Analyzing Supervised Machine Learning Models for the Prediction of Endoleaks Following Endovascular Aortic Aneurysm Repair

LOMA LINDA UNIVERSITY HEALTH

Sharon C. Kiang, MD[1], Daniel Roh, B.S.[2], Andrew Cabrera, B.S.[2], Ryan Mezan, MD[2], Sina Asaadi, MD[2], Zion Shih, B.S.[1], Jeffrey J. Siracuse, MD[3], Caitlin W. Hicks, MD[4], Roger T. Tomihama, MD[2].

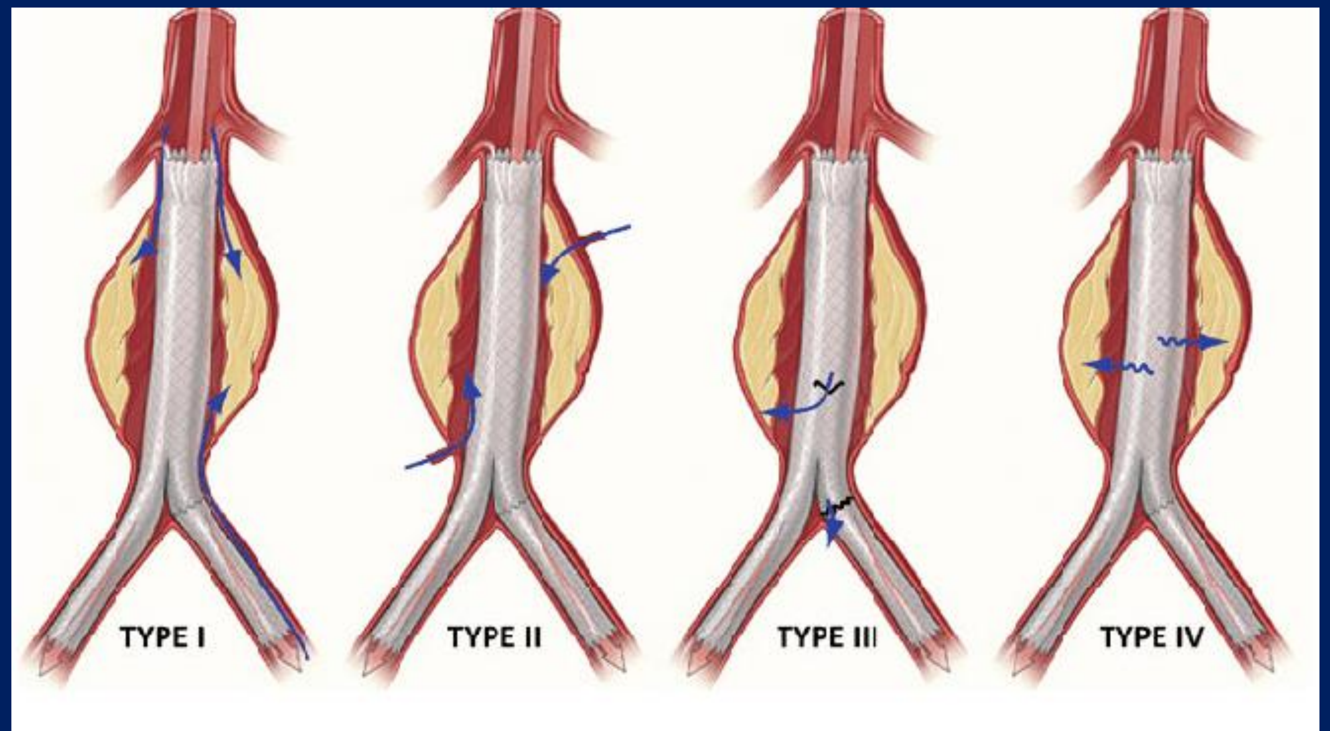
[1]Loma Linda Veterans Healthcare System, Loma Linda, CA, USA, [2]Loma Linda University Medical Center, Loma Linda, CA, USA, [3]Boston University, Chobanian and Avedisian School of Medicine, Boston, MA, USA, [4]Johns Hopkins University School of Medicine, Baltimore, MD, USA.

