Models to Predict Pulmonary Embolism for Patients with Asthma Exacerbation

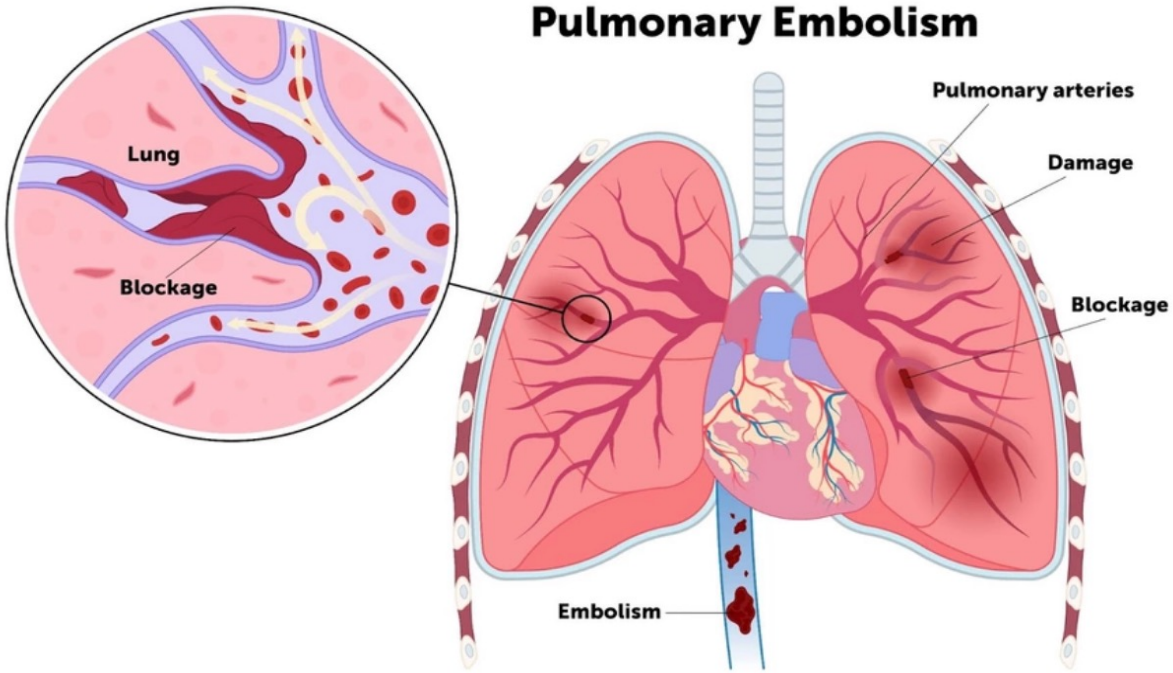
David Hu

April 22, 2024

Introduction

- Blood flow is blocked to an artery in the lungs
  - Caused by a blood clot that comes from the deep veins in the legs, called deep vein thrombosis (DVT)
- ~900,000 people/year affected

  ~60,000 to 100,000 die
- Diagnostic Tests
  - D-Dimer
  - CTA



Data

- June 2011 and October 2018 at University of Florida (UF) Health System, Gainesville, Florida
- 700 Patients
    - Asthma Exacerbation
    - Subjected to CTA for suspected PE
- 22 clinical and demographical variables measured

Methods (Models)

- Backward Stepwise Logistic Regression
    - Add additional features from Random Forest
- Random Forest (RF)
- Support Vector Machine (SVM)
- Naïve Bayes

Data Description & Preprocessing

- 136 instances of PE
- Contraceptives, Fractures/Anesthesia, Hemoptysis dropped
- 80% training, 20% testing
- 10-fold cross validation
- Done in R

Methods (Evaluation Metrics)

