

# Comparison of algorithms for the classification of coronary artery disease using protein multiplex panel



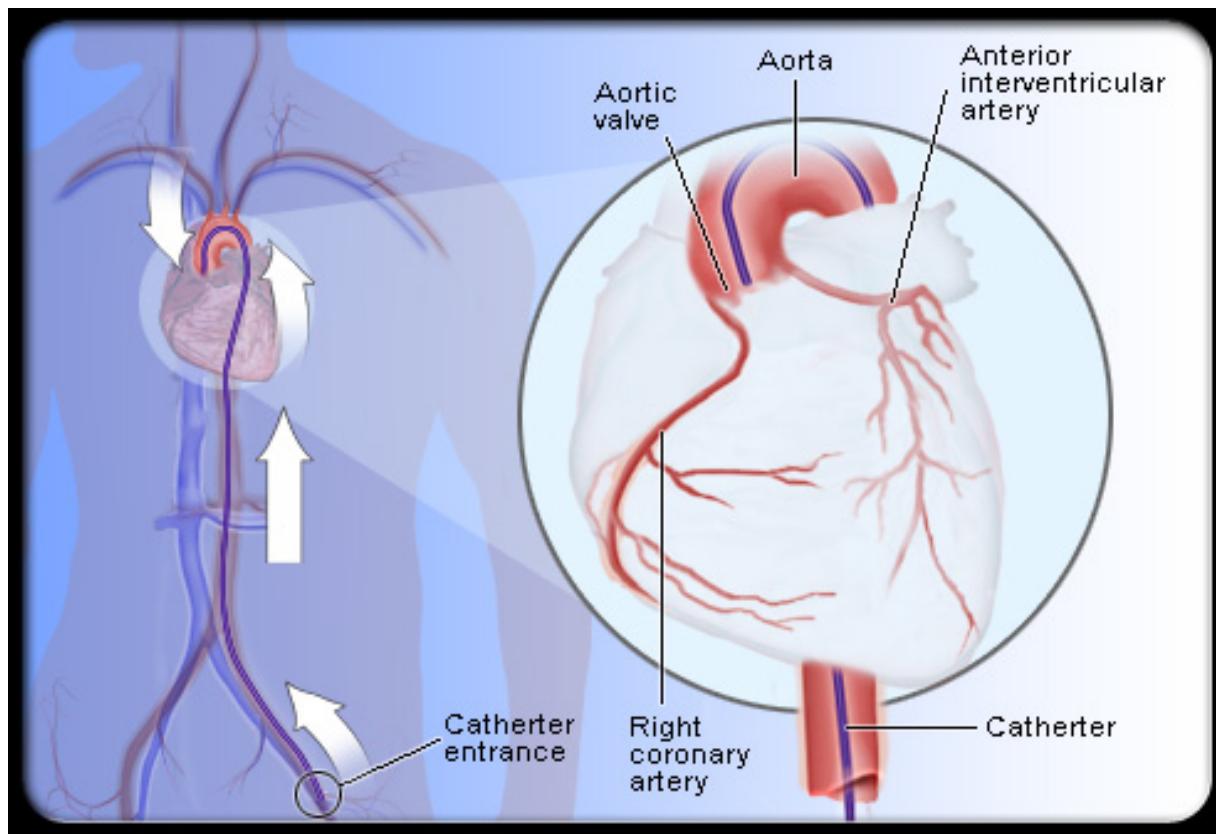
Frank Lin, John-Krill Burger, Prateek Tandon, Vineet Joshi

# Introduction

- + Coronary Artery Disease (CAD) is the number one cause of death globally.
- + Causes over 800,000 deaths per year in the US.

# Coronary Angiography

- + Gold standard for assessment of anatomic coronary disease



# Angiography Disadvantages

- + Mean cost of \$10,880
- + Invasive
- + Risk of using contrast media and ionizing radiation

Table 1.

Risk of Cardiac Catheterization and Coronary Angiography (No. of Patients = 59,792)

	%
Mortality	0.11
Myocardial infarction	0.05
Cerebrovascular accident	0.07
Arrhythmia	0.38
Vascular complications	0.43
Contrast reaction	0.37
Hemodynamic complications	0.26
Perforation of heart chamber	0.03
Other complications	0.28
Total of major complications	1.70

legend Modified with permission from Noto et al. (13).

