Predicting the Onset of Paroxysmal Atrial Fibrillation: The Computers in Cardiology Challenge 2001

1Harvard-M.I.T. Division of Health Sciences and Technology, Cambridge, MA, USA
2Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, MA, USA
3Northwestern University, Evanston, IL, USA

Abstract

The advent of pacing techniques for preventing the onset of atrial arrhythmias motivates the development of accurate predictors of these arrhythmias, and of paroxysmal atrial fibrillation (PAF) in particular. The goals of the second annual Computers in Cardiology Challenge were to determine if segments of ECG that do not include PAF contain information sufficient (1) to distinguish subjects at risk of PAF from others not at risk, and (2) to predict imminent PAF in at-risk subjects. Via PhysioNet, 18 teams of participants studied training and test databases containing two half-hour ECG recordings from each of 100 subjects (of whom 53 experienced PAF immediately following one of the two recordings). Results indicate that roughly 80% of the subjects can be correctly classified (as at-risk or not), and that imminent PAF can be predicted in roughly 80% of subjects at risk. The most successful approaches were based on analysis of the incidence of premature atrial complexes (PACs) and P-wave variability.

1. Introduction

Following the success of the first Computers in Cardiology Challenge[1], we have introduced a new challenge: to characterize changes in the surface electrocardiogram (ECG) immediately prior to the onset of paroxysmal atrial fibrillation (PAF), in an effort to develop a reliable and fully automated method for predicting the arrhythmia.

No such methods have previously been shown to be reliable. Twelve-lead electrocardiograms[2], signal-averaged P-wave morphology[3, 4], R-R interval dynamics[5, 6], and atrial ectopy[6, 7] have all been studied as possible predictors of the onset of PAF. Sensitive and specific non-invasive markers predicting the onset of PAF have not yet been determined or independently validated, however. Given recent advances in clinical electrophysiology, a prediction tool that would allow for detection of imminent atrial fibrillation may have future therapeutic ramifications.

Atrial fibrillation (AF) is the most common major cardiac arrhythmia. In the United States alone, it affects an estimated 2.2 million people, and is expected to rise; a recent study[8] estimated the population over the age of 65 to be 22 million in the elderly. AF is associated with a number of secondary complications, including stroke, congestive heart failure, and heart failure. Management of AF involves strategies to prevent recurrence, including drug therapy and electrical cardioversion. Paroxysms of atrial fibrillation may be more sustained among patients with atrial fibrillation, and that sustained episodes of AF are at higher risk of thromboembolic events.

The development of PAF is clinically significant, as the purpose of electrical cardioversion or other interventions to terminate atrial fibrillation, in the absence of other arrhythmia treatment, remains unknown. Preliminary studies have indicated that acute suppression of atrial fibrillation may be associated with the loss of sinus rhythm[9]. Atrial fibrillation may be induced or maintained by rapid atrial pacing, and there is evidence that chronic atrial pacing may have a detrimental effect on atrial function. Anticoagulation therapy is an important aspect of atrial fibrillation management, as patients with persistent atrial fibrillation are at higher risk of thromboembolic events.

2. The PAF Challenge

Participants used algorithms created for the Challenge to determine the earliest possible prediction of AF onset, based on ECG recordings from normal subjects. The database, available at http://www.physionet.org/challenge/afdb/, consists of excerpts of 24-hour Holter ECG recordings, and is designed to test algorithms for predicting AF with the least possible delay.

The PAF Challenge database contains 100 full two-channel recordings, each 24 hours long, with clinical annotations of AF onset. Participants were required to submit predictions of AF onset for each subject, along with confidence scores indicating the likelihood of AF onset. Participants were not allowed to use clinical annotations for their analysis.

The Challenge was judged on two criteria: (1) accuracy of prediction, and (2) timeliness of prediction. Accuracy was measured by comparing the predicted AF onset time with the actual AF onset time, as annotated by the Challenge organizers. Timeliness was measured by the time difference between the predicted AF onset and the actual AF onset.

The Challenge was open to all researchers and clinicians, and was designed to promote the development of algorithms for the early prediction of AF onset. The Challenge was judged on two criteria: accuracy and timeliness of prediction.

The Challenge was judged on two criteria: accuracy of prediction, and timeliness of prediction. Accuracy was measured by comparing the predicted AF onset time with the actual AF onset time, as annotated by the Challenge organizers. Timeliness was measured by the time difference between the predicted AF onset and the actual AF onset.

The Challenge was open to all researchers and clinicians, and was designed to promote the development of algorithms for the early prediction of AF onset. The Challenge was judged on two criteria: accuracy and timeliness of prediction.
vided into a learning set and a test set of equal size. The
database includes the digitized ECG signals (sampled at
128 Hz per signal, with 12-bit resolution) and a set of unaud-
dited, automatically-generated QRS annotations (contain-
ing times of occurrence only; no attempt was made to dif-
ferentiate ectopic beats from normal beats, or to correct oc-
casional detection errors). Participants were free to use the
annotations provided or to create their own based on the
digitized ECG signals.