Introduction

- An abdominal aortic aneurysm (AAA) is a dilation in the abdominal aorta due to a weakening in the aortic wall
- Untreated AAA have an increased rupture risk when untreated
- After AAA repair, there are patients that require surveillance for a complication called an endoleak
  - Where blood leaks around the graft into the aneurysm
- AAA are often found incidentally on imaging exams.
- Surveillance is an important part of monitoring disease progression.
- Automated detection may reduce errors in diagnosis and surveillance.
- Purpose: Using supervised machine learning algorithms, we aim to predict the occurrence of endoleaks after endovascular abdominal aortic aneurysm repair.

Figure 1

Figure 1. Types of endoleaks post-endovascular repair of abdominal aortic aneurysms.



Computational Analysis of Displacement Forces Acting on Endografts Used to Treat Aortic Aneurysms - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Schematic-representation-of-the-endoleak-types-following-AAA-endovascular-repair_fig4_226789359 [accessed 6 Jan, 2024]

Methods

- Retrospective study analyzing the demographic and clinical data from 72 patients with endoleaks and 72 patients propensity matched controls
- 7 machine learning models were used to evaluate optimal model for analyzing this data:
  - Logistic Regression
  - Random Forest Classifier
  - Gradient Boosting Classifier
  - Decision Tree
  - Support Vector Machine
  - Gaussian Naïve Bayes
  - Multi-Layer Perceptron Model
- 64 clinical and demographic variables were used as the feature matrix
- Permutation Feature Importance:
  - ELI5 was conducted to calculate significance of all estimators of a target variable in a dataset.
  - Top 10 features were used to create the ROC for each model
- Sensitivities, specificities, confusion matrices, and AUC were calculated for each model

Results

- Using PFI, we identified and isolated 10 clinical data variables that significantly impacted the prediction of the models to improve the AUC of the models
  - ARB
  - ASA (Aspirin)
  - Statin
  - Insulin
  - Valvular Heart Disease
  - COPD
  - Chronic Kidney Disease
  - History of Smoking
  - Carotid Artery Stenting
  - Percutaneous Coronary Intervention
- Average AUC improvement = 0.123.
- The Random Forest Classifier was the best-performing model
  - AUC improvement from 0.53 to 0.67
- The Support Vector Machine Classifier was the least accurate model
  - AUC decrease from 0.49 to 0.35

Table 1

Model	Initial AUC	Sensitivity (%)	Specificity (%)	Accuracy (%)	Post PFI AUC
Logistic Regression	0.502	40	73.7	54.5	0.568
Random Forest Classifier	0.531	48	73.7	59.1	0.672
Gradient Boosting Classifier	0.559	68	36.8	54.5	0.563
Decision Tree	0.454	48	68.4	56.8	0.649
Support Vector Machine	0.487	52	73.7	61.4	0.354
Gaussian Naive Bayes	0.503	16	89.5	47.7	0.604
Multi-Layer Perceptron	0.58	20	84.2	47.7	0.596

