A Neonatal Apnoea Monitor for Resource-Constrained Environments

Jonathan Daly[1], Violeta Monasterio[1,2,3], Gari D Clifford[1]

[1] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK
[2] CIBER de Bioingenieria, Biomateriales y Nanomedicina (CIBER-BBN), Zaragoza, Spain
[3] Aragon Institute of Engineering Research, Universidad de Zaragoza, Zaragoza, Spain

Abstract

A prototype Android application was designed to monitor for apnoea in neonates using a smartphone. The application receives data from a wireless pulse oximeter and uses machine learning techniques to detect apnoea. Distribution of the system requires only the pulse oximeter and a current mid-range smartphone. This work builds on previous research, but with a particular focus on classifying events accurately using a reduced set of information appropriate to a resource-constrained environment. This information consists only of the photoplethysmogram (PPG) and a set of PPG-derived physiological variables including heart rate and respiration rate. Various methods using the Support Vector Machine (SVM) were assessed using data from 27 annotated stays in a neonatal intensive care unit, divided approximately in half into training and test data. The best approach was found to be a combination of a feature selection method based on mutual information and an SVM with a radial basis function kernel, producing a classifier with a sensitivity of 98.7%, a specificity of 62.2% and a balanced accuracy of 80.5% on a training set of 796 events, and a sensitivity of 76.9%, a specificity of 52.0% and a balanced accuracy of 64.4% on a test set of 663 events.

1. Introduction

Apnoea (cessation of breathing) is a common issue with patients in the Neonatal Intensive Care Unit (NICU), particularly in infants born prematurely. Apnoea causes desaturation of oxygen in arterial blood, which can result in permanent damage to vital organs if the apnoea is not detected and resolved [1]. Although neonates in developed countries have access to abundant medical resources, those born in developing countries are far less likely to receive a high level of care. The ability to detect and intervene during an apnoeic event in a resource-constrained environment would therefore be of great benefit in such regions.

Neonates in developed countries are monitored in the NICU using a set of physiological waveforms. A conventional apnoea alarm sounds when a physiological signal crosses some threshold, for example when arterial oxygen saturation (SpO_2) drops below 90%. Because the measured signals are prone to noise (such as movement artefact) the resulting alarm has a very high rate of false positives. In some studies over 90% of alarms have been observed to be clinically insignificant [2].

Recently, more accurate ICU alarms have been created which utilise data fusion from multiple sources to greatly reduce the rate of false alarms in critical care units. A study by Monasterio *et al.* [3] used machine learning techniques to reduce the false alarm rate in the NICU. The authors used a feature selection method based on mutual information and combined it with a support vector machine (SVM) with a radial basis function, producing a classifier with a sensitivity of 86.2% and a specificity of 91.4% on the test data. Monasterio *et al.* emphasised the need for high-quality reference data, stressing the importance of representative training data and accurate annotation.

The approach used by Monasterio *et al.* [3] produced an apnoea monitor with considerably better performance than the traditional method of NICU monitoring, but used a full set of ICU waveforms to do so. The aim of this research, however, was to create an apnoea monitor for resource-constrained environments based on a significantly reduced set of signals (and therefore information). In order to represent such an environment this research describes the creation of a method of false alarm detection using only the photoplethysmogram (PPG) and PPG-derived variables.

An Android application was therefore created to acquire the PPG waveform from an external Bluetooth pulse oximeter and compute a series of physiological variables. During a suspected apnoeic episode these are passed to a classifier, and an alarm is sounded if the event is determined to be a real apnoea (rather than one caused by noise or movement artefact).

2. Materials and methods

2.1. Dataset

This study used physiological variables calculated from the PPG waveform recorded during 27 randomly-selected NICU stays. The data were obtained from Physionet's MIMIC II (Multi-parameter Intelligent Monitoring in Intensive Care II) database, which was created to assist with the development of intelligent patient monitoring in "critical care" environments such as the NICU [4].

The annotation process described by Monasterio *et al.* [3] was used to produce a reference data set of 1616 desaturation events ($SpO_2 < 90\%$). Within this set, 316 events were annotated as true positives and 1300 as false positives, giving an overall NICU false alarm rate of 80.4%. The reference set was then randomly divided into a training set of 796 events and a test set of 663 events, ensuring that all NICU visits were mutually exclusively allocated to either the training or test sets.

