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

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

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.

I. 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 [1, 2]. According to some sources, over $1 billion is spent annually on this problem in the US alone [2]. 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 [3].

A method for the identification of those patients who are at highest risk for AF onset is greatly needed [3]. 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. For instance, Mathews *et al.* presented a risk stratification method after analyzing the usual demographic, operative, and post-op features. They found important factors to be age, history of AF, the presence of chronic obstructive pulmonary disease, and post-op withdrawal of beta blockers (BB). These were taken and combined into a weighted multivariable predictor which yielded an area under the curve (AUC) on its receiver operating characteristic (ROC) curve of 77% [4]. Chandy *et al.* did a similar study but also included some basic electrocardiogram (ECG) features. Their analysis found a relationship between post-op AF and age, body surface area, and an increase in P-wave dispersion [5]. 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.

II. METHODOLOGY

A. 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. Most subjects were male (99.2%), consistent with the general VA population.

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. The total set of variables consisted of those collected by the VA Continuous Improvement in Cardiac Surgery Program (CICSP) with the addition of fields for the classes of medications prescribed at the time of surgery. A list of CICSP variables and their definitions are available at http://www1.va.gov/health/cscc/define.htm. Missing data constituted 22.5% of the total data fields.

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 (m²)	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	

B. 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 (discussed in further detail in the next section) allowing probabilities of the existing data to be used in the completion process.

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]

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.

C. 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 [6]. 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.

In rough sets (RS), the knowledge is represented using an information system I, described by the pair (U, A), where U is the set of objects and A is the set of attributes. The set U can be thought of as samples—in this particular case, as patients—whereas A is the set of measures that describe a given sample (*e.g.*, weight, height, age, *etc*.). The easiest way to think of I is as a matrix as shown in Table III, where the rows are the objects (subjects) and the columns are attributes (features). The attributes in A are known as conditional attributes. An information system $I = (U, A \cup d)$ is an information system where the object is labeled (from a pattern recognition insight, this is the target). In this work, the decision attribute denotes whether a patient was diagnosed with AF or not.

Every subset of attributes B of A is characterized by indiscernibility, better denoted as the B-indiscernibility mathematically expressed as

$$IND(B) = \left\{ (o_i, o_j) \in U^2 : \forall a \in B \mid a(o_i) = a(o_j) \right\}, \quad (1)$$

where o_i and o_j are two given objects, a denotes an attribute, and IND(B) is an equivalence relation and

$$IND(B) = \bigcap_{a \in B} IND(a). \quad (2)$$

If objects (samples) o_i and o_j meet (1), they are said to be indiscernible (*i.e.*, cannot to be distinguished or discriminated). For example, let $B = \{G, D\}$, then $B_\Omega = \{\{P1, P3\}, \{P2, P5\}, \{P4\}\}$, where B_Ω is the family of all equivalent classes (*i.e.*, cannot be discerned by using attributes in B) using subset B. Now, letting $B \subseteq A \cup d$ and $O \subseteq U$, we can determine the sets B-lower (B_*) and B-upper (B^*) approximations as

$$B_* X = \bigcup \left\{ Y \in B_\Omega \mid Y \subseteq X \right\} \text{ and}$$
$$B^* X = \bigcup \left\{ Y \in B_\Omega \mid Y \cap X \neq \varnothing \right\}, \quad (3)$$

where $B_* X$ is the set of all objects of U that can definitely be categorized as objects of subset X using the attributes in B and $B^* X$ is the set of all the objects in U that are possible

objects of subset X using attributes in B. The subtraction between B^*X and B_*X is called the B-boundary, denoted as BX, and it is those that cannot be categorized as either objects of X or its complement based on the attributes in B. It is denoted as BX. Recalling the previous example, and letting $X = \{P2, P4\}$, the B_*X set is $\{P4\}$, $B^*X = \{P2, P4, P5\}$, and $BX = \{P2, P5\}$.

TABLE III
DECISION TABLE

Objects	gender (G)	diabetes (D)	smoker (S)	ill (IL)
P1	F	Yes	Yes	Yes
P2	M	No	No	No
P3	F	Yes	No	Yes
P4	M	Yes	No	No
P5	M	No	Yes	No

Ideally, we are searching for the minimal subset of attributes that can categorize the objects correctly. For the hypothetical example in Table III, it can be noted that we determine correctly whether a patient is ill by looking only at the gender attribute, making the other attributes redundant. The attribute {G} is called 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 example, *if* G = F, *then* IL = Yes, *else* IL = No. For space limitation reasons, more details of rough set theory are beyond the scope of this paper.

For the experiments, we used the ROSETTA software, C++-based software designed by A. Øhrn [7]. 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 [6], in this work, we use a genetic algorithm, given that the number of possible reducts is approximately 1.3×10^{25}.

D. 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 [7]. 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.

III. RESULTS

The multivariate regression model, when discriminating those patients who would and would not develop post-op
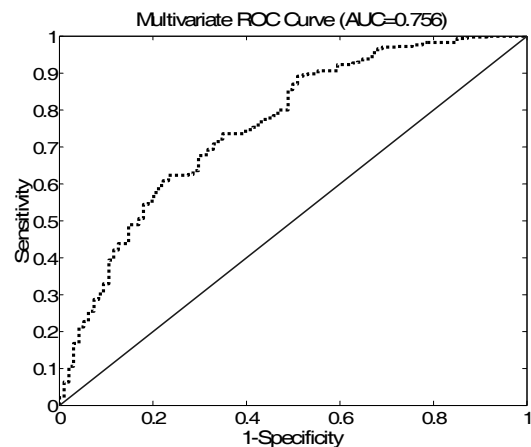

Figure 1. The multivariate classifier ROC curve shown with the dotted line and random classification being the solid diagonal line.

