From Single Biomarkers to Combinations: Identifying Optimal Cutoffs for Prostate Cancer Detection

Xue Qin, Qiwei He (Supervisor), Gina D’Angelo (Mentor)
Georgetown University, MS.

Abstract

- Prostate Cancer is one of the most common cancer types and it’s essential to develop methods for early disease detection. Biomarkers are essential in helping to detect the presence of a disease at an early stage and certain biomarkers can indicate Individual’s risk of developing a disease. Compare to individual biomarker cutoff, finding biomarker combination cutoffs are essential to increase the accuracy of the diagnosis outcomes.
- This study will focus on comparing various methods for determining biomarker cutoffs, evaluating the predictive results from each, and identifying the method that shows the most potential for accurately estimating cutoffs to aid in diagnostic detection.

Introduction

- While many current approaches focus on finding a single biomarker cut-off, there is a need for methods that evaluate multiple biomarkers simultaneously to improve predictive accuracy and clinical outcomes. The traditional method for identifying the individual biomarker cutoff involved applying the ROC-based methods, like Youden Index and Point Closest to (0,1). Some other machine learning methods can find the biomarker combination cutoffs.
- We used Caret PSA Dataset from Diagnostic and Biomarkers Statistical (DABS) Center. two biomarkers were used: Total prostate-specific antigen (**TPSA**) and free-to-total PSA (**FPSA**)

Methodology

Youden Index

The optimal cutoff is the point where the Youden Index is maximized

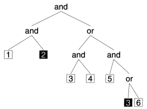
$$J = Sensitivity + specificity - 1 = recall1 + recall0 - 1$$

Point closest to (0,1)

For each cutoff value, the model calculates the Euclidean distance from the point on the ROC curve to the ideal point (0,1)

$$Distance = (1-Sensitivity)^2 + (0-(1-Specificity))^2$$

Logic Regression



Searches for combinations of conditions using logical operators (AND, OR, NOT) to predict the outcome.

Logistic Regression

The log-odds of the outcome are modeled as a linear combination of the biomarkers

$$log\ odds = \beta 0 + \beta 1 \times biomarker^1 + \beta 2 \times biomarker^2 + \ldots$$

Decision Tree

Works by recursively splitting the data into subsets based on the values of the biomarkers, and the splitting process continues until it creates a tree structure where each branch leads to a specific outcome classification.

Results Analysis - Individual Biomarker

Youden Index & Point- Closest to (0,1)

Biomarker	Index	Method	Optimal Cutpoint	Index Results	Accuracy	Sensitivity	Specificity	AUC
fpsa	Youden Index	maximize_metric	0.37	0.41	66.76%	81.66%	59.25%	77.35%
	Distance to (0,1)	minimize_metric	0.47	0.42	71.01%	68.56%	72.25%	77.35%
tpsa	Youden Index	maximize_metric	2.46	0.53	77.01%	75.11%	77.97%	83.75%
	Distance to (0,1)	minimize_metric	2.46	0.33	77.01%	75.11%	77.97%	83.75%

Table 2: *Youden Index and Point-Closest-to (0,1) Method Comparison*

- For biomarker tpsa, accuracy is identical (77.01%) under both methods.
- for biomarker fpsa, the Point-Closest-to-(0,1) method achieved higher accuracy (71.01%) compared to the Youden Index Method (66.76%)
-the tpsa would be a better predictor than fpsa while individually predicting the diagnosis outcome.

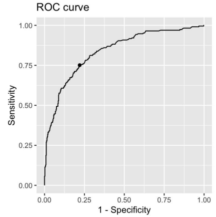


Figure 1: *Youden Index ROC Curve*

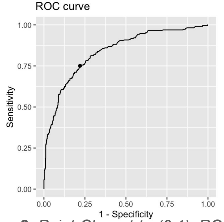


Figure 2: *Point-Closest-to-(0,1) ROC Curve*

Results Analysis - Biomarker Combination

Logic Regression

- Around **82%** overall accuracy score
- ‘AND’ Condition: fpsa > 0 ∧ tpsa > 3.61
- ‘OR’ Condition: fpsa > 1.10 ∨ tpsa >3.61
- ‘AND, OR Combination’ Condition: ((tpsa ≥ 3.21 ∧ tpsa < 6.27) ∧ fpsa > 0) ∨ (tpsa > 6.27)

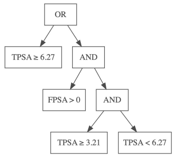


Figure 3: *Logic Tree*

Logistic Regression

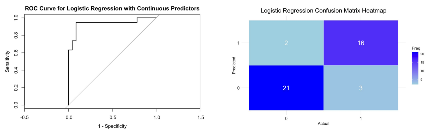


Figure 4: *Logistic Regression Heatmap and ROC Curve*

- Around **88.09%** overall accuracy score
- high sensitivity of **91.3%** and specificity of **89.5%**
- high AUC score from the ROC curve

Decision Tree

- Around **83.3%** accuracy score and High AUC value of **89.9%**
- nodes with tpsa < 7.2 and fpsa >= 0.53 played more significant roles in classifying observations into positive classes.

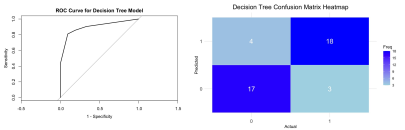


Figure 5: *Logistic Regression Heatmap and ROC Curve*

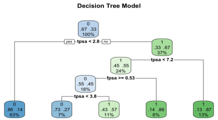


Figure 6: *Decision Tree*

Conclusion

- Biomarker TPSA plays more important role to predict outcome.
- Biomarker combination cutoffs predict more accurate outcome than individual biomarker cutoff.
- Logistic regression is the best model for predicting the outcome based on the two biomarkers, TPSA & FPSA.

GEORGETOWN UNIVERSITY