# Classification Methods for Atrial Fibrillation Prediction after CABG

S. Sovilj[1], R. Magjarević[1] and G. Rajsman[2]

[1] University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia
[2] Clinical Hospital Center Zagreb, Department of Cardiac Surgery, Zagreb, Croatia

*Abstract*— **The aim of this study is to compare different methods for the classification type problems specifically in predicting Atrial Fibrillation (AF) after Coronary Artery Bypass Grafting (CABG). The prediction/classification model tends to predict a categorical dependent variable (which determines the belonging of a patient to a group of patients that have or to a group of patients that have not developed AF), by one or more continuous and/or categorical predictor variables derived from the patients' history, ECG and in particular from the P wave. We have obtained the parameters from continuously recorded ECG after the surgery.**

*Keywords*— **classification, prediction, atrial fibrillation, CABG, discriminant analysis, decision tree.**

## I. Introduction

Atrial fibrillation (AF) is the most common postoperative arrhythmia after CABG and occurs in 30 – 40 % patients [1]. AF may cause different complications like hemodynamic changes, cerebral and other thrombo-embolisms etc., all potentially dangerous for patients. In previous studies which tried to find predictors of AF in similar groups of patients, authors analyzed relatively short segments of multichannel ECG recorded and included into analysis also some other measured physiological parameters or data from patient's history [2, 3]. In these studies the authors have not agreed on either parameters which should be recorded or on their value which could be used in clinical practice as a general procedure. Therefore we decided to analyze the parameters of the patients' ECGs recorded continuously after the surgery (while the patients are still in the ICU). We have enlarged the number of parameters from the ECG that entered the analysis and we used less demanding instrumentation for ECG acquisition.

The aim of this study is to develop a statistical classification model for prediction of AF based on parameters obtained from the patients' ECGs which were continuously acquired after the CABG. The early risk assessment for AF, several hours before its beginning, would result in timely medication treatment of patients prone to AF and would reduce the incidence of arrhythmia. The group of patients that does not show AF predictors could be excluded from prophylactic anti-arrhythmia medication, thus reducing possible drug contraindications.

## II. Methods

In the period from 2005 to 2006, we continuously recorded the standard II lead ECG of fifty patients in a period of typically 48 hours after CABG or until the onset of AF. The ECG was recorded with HP Patient Monitor 78330A and digitized using a standard ADC card (Measurement Computing CIO-DAS08/JR) embedded into a PC. The sampling frequency was 1kHz and the amplitude resolution 12-bit. Electrocardiograms were segmented with a QRS and P wave detector based on the wavelet transformations [2] and numerous parameters were calculated especially in the P wave segment of ECG.

Every recorded ECG was divided in time segments of one-hour duration. In each one-hour ECG record, the number of detected QRS segments had to be more than 2000 in order to exclude those segments which had a lot of artifacts superimposed to the signal. Furthermore the records in which a number of detected P waves was less than 75% of the total number of detected QRS segments in current hour were also excluded from analysis.

We made statistical analysis of the data obtained from the recorded ECG segments and identified parameters which best discriminate the two patient groups (AFP and NAFP). Different AF prediction/classification models were proposed and we compared their accuracy and possible applicability: linear discriminant analysis models, classification and regression tree (C&RT) and CHAID statistical tree models, Boosting Tree Classifier models and Binomial Logit Regression models. For the statistical analysis Statistica 7, StatSoft Inc. was used [4].

### *A. Measured ECG parameters*

After the QRS and P wave detection in recorded ECG segments, 110 different parameters dealing predominantly with atrial activity were measured or calculated. Only those that were considered important in the discrimination between the groups are presented in Table 1 though later analysis excluded some of them from the models. An hourly average value and standard deviation for each of these parameters was calculated. The parameters were aligned in three categories: Time parameters, Wavelet parameters and Other parameters, as presented in Table 1.

