Improving mortality models in the ICU with high-frequency data

James Todd[a], Adrian Gepp[a,*], Brent Richards[b], Bruce James Vanstone[a]

[a] Bond Business School, Bond University, Gold Coast, Queensland, Australia
[b] Gold Coast University Hospital, Gold Coast, Queensland, Australia

ARTICLE INFO

ABSTRACT

Background: Assessment of the performance of Intensive Care Units (ICU) is of vital importance for an effective healthcare system. Such assessment ensures that the limited resources of the healthcare system are allocated where they are most needed. Severity scoring systems are employed for this purpose and improving these systems is a continuing area of research which has focused on the use of more complex techniques and new variables.

Objectives: This paper investigates whether scoring systems could be improved through use of metrics which better summarise the high frequency data collected by automated systems for patients in the ICU.

Methods and Data: 3128 admissions to the Gold Coast University Hospital ICU are used to construct three logistic regressions based on the most widely used scoring system (APACHE III) to compare performance with and without predictors leveraging available high frequency information. Performance is assessed based on model accuracy, calibration, and discrimination. High frequency information was considered for existing pulse and mean arterial pressure physiology fields and resulting models compared against a baseline logistic regression using only APACHE III physiology variables.

Results: Model discrimination and accuracy were better for models which included high frequency predictors, with calibration remaining good in all cases. The most influential high frequency summaries were the number of turning points in a patient's mean arterial pressure or pulse in the first 24 h of ICU admission.

Conclusions: The findings indicate that scoring systems can be improved by better accounting for high frequency data.

1. Introduction

Assessment of the performance of intensive care units (ICUs) is vital to ensure underperforming units are identified and the limited resources of the healthcare system are allocated where they are most needed [1]. To measure the performance of an ICU on an equitable basis, the mix of patients presenting must be adjusted to avoid penalising those ICUs servicing higher risk patients. This is done with scoring systems assessing the severity of a patient's condition using information such as acute physiological status, age, health history, and reason for admission [2–6]. These systems quantify the risk associated with patients presenting to ICUs using severity scores and predicted risk of mortality. These systems have been implemented in hospitals around the world and further improving them represents an ongoing area of research.

There has been a great deal of literature studying ICU scoring systems since the first system was published in 1981 [7]. This research has been motivated both by the value of better performance measures as well as the deteriorating performance of adopted scoring systems over time due to changes in practice and technology [2].

There are presently three widely used and researched scoring systems, with each having gone through several iterations since their original development in the 1980s [7–9]. These are the Acute Physiology and Chronic Health Evaluation (APACHE), Simplified Acute Physiology Score (SAPS), and Mortality Probability Model (MPM) [3]. The latest versions of each are the APACHE IV [2], the SAPS III [10], and the MPM_0 III [11]. Region-specific models have also been proposed in recent years [12,13].

The research relating to the improvement of scoring systems and risk adjustment can, for the most part, be broken into two streams. The first stream is where improvements are made within the currently used framework of logistic regression. Improvements have been achieved by increasing the number of variables used in existing models as well as by increasing the sophistication of the modelling process. This research stream is most evident when considering how the three major systems and region-specific models have developed since the original models

were put forward [2,5,9–12,14,15]. Motivated in part by the increasing volumes of data available, advanced statistical methods have been used in determining data transformations [2,10,14,16] and variable selection [2,10,11,16,17], while model equations have recently begun including interaction terms and cubic splines [2,5,11,12].

The second stream of research has moved away from traditional techniques towards those potentially better suited for the large data sets characterising the current big data paradigm. Research in this stream has considered various machine learning techniques such as artificial neural networks (ANNs) [18–25], decision trees [24,25], support vector machines (SVMs) [25], and ensemble methods[26,27]. While ANNs and SVMs are not clearly superior or inferior to logistic regression [18–22,24,25], positive results have been found for decision trees and an ensemble method [24–26]. The usage of new techniques has yet to filter through to practice, however, and logistic regression remains the dominant modelling approach implemented in hospitals.

While the two research themes described appear distinct in their approaches to improving model performance, there is a key commonality. Both themes have inherited the low-frequency approach to physiology variable representation used in the high-cost data environment of the 1980s when such systems were first developed.

