Risk Prediction and Diagnostics of Cardiovascular Diseases

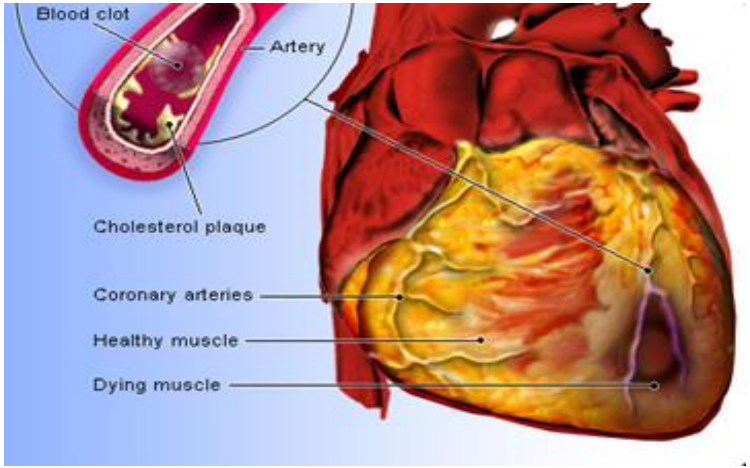
Igor Putrenko

Insight Health Data Science Program

San Francisco, CA
October 2017

Cardiovascular Diseases (CVD)

- Annual direct medical expenditures (by 2030) – $818 billion
- One in three U.S. adults has one or more types of CVD
- Optimized diagnostic process can substantially decrease the costs

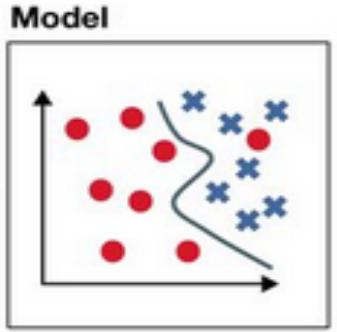


Can we predict CVD type based on lab tests?

Patient

- Physical assessment
- ECG
- Stress Test
- Ultrasound
- X-ray
- MRI
- **Lab Tests**
- Angiography

Dx



Dx

DATA SOURCE

Beth Israel Deaconess Medical Center

- MIMIC-III, critical care database
- 38,597 patients
- Median age is 65.8 years (52.8–77.8)

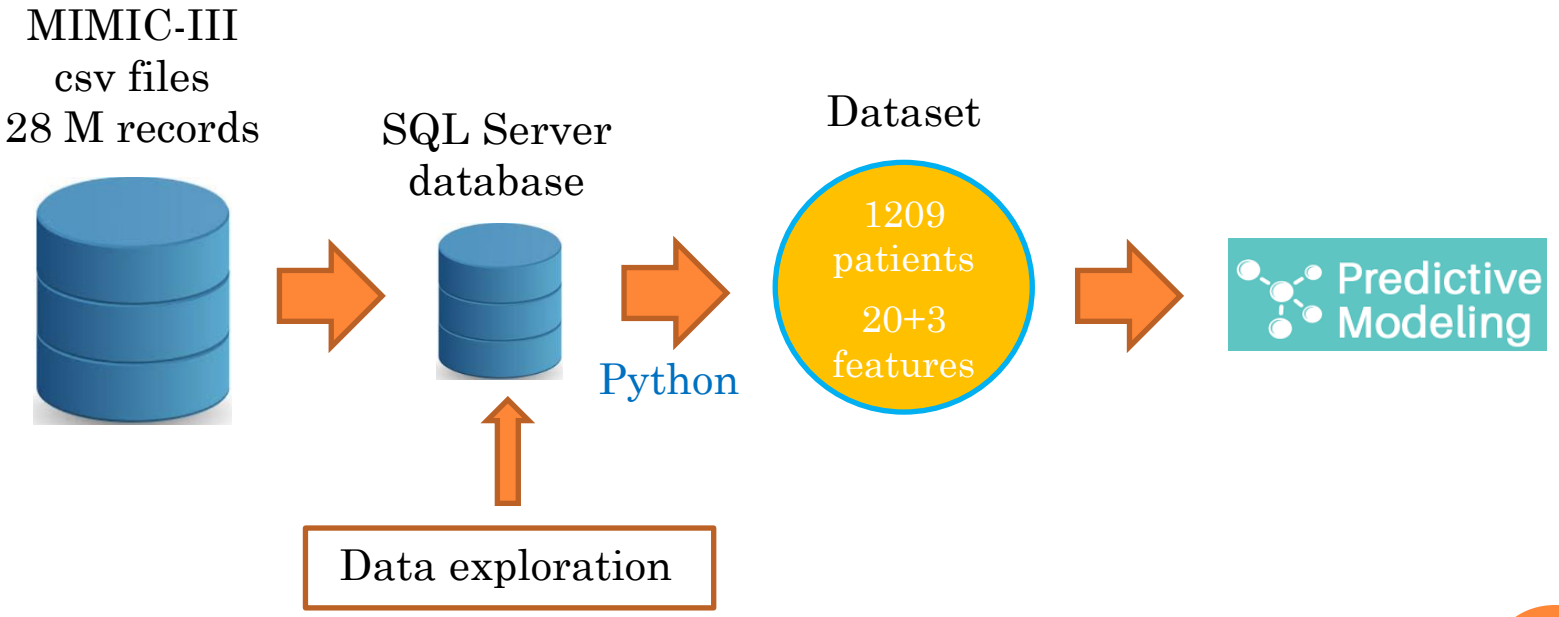
Recordings
Images
Lab Tests
Demographics
Notes and Reports