2.1. Learning set

The learning set consists of 50 record sets. Each record
set contains two 30-minute records, and two 5-minute "con-
tinuation" records. All four records in each record set are
excepts of longer continuous ECG recordings of a single
subject; the 50 record sets come from 48 different subjects.

Twenty-five of these record sets (those with names be-
inging with 'p') come from subjects who have paroxysmal
atrial fibrillation (PAF). One record in each pair of consec-
utively numbered 30-minute records contains the ECG im-
mediately preceding an episode of PAF, which can be ver-
ified by examining the associated continuation record. The
other 25 records in each pair contains the ECG dur-
ing a period that is distant from any episode of PAF (there
is no PAF during the 45-minute periods before the begin-
ing or after the end of the 30-minute record); the associ-
ated continuation record can be used to verify that at least
the next five minutes do not contain PAF.

The other 25 record sets (those with names beginning
with 'n') come from subjects who do not have documented
atrial fibrillation, either during the period from which the
records were excerpted or at any other time. These sub-
jects include healthy controls, patients referred for long-
term ambulatory ECG monitoring, and patients in intensive
care units.

2.2. Test set

The test set is similarly constructed of 50 record sets
(from 50 different subjects); unlike the learning set, there
are no continuation records. The test set records are named
01, 02, ..., 1100. As in the learning set, pairs of consec-
utively numbered records come from the same long-term
ECG recording of a single subject. Participants in the chal-
lenge were told that between 20 and 30 of the record sets in
the test set come from subjects with PAF; the actual number
(28) was not disclosed until the conclusion of the challenge.
The remaining 22 record sets come from subjects without
PAF.

3. Organization of the Challenge

The challenge was announced, and the PAF Prediction
Challenge Database was posted, on 1 March 2001. Particip-
ants used the learning set to develop and optimize fully-
automated methods for classifying the record sets and for
predicting imminent PAF. Beginning on 21 April, partic-
pants were able to submit their algorithms' classifications
of the test set to an automated scorer via PhysioNet.

In event 1 (PAF screening), the challenge is to clas-
sify the record sets (subjects) in the test set into PAF and
non-PAF groups (in other words, can individuals at risk of
PAF be identified within a larger population, based on their
ECGs?). Results submitted for event 1 were given a nu-
merical score from 0 to 50 based on the number of correct
classifications. No distinction was made between false neg-
ative and false positive classification errors.

In event 2 (PAF prediction), the goal is to identify which
record in each pair immediately precedes PAF, for those
subjects in the PAF group (in other words, is the imminent
onset of PAF predictable in an individual known to be at
risk of PAF?). Since the same records were used for both
events, participants were not told which subjects belonged
to the PAF group; rather, they were asked to choose one
record from each pair as "likely to be followed by PAF".
Entries were given numerical scores based on the number
of correct identifications; in order to disguise the number of
PAF subjects, a point was awarded for each non-PAF sub-
ject, so that the scores ranged from 22 to 50 (although the
minimum possible score was not known to the participants).
The event 2 scores reported in the next section have been
corrected by subtracting 22 in each case.

Entrants were permitted to submit multiple sets of results
in order to permit experimentation with variations of their
algorithms. To discourage attempts to discover the correct
classifications by repeated submission of results, particip-
ants were allowed to submit up to six entries (total for both
events) without restriction, but were required to wait
between any additional entries for a period that began at
one day and doubled with each subsequent entry.

Participants who wished to present their work in these
proceedings needed to obtain initial results no later than 1
May 2001. Entries continued to arrive throughout the sum-
mer; the final standings were determined from the entries
submitted before noon GMT on 21 September.

4. Results

In each event, the most successful participants achieved
scores of roughly 80% of a perfect result. Tables 1 and
2 show the best score achieved by each of the top-scoring
entrants, together with the date when they first achieved that
score and the number of entries needed to do so.

114
Table 1. Final standings for event 1 (PAF screening).

<table>
<thead>
<tr>
<th>Score</th>
<th>Entrant</th>
<th>Date</th>
<th>Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>41/50</td>
<td>G Schreier, P Kastner, and W Marko</td>
<td>17 Sep</td>
<td>8</td>
</tr>
<tr>
<td>82%</td>
<td>Austrian Research Centers Seibersdorf, Graz, Austria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40/50</td>
<td>W Zong and RG Mark</td>
<td>12 Sep</td>
<td>7</td>
</tr>
<tr>
<td>80%</td>
<td>Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA (unofficial entry)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37/50</td>
<td>R Sweeney and colleagues</td>
<td>8 May</td>
<td>3</td>
</tr>
<tr>
<td>74%</td>
<td>Guidant Corp., St. Paul, MN, USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36/50</td>
<td>C Maier, M Bauch, and H Dickhaus</td>
<td>19 Sep</td>
<td>2</td>
</tr>
<tr>
<td>72%</td>
<td>University of Applied Sciences, Heilbronn, Germany</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33/50</td>
<td>C Marchesi and M Paoletti</td>
<td>27 Apr</td>
<td>1</td>
</tr>
<tr>
<td>70%</td>
<td>Università di Firenze, Firenze, Italy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34/50</td>
<td>KS Lynn and HD Chiang</td>
<td>28 Apr</td>
<td>6</td>
</tr>
<tr>
<td>69%</td>
<td>Cornell University, Ithaca, NY, USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33/50</td>
<td>CC Yang</td>
<td>21 Apr</td>
<td>4</td>
</tr>
<tr>
<td>66%</td>
<td>National Yang-Ming University, Taipei, Taiwan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33/50</td>
<td>JA Kors</td>
<td>10 Jul</td>
<td>2</td>
</tr>
<tr>
<td>66%</td>
<td>Erasmus University, Rotterdam, The Netherlands</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32/50</td>
<td>P de Chazal and C Heneghan</td>
<td>13 Sep</td>
<td>1</td>
</tr>
<tr>
<td>64%</td>
<td>University of New South Wales, Sydney, Australia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32/50</td>
<td>R Loesch</td>
<td>14 Sep</td>
<td>6</td>
</tr>
<tr>
<td>64%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Final standings for event 2 (PAF prediction).

