2012 PhysioNet Challenge: An Artificial Neural Network to Predict Mortality in ICU Patients and Application of Solar Physics Analysis Methods

Tom J Pollard[1,2], Louise Harra[1], David Williams[1], Steve Harris[3], Demetrio Martinez[1], Kevin Fong[2]

[1]Mullard Space Science Laboratory, University College London, UK
[2]University College Hospital, London, UK
[3]London School of Hygiene and Tropical Medicine, London, UK

Abstract

Advances in technology are improving the quality and quantity of data available in ICU, creating opportunities for the development of patient-specific predictive models to support clinical decision-making.

The 2012 PhysioNet Computing in Cardiology Challenge is to develop a patient-specific model for predicting in-hospital mortality using data collected during the first 48 hours of an ICU stay.

Our approach was to develop an algorithm incorporating an artificial neural network trained on features extracted from the patient data. We explored the stability of vital signs such as heart rate and blood pressure with a method previously used to detect frequency and intensity of solar 'nanoflares'.

The ability of the resulting model to predict outcomes of patients was evaluated. The model was most successful in Event 2 of the Challenge, receiving a score of 22.83 for Set B and 38.23 for Set C. For the model to be clinically useful and to improve on existing scoring systems such as SAPS, further work is needed.

1. Background

The past decade has seen rapid developments in computing power and digital networking capability, creating new opportunities for improving patient care. Advanced patient monitoring systems are now commonplace in hospitals, particularly in intensive care units (ICUs) where patients require close observation.

Despite these developments, the potential of digital information is not being fully realised. Humans are not able to fully process the large volumes of data available, while monitoring and alarm systems often operate independently meaning complex interactions between vital signs may go unnoticed.

Models commonly used to predict outcomes such as SAPS and APACHE provide useful insights over a population, but are too general to provide patient-specific estimation of outcome and risk [1,2].

Efforts to bring together data from multiple data sources systems, such as MIMIC II, present opportunities to develop more sophisticated models that are able to estimate patient-specific outcomes to assist the caregiver [3].

The PhysioNet/Computers in Cardiology Challenge 2012 calls for participants to develop models for patient-specific prediction of in-hospital mortality using data collected during the first 48 hours of an ICU stay [4,5].

Given the complexity of the data, we took a machine learning approach to the Challenge. Features were extracted from the patient data and used to train an artificial neural network (ANN).

Several methods were used to extract features from the data. The teams background in solar physics led to the application of a method used previously for detecting solar 'nanoflares' because of the similarity between solar data and time series patient data.

2. Method

2.1. Challenge data

Participants in the Challenge were provided with two datasets (Set-A and Set-B), each containing physiological data from 4000 ICU stays. Patient outcomes were available for Set-A (the training set), but patient outcomes for Set-B were not shared with participants. A third dataset (Set-C) was withheld from participants and was used by the organisers in independently scoring the final Challenge entries.

Up to 6 general descriptors were available for each patient stay (for example, ICU type, age, and gender), along with up to 37 time series variables (for example, heart rate, temperature, and respiratory rate). Only stays of 48 hours were included in the data and not all variables were available in all cases.

The Challenge consisted of two events: Event 1 required participants to submit an algorithm that outputs a prediction of in-hospital mortality (survival or non-survival) for each record. The algorithm was scored on its correct classification rate, which was evaluated by taking the 'best' result of sensitivity and positive predictivity. Event 2 required the algorithm to estimate risk of in-hospital death for each patient, with scoring based on performance against a range-normalised Hosmer-Lemeshow test to assess 'goodness of fit' of the predictive model.

2.2. Preprocessing

Data were processed to remove extreme outliers and obvious errors using a filter of normal values for each parameter. Missing information was replaced with mean values for the patient population, and body mass index was calculated from height and weight.

Clinical experience suggested that the nature of patients would vary greatly between the different ICU types and this was confirmed in the data. Patients in post-surgery recovery units for example were more likely to be mechanically ventilated. We therefore separated the dataset by ICU type prior to ANN training.

2.3. Features and network training

Features were extracted from the patient records using a variety of methods. These features included the mean values of parameters over 48 hours; the variance of time series data, and the moments of the gradients between time series values. Given the group's background in solar physics, we also took the opportunity to apply a method previously used by Harra et al for detection of solar nanoflares (Figure 1) [6].

One of the major science goals in solar physics is to determine how the solar atmosphere is heated and maintained at a higher temperature than the Sun's surface. One theory is that this temperature difference is a result of atmospheric heating by small-scale energy release events known as nanoflares.

Nanoflares are short-lived, lasting for tens of seconds to a few minutes, and occur as a result of collisions of magnetic fields. They are detected by looking for intensity increases observed in the high-energy end of the spectrum in ultraviolet or X-ray emission.