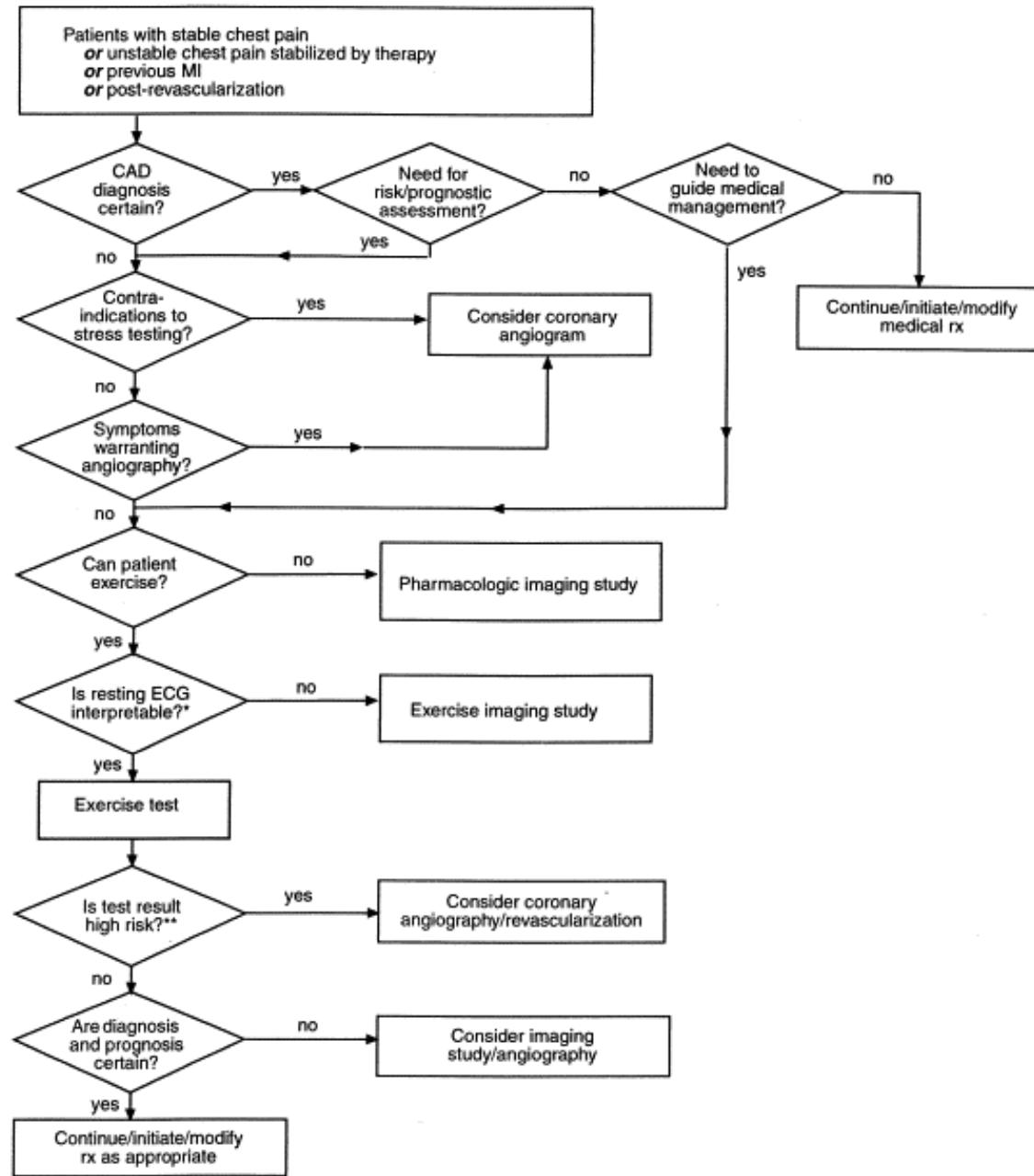


Table 4.

