Proceedings of the 29th Annual International
Conference of the IEEE EMBS
Cité Internationale, Lyon, France
August 23-26, 2007.

Diagnosis of Hemorrhage in a Prehospital Trauma Population Using Linear and Nonlinear Multiparameter Analysis of Vital Signs

Liangyou Chen, Andrew T. Reisner, Thomas M. McKenna, Andrei Gribok, and Jaques Reifman

***Abstract*—In this study, we analyzed a dataset of time-series vital-signs data collected by standard Propaq travel monitor during helicopter transport of 898 civilian trauma casualties from the scene of injury to a receiving trauma center. The goals of the analysis are two fold. First, to determine which combination of the automatically-collected and -qualified vital signs provides the best discrimination between casualties with and without major hemorrhage. Second, to determine whether nonlinear classifiers provide improved discrimination over simpler, linear classifiers. Major hemorrhage is defined by the presence of injuries consistent with hemorrhage in casualties who received one or more units of blood. We randomly selected a subset of the casualties to train and test the classifiers with multiple combinations of the vital-signs variables, and used the area under the receiver operating characteristic curve (ROC AUC) as a decision metric. Based on the results of 100 simulations, we observe that: (i) the best two features obtained are systolic blood pressure and heart rate (mean AUC = 0.75 from a linear classifier), and (ii) the use of nonlinear classifiers does not improve discrimination. These results support earlier findings that the interaction of systolic blood pressure and heart rate is useful for the identification of trauma hemorrhage and that linear classifiers are adequate for many real-world applications.**

I. Introduction

TRAUMA is the leading cause of death for Americans ages 1 through 44 years [1], terminating lives long before their expected lifespan. Most of these deaths are due to hemorrhage or major head trauma [2, 3]. Unlike the latter, the former is often treatable, so early and accurate diagnosis of significant hemorrhage is of great importance. Ideally, hemorrhage in trauma casualties would be diagnosed at the scene of injury. This information would be very useful for triage (i.e., prioritization of casualties based on injury severity and determination of whether to send the casualty to a specialized trauma center or a local medical facility), resource mobilization (e.g., activation of trauma teams at a receiving trauma center), and therapeutic decision-making.

With the ultimate goal of developing triage and diagnostic decision-aid methods, we reviewed a dataset of time-series vital-signs data measured by a standard Propaq travel monitor during helicopter transport of civilian trauma casualties from the scene of injury to a receiving trauma center. Consistent with this long-term goal, in this study we determine: (a) the most discriminatory combination of vital signs (i.e., features) for classification of casualties in terms of hemorrhage versus no hemorrhage, and (b) if nonlinear methods improve classifier performance.

Manuscript received March 29, 2007. This work was supported by the Combat Casualty Care and the Military Operational Medicine research programs of the U.S. Army Medical Research and Materiel Command (USAMRMC), Fort Detrick, Maryland. The investigators adhered to the policies for protection of human subjects as prescribed in Army Regulation 70-25, and the research was conducted in adherence with the provisions of 45 CFR Part 46.

L. Chen is with the Bioinformatics Cell, Telemedicine and Advanced Technology Research Center (TATRC), USAMRMC, Frederick, MD 21702 (e-mail: lchen@bioanalysis.org).

A. T. Reisner is with the Massachusetts General Hospital Department of Emergency Medicine, Boston, MA 02114 and with the Bioinformatics Cell, TATRC, USAMRMC, Frederick, MD 21702 (e-mail: areisner@partners.org).

T. M. McKenna is the Deputy Director of the Bioinformatics Cell, TATRC, USAMRMC, Fort Detrick, MD 21702 (e-mail: tmckenna@bioanalysis.org).

A. Gribok is with the Bioinformatics Cell, TATRC, USAMRMC, Frederick, MD 21702, and with the Nuclear Engineering Department of the University of Tennessee, Knoxville (e-mail: agribok@bioanalysis.org).