Data Archive

De-identification
Date shifting
Format conversion

MIMIC-III

Data pipeline



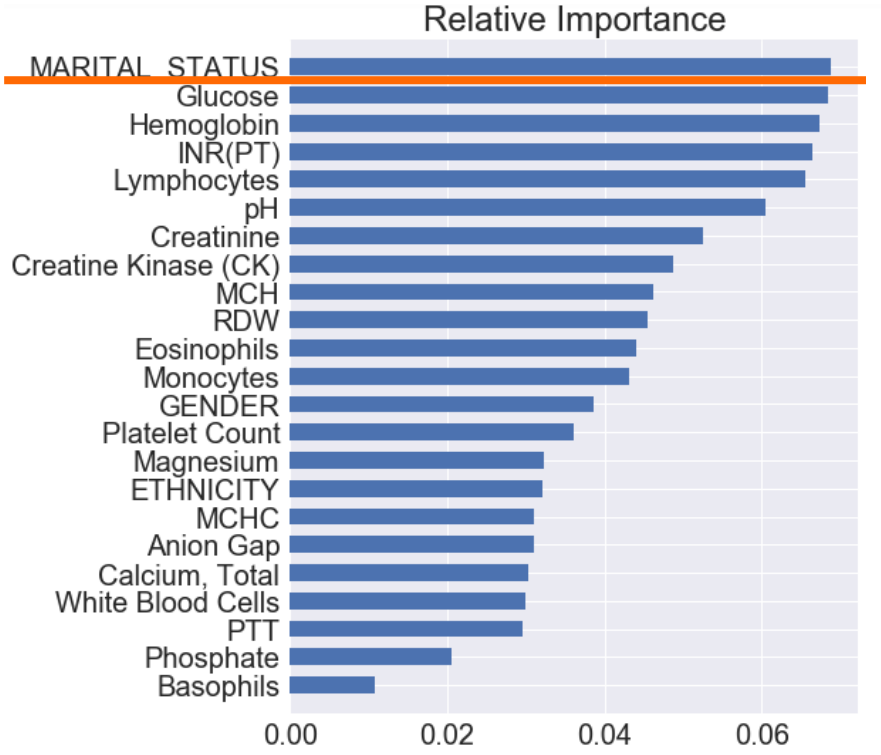
Predictive modeling

- 6 CVD categories
- Balanced classes
- Cross-validation (5-fold)
- Grid search

- Gradient Boosting
- Random Forest
- Artificial Neural Network

Accuracy: 39% (vs 16.7% chance)