Table 1 Measured parameters.

| Parameter | Description |
|---|---|
| **Time parameters:** | |
| *PonPoffAVR, PonPoffDEV* | P wave duration (from P wave onset to P wave offset) |
| *PonPpeakAVR, PonPpeakDEV* | 1st half P wave duration (from P wave onset to P wave peak) |
| *PpeakPoffAVR, PpeakPoffDEV* | 2nd half P wave duration (from P wave peak to P wave offset) |
| *PpeakRpeakAVR, PpeakRpeakDEV* | PR interval duration (from P peak to R peak) |
| *PonQonAVR, PonQonDEV* | PQ interval duration (from P onset to Q onset) |
| *RRAVR, RRDEV, HRAVR, HRDEV* | RR interval duration and Heart Rate |
| **Wavelet parameters:** | |
| *Pslope1AVR5, Pslope1DEV5* | P wave rising slope (value of wavelet coefficient detected at the 5th wavelet scale) |
| *Pslope2AVR5, Pslope2DEV5* | P wave falling slope (value of wavelet coefficient detected at the 5th wavelet scale) |
| *Pslope1Pslope2AVR5, Pslope1Pslope2DEV5* | duration between points of highest and lowest P wave slope |
| *WenergyAVR5, WenergyDEV5* | energy measured between P wave onset and offset at 5th wavelet scale |
| *relWenergyAVR5* | relative P wave energy (ratio between energy at 5th wavelet scale and total) |
| *Wentropy* | measure of P wave energy dispersion at different wavelet scales |
| *relPslope1Pslope2AVR, relPslope1Pslope2DEV* | relative ratio between rising and falling P wave slope |
| **Other parameters:** | |
| *ampAVR, ampDEV* | P wave amplitude |
| *AonoffAVR, AonoffDEV* | Surface area below P wave |
| *RecordHour* | number of hour after CABG |
| *PpeakRpeakAVR_RR, PpeakRpeakDEV_RR,* | PR interval duration normalized with RR interval |
| *PonQonAVR_RR, PonQonDEV_RR,* | PQ interval duration normalized with RR interval |
| *PonPoffAVR_RR, PonPoffDEV_RR* | P wave duration normalized with RR interval |

suffix AVR notes mean value of measured parameter in 1 hour period
suffix DEV notes standard deviation of measured parameter in 1 hour

Statistical analysis included 360 hours of ECG of patients who developed AF and 1003 hours of ECG of patients who did not develop AF.

Approximately two third of the cases (930 hours) were randomly selected and entered in the analysis as a learning sample and the remaining third of the cases (433 hours) was used for a cross-validation and was treated as a testing sample. Prior probability of the class size was estimated from the learning sample.

### *B. General discriminant analysis model*

General Discriminant Analysis (GDA) is a method for building a multivariate linear model used to determine the variables that discriminate between two or more naturally occurring groups. A categorical dependent (criterion) variable labeled AF determines the belonging of a patient to either the group that has developed (PAF) or has not developed AF (nPAF) and it was predicted with more continuous independent (predictor) variables using the model obtained by GDA.

Using GDA, a model for prediction of AF based on a number of parameters measured from ECG was built (labeled the model GDA1). In a forward stepwise analysis (in every step) all variables were evaluated in order to determine which contribute the most to the discrimination between two groups and were included into the model step by step. In each step, variables that had a statistical significance $p < 0.05$ in the discrimination entered into the model GDA1 while the others were removed from the discriminant function.

Finally, 16 variables were entered in the discriminant function. Variables that contribute the most to the discrimination are given in order of their importance: *RRAVR, PonPoffAVR,, PonQonAVR, PpeakRpeakAVR, Pslope2AVR5, Pslope2DEV5, PonPoff23AVR* (for more the detailed description of noted predictors see Table 1).

It is still difficult to interpret the model and to explain why the observations are classified or predicted in a particular manner. Particularly, GDA1 model assumes a linear relation between predictor variables and dependent variables. The quality measures for the model GDA1 are presented in the Table 2.

Table 2 Quality measures for GDA1 model (TP – true positive, FN – false negative, TN – true negative, FP – false positive).

| | learning sample | testing sample |
|---|---|---|
| TP cases | 138 | 47 |
| FN cases | 109 | 65 |
| TN cases | 637 | 289 |
| FP cases | 46 | 32 |
| sensitivity | 55,9% | 42,0% |
| specificity | 93,3% | 90,0% |
| positive predictivity | 75,0% | 59,5% |
| negative predictivity | 85,4% | 81,6% |
| accuracy | 83,3% | 77,6% |

### *C. General classification and regression tree model*

General Classification and Regression Tree algorithm (GC&RT) is used to build a classification tree for predicting

a categorical predictor variable. GC&RT algorithm determines a set of *if-then* logical, univariate split conditions and tries to achieve maximal possible accuracy for the prediction. Tree models are nonparametric and nonlinear and can reveal a non-monotonic relationship between the variables using multiple splits on the same variable. The interpretation of results summarized in a tree is very simple and this simplicity is useful for a rapid classification and also for physiological evaluation and interpretation [4].

