Predicting the Occurrence of Acute Hypotensive Episodes: The PhysioNet Challenge

F Chiarugi[1], I Karatzanis[1], V Sakkalis[1], I Tsamardinos[1,2],
Th Dermitzaki[3], M Foukarakis[4], G Vrouchos[5]

[1]Institute of Computer Science, Foundation for Research and Technology - Hellas, Heraklion, Crete, Greece
[2]Computer Science Department, University of Crete, Heraklion, Crete, Greece
[3]Medical School, University of Crete, Heraklion, Crete, Greece
[4]Cardiology Department, Venizeleio-Pananeio Hospital, Heraklion, Crete, Greece
[5]ICU-CCU Department, Venizeleio-Pananeio Hospital, Heraklion, Crete, Greece

Abstract

The PhysioNet Challenge 2009 addresses the prediction of acute hypotensive episodes (AHEs), which are serious clinical events since they could result in multiple organ failure and eventually in death. This objective is pursued with two different events: a) event 1: the separation of records with critical AHE (subgroup H1) in the forecast window (FW), the one-hour period immediately following a specified time T_0, from records from patients with no documented AHEs at any time during their hospital stay (subgroup C1) and b) event 2: the separation of records with an AHE in the FW (group H) from records without any AHE in the FW (group C).

Both events have been approached, using a subset of information common to the whole dataset, through the extraction of significant features from the last hours before T_0 of the ABP and HR time series, linearly interpolated in the empty intervals and processed with a median filter for suppressing most artifacts. Decision tree classifiers based on these features have been designed for event 1 and 2, having better performances than classifiers based on support vector machine. The H1/C1 classifier (event 1) correctly classified all cases of the learning set (15 H1, 15 C1) producing also a perfect score on test set A. The H/C classifier (event 2) correctly classified 91.67% of the cases in the training set (30 H, 30 C) and obtained a score of 75% on test set B.

1. Introduction

Acute hypotensive episodes (AHEs) are serious clinical events in intensive care units (ICU), since could result in multiple organ failure and eventually in death. The AHE prediction would enable physicians to respond timely and to improve patient outcome.

In this paper the approach adopted for predicting the AHE occurrence using the dataset made available by Physionet/Computers in Cardiology for the Challenge 2009 [1] is described. In this dataset an AHE is defined as any period of 30 minutes or more during which at least 90% of the mean arterial blood pressure (ABP) measurements were at or below 60 mmHg.

In this challenge two goals are pursued: a) to separate the records with a critical AHE in the forecast window (FW), the one-hour period immediately following a specified time T_0, from records from patients with no documented AHEs at any time during their hospital stay (event 1) and b) to separate the records with an AHE in the FW from the records without any AHE in the FW (event 2).

2. Methods

The information available in the ICU is usually very huge and heterogeneous. The PhysioNet Challenge 2009 dataset corresponds to the previous point and consists for each record of: clinical data entered into the ICU medical information systems (records of diagnosis, observations, measurements, and interventions/therapy performed in the ICU); ECG and ABP signals sampled at 125 Hz (some records may include up to six additional signals); time series of vital signs sampled once per minute or once per second including heart rate and mean, systolic, and diastolic ABP (most records include a variety of additional vital-sign time series, most often including respiration rate and oxygen saturation).

The training set is composed by 60 records. Each one has the FW starting at a time T_0 and is assigned to a group H or C and to a subgroup H1, H2, C1, or C2. Records in group H contain an AHE starting in the FW with patients in subgroup H1 (critical AHE) receiving pressor medication and patients in H2 (non critical AHE) not. Records in group C contain no AHE within the FW with patients in subgroup C1 without any AHE during their hospital stay and patients in C2 with AHE outside the FW. Group population is totally balanced: each group is composed by 30 records and each subgroup by 15. Two test sets are available for testing the developed algorithms. Test set A, for event 1, consists of 10 records. Five of these belong to subgroup H1 and the other five to subgroup C1. Test set B, for event 2, consists of 40 records. Between 10 and 16 of these belong to group H and the others to group C.

From a clinical point of view, vital signs older than 24 hours compared to $\mathbf{T_0}$ have been considered not significant in the estimation of possible AHE occurrences. Instead of statistically treating missing data situations, the adopted approach has been based on the subset of available information common to the whole dataset. This method has been considered more realistic and easily applicable or adaptable to most ICU settings where this core information is usually available while more extended information might not be present. Furthermore, an analysis of the available data revealed that the high-resolution signals (sampled at 125 Hz) contained several intervals of missing data, while the low-resolution data (time series) were more reliable with very few occurrences of empty intervals. For this reasons, initially only time series have been considered in the algorithm, leaving the possibility of using clinical data, which may not have a reliable time annotation for the registered events according to the challenge organizers, to a more advanced phase of development.