<table>
<thead>
<tr>
<th>Score</th>
<th>Entrant</th>
<th>Date</th>
<th>Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>22/28</td>
<td>W Zong and RG Mark</td>
<td>1 May</td>
<td>1</td>
</tr>
<tr>
<td>79%</td>
<td>Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA (unofficial entry)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20/28</td>
<td>G Schreier, P Kastner, and W Marko</td>
<td>19 Aug</td>
<td>2</td>
</tr>
<tr>
<td>71%</td>
<td>Austrian Research Centers Seibersdorf, Graz, Austria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19/28</td>
<td>P de Chazal and C Heneghan</td>
<td>28 Apr</td>
<td>1</td>
</tr>
<tr>
<td>68%</td>
<td>University of New South Wales, Sydney, Australia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19/28</td>
<td>C Maier, M Bauch, and H Dickhaus</td>
<td>11 Sep</td>
<td>3</td>
</tr>
<tr>
<td>68%</td>
<td>University of Applied Sciences, Heilbronn, Germany</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18/28</td>
<td>KS Lynn and HD Chiang</td>
<td>29 Apr</td>
<td>2</td>
</tr>
<tr>
<td>64%</td>
<td>Cornell University, Ithaca, NY, USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17/28</td>
<td>P Langley, D di Bernardo, J Allen, E Bowers, F Smith, S Vecchietti, and A Murray</td>
<td>30 Apr</td>
<td>1</td>
</tr>
<tr>
<td>61%</td>
<td>Freeman Hospital, Newcastle upon Tyne, UK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17/28</td>
<td>D Gamberger and T Smuc</td>
<td>23 Aug</td>
<td>2</td>
</tr>
<tr>
<td>61%</td>
<td>Rudjer Boskovic Institute, Zagreb, Croatia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16/28</td>
<td>CC Yang</td>
<td>23 Apr</td>
<td>1</td>
</tr>
<tr>
<td>57%</td>
<td>National Yang-Ming University, Taipei, Taiwan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16/28</td>
<td>R Sweeney and colleagues</td>
<td>8 May</td>
<td>1</td>
</tr>
<tr>
<td>57%</td>
<td>Guidant Corp., St. Paul, MN, USA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15/28</td>
<td>L Almarro</td>
<td>30 Apr</td>
<td>1</td>
</tr>
<tr>
<td>54%</td>
<td>UPV, Valencia, Spain</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. Conclusions

The questions posed by the challenge are quite difficult. The best-performing algorithms were based on incidence of isolated atrial premature complexes and on P-wave variability. Methods based on time-domain and frequency-domain analysis of heart rate variability were somewhat less successful but show promise.

Further study is needed to determine if combinations of these diverse strategies can yield further improvements in PAF screening or prediction accuracy.

As we observed in the first Computers in Cardiology Challenge, by providing open access to a well-chosen data set, it is possible to stimulate rapid progress on significant research questions, with the participation of both specialist and non-specialist investigators. Notwithstanding the difficulty of predicting PAF, the results of the challenge now make it clear that a variety of approaches can succeed in doing so in most subjects. The use of a common database for evaluation of these diverse approaches permits straightforward and objective comparisons of their merits.

Acknowledgments

The Computers in Cardiology Challenges have been conducted using the facilities of PhysioNet, a public service of the Research Resource for Complex Physiologic Signals, which is supported by a grant from the National Center for Research Resources of the US National Institutes of Health (P41 RR13622). We thank Gerald Parents and the board of Computers in Cardiology for their continuing and enthusiastic support. GBM thanks his coauthors for their generous contributions of carefully selected data to the PAF Prediction Challenge Database. Thanks also to Isaac Henry, Christoph Maier, Joseph Mietus, and Juan Millet for timely and valuable feedback on the database. Finally, we thank the 16 teams of participants from 10 nations, who accepted this difficult challenge; ten of them report their results elsewhere in this volume.

References


Address for correspondence:

George B. Moody
MIT Room E25-505A, Cambridge, MA 02139 USA.
george@mit.edu