Journal of Clinical Monitoring and Computing (2019) 33:39–51
https://doi.org/10.1007/s10877-018-0139-y

ORIGINAL RESEARCH

Forewarning of hypotensive events using a Bayesian artificial neural network in neurocritical care

Rob Donald[1] · Tim Howells[2] · Ian Piper[3] · P. Enblad[4] · P. Nilsson[4] · I. Chambers[5] · B. Gregson[6] · G. Citerio[7] · K. Kiening[8] · J. Neumann[8] · A. Ragauskas[9] · J. Sahuquillo[10] · R. Sinnott[11] · A. Stell[12] on behalf of the BrainIT Group

Received: 22 July 2017 / Accepted: 14 March 2018 / Published online: 24 May 2018

Abstract

Traumatically brain injured (TBI) patients are at risk from secondary insults. Arterial hypotension, critically low blood pressure, is one of the most dangerous secondary insults and is related to poor outcome in patients. The overall aim of this study was to get proof of the concept that advanced statistical techniques (machine learning) are methods that are able to provide early warning of impending hypotensive events before they occur during neuro-critical care. A Bayesian artificial neural network (BANN) model predicting episodes of hypotension was developed using data from 104 patients selected from the BrainIT multi-center database. Arterial hypotension events were recorded and defined using the Edinburgh University Secondary Insult Grades (EUSIG) physiological adverse event scoring system. The BANN was trained on a random selection of 50% of the available patients (n = 52) and validated on the remaining cohort. A multi-center prospective pilot study (Phase 1, n = 30) was then conducted with the system running live in the clinical environment, followed by a second validation pilot study (Phase 2, n = 49). From these prospectively collected data, a final evaluation study was done on 69 of these patients with 10 patients excluded from the Phase 2 study because of insufficient or invalid data. Each data collection phase was a prospective non-interventional observational study conducted in a live clinical setting to test the data collection systems and the model performance. No prediction information was available to the clinical teams during a patient's stay in the ICU. The final cohort (n = 69), using a decision threshold of 0.4, and including false positive checks, gave a sensitivity of 39.3% (95% CI 32.9–46.1) and a specificity of 91.5% (95% CI 89.0–93.7). Using a decision threshold of 0.3, and false positive correction, gave a sensitivity of 46.6% (95% CI 40.1–53.2) and specificity of 85.6% (95% CI 82.3–88.8). With a decision threshold of 0.3, > 15 min warning of patient instability can be achieved. We have shown, using advanced machine learning techniques running in a live neuro-critical care environment, that it would be possible to give neurointensive teams early warning of potential hypotensive events before they emerge, allowing closer monitoring and earlier clinical assessment in an attempt to prevent the onset of hypotension. The multi-centre clinical infrastructure developed to support the clinical studies provides a solid base for further collaborative research on data quality, false positive correction and the display of early warning data in a clinical setting.

Keywords Traumatic brain injury · Neuro-intensive care · Bayesian prediction · Clinical study results

1 Introduction

One of the primary aims in treating emergency admission patients with physical trauma is to stabilise vital physiological systems [1]. It is also recognised that patients who have suffered a traumatic brain injury (TBI) are at risk of having their "primary" brain damage exacerbated from a class of complications during neurocritical care (NCC) known as secondary insults [2]. The secondary insults can be both intracranial and systemic. The intracranial insults predominantly result from high intracranial pressure caused by mass lesions and ischaemic mechanisms resulting in brain swelling. One of the most critical types of systemic secondary insult is arterial hypotension, defined as abnormally low arterial blood pressure. Research has shown [3] that increased incidence of hypotensive episodes is related to poor neurological outcome. The physiological vital functions are extensively monitored in critical care and alarms

✉ Ian Piper
ian.piper@brainit.org

Extended author information available on the last page of the article

are sent out when a secondary insult, e.g. an hypotensive event, occurs so effective treatment can be initiated immediately without delay to reduce the development of secondary injuries. The ideal situation in the future would be to have a monitoring system that sends out alarms before a secondary insult will take place so relevant precautions can be instituted to avoid the secondary event completely.

The overall aim of this study was to get proof of the concept that advanced statistical techniques (machine learning) are methods that are not only able to provide early warning of impending hypotensive events before they occur, but can do so prospectively in a live neuro-critical care environment. This project had two specific aims: to develop a Bayesian artificial neural network (BANN) based system for prediction of hypotensive events; and to run a prospective multi-centre clinical study to test the system in a real clinical environment. For the clinical study a secure software and hardware infrastructure was developed and used for collection of anonymised data from hospital intensive care units (ICUs).

2 Materials and methods

2.1 Patient data

In 1997 an international group of clinicians and clinical scientists set up the BrainIT consortium [4] with an aim to collect high resolution physiological and clinical data from TBI patients in a standardized manner, and to use that data to undertake further research in this domain through hypothesis generation as well as the development and testing of new data analysis methods. The group defined a core data set [5] and then collected a database of over 200 patients containing demographic data (e.g. age, gender, time of injury, initial admission data both at pre-neurosurgical and neurosurgical admission), episodic data (e.g. treatments given, lab results from routine blood samples and nursing procedures) and minute-by-minute physiological data from bedside monitors [6]. The BANN was constructed using data from 104 patients from the BrainIT database, which was recruited from 22 European neuro-intensive care centers between March 2003 and June 2005 [6, 7].

To support a pilot study of the prediction system, 30 patients were prospectively recruited between September 2009 and October 2010 from six BrainIT centres (Prospective Phase I Dataset): Uppsala, Sweden; Glasgow, Scotland; Vilnius, Lithuania; Heidelberg, Germany; Monza, Italy; and Barcelona, Spain. Then in support of this data, 49 more patients were prospectively recruited from the same centers between December 2010 and July 2011 (Prospective Phase II dataset). The Phase II prospective study used a "Hanfelt" two stage design consisting of an initial collection of 13 patients (Phase II—Stage 1) followed by a subsequent 36 patients (Phase II—Stage 2). The Hanfelt design is used to minimise the recruitment of patients in Phase II trials [14]. See Fig. 1 outlining the design and timing of the clinical studies.

In the final analysis reported here the initial Phase I and repeat Phase II cohorts were combined. Ten patients had to be excluded due to excessive missing or invalid data (seven patients did not have the required minimum of 24 h

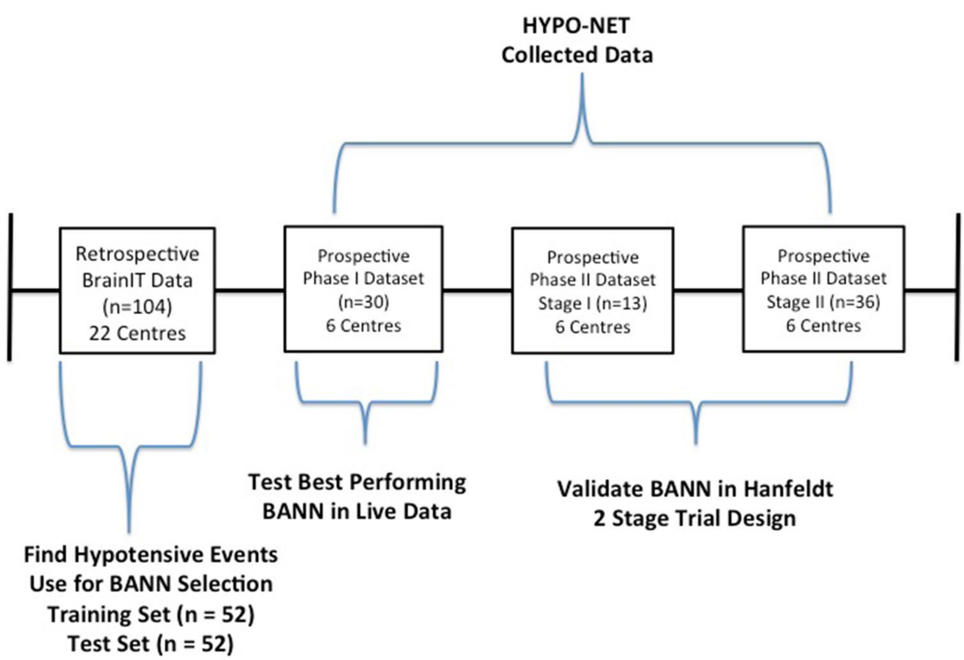


Fig. 1 Graphical representation of the clinical study design broken down by testing stage