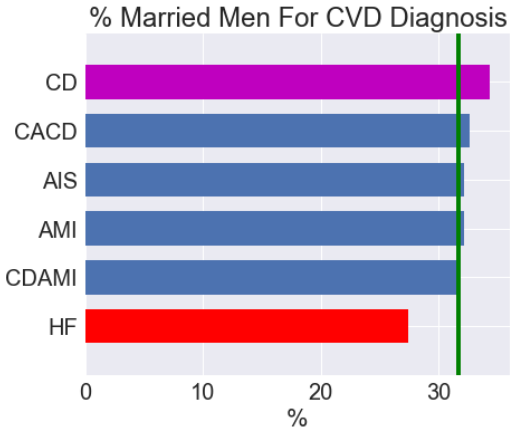
Feature Importance

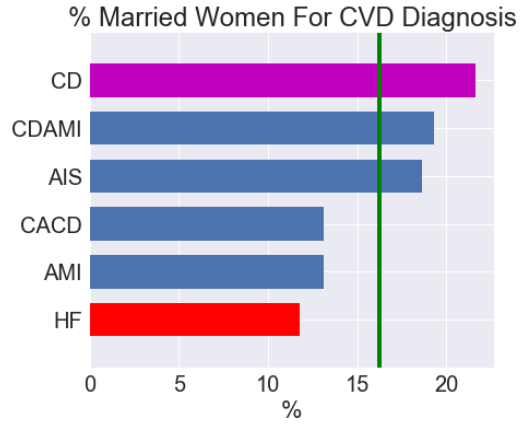


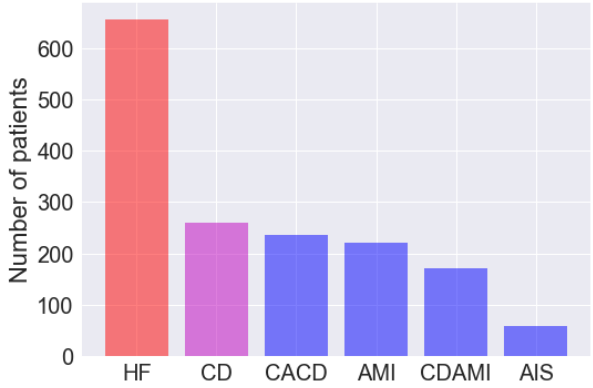
Identified features are indicators of:

- status of immune system
- chronic diseases (kidneys, liver, heart)
- chronic infection
- genetic factors
- psychological well-being – ?

SHOULD YOU GET MARRIED?







Conclusion

The developed model can improve CVD diagnostic process:

- Risk assessment
- Early detection
- Decision support
- Cost reduction
- Faster diagnostics

Clients: clinicians, health agencies, insurance companies

About me

- Ph.D. in Biochemistry
- Academia: electrophysiology, brain & pain research
- Biotech industry: pain drug discovery
- IT: business data and systems analysis
- Concentrate on Data Science