## Appropriateness of Angiography: Range of Findings in Literature

Author	Study Cohort	Methods	Results
Bernstein et al. (52)	1,335 patients 15 hospitals in New York	Rand method	Inappropriate, 4%
Mozes and Shabtai (53)	499 patients 2 hospitals in Israel	Consensus panel	Inappropriate, 58%
Hampton et al. (54)	3 centers in England	Expert panel	Inappropriate, 10%–28%
McGlynn et al. (55)	1,336 patients 15 hospitals in New York; 15 in Canada	Rand method in United States Consensus panel in Canada	Inappropriate, 4%–10%
Bengtson et al. (56)	831 patients in Sweden	Expert panel	Inappropriate, 2%
Roos et al. (57)	351 patients in Canada	Expert panel	In Canada, inappropriate, 9%
	1,677 patients in United States		In United States, inappropriate, 15%–18%

# Alternatives

- + Gene expression classifier
  - + Rosenberg et al. developed a 23 gene blood-based expression test
- + Protein markers
  - + Troponin T for myocardial infarction
  - + NT-pBNP for acute coronary syndrome
  - + CRP for future cardiovascular events
- + Protein multiplex panels
  - + LaFramboise et al. measured concentration of 24 proteins

# Algorithm to beat

- + Scoring function:

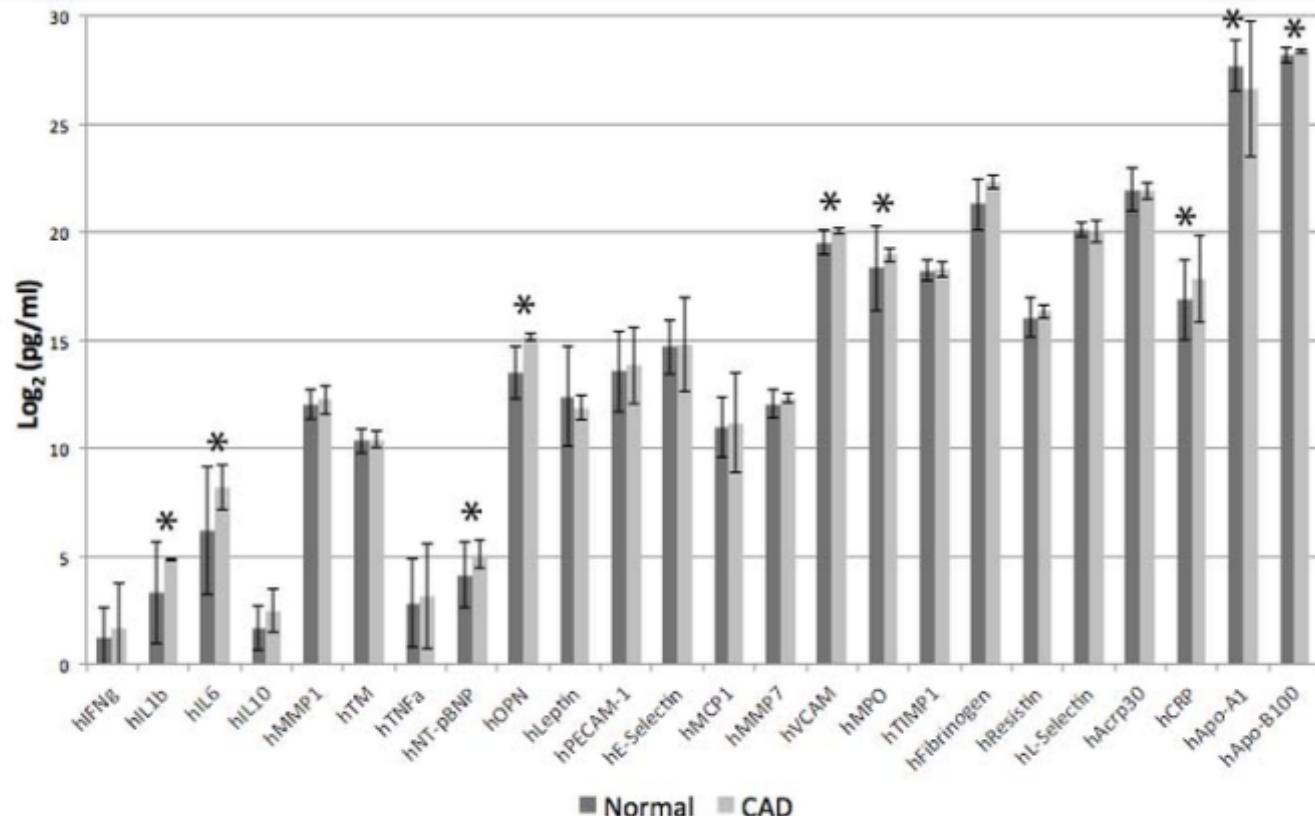
$$S(\rho) = A_0 + \sum_{i=1}^N A_i \ln(M_i(\rho) + C_i)$$