We applied a routine for detection of nanoflares to the patient dataset, since it was clear that in some parameters such as heart rate there were sudden increases and decreases during the 48 hour time period. We aimed to assess the number and intensity of these episodes during the stay as a measure of patient stability.

'Nanoflares' were detected firstly by determining a background level by fitting a line through the local points of minima. A minimum was defined as a point with three points on either side having values of emission measure greater than it. Peaks were defined as the points with a lower point on one side and two lower points on the other side, and in addition having flux at least 10% above the background level.

Combinations of features were selected as inputs for ANNs. Entry 1 incorporated mean averages of variables over the 48 hours, such as mean heart rate and mean glucose. All other entries incorporated mean averages as well as additional features, such as the mean height of maxima and variance from the mean.

Matlab's pattern recognition network function, `patternnet`, was used to generate feedforward ANNs for each ICU type [7]. These networks were then trained with scaled conjugate gradient backpropagation (Matlab's `trainscg` function) to classify features according to target patient outcomes [8].

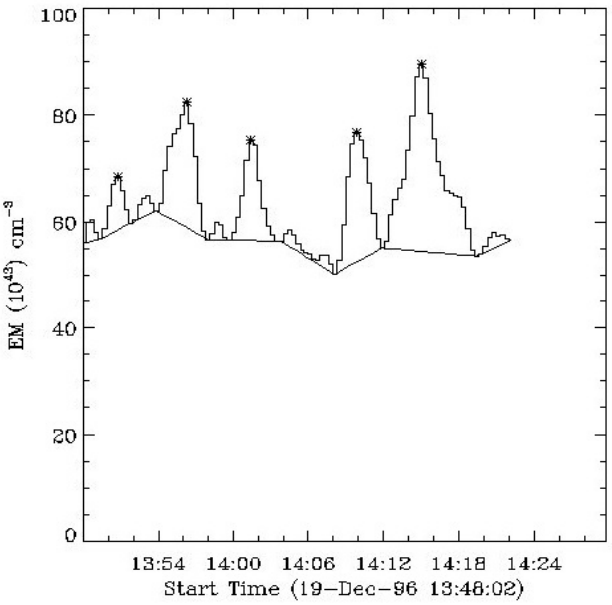

Figure 1. Example of solar nanoflares. The asterisks show the defined brightenings. The solid line shows the derived background. Reproduced from Harra et al, 2001 [6].

To train the networks, datasets were separated into 70% for training, 15% for validation, and 15% for testing. Performance of the networks was assessed according to the mean of square error on the validation samples. Training ended when the mean square error of the validation samples increased, indicating that generalization was no longer improving the network.

ANNs were independently trained for each ICU type, and in all cases a three-layer network (input layer, hidden layer, and output layer) was used, where the number of

neurons in the hidden layer was between 60% and 80% of the number of inputs (Figure 2).

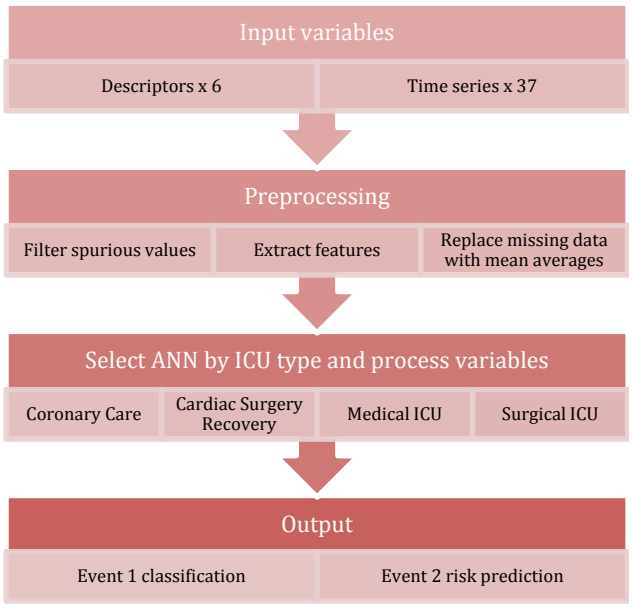

Figure 2. Summary of algorithms entered for the Challenge. Each algorithm processes input variables for individual records and outputs a classification for Event 1 (survival/non-survival) and an estimated risk for Event 2.

3. Results

The nanoflare detection routine was applied to variables including heart rate and blood pressure and successfully detected peaks in a number of cases (Figure 3). However, the limited time resolution of the available data meant that the method could only be applied to a limited number of records.

The ANN for Entry 1 took into account mean values of parameters over the 48 hours, while the ANN for entry 5 took into account a range of features such as the average height of maxima for mean arterial pressure and variance from the mean for blood oxygen saturation. For records where the time resolution was too low to extract features, the mean average value for the population was used.