of continuous monitoring data and three patients from one centre had technical data capture issues resulting in unrecoverable corrupted data), leaving a total of 69 patients for analysis. Of these 52 (75%) were male and 17 (25%) female. In terms of patient demographics, the median age was 40 years, range 16 to 83. There were two main types of injury: fall (30) and road traffic accident (25). The remaining 14 injuries were in the categories: pedestrian (4); unknown (4); assault (3); and sports related (3). The GCS scores, where complete, were obtained from two time-points: at admission to the pre-neurosurgery hospital, median GCS = 8 (range 3–15; complete n = 51); at admission to the neurosurgery ICU, median GCS = 8 (range 3–14; complete n = 24).

2.2 Hypotension definition

Although other studies have assessed machine learning approaches to hypotension prediction [19–25], these use hypotension definitions that are markedly more severe in terms of the depth of hypotension and their combined duration (often > 30 min). In the management of patients with TBI there is compelling evidence that even "minute by minute" hypotension events cumulatively can influence patient outcome [2]. Thus, in order to quantify the occurrence of hypotension the Edinburgh University Secondary Insult Grading (EUSIG) system was used, which defines hypotension as systolic arterial pressure (BPs) ≤ 90 mmHg or mean arterial pressure (BPm) ≤ 70 mmHg sustained for at least 5 min. This definition has been shown to be a clinically relevant measure when assessing the burden of hypotension [2]. The EUSIG definition requires that blood pressure return to above threshold for at least 5 min before the hypotension event is considered to have ended. For the purposes of these studies we have increased this "switch off time" to 15 min, because we found that otherwise we created clusters of closely related events that were better considered as a single episode.

2.3 Bayesian artificial neural networks (BANN)

The term artificial neural network incorporates a variety of techniques, but most often (and in the present case) describes a form of nonlinear regression. Neural networks are very general and powerful models, capable of modelling any continuous function [8]. They have consistently proved to be among the best techniques for difficult problems in statistical pattern recognition [9, 10]. Bayesian analysis can be used to estimate error bars for neural network predictions, and to mitigate problems like over-fitting that can arise due to the power of the models [11, 12]. We have adapted a Bayesian system designed and implemented by Radford Neal that utilizes a hybrid Monte-Carlo technique [12].

2.4 BANN construction and testing

The inputs to the model were built from a series of moving windows (based upon minute by minute data sampling) configured across the data.

Inputs for the neural network were taken from the demographic database or formed by extracting features such as mean, standard deviation, and slope from the physiological data over a time window or a series of windows prior to the prediction. The slope feature is the slope of a linear trend line with a least-squares fit over the data in the given time window. A set of BANNs were constructed using clinical parameters selected through a combination of clinical guidance and exploratory modelling for the selection of input features.

A cohort of suitable patients (n = 104) having more than 24 h of data for heart-rate and mean and systolic blood pressure was extracted from the BrainIT database [6]. A total of 33 models were trained on the data from 52 randomly selected patients (the 'training' set) and tested on the remainder (the unseen 'test' set) all 33 models and their parameters are summarised in Appendix. This was a conservative approach and possibly sub-optimal given available techniques such as cross validation to make best use of a relatively small data set. However, given the cautious nature of the clinical community, one which is more likely to feel comfortable with a prediction model based on incorporating pertinent physiological parameters derived from clinical guidance, it was felt that this approach to creating a model for early warning of hypotensive events would be easier to gain acceptance clinically.

The model with the best performance on the test set and good coverage with respect to completeness of data within the BrainIT database was selected as the predictive model to be used in the Phase I and repeat Phase II clinical studies. The model selected used demographic parameters of age and gender with the physiological variables of heart rate, systolic and mean blood pressure. Models including parameters other than blood pressure and heart rate were tested. Additional parameters (e.g. ETCO2, temperature, SAO2, admission time to ICU) were tried but they did not produce models with better sensitivity. Often, due to missing data, these models could only be trained and tested on subsets of the complete training and test sets.

2.5 Clinical study methodology

Figure 1 below summarises the clinical study design. A Phase 1 study of 30 patients was conducted to ensure that the data acquisition and prediction software could be installed in the clinical environment and produce predictions on live clinical data being collected from the patient monitors, considering all the data quality issues that are present in an operational setting. Details of the participating hospitals and their hardware and software infrastructure can be found in [13].

The Phase I study was successful, with the prediction results matching the results from the test on the retrospective data in the BrainIT database. Therefore we began a repeat study (Phase 2), using the methods detailed in the paper by Hanfelt [14] which is a modification of the paper by Simon [15].

2.6 Sample size analysis

Our end-point was defined as the sensitivity of true predictions of hypotensive events. In order to minimise the recruitment of patients, we opted for an oftenused Phase II trial design based upon the "Hanfelt design" modified from the method first described by Simon et al. This required a stage 1 sample size of 13 and, assuming a positive outcome in stage 1 of the study, a second stage of 36 patients to give an overall minimum sample size of 49. The sample size calculations and supporting tables are fully described in the referenced paper by Hanfelt et al. [14].

2.7 False positive correction

To minimise false positives, two checks were performed on the arterial blood pressure readings: (i) *BP pulse* < 10 mmHg, if the pulse pressure (BPs-BPd) is < 10 mmHg, this potentially indicates a fault within the pressure measuring system resulting as often found in an overdamped trace and therefore any warnings that occurred under this condition are ignored. (ii) *BPm* > 90 mmHg, if the mean arterial blood pressure is > 90 mmHg when the warning was generated a hypotensive event is unlikely to occur in the near future and therefore this warning can be discounted. The false positive correction (FPC) was not used during BANN model development. The various models developed during the research phase were assessed on their sensitivity/specificity scores. While true the post-hoc processing influences the real-world results, it is valid as exploratory analyses of how to improve the clinical applicability for future projects. In reality, the FPC corrections although improving the prediction performance, did not make a significant improvement at the decision thresholds used.

2.8 ROC confidence intervals

Confidence intervals were calculated for sensitivity and specificity estimates using the bootstrap methodology described in [16].

2.9 HYPO-NET

HYPO-NET—a previously developed data collection framework was used for testing of the BANN models in real-time in a live clinical environment [13] (Fig. 2). To be clinically useful, not only do predictive models need to be developed with appropriate prediction accuracy, but it must also be shown that it is feasible to develop a data collection platform capable of running these BANN models in real time in a live clinical setting. It is one thing to show that predictive models have good accuracy on cleaned retrospective data and another to show that the same models perform equally well in a live clinical environment particularly one as technically challenging as an intensive care unit.

2.10 Study ethics

Each local hospital ethics and R&D committee's approved the patient recruitment protocol and consent documentation. As patients in coma are considered "Vulnerable Individuals", Local ethics committee approval was obtained at each centre. Patient "assent" was required for every patient by interview with patients relatives or next of kin.