- Sensitivity
- Specificity
- AUC-ROC

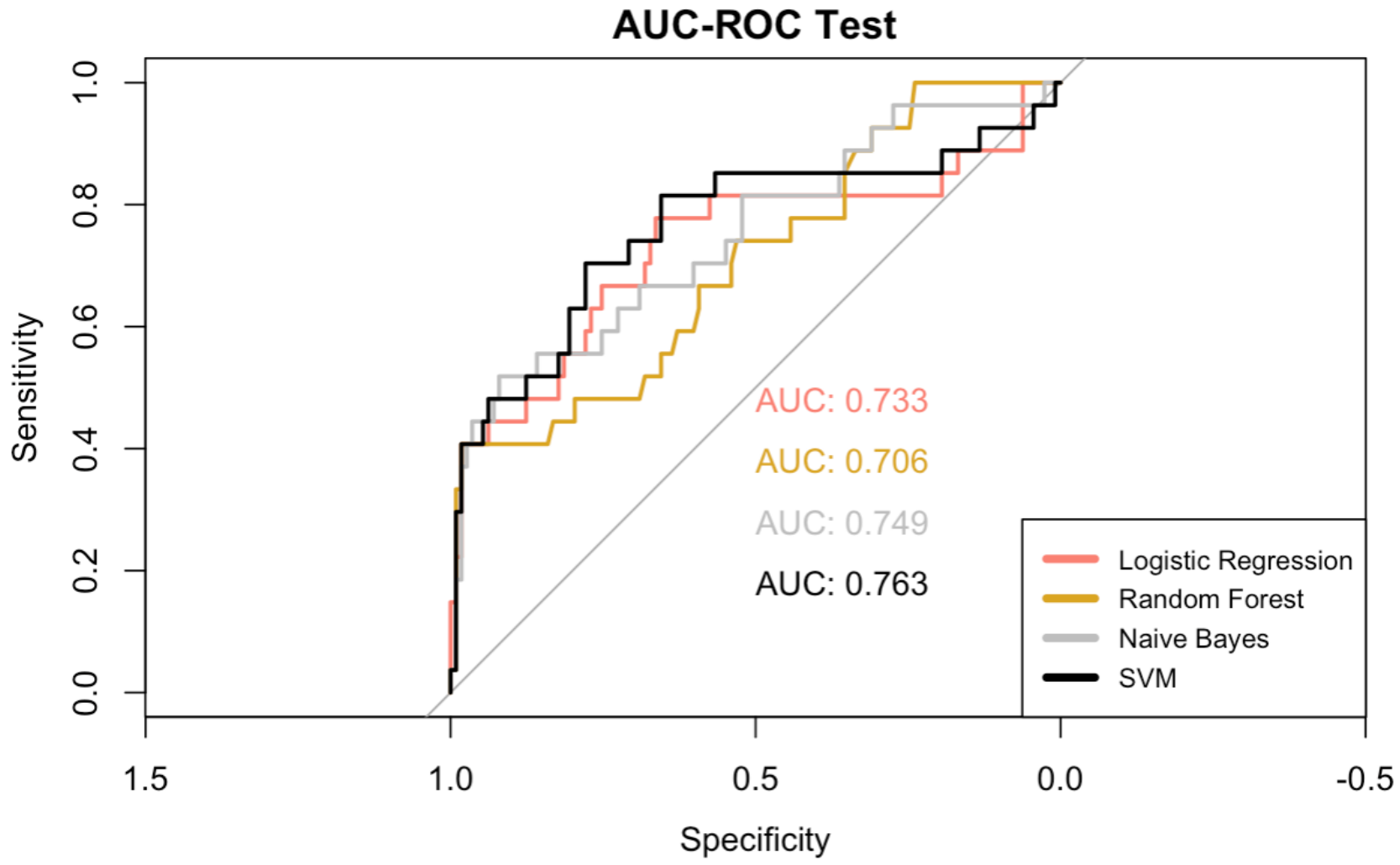
Threshold Values

- Each model outputs a probability of PE
- Chose threshold that achieves 80% sensitivity on training set

Results

Method	Sensitivity	Specificity	AUC-ROC
Logistic Regression	.81	.56	.73
Random Forest	.78	.44	.71
SVM	.81	.6	.75
Naïve Bayes	.7	.6	.76

Results


AUC-ROC Test
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1.5
1.0
0.5
0.0
-0.5
Specificity
AUC: 0.733
AUC: 0.706
AUC: 0.749
AUC: 0.763
Logistic Regression
Random Forest
Naive Bayes
SVM

Discussion (Results)

- Best Model: Logistic Regression
- “PE History”, “stroke/TIA”, “hypertension”, “BMI”, “HR”, “Age”
- Issues
  - Low specificity (over-diagnose)
  - Generalizability to asthma exacerbation population
  - No laboratory measures

Discussion (Next Steps)

- Imbalanced Data
  - Resampling Methods
  - Cost Function
- External Validation
- Hyperparameter tuning
- Other ML Models

Thank You!/Questions?