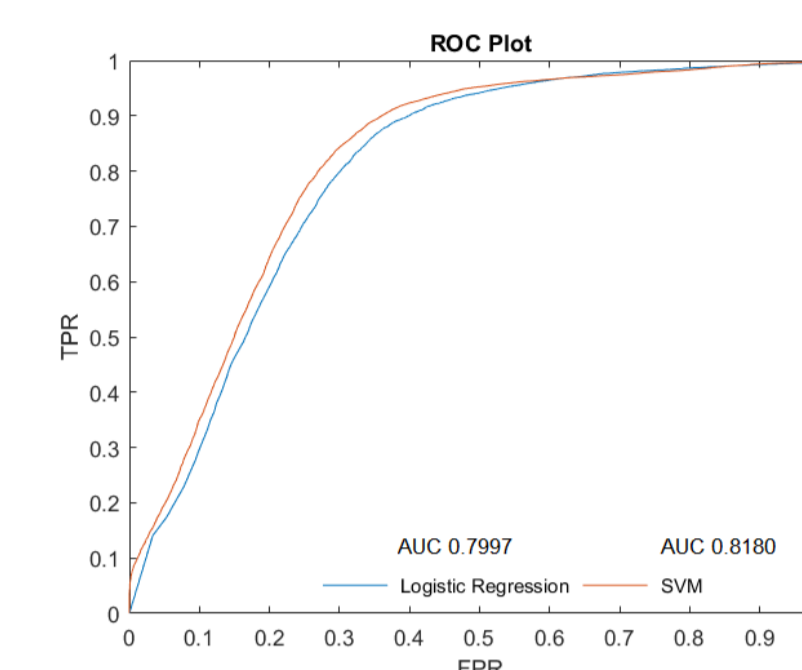
Slice with Tumor



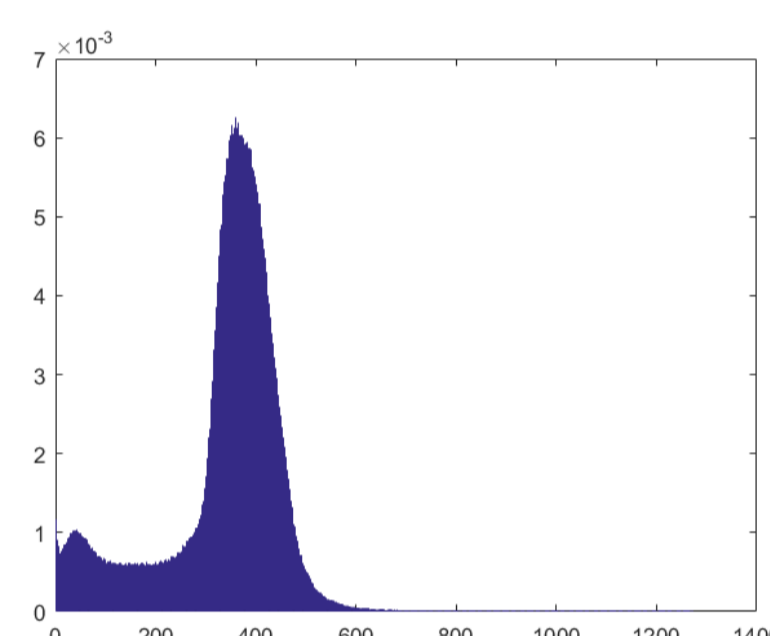
SVM Error



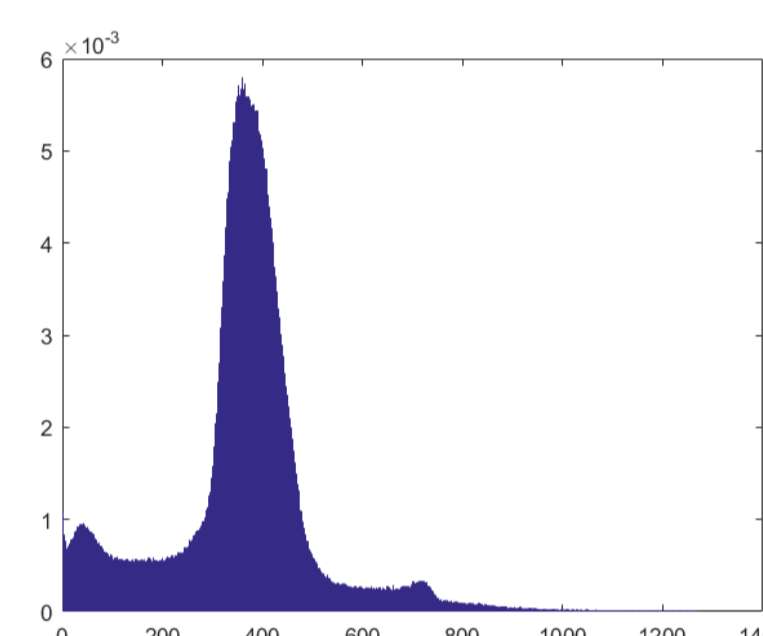
SVM Sens/Spec



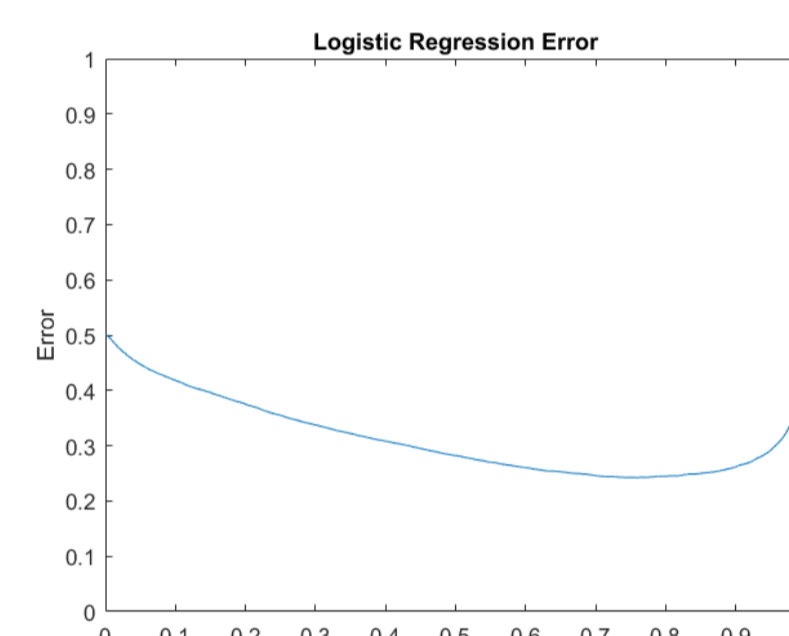
ROC



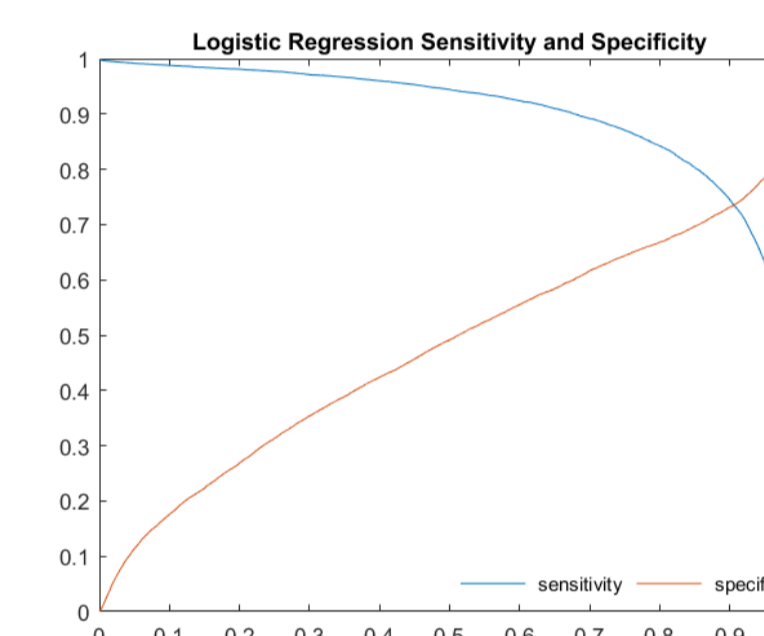
Hist. no Tumor



Hist. with Tumor



LR Error



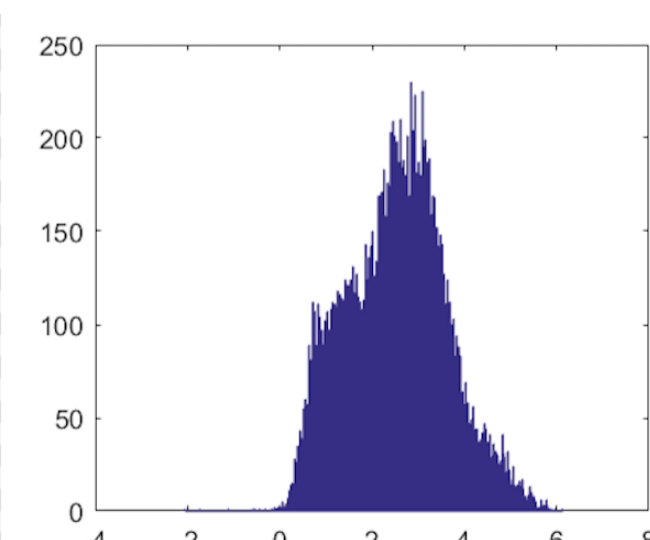
LR Sens/Spec



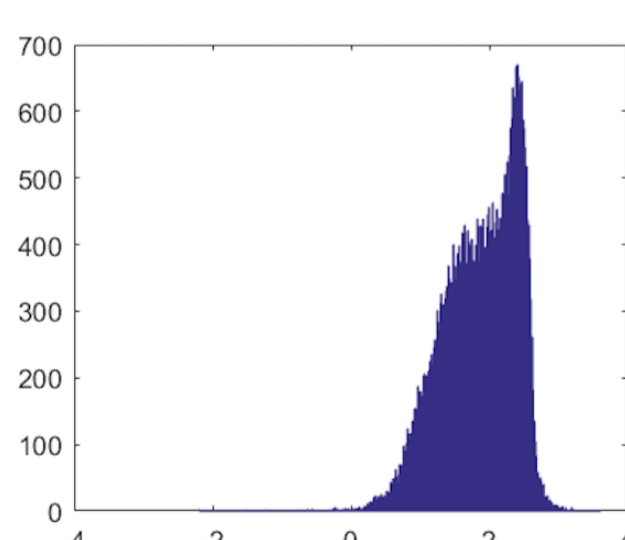