3 Results

3.1 Model selection

The model selected for prospective testing is described in Table 1 below. Two demographic variables; age and sex are included. Additionally seven physiological features based on mean and systolic blood pressure and heart rate were extracted from two sub-windows of 15 min of data starting at 15 and 30 min before the prediction time. This gives a total of 16 inputs to the neural network for each prediction.

3.2 Predictive accuracy: bann test set

In the next sections we present results using a standard visual technique to assess the trade off between true positive identifications of episode starts and false alarms, a plot known as a receiver operating characteristic (ROC) curve. The ROC curve plots the specificity (false positive rate) on the x-axis and the sensitivity (true positive rate) on the y-axis. The cross markings on the plot indicate probability decision thresholds and move from a decision setting of probability zero (we will accept all predictions) at the top right of the plot, to a probability of one at the bottom left of the plot (we would reject all predictions).

The ROC curve for the BANN test set results (data from the retrospective BrainIT database) is shown in Fig. 3. A review of the prediction results suggested that the BANN

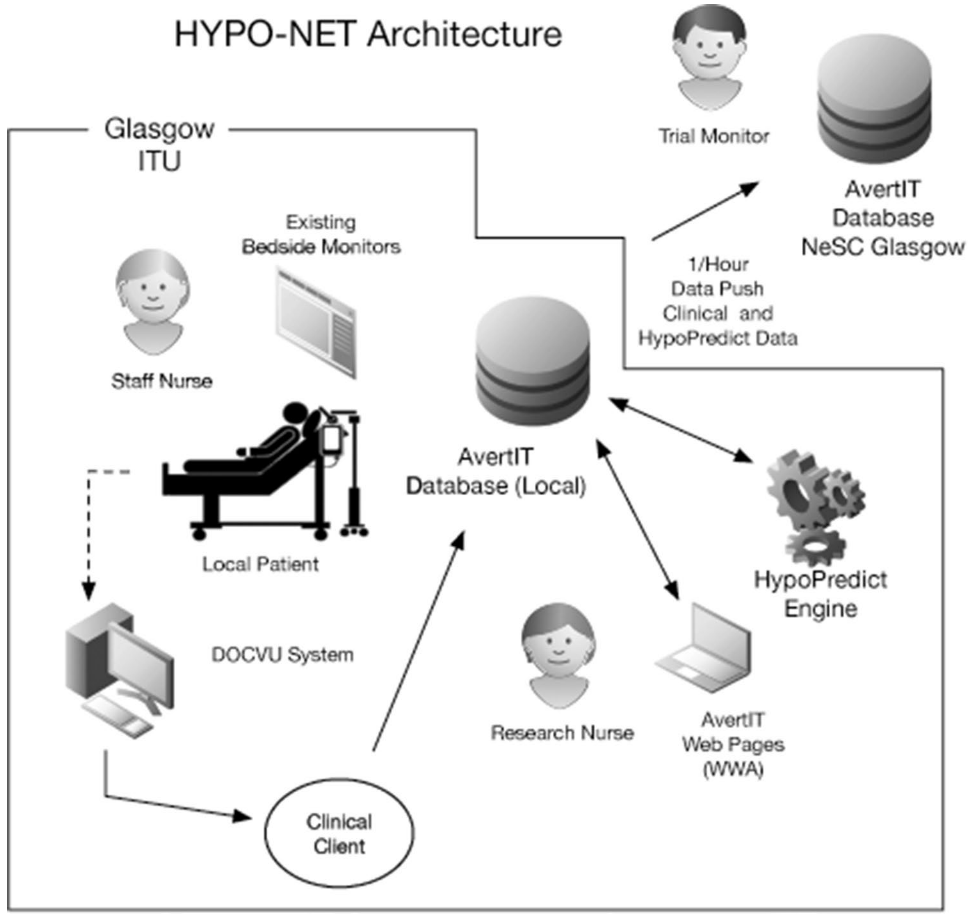


Fig. 2 HYPO-NET system architecture in one of six BrainIT centres. Staff ICU nurse log patient's onto local information systems. Research Nurse log patient's onto a patient study via the ward web app (WWA). This triggers "Clinical Client" software to parse clinical, laboratory and monitoring data from local hospital based systems and push data onto a local HypoNet SQL database. Non-hospital Research data is entered via the WWA program. Analysis software calculates project specific derived indices from data acquired from the local SQL database. Once per hour "Data Push" software anonymises local data and pushes clinical, monitoring and status log files out of hospital firewalls and up to a central SQL database sitting at the National eScience (NeSC) centre in Glasgow University. A trial monitor can view and run queries on HypoNet data arriving from all six BrainIT centres in quasi-real time

Table 1 BANN model parameters

Variable	Feature
Age	Demographic
Sex	Demographic
Mean arterial blood pressure	Mean
Mean arterial blood pressure	Slope
Systolic arterial blood pressure	Mean
Systolic arterial blood pressure	Slope
Heart rate	Mean
Heart rate	Slope
Heart rate	Standard deviation

Input features for the model selected for the clinical studies. The mean, slope and standard deviation features were calculated over two sequential time windows for the seven physiological variables: from 30 to 15 min before the prediction time, and from 15 min before up to the prediction time. This produced two demographic and 14 physiological variables as inputs to the neural network

could achieve a target for sensitivity of 30% with specificity at around 90%. Using the ROCR (R) package, an area under curve value of 0.74 was generated.

3.3 Predictive accuracy: clinical studies

After the encouraging results of the model on the BANN test set, the project moved to the clinical environment with a prospective study collecting data using HYPO-NET on 30 patients (Phase 1). The ROC curve for this study is shown on the left hand side of Fig. 4. The results for the clinically interesting region of false positive rate 0.3–0.4 are, unusually, slightly better than predicted on the test set despite being conducted in a live clinical setting. Improving upon a training model performance in a clinical setting is unusual and so we felt it was important to assess the repeatability of this result. Thus, the next task was to see if we could repeat these results on another group of patients (Phase 2). Using a study design given in Hanfelt [14], 49 patients were recruited. Data from ten patients was excluded due to problems with data collection. The results of the repeat study (Phase 2) are shown on the right hand side of Fig. 4. Again, the results are slightly better than the model test set and similar to the Phase 1 study.

We note that the AUC scores for the two clinical studies are less than the training set. This is a criticism of using the AUC metric and is noted in the literature [34, 35] that there are problems in comparing classifier performance

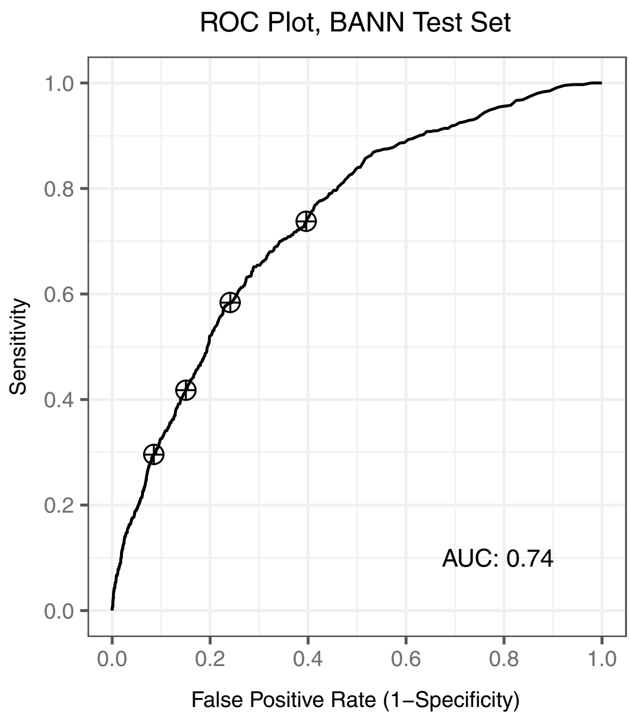


Fig. 3 Predictive accuracy from BANN test set. The decision threshold settings are marked with the crossed circles. The decision thresholds shown start at 0.1 at the upper most mark and increase to 0.4 in the lower left part of the trace. Using the ROCR (R) package, an area under curve value of 0.74 was generated

particularly when the ROC curves cross. For our purposes we are interested in false positive rates of between 0.3 and 0.4.