The entries were most successful in Event 2, with the best scores being gained by entry 1 and entry 5 (Table 1). Entry 5 was placed 7th overall in the final scores based on Set-C.

Table 1. Scores for competition entries. In Event 1 an ideal score is 1. In Event 2 an ideal score is 0.

Entry	Event 1 Score	Event 2 Score
Entry 1 (Set-B)	0.27	22.83
Entry 5 (Set-B)	0.24	23.03
Entry 5 (Set-C)	0.23	38.23
SAPS-I Matlab (Set-C)	0.31	68.58
SAPS-I in C (Set-C)	0.31	35.21

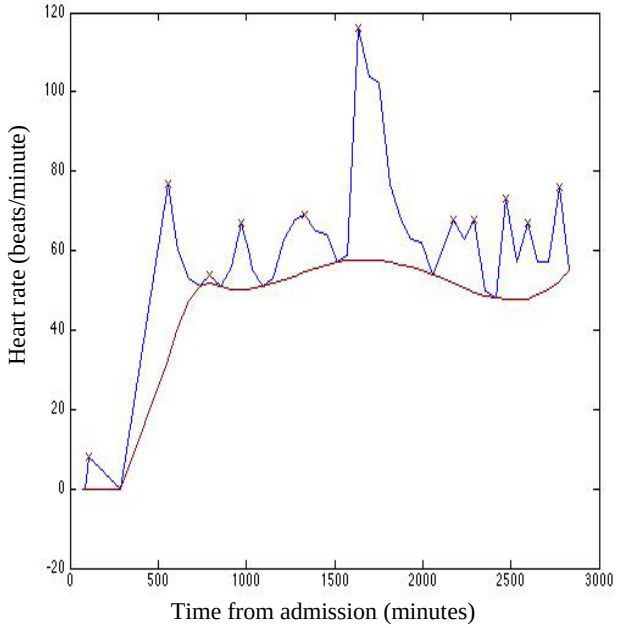

Figure 3. Example of the nanoflare detection routine applied to heart rate activity in a patient. The asterisks show the defined peaks. The solid red line shows the derived background.

4. Discussion

We took a simplistic machine learning approach and achieved reasonable performance in Event 2, but further work is needed to achieve a useful patient-specific predictive model. Currently our algorithms offer little or no predictive improvement on the SAPS-I algorithm.

Assessing patient stability through the application of the nanoflare routine would benefit from further exploration. Work is needed to identify whether this approach has any value in prediction of patient outcomes. In its current state the nanoflare routine is of limited benefit and is likely to mischaracterise stability in records with limited time resolution, where there are too few data

points to identify maxima and minima.

One of the benefits of scoring models such as SAPS-I is that they are straightforward to apply at the bedside and they are relatively easy to interpret. In contrast, the inner workings of machine learning approaches such as the one used here are less clear to the user.

Given that our collaboration is new, the progress made here is a good starting point and provides a basis for developing more successful predictive algorithms. Our focus will now be on building an archive of ICU and theatre data at University College Hospital and continuing to explore models for predicting outcome in patients.

Acknowledgements

The authors would like to thank University College London and the Mullard Space Science Laboratory for supporting our work. TJP is grateful for the award of a UCL impact studentship jointly supported by the Institute of Child Health. We would also like to thank Tom Bashford for helpful clinical advice.

References

[1] Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, Mercier P, Thomas R, Villers D. A simplified acute physiology score for ICU patients. Critical Care Medicine 1984: 12(11):975-977.

[2] Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE - acute physiology and chronic health evaluation: A physiologically based classification system. Critical Care Medicine 1981: 9(8):591-597.

[3] Celi LAG, Lee J, Scott DJ, Panch T, Mark RG. Collective experience: a database-fuelled, inter-disciplinary team-led learning system. J Comput Sci Eng, 2012: 6(1):51–59.

[4] Moody GB. The Physionet/Computing in Cardiology Challenge 2012. Computing in Cardiology 2012 [in press]

[5] Physionet Challenge 2012: http://www.physionet.org/challenge/2012/

[6] Harra LK, Gallagher PT, Phillips KJH. Characteristics of quiet Sun cell and network brightenings. Astronomy and Astrophysics 2000; 362, 371–378

[7] Matlab Neural Network Toolbox: http://www.mathworks.co.uk/products/neural-network/

[8] Möller MF. Neural Networks. 1993; 6: 525–533

Address for correspondence:

Tom J Pollard
University College Hospital
Centre for Anaesthesia Research Office
3rd Floor Maple Link Bridge
235 Euston Road
London NW1 2BU
tom.pollard.11 <at> ucl.ac.uk