Comparing Machine-Learning Methods for the Prediction of Major Adverse Limb Events and Mortality after a Percutaneous Intervention

Laura E. Gressler[1,2], MS; Danica Marinac-Dabic[2], MD, PhD, MMSc, FISPE[2]; Susan dosReis[1], PhD, FISPE; Philip Goodney, MD, MS[3]; C. Daniel Mullins[1], PhD; Fadia T. Shaya[1], PhD
[1] University of Maryland School of Pharmacy, Department of Pharmaceutical Health Services Research, Baltimore, MD
[2] Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD, USA
[3] Section of Vascular Surgery, Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA

Abstract

OBJECTIVES: The objective was to formulate, test, and compare the performance of regression-based and machine learning models in the prediction major adverse limb events (MALE) and mortality among patients receiving treatment for lower extremity peripheral artery disease (PAD).
METHODS: Patients undergoing atherectomy, stent, and combination stent atherectomy for lower extremity PAD were identified in the Vascular Quality Initiative registry. Thirty-nine variables summarizing demographic, medical history, pre-operative, indication-specific, and procedure-specific characteristics were utilized to predict MALE and mortality events. For both events, we compared the performance of four different prediction models: a generalized linear model (GLM), a Least Absolute Shrinkage and Selection Operator (LASSO) regularized GLM, a gradient boosted decision tree, and random forest model. The area under the curve (AUC) evaluated the effectiveness of each prediction model. For validation purposes, 5-fold cross-validation was repeated three times. Pairwise comparisons of the receiver operating characteristic curves (ROC), sensitivity, and specificity measures with Bonferroni adjustment for multiple testing applied were performed to compare the models' performance.
RESULTS: Among 15964 identified patients, a MALE occurred in 26.02% of patients, and death occurred in 18.82% of patients. The most effective predictive model for MALE, as determined by the AUC, was the gradient boosted decision tree (AUC= 0.7539) followed by the LASSO regulated GLM (AUC= 0.749). The most effective predictive model for mortality was the LASSO regularized GLM (AUC=0.7930) followed by the GLM model (AUC=0.7922). The GLM, LASSO regularized GLM model, and gradient boosted decision tree produced similar ROC.
CONCLUSIONS: All models showed acceptable discrimination, with an AUC greater than 0.7, when predicting MALE and mortality among patients receiving treatment for lower extremity peripheral artery disease. The machine learning techniques outperformed traditional regression-based techniques and can be leveraged to generate robust predictive models within the clinical space of lower extremity PAD.

Introduction

- Regulatory bodies and clinicians have increasingly accepted and leveraged the use of real word evidence (RWE) generated from real-world data (RWD) to inform regulatory and clinical-decision making.[6–8]
- RWD can not only be used to retrospectively assess the effectiveness and safety of revascularization procedures but can also be used to formulate models that help determine risk factors and probabilities of successful revascularization among patients needing percutaneous treatment.
- Registries are an important source of RWD and have been continuously used to inform clinical and regulatory decision making.[6,13]
- The results from robust predictive models may help regulatory processes by informing the future clinical studies for labeling or indications for use for a device, and clinical guidelines for the appropriate use of devices among patients.
- Machine learning models may overcome a number of limitations associated with traditional regression-based models.
- Lower extremity peripheral artery disease (PAD) refers to the buildup of plaque within the peripheral arteries.[1,2]
- The objective of this study was to formulate and test a model used to predict major adverse limb events (MALE) and mortality among patients receiving treatment for lower extremity PAD.

Materials and Methods

- This study utilized data from the Vascular Quality Initiative (VQI) registry.
- Included patients received atherectomy or stenting for symptomatic arterial occlusive disease in a lower extremity non-aortic vein between January 2010 and September 2018.
- Patients were excluded who (1) previously received a stent or atherectomy procedure to mitigate potential misclassification of adverse events to the current procedure observed in the dataset, (2) experienced an endpoint of interest because the focus of this study was on new events and not recurrent events, or (3) did not experience a 3 years of follow-up.
- The study cohort was stratified into three exposure groups: (1) atherectomy alone, (2) stent placement alone, and (3) combination of stent placement and atherectomy
- The primary endpoint of interest was a major adverse limb event (MALE) which encompasses above the ankle amputations and major reinterventions, such as a bypass graft revision and mortality.
- A total of 39 variables describing demographic, medical history, pre-operative, indication, procedure, and discharge-related characteristics of the included population were included in the predicted models.
- The MALE and mortality prediction models were built using logistic regression, logistic regression with the Least Absolute Shrinkage and Selection Operator (LASSO) regularization, gradient boosting, and random forest models
- The final predictive model with the greatest area under curve (AUC) within each model derivation method was selected.
- Given that the outcomes are categorical variables, sensitivity, specificity, the positive predictive values, and the negative predictive values were calculated to further describe the generated predictive models.
- Variable importance was used to identify the twenty most significant predictors within each model
- For validation purposes, 5-fold cross-validation was repeated three times.
- To compare the performance of the models, pairwise comparisons with Bonferroni adjustment for multiple testing applied to receiver-operating curves (ROC), sensitivity, and specificity measures were performed.

