Optimizing Claims Codes to Identify Claudication and Chronic Limb-threatening Ischemia Using a Machine-learning Approach

Sanuja Bose, MD, MPH[1] Sharon C. Kiang, MD[2], Daniel Roh[2], Jialin Mao, MD, PhD[3], Andrew Cabrera, David P. Stonko, MD, MS[1], James H. Black, III, MD[1], Jesse A. Columbo, MD, MS[4], Philip P. Goodney, MD[4], Leigh Ann O'Banion, MD[5], Roger T. Tomihama, MD[2], Caitlin W. Hicks, MD, MS[1]

[1]Johns Hopkins University School of Medicine , [2]Loma Linda University Medical Center, [3]Weill Cornell Medical College, [4]Dartmouth-Hitchcock Medical Center, [5]University of California San Francisco

Background

- The accuracy of using contemporary claims codes to identify patients with peripheral artery disease has been questioned.
- We aimed to validate a predefined set of ICD-10 codes that have been used to identify patients with claudication and chronic limb-threatening ischemia (CLTI), and to optimize their diagnostic accuracy using a supervised machine-learning (ML) approach.

Methods

- Design: Retrospective observational study
- Database: Vascular Quality Initiative Vascular Implant Surveillance and Interventional Outcomes Network (SVS VQI-VISION)
- Study Period: Jan 2016 - Dec 2019
- Population:
  - Inclusion: All patients who underwent a peripheral vascular intervention for claudication or CLTI (defined by VQI)
  - Exclusion: Patients with acute limb ischemia or unknown leg symptoms or side of surgery
- Statistical Analysis:
  - Traditional logistic regression
    - Primary code position
    - Any code position
  - ML Logistic Regression Classifier
  - ML Random Forest Classifier
  - ML Gradient Boosting Classifier
  - ML Decision Tree Classifier
  - ML Gaussian-Naïve Bayes
  - ML Multi-Layer Preceptron

Results

- 54,180 patients (mean age 71.9, 41.0% female, 16.9% non-Hispanic Black) underwent a peripheral vascular intervention (VQI-VISION)
  - 20,768 (38.3%) had claudication per VQI
  - 33,411 (61.7%) had CLTI per VQI