J. Reifman is a Senior Research Scientist and Director of the Bioinformatics Cell, TATRC, USAMRMC, ATTN: MCMR-ZB-T, Bd. 363 Miller Drive, Fort Detrick, MD 21702 (corresponding author; phone: 301-619-7915; fax: 301-619-1983; e-mail: jaques.reifman@us.army.mil).

II. Methods

A. Dataset

This study is based on discrete attribute data and physiologic time-series data collected from 898 trauma casualties during and after transport by helicopter service from the scene of injury to the Level-I unit at the Memorial Hermann Hospital in Houston, Texas [4]. Approximately 10% of the casualties suffered penetrating trauma and 90% blunt trauma. Mortality was 10% overall. The time-series variables were measured by Propaq 206EL vital-signs monitors and downloaded to an attached personal digital assistant. The variables consist of electrocardiogram, photoplethysmogram, and respiratory waveform signals recorded at various frequencies, and their corresponding monitor-calculated variables, such as heart rate (HR), respiratory rate (RR), and oxygen saturation of arterial hemoglobin (SaO2), recorded at 1-second intervals. In addition, systolic (SBP), mean, and diastolic (DBP) blood pressures were collected intermittently at multi-minute intervals. The casualties' attribute data include discrete information, such as demographic data, injury description, and treatments. There are over 100 variables of this type for each patient.

B. Outcomes

In this study, the outcome of interest is major hemorrhage, which we defined as the requirement for transfusion of one or more units of red cells *and* an explicit hemorrhagic injury (laceration of solid organs *or* hematoma in the abdomen *or* hemothorax *or* explicit vascular injury and operative repair *or* limb amputation), identified by a text search of the casualties' injury description fields.

C. Inclusion/exclusion criteria

Records with missing variables were excluded. We analyzed the 492 case records with each of the five vital signs (SBP, DBP, HR, RR, SaO2) present during the 5-to-7-minute interval of their transport to the trauma center, of which 55 were identified, based on the above criteria, as having major hemorrhage. The rationale for the 5-to-7 minute interval is addressed in Section III, Results and Discussion. We excluded patients who received one or more units of red cells *without* an explicit hemorrhagic injury because of ambiguity in whether or not they truly required blood replacement therapy.

D. Processing of vital signs data

All of the calculations were based on 2-minute time windows. We used data quality algorithms [5, 6] to identify 5 seconds of the most reliable vital-signs data within that time window and used the average of those 5 seconds.

E. Linear and nonlinear classifiers

Linear and nonlinear classifiers were used for feature evaluation and patient outcome classification. The linear classifier employed a linear discriminant function $f = \boldsymbol{w}^{T}\boldsymbol{x} + w_0$, where the vector of coefficients $\boldsymbol{w}$ and the coefficient w_0 were learned from a training dataset, to evaluate a given input feature vector $\boldsymbol{x}$ against two classes (hemorrhage versus non-hemorrhage), and assigned $\boldsymbol{x}$ to one of the two classes based on the decision $f(\boldsymbol{x}) > \theta$, where θ is a chosen decision threshold.

Two nonlinear classifiers, leading to nonlinear decision boundaries, were evaluated: a feedforward artificial neural network (ANN) and a support vector machine (SVM). A three-layer ANN, with two hidden nodes and one output node, was trained with a conjugate gradient algorithm [7, 8]. We tried different number of hidden nodes, spanning the range from 2 to 20, and found no significant difference in performance. A nonlinear sigmoid activation function was used to map all network nodes. The LibSVM library [9] was used to train the SVM classifiers, with a radial basis kernel used as the nonlinear mapping function. An independent cross-validation dataset was randomly selected from 40% of each training dataset to obtain a stopping criterion for the ANN as well as to obtain the optimum γ and C parameters for the SVM kernel function. Both training and testing datasets were normalized to have zero mean and unit variance.