A misclassification cost was assumed equal for both PAF and nPAF groups and a priori knowledge about the sizes of groups was estimated based on analyzing sample size and was used by GC&RT algorithm. The classification tree labeled CRT1 was designed (Fig. 1) and the classification properties of the obtained model are presented in the Table 3.

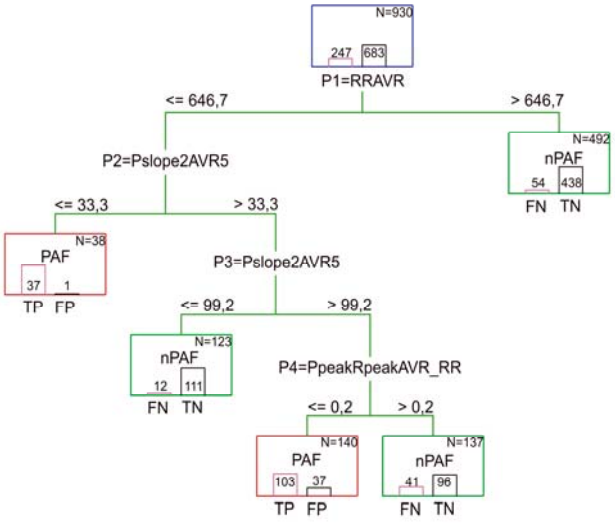


Fig. 1 CRT1 classification tree model. N denotes total number of cases that entered the classification tree at a certain node, PAF denotes cases (hours) the tree classified as belonging to a patient prone to AF, nPAF denotes cases (hours) the tree classified as belonging to a patient not prone to AF, Pi denotes the name of the predictor (see Table 1 for description), TP, TN, FP, FN same abbreviation as in Table 2, numbers above TP, TN, FP, FN designate the validity of the classification in a particular node.

### *D. General CHAID model*

*CHI*-squared Automatic Interaction Detector (CHAID) represents one of the oldest algorithms for classification tree design. The CHAID classification trees do not have to be binary and can have more than two branches in a single node. Because of simplicity the CHAID design algorithm can be used for the analysis of very large date sets. The CHAID algorithm first divides a continuous predictor into a number of categories with about equal number of observations so it becomes a categorical predictor. For each predictor a pair of category that is least significantly different in Pearson Chi-square test, with respect to the dependent variable for the classification, is determined. The algorithm then makes a choice of the predictor variable that yields the most significant split. Terminal nodes are defined as the points were no more splits can be performed because p-value for selected predictor is greater than some pre-adjusted smallest value.

Table 3 Quality measures for CRT1 model.

| | learning sample | testing sample |
|---|---|---|
| TP cases | 140 | 42 |
| FN cases | 107 | 70 |
| TN cases | 645 | 291 |
| FP cases | 38 | 30 |
| sensitivity | 56,7% | 37,5% |
| specificity | 94,4% | 90,7% |
| positive predictivity | 78,7% | 58,3% |
| negative predictivity | 85,8% | 80,6% |
| accuracy | 84,4% | 76,9% |

The cases that have missing data in any of predictor variables are excluded from the analyses. Quality measures of the CHAID classification tree model labeled CHAID1 is presented in Table 4.

Table 4 Quality measures for CHAID model labeled CHAID1.

| | learning sample | testing sample |
|---|---|---|
| TP cases | 180 | 65 |
| FN cases | 67 | 47 |
| TN cases | 515 | 238 |
| FP cases | 168 | 81 |
| sensitivity | 72,9% | 58,0% |
| specificity | 75,4% | 74,6% |
| positive predictivity | 51,7% | 44,5% |
| negative predictivity | 88,5% | 83,5% |
| accuracy | 74,7% | 70,3% |

### *E. Boosting Tree Classifiers Model*

Gradient Boosting Trees is a method that repeatedly applies predictive function in the series and weights the output of each function so that the total error of the prediction is minimized. The predictive accuracy of such a series greatly exceeds the accuracy of the base function used alone. After the first tree is designed, the residuals (error values) from the first tree are fed into the second tree which attempts to reduce the error. This process is repeated

through a chain of successive trees. The final predicted value is formed by adding the weighted contribution of each tree. Models designed as an additive series of trees are among the most accurate and they achieve better results than any other known modeling technique [4] and our results in Table 5 support that statement.