2.2. Physiological variable extraction

All physiological variables were derived from the PPG. A physiologically relevant period of 300s leading up to each desaturation was split into 20 non-overlapping 15s time-windows. Four groups of variables were calculated:

1. *Variables related to PPG-derived oxygen saturation:* For each window the minimum value of SpO_2 was computed (denoted by $min\ SpO_2$) and ordinary least squares (LS) regression was used to calculate the SpO_2 gradient (ΔSpO_2). Robust LS regression was found to be unnecessary due to the use of signal quality indicators [3].

2. *Variables related to PPG-derived heart rate:* The peaks of the PPG (corresponding to systole during the heartbeat) were detected using a modified version of an open-source beat detector for arterial blood pressure signals developed by Zong *et al.* [5]. A time and amplitude threshold adjustment to the PPG beat width and height created by Li and Clifford [6] was included. The heart rate was then calculated from the median peak-to-peak interval in the 15s window. The minimum instantaneous heart rate ($min\ HR_{PPG}$) in each window was calculated since we expected a bradycardic response during apneoa. Furthermore, the gradient of the heart rate (ΔHR_{PPG}) was calculated over the 15s window using a LS fit to the instantaneous heart rate vector, since we also expected to observe a strong negative gradient during the bradycardic response.

3. *Variables related to PPG-derived respiration rate:* A time series of differences between peaks and troughs of each beat was formed for each 15s window. An auto-regressive modelling approach was used to estimate the dominant frequency (per Nemati *et al.* [7]), which was taken to be the respiration rate of the segment. The variable was denoted by $min\ RR\_PDR_{RS}$.

4. *Variables related to PPG signal quality:* Several Signal Quality Indices (SQIs) were used to provide a measure of trustworthiness of the PPG signal, such as the spectral purity [7] of the PPG (SQI_PPG) and the PPG-derived respiratory signal (SQI_PDR_{RS}). We also included the four methods of Li *et al.* [6] who matched an average beat template (in each window) to all the beats in a window using cross correlation directly (SQI_PPG_{DM}), with linear resampling of each beat (SQI_PPG_{LR}), with dynamic time warping to stretch each beat (SQI_PPG_{DTW}), and with detection of signal saturation or 'clipping' (SQI_PPG_{CD}). Minimum, maximum, mean and median values were determined for each of these latter four variables. Lastly the Hjorth parameters as described by Gil *et al.* [8] were included, denoted by SQI_PPG_{H1} and SQI_PPG_{H2}.

In order to deal with missing data, the SQIs were set to zero when no source data was available, producing the full data set. A dimensionally-reduced feature set was created by performing univariate Receiver Operating Characteristic (ROC) analysis to choose the optimum evaluation interval for each physiological variable. Each variable was time-windowed every 15s and the time period with the highest area under the ROC curve (AUC_{opt}) was chosen to be used in the final algorithm. A feature was created for each variable by taking the value during the optimum evaluation interval. The resulting list of features is shown in Table 1 together with each variable's AUC_{opt} and corresponding window.

2.3. Feature selection and classification

In order to reduce the dimensionality of the classification problem to remove co-linear and confounding variables and prevent over-fitting, a feature selection method was used to identify a parsimonious subset. The 'minimum Redundancy, Maximum Relevance' (mRMR) method [9] was chosen for this purpose. The mRMR algorithm seeks to select a subset of features that best correlate to a target classification variable. The selection is subject to the constraint that the chosen features contain the minimum amount of redundant information. The algorithm outputs a score for each feature, and then ranks them.

The n highest-ranked features were then presented to a Radial Basis Function (RBF) Support Vector Machine (SVM) classifier for training. The RBF-SVM has two parameters, the degree of regularisation in the SVM (C) and the size of the RBF (σ), which were optimised by cross-validation when training the SVM.

The process of training the SVM using the rankings produced by feature selection was as follows:

1. For $n = 1...25$, the n highest-ranked features for each event were selected by the mRMR algorithm.

#	Variable	AUC_{opt}	Win
1	$min\ SpO_2$	0.552	19
2	ΔSpO_2	0.509	18
3	SQI_PDR_{RS}	0.663	5
4	$min\ RR\_PDR_{RS}$	0.634	3
5	SQI_PPG	0.782	3
6	ΔHR_{PPG}	0.569	1
7	$min\ HR_{PPG}$	0.645	1
8	$min\ SQI\_PPG_{DM}$	0.588	1
9	$max\ SQI\_PPG_{DM}$	0.585	1
10	$mean\ SQI\_PPG_{DM}$	0.634	1
11	$median\ SQI\_PPG_{DM}$	0.647	1
12	$min\ SQI\_PPG_{LR}$	0.615	1
13	$max\ SQI\_PPG_{LR}$	0.579	1
14	$mean\ SQI\_PPG_{LR}$	0.637	1
15	$median\ SQI\_PPG_{LR}$	0.640	1
16	$min\ SQI\_PPG_{DTW}$	0.603	1
17	$max\ SQI\_PPG_{DTW}$	0.592	1
18	$mean\ SQI\_PPG_{DTW}$	0.638	1
19	$median\ SQI\_PPG_{DTW}$	0.636	1
20	$min\ SQI\_PPG_{CD}$	0.596	1
21	$max\ SQI\_PPG_{CD}$	0.575	1
22	$mean\ SQI\_PPG_{CD}$	0.642	1
23	$median\ SQI\_PPG_{CD}$	0.609	1
24	SQI_PPG_{H1}	0.614	1
25	SQI_PPG_{H2}	0.634	1

Table 1. The dimensionally-reduced set of features. '#' is the feature number, 'Win' is the chosen evaluation period (where 1 is the closest window to the desaturation), and 'AUC_{opt}' is the area under the ROC for that window.

2. A grid search was performed to find the optimum values for the parameters C and σ. These values were chosen by using 10-fold cross-validation to evaluate the classifiers produced (with 'balanced accuracy', the mean of sensitivity and specificity, being used to assess performance).
3. The full training set was used, together with the chosen values of C and σ, to train an RBF-SVM. The trained SVM was then used to classify the unseen test set.

2.4. Android application

The mobile phone application was created for the Android platform. Since the application is targeted at both clinicians and non-clinicians, it was decided that the data should be displayed simply, alongside a series of graphs to show recent trends in the patient's history (Figure 1).

The application takes input from multiple sources, including the external pulse oximeter connected via Bluetooth, the smartphone's internal microphone and its internal accelerometer (which could also be used in the future to add additional information relating to respiration and movement). Signals derived from these inputs (heart rate, SpO_2, audio and actigraphy) are displayed as part of the user interface, and are also written to a series of text files stored on the phone. When a desaturation occurs, the required physiological variables are derived and passed into an SVM, which classifies the event as either true or false. Following a true event, the application sounds an alarm; this alarm can be audible and visible (as with a conventional bedside monitor), can initiate a phone call to a prespecified number (so that the receiver can listen in to the activity, and/or review the data remotely) or can send an SMS (text message) containing details of the event.

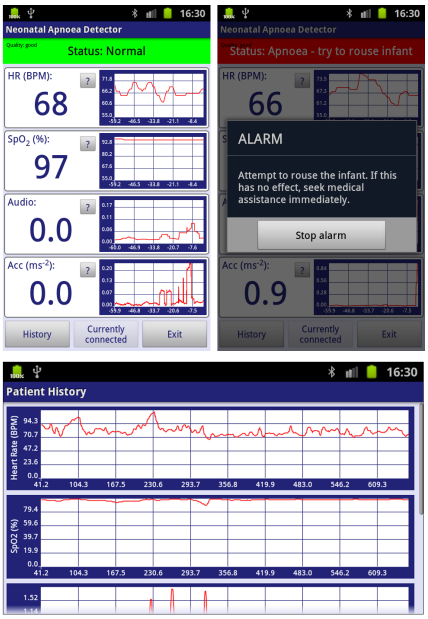

Figure 1. Screenshots from the developed Android application with a normal condition (upper left), an alarm condition (upper right) and historical data tracing (lower panel). A video demonstrating of the application can be found at http://www.youtube.com/watch?v=KdFflpxseDI.

3. Results

The best-performing classifier using the full feature set achieved a sensitivity of 98.7%, a specificity of 62.2% and a balanced accuracy of 80.5% on the training set. On the independent test set, it achieved a sensitivity of 76.9%, a specificity of 52.0% and a balanced accuracy of 64.4%.

The best-performing classifier using the dimensionally-reduced feature set used the eight highest-ranked features selected using mRMR as shown in Table 2. Using the training set it achieved a sensitivity of 98.1%, a specificity of 13.3% and a balanced accuracy of 55.7%. On the test set it achieved a sensitivity of 67.9%, a specificity of 57.7% and a balanced accuracy of 62.8%.

mRMR ranking	Variable
1	$mean\ SQI\_PPG_{LR}$
2	ΔHR_{PPG}
3	$min\ SpO_2$
4	SQI_PPG
5	SQI_PDR_{RS}
6	ΔSpO_2
7	$min\ RR\_PDR_{RS}$
8	SQI_PPG_{H2}

Table 2. The features used in the best-performing classifier trained on the reduced feature set.