F. Classifier evaluation

For each of the 31 possible combinations of the five basic vital signs, using one, two, three, four and five variables, we performed simulations using the linear, ANN and SVM classifiers. Each simulation was performed 100 times (trials) and the results averaged. In each trial, we randomly selected 54 casualties from the original dataset for training and 54 for testing, and balanced the two outcome classes with 27 casualties per hemorrhage/no hemorrhage class. In doing so, we minimize the chance of generating a bias classifier at the expense of smaller training/testing data sets.

A receiver operating characteristic (ROC) curve is used to describe the sensitivity versus specificity of the classifiers as a function of decision threshold θ. The area under the ROC curve (ROC AUC), calculated through trapezoidal integration of 50 evenly-spaced decision thresholds spanning the entire output range, is used as a metric for comparing the classification power of different feature combinations and linear versus nonlinear classifiers. No formal statistical testing between ROC curves was undertaken because there were a very large number of comparisons.

III. Results and Discussion

A. Basic vital signs

For the linear classifiers, any combination containing SBP leads to an AUC $\geq$ 0.71, and any combination lacking SBP leads to an AUC $\leq$ 0.70. SBP is a classic vital sign whose utility in the early detection of traumatic hemorrhage has been questioned [10, 11]. Similarly, it has been observed that an elevated heart rate is not a reliable indicator of hypovolemic shock in trauma patients [12, 13]. We found that HR alone is a poor indicator of hemorrhage (AUC = 0.66), but HR does add additional information when used in conjunction with SBP. In this dataset, SBP is the best single feature, while the combination of SBP and HR yields the highest AUC for two features with AUC = 0.75 (Table I). Adding another feature, SaO2, only slightly improves the AUC to 0.76. Considering the benefit of using fewer variables, we selected SBP and HR as the best feature combination.

One notable distinction of the current methodology is that, unlike other trauma registries (e.g., [10]) our vital-signs data were automatically archived and subsequently automatically analyzed, without on-line human oversight. Vital-signs data are famous for being noisy and frequently unreliable, so we relied on automated data quality algorithms to select the best data and processed merely 5-second excerpts of such data. Our results are diagnostically similar to prior analyses of vital signs and hemorrhage, with the ramification that we may be able to develop fully automatic analysis of vital-signs data for real-time detection of significant hemorrhage (and presumably for other pathophysiological states too).

One important limitation to real-world implementation of automated diagnosis is the issue of data availability.