The original APACHE system was developed in 1981 and produced an integer score representing the severity of patients' physiological condition [7]. The physiology score was calculated by considering many different underlying physiology variables. For each physiology variable, only the most extreme value observed in the first 24 h of ICU admission is used. These values are then discretised by comparing them to variable-specific ranges, and integer scores assigned depending on the range in which the values fell. The overall physiology score is then calculated by summing the scores for each variable.

While ICU mortality models have adapted to the current data-rich environment with increasingly data-driven techniques, advancements in variable representation have not been made. Systems continue to use only the most extreme observations, thereby ignoring the additional information from physiology variables now recorded at higher frequencies. Pulse and mean arterial pressure are, for example, recorded at minutely frequencies. Thus, our research question is:

Can estimates of mortality risk be improved using high frequency data recorded for the currently used physiology fields in the widely used APACHE III scoring system?

To answer this question, the APACHE III system is used as it is the model currently used in Australia, where data for this study was collected. While the most recent version is APACHE IV, it was developed in the US and requires data not routinely collected in Australian ICUs. A model using variables from APACHE III is compared with models that incorporate new variables based on high frequency data. Logistic regression is used because it is the most popular technique and currently used in all APACHE systems. While machine learning techniques are increasingly being considered in the literature, their use in this case would make it difficult to isolate differences in performance due to the new variables, rather than due to a difference in modelling technique.

As the technique is remaining constant, any differences in model performance are attributed to leveraging more information from high frequency data.

Current research has not yet made use of high frequency data, so this paper investigates the potential value of using data from automated, high frequency recording systems to complement existing research themes. Such an investigation has not yet been carried out in the literature and provides a novel avenue and logical next step for improving model performance. While a global ICU model is used as the benchmark, results are also expected to be relevant to models developed for more narrowly defined cohorts of ICU patients where variable construction for high frequency fields is based on only the most extreme observations.

2. Materials and methods

2.1. Data sources

The data used in the project consisted of 3128 ICU admissions to the Gold Coast University Hospital (GCUH), a large public teaching and referral hospital in Australia. Admissions occurring between 28-Sep-2013 and 31-Jan-2017 were included in the study, with data for each admission being sourced from two separate databases.

The first is the MetaVision database, which is an automated electronic clinical information system used in the ICU to record all information being captured for admitted patients, either directly or through a connected system.

The second database is used by the GCUH for reporting to the Australia and New Zealand Intensive Care Society (ANZICS) and is known as AORTIC. This database contains APACHE III scores for each ICU admission.

2.2. Data processing

The data extracted required further processing before being appropriate for use. Processing involved either the removal of entire admissions and all data associated with them, or the removal of select values.

2.2.1. Excluding admissions

The process of obtaining a set of admissions common to both MetaVision and AORTIC is shown in Table 1. Admissions were removed as per the exclusion criteria employed in the development of APACHE III or if there was insufficient data for calculating APACHE III scores or new variables.

2.2.2. Excluding values

The removal of erroneous values, likely from data entry errors, was performed by considering acceptable ranges for each field. Acceptable ranges were based on the Adult Patient Database Data Dictionary V5.4 [28] as shown in Table 2. The database associated with these guidelines

Table 1
: Included admissions.

Data	Step ID	Description	Admissions
AORTIC	S01	Extract admissions from AORTIC data	10515
AORTIC	S02	Remove admissions with patients aged under 16 years of age	10462
AORTIC	S03	Remove admissions where APACHE III scores have not been calculated	10461
MetaVision	S04	Extract admissions from MetaVision data	8809
MetaVision	S05	Remove admissions where no discharge had occurred by 31-01-2017	8780
MetaVision	S06	Remove admissions where the patient has previously been admitted to the ICU.	7485
MetaVision	S07	Remove admissions where the ICU stay was under 4 hours	7277
Joint Data	S08	Isolate common admissions from both AORTIC and MetaVision datasets	4416
Joint Data	S09	Remove admissions with outcomes other than "died" or "home"	3592
Joint Data	S10	Remove admissions with insufficient data to calculate APACHE III scores or new variables	3128

Table 2
: Acceptable ranges for physiology variables.