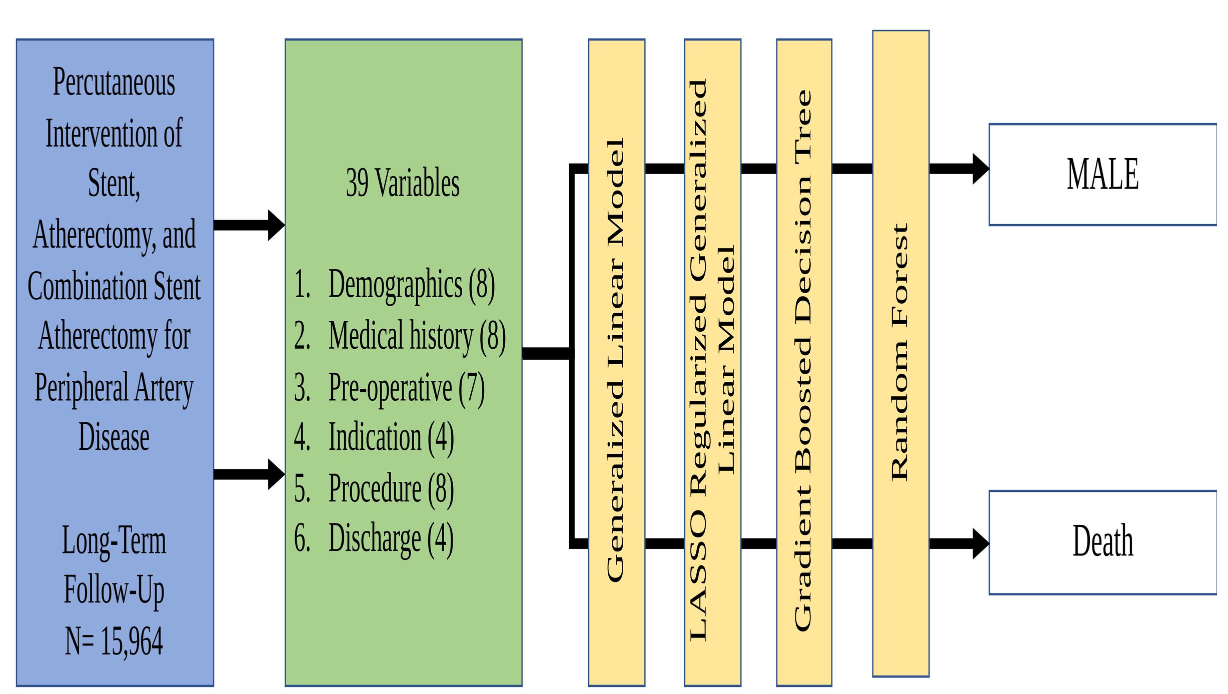

Figure 1. Methods Summary of models and outcomes assessed at 3 years

Results and Discussion

- In the VQI database, 98868 patients underwent an atherectomy or stent placement procedure from January 2010 through September 2018
  - After applying the exclusion criteria, 46108 patients remained with 15964 patients experiencing follow-up for three years

Outcome: Major Adverse Limb Event	Generalized Linear Model	LASSO Regularized Generalized Linear Model	Gradient Boosted Decision Tree	Random Forest
Accuracy; (95% CI)	0.7663 (0.7512, 0.7809)	0.7663 (0.7512, 0.7809)	0.7675 (0.7525, 0.7821)	0.7669 (0.7519, 0.7815)
NIR	.74	.74	.74	.74
P-value [Accuracy>NIR]	<.0001	<.0001	<.0001	<.0001
Kappa	.2645	.261	.244	.2216
AUC	.7483	.7492	.7539	.7350
Sensitivity	.28193	.27590	.24337	.20964
Specificity	.93649	.93861	.95174	.96274
Positive Predictive Value	.60937	.61230	.63924	.66412
Negative Predictive Value	.78775	.78673	.78164	.77611

Table 2. Summary of Model Results for the Outcome Major Adverse Limb Event*NIR= No Information Rate; AUC=Area Under the Curve