TABLE I
CLASSIFIER PERFORMANCE FOR DIFFERENT COMBINATIONS OF VITAL SIGN VARIABLES

Variables	Area under the ROC Curve		
	Linear	ANN	SVM
SBP	**0.71±0.05**	0.70±0.06	0.68±0.07
HR	0.66±0.07	0.65±0.07	0.64±0.07
DBP	0.58±0.10	0.56±0.10	0.58±0.08
RR	0.57±0.10	0.56±0.10	0.57±0.08
SaO2	0.56±0.08	0.54±0.08	0.55±0.07
HR,SBP	**0.75±0.06**	0.74±0.06	0.72±0.07
SBP,DBP	0.73±0.06	0.71±0.08	0.69±0.07
SaO2,SBP	0.73±0.05	0.71±0.08	0.69±0.08
RR,SBP	0.71±0.06	0.70±0.07	0.69±0.08
HR,DBP	0.68±0.07	0.67±0.07	0.66±0.08
HR,SaO2	0.67±0.06	0.66±0.07	0.65±0.08
HR,RR	0.66±0.06	0.64±0.09	0.63±0.08
RR,DBP	0.62±0.09	0.60±0.08	0.59±0.08
SaO2,DBP	0.60±0.10	0.59±0.09	0.60±0.08
RR,SaO2	0.59±0.09	0.59±0.10	0.60±0.09
HR,SBP,SaO2	**0.76±0.05**	0.75±0.07	0.73±0.07
HR,SBP,DBP	0.75±0.06	0.74±0.06	0.73±0.07
HR,RR,SBP	0.74±0.06	0.73±0.06	0.72±0.07
SaO2,SBP,DBP	0.74±0.06	0.71±0.07	0.70±0.08
RR,SBP,DBP	0.73±0.06	0.70±0.08	0.69±0.07
RR,SaO2,SBP	0.72±0.06	0.71±0.07	0.70±0.07
HR,SaO2,DBP	0.70±0.07	0.68±0.08	0.68±0.08
HR,RR,DBP	0.67±0.07	0.66±0.07	0.66±0.06
HR,RR,SaO2	0.67±0.07	0.65±0.09	0.64±0.08
RR,SaO2,DBP	0.62±0.09	0.61±0.09	0.60±0.08
HR,SaO2,SBP,DBP	0.75±0.06	0.73±0.07	0.73±0.07
HR,RR,SaO2,SBP	0.74±0.06	0.73±0.07	0.72±0.08
HR,RR,SBP,DBP	0.74±0.06	0.73±0.06	0.72±0.07
RR,SaO2,SBP,DBP	0.73±0.06	0.69±0.07	0.69±0.08
HR,RR,SaO2,DBP	0.69±0.07	0.66±0.09	0.67±0.08
HR,RR,SaO2,SBP,DBP	0.74±0.06	0.73±0.07	0.73±0.07

Mean ± standard deviation of the ROC AUC for various combinations of vital signs, classified using (a) linear, (b) artificial neural network (ANN), and (c) support vector machine (SVM) classifiers is reported. Each result is from 100 trials of randomly selected testing datasets, each consisting of 54 casualties with outcome classes equally balanced between casualties with and without hemorrhage. The best combinations of one, two and three variables are shown in bold.

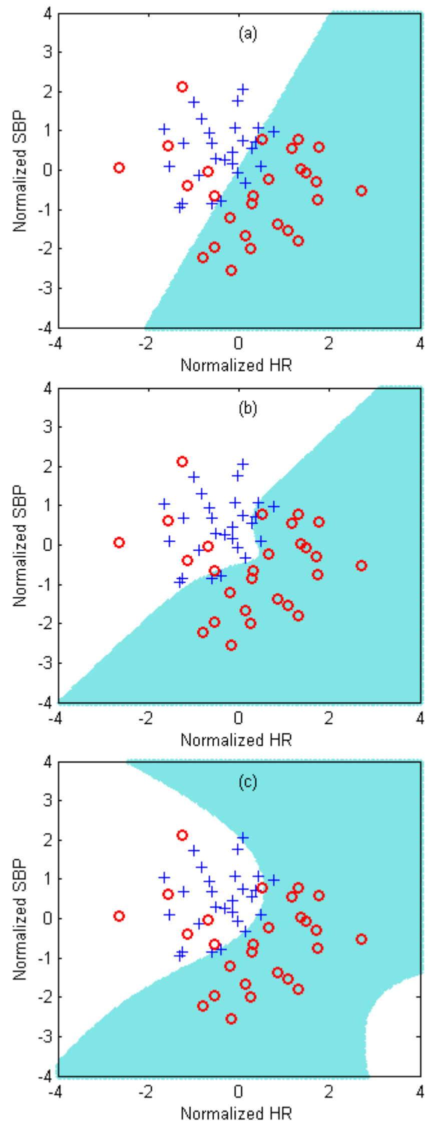

Fig. 1. Typical decision boundaries for discriminating hemorrhage (red circle, shaded areas) and non-hemorrhage (blue crosses, non-shaded areas) casualties of a testing dataset. These boundaries were obtained by (a) Linear, (b) ANN, and (c) SVM classifiers.