Figure 1

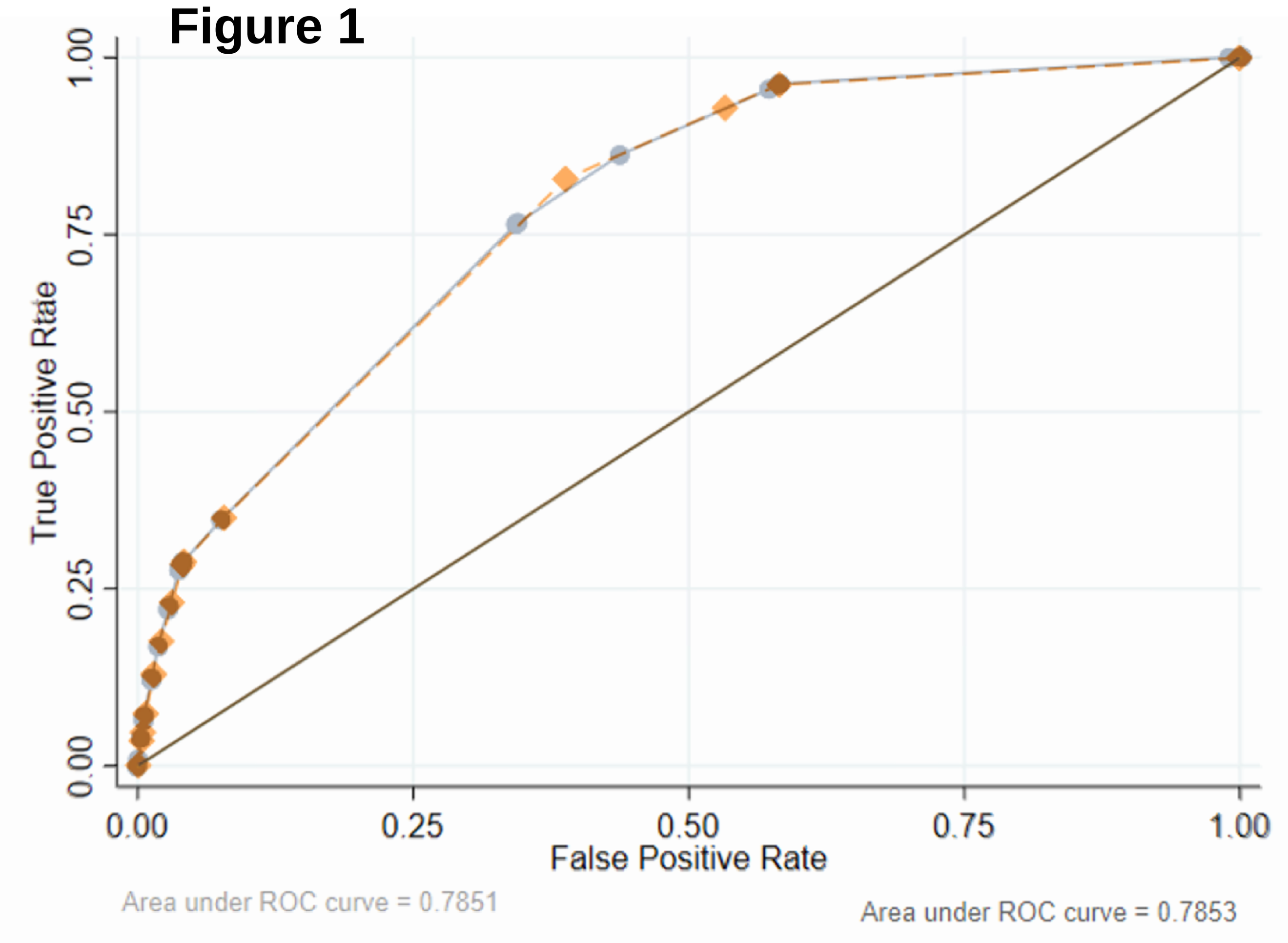

Figure 2

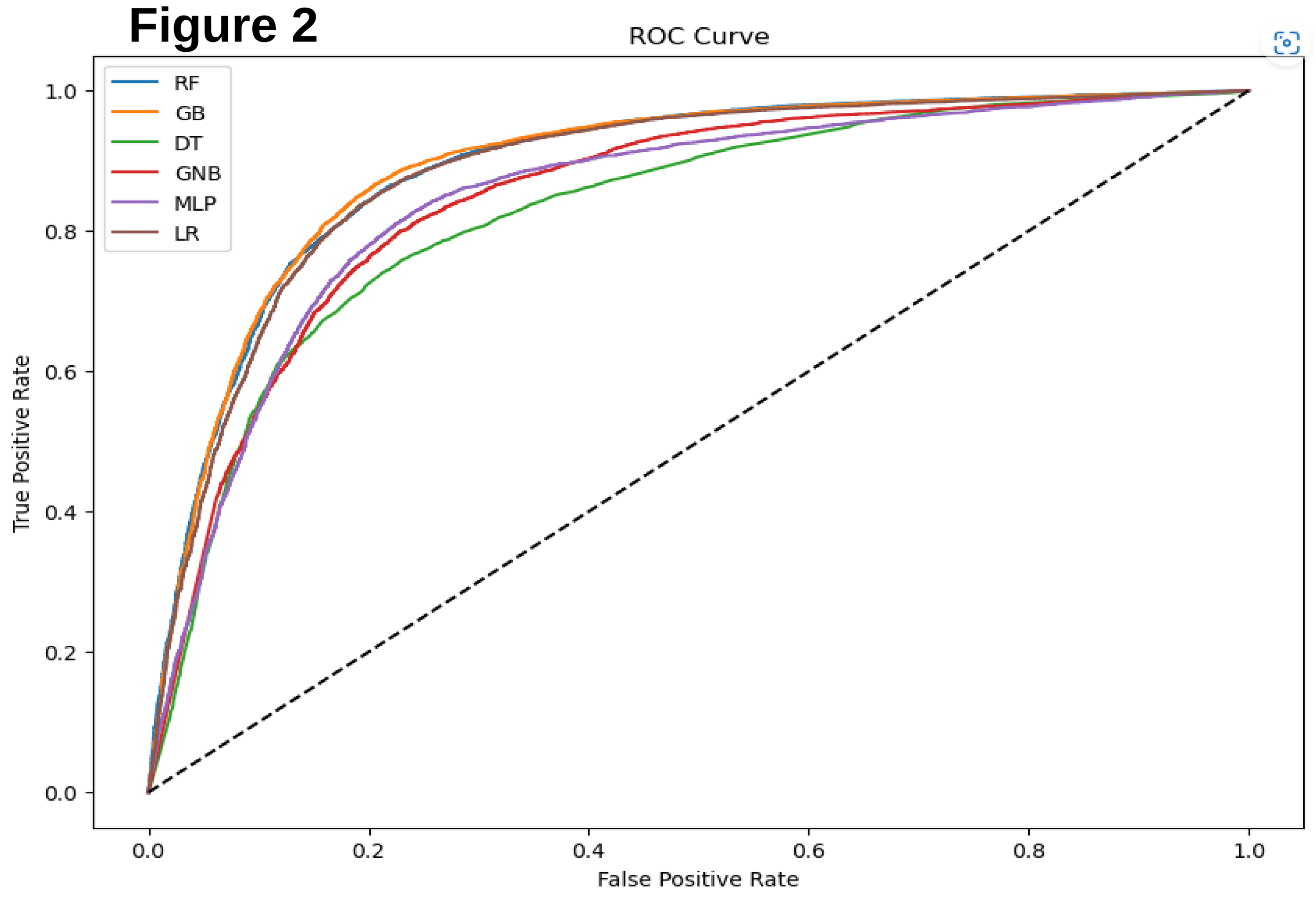

Figure 1. ROC curves for distinguishing claudication and chronic limb-threatening ischemia using traditional logistic regression models, for both primary codes (orange diamonds) and codes in any position (blue circles).

Figure 2. ROC curves for distinguishing claudication and chronic limb-threatening ischemia using several supervised machine-learning models.

Table. Performance statistics of various diagnostic models using ICD-10 diagnostic codes to distinguish claudication and chronic limb-threatening ischemia.

Type of Model	Sensitivity (%)	Specificity (%)	Total Agreement (%)	AUC	F1 Score
Predetermined Code List					
Primary Code Position	65.5	89.1	76.4	—	—
Any Code Position	80.9	81.9	81.3	—	—
Logistic Regression Models					
Primary Code Position	96.2	41.8	75.4	0.7853	—
Any Code Position	95.4	42.8	75.3	0.7851	—
ML Models – Primary Code Position					
Logistic Regression	92.56	55.27	78.27	0.8458	0.84
Random Forest Classifier	77.69	74.72	76.55	0.8478	0.80
Gradient Boosting Classifier	93.40	54.02	78.30	0.8470	0.84
Decision Tree Classifier	75.06	77.56	76.59	0.8475	0.80
Gaussian Naïve Bayes	13.18	97.16	45.38	0.6640	0.23
Multi-Layer Perceptron Model	13.18	97.16	78.29	0.8451	0.84
ML Models – Any Code Position					
Logistic Regression	87.15	77.03	83.29	0.8905	0.87
Random Forest Classifier	82.29	81.61	82.03	0.8921	0.85
Gradient Boosting Classifier	88.62	77.07	84.22	0.8278	0.87
Decision Tree Classifier	88.62	77.07	76.06	0.8482	0.80
Gaussian Naïve Bayes	45.54	92.91	63.59	0.8466	0.61
Multi-Layer Perceptron Model	83.26	75.23	80.20	0.8849	0.84

Discussion

- Performance statistics show the predetermined sets of ICD-10 codes used to identify CLTI and claudication in administrative databases perform on par with validated registry data
- Supervised ML models can be used to further optimize accuracy and precision of distinguishing between claudication and CLTI if necessary

Limitations

- Registry data used as gold standard
- Data only tested
- Limited accessibility of ML technology

Conclusions

- Previously published sets of ICD-10 codes can accurately discriminate between between claudication and CLTI.