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

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

## Abstract

**Introduction:** Diagnosis of coronary artery disease (CAD) by coronary angiography, the current gold standard for CAD detection, results in over 500,000 negative tests per year in the US alone. Better classification of patients by non-invasive methods prior to angiography would reduce the number of patients having to undergo the invasive and costly angiography.

**Objective:** The aim of the current project is to improve upon the algorithms used in previous studies to discern patients at low-risk for CAD.

**Methods:** Blood samples were collected at the University of Pittsburgh from 226 patients who entered the emergency room with chest pain symptoms and were referred for coronary angiography. Samples were interrogated for the concentration of 24 different cardiovascular-related proteins and the data was separated into two groups based on whether the patients were diagnosed with coronary artery disease or released without need of further treatment. Based on the protein data, we developed logistic regression and voted perceptron classifiers tuned to achieve the greatest specificity while maintaining greater than 90% sensitivity and evaluated their performance by ROC analysis.

**Results:** Using the full set of 24 protein markers, the logistic regression classifier (AUC = .7728) and the perceptron classifier (AUC = .8208) had comparable performance with 66% specificity at 90.5% sensitivity and 66.4% specificity at 92.5% sensitivity respectively. However, using the perceptron classifier with a panel of 9 proteins (IL1b, IL6, NT-pBNP, OPN, VCAM, MPO, CRP, Apo-A1, Apo-B100) resulted in the best overall classification, yielding .871 AUC and 65% specificity at 95% sensitivity.

**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.

# 1. Introduction

Heart disease/Coronary Artery Disease (CAD) is the number one cause of death globally, and causes over 800,000 deaths per year in the US [1]. The current standard for evaluating patients presenting CAD symptoms is to perform coronary angiography by cardiac catheterization [2]. This procedure involves inserting a catheter into the arm and threading it into the coronary arteries. The catheter then releases dye into the bloodstream and x-rays are used to visualize the dye traveling through the arteries [2]. Unfortunately, CAD does not have explicit symptoms and only about half of the patients recommended to undergo angiography are found to have CAD [3]. Therefore, over 500,000 people per year in the US are unnecessarily subjected to a procedure which is invasive, expensive, and exposes them to ionizing radiation and contrast media (dye). If there were non-invasive diagnostic tools that could identify a fraction of patients presenting CAD symptoms as non-risk (negatives), then we could reduce the economic and medical burden created by subjecting patients with normal coronary arteries to unnecessary angiographies.

There have been many recent attempts at identifying CAD using mRNA or microRNA signatures [6,10]. These studies have yielded interesting findings, such as the correlation of three microRNAs (miR-134, miR-198, and miR-370) with CAD [5] and a 23 gene classifier for obstructive CAD [9]. However, these results have yet to be implemented clinically, largely due to the fact that the measurement of mRNA and microRNA requires PCR amplification and is limited to relative comparisons. Without the ability to establish an absolute standard for the expression level that constitutes a patient as being at risk for CAD, the clinical extensibility is limited. Alternatively, proteins are present at abundant levels in the blood, allowing for direct measurement and absolute quantification using spike-ins. This advantage is evidenced by the widespread clinical use of proteins such as troponin T to identify myocardial infarction [8] and C-reactive protein to predict future cardiovascular events [7], but there has not yet been a protein biomarker identified for predicting CAD.

In the current project, we combine the strong point of gene expression, assessing multiple pathways to increase chances of identifying a complex disease, with the strong point of protein biomarker tests, a more robust and potentially diagnostic test, and create a CAD classifier based on a panel of protein biomarkers. We begin by using the logistic regression method employed by Rosenberg et al. [9] and then compare it to other means of classifying, such as voted perceptron and support vector machines. Since we aim to rule out CAD for a percentage of patients and save them from having to undergo angiography, the best classifier is the one that identifies the greatest number of true negatives while minimizing the number of false negatives. A non-invasive biomarker test would be desirable even if it ruled out CAD for a small percentage of patients because the cost and risk involved in a multiplex blood test is fractional in comparison to angiography.