Initially, we tried to analyze data within the first two minutes of each record (e.g., the 0-to-2 minute window); however, the vital-signs data streams were coming on-line in that time period (perhaps as the caregivers attached new sensors one-by-one). It was not until after 5 minutes that an acceptable number of subjects had complete vital signs availability (HR, SBP, DBP, SaO2 and RR): i.e., 492 out of 898 subjects had complete data availability within the 5-to-7 minute window. The prevalence of incomplete pre-hospital vital-signs data has been previously described [14], and poses a limitation to algorithms requiring a multitude of input parameters.

B. *Nonlinear classifiers*

For each of the 31 combinations of variables, the linear classifiers performed slightly better than the ANNs, which in turn performed marginally better than the SVMs (Table I). Figure 1 illustrates the discrimination of casualties with hemorrhage (red circles) versus non-hemorrhage (blue crosses) using normalized values of HR and SBP from a typical testing trial dataset. Figure 1(a) shows the results of a linear classifier and associated linear decision boundary, while Figs. 1(b) and 1(c) show the nonlinear boundaries obtained with an ANN and a SVM classifier, respectively. The nonlinear boundaries do not discriminate the two classes better than the linear boundary. The slightly degraded overall performance of the ANN and SVM classifiers might be explained by the limited amount of training data, which could have prevented classifier optimization through cross-validation. Interestingly, in this problem, the ANNs usually produced close-to-linear boundaries. The SVMs showed the best performance (matching the linear one) if we abandoned cross-validation and fixed the parameter γ of the radial basis kernel at a very small value ($\gamma < 1e\text{-}6$).

C. Derived physiologic parameters

There has been historic interest in physiologic variables that are derived from basic vital signs. These variables are intended to be more indicative of the underlying physiologic state than the raw vital signs. An example is the shock index (SI), which is defined by the ratio of a patient's SBP relative to the HR. The SI is intended to overcome the diagnostic limitations of SBP and HR alone. Another example is the pulse pressure (PP), which is defined as the difference between systolic and diastolic blood pressures. PP has attracted interest as a potentially superior predictor of casualty severity [15].

We computed ROC AUCs for linear classification using combinations of two to four features of both basic vital-signs variables and the variables derived from them, which included PP, SI, hemorrhage index (HR*RR/PP), and RR/HR. The results were obtained using a procedure similar to the one discussed above, where each simulation consisted of 100 randomly selected trials with balanced outcome classes and an equal number of data points used for training and testing.

The SI was the single best derived feature, with an AUC = 0.77. Pulse pressure alone offered an AUC = 0.73. The other combinations involving two or more derived parameters yielded AUCs of 0.75-0.77. The results suggest that derived features do not provide added classification power compared to inputting the same basic variables separately. For instance, the AUC for the SI is 0.77 as compared to 0.75 when the linear classifier inputs are SBP and HR. The results also suggest that PP does not contain additional information than the combination of SBP and DBP.

IV. Conclusions and Future Work

In this retrospective exercise, we discovered that the linear combination of SBP and HR leads to good performance (AUC = 0.75) in the detection of major hemorrhage in a civilian prehospital trauma population, comparable to the well-known shock index, and superior to other vital signs (DBP, RR, and SaO2). The use of nonlinear classifiers does not improve performance. The use of derived physiologic metrics scarcely improves performance.

Of note, our results are based on fully automatic data processing and a widely-used monitoring device, the Propaq monitor. Data were filtered by automated algorithms to promote analysis based on accurate vital signs. It may be practical to incorporate such algorithms in transport monitors for real-time use. One challenge is that of data availability, as only 492 of the 898 subjects had all five basic vital-signs variables available in the two-minute window of interest, suggesting that a solution involving an “ensemble” of classifiers, with each classifier in the ensemble requiring only a subset of the vital signs, may be more robust than one involving a single classifier requiring simultaneous availability of all vital signs. In future work, accurate detection of hemorrhage may be enhanced if additional physiologic information is used, such as trend data, information from photoplethysmograms, and heart rate variability information.