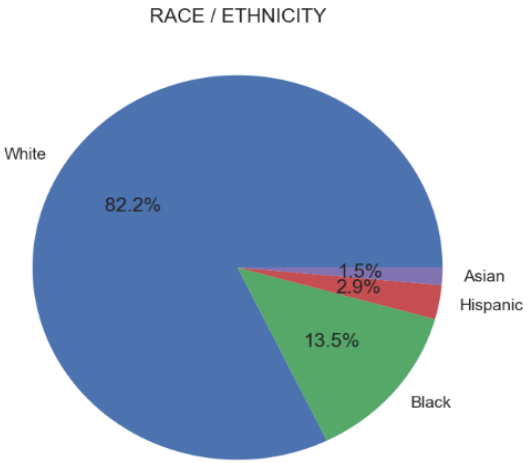
Backup Slides

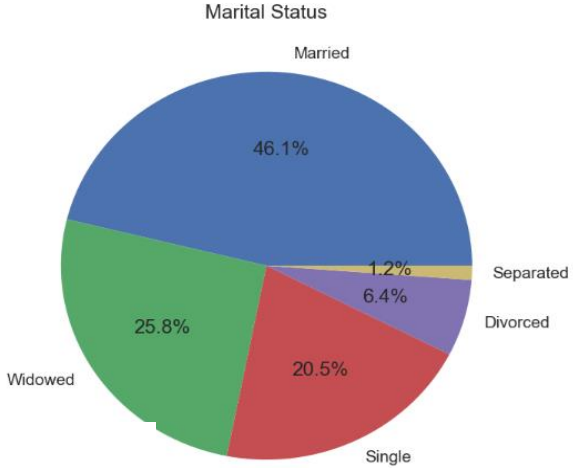
CVD Categories

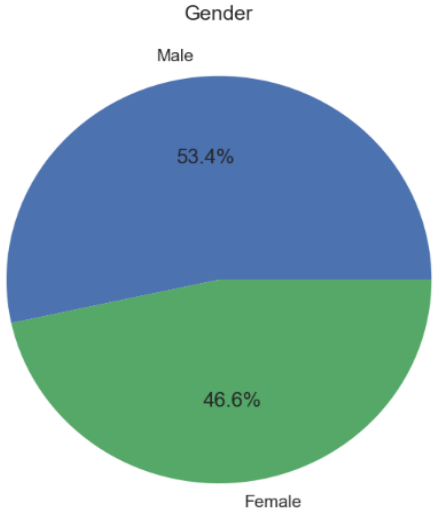
Patients selected with the following Diagnosis Related Groups (DRGCODES) used by the hospital for billing purposes

- ACUTE ISCHEMIC STROKE
- ACUTE MYOCARDIAL INFARCTION (AMI)
- CARDIAC ARRHYTHMIA & CONDUCTION DISORDERS
- HEART FAILURE
- CIRCULATORY DISORDERS EXCEPT AMI
- CIRCULATORY DISORDERS WITH AMI