Physiology Variable	Minimum	Maximum	Units
Pulse	0	300	bpm
Mean Arterial Blood Pressure	0	300	mmHg
Temperature	20	46	°C
Respiratory Rate	0	80	bpm
PaO2	15	720	mmHg
FiO2	0.21	1	NA
PaCO2	5	250	mmHg
pH	6.3	8.5	NA
Haematocrit	0.05	0.75	NA
White Blood Cell Count	0	300	10^9/L
Creatinine	10	2500	μmol/L
Urine Output	0	30000	ml
Blood Urea Nitrogen	0.5	100	mmol/L
Sodium	100	215	mmol/L
Albumin	5	65	g/L
Bilirubin	1	1200	μmol/L
Glucose	0	90	mmol/L
Glasgow Coma Scale: Eyes	1	4	NA
Glasgow Coma Scale: Verbal	1	5	NA
Glasgow Coma Scale: Motor	1	6	NA
Glasgow Coma Scale: Total	3	15	NA

is highly reputable and has been used in previous studies to assess ICU performance in Australia and New Zealand [1]. Any values outside of acceptable ranges were removed. Because of clustering of values at the extremes, the maximum and minimum allowable values (0 and 300 beats per minute) were also removed for the pulse field after consultation with a senior clinician. Values were also removed for the fields if they were obviously extreme relative to preceding and subsequent values for the same admission. Specifically, observations for the high-frequency pulse and mean arterial pressure fields were classified as erroneous and removed if they were at least 30 units more extreme than any of the ten neighbouring observations. This rule was decided on in consultation with a senior clinician and removed 0.046% of all pulse observations and 0.38% of mean arterial pressure observations.

2.3. Variable calculation

After data extraction and processing, the variables used to calculate components of the APACHE III system and the new variables summarising the high frequency data were recorded for pulse and mean arterial pressure fields. The pulse and mean arterial pressure physiology variables were selected for this first study of high frequency data on the advice of a senior clinician and because they are commonly collected at minutely frequencies. The components from the APACHE III system are the Age Score, Chronic Health Evaluation Score, and Acute Physiology Score (APS). The process of calculating these scores for the APACHE III system is detailed in its accompanying publication [16] and is not further described in this paper. The modification of the original APACHE III system to incorporate the new variables is illustrated in Appendix A.

The five new variables add additional information about the distribution of high frequency data for the pulse and mean arterial pressure fields over the first 24 h of ICU admissions. As the first study to incorporate information from high frequency data, the first four variables are commonly used distributional descriptors based on the first four moments of the distribution. These are the average, standard deviation, skewness, and the kurtosis of the observed values over the first 24 h of ICU admission. A slight modification is made to the standard calculations to be consistent with the emphasis that the APACHE III system places on deviations from normal values (where more abnormal values are assigned higher scores); the new distributional descriptor variables are calculated for the observed values minus a constant value. The constant represents a normal value for the physiology field: 75

beats per minute and 90 mmHg for the pulse and mean arterial pressure fields respectively. These values are the median of the range of score that are assigned zero (normal) by the APACHE III system. The formulae employed for these four variables are shown below:

$$x_i = Obs_i - Constant \tag{1}$$

$$Average(x) = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{2}$$

$$Standard\ Deviation(x) = \frac{1}{n-1} \times \sum_{i=1}^{n} (x_i - Average(x))^2 \tag{3}$$

$$Skewness(x) = \sqrt{n} \times \frac{\sum_{i=1}^{n} (x_i - Average(x))^3}{\left(\sum_{i=1}^{n} (x_i - Average(x))^2\right)^{3/2}} \times \left(1 - \frac{1}{n}\right)^{3/2} \tag{4}$$

$$Kurtosis(x) = n \times \frac{\sum_{i=1}^{n} (x_i - Average(x))^4}{\left(\sum_{i=1}^{n} (x_i - Average(x))^2\right)^2} \times \left(1 - \frac{1}{n}\right)^2 - 3 \tag{5}$$