# 2. Methods

## 2.1 Samples

Oscar Marroquin, Suresh Mulukutla, and Dennis McNamara at the University of Pittsburgh, Division of Cardiology, Department of Medicine, collected blood samples from 226 patients who were referred to them with CAD symptoms. Angiography revealed that 125 of the patients had significant arterial obstructions and required further interventional therapy, while the other 101 patients appeared normal. The data is separated into two groups, positive and negative, according to these results. Other clinical risk factors such as tobacco use, diabetes, cholesterol level, current medications, age, sex, race, and body mass index were also included with the samples.

## 2.2 Proteomics Assay

Blood samples were interrogated for the concentration of 24 different proteins (IFNg, IL1b, IL6, IL10, MMP1, MMP7, TM, TNFa, NT-pBNP, OPN, Leptin, PECAM-1, E-Selectin, MCP1, VCAM, MPO, TIMP1, Fibrinogen, Resistin, L-Selectin, Acrp-30, CRP, Apo-A1, Apo-B100) using the Searchlight Protein Array System. This system measures absolute quantities using a five-point standard curve so there is no need for across batch normalization. For a detailed description of the proteomics assay, see LaFramboise et al. [4].

At this point, our group received the data from the University of Pittsburgh in the form of the concentrations of 24 proteins for all 226 patients. The only alterations to the data were averaging the two replicates that existed for each protein concentration and thresholding the proteins that existed in concentrations beyond the range of the standard curve. This dataset, comprised of 226 patients with clinical information and concentrations of 24 proteins, can be found in the supplementary file “data.xlsx.”

## 2.3 Algorithm Development

The details of the protocol we followed are as follows:

1. We transform the data into a matrix where each row represents a patient and each column entry is an attribute representing a normalized concentration for each marker. We also created a vector containing the known labels for each patient in the same order as the attribute matrix.

2. The normalized values are calculated by dividing marker concentration in each cell against the median value for that marker’s concentration across the column.

3. We randomly shuffled the patients in the attributes and labels making sure that their relative orders are unchanged. Then we divided the attributes and labels into two equal training and test sets.

4. We used 3 model classifiers (Logistic Regression, Voted Perceptron and Support Vector Machines) to calculate the accuracy of predictions on the test set after training them with the training set using 3 classifiers. We wrote a matlab code for classification using logistic regression and voted perceptron.

5. We predict the labels as:
    a. Logistic function for logistic regression: y = P(x) = 1/(1+e^(-wx)), where w is the trained classifier
    b. Sigmoid function for perceptron: *y* = sign(v.x), where v is the trained classifier.

6. For the SVM classification, we used the built-in MATLAB SVM function with a linear as well a Gaussian kernel function [11]. Tuning the parameters for each of the functions did not yield satisfactory results for precision and recall. The linear kernel function gave stable results with recall ~0.7 while the Gaussian function performed with high instability with recall ranged from 0.4 to 1.0 (where it classified everything as N).

7. We compared the precision, recall and accuracy for the 3 classifiers from the above training and test sets. (all values averaged over 5 runs)

| | |
|---|---|
| **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** |

8. Since our goal was to generate a classifier that gives the greatest number of true negatives while minimizing the number of false negatives, we modified our learning algorithm in the voted perceptron and the classification threshold in the logistic regression classifier to create a bias against false negative predictions. We manually tested the parameters until a consistent recall of over 90% was obtained while keeping the precision and accuracy around 65-70% for both the classifiers. (This improvement in accuracy did come about at the expense of some decrease in precision and recall, but we still did manage to get a significant value for both).

| | |
|---|---|
| **Logistic Regression:** | **P:0.646, R:0.927, A:0.680** |
| **Voted Perceptron:** | **P:0.653, R:0.925, A:0.700** |

*(all values averaged over 5 runs)*

9. Since the SVM did not improve the recall without maintaining significant precision and accuracy, we dropped the model from further analysis.

10. In the next step we sought to find the robustness of the two remaining classifiers. We performed a 5-fold cross validation for each classifier using both biased and naive algorithms. We repeated 5-fold cross-validation 5 separate times and noted the average values of precision, recall and accuracy for each classifier.