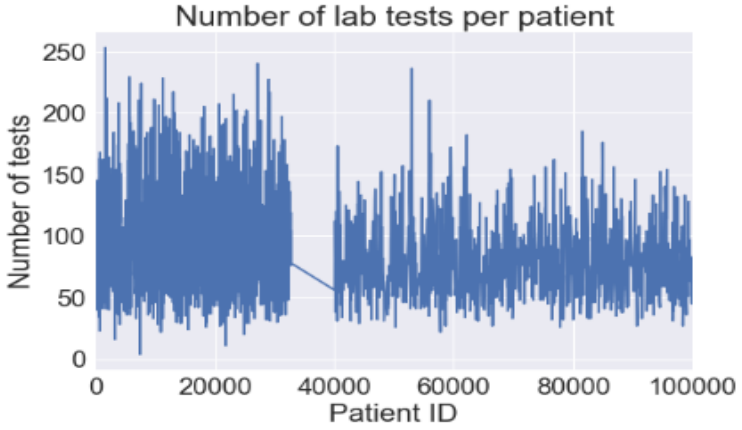
Demographic Data

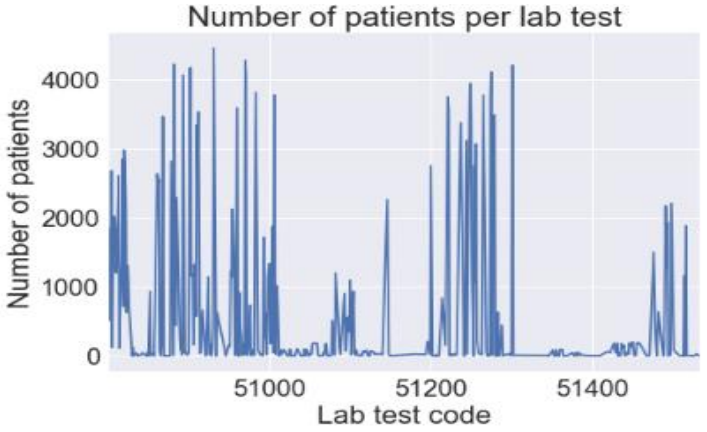


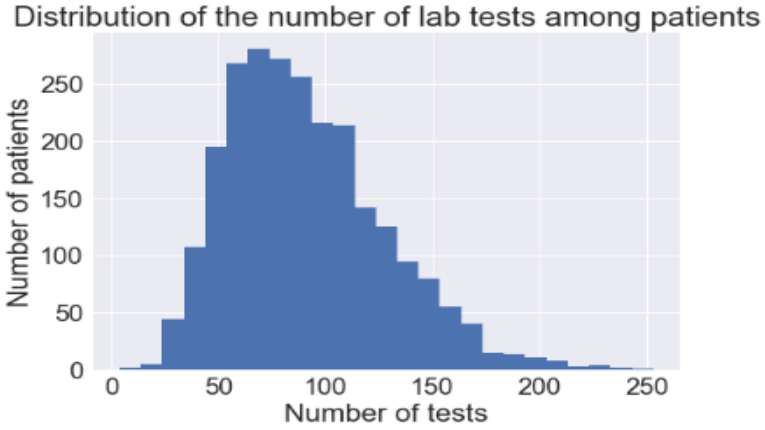




Dimension Reduction – Selecting Lab Tests

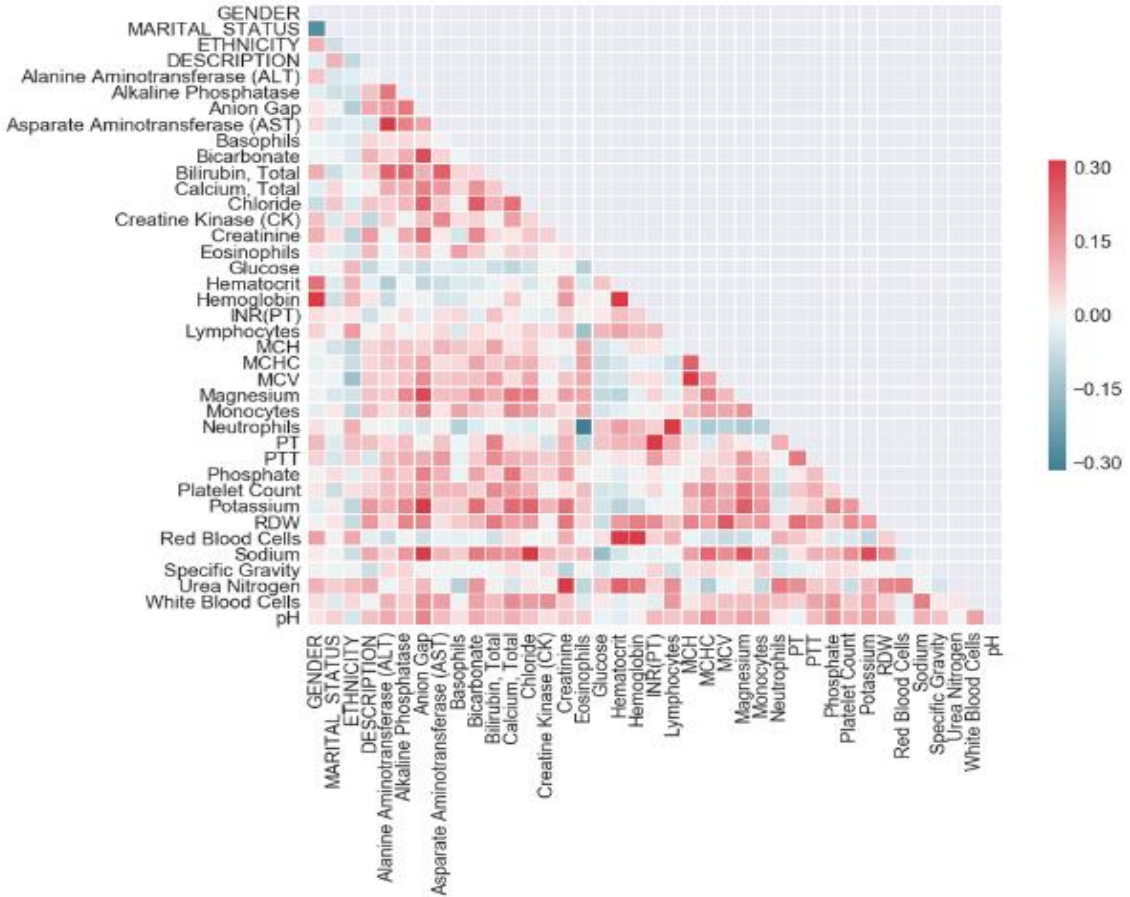




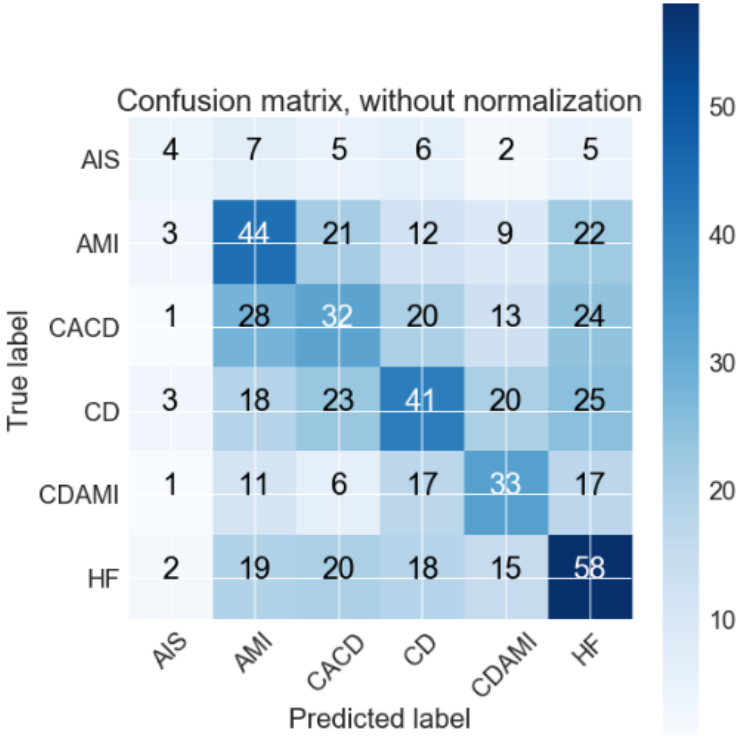


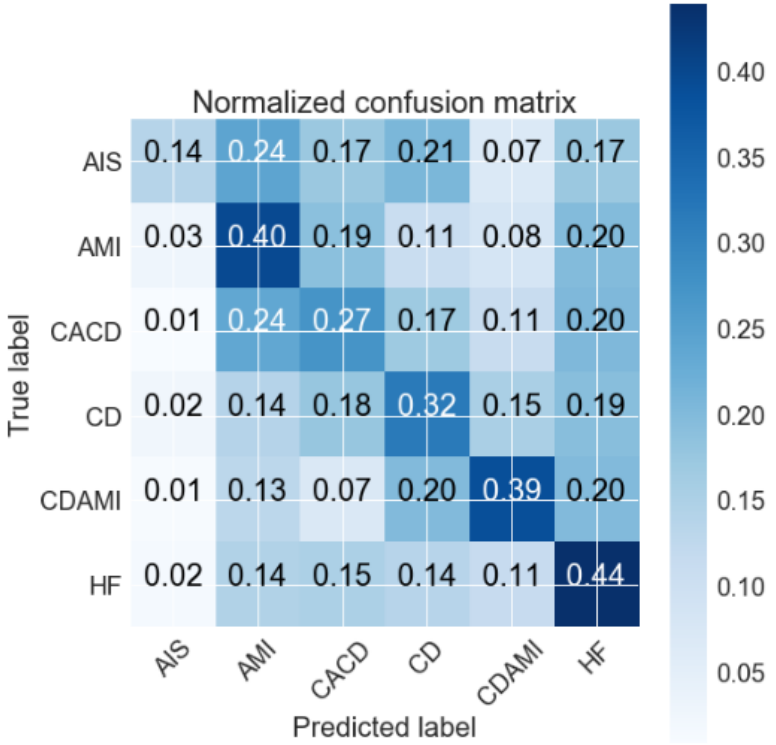
Total number of CVD patients	**2455**
Total number of lab tests	**403**
Feature selection – missing values threshold (patients/test)	**2155**
Lab tests (features) passed	**35**
Final number of selected patients	**1209**

Identifying Multicollinearity



Confusion Matrix





Precision, Recall, and F1-score

	precision	recall	f1-score	support
0	0.29	0.14	0.19	29
1	0.35	0.40	0.37	111
2	0.30	0.27	0.28	118
3	0.36	0.32	0.34	130
4	0.36	0.39	0.37	85
5	0.38	0.44	0.41	132
avg / total	0.35	0.35	0.35	605