PR Plot

SVM is better than logistic regression. We are doing 30% better than naïve (guessing at random would lead to a 50% error.)

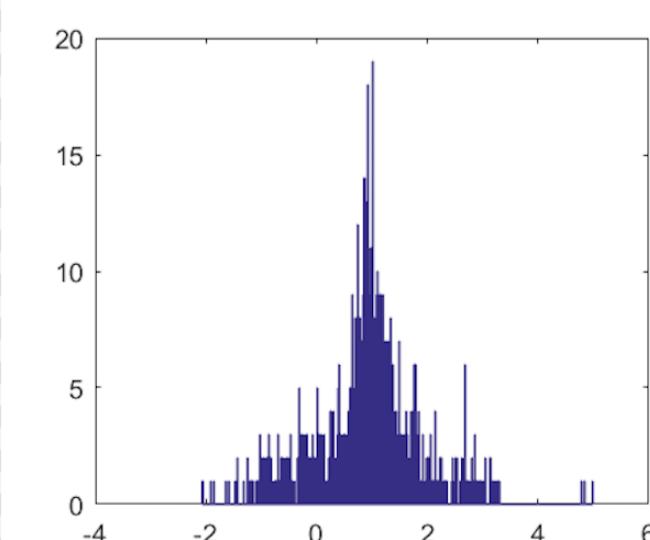
HISTOGRAMS



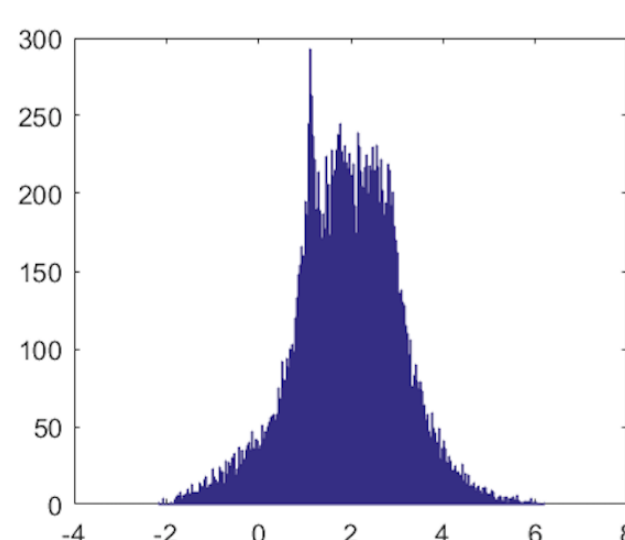
Necrosis



Edema



Non Enhancing



Enhancing

VALIDATION OF THE MODEL

Determining the low grade patients is as important as determining the high grade ones. The sensitivity and specificity curves for LR cross at a threshold value of 0.9060 with an error of 0.2649. The confusion matrix for LR is:

	real false	real true
predicted false	11.77	4.24
predicted true	4.23	11.76

The sensitivity and specificity curves for SVM cross at a decision value of 0.6600 with an error of 0.2448. The confusion matrix for SVM is:

	real false	real true
predicted false	12.07	3.91
predicted true	3.93	12.09