The primary disadvantage of the Boosting Tree is that the model is complex and cannot be visualized like a C&RT or CHAID models.

Table 5 Quality measures for Boosting Tree model labeled BOOST1.

| | learning sample | testing sample |
|---|---|---|
| TP cases | 297 | 91 |
| FN cases | 144 | 38 |
| TN cases | 741 | 272 |
| FP cases | 25 | 26 |
| sensitivity | 67,3% | 70,5% |
| specificity | 96,7% | 91,3% |
| positive predictivity | 92,2% | 77,8% |
| negative predictivity | 83,7% | 87,7% |
| accuracy | 86,0% | 85,0% |

After building the boosting tree, predictor statistics can be calculated. The predictors are given in order of importance: *relPslope1Pslope2AVR5, WenergyAVR4, Pslope2AVR5, WenergyAVR5, RRAVR, Pslope2DEV5, PonQonAVR4, PonPoffAVR_RR* (for more the detailed description of noted predictors see Table 1).

### *F. Binomial Logit Regression Models*

Binomial Logit Regression Model estimates the relationship between more continuous independent variables (predictors) with the binary dependent variable which specifies the case belonging to the class. The cases that belong to the class of patients who have developed AF were coded with 1 and the cases that belong to the class of patients who have not developed AF were coded with 0.

Table 6 Quality measures for Binomial Logit model labeled LOGIT1.

| | learing sample | testing sample |
|---|---|---|
| TP cases | 156 | 55 |
| FN cases | 91 | 57 |
| TN cases | 626 | 293 |
| FP cases | 57 | 28 |
| sensitivity | 63,2% | 49,1% |
| specificity | 91,7% | 91,3% |
| positive predictivity | 73,2% | 66,3% |
| negative predictivity | 87,3% | 83,7% |
| accuracy | 84,1% | 80,4% |

The predicted values for the dependent variable is never less than or equal to 0, or greater than or equal to 1, regardless of the values of the independent variables because of logit or logistic transformation [4]. After building the logit regression model (labeled LOGIT1) who quality measures are presented in Table 6, a test of significance for the predictors was performed and they are presented in order of importance for the model: *PonPoffAVR5, PonPoffAVR_RR, relPslope1Pslope2AVR5, Pslope2DEV5, Pslope2AVR5, PonQonAVR4* (for more detailed description of noted predictors see Table 1).

## III. Conclusions

In our previous work we found that several additional P wave ECG parameters may be relevant for early CABG AF prediction [1]. However, manipulating with a large number of parameters does not allow easy and simple decision making and demands formal blind studies. We have evaluated five classification tree models on our data samples. The Boosting Trees classification showed best results, i.e. highest overall sensitivity and accuracy, which might have been expected due to the highest complexity. Our results are comparable with the results obtained in previous studies and we expect that inclusion of additional parameters from the patient history will improve the classification.

## Acknowledgment

This study was supported by Ministry of Science, Education and Sport of the Republic of Croatia under grant no. 036-0362979-1554.

## References

1. Sovilj S, Rajsman G, Magjarević R. Multiparameter Prediction Model for Atrial Fibrillation after CABG. Detector. Proc. Computers in Cardiology 2006; 33:489-92.
2. Sovilj S, Rajsman G, Magjarevic R. Continuous Multiparameter Monitoring of P Wave Parameters after CABG Using Wavelet Detector. Proc. Computers in Cardiology 2005; 32:945-8.
3. Poli S, Barbaro V, Bartolini P, Calcagnini G, Censi F. Prediction of AF from surface ECG: review of methods and algorithms. Ann 1st Super Sanita 2003; 39(2).
4. StatSoft Inc. at http://www.statsoft.com

Author: Siniša Sovilj
Institute: University of Zagreb Faculty of Electrical Engineering
Street: Unska 3
City: Zagreb
Country: Croatia
Email: sinisa.sovilj@fer.hr