Combining the prediction data from the prospective pilot (Phase 1) and repeat pilot (Phase 2) studies produces the ROC curve plotted in Fig. 5.

Figure 6 shows the ROC curve over the decision threshold range of 0.2–0.5 with point estimates and confidence intervals as error bars on the results trace. It can be seen from Fig. 6 that using a decision threshold of 0.4, with false positive techniques applied, the target sensitivity (> 30%) and specificity (> 90%) was achieved.

3.4 Precision recall plot: clinical data

An alternative presentation of the predictive performance of the model is shown in Fig. 7. This plot shows precision aka: positive predictive value (PPV), defined as true positives/(true positives + false positives) on the y axis versus recall aka: sensitivity, defined as true positives/(true positives + false negatives) on the x-axis. Although Fig. 7 indicates low values of precision in the likely operating range of 0.3–0.4 for recall (sensitivity), the plot, similar to the ROC presentation also suggests that > 30% of the warnings would be clinically relevant.

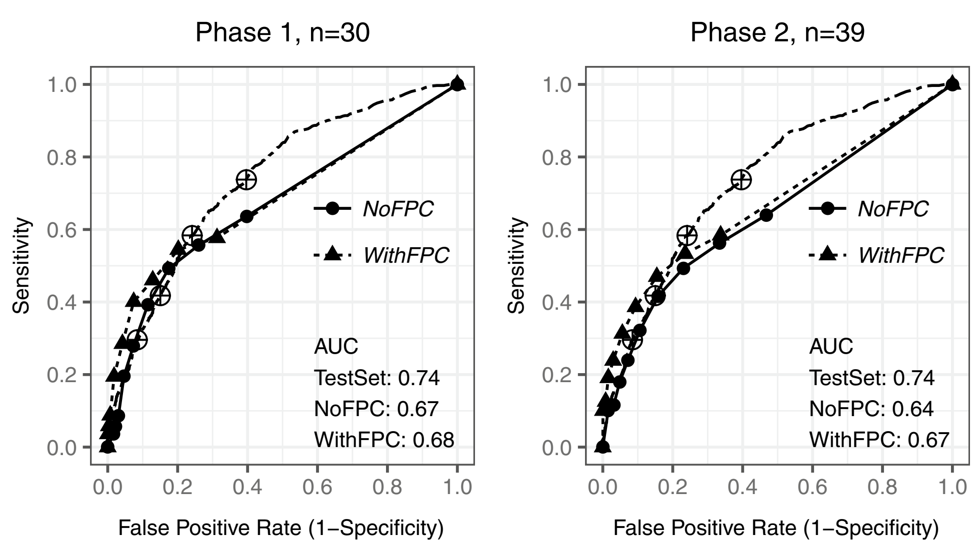
3.5 Episode warning time: clinical data

Finally, Fig. 8 below presents the warning time for patients where the system was able to provide early warning of the onset of hypotension.

4 Discussion

It is known that the cumulative duration of arterial hypotensive events is a prognostic factor for their clinical outcome following a head injury [2, 3]. The importance of reducing the burden of hypotension is not restricted to the TBI population; for with the medical management of sepsis, it has been shown that duration of hypotension before initiation

Fig. 4 ROC plots from clinical studies. The solid line with filled circles show the results obtained with no false positive correction (FPC) applied. The dashed line with filled triangles show the results when FPC techniques are used. This results in a better sensitivity for a given decision threshold. The BANN test set results are shown with the dashed dot line and crossed circles for comparison



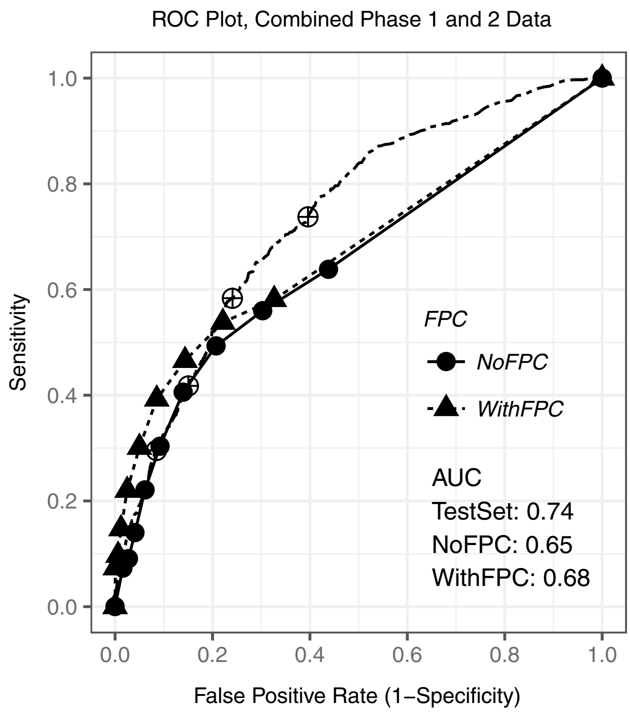


Fig. 5 ROC plot from clinical data. The decision threshold markers indicate settings from 0 in the top right to 1 at the bottom left. The step size is 0.1. Decision threshold markers: crossed circles for BANN test set data; solid circle for clinical data with no false positive correction (FPC); solid triangle for clinical data with false positive correction

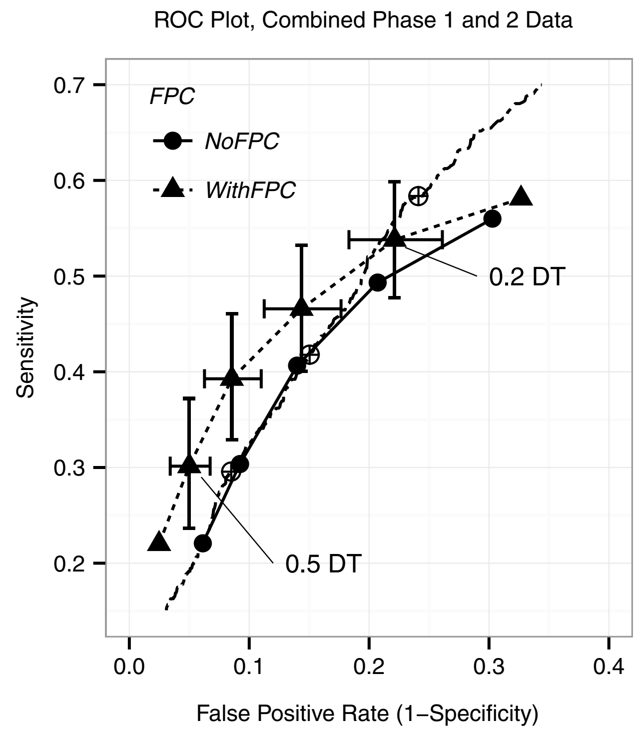


Fig. 6 ROC plot with error bars. Using combined phase 1 and 2 clinical data, this plot is a zoomed in version of Fig. 4 showing error bars for the decision threshold (DT) settings of 0.2–0.5

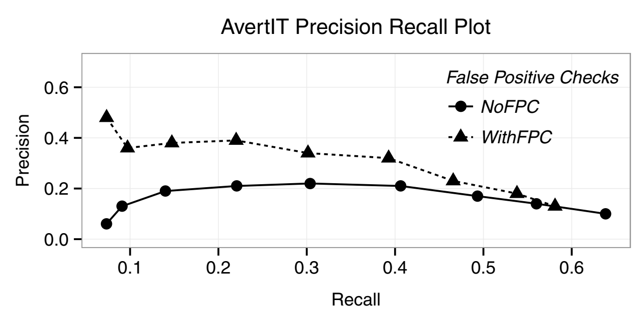


Fig. 7 Precision recall plot. An alternative visualization of prediction performance. This plot indicates low values of precision in the likely operating range of 0.3–0.4 for recall (sensitivity), the plot, similar to the ROC presentation also suggests that > 30% of the warnings would be clinically relevant

of effective antimicrobial therapy is the critical determinant of survival [17].