For each record of the training set, the low-resolution data (sampled once per minute) have been used and the following time series have been extracted: heart rate (HR), systolic ABP (ABPS), mean ABP (ABPM) and diastolic ABP (ABPD) with the samples in the last 10 hours before T_0. A fifth time series (MAP) has been obtained with the mean ABP from the start of the record to T_0 (if less than 24 hours) or in the last 24 hours before T_0, These 5 time series have been linearly interpolated in order to fill in the very rare (in the last 10 hours) empty intervals and then processed with a median filter (width = 10 samples) for reducing most artifacts.

In each record of the training set also 2 hours after T_0 are available so that the AHE can be easily noticed and the vital sign trend around T_0 can be analyzed. An example of that is reported in Figure 1 which displays a record with an AHE in the FW. In Figure 2 the ABPM time series (after linear interpolation and median filtering) in the interval from 10 hours before T_0 to 2 hours after T_0 are shown for all the records of each subgroup H1, H2, C1 and C2.

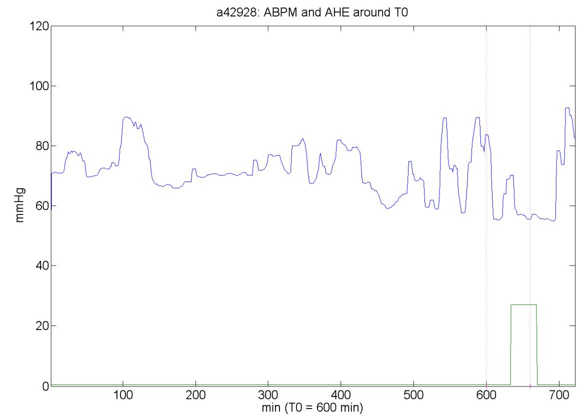
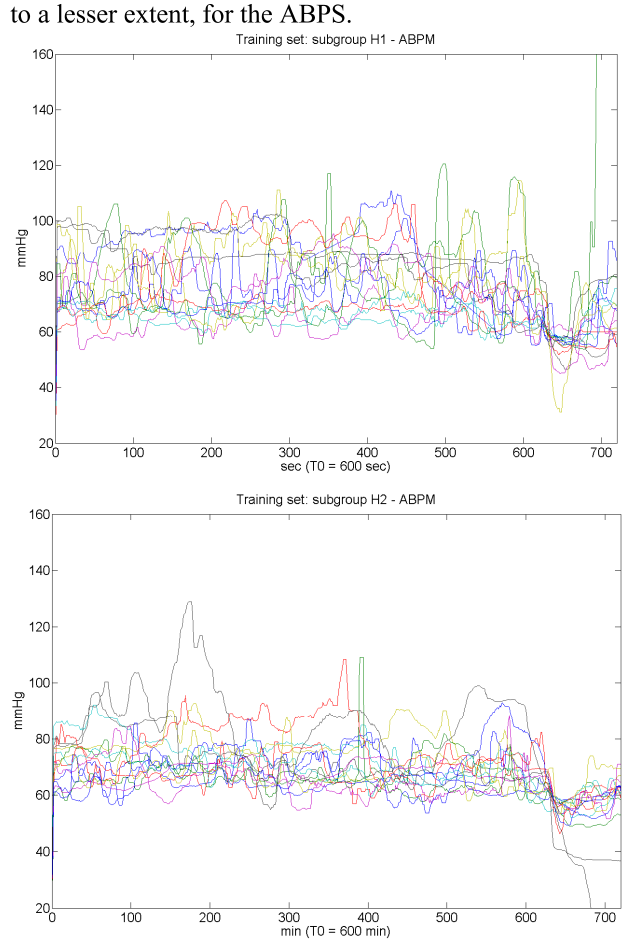


Figure 1. The ABPM time series from T_0 – 10 hours to T_0 + 2 hours for one record of the training set (subgroup H1). For this record an AHE (identified by the rectangular pulse) starts at about t = 630 min, inside the FW (600-660).

It can be noticed, from Figure 2, that the average value of ABPM in subgroup C1 is significantly higher than in the other subgroups. A similar consideration can be made for the ABPD, which is quite similar to the ABPM, and, to a lesser extent, for the ABPS.



