CoSIBS – the colorado summer institute in biostatistics

colorado school of public health | UNIVERSITY OF COLORADO · COLORADO STATE UNIVERSITY · UNIVERSITY OF NORTHERN COLORADO

Machine learning/artificial intelligence (ML/AI) applications in predicting olfactory outcomes for surgically treated CRS patients

Felipe Delgado[a], Van Hovenga[a], Noah Kim[a], Lydia Lo[a], Ingrid Shu[a], Krithika Suresh PhD[a], Vijay Ramakrishnan MD[b]

[a]Colorado School of Public Health, [b]University of Colorado Department of Otolaryngology

Background

Chronic rhinosinusitis (CRS) is a highly prevalent disease that is defined by symptoms of smell loss, congestion, pressure, and drainage. CRS inflicts a severe quality-of-life (QOL) impairment due to local sinonasal symptoms, its compounding role in other respiratory diseases, and general health effects (e.g., sleep dysfunction, depression, productivity loss). An individual patient's decision to pursue aggressive treatments, such as surgery, is a complex one that balances risks and alternatives, disease burden, and expectations for improvement.

We aimed to develop a clinical prediction model for predicting the improvement in the **S**ino-**n**asal **O**utcome **T**est (SNOT-22) [1], a QOL measure, following surgery. The goal is for clinicians to be able to use this prediction model to help patients make an informed personalized decision about whether to undergo surgery.

Methods

The data source is n=387 adults with CRS treated with surgery who were enrolled into a prospective, multicenter, observational CRS treatment outcomes study performed from 2011-2016 and underwent pre- and post-treatment clinical assessments.

Our primary outcome is change in SNOT22 from baseline after surgery. We defined this as both continuous (difference in SNOT22 score) and binary (change of ≤-9 vs. >-9). A higher SNOT22 score (0-110 range) indicates worse sinonasal QOL, so a decrease in score indicates improvement.

To assess the performance of the binary prediction models, we used "Area Under the ROC Curve" (AUC), sensitivity (true positive rate), and specificity (true negative rate) [2]. For the continuous prediction models, we looked at residual mean square error (RMSE), mean absolute error (MAE), and R-squared [2]. The best-fitting model was selected based on highest AUC (binary outcome) and lowest RMSE (continuous outcome). Missing values for predictors were imputed using random forest (missForest function in R) [3].

We performed internal validation using repeated k-fold cross validation, which repeats n times the process of k-fold cross validation. Cross validation is a resampling method that divides the data into k folds and fits a chosen statistical method multiple times by using different subsets of given data. We chose this approach due to the small sample size in which apparent performance estimates can be optimistic [4]. We chose the standard k=10 folds and n=10 repeats for computational expedience. Performance metrics were then averaged across 100 models.

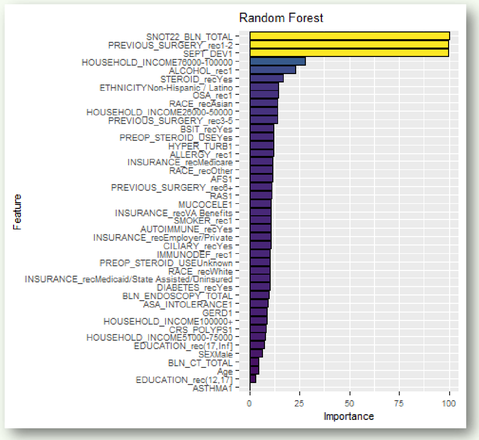
Down sampling was used as a cross validation control for the binary models to adjust for imbalanced classification (n = 321, 83% for "improvement" versus n = 66, 17% for "no improvement").

Model building and performance were conducted in R software using the caret package [5].

