Prediction of Atrial Fibrillation following Cardiac Surgery using Rough Set Derived Rules

Matthew C. Wiggins[1], Hiram A. Firpi[1], *Member, IEEE*, Raul R. Blanco[2], Muhammad Amer[2], and Samuel C. Dudley, Jr.[2]

[1]Georgia Institute of Technology, Atlanta, GA · [2]Emory University and the Veterans Affairs Medical Center, Atlanta, GA

Abstract—Atrial fibrillation (AF) and flutter are common following cardiac surgery, increasing costs and morbidity. Cardiologists need a method to discern those patients who are at high risk for this arrhythmia in order to attempt to treat them by either pharmacologic or non-pharmacologic means. We performed a retrospective analysis of 377 CABG patients, of which 94 developed AF post-operatively. Feature selection and AF occurrence prediction was performed using a multivariate regression model, and two rough set derived rule classifiers. The rough set derived feature subset performed best with an accuracy of 87%, a sensitivity of 58.5%, and a specificity of 96.5%. This shows the importance of testing feature subsets, thereby discouraging the practice of simply combining the best individual predictors. The utility of rough set theory in prediction of cardiac arrhythmia is also validated.

INTRODUCTION

Roughly thirty percent of cardiothoracic surgery patients develop atrial fibrillation (AF) or flutter prior to discharge, increasing the risk of stroke, prolonging hospital stay, and increasing the overall cost of the procedure. According to some sources, over $1 billion is spent annually on this problem in the US alone. Current pharmacologic and non-pharmacologic means of AF prevention are suboptimal, and their side effects, expense, and inconvenience limit their widespread use in all patients.

A method for the identification of those patients who are at highest risk for AF onset is greatly needed. This would allow application of the prophylactic measures to the subset at risk, preventing low risk patients from suffering unwanted side effects and reducing overall costs.

Traditionally, the clinical profession has used univariate statistics to analyze demographic, pre-operative (pre-op), operative, and post-operative (post-op) data to find predictors of AF following cardiac surgery. Often, these studies make no attempt to validate their predictive model nor offer an accuracy measure. Overall, previous methods have yielded somewhat disappointing results with limited applicability to the clinical setting. Therefore, the best predictive model parameters are still unclear. Using a database of 377 consecutive cardiac surgery patients, we compare the standard multivariate analysis of the clinical data with a set of rules developed through the use of rough set theory to predict the incidence of post-op AF.

METHODOLOGY

Patient Population

Data was obtained for all patients who underwent cardiothoracic surgery at the Atlanta Veterans Affairs (VA) Medical Center between January 2000 and March 2005 under a protocol approved by the Emory University IRB. The diagnosis of post-op AF (94 of 377 subjects) was based on review of notes and ECGs.

Eighty-eight variables were collected including patient demographics, current medical conditions, pre- and post-op medications, echocardiogram results, electrocardiogram results, coronary angiogram, and pre-op laboratory results including the use of pre-op peroxisome proliferator activator receptor γ (PPAR) agonists, a patient's NYHA functional class, the European Society of Cardiology's SCORE risk measure, and the physician's estimate of operation mortality.

TABLE I
SAMPLE CHARACTERISTICS WITH A P-VALUE ≤ 0.1

Characteristic	No AF (n=283)	AF (n=94)	Total (n=377)	p-value
Age (years)	61.1 ± 9	65.9 ± 9	62 ± 9	1.8E-5
Body Surface Area (m^2)	2.03 ± 0.19	2.06 ± 0.165	2.04 ± 0.18	0.10
PPAR (%)	0.4	2.1	0.8	0.09
Pre-op AF/AFL (%)	1.1	10.6	3.5	1.1E-5
Employment Status (%)				0.08
Full-time	24.5	16	22	
Part-time	1.8	2.1	2	
Retired	42.6	44.7	43	
Unemployed	5.4	5.3	5	
Other	25.6	31.9	27	
Cardiomegaly (%)	13.4	25.5	18.8	0.0061
Current Smoker (%)	41.3	26.6	37.7	0.011
Smoking History (%)				0.0089
1	23.6	19.2	22	
2	45.3	28.8	41	
3	4.7	8.2	6	
4	26.4	43.8	31	
Prior heart surgery (%)	1.1	8.5	2.9	0.0002
Functional Class (%)				0.04
I	46.3	54.3	48	
II	28.3	29.8	29	
III	20.5	13.8	19	
IV	4.9	2.1	4	
Digoxin Use (%)	2.5	6.4	3.4	0.07
Left Anterior Descending Stenosis	71.2 ± 28.4	64.4 ± 31.7	69.4 ± 29.4	0.07
MV Regurgitation (%)				0.04
None	81.9	68.2	78	
Mild	12.6	22.4	15	
Moderate	2.9	4.7	3	
Severe	2.5	0	3	
Physician Risk Estimate	4.34 ± 2.31	5.05 ± 3.06	4.52 ± 2.53	0.039
Mammary Artery Usage	90.1	81.9	88.1	0.034
Complications (%)				0.043
None	33.9	15.4	29.1	
One	60.7	79.5	65.6	
Two	5.4	5.1	5.3	