**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 |
| **0.6599** | **0.9049** | **0.682** | **0.78833** | **0.783715** | **0.758** |

**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 |
| **0.664315** | **0.924936** | **0.699** | **0.731455** | **0.801835** | **0.726** |

**Table 1.** Results of biased and unbiased classifiers.

11. The results from the 5-fold cross-validation clearly suggest that the classifiers are consistent and robust enough for both biased and naive classifications and there is a clear improvement in the recall after adding a bias (with some decrease in precision and accuracy). We then set out to further evaluate and compare the classifiers using a "Receiver Operating Characteristic" (ROC) curve. We generated each point in the plot by changing the threshold for classification from a complete bias for 'positives' to a complete bias for 'negatives' during the testing operation for both the voted perceptron and logistic regression classifiers. We calculated true positive rate and false positive rate for each value of the threshold and came up with 20 points that were used to build the ROC curve using matlab. We then calculated the area under the curve for each classifier to measure their performance. The area under curve for both exceeded 0.5.

    **LR = 0.7728 Perceptron = 0.8208**

12. To compare the performance between the two classifiers, we compare the area under their ROC curves. The area under the voted perceptron was found to be slightly better than the logistic regression classifier. To further quantify and derive a statistical significance, we used paired two-tailed t-test for difference in errors by the two classifiers over same sets of data. We did this by again using 5-fold cross-validation and

calculating the error in classification for each held-out set by both the classifiers. We then calculate the difference in error by both the classifier for all five held-out sets and use this to calculate a p-value from paired two-tailed t-test. The null hypothesis (H0) used here was that the E-value of error difference should be zero. Under this hypothesis we get a p-value of **0.846537,** which shows that the current null hypothesis is highly likely and we cannot say with confidence that one classifier always performs better than the other. This can also be inferred from the average 5-fold cross validation results for P,R and A which are also very similar.

**Paired t-test (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**

**Table 2.** Classifier performance comparison by a paired t-test

13. Next, we set out to look at the best set of attributes that should be enough to classify for the disease. We have been using all 24 attributes (A-24) for classification up to this point. We selected attributes based on significance value of student's t-test between normal and CAD groups. We selected the best 9 attributes (MS-9) based on highest p-values and plotted a ROC curve for both classifiers with them. We then did a similar analysis with best 5 (MS-5) attributes, best 2 (MS-2) attributes, and 2 attributes established from study (MIT-2) [4]. We observed that the maximum area under the curve was obtained using best 9 attributes in both the classifiers (refer figure ****).

14. In order to establish the statistical significance of the result, we repeated this process 100 times for the best 9 attributes and obtained the mean and maximum value for the area under the ROC curve. As a control we also repeated a similar process 100 times, but this time area was calculated using 9 randomly selected attributes for each iteration. Both, mean and maximum values for area under the ROC curve were found lesser in case of control samples.

15. We further performed a paired t-test between the area obtained using the 9 best attributes and using 9 randomly selected attributes over the 100 experiments. We obtained a p-value of **5.413E-32**, showing that the 9 best attributes performed consistently better than 9 randomly selected samples, signifying their importance in this classification.

16. Finally, we calculated the average Precision, Recall and Accuracy by 5 fold cross-validations using the 9 best attributes and repeated this procedure 5 times for both

biased classifiers. The final average P,R and A was calculated and compared to the 5-fold CV values obtained for both biased classifiers with all 24 attributes. We found a further increase in Recall to 95% with the Accuracy and Precision still between 65-70%. This improved our performance as it further minimized the False negatives, still giving a decent number of True negatives on using the best 9 attributes.

**Perceptron (A-24) = P:0.653, R:0.925, A:0.700**
**Perceptron (MS-9) = P:0.6499, R: 0.9478, A: 0.685**

**Logistic Regression (A-24) = P:0.646, R:0.927, A:0.680**
**Logistic Regression (MS-9) = P:0.624 , R:0.9497, A:0.655**
*(all values averaged over 5 runs)*

The Matlab code for creating and testing each classifier can be found in the supplementary zip file "Matlab Codes."

## 3. Results

Blood samples from patients who underwent coronary angiography are divided into two groups, CAD and Normal, based on whether the angiography revealed coronary artery disease requiring further intervention (CAD), or normal blood flow requiring no further treatment (Normal). Among the 24 protein markers, we found there were 9 proteins (IL1b, IL6, NT-pBNP, OPN, VCAM, MPO, CRP, Apo-A1, Apo-B100) that were significantly different ($P < .01$, unpaired Student's t test ) between groups (Figure 1).