- +  $M_i(p)$  is concentration of  $i$ th marker for patient  $p$
- + Optimization of parameters using MCMC sampling
- + 58.5% specificity at 95% sensitivity (AUC = .84)

# Sample Collection

- + University of Pittsburgh, Division of Cardiology, Department of Medicine
- + Blood samples from 226 patients with CAD symptoms.
  - + 125 with significant arterial obstructions
  - + 101 patients normal
- + Blood samples were interrogated for the concentration of 24 different proteins using the Searchlight Protein Array System.

# SearchLight Proteomics Assay



**Figure 1.** Protein concentration (Mean±Std.dev.) for each of the 24 markers in  $\log_2(\text{pg/ml})$ . There are 96 samples in the normal group and 121 samples in the CAD group.

\* Comparisons with  $P < .01$  by unpaired Student's t test

# Methods/Results

- + We transformed the data into a matrix with normalized values for each marker concentration as attributes (columns) while rows represented patients.
- + Created label vector representing each patient.
- + Shuffled the data randomly and divided into two equal training and test sets.

# Methods/Results

- + We used 3 classifiers to make initial predictions
  - + Logistic Regression (Coded in Matlab)
  - + Voted Perceptron (Coded in Matlab)
  - + Support Vector Machines (built in function)
- + Compared the precision, recall, and accuracy for the 3 classifiers. Logistic Regression and Voted Perceptron were ideal.
  - + SVM: P:**0.740**, R:**0.712**, A:**0.720**
  - + Logistic Regression: P:**0.754**, R:**0.754**, A:**0.720**
  - + Voted Perceptron: P:**0.810**, R:**0.770**, A:**0.750**

# Methods/Results

- + Tested for different values of training parameters until a consistent recall of over 90% was obtained, while keeping the precision and accuracy around 65-70% for both the classifiers.
  - + Perceptron: Changed parameters so that learning algorithm penalized False Negatives more strongly than False Positives
  - + LR: Changed the classification threshold for testing so classifier only make negative when it is more certain.
- + **Logistic Regression: P:0.646, R:0.927, A:0.680**
- + **Voted Perceptron: P:0.653, R:0.925, A:0.700**

# Methods/Results

- + Performed 5-fold cross validation for each classifier (both biased and unbiased) using both biased and naive algorithms. We repeated this step 5 times and reported the results.

LOGISTIC REGRESSION

BIASED			UNBIASED		
precision	recall	accuracy	precision	recall	accuracy
0.633933	0.907093	0.65	0.802398	0.787288	0.765
0.66032	0.908	0.685	0.79197	0.769771	0.755
0.637062	0.885279	0.655	0.803929	0.803508	0.775
0.686679	0.922551	0.715	0.783382	0.733745	0.735
0.681473	0.901587	0.705	0.759973	0.824263	0.76
<b>0.6599</b>	<b>0.9049</b>	<b>0.682</b>	<b>0.78833</b>	<b>0.783715</b>	<b>0.758</b>

PERCEPTRON

BIASED			UNBIASED		
precision	recall	accuracy	precision	recall	accuracy
0.641215	0.924497	0.68	0.730392	0.770197	0.71
0.663908	0.904911	0.685	0.68694	0.789444	0.68
0.678287	0.955849	0.725	0.715241	0.745367	0.7
0.672975	0.929241	0.71	0.72221	0.774938	0.71
0.665188	0.910185	0.695	0.749579	0.809916	0.74
<b>0.664315</b>	<b>0.924936</b>	<b>0.699</b>	<b>0.731455</b>	<b>0.801835</b>	<b>0.726</b>