There is no universally accepted definition of arterial hypotension used during the clinical management of TBI patients. Bijker et al. found almost 50 definitions being quoted in the surgical literature [18], many of these use hypotension definitions that are markedly more severe in terms of the depth of hypotension and their combined duration (often > 30 min) than those considered relevant in the neuro-critical care of TBI patients. We carried out our own survey of the current practices within the participating centres, which indicated that both systolic and mean arterial pressure should be monitored and considered in the definition of hypotension and therefore it was decided to use the published Edinburgh University Secondary Insult Grading (EUSIG) definition [2], which requires both measurements.

Our results and the predictive variables selected (Table 1) are generally consistent with previous work. A study of hypotensive episodes during general anesthesia found that baseline mean arterial pressure and patient age were significant predictors of outcome [6]. Several studies have found that heart-rate variability can be used as a predictor of hypotension [19]. The standard deviation of the 1-min trend values of heart rate incorporated in our model is a crude measure of heart-rate variability.

The choice of a neural network model is supported by the results of a competition sponsored by the PhysioNet group [20] that tested the ability of various models to predict hypotension in intensive care patients. A neural network entry won this challenge [21], although optimized rule-based approaches also worked well [22, 23]. An earlier study also demonstrated that a neural network model was an effective

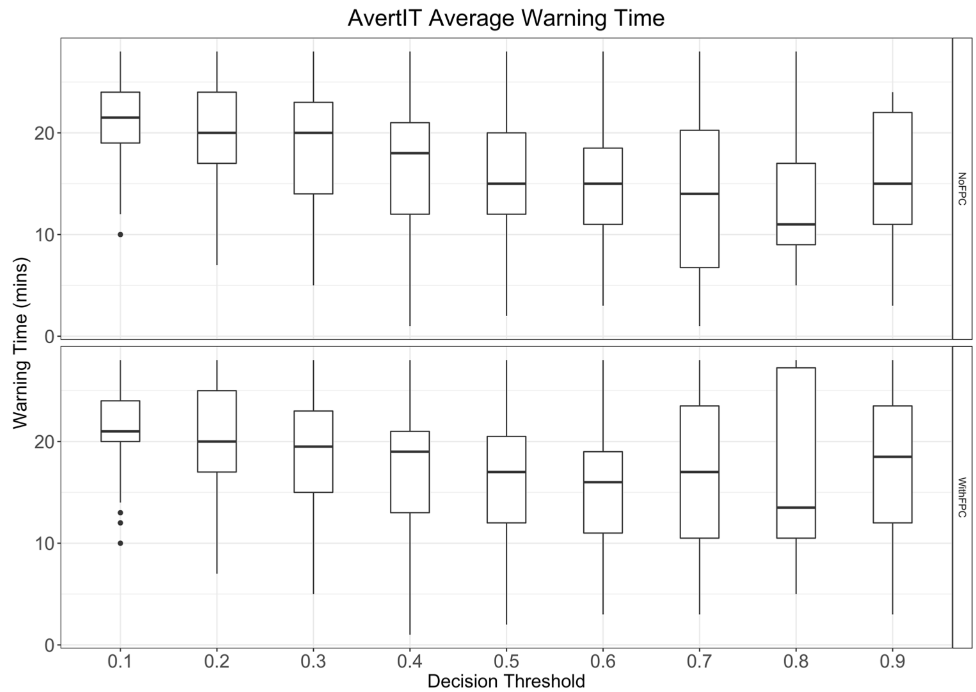


Fig. 8 Box plot of warning time for varying decision thresholds. Note that in the clinically relevant decision threshold region of 0.3–0.4, the average hypotension warning time was 15 min or greater

predictor of hypotension during haemodialysis treatment [24].

A distinguishing feature of this study is that predictions are produced and evaluated continuously on a minute-by-minute basis. Previous studies have made a single prediction for each patient based on demographic and physiological indicators at a point in time either immediately preceding a specific procedure [24–26], or chosen by the competition organizers [20–23]. The single prediction was then evaluated based on the occurrence of hypotension during the procedure (in the former case) or during the entire remaining monitoring period (in the latter case). This makes a meaningful comparison with previous results difficult, but it also makes our results of special interest in terms of fielding a practical intensive care system. A second distinguishing feature of this work is that a real-time system (HYPO-NET) has been implemented and tested running live in intensive care units in a multi-center study. Although developed in patients with brain injury, there is every possibility that this technology will also find application in general intensive care where the optimal management of the patient's blood pressure is ubiquitous and paramount.

One could argue that a sensitivity of only 35–40% is low. However the ability to pick up even one potentially emerging hypotensive event in three before it has happened is clinically important. We know from our research that on average a patient will have 8–10 hypotensive events during their management in the ICU. The system will forewarn accurately 3–4 of these events and give the nursing and medical staff more time to assess the patient and potentially prevent the event from occurring, or, to treat it sooner and less aggressively. In the medical management of sepsis, it has been shown that with each hour delay in treatment of hypotension leads to an increase in in-hospital mortality of 7.6% [17]. Another potential clinical outcome of this work is that fostering earlier and less aggressive intervention will in turn lead to a quantifiable reduction in the duration of average patient stay, with associated reduction in cost of intensive care management (in 2005, typically estimated at c. £1200/patient/day) [27]. Another point worth discussion is with regards to the poorer performance of the system at the upper sensitivity ranges The BANN ROC is a more simplistic score of exact prediction of a hypotensive event in 15 min time whereas the scoring of the clinical data uses a more complex decision process. Ultimately, our focus upon the decision threshold of 0.3–0.5 is driven by clinical pragmatism of requiring an acceptably low false-positive rate and thus the poorer performance at higher sensitivities although clearly desirable is not relevant.

There is no doubt that the technology could be improved further by research into reducing the influence of false-positives (when the device claims a hypotensive event is about to occur but then it does not). In this study we have not attempted to distinguish between good and bad false-positives. Many predictions recorded as false positive closely precede, by just a few minutes, a true positive prediction, for example. Further clinical studies are needed in which the clinical staff can classify an alarm as a valid warning or a nuisance. A major obstacle to reducing nuisance alarms is the problem of artifactual and missing data. Blood pressure measurement systems are sensitive to patient movement such as that produced by nursing and medical procedures (e.g.: blood sampling, routine patient hygiene or pressure area care) that either interrupt recording or cause movement of the recording system. Automatic false positive correction is challenging. Many authors agree on the need for systems that can automatically detect and remove non-physiological artefact, yet due to the inherent plethora of causes of artefact within a clinical environment such as an intensive care unit, a robust general purpose system does not exist. Some authors (e.g.: Williams et al.) have successfully designed and tested dynamic linear models trained on specific artefact types (such as those caused by blood sampling from the arterial pressure line), but many of these approaches require relevant centre specific training data [29].