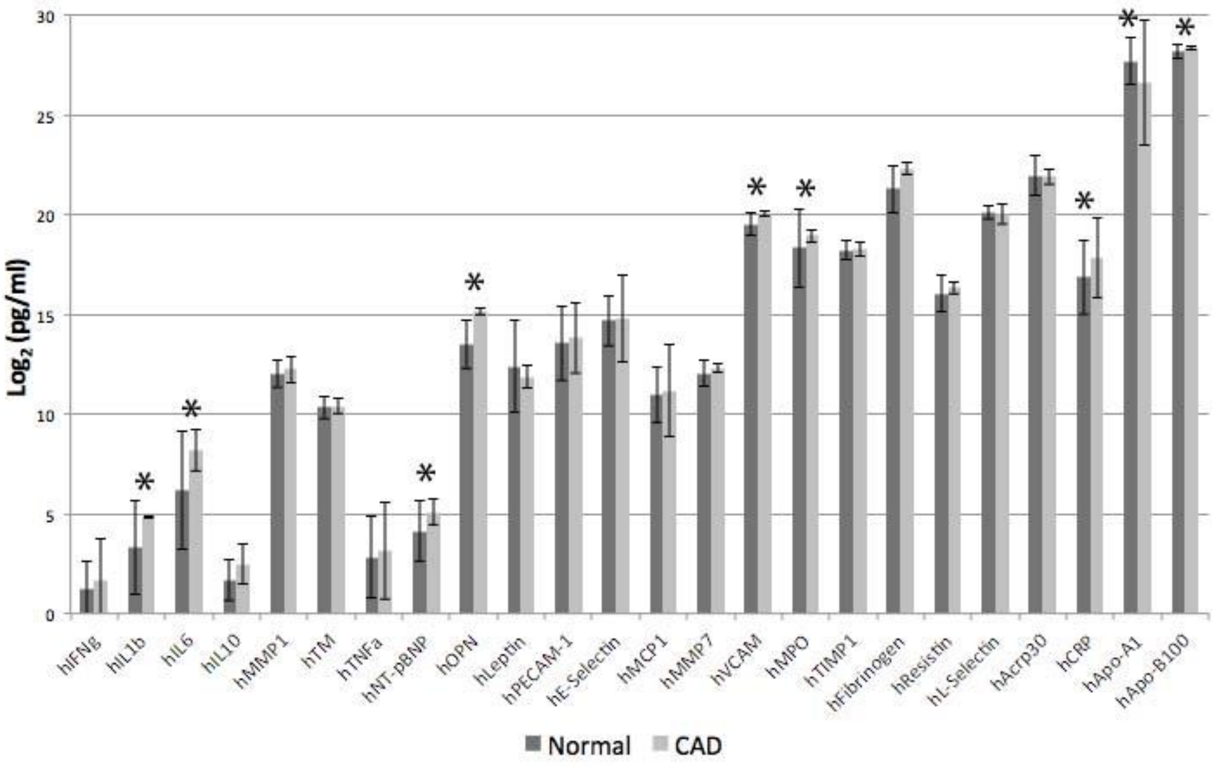

**Figure 1.** Protein concentration (Mean±Std.dev.) for each of the 24 markers in $\log_2$(pg/ml). There are 96 samples in the normal group and 121 samples in the CAD group. *Comparisons with P < .01 by unpaired Student's t test*

While there was a group of significant markers, there was not a single individual marker that was able to discriminate between the groups. For example, the most significant marker, OPN (P < .0001), ranged from 1138.8 to 118203.2 pg/ml in the normal group and from 2952.3 to 204887.6 pg/ml in the CAD group. This overlapping interval, 2952.3 to 118203.2, contains 93.5% of all samples, making it possible to only definitively classify 5 samples as Normal and 9 samples as CAD. However, receiver operating characteristics (ROC) indicate that logistic regression and voted perceptron methods of classification are able to effectively classify Normal and CAD samples using all 24 markers (Figure 4).

