Robust Prediction of Patient Mortality from 48 Hour Intensive Care Unit Data

Luigi Yuri Di Marco[1], Marjan Bojarnejad[1], Susan T King[2], Wenfeng Duan[1], Costanzo Di Maria[2], Dingchang Zheng[1], Alan Murray[1], Philip Langley[1]

[1]Newcastle University, Newcastle upon Tyne, UK
[2]Freeman Hospital, Newcastle upon Tyne, UK

Abstract

The aim of this study was to develop a new algorithm to predict individual patient mortality with improved accuracy with respect to established methods from data collected over the first 48 hours of admission to the Intensive Care Unit.

A binary classifier was developed to participate in Event 1 of the PhysioNet/Computing in Cardiology Challenge 2012. The algorithm development was undertaken using only posterior knowledge from the training dataset (Set-A), containing 41 demographic and clinical variables from 4000 ICU patients.

For each variable a feature was defined as the average (across all available measurements of the given variable) likelihood of being part of the "survivors" group.

To select features with highest discrimination ability ("survivors" vs. "non-survivors"), a forward sequential selection criterion with logistic cost function was adopted and repeated for cross-validation on N (=10) "leave M-out" (M=50%) random partitions of Set-A. Features that were selected in more than one partition were considered (#Feat = 32). A logistic regression model was used for classification. The score was defined as the lowest between sensitivity and positive predictive value in classification.

The proposed method scored 54.9% on Set-A and 44.0% on the test set (Set-B), outperforming the established method SAPS-I (29.6% on Set-A, 31.7% on Set-B).

1. Introduction

The need to compare the efficacy of intervention and treatment in Intensive Care Unit (ICU) populations has led over the past three decades to the development of methods for prediction of mortality rates [1]. Controlling for differences in severity of illness as well as demographic factors (such as age, gender, ethnicity) and clinical variables represents a key factor in the prediction process [1,2]. This topic was the objective of the Physionet Challenge 2012 [1].

One of the first exhaustive studies on this topic dates back to 1981 when the Acute Physiology and Chronic Health Evaluation (APACHE) was proposed as a tool for ICU outcome prediction [3]. This method utilised 34 physiological measurements recorded during the first 32 hours from admission to ICU. A value ranging from 0 to 4 was assigned to each variable according to its degree of abnormality, and the final APACHE score was the sum of weights assigned to each recorded measurement. Also a pre-admission health status was considered for the final outcome prediction. Further assessments highlighted the unavailability in many practical cases of one or more of those 34 physiological variables and that the assumption of normal values for these missing data led to biased results. Subsequent studies focused on selecting a suitable subset of the original variables which could be routinely available [2,4-11]. In particular, the Simplified Acute Physiology Score (SAPS) I proposed a set of 13 physiological variables including age as an important predictor [2]. The study which led to the APACHE II system pointed out the importance of combining the severity classification with precise clinical diagnosis [5]. It showed that patients with the same severity score had different likelihood of ICU survival depending on the type of ICU in which they were hospitalised and the disease diagnosed. A worldwide study proved that the accuracy of the same method varied between geographic areas, probably due to the presence of different conditions and availability of different therapies in each area [9,10]. In spite of the various methodologies in the literature, the accuracy of these predictive models is subject to continuous revision because of the onset of new conditions and development of new therapies. For example, studies using APACHE III [6], SAPS II [8], and mortality probability models [7] within independent ICUs reported a predicted mortality that was significantly

different from the observed [12,13]. Hence these models need to be periodically re-evaluated or replaced by new, more adequate approaches.

The aim of this study was to develop a new algorithm to predict individual patient mortality with improved accuracy with respect to an established method (SAPS-I [2]) from data collected over the first 48 hours of admission to the ICU (Physionet Challenge, Event I).

2. Methods

2.1. Datasets

A training set (Set-A) containing 4000 recordings from the first 48 hours of patients' admission to the ICU was made available including the outcome for each recording ("survivor" vs. "non-survivor"). Two test sets (Set-B, Set-C) of 4000 recordings each were assigned for scoring. The data of Set-B were made available (without classification outcome), whereas Set-C was not disclosed (only used to determine the final score for the Challenge). The data provided consisted of a collection of demographic and clinical variables. Each clinical variable was presented as a series of measurements, each one marked with the time of acquisition. Variables could be recorded any number of times (missing measurements were also possible).

Each record contained up to six demographic variables, referred to as "general descriptors" (Record ID, age, gender, height, ICU admission type, weight) and up to 36 clinical variables, measured an arbitrary number of times during the first 48 hours of ICU admission.

The score was defined by the Organisers as the lowest between sensitivity and positive predictive value in classification ("survivors" vs. "non-survivors") of the given dataset.