Data Analysis

Characteristics of the study groups were compared using student t-tests for the continuous variables and Chi-square tests for the binary variables. The variables having a p-value < 0.1 and their characteristics are listed in Table I. All continuous variables are presented as mean plus or minus the standard deviation. Discrete variables are presented in the percentage of the population having a specific value.

Missing data points were filled using the conditioned mean/mode tool in the ROSETTA software allowing probabilities of the existing data to be used in the completion process. Several variables, listed in Table II, were discretized manually based on heuristically determined thresholds. **Following univariate analysis, variables with a p-value < 0.1 were placed into a multivariate regression model** to predict the occurrence of AF.

TABLE II
MANUAL DISCRETIZATION FOR UNIVARIATE SIGNIFICANT VARIABLES

Characteristic	Discretization
Age	[40,50), [50,60), [60,70), [70,80), [80,90)
Body Surface Area	[0,1.8), [1.8,2.2), [2.2, ∞)
Employment Status	1, [2,3], [4,5]
Smoking History	[1,2], [3,4]
Functional Class	[1,2], 3, 4
LAD Stenosis	[0,50), [50, 75), [75, 100]
Physician Risk Estimate	[1,2], 3, [4,6], [7,15]

Rough Set Theory

Rough sets theory, proposed by Z. Pawlak early in the 1980s, is a paradigm that can capture vagueness and uncertainty in a given data set. Rough sets have been applied to many different problems from medical prognostics to mechanical fault diagnostics. Although, rough set theory has its own terminology and background theory, if viewed from a pattern recognition standpoint, rough sets can be seen as a feature selector and classifier where the objective is a subset of features that can discriminate the patients that will develop AF. From this subset of features, a classifier is designed.

Ideally, we are searching for the minimal subset of attributes that can categorize the objects correctly, a subset referred to as a reduct. For a complex problem, there may be many of these minimal reduct sets. Once reducts are obtained, a set of *if-then* rules can be set to create a classifier.

For the experiments, we used the ROSETTA software, C++-based software designed by A. Øhrn. The program has embedded several routines to dicretize the attributes, find the reducts, and filter variables, reducing the number of rules produced at the end of the evaluation. It also allows the flexibility to validate the classifier using a validation data set. Although, there are well-established procedures to find the reducts using a discernibility matrix and function in this work, we use a genetic algorithm, given that the number of possible reducts is approximately 1.3×1025.

Rough Set Experiments

We perform two experiments using rough set theory to compare with the multivariate regression model. **The first experiment (Exp. I) uses the sample characteristics, which were found to have a p-value < 0.1** for the initial set of attributes, *A*. **The second experiment (Exp. II) uses all the characteristics from the database**. The same parameters were used for the genetic algorithm including a hitting set fraction of 0.8. Any discretizations besides those in Table II were done using entropy based binning as performed in the ROSETTA software. Rough set theory was then used to find the best subset and the resulting rules.