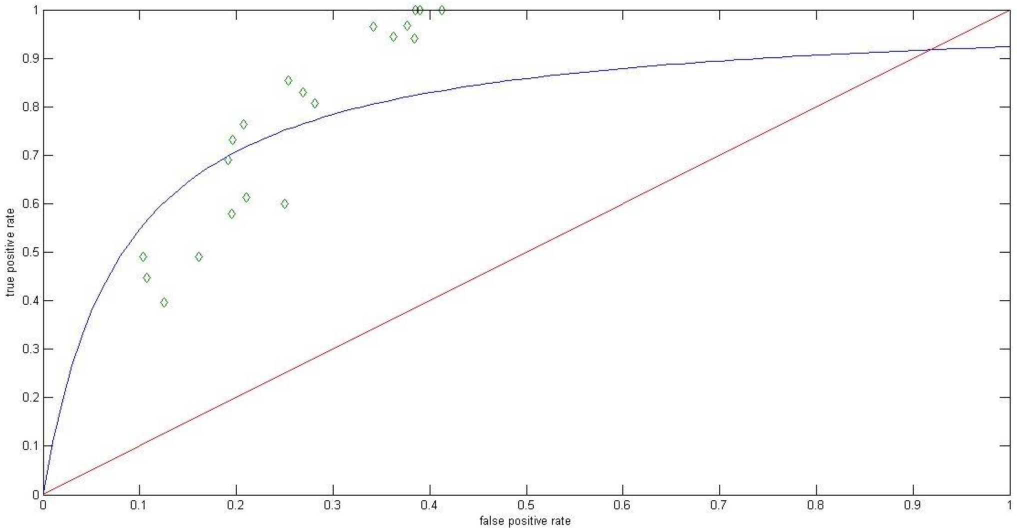

**Figure 2.** Logistic Regression ROC

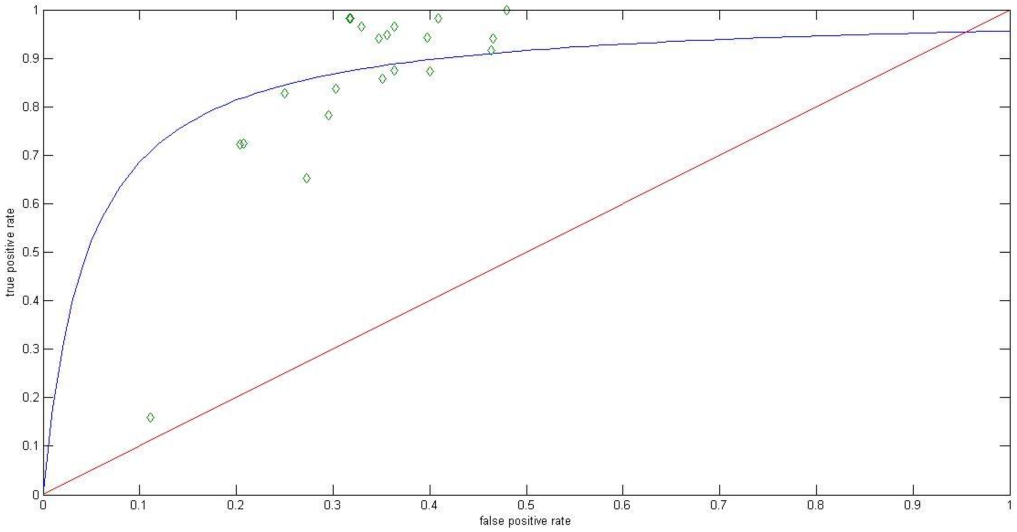

**Figure 3.** Perceptron ROC

**Figures 2,3.** “Receiver Operating Characteristic” (ROC) curve for Logistic regression and Voted perceptron . Each point in the plots was obtained by changing the threshold for classification from a complete bias for ‘positives’ to a complete bias for ‘negatives’ during the testing operation for both classifiers.