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%.

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 IV along with their associated outcomes, give an accuracy of 87% with a sensitivity of 58.5% and a specificity of 96.5%.

IV. DISCUSSION

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 [3, 4, 8-10], 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.

V. CONCLUSION

Among the classifiers tested, Experiment I had the highest accuracy, but also had an unacceptably low coverage of the dataset, making it of little clinical use. Experiment II, being derived from the entire database of variables, had a high coverage while still performing well with an accuracy of 87%, a sensitivity of 58.5%, and a specificity of 96.5%. This method seems to identify, with high accuracy, those patients who do not need AF prophylaxis.

The variables selected using univariate analysis and those selected for the best overall set differ significantly, though smoking history appears prominently in all models. This variable selection contrast shows the importance of testing feature subsets and discourages the practice of simply combining the best individual predictors. In this case, the most predictive set of features for post-op AF was total cholesterol, smoking history, presence of coronary artery disease, the European Society of Cardiology's SCORE risk index, and the presence of mitral valve regurgitation.

One limitation of this study is the missing data in this data set that had to be filled, which could have changed the overall data topography. This situation, though not ideal, is typical in this type of medical study. The discretization thresholds for variables are very important to this type of classifier and, though selected with care, might not have yielded maximum classification potential. We plan to implement several methods for the optimum discretization thresholds as well as try several new methods for the determination of the reducts.

In future work, a Bayesian network will be investigated for classification in order to provide the doctor with probabilistic reasoning on the prediction output. This is information that rule-based classifiers do not provide. Additionally, we will be implementing these methods with the inclusion of the features derived from the patients' ECG data. This includes the segmentation of the ECG signal into its morphological components after the identification of its fiducial points. Then, time, frequency, wavelet, symbolic, and information domain features can be calculated on each segments as well as the whole. In addition, typical heart rate variability measures can be added to determine the sympathetic nervous system's deviation prior to AF [11]. These ECG features combined with the demographic and clinical measures promise to offer an abundance of information previously missed in piecemeal investigation, thereby resulting in a user-friendly, robust, AF risk stratification architecture which could change the field of medical diagnostics.

REFERENCES

[1] R. P. Villareal, R. Hariharan, B. C. Liu, B. Kar, V. V. Lee, M. Elayda, J. A. Lopez, A. Rasekh, J. M. Wilson, and A. Massumi, "Postoperative atrial fibrillation and mortality after coronary artery bypass surgery.," *Journal of the American College of Cardiology.*, vol. 43, pp. 742-8, 2004.

[2] J. S. Steinberg, "Postoperative atrial fibrillation: a billion-dollar problem," *Journal of the American College of Cardiology*, vol. 43, pp. 1001-1003, 2004.

[3] T. Hakala and A. Hedman, "Predicting the risk of atrial fibrillation after coronary artery bypass surgery.," *Scandinavian Cardiovascular Journal.*, vol. 37, pp. 309-15, 2003.

[4] J. P. Mathew, M. L. Fontes, I. C. Tudor, J. Ramsay, P. Duke, C. D. Mazer, P. G. Barash, P. H. Hsu, and D. T. Mangano, "A multicenter risk index for atrial fibrillation after cardiac surgery," *JAMA*, vol. 291, pp. 1720-1729, 2004.

[5] J. Chandy, T. Nakai, R. J. Lee, W. H. Bellows, S. Dzankic, and J. M. Leung, "Increases in P-wave dispersion predict postoperative atrial fibrillation after coronary artery bypass graft surgery," *Anesthesia & Analgesia*, vol. 98, pp. 303-10, 2004.

[6] Z. Pawlak, *Theoretical Aspects of Reasoning about Data*: Kluwer Academic Publishing, 1991.

[7] A. Øhrn, "Discernibility and Rough Sets in Medicine: Tools and Applications," Norwegian University of Science and Technology, 1999.

[8] D. Amar, W. Shi, J. Hogue, Charles W., H. Zhang, R. S. Passman, B. Thomas, P. B. Bach, R. Damiano, and H. T. Thaler, "Clinical prediction rule for atrial fibrillation after coronary artery bypass grafting," *Journal of the American College of Cardiology*, vol. 44, pp. 1248-1253, 2004.

[9] M. Funk, S. B. Richards, J. Desjardins, C. Bebon, and H. Wilcox, "Incidence, Timing, Symptoms, and Risk Factors for Atrial Fibrillation After Cardiac Surgery," in *American Journal of Critical Care*, vol. 12: American Association of Critical Care Nurses, 2003, pp. 424-433.

[10] A. G. Zaman, R. A. Archbold, G. Helft, E. A. Paul, N. P. Curzen, and P. G. Mills, "Atrial Fibrillation After Coronary Artery Bypass Surgery : A Model for Preoperative Risk Stratification," *Circulation*, vol. 101, pp. 1403-1408, 2000.

[11] M. C. Wiggins, E. P. Gerstenfeld, G. Vachtsevanos, and B. Litt, "Electrogram Features are Superior to Clinical Characteristics for Predicting Atrial Fibrillation After Coronary Artery Bypass Graft Surgery," *Journal of the American College of Cardiology (JACC)*, vol. 47, pp. 12A-13A, 2006.