Acknowledgement

The authors wish to express their gratitude to COL John Holcomb and Dr. Jose Salinas of the U.S. Army Institute of Surgical Research, Fort Sam Houston, San Antonio, Texas, who provided the trauma patient data.

Disclaimer

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or the U.S. Department of Defense.

References

[1] D. L. Hoyert, T. J. Mathews, F. Menacker, D. M. Strobino, and B. Guyer, "Annual summary of vital statistics: 2004," *Pediatrics*, vol. 117, pp. 168-83, 2006.

[2] A. Sauaia, F. A. Moore, E. E. Moore, K. S. Moser, R. Brennan, R. A. Read, and P. T. Pons, "Epidemiology of trauma deaths: a reassessment," *J Trauma*, vol. 38, pp. 185-93, 1995.

[3] R. Peng, C. Chang, D. Gilmore, and F. Bongard, "Epidemiology of immediate and early trauma deaths at an urban Level I trauma center," *Am Surg*, vol. 64, pp. 950-4, 1998.

[4] J. B. Holcomb, S. E. Niles, C. C. Miller, D. Hinds, J. H. Duke, and F. A. Moore, "Prehospital physiologic data and lifesaving interventions in trauma patients," *Mil Med,* vol. 170, pp. 7-13, 2005.

[5] C. Yu, Z. Liu, T. McKenna, A. Reisner, and J. Reifman, "A Method for Automatic Identification of Reliable Heart Rates Calculated from ECG and PPG Waveforms," *JAMIA,* vol. 13, pp. 309-320, 2006.

[6] L. Chen, T. M. McKenna, A. T. Reisner, and J. Reifman, "Algorithms to qualify respiratory data collected during transport of trauma patients," *Physiological Measurement*, vol. 27, pp. 797-816, 2006.

[7] S. Haykin, *Neural Networks: A Comprehensive Foundation,* 2nd ed. New Jersey: Prentice-Hall, Inc., 1999.

[8] M. Hanke. *Conjugate Gradient Type Methods for Ill-Posed Problems*, Chapman & Hall/CRC, 1995.

[9] C. C. Chang and C. J. Lin, "LibsVM: a library for support vector machines," 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[10] J. K. Parks, A. C. Elliott, L. M. Gentilello, and S. Shafi, "Systemic hypotension is a late marker of shock after trauma: a validation study of Advanced Trauma Life Support principles in a large national sample," *Am J Surg*, vol. 192, pp. 727-31, 2006.

[11] J. W. Davis, I. C. Davis, L. D. Bennink, J. F. Bilello, K. L. Kaups, and S. N. Parks, "Are automated blood pressure measurements accurate in trauma patients?," *J Trauma*, vol. 55, pp. 860-3, 2003.

[12] H. S. Snyder and S. J. Dresnick, "Lack of tachycardic response to hypotension in penetrating abdominal injuries," *J Emerg Med*, vol. 7, pp. 335-9, 1989.

[13] G. P. Victorino, F. D. Battistella, and D. H. Wisner, "Does tachycardia correlate with hypotension after trauma?" *J Am Coll Surg*, vol. 196, pp. 679-84, 2003.

[14] M. Raux, M. Thicoipe, E. Wiel, E. Rancurel, D. Savary, J. S. David, F. Berthier, A. Ricard-Hibon, F. Birgel, and B. Riou, "Comparison of respiratory rate and peripheral oxygen saturation to assess severity in trauma patients," *Intensive Care Med*, vol. 32, pp. 405-12, 2006.

[15] W. H. Cooke, J. Salinas, V. A. Convertino, D. A. Ludwig, D. Hinds, J. H. Duke, F. A. Moore, and J. B. Holcomb, "Heart rate variability and its association with mortality in prehospital trauma patients," *J Trauma,* vol. 60, pp. 363-70; discussion 370, 2006.