Outcome: Major Adverse Limb Event	ROC		Sensitivity		Specificity	
	Difference	p-value	Difference	p-value	Difference	p-value
GLM vs. LASSO	-0.000989	1.00	-0.002540	1.00	0.003416	1.00
GLM vs. XGB	-0.004355	1.00	-0.017852	<0.001	0.039518	<0.001
GLM vs. RF	0.005895	0.01956	-0.026461	<0.001	0.082932	<0.001
LASSO vs. XGB	-0.003366	1.00	-0.015312	<0.001	0.036102	<0.001
LASSO vs. RF	0.006884	0.64787	-0.023920	<0.001	0.079517	<0.001
XGB vs. RF	0.010251	0.03996	-0.008609	<0.001	0.043414	<0.001

Table 3. Difference in ROC, Sensitivity, and Specificity with Bonferroni Adjusted p-values for the Outcome Major Adverse Limb Event

- In the prediction of MALE, the gradient boosted model generated the greatest accuracy of 0.7675 (95%CI:0.752-0.782), the greatest AUC of 0.754, a sensitivity of 0.243, and specificity of 0.952.
- Common leading predictors among all models were procedure setting, artery type, age category, primary insurance type, pre-operative conditions, pre-operative medications, and number of arteries treated. Stent as a treatment type was only identified as a leading predictor in the gradient boosted model.
- The difference in ROC statistic was significantly different when comparing the random forest model to the GLM model and gradient boosted model.
- When predicting mortality, the LASSO regularized generalized linear Common leading predictors among all models were age category, primary insurance type, BMI category, pre-operative diabetes, pre-operative dialysis, a history of congestive heart failure, procedure setting, leg symptoms, and pre-operative smoking. Treatment type was not identified as a leading predictor in any of the models.
- Pairwise comparisons with Bonferroni adjustment for multiple testing showed that the GLM, LASSO regularized GLM model, and gradient boosted decision tree produced similar ROC.

Outcome: Mortality	Generalized Linear Model	LASSO Regularized Generalized Linear Model	Gradient Boosted Decision Tree	Random Forest
Accuracy; (95% CI)	0.8305 (0.817, 0.8434)	0.8308 (0.8174, 0.8437)	0.834 (0.8209, 0.847)	0.8299 (0.8164, 0.8428)
NIR	0.8117	.8117	.8117	0.8117
P-value [Accuracy>NIR]	.003	.003	<.001	.004
Kappa	.3156	.3120	.3150	.2096
AUC	.7922	.7930	.7881	.7808
Sensitivity	.30283	.29617	.28785	0.16306
Specificity	.95291	.95484	.96102	0.98456
Positive Predictive Value	.59868	.60339	.63139	0.71014
Negative Predictive Value	.85492	.85399	.85332	0.83530

Table 4. Summary of Model Results for the Outcome Mortality*NIR= No Information Rate; AUC=Area Under the Curve

Outcome: Mortality	ROC		Sensitivity		Specificity	
	Difference	p-value	Difference	p-value	Difference	p-value
GLM vs. LASSO	-0.001190	1.0	-0.0029583	.5044	0.017350	0.02159
GLM vs. XGB	0.003138	1.0	-0.0033766	.2262	0.022461	0.03034
GLM vs. RF	0.012734	1.0	-0.0260095	1.00	0.152104	1.00
LASSO vs. XGB	0.004328	0.058074	-0.004183	<.001	0.005111	<.001
LASSO vs. RF	0.013924	0.008389	-0.0230512	<.001	0.134753	<.001
XGB vs. RF	0.009595	0.261462	-0.0226329	<.001	0.129643	<.001

Table 5. Difference in ROC, Sensitivity, and Specificity with Bonferroni Adjusted p-values for the Outcome Mortality

- A predictive model that can detect important latent relationships between predictor variables and that is not hindered by strong assumptions regarding the functional form of the relationship between the available predictors and outcomes of interest may inform clinical and regulatory decision making.
- A robust and validated predictive model may serve as an additional clinical decision-tool that may identify high-risk patients.
- Identified high-risk patients may benefit from additional post-operative monitoring and can also inform communications between clinicians and patients. Informing patients of their increased risk may allow them to implement lifestyle changes.
- Predictive models may support regulatory decision making by promptly identifying complex patient populations where devices or a combination of devices may not be performing as intended.

Conclusion

- All evaluated models predicting MALE and mortality following a revascularization procedure showed acceptable discrimination in determining events.
- This study supports the use of predictive modeling within the clinical space of lower extremity peripheral artery disease.
- Future machine learning models may employ additional data and linkage to other data sources to further inform and increase the generated predictive models' discriminatory ability.