Models

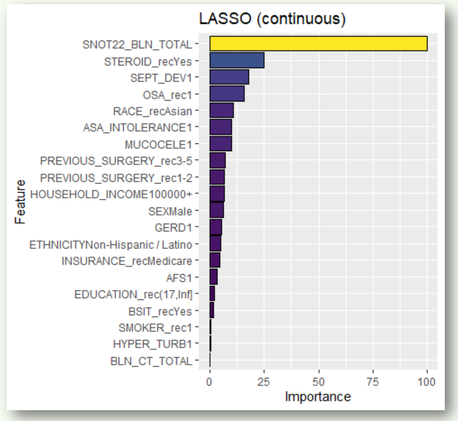
Method	Description
Adaptive Boosting (AdaBoost)	Uses a linear combination of shallow decision trees to perform classification. Each tree is adapted to emphasize performance on subsets of the data that were misclassified by previous trees through dynamic error weighting. Hyperparameters: Number of Trees = 20
Conditional Random Forest	Utilizes random forest to apply conditional inference trees as base learners
Multiple Linear Regression	Linear regression model with all predictors, no variable selection performed
Multivariate Adaptive Regression Splines (MARS)	Applied similarly to traditional logistic regression but uses stepwise methods for variable selection and allows for nonlinear and interaction effects.
Least Absolute Shrinkage and Selection Operator (LASSO) Regression, Ridge Regression	Penalizes multiple linear regression (MLR) to reduce variable coefficients. Additionally, LASSO does variable selection. Hyperparameter: lambda
Partial Least Squares (PLS)	Essentially a "supervised" alternative to PCA since it uses information about the outcome to make the transformations Hyperparameter: ncomp (# of components)
Principal Components Analysis (PCA)	Transforms predictors into a smaller number of components to effectively reduce dimension Hyperparameter: ncomp (# of components)
Stepwise Logistic Regression	Selects important variables in a stepwise fashion to build the best performing logistic regression model
Support Vector Machine/Regression (SVM)	Creates a line or hyperplane to separate data into classes

Variable Importance



Binary Outcome

Random Forest



Continuous Outcome

LASSO

****variables with 0 importance omitted***

Results

The best prediction models were Conditional Random Forest for the binary outcome and LASSO regression for the continuous outcome.

Binary Outcome

Model	AUC	Sensitivity	Specificity
Conditional Random Forest	0.723	0.648	0.682
Ridge Regression	0.720	0.210	0.943
LASSO	0.716	0.250	0.922
Logistic Regression	0.703	0.254	0.915
MARS	0.703	0.633	0.655
Support Vector Classification	0.684	0.642	0.609
AdaBoost	0.674	0.555	0.689
Stepwise Regression	0.673	0.630	0.642

Continuous Outcome

Model	RMSE	MAE	R^2
LASSO	16.85	13.25	0.37
Ridge	17.14	13.48	0.34
Partial Least Squares	17.35	13.53	0.34
Linear Regression	17.45	13.74	0.33
MARS	17.68	13.79	0.32
Support Vector Regression	17.74	13.93	0.30
Principal Components Regression	17.96	14.09	0.29

Discussion

The models we fit do not have great predictive performance. The collection and inclusion of additional variables may help build better predictive models and be elucidating for our research question.

Due to the time-limited nature of our project, we did not perform sensitivity analyses or mixed modeling to fill in missing outcomes, which future research may benefit from. Future steps can also include exploring predictive performance in particular subgroups.

References and Acknowledgment

Thank you to CoSIBS for this opportunity and Dr. Vijay Ramakrishnan for providing the project concept and data.

[1] Kennedy JL, et al. (2013) *Ann Allergy Asthma Immunol*. 111(4):246-251.e2. doi:10.1016/j.anai.2013.06.033
[2] Steyerberg EW, et al. (2010) *Epidemiology*. 21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2
[3] Stekhoven, DJ, et al. (2012) *Bioinformatics*. *28*(1), 112-118.
[4] Steyerberg EW, Harrell FE Jr. (2016) *J Clin Epidemiol*. 69:245-247. doi:10.1016/j.jclinepi.2015.04.005
[5] Kuhn, M. (2008). *Journal of statistical software*. *28*(1), 1-26.