# Methods for Discriminating Pre-Ectopic Sinus Beats

S Cavalcanti[1], S Lodi[0], G Moro[1], M Samorani[2], C Sartori[0], S Severi[1]

[0]DEIS, Università degli Studi di Bologna, Italy
[1]DEIS, Università degli Studi di Bologna, Cesena, Italy
[2]Leeds School of Business, University of Colorado at Boulder, CO, USA

## Abstract

*One hour, high resolution 12-lead ECG recordings were digitally obtained from 10 patients with frequent isolated premature ventricular contraction (PVCs >100/hr). Recordings were processed and single beats were automatically annotated. For each lead one record per beat was generated with 102 quantitative attributes of ECG waveform For each recording, two sets of beats were randomly sampled: 200 isolated regular beats (S1) and 200 isolated premature ventricular beats (S2). In isolated premature ventricular beats the 3 preceding and 3 following beats were normal sinus beats. For each of the 200 ectopic beats, a new record having as attributes the maximum, minimum and average value of every original attribute of the 3 preceding sinus beats. Likewise, we randomly sampled 200 sinus beats such that the 3 preceding and following beats were regular as well, and, for each of the 200, created a record in the same fashion. Every time a beat was sampled, the 3 beats preceding it and the 3 beats following it were deleted from the data set, as well as the sampled beat. Thus, every record of the first set represents a pre-ectopic beat, whereas every record of the second set a non pre-ectopic beat. A two-tailed unpaired Student's T-test and other experiments not only showed that there is a statistically significant difference between the two sets, but also that, for every patient, one attribute of the new data set allows to discriminate between pre-ectopic beats and non pre-ectopic beats in 80% of the cases. These results indicate that detectable signs of incoming PVC are present in pre-ectopic sinus beats.*

## 1. Introduction

Isolated ectopic beats are usually regarded as a sudden unpredictable event having no correlation with pre-ectopic sinus beats. Actually, non homogeneous excitability underlying sinus beats may prepare in short-term the cardiac tissue to originate an isolated ectopic beat. To ascertain whether a triggered activity preludes the ectopic beat, we assessed the statistical differences between ECG waveform of sinus beats preceding a premature ventricular extrasystole beat (pre-ectopic) with respect to non pre-ectopic sinus beats.

In order to assess if there is a statistically significant difference between pre-ectopic and non pre-ectopic beats, a data set is built whose observations are pre-ectopic and non pre-ectopic beats, and whose attributes have been obtained by massaging the attributes obtained by the software MISHA (Mortara Instrument). Then, an in-depth statistical analysis performed on these data sets shows that one single attribute is sufficient to discriminate between pre-ectopic and non pre-ectopic sinus beats, and that it is possible to accurately predict whether an unseen sinus beat is pre-ectopic or not.

## 2. Data set construction

One hour, high resolution 12-lead ECG recordings were digitally obtained from 10 patients with frequent isolated premature ventricular contraction (PVCs >100/hr). Recordings were made on a 12-lead Mortara H-12 recorder and transferred to a personal computer. Digital recordings were processed by the software package MISHA (Mortara Instrument) in order to automatically annotate single beats and to extract for each lead one record per beat with quantitative attribute of ECG waveform.

The following attributes were automatically determined beat-to-beat for each lead: T wave amplitude (TA) and time of peak (TP), T wave upslope (TU) and downslope (TD), ST-segment offset (ST), Q wave amplitude (QA), R wave amplitude (RA) and S wave amplitudes (SA), RR interval (RR). QT interval duration (QT), QT interval correct to RR (QTc), QT correct according to the Bazett's formula (QTcbaz), first and second eigenvalues of T-wave (E1, E2) were derived for each cluster of 8 beats and were assigned to each beat corresponding to the previous cluster. Hence the number of attributes obtained by MISHA on each single beat were 102.

For each recording, two sets of beats were considered: isolated regular beats ($S_1$) and isolated premature ventricular beats ($S_2$). Here, *isolated* means that the 3

preceding and the 3 following beats were regular sinus beats. Then, in order to have a balanced and big enough data set, we randomly sampled 200 beats for each of the two sets. If the total number of premature ventricular beats was less than 200, we considered all the available ones, and we sampled the same quantity from the set of isolated regular sinus beats. Every time a beat was sampled, the 3 beats preceding it and the 3 beats following it were deleted from the data set, as well as the sampled beat. The preceding beat of each of these 400 beats is either a pre-ectopic beat or a non pre-ectopic beat, and it becomes an observation of the modified data set.