The fifth variable calculated for pulse and mean arterial pressure fields was the proportion of turning points observed in the recorded values in the first 24 h of ICU admission. This variable was expected to provide an indication of the stability of a patient's condition while incorporating the longitudinal nature of the data. Turning points were defined as any values which were larger or smaller than the four preceding and four subsequent observations. The formula for this variable is given by the following equation:

$$Prop.\ TP = \frac{1}{n} \sum_{i=5}^{n-4} I(x_i = \max(x_{i-4},...,x_{i+4}) OR x_i = \min(x_{i-4},...,x_{i+4})) \tag{6}$$

Where $I(...)$ is an indicator function taking value one if the conditions in the brackets are true and zero otherwise.

2.4. Model construction

All models are constructed using AORTIC data for the three variables from the APACHE III system and MetaVision data for new variables based on high frequency data. Three different models are estimated as summarised in Table 3. The first model uses only the three original score variables from the APACHE III and is referred to as the Baseline model. The second and third models additionally use the five variables that leverage high frequency data respectively for pulse and mean arterial pressure. As the same model design and data will be used in estimating the models, differences in performance can be solely attributed to the differences in the variables included.

Before estimating the models, 50% of ICU admissions were randomly assigned to a training set. Table 4 describes the training, testing and aggregate data. The training data were used for model estimation, while the test data were used to assess model performance.

2.5. Model assessment

Models are evaluated on test data using performance metrics commonly employed in the literature. The first of these metrics is the accuracy of the models in assigning class labels of "survivor" and "death", measured both overall and for each class. Second, model calibration is assessed using standardised mortality ratios (SMRs) and the Hosmer-Lemeshow Ĉ statistic. Lastly, model discrimination is measured using the Receiver Operating Characteristic (ROC) curve. Comparisons between model discrimination were made through the standard approach of computing the area under the ROC curve (AUC). On recommendation from a reviewer, the Net Reclassification Index (NRI) of the models were also computed.

Table 3

: Model variables.

	Variables	Baseline	Baseline + Pulse	Baseline + MAP
Original	Acute Physiology Score	X	X	X
	Age Score	X	X	X
	Chronic Health Evaluation Score	X	X	X
New Pulse Variables	Average		X	
	Standard Deviation		X	
	Kurtosis		X	
	Skewness		X	
	Turning Points		X	
New Mean Arterial Pressure Variables	Average			X
	Standard Deviation			X
	Kurtosis			X
	Skewness			X
	Turning Points			X

Table 4

: Admissions and class counts.

	Admissions	Survivors	Deaths
Train Data	1564	1399	165
Test Data	1564	1409	155
All	3128	2808	320

3. Results and discussion

3.1. Variables used

The five new variables that leverage high frequency data have not been considered in previous research and so summary histograms are provided in Appendix B. Similar distributions are found for each summary variable for both pulse and mean arterial pressure, though differences in spread can be seen.

3.2. Estimated models

The three logistic regression models estimated are shown in Table 5, where ** and *** denote statistical significance at the 1% and 0.1% levels respectively. For the Acute Physiology, Age and Comorbidity scores, all coefficients were associated with p-values below 0.1%, reflecting their prognostic value within the APACHE framework. Of greater interest are the coefficients of the new variables that incorporate additional information about pulse or mean arterial pressure.

Table 5

: Model coefficients and P-values.

Variable	Baseline Coefficient (P-Value)	Baseline + Pulse Coefficient (P-Value)	Baseline + MAP Coefficient (P-Value)
Intercept	−8.3883 (< 2E-16***)	−7.4933 (< 2E-16***)	−4.8532 (7.22E-10***)
Acute Physiology Score	0.0626 (< 2E-16***)	0.0558 (< 2E-16***)	0.0558 (< 2E-16***)
Age Score	0.0574 (0.0003***)	0.0683 (6.30E-05***)	0.0687 (5.83E-05***)
Comorbidity Score	0.0895 (0.0004***)	0.0923 (0.0005***)	0.0903 (0.0008***)
Average (as per Eq. (2))		0.0040 (0.5568)	0.0051 (0.5737)
Standard Deviation (as per Eq. (3))		0.0731 (0.0016**)	0.0079 (0.5338)
Skewness (as per Eq. (4))		−0.0497 (0.6210)	−0.0441 (0.5410)
Kurtosis (as per Eq. (5))		−0.0160 (0.1642)	0.0013 (0.7619)
Turning Points (as per Eq. (6))		−11.5032 (6.07E-06***)	−23.4127 (4.93E-08***)