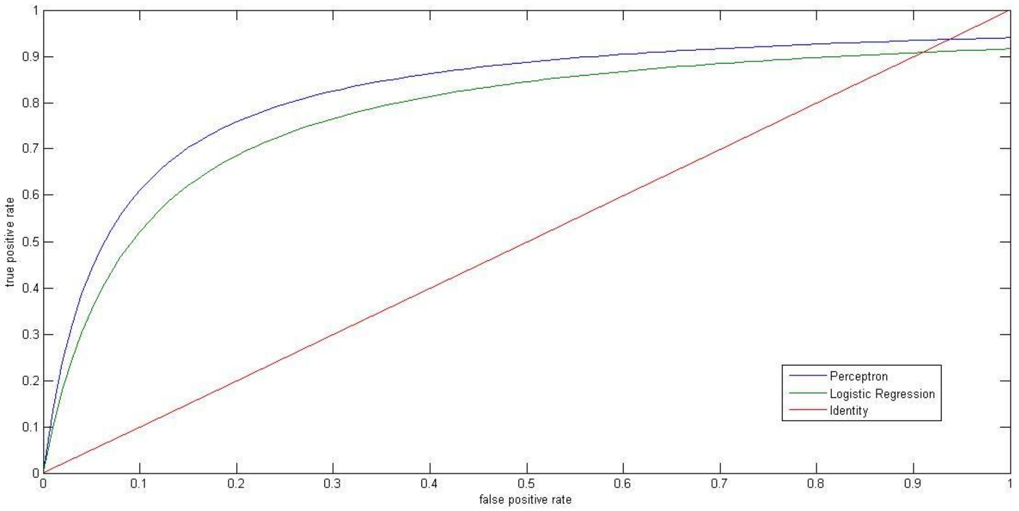

**Figure 4.** ROC for the perceptron and logistic regression classifiers using all 24 protein markers. The logistic regression (AUC = .7728) achieved specificity of .6599 at .9049 sensitivity and the perceptron (AUC = .8208) achieved specificity of .6643 at .9249 sensitivity.

Using 24 markers, the perceptron classifier had a greater area under the curve (AUC = .8208) than the logistic regression classifier (AUC = .7728) (Figure 2). When interrogating the specificities of the classifiers at high sensitivities (true positive rate > .9) necessary to make the test clinically useful, we found that the specificity of the perceptron classifier (.6643) was greater than that of the logistic classifier (.6599). These curves were generated using 5 fold cross validation for 5 trials and averaging the results.