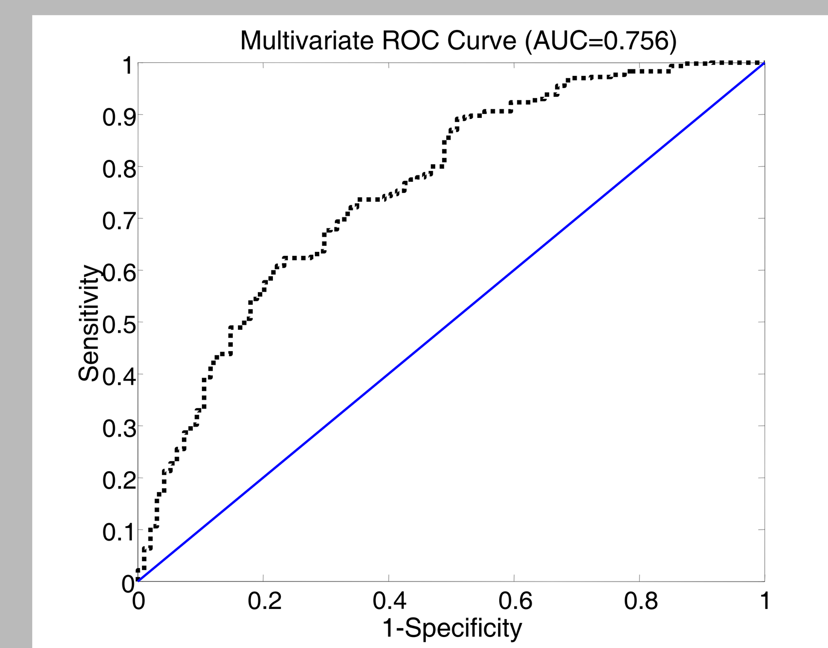
TABLE III
PROMINENT RULES AND ASSOCIATED AF OCCURRENCES FOR EXPERIMENT II

Total Cholesterol	*Smoking History*	*CAD*	*SCORE*	*M. Valve Regurg.*	*Occurrence of AF (0/1)*
[176,213)	[1, 2]	3	1	1	24 / 3
[0, 171)	[1, 2]	3	1	1	18 / 2
[0, 171)	[1, 2]	2	1	1	15 / 3
[176, 213)	[1, 2]	2	1	1	15 / 0
[0, 171)	[3, 4]	3	1	1	8 / 6
[0, 171)	[1, 2]	3	3	1	12 / 2
[176, 213)	[3, 4]	3	1	1	11 / 2
[0, 171)	[1, 2]	3	2	1	4 / 7
[176, 213)	[1, 2]	3	3	1	7 / 1

RESULTS

The multivariate regression model, when discriminating those patients who would and would not develop post-op AF, found the most important coefficients to be age, body surface area, smoking history, and NYHA functional class. The overall model yielded the ROC curve in Figure 1, having an area under the curve of 75.6%. Given a threshold found as the maximum of the product of the sensitivity and the specificity, this classifier has an accuracy of 53% with a sensitivity of 51% and a specificity of 54%.

Figure 1. The multivariate classifier ROC curve shown with the dotted line and random classification being the solid diagonal line.



The important variables in **Experiment I** were age, body surface area, employment status, smoking history, smoking status, left anterior descending artery stenosis, and physician estimated mortality risk. This made for 243 rules, of which nine rules cover 14.5% of the population. The overall rule set gives an accuracy of 91.5% with a sensitivity of 68.1% and a specificity of 99.3%.

Experiment II found the important variables to be the patient's total cholesterol, smoking status, presence of coronary artery disease, SCORE risk value, and mitral valve regurgitation. Given their discrete values, this comes to 138 rules, of which nine rules cover 36.6% of the population. These rules, in Table III along with their associated outcomes, give an accuracy of 87% with a sensitivity of 58.5% and a specificity of 96.5%.

DISSCUSION

In the multivariate regression model, the most important coefficients were age, body surface area, smoking history, and NYHA functional class. Using our data, this standard approach resulted in similar risk predictors to previous studies, suggesting that our data represented a typical clinical experience.

Although Experiment I has a high accuracy, the number of rules created was too large (*i.e.*, a low coverage) to be clinically useful. Experiment II found a smaller rule set (*i.e.*, higher coverage) so that more patients were covered per rule. This is much better for use in a clinical setting. The characteristics found important in Experiment II were **total cholesterol, smoking history, presence of coronary artery disease, the SCORE risk index, and mitral valve regurgitation**. Notice, several of these variables do not have p-values < 0.1. Comparison of the two approaches indicates that those **features with the best individual predictive power may not combine to make the best overall classifier**, and a more thorough search of feature subsets should be attempted.

Georgia Institute of Technology