For the common distributional descriptors, the associated coefficients were mostly not significantly different to zero even at the 10% level. The sole exception to this was the standard deviation of the pulse data, which was significant at the 1% level. In contrast, the turning points variable was highly significant at the 0.1% level in both models in which it was used. The negative weighting on this variable indicates that observing higher proportions of turning points is associated with a decreased chance of death. A possible explanation for this may be that a high proportion of turning points is associated with natural oscillations around a normal value, while a low proportion of turning points is associated with a steady decline in patient condition.

3.3. Model performance

Model performance is shown in the tables below in terms of accuracy (Table 6), calibration (Tables 7–8) and discrimination (Table 9). Similar accuracy was observed overall all models, driven by excellent performance in capturing patients who survive. Performance in the minority class is naturally lower for all three models as well. Across all accuracy measures, the models with new variables outperformed the Baseline model. The NRI for both models, as compared to the Baseline model, were not significant (Baseline + Pulse: NRI = 0.0136, P-Value = 0.5991; Baseline + MAP: NRI = 0.0107, P-Value = 0.5829).

In terms of calibration, both overall and for different risk groups, Tables 7 and 8 demonstrate that all models are acceptable. Predicted risk is not significantly different from actual risk of mortality.

Discrimination was shown in Table 9 to be excellent in two of the three models, with AUCs greater than 90%. The Baseline + Pulse model underperformed the Baseline, but the best performance was achieved with the Baseline + MAP model. The AUC of the Baseline + MAP model represents an improvement of almost 1% compared to the Baseline model. The improved discrimination of the Baseline + MAP model over the Baseline model was found to be statistically significant at the 5% level (P = 0.03297, one-sided test) using the DeLong test for correlated ROC curves [29].

4. Discussion and conclusion

This paper set out to assess whether high frequency data, which is increasingly prevalent in modern ICUs, could be used to improve ICU scoring systems. Using models based on the APACHE III score, the

Table 6

: Model accuracy.

	Baseline	Baseline + Pulse	Baseline + MAP
Accuracy (Overall)	91.43%	91.62%	91.88%
Accuracy (Deaths)	34.84%	36.13%	35.38%
Accuracy (Survivors)	97.66%	97.73%	98.08%

Table 7
: Standardised mortality ratios.

	Baseline	Baseline + Pulse	Baseline + MAP
SMR	0.9646	0.9886	0.9718
P-Value[1]	0.34518	0.46429	0.38031
95% Confidence Interval	0.85485–1.08848	0.8784–1.11266	0.86181–1.09593

[1] The p-values in Table 8 were calculated using an accurate approximation to the exact Poisson test, as described by Breslow and Day [30].

Table 8
: Hosmer-Lemeshow Ĉ tests.

	Baseline	Baseline + Pulse	Baseline + MAP
Chi-Squared Statistic	6.7981	6.5556	11.1747
P-Value	0.5586	0.5852	0.1920

Table 9
: Area under the receiver operating characteristic curve.

	Baseline	Baseline + Pulse	Baseline + MAP
AUC	90.45%	89.30%	91.40%

results demonstrate that this is possible. The addition of new variables summarising the available high frequency data for two currently used physiology variables (pulse and mean arterial pressure) resulted in improvements in model performance. Further, improvements on relevant measures did not come at the cost of systemic deterioration in other areas of model performance.

Models both with and without the new variables were found to be adequately calibrated, with no significant differences between actual and predicted risk either overall or in risk sub-groups. In this respect no model was found to be better or worse than others. All models were also found to have very good discrimination, with the Baseline and the Baseline + MAP models having AUCs above 90%. The Baseline + MAP model was best in terms of AUC, with a significant improvement on the Baseline model of almost 1%. Improved classification accuracy was also observed for both models using the new metrics, but compared to the baseline model the reclassification improvement was not significant. Overall, the new metrics resulted in a significant discrimination improvement for the Baseline + MAP model, non-significant accuracy improvements for both models, and retained good calibration in both cases.