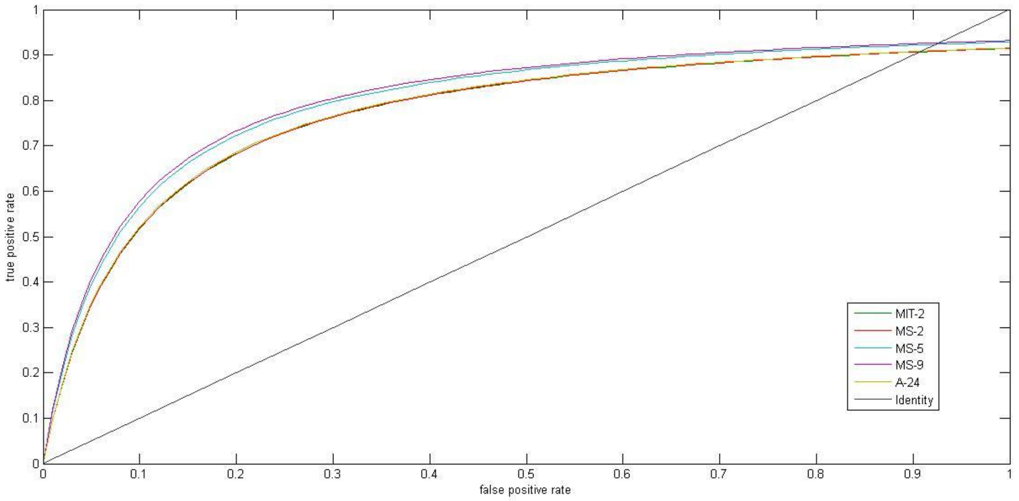

**Figure 5.** ROC for Logistic Regression Classifier (MIT-2, MS-2, MS-5, MS-9, A-24)

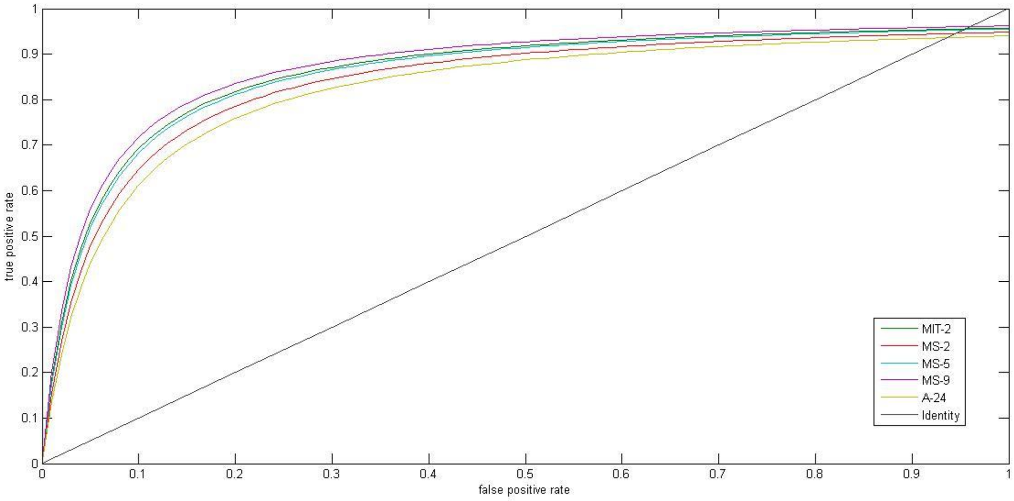

**Figure 6.** ROC for Perceptron Classifier (MIT-2, MS-2, MS-5, MS-9, A-24)

**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

**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

# 4. Discussion

Each of the 24 proteins tested were originally selected for having a known cardiovascular-related function or for being currently used individually, such as NT-pBNP and CRP, as risk factors for cardiovascular events. We found that each of our selected panels (A-24, MS-9, MS-5, MS-2) using either classification method, logistic or perceptron, do a better job classifying the Normal and CAD patients than any of the 24 proteins can do alone. The best overall classification method we identified was using the perceptron classifier with the 9 proteins selected in the MS-9 panel. However, while the perceptron achieved a higher AUC than logistic regression on average, individual results depict more variability in performance for perceptron model. Therefore, more samples are required to delineate which classifier works better across samples from multiple locations.

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. With that in mind, we tried to reduce the number of markers as much as possible without losing substantial sensitivity and specificity. We found that reducing the number of markers from 9 to 5 in panel MS-5, caused a decrease in the AUC (.871 to .855 using perceptron and .804 to .798 using logistic) but these values were comparable enough that if it reduced the cost of the test, it might be worth the small loss in accuracy. However, decreasing the number of protein markers further from 5 to 2, as seen in MS-2, resulted in a drop off in classification accuracy that more closely resembled what we see using all 24 markers (Table 3).

The modification of classification algorithms to be biased towards calling false positives rather than false negatives proved to be critical if either of these algorithms is ever to be successful clinically. Using the biased algorithms we optimize the accuracy while holding the sensitivity at over 90%. Admittingly, 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. Therefore, in the present day scenario, 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. Used in this way, a 95% sensitivity test could still be a highly beneficial addition to a cardiologist's toolbox.

Our MS-9 perceptron classifier outperformed the 23 gene classifier of Rosenberg et al. [9] in terms of overall AUC (.871 vs. .72) and specificity at 95% sensitivity (66.4% vs. ~20%). The use of stable and inexpensive to quantify proteins in our classifier rather than rapidly degrading mRNA also makes our classifier more likely to provide robust results in a clinical environment. We were also able to improve upon the protein classifier of Laframboise et al. [4], where they observed a maximum AUC of .85 using a 3 marker panel (OPN, Resistin, Apo-B100) or a maximum specificity of 58.5% at 95% using a 4 marker panel (IFNgamma, OPN, MMP7, MPO). We were able to increase both the AUC and specificity using our biased perceptron classification versus their MCMC sampling technique. However, we are using a different dataset than Laframboise et al. and to make a final decision about which classifier is better, we would have to run them both on the same dataset. The comparison of our classifier to the Rosenberg et al. study is also approximate because their study was over multiple locations and our accuracy is likely to drop if we were to also include multiple locations. Regardless, we have accomplished our goal of improving upon the classification of these studies given our data.