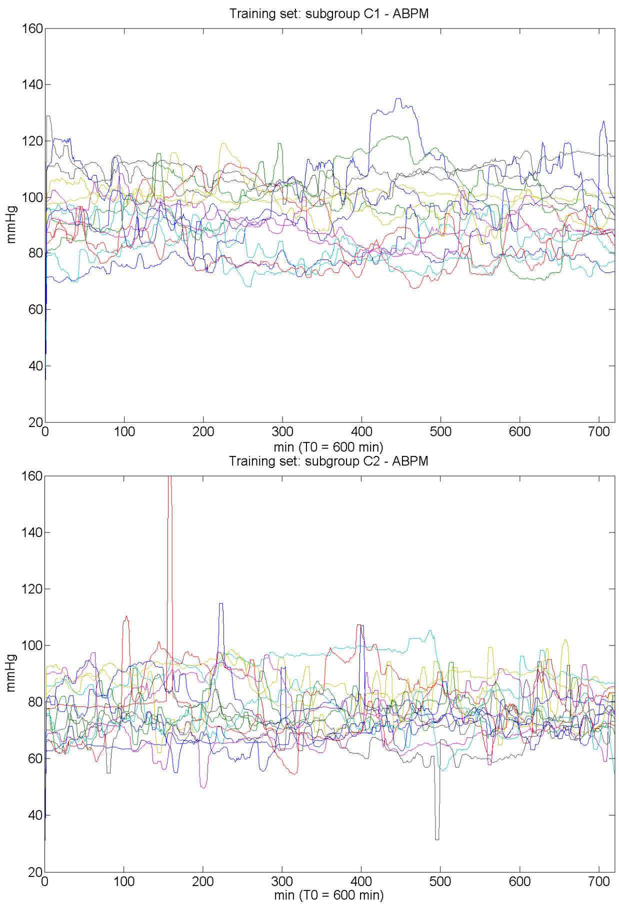


Figure 2. The ABPM time series (after linear interpolation and median filtering) from T_0 – 10 hours to T_0 + 2 hours for all the records of the H1, H2, C1 and C2 subgroups of the training set.

From these time series several features have been extracted and tested for their applicability in building the H1/C1 classifier (event 1) and the H/C classifier (event 2) based on support vector machine or on a decision tree. The approach with the support vector machine has been tested using the GEMS software (version 2.0.2) [2, 3].

In the case of event 1, the most significant features have been ABPS_l5h, ABPM_l5h, ABPD_l5h, ABPS_l1h, ABPM_l1h, ABPD_l1h, which are respectively the mean value of the systolic, mean and diastolic ABP in the last 5 hours before T_0 and the mean value of the systolic, mean and diastolic ABP in the last hour before T_0. Additionally, the presence of micro-episodes (shAHE_l24h), from the beginning of the record to T_0 (if less than 24 hours) or in the last 24 hours before T_0, has been considered. A micro-episode is defined as an AHE, but in an interval of only 20 minutes (any period of 20 minutes or more during which at least 90% of the mean ABP measurements were at or below 60 mmHg). In Table 1 the mean values of the selected features are shown for the records of the training set.

The use of a linear classifier based on support vector machine working on these 7 selected features has been tested with GEMS obtaining a max accuracy of 83.33% on the training set with the leave-one-out approach. In the attempt of obtaining better performances, a different approach based on a decision tree structure has been then adopted.

From Table 1, it can be observed how the values of all features in subgroup H1 were significantly lower than the same values in subgroup C1. Moreover, while the features are not depending from the time in subgroup C1, they are a little lower in the last 1 hour than in the last 5 hours in subgroup H1. Furthermore, shAHE_l24h was 0 for all the records of subgroup C1, while higher than zero for 8 out of 15 records of subgroup H1.

Table 1. The values of the selected features in subgroups C1 and H1 of the training set.

	C1	H1
ABPS_l5h	128.7 +/- 20.6	122.0 +/- 16.6
ABPM_l5h	93.0 +/- 12.1	77.5 +/- 7.3
ABPD_l5h	70.2 +/- 10.8	55.5 +/- 4.9
ABPS_l1h	128.6 +/- 19.2	115.8 +/- 14.4
ABPM_l1h	93.0 +/- 11.4	72.9 +/- 8.9
ABPD_l1h	70.3 +/- 10.6	52.0 +/- 5.8

The implemented algorithm is described below:

```
if (shAHE_l24h > 0)
    group1 = 'H1';
elseif (((ABPM_l5h >= 75) && (ABPD_l5h >= 60)) && ((ABPM_l1h >= 75) && (ABPD_l1h >= 60)))
    group1 = 'C1';
elseif (((ABPM_l5h >= 70) && (ABPD_l5h >= 50) && (ABPM_l1h >= 70) && (ABPD_l1h >= 50)) && ((ABPS_l5h - ABPM_l5h) <= 1.2 * (ABPM_l5h - ABPD_l5h)) || ((ABPS_l1h - ABPM_l1h) <= 1.2 * (ABPM_l1h - ABPD_l1h)))
    group1 = 'C1';
elseif ((ABPD_l5h <= 55) && (ABPD_l1h <= 55))
    group1 = 'H1';
elseif (((ABPM_l5h <= 70) && (ABPD_l5h <= 60)) || ((ABPM_l1h <= 70) && (ABPD_l1h <= 60)))
    group1 = 'H1';
elseif (((abs(ABPM_l5h - ABPM_l1h) / ABPM_l5h > 0.05)) && (abs(ABPD_l5h - ABPD_l1h) / ABPD_l5h > 0.05))
    group1 = 'H1';
else
    group1 = 'C1';
end
```