# Methods/Results

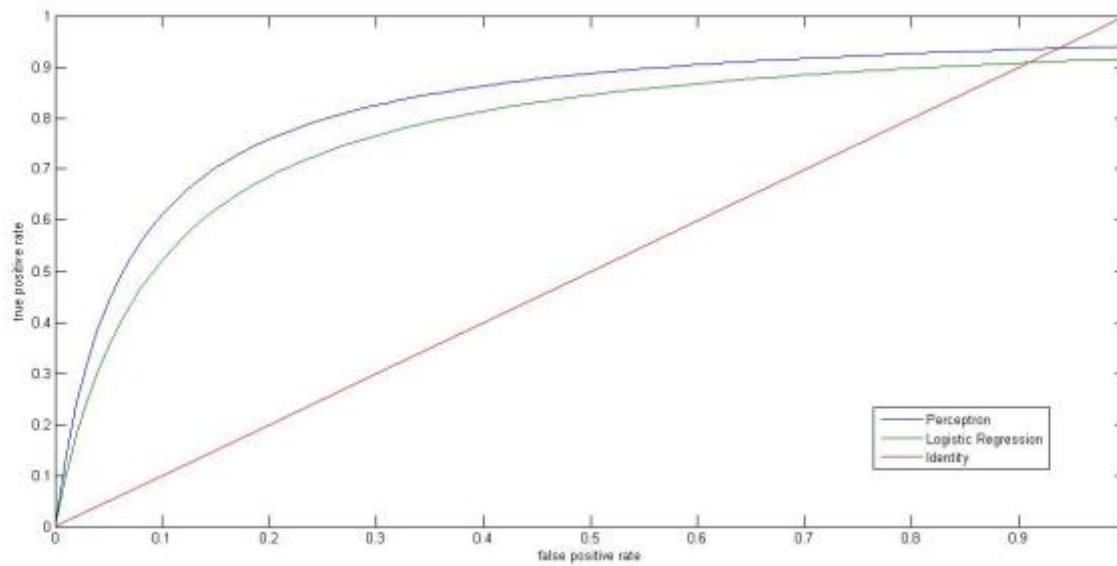
- + Comparing the performance of the Logistic Regression and Voted Perceptron Classifier
  - + We performed a paired two-tailed t-test for difference in errors by the two classifiers over same sets of data to quantify and derive a statistical significance
  - + Used 5-fold cross-validation and calculating the error in classification for each held-out set by both the classifiers.

Paired t-test with degree of freedom = 4		
Error_LR	Error_Per	Error_LR - Error_Per
0.275	0.3	-0.025
0.325	0.275	0.05
0.35	0.3	0.05
0.3	0.325	-0.025
0.35	0.375	-0.025

- + p-value = 0.846537 (doesn't tell much)

# Methods/Results

- + Next we compared performance between the two classifiers using area under their “Receiver Operating Characteristic” ROC curves.



- + AUC for LR < Perceptron and the curve also demonstrates that for the same sensitivity Perceptron gives a better specificity

# Methods/Results

- + Have been using 24 attributes/proteins for classification.
- + We wanted to find the best and smallest set of attributes that would be enough to classify for the disease.
- + Selected attributes based on p value of t-test between normal and CAD groups.
- + Groups
  - + A-24 = All 24 attributes
  - + MS-9 = 9 attributes with best p-values
  - + MS-5 = 5 attributes with best p-values
  - + MS-2 = 2 attributes with best p-values
  - + MIT-2 = 2 attributes from a previous study at MIT

# Methods/Results

- + Analyzed ROC plots of all groups using the Logistic Regression and Perceptron Classifier.

Area Under Curve		
Dataset	Logistic Regression	Perceptron
A-24	0.7728	0.8208
MS-9	0.8036	0.8708
MS-5	0.7975	0.8547
MS-2	0.771	0.8382
MIT-2	0.7705	0.8595