The problem of invalid data may have caused the incidence of hypotension to be underestimated in this study. Analysis of the dataset for the presence of hypotension events using the EUSIG classification has detected 2081 hypotension events from 100 patients with full BP systolic, diastolic and mean data parameters. In this dataset, additionally there were 600 "Abandoned events" defined as where an event was in the "Active" state but the next readings did not contain valid measurements for BPs, BPd or BPm. There were also 47 "Invalid Events" defined as where the event ran to completion but there were missing time blocks of data [28]. Thus taking the worst-case scenario that all abandoned and invalid events would have become hypotension events indicates that 31% of potential hypotension events are lost to detection and potentially to medical treatment. This is compelling evidence for the need for advanced statistical methods for coping with artifactual and missing data. Closely aligned and following on from this study, Williams et al. developed and tested a descrimitive switching linear dynamical system model to automatically detect both arterial pressure artefact due to line interruption during blood sampling and pressure line damping events [29]. Work such as this renders the ability to run predictive models even on physiological data channels known to be susceptible to artefact, more feasible.

Another area that could benefit from further research is that concerning user interface. How best to present critical warnings/predictions to intensive care medical staff? ICU staff are already bombarded with hundreds of buzzers, bells and lights from a plethora of medical devices surrounding the patient. The technology used in this study and its ability to predict hypotension before it has emerged is only part of the challenge. The other part is how to ensure the medical staff respond to the increased risk warning and adhere to an expected guideline or response which can be quantified as part of a future randomised clinical trial on the technology used in this study.

The technology used in this study has been shown to be capable of forewarning arterial hypotension events with > 15 min of warning. By adjusting the decision threshold within the range 0.3–0.5, which could be carried out on a patient-by-patient basis, an operational system could expect to provide warning to a clinical team somewhere between 10 and 15 min in advance of a hypotensive episode. Adjustment of the decision threshold would, of course, have effects on system sensitivity and specificity. As an example, with a decision threshold of 0.4, a sensitivity of 39.3% (95% CI 32.9–46.1) and specificity of 91.5% (95% CI 89.0–93.7) was achieved. The average warning at this setting was 14 min. This result has been obtained from two independent patient cohorts using real-time data collection in a live clinical environment. Before a randomized controlled trial can be designed to test this technology to assess if the warnings delivered influence medical management and patient outcome, two areas of further research are warranted: (a) research towards implementing artefact detection and removal from BP data such as that developed by Williams et al. [29] and (b) research on optimising the display of hypotensive warnings to medical staff using BANN prediction data. There is no doubt that a revolution is currently taking place in healthcare; technology is enabling the creation and collection of vast amounts of patient data, as part of a "Big Data" phenomenon [30]. This rapid expansion of data has the potential to transform clinical practice, no more so than in the management of patients within critical care [31–33]. Critical care is arguably one of the most technology led medical domains; sophisticated patient monitoring systems can capture large volumes of streaming physiological data which continuously report on a patient's status and are often combined with patient data captured in electronic health records. The potential clinical benefit of harnessing this large amount of physiological data available will inevitably require the real-time application of machine learning technology able to run unattended within a live clinical environment. The success of such an approach will require the collaboration of clinicians, computer scientists and technologists and will depend upon their shared understanding of each others understanding of the areas of unmet need within their own domains. Machine learning algorithms are crucially dependant on the quality of the data available for model training and a solid understanding of the requirements of the busy clinician. This work has demonstrated that machine learning algorithms can be developed and used prospectively within a live critical care environment. The next crucial step is to demonstrate their robustness and clinical effectiveness.

5 Conclusions

Proof was gained of the concept that advanced statistical techniques (machine learning) are methods that are able to provide early warning of impending hypotensive events before they occur and that they can be used effectively in a live clinical setting, further prospective multi-centre clinical work is required to demonstrate the robustness of running such a model in a live clinical setting.

Acknowledgements Funding: European 2020 IST Programme - IST-2007-217049; Investigators and participating centers: Barcelona, Spain: Prof. Sahuquillo; Cambridge, UK: Prof. Pickard; Edinburgh, UK: Prof. Whittle; Glasgow, UK: Mr. Dunn; Gothenburg, Sweden: Dr. Rydenhag; Heidelberg, Germany: Dr. Kiening; Iasi, Romania: Dr. Iencean; Kaunas, Lithuania: Prof. Pavalkis; Leipzig, Germany: Prof. Meixensberger; Leuven, Belgium: Prof. Goffin; Mannheim, Germany: Prof. Vajkoczy; Milano, Italy: Prof. Stocchetti; Monza, Italy: Dr. Citerio; Newcastle upon Tyne, UK: Dr. Chambers; Novara, Italy: Prof. Della Corte; Southampton, UK: Dr. Hell; Uppsala, Sweden: Prof. Enblad; Torino, Italy: Dr. Mascia; Vilnius, Lithuania: Prof. Jarzemaskas; Zu¨rich, Switzerland: Prof. Stocker.

BrainIT Steering Group members: IR Chambers, Newcastle upon Tyne; G Citerio, Monza; P Enblad, Uppsala; BA Gregson, Newcastle upon Tyne; T Howells, Uppsala; K Kiening, Heidelberg; J Mattern, Heidelberg; P Nilsson, Uppsala; I Piper, Glasgow; A Ragauskas, Kaunas; J Sahuquillo, Barcelona; YH Yau, Edinburgh.