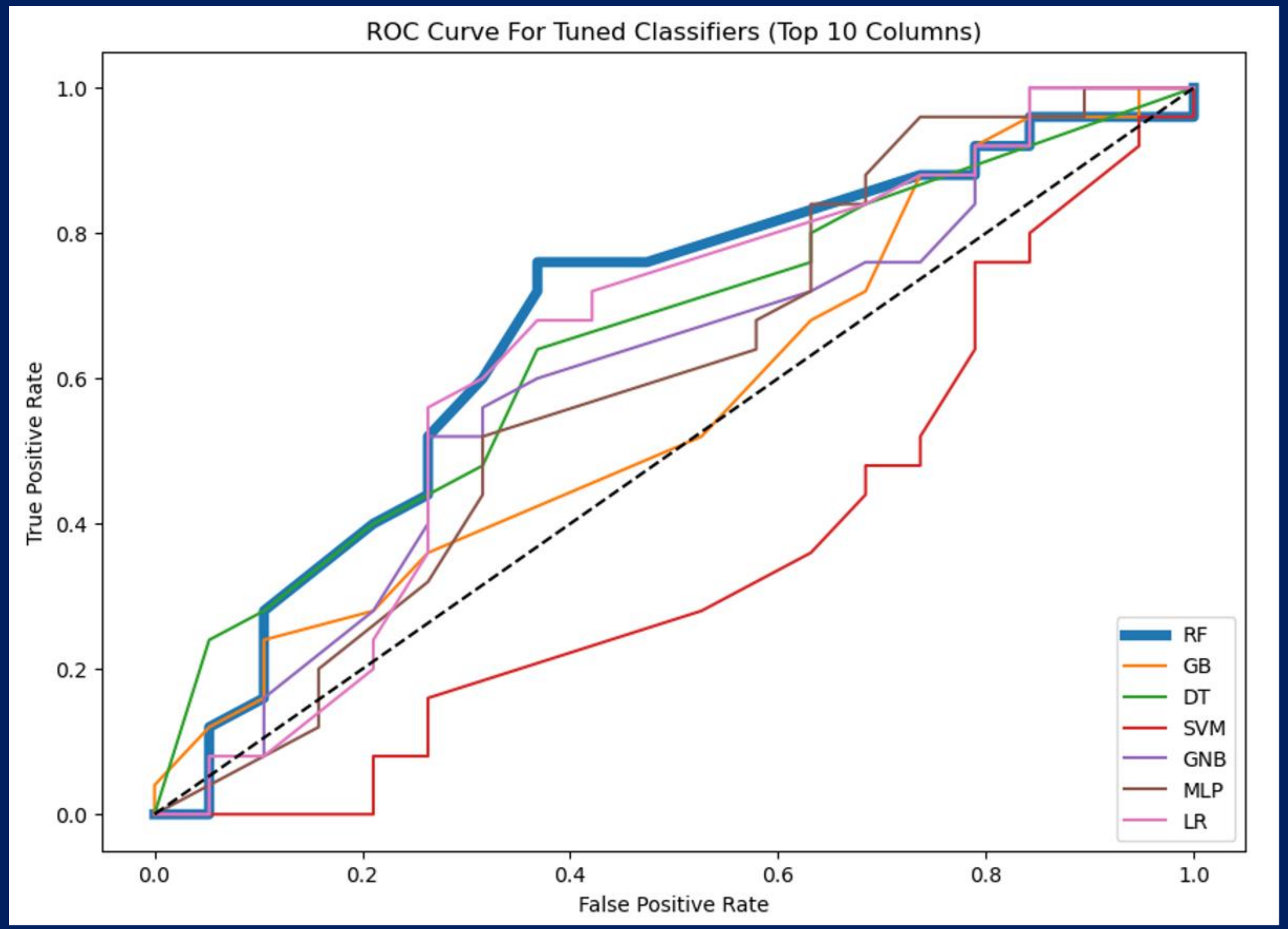
Table 1. Performance metrics for machine learning algorithms. AUC: Area Under Receiver Operator Curve. AUC was calculated before and after Performance Feature Importance was used to identify top 10 most significant clinical variables. Sensitivity, specificity, and accuracy were calculated based on the receiver operator curve calculated after the Permutation Feature Importance.

Discussion

- Limitations:
  - Small sample size
  - Single institution, retrospective analysis
- Future Work:
  - Combine computed tomography angiography (CTA) imaging and clinical data with deep learning to predict complication of endoleak prior to aortic endovascular repair.

Figure 2

Figure 2. ROC curve of supervised machine learning models used to predict endoleaks after being optimized with permutation feature importance.



Conclusion

- The Random Forest Classifier was the best-performing model, with an improvement in sensitivity of 4%, an improvement in specificity of 11%, and an AUC improvement from 0.39 to 0.65.
- Permutation feature importance (PFI) allowed for selecting clinical variable machine learning algorithms were later able to optimize for predictive capacity in detecting endoleaks.
- Machine learning algorithm used to determine PFI will affect the clinical data points selected for optimization.