2.2. Feature definition

The "Record ID" information was discarded from this study as it was not relevant to the topic. For each of the remaining variables, all measurements from all recordings of Set-A were collected to generate two separate distributions, for the "survivors" (S) and the "non-survivors" (NS) groups.

For each distribution $D_j(S)$, $D_j(NS)$ of the *j*th variable (j=1,…,41), the 5^{th} – 95^{th} percentile range was calculated and divided into 10 equally sized bins: $\Delta D_j(S)$, $\Delta D_j(NS)$.

For each measurement v_{jk} of variable v_j, the relative *posterior* probability P_j^k of such measurement being from a "survivor" (i.e. being from the distribution $D_j(S)$) was defined as:

$$P_j^k = \frac{p_S}{p_S + p_{NS}}$$
$$p_S = \frac{\#v \in \{v_{jk} \pm \Delta D_j(S)\}}{\#v \in D_j(S)} \quad (1)$$
$$p_{NS} = \frac{\#v \in \{v_{jk} \pm \Delta D_j(NS)\}}{\#v \in D_j(NS)}$$

where "#" denotes "number of". By definition $0 \le P_j^k \le 1$. For each variable, a feature F_j was defined:

$$F_j = \underset{k}{mean}\{P_j^k\} \quad (2)$$

If no measurement was available for the *j*th variable (k=0), F_j was assigned to the default value of 0.5 (equal likelihood of sample being drawn from "survivor" or "non-survivor" distribution).

2.3. Feature selection

Set-A was randomly partitioned N (=10) times following a "leave-M out" (M=50%) approach for cross-validation (CV), preserving the "survivors"-to-"non-survivors" ratio. For each partition, a forward sequential selection (FSS) with logistic cost function was adopted, with the goal of determining the features with highest discrimination ability. Features that were selected in two or more partitions, constituted the final set.

2.4. Classification

Binary classification of recordings ("survivors" vs. "non-survivors") was done using a supervised-learning approach. Several classifier models were implemented and compared: support vector machine with radial basis function kernel (SVM), linear discriminant (LDA), naïve-Bayesian (NB), and logistic (LRM). The last was eventually chosen as the one showing the best performance.

The chosen classifier was based on a logistic model having the selected features as regressors. The classifier was trained (model parameters were estimated) on Set-A. The output of LRM was compared against an empirical threshold THR_C to determine the classification outcome. THR_C was tuned in order to maximize the score on Set-A after the model's (regression) parameters had been estimated.

3. Results

3.1. Feature selection and cross-validation

FSS and CV yielded a total of 32 features. Table 1 shows the selected features and the number of occurrences (N_O) across the CV partitions ($2 \le N_O \le 10$).

Table 1. Occurrence of selected features in CV partitions.

N_O	Features
10	Age, Glasgow Coma Scale, Temperature, Blood Urea Nitrogen, Glucose
9	Weight, Sodium, White Cell Count, Bilirubin, Cholesterol
8	Height, ICU Admission Type, Troponin I, Troponin T
7	$PaCO_2$, Resp. Rate, Heart Rate, Hematocrit, Albumin, ALP Transaminase
6	SaO_2, NI Systolic BP, Magnesium, Platelets, Lactate
5	PaO_2
4	Systolic BP
3	pH, Mean Art. BP, NI Diastolic BP, ALT Transaminase, AST Transaminase
2	(none)

N_O: Number of occurrences across CV partitions

Figure 1 illustrates the performance of various classifiers on Set-A partitions used for CV. The logistic classifier (LRM) exhibited the highest performance on the test sets (XS) of the partitions.

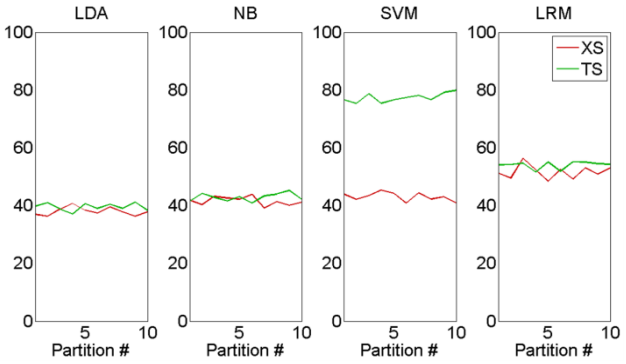

Figure 1. Performance of different classifiers on training (TS, green trace) and test (XS, red trace) set, for each partition of CV.

3.2. Mortality prediction

The ability of the proposed algorithm (FSS-LRM) to predict mortality of ICU patients based on information collected in the first 48 hours was evaluated using an established method (SAPS-I) for comparison.