BrainIT Group – Data Contributors: Professor Per Enblad, Department of Clinical Neurosciences, Section of Neurosurgery, Uppsala University Hospital, S-75185 Uppsala, Sweden. Phone: +46 18 611 0000 Fax: +46 18 558617 Email: Per.enblad@neurokir.uu.se;Dr Karl Kiening, Department of Neurosurgery, Im Neuenheimer Feld 672, 69120 Heidelberg, Germany Phone: +49 30 4505 60724, Fax: +49 30 4505 6072, Email Karl.Kiening@med.uni-heidelberg.de; Dr Giuseppe Citerio, Dipartimento di Anestesia e Rianimazione, Azienda Ospedaliera, Ospedale San Gerardo, Via Donizetti 106, 20052 Monza (Milano), Italy. Phone: +39 029 233 3293, Fax: +39 029 233 3293, Email: g.citerio@hsgerardo.org; Professor Juan Sahuquillo, Department of Neurosurgery, Neurotraumatology Research Unit, Institut Catala de la Salut, Paseo Vall d'Hebron, 119-129, 08035 Barcelona, Spain. Phone: +34 93 489 3512, Fax: +34 93 489 3513, Email: sahuquillo@wanadoo.es; Professor Arminas Ragauskas, Telematics Science Laboratory, Kaunas University of Technology, Studentu 50-449, LT3031 Kaunas, Lithuania. Phone: +370 7 736897, Fax: +370 7 736 6869, Email: telematics@tef.ktu.lt; Dr Iain Chambers, Regional Medical Physics Department, Newcastle General Hospital, Westgate Road, Newcastle upon Tyne, NE4 6BE, UK. Phone: +44 191 273 8811, Fax: +44 191 226 0970, Email: Iain.Chambers@nuth.northy.nhs.uk; Professor Ian Whittle, Department of Clinical Neurosciences, Western General Infirmary, Crewe Road, Edinburgh, UK. Phone: +44 131 537 2103, Fax: +44 131 537 2561, Email: irw@skul.dcn.ed.ac.uk; Professor David Wyper, Department of Clinical Physics, Institute of Neurological Sciences, Southern General Hospital, 1345 Govan Road, Glasgow G51 4TF, UK. Phone: +44 (0) 141 201 2105, Fax: +44 (0) 141 201 4127, Email: Gpca12@udcf.gla.ac.uk; Dr Michael Kiefer, Department of Neurosurgery, Universitatsklinik des Saarlandes, Kirrbergerstrasse, Gebaude 90, D-66421 Homburg-Saar, Germany. Phone: +49 7841 184469, Fax: +49 6841 164480, Email: kiefer@web.de; Professor Peter Vajkoczy, Department of Neurosurgery, University Hospital Mannheim, Theodor-Kutzer-Ufer 1-3, D68167 Mannheim, Germany. Phone: +49 621 3832360, Fax: +49 621 3832004, Email: Peter.vajkoczy@nch.ma.uni-heidelberg.de; Dr Dirk de Jong, Department of Neurosurgery, University Hospital Rotterdam, Box 2040, Dr Molewaterplein 40, 3015GD Rotterdam, The Netherlands. Phone: +31 10 4087825, Fax: +31 10 4089452, Email: dejong@neuro.fgg.eur.nl; Professor Flemming Gjerris, University Clinic of Neurosurgery, Rigshospitalet, Blegdamsvej 9, DK 2100 Copenhagen, Denmark. Phone: +45 35 452390, Fax: +45 35 456575, Email: fgjerris@rh.dk; Professor Nino Stocchetti, Servizio di Anestesia e Rianimazione, Terapia Intensiva Neuroscienze, Via Francesco Sforza 35, 20122 Milano, Italy. Phone: +39 02 550 35517, Fax: +39 02 599 02239, Email: stocchet@polic.cilea.it; Dr Luciana Mascia, Department of Anaesthesia and Intensive Care Medicine, Ospedale Molinette, Corso Bramante 88, 10126 Torino, Italy. Email: Luciana.mascia@unito.it; Professor Dr Reto Stocker, Division of Surgical Intensive Care, Universitasspital Zurich, B-HOF109, Ramistrasse 100, 8091 Zurich, Switzerland. Phone: +41 1 255 2376, Fax: +41 1 255 3172, Email: Reto.stocker@chi.usz.ch; Professor Jan Goffin, Department of Neurosurgery, University Hospital Gasthuisberg, Herestraat 49, 3000 Leuven, Belgium. Phone: +32 16 344290, Fax: +32 16 344285, Email: Jan.goffin@uz.kuleuven.ac.be; Dr Bertil Rydenhag, Institute of Clinical Neuroscience, Department of Neurosurgery, Sahlgrenska University Hospital, Bla Straket 7, van 5, SE 413-45 Gothenburg, Sweden. Phone: +45 31 3421578, Fax: +45 31 416719, Email: Bertil.rydenhag@neuro.gu.se; Professor John Pickard, Academic Neurosurgical Unit, Addenbrookes Hospital, Level 4, A Block, Box 167, Hills Road, Cambridge CB2 2QQ, UK. Phone: +44 01223 336946, Fax: +44 01223 3216926, Email: jdpsecretary@medschl.cam.ac.uk; Dr Sue Hill, Department of Anaesthetics, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK. Phone: +44 02380 796401, Fax: +44 02380 794843, Email: sahneuro@aol.com; Professor Garth Cruickshank, University Department of Neurosurgery, Queen Elizabeth Hospital, Edgbaston, Birmingham B15 2TH, UK. Phone: +44 0121 697 8225, Fax: +44 0121 697 8248, Email: g.s.cruickshank@bham.ac.uk; Dr Stefan Iencean, Department of Neurosurgery, SF Treime Hospital, 2 Ateneului Street, 6600 Iasi, Romania. Email: mirceasteffan@yahoo.com; Mr Lawrence Watkins, Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, Queen Square, London WC1N 3BG, UK. Phone: +44 0207 837 3611, Email: l.watkins@ion.ucl.ac.uk.

Compliance with ethical standards

Conflict of interest We declare that we have no conflicts of interest.

Appendix

See Table 2.

Table 2 Table listing the 33 models and input parameters selected for assessment based upon clinical guidance and literature search

Date	Group	Name	Duration h:min:s	Sensitivity (%)	Specificity (%)	PPV (%)
2008-10-09	30_15	BasicWithETCO2	01:14:11	43.9	85.1	23.1
2008-10-16	30_15	BasicSlopeVar550Nets	15:56:45	43.3	88.3	31.8
2008-10-09	30_15	BasicWithSlope	05:27:44	42.1	85.0	26.3
2008-10-07	25_10	Basic	03:01:08	41.4	91.1	38.1
2008-10-13	30_15	BasicWithSlope16HL	02:43:39	41.3	85.4	26.6
2008-10-16	25_10	BasicSlopeVar	07:32:14	41.2	92.9	43.3
2008-10-13	30_15	BasicSlopeVar	07:05:21	39.6	90.2	33.7
2008-09-19	30_15	Basic	03:05:05	38.4	88.6	28.0
2008-10-13	25_10	BasicSlopeETCO2	03:07:07	37.0	90.5	29.4
2008-10-22	40_15	BasicWithSlopeSingleStart	08:51:31	36.0	91.9	35.0
2008-10-03	30_15	BasicWithVar	04:53:32	34.8	93.4	40.3
2008-10-10	30_15	BasicWithSlope64HL	10:08:03	32.9	89.0	27.7
2008-10-15	30_15	BasicSlopeVarT1	09:34:31	32.6	92.9	35.9
2008-10-16	30_15	BasicSlopeVarSaO2	11:01:24	32.3	92.2	33.3
2008-10-20	30_15	BasicSlopeVarAPA	10:54:36	31.7	91.3	31.4
2008-10-08	30_15	BasicNoDemog	02:47:40	31.4	94.4	41.5
2008-10-13	30_15	BasicSlopeSingleStart	05:38:54	31.1	92.9	36.3
2008-09-19	30_15	Minimal	02:38:18	30.7	93.8	38.4
2008-10-16	30_15	BasicSlopeVar15PC	05:56:57	30.3	92.1	38.1
2008-10-07	35_20	Basic	02:58:20	28.6	93.2	34.0
2008-10-22	40_15	BasicWithSlope	09:14:31	27.3	93.1	31.6
2008-10-24	30_15	BWSlope30win	05:26:53	25.6	95.1	38.2
2008-09-21	30_15	MiniVariance	02:54:09	24.5	93.9	35.2
2008-09-12	60_15	AllPhys	Unknown	22.4	94.2	27.6
2008-10-06	45_30	BasicWithVar	05:01:13	21.9	91.2	22.5
2008-10-22	50_15	BasicWithSlope	13:47:41	18.5	98.3	55.2
2008-10-22	50_15	BasicWithSlopeSingleStart	13:01:40	15.0	96.3	31.6
2008-09-11	60_15	Lots-01	04:19:32	14.5	98.0	39.3
2008-09-18	90_30	PlusDemog	11:35:47	7.0	93.1	6.3
2008-09-15	120_30	Basic	04:52:08	5.8	99.1	38.2
2008-09-17	90_30	First	05:55:07	4.2	99.2	20.4
2008-09-10	60_15	Basic	29:20:52	1.7	99.9	28.6
2008-10-23	30_20	Minimal	02:22:34	0.7	99.8	31.3

There are 33 runs in the table, 6 above 40% sensitivity and 13 above 30% sensitivity. The two groups investigated the most were 30_15 and 25_10. The first digit indicates the maximum start point for a sub-window and the second digit gives the minimum start window and is therefore an indication of the prediction window. For example the 30_15 group is using data which starts at 15 min before an event in the training/test database but also uses a sub-window that starts at 30 min before an event. In order to determine the actual amount of data that is required before a prediction could begin using this model you also have to consider the sub-window size which is defined in the input vector definition file. In the case of this model the sub-window size is 15 min, therefore 45 min of data are required (15 + 30)