Table 3. Area Under Curve for different datasets

# Methods/Results

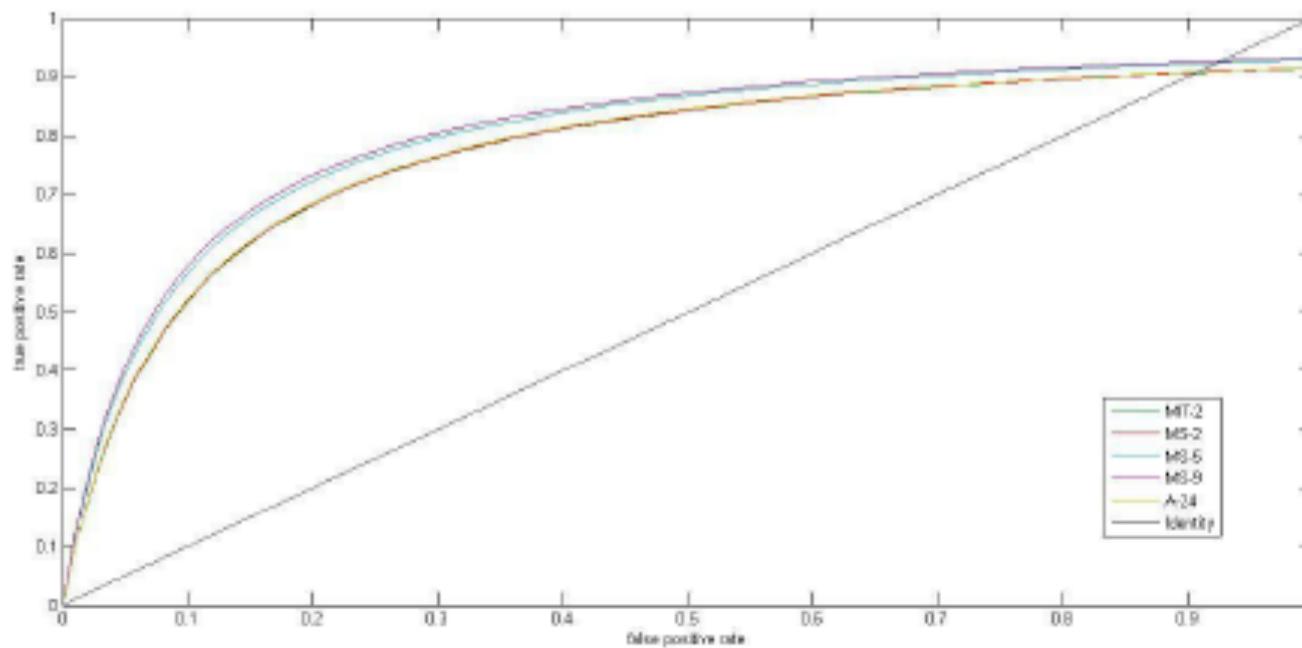


Figure 5. ROC for Logistic Regression Classifier (MIT-2, MS-2, MS-5, MS-9, A-24)