4. Discussion and conclusions

The best-performing classifier achieved a balanced accuracy of 64.4% in testing. This classifier, however, was trained using the full 500-feature set, and not the reduced 25-feature set produced using ROC analysis (see section 2.2). If implemented using the phone application, the use of this full data set would require much more memory and processing time than the dimensionally-reduced set would. Classifiers which use the dimensionally-reduced feature set are therefore preferred for the purposes of real-time classification (as part of the Android application). The best classifier trained on the reduced set achieved a balanced accuracy of 62.8% in testing. Ranking features and training SVMs using the reduced set led to a considerably faster execution time with only a small drop in performance.

Existing NICU apnoea monitors produce almost no false negatives but many false positives, sounding an alarm for all desaturations. Using existing NICU methods, monitoring the NICU data used in this study would have resulted in a sensitivity of 100.0%, but a specificity of 0.0% on the test set. The classifiers produced here are therefore a clear improvement over simple apnoea alarms.

Table 2 indicates that both signal quality and physiological measurements are useful in reducing false alarms and identifying real alarm conditions. The results show, however, that the PPG waveform and its derived variables are insufficient to detect apnoea in a reliable and robust way.

Balanced accuracies of 60-62% suggest that the trade-off between sensitivity and specificity is too costly when using a single channel of information (the PPG). Further features may be required, most likely derived from other channels of information; these could include accelerometry, audio and even video, from which movement and respiration (breath sounds) can be evaluated. All of these parameters can easily be captured using the mobile phone.

The use of a smartphone allows the creation of a compact and portable system which uses existing infrastructure and functional supply chains (both rarities in most developing countries), and which can take advantage of the rapidly growing smartphone market in these regions [10].

In summary, the results of this research show that the use of the PPG alone is not sufficient for clinical classification of apnoea. They suggest, however, that with further work a cheap and accurate neonatal apnoea monitor could be produced for a resource-constrained region. Given the lack of alternatives in resource-constrained environments, there is much potential for leveraging the extensive computational, sensing and network capabilities of modern phones to address the lack of facilities for continuous physiological monitoring.

Acknowledgements

This work was supported by The John Fell Fund and the University of Oxford (UK), CIBER de Bioingenieria, Biomateriales y Nanomedicina through Instituto de Salud Carlos III and Fondo Europeo de Desarrollo Regional (Spain), Projects TEC2010-21703-C03-02 of Ministerio de Ciencia e Innovacion (MICINN), and GTC T-30 from DGA (Spain).

References

[1] Aggarwal R, Singhal A, Deorari AK, Paul VK. Apnea in the newborn. Indian J Pediatr 2001;68(10):959–962.

[2] Aboukhalil A, Nielsen L, Saeed M, Mark RG, Clifford GD. Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. J Biomed Inform 2008; 41(3):442–451.

[3] Monasterio V, Burgess F, Clifford GD. Robust classification of neonatal apnoea-related desaturations. Physiol Meas 2012;33:1503–1516.

[4] Saeed M, Villarroel M, Reisner AT, Clifford GD, Lehman L, Moody G, Heldt T, Kyaw TH, Moody B, Mark RG. Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database. Crit Care Med 2011;39(5):952–960.

[5] Zong W, Heldt T, Moody GB, Mark RG. An open-source algorithm to detect onset of arterial blood pressure pulses. Comput Cardiol 2003;30:259–262.

[6] Li Q, Clifford GD. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. Physiol Meas 2012;33:1491–1501.

[7] Nemati S, Malhotra A, Clifford GC. Data fusion for improved respiration rate estimation. EURASIP J Adv Sig Pr 2010;2010:1–10.

[8] Gil E, Vergara JM, Laguna P. Detection of decreases in the amplitude fluctuation of pulse photoplethysmography signal as indication of obstructive sleep apnea syndrome in children. Biomed Signal Proces 2008;3:267–277.

[9] Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE T Pattern Anal 2005;27:1226–1238.

[10] Wireless Intelligence. Global mobile connections surpass 5 billion milestone, Jul 2010. https://www.wirelessintelligence.com/print/snapshot/100708.pdf. Accessed 04/05/2012.

Address for correspondence: gari@robots.ox.ac.uk