References

1. American College of Surgeons. Advanced trauma life support. 2011.
2. Jones P, Andrews P, Midgley S, Anderson S, Piper I, Tocher J, Housley A, Corrie J, Slattery J, Dearden M, Douglas Miller J. Measuring the burden of secondary insults in head-injured patients during intensive care. J Neurosurg Anesthesiol. 1994;6:4–14.
3. Marmarou A, Anderson RL, Ward JD, et al. Impact of ICP instability and hypotension on outcome in patients with severe head trauma. J Neurosurg. 1991;75:S59–66.
4. BrainIT Consortium. BrainIT. 2007.
5. BrainIT. Brain-it core dataset, manual of operations, data element definitions. 2009.
6. Piper I, Chambers I, Citerio G, Enblad P, Gregson B, Howells T, Kiening K, Mattern J, Nilsson P, Ragauskas A, Sahuquillo J, Donald R, Sinnott R, Stell A. The brain monitoring with information

technology (BrainIT) collaborative network: EC feasibility study results and future direction. Acta Neurochir. 2010;152(11):1859–71. https://doi.org/10.1007/s00701-010-0719-1.
7. AvertIT. Avert-it project (FP7-217049-AVERT-IT). 2008.
8. Funahashi K. On the approximate realization of continuous mappings by neural networks. Neural Networks. 1989;2(3):183–192.
9. Bishop CM. Neural networks for pattern recognition. Oxford University Press; 1995.
10. Ripley BR. Pattern recognition and neural networks. Cambridge University Press; 1996.
11. MacKay DJC. Probable networks and plausible predictions—a review of practical bayesian methods for supervised neural networks. Network. 1995;6:469–505.
12. Neal RM. Bayesian learning for neural networks. Springer, Lecture Notes in Statistics. 1996.
13. Stell A, Sinnott R, Jiang J, Donald R, Chambers I, Citerio G, Enblad P, Gregson B, Howells T, Kiening K, Nilsson P, Ragauskas A, Sahuquillo J, Piper I. Federating distributed clinical data for the prediction of adverse hypotensive events. Philos Trans R Soc. 2009;367(1898):2679–90.
14. Hanfelt JJ, Slack RS, Gehan EA. A modification of simon's optimal design for phase ii trials when the criterion is median sample size. Control Clin Trials. 1999;20:555–66.
15. Simon R. Optimal two-stage designs for phase ii clinical trials. Control Clin Trials. 10:1–10, 1989.
16. Crawley M. The R book (Second Edition). New York: Wiley; 2012.
17. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, Suppes R, Feinstein D, Zanotti S, Taiberg L, Gurka D, Kumar A, Cheang M. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med. 2006;34(6):1589–96.
18. Bijker JB, vanKlei WA, Vergouwe Y, .Eleveld DJ, van Wolfswinkel L, Moons KGM, Kalkman CJ. Intraoperative hypotension and 1-year mortality after noncardiac surgery. Anesthesiology. 2009;111(6):1217–26.
19. Smiley R. Fast fourier transforms as prophecy: predicting hypotension during spinal anes-thesia. Anesthesiology. 2005;102(6):1079–80.
20. Moody GB, Lehman LH. Predicting acute hypotensive episodes: the10th annual physionet/computers in cardiology challenge. Comput Cardiol. 2009;36:541–4.
21. Henriques JH, Rocha TR. Prediction of acute hypotensive episodes using neural network multi-models. Comput Cardiol. 2009;36:549–52.
22. Chen X, Xu D, Zhang G, Mukkamala R. Forecasting acute hypotensive episodes in intensive care patients based on a peripheral arterial blood pressure waveform. Comput Cardiol. 2009;36:545–8.
23. Mneimneh MA, Povinelli RJ. A rule-based approach toward the prediction of acute hypotensive episodes. Comput Cardiol. 2009;36:557–60.
24. Gabutti L, Vadilonga D, Mombelli G, Burnier M, Marone C. Artificial neural networks improve the prediction of kt/v, follow-up dietary protein intake and hypotension risk in haemodialysis patients. Nephrol Dial Transplant. 2004;19:1204–11.
25. Reich DL, Hossain S, Krol M, Baez B, Patel P, Bernstein A, Bodian CA. Predictors of hypotension after induction of general anesthesia. Anesth Analg. 2005;101(3):622–8.
26. Lehman L, Saeed M, Moody G, Mark R. Similarity-based searching in multi-parameter time series databases. Comput Cardiol. 2008;35:653–6.
27. Dasta JF, McLaughlin TP, Mody SH, Piech CT. Daily cost of an intensive care unit day: the contribution of mechanical ventilation. Crit Care Med. 2005;33(6):1266–71.
28. Donald R, Howells T, Piper I, Chambers I, Citerio G, Enblad P, Gregson B, Kiening K, Mattern J, Nilsson P, Ragauskas A, Sahuquillo J, Sinnott R, Stell A. Trigger characteristics of EUSIG-defined hypotensive events. Acta Neurochir Suppl. 2012;114:45 – 9.
29. Williams CKI, Georgatzis K, Hawthorne C, McMonagle P, Piper I, Shaw M, Lal P. Detecting artifactual events in vital signs monitoring data. In: Clifton DA, editor. Machine learning for healthcare technologies; 2016. http://homepages.inf.ed.ac.uk/ckiw/projects/adult_icu/CSOreport230915.pdf.
30. Ghassemi M, Celi L,A, Stone J. State of the art review: the data revolution in critical care. Crit Care. 2015;19:118.
31. IBM InfoSphere, Big Data Help Toronto Hospital Monitor Premature Infants. http://www.eweek.com/enterprise-apps/ibm-infosphere-big-data-help-toronto-hospital-monitor-premature-infants.
32. Docherty AB, Lone NI. Exploiting big data for critical care research. Curr Opin Crit Care. 2015;21(5):467–72.
33. Celi LA, Mark RG, Stone DJ, Montgomery RA. "Big Data" in the intensive care unit. Closing the data loop. Am J Respir Crit Care Med. 2013;187(11):1157–60.
34. Figini S, Maggi M. Performance of credit risk prediction models via proper loss functions. Technical report, Universita di Pavia, Department of Economics and Management, 2014.
35. Hand D. Assessing the performance of classification methods. Int Stat Rev. 2012;80(3):400–14. https://doi.org/10.1111/j.1751-5823.2012.00183.x.

Affiliations

Rob Donald[1] · Tim Howells[2] · Ian Piper[3] · P. Enblad[4] · P. Nilsson[4] · I. Chambers[5] · B. Gregson[6] · G. Citerio[7] · K. Kiening[8] · J. Neumann[8] · A. Ragauskas[9] · J. Sahuquillo[10] · R. Sinnott[11] · A. Stell[12] on behalf of the BrainIT Group

Rob Donald
rob@statsresearch.co.uk

1 Stats Research Ltd, Dingwall, Scotland, UK

2 Department of Neurosurgery, Uppsala University Hospital, Uppsala, Sweden

3 Institute of Neurological Sciences, Queen Elizabeth University Hospital, Clinical Physics, Glasgow, Scotland, UK

4 Section of Neurosurgery, Uppsala University, Uppsala, Sweden

5 Department of Medical Physics, James Cook University Hospital, Middlesbrough, UK

6 Neurosurgical Trials Group, Newcastle University, Newcastle upon Tyne, UK

7 Hospital San Gerardo, Neurorianimazione, Monza, Italy

8 Department of Neurosurgery, Ruprecht-Karls-Universitat Hospital, Heidelberg, Germany

9 Kaunas University of Technology, Kaunas, Lithuania

[10] Department of Neurosurgery, Vall d'Hebron University Hospital, Barcelona, Spain

[11] Department of Information Systems, University of Melbourne, Parkville, Australia

[12] Department of Clinical Physics, University of Glasgow, Glasgow, Scotland, UK