# Methods/Results

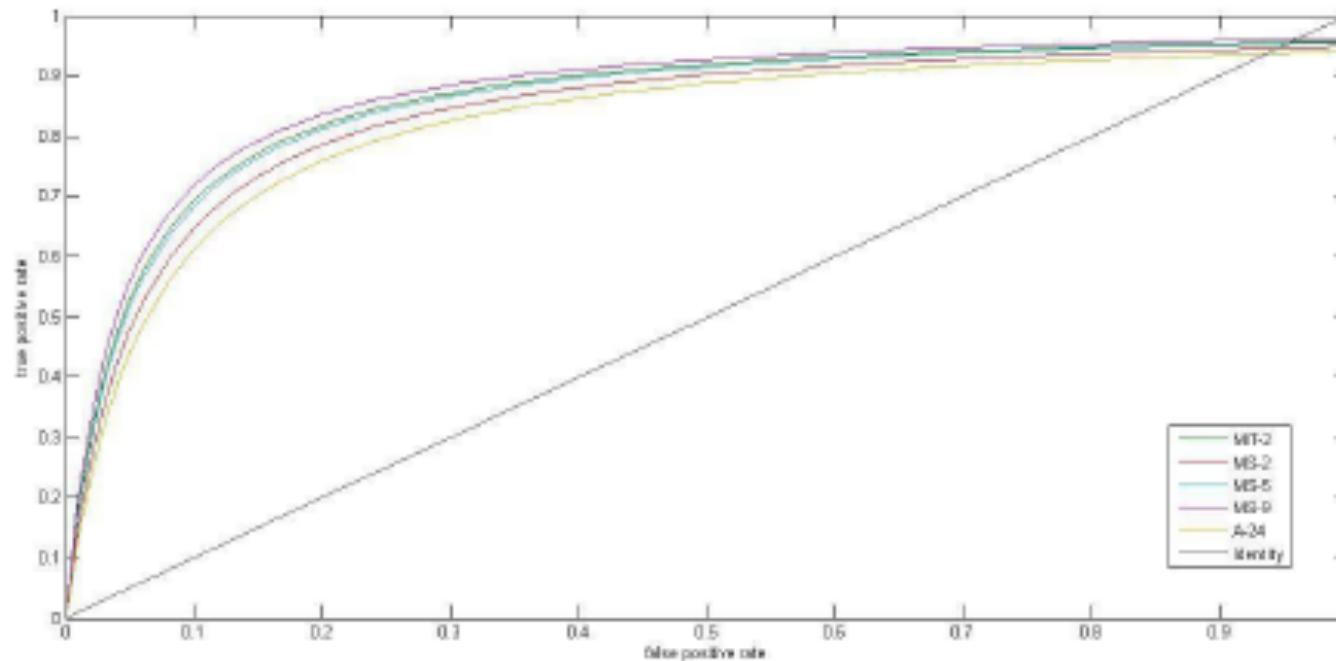


Figure 6. ROC for Perceptron Classifier (MIT-2, MS-2, MS-5, MS-9, A-24)

# Methods/Results

- + To establish significance, we repeated the process 100 times for MS-9.
- + As a control, we repeated a similar process but used 9 randomly sampled attributes for each iteration.
- + Performed a paired t-test between the MS-9 and Random 9.
  - + P-value =  $5.413E-32$
- + Ms-9 group

# Methods/Results

- + 5-fold cross validations using MS-9 for both biased classifiers.
- + Repeated 5 times and found average Precision, Recall, Accuracy.
- + Increase in Recall (95%) while Precision and Accuracy remained between 65%-70%.

# Methods/Results

5 Fold Cross Validation (Avg. 5 iterations)			
Method	Precision	Recall	Accuracy
Unbiased Logistic Regression	0.788330246	0.783714901	0.758
Biased Logistic regression	0.6599	0.9049	0.682
Biased L.R. (MS-9)	0.624	0.9497	0.655
Unbiased Perceptron	0.731454537	0.801834978	0.726
Biased Perceptron	0.664314586	0.924936471	0.699
Biased Perceptron (MS-9)	0.6499	0.9478	0.685

Table 4. Cross validation results for different classifier biasing

# Discussion

- + The best overall classification method we identified was using the perceptron classifier with the 9 proteins selected in the MS-9 panel.
- + Our MS-9 perceptron classifier outperformed the 23 gene classifier of Rosenberg in terms of overall AUC (.871 vs. .72) and specificity at 95% sensitivity (66.4% vs. ~20%).
- + Since the ultimate goal is to create a blood test for patients entering the ER with chest pain, using fewer markers could substantially reduce the cost of the test and impact its potential for widespread use.

# Discussion

- + The best sensitivity achieved, 95% by the perceptron MS-9 classifier, would still create too many false negatives to be used as a substitute for all other clinical tests.
- + The test would be used as an initial filtering method to prioritize patients being referred for angiography and false negatives would be avoided by augmenting the test with the routine methods of investigation of family and medical history, Electrocardiograms (EKG), Stress testing, Echocardiography (ECHO), other blood tests, and Electron-Beam Computed Tomography.

# Discussion

- + The best sensitivity achieved, 95% by the perceptron MS-9 classifier, would still create too many false negatives to be used as a substitute for all other clinical tests.
- + The test would be used as an initial filtering method to prioritize patients being referred for angiography and false negatives would be avoided by augmenting the test with the routine methods of investigation of family and medical history, Electrocardiograms (EKG), Stress testing, Echocardiography (ECHO), other blood tests, and Electron-Beam Computed Tomography.

# Conclusions

	AUC	Sensitivity	Specificity
MS-9 perceptron	.871	95%	66.4%
Aleksey Lomakin 4 protein panel	.84	95.7%	58.4%
Rosenberg 23 gene	.721	95%	~20%

# Conclusion

- + To our knowledge, our 9 protein perceptron classifier is the only protein CAD classifier capable of achieving 65% specificity while maintaining 95% sensitivity.
- + This supports the idea that a low cost, low risk blood test for CAD could be implemented as a clinically useful tool in the near future.