The first condition states that if there have been one or more micro-episodes in the last 24 hours the record belongs to group H1. The second condition states that record with mean ABPM in the last hour and in the last 5 hours greater than 75 mmHg with mean ABPD greater than 60 mmHg belong to C1. The third condition has been introduced in order to identify the borderline records which have the ABPM more tied to the ABPS than to the ABPD. In such cases even a low value of ABPD is not sufficient to drive the ABPM to critical low values, thus these records still belong to C1. The remaining records with mean ABPD in the last hour and in the last 5 hours lower than 55 mmHg or with mean ABPM lower than 70 mmHg and mean ABPD lower than 60 mmHg in the last hour or in the last 5 hours belong to H1. The last condition identifies the time dependency of the mean ABPM and ABPD for subgroup H1 while in subgroup C1 such values are more stationary.

In the case of event 2, additional features have been added and the implemented algorithm has been still based on a decision tree structure. In this case, a non linear classifier has been considered even more necessary being the groups H and C more overlapped in the feature space than the subgroups H1 and C1.

The additional features considered for the H/C classifier have been: result of the H1/C1 classification, the max and min of the autocovariance function for the last hour of the ABPM, ABPD and HR time series, the percentage of ABPS higher than 110 and 120 mmHg in the last hour, the percentage of ABPS lower than 95 mmHg in the last hour, and finally the percentage of ABPM lower than 65 mmHg in the last hour.

3. Results

For the H1/C1 classifier (event 1), in the learning set (15 H1, 15 C1), all cases were correctly classified. After a first attempt on 16/04/09 obtaining a score of 80% on the event 1 test set (10 records), the algorithm was improved to obtain the one described in the previous section. This algorithm had the same performances on the learning set (100% accuracy) and produced, in the second try on 27/04/09, a perfect score on test set A.

The first version of the algorithm for the H/C classifier (event 2) was able to classify correctly 83.33% of the cases in the training set (30 H, 30 C). The algorithm assessment on the event 2 test set B obtained, at the first try on 27/04/09 a score of 75%. The algorithm was extended with further features, as described in the previous section, improving the accuracy on the training set (91.67%) but with the same results on test set B.

4. Discussion and conclusions

The excellent results in the event 1 strongly indicate the designed method could be suitable for clinical application. However, a larger dataset should be used in order to investigate further the algorithm and to better assess its performances. For such purpose the Mimic II database [4] could be exploited, once identifying the eventual AHEs in the records, extracting excerpts of such records and assigning them to subgroups H1, H2, C1 or C2.

The results obtained in event 2 are still satisfactory but lower than the ones obtained in event 1. It is opinion of the authors that these results could be further improved. In this study, some attempts of extracting new features for the event 2 have been already made. High-resolution data, after properly solving the problem of some intervals of missing data, might be used for the extraction of features from the spectral analysis of heart rate variability [5] in the last hours before T_0. Such approach has been tried in order to verify some possible differences between the sympathetic and parasympathetic indexes, but unfortunately without any significant result.

At this moment, the best approach to follow for the improvement of the results in event 2 seems to be the use of the clinical data, even if some problems exist about the reliability of the time references in such data, according to the challenge organizers. A first analysis, based on data independent from the time references, has already indicated that older patients and intubated patients have a higher probability to develop an AHE.

References

[1] PhysioNet/Computers in Cardiology Challenge 2009: Predicting Acute Hypotensive Episodes. http://www.physionet.org/challenge/2009/.

[2] GEMS: Gene Expression Model Selector. http://www.gems-system.org/.

[3] Statnikov A, Aliferis CF, Tsamardinos I, Hardin D, Levy S. A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis; Bioinformatics. 2005 Mar 1;21(5):631-43.

[4] The Mimic II Waveform Database. http://www.physionet.org/physiobank/database/mimic2db/.

[5] Pelosi G, Emdin M, Carpeggiani C, Morales MA, Piacenti M, Dattolo P, Cerrai T, Macerata A, L'Abbate A, maggiore Q. Impaired sympathetic response before intradialytic hypotension: a study based on spectral analysis of heart rate and pressure variability. Clin Sci 1999;96:23-31.

Address for correspondence

Franco Chiarugi
Institute of Computer Science (ICS),
Foundation for Research and Technology - Hellas (FORTH),
P.O. Box 1385, N. Plastira 100, Vassilika Vouton
GR 700 13 Heraklion, Crete, Greece
Tel.: +30 (2810) 391658
Fax: +30 (2810) 391428
E-Mail: chiarugi@ics.forth.gr