The attributes of these beats are built in order to include as much information as possible of the preceding beats. In fact, including only the attributes generated by MISHA would describe the beats individually, ignoring the information of the current *trend* of beats. In order to capture this important piece of information, we created, for each of the 400 beats, a new record having as attributes the maximum, minimum and average value of every original attribute computed among the 2 preceding sinus beats and the current sinus beat. The total number of attributes obtained is then 306.

## 3. Experiments and results

For each patient, for each attribute, we considered the two univariate data sets corresponding to $S_1$ and $S_2$, and we performed a two-tailed unpaired Student's T-test in order to assess whether their true means were significantly different. Table 1 shows, for each patient, for each level of significance, the number of attributes that are statistically different – we call them *discriminating attributes*.

| Patient | Discriminating attributes | | |
|---|---|---|---|
| | **α=0.05** | **α=0.01** | **α=0.0001** |
| 1 | 118 | 83 | 32 |
| 2 | 205 | 180 | 155 |
| 3 | 164 | 146 | 106 |
| 4 | 197 | 168 | 120 |
| 5 | 127 | 104 | 64 |
| 6 | 35 | 7 | 0 |
| 7 | 87 | 49 | 19 |
| 8 | 270 | 258 | 229 |
| 9 | 99 | 77 | 18 |
| 10 | 167 | 129 | 26 |

To confirm the significance of the most discriminating attribute (MDA), i.e. the one with the lowest p-value, we set up a classification procedure and measured both its classification and its prediction accuracies. For each patient we considered only the most discriminating attribute, and we computed a discriminant threshold *dt* as follows:

$$dt = \frac{\mu_1 \cdot \sigma_2 + \mu_2 \cdot \sigma_1}{\sigma_1 + \sigma_2}$$

Where $\mu_1$ is the mean of $S_1$, $\mu_2$ is the mean of $S_2$, $\sigma_1$ is the standard deviation of $S_1$, $\sigma_2$ is the standard deviation of $S_2$. In practice, *dt* is between $\mu_1$ and $\mu_2$, and is closer to the mean of the set having the smaller standard deviation. Suppose, without loss of generality, $\mu_1 < \mu_2$; then, given the attributes of a sinus beat, if the value of the most discriminant attribute is greater than *dt*, the beat is classified as pre-ectopic because it is on the side of $\mu_2$, otherwise it is classified as non pre-ectopic.

The table below shows, for each patient, the number of beats in the data set (≤400), the MDA, and the classification accuracy achieved over the considered sinus beats.

| Patient | Beats | MDA | Accuracy |
|---|---|---|---|
| 1 | 124 | TD/V6(ms)_min | 78.23% |
| 2 | 400 | TP/I(ms)_min | 95.50% |
| 3 | 380 | TU/II(ms)_min | 94.74% |
| 4 | 360 | TD/I(ms)_min | 94.72% |
| 5 | 168 | TD/II(ms)_min | 86.90% |
| 6 | 34 | TD/V5(ms)_avg | 73.53% |
| 7 | 74 | QTc(ms)_avg | 79.73% |
| 8 | 400 | QA8(uV)_max | 85.00% |
| 9 | 198 | E1-T_min | 59.09% |
| 10 | 240 | TD/V5(ms)_avg | 60.42% |

As noted in [1], the apparent good accuracy obtained over the entire data set may be different (greater) than the accuracy obtained predicting the class of new, unseen, sinus beats. The presence of this effect, which is known as overfitting, can be measured by a *k*-fold cross validation, which consists in partitioning the entire data set into *k* different sets, and performing k different classification tests by alternatively using the union of *k*-1 subsets as training set, and the remaining subset as test set. The average accuracy obtained in predicting the class of the observations in the *k* tests is called cross validation accuracy (CVA), and the standard deviation of the k accuracies obtained is called Accuracy Standard Deviation (ASD). In our case, each test consists in finding the MDA considering only (*k*-1)/*k* of the beats, computing the corresponding discriminant threshold, and using it to predict the class of the remaining 1/*k* of the beats. The CVA is the total fraction of beats that are correctly predicted. Since *k* T-tests are performed for every patient, there may be more than one distinct MDAs (at most *k*).