The result of the current project strongly supports the notion that a multiplex protein classifiers can be used to accurately discern patients at low-risk for CAD from those at high-risk that should undergo coronary angiography. Given more time and resources we could evaluate the potential of other classifiers and further refine our current model. We initially tested a SVM classifier, but the number of parameters was excessive for the amount of data we have. However, given larger datasets it may be worth investigating the fine tuning of SVM's parameters to achieve greater precision while maintaining recall. Another possibility is to transform our classification system from outputting a binary result to outputting multiple risk levels that physicians could incorporate into their decision making, similar to the Framingham Coronary Heart Disease Risk Scores. Given more data we could also better identify the confidence with which a particular patient is classified as positive or negative and differentiate more clearly between algorithms. As the ageing and obesity demographics continue to grow, answering these questions will likely prove to be highly beneficial in reducing the clinical burden of diagnosing coronary artery disease.

## 5. References

1. World Health Organization, **Cardiovascular Diseases Fact Sheet.** (http://www.who.int/mediacentre/factsheets/fs317/en/)

2. National Heart, Lung, and Blood Institute. (http://www.nhlbi.nih.gov/health/health-topics/topics/ca/)

3. Patel MR, Peterson ED, Dai D, Brennan JM, Redberg RF, Anderson HV, Brindis RG, Douglas PS: **Low diagnostic yield of elective coronary angiography.** *N Engl J Med* 2010, 362:886-895.

4. LaFramboise WA, Dhir R, Kelly LA, Petrosko P, Krill-Burger JM, Sciulli CM, Lyons-Weiler MA, Chandran UR, Lomakin A, Masterson RV, Marroquin OC, Mulukutla SR, McNamara DM. **Serum protein profiles predict coronary artery disease in symptomatic patients referred for coronary angiography**. *BMC Med*. 2012 Dec 5;10:157. doi: 10.1186/1741-7015-10-157.

5. Hoekstra M, van der Lans CA, Halvorsen B, Gullestad L, Kuiper J, Aukrust P, van Berkel TJ, Biessen EA. **The peripheral blood mononuclear cell microRNA signature of coronary artery disease**. *Biochem Biophys Res Commun.* 2010 Apr 9;394(3):792-7.

6. Kinet V, Halkein J, Dirkx E, Windt LJ**. Cardiovascular extracellular microRNAs: emerging diagnostic markers and mechanisms of cell-to-cell RNA communication**. *Front Genet.* 2013 Nov 12;4:214.

7. Zebrack JS, Muhlestein JB, Horne BD, Anderson JL; **Intermountain Heart Collaboration Study Group. C-reactive protein and angiographic coronary artery disease: independent and additive predictors of risk in subjects with angina**. *J Am Coll Cardiol.* 2002 Feb 20;39(4):632-7.

8. Collinson P, Gaze D, Goodacre S. **Comparison of contemporary troponin assays with the novel biomarkers, heart fatty acid binding protein and copeptin, for the early confirmation or exclusion of myocardial infarction in patients presenting to the emergency department with chest pain**. *Heart.* 2013 Nov 22. doi: 10.1136/heartjnl-2013-304716.

9. Rosenberg S, Elashoff MR, Beineke P, Daniels SE, Wingrove JA, Tingley WG, Sager PT, Sehnert AJ, Yau M, Kraus WE, Newby LK, Schwartz RS, Voros S, Ellis SG, Tahirkheli N, Waksman R, McPherson J, Lansky A, Winn ME, Schork NJ, Topol EJ; **PREDICT (Personalized Risk Evaluation and Diagnosis in the Coronary Tree) Investigators. Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients.** *Ann Intern Med.* 2010 Oct 5;153(7):425-34.

10. Hsu J, Smith JD. **Genome-wide studies of gene expression relevant to coronary artery disease**. *Curr Opin Cardiol.* 2012 May;27(3):210-3.

11. Hongzong S, Tao W, Xiaojun Y, Huanxiang L, Zhide H, Mancang L, BoTao F. **Support vector machines classification for discriminating coronary heart disease patients from non-coronary heart disease.** *West Indian Med J.* 2007 Oct;56(5):451-7.