Table 2 shows the performance (Event I score) on three datasets (Set-A and Set-B, revised 8 May 2012) of the proposed method, compared to the reference (SAPS-I).

Table 2. Performance of algorithms (Challenge Event I).

		Score [%]		
Algorithm	#Features	Set-A	Set-B	Set-C
SAPS-I[(a)]	15	29.6	31.7	31.3
FSS-LRM	32	54.9	44.0	44.6

[(a)]implementation by Physionet Challenge 2012 Organizers [1].

Table 3 compares the performance of the proposed method with SAPS-I, using only features (F_j) from the SAPS-I variable set.

Table 3. Performance of proposed algorithm (using only SAPS features) on Set-A.

Algorithm	Observation Interval [hours]	Score [%] Set-A
SAPS-I	0-24	29.6
FSS-LRM	0-24	13.9
	0-48	49.3

4. Discussion

In this study a new algorithm was presented to predict mortality of individual ICU patients based on information collected from the first 48 hours of admission. A set of 32 features was extracted from the original set of 41 variables ("Record ID" was excluded), for binary classification ("survivor" vs. "non-survivor").

Features were defined as the posterior probability ratio of being a member of the "survivors" group versus the "non-survivors", based on posterior knowledge from the training set Set-A.

The (frequent) cases of missing measurements for any of the 41 variables were handled following a simple intuitive approach of assigning the feature value (F_j) to "equal likelihood" of the sample being drawn from the "survivors" or the "non-survivors" group.

Of the 15 variables (features) used in SAPS-I, 11 were also selected by FSS (age, systolic blood pressure, respiratory rate, Glasgow Coma Scale, heart rate, temperature, blood urea nitrogen, glucose, hematocrit, sodium, white cell count).

The proposed method (FSS-LRM) exhibited higher performance compared to the reference SAPS-I (Table 2), both in the training and test sets. Both methods are based on posterior probability. However, FSS-LRM estimates posterior probability based only on data distributions from the available measurements. As shown in Table 3, increasing the observation window from the first 24 to 48 hours substantially increases the classification performance, indicating the presence of relevant

information beyond the time frame considered in SAPS-I.

On the other hand, the dramatically lower performance of FSS-LRM with respect to SAPS-I for classification based on the first 24 hours also suggests that the data distributions $D_j(S)$, $D_j(NS)$ require an adequately large sample size to yield a reliable estimate of the posterior probability ratio P_j^k as the size of the bins $\Delta D_j(S)$, $\Delta D_j(NS)$ depends on the spread of the distributions, which generally increases with increasing number of observations.

Comparing Table 2 with Table 3, it can be seen that additional features (selected by FSS) further contributed to increasing the classification performance (49.3% vs. 54.9%).

In conclusion, the proposed method based on the established approach of binary classification inspired by posterior probability, corroborates the hypothesis that increasing the observation period from (the first) 24 to 48 hours, and adding relevant clinical variables, may improve accuracy of ICU patient mortality prediction.

References

[1] Physionet Challenge 2012. www.physionet.org/challenge/2012

[2] Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D et al. A simplified acute physiology score for ICU patients. Crit Care Med 1984; 12(11):975-77.

[3] Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE – acute physiology and chronic health evaluation: a physiologically based classification system. Crit Care Med 1981; 9(8):591-7.

[4] Lemeshow S, Teres D, Pastides H, Avrunin JS, Steingrub JS. A method for predicting survival and mortality of ICU patients using objectively derived weights. Crit Care Med 1985; 13(7):519-25.

[5] Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985; 13(10):818-29.

[6] Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest 1991; 100(6):1619-36.

[7] Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care units patients. JAMA 1993; 270(20):2478-86.

[8] Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. JAMA 1993; 270(24):2957-63.

[9] Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3 – From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. Intensive Care Med 2005; 31:1336-44.[10] Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, et al. From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intens Care Med 2005; 31:1345-55.

[11] Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med 2006; 34(5):1297-310.

[12] Zimmerman JE. Comparing ICU populations: Background and current methods. *In* Evaluating Critical Care. Sibbald WJ, Bion JF (Eds). New York, Springer Verlag, 2001, pp 121-39.

[13] Glance LG, Osler TM, Dick AW. Identifying quality outliers in a large, multiple-institution database by using customized versions of the Simplified Acute Physiology Score II and the Mortality Probability Model II0. Crit Care med 2002; 30:1995-2002.

Address for correspondence.

Luigi Yuri Di Marco
Cardiovascular Physics and Engineering Research Group
Institute of Cellular Medicine
Newcastle University, NE1 7RU, Newcastle upon Tyne, UK
luigi.di-marco@ncl.ac.uk