We performed a 5-fold cross validation for each patient and reported the results in the following table, which show the CVA, the ASD, and the MDAs found for each patient. The frequency of each attribute is shown in parentheses.

| Patient | CVA | ASD | MDAs |
|---|---|---|---|
| 1 | 76.57% | 3.78% | TD/V6(ms)_min (3)<br>TD/III(ms)_min (1)<br>E1-T_min (1) |
| 2 | 87.5% | 5.08% | TU/I(ms)_min (4)<br>TP/I(ms)_avg (1) |
| 3 | 92.63% | 4.01% | TU/II(ms)_avg (4)<br>TU/II(ms)_min (1) |
| 4 | 90.56% | 5.14% | TD/I(ms)_min (2)<br>TU/I(ms)_min (2)<br>TP/I(ms)_min (1) |
| 5 | 84.47% | 6.6% | TD/II(ms)_min (2)<br>TD/AVR(ms)_min (1)<br>TD/V2(ms)_min (1)<br>TP/II(ms)_min (1) |
| 6 | 56.19% | 16.63% | TD/V5(ms)_avg (1)<br>TU/V4(ms)_max (1)<br>TP/V5(ms)_avg (1)<br>RR (ms)_max (1)<br>TD/II(ms)_avg (1) |
| 7 | 76.86% | 8.46% | QTc(ms)_avg (3)<br>QTc(ms)_min (1)<br>QTc(ms)_max (1) |
| 8 | 86.50% | 5.41% | QA5(uV)_max (4)<br>QA5(uV)_avg (1) |
| 9 | 57.01% | 8.76% | E1-T_min (2)<br>TA/II(uV)_min (1)<br>RA12(uV)_avg (1)<br>TD/V6(ms)_avg (1) |
| 10 | 57.92% | 6.81% | TA/V2(uV)_avg (1)<br>ST/V6(uV)_max (1)<br>TA/V2(uV)_min (1)<br>E1-T_avg (1)<br>TD/V5(ms)_avg (1) |

The results show that the overfitting is very limited because the CVA is always very close to the training set accuracy; using only one attribute as classifier certainly helps to keep the overfitting low, but higher accuracies can be expected by using more complex classifiers, such as decision trees, neural networks, etc…

Furthermore, the results show an interesting pattern: the tests that leads to the best CVAs generally use fewer attributes than the other tests. For example, 4 out of the 5 tests for patient 3 use the same attribute TU/II(ms)_avg, and the CVA obtained is 92.63%; on the other hand, all tests for patient 10 use a different attribute, and the CVA is 57.92%.

## 4. Discussion and conclusions

Present analysis is based on the assumption that isolated ventricular premature beats are due to triggering sinus beats. To validate this assumption we compared the sinus beat preceding ectopic beat, with sinus beat far form ectopic one. Consistent with the initial hypothesis, we found that detectable signs of incoming PVC are present in pre-ectopic sinus beats.

Except one patient (pat. 9) measures on T wave are able to discriminate between pre-ectopic and non pre-ectopic sinus beat. In 5/10 patients, T wave duration (TD) was the most discriminating attribute, denoting that ventricular repolarization in sinus beat is most sensitive phase of cardiac cycle to predispose to premature ventricular beat. In fact, a prolonged ventricular repolarization in sinus beats may prepare in short-term the cardiac tissue to originate an isolated ectopic beat. According with this observation, [2] show that adrenergic stimulation can facilitate the induction of arrhythmias.

Changes in RR duration preceding ventricular ectopic beats have been used to identify clinical subsets of patients [3]. In present analysis, RR interval was not a relevant discriminating attribute but it remain confirmed that analysis of sinus beats may serve as a basis for regarding the treatment and prognosis of ventricular ectopic beats.

## Acknowledgements

We would like to thank Mortara Rangoni Europe for providing the 12-led digital recorder and the MISHA software.

## References

[1] Han J, Kamber M. Data Mining. Concepts and Techniques. San Francisco: Morgan Kaufmann Publishers, 2006.

[2] Stein KM, Karagounis LA, Markowitz SM, Anderson JL, Lerman BB. Heart rate changes preceding ventricular ectopy in patients with ventricular tachycardia caused by reentry, triggered activity, and automaticity. Am Heart J. 1998, 136(3), 425-34.

[3] Sapoznikov D, Luria MH, Gotsman MS. Changes in sinus RR interval patterns preceding ventricular ectopic beats: assessment with rate enhancement and dynamic heart rate trends. Int. J. Cardiol 1999, 69(2):217-24.