These findings indicate that high frequency data for physiology fields can be better leveraged to improve estimates of mortality risk when compared with the currently used APACHE III scoring system. Furthermore, it can be expected that more value can be extracted from these fields than shown in this paper for two reasons. First, as an initial investigation this paper chose common distribution descriptors as metrics, and there is large scope to further investigate alternative functions of the underlying data more strongly related to outcomes. Model performance can be expected to be improved beyond that demonstrated in this paper by adopting a more rigorous approach to constructing new metrics that use higher frequency information. Secondly, variable selection techniques were not employed to remain consistent with the model using the APACHE III variables. Adding five new variables to a logistic regression increases the complexity of the model and potentially the noisiness of predictions on unseen data. This is particularly relevant as the new metrics were all functions of the same underlying data and so likely had some overlapping information content. Despite this, an improvement was still observed after adding in the new metrics. This indicates that the benefit of better leveraging the high frequency data available outweighed the additional noise added to the model.

Employing a more sophisticated approach to variable inclusion and model construction would reduce the effect of noise while retaining the predictive power of the new variables, resulting in improved model performance.

The key finding of this paper is that model performance can be improved by better leveraging the high frequency physiology data being collected in ICUs. This contribution represents a new third research stream that can be pursued for model improvement. This new approach represents a new avenue of research and complements the two existing research streams of increasing logistic regression model complexity and the use of data-driven techniques.

Overall, this study validates the approach of using high frequency information in ICU scoring systems. This study found that the most valuable variable considered for high frequency data was the proportion of turning points observed, which provides some insight into the stability of a patient's condition during the first 24 h of their ICU stay. Future research should aim to build upon this research by considering patient populations of multiple hospitals, utilising high frequency information for variables other than pulse and mean arterial pressure, and determining whether alternative metrics could further improve performance. Further, differences between approaches taken in ICU scoring systems may lead to variable value derived from better utilising high frequency information. The latest iteration of the SAPS model considers only the first hour of ICU admission while the MPM systems can consider several time frames but considers only binary variables. The value of high frequency metrics in the context of other ICU scoring systems should also be an area of future research.

Data statement

The data used in this study are available from Gold Coast Hospital and Health Service, but restrictions apply to the availability of these data, which were used under licence for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Gold Coast Hospital and Health Service.

Author statement

JT was responsible for the manipulation and analysis of data and was the primary writer. The data acquisition was performed by BR, who also provided the context of the problem based on his clinical expertise. AG, BV, and JT were responsible for conceiving the study design. AG and BV also provided critical revision of this paper and the analysis used.

Summary Points
What is already known?

- There are currently a wide range of scoring systems in use around the world for assessing ICU performance on an equitable basis.
- Research seeking to further improve the performance of these scoring systems has focused on increasing sophistication within the logistic regression framework or through adoption of machine learning methods.
- In all models, physiological predictors are based only on the most extreme values observed during the initial ICU admission.

What has this paper added to the body of knowledge?

- This is the first paper to consider alternative summaries of physiological variables instead of using only the most extreme values for predicting mortality in the ICU
- The results of this study demonstrate that such summaries can leverage additional information content not captured in current approaches, leading to improvements in mortality models through an avenue distinct from current research themes.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijmedinf.2019.07.002.

References

[1] J. Kasza, J.L. Moran, P.J. Solomon, Evaluating the performance of Australian and New Zealand intensive care units in 2009 and 2010, Stat. Med. 32 (21) (2013) 3720–3736.

[2] J.E. Zimmerman, A.A. Kramer, D.S. McNair, F.M. Malila, Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients, Crit. Care Med. 34 (5) (2006) 1297–1310.

[3] M.J. Breslow, O. Badawi, Severity scoring in the critically ill: part 1–interpretation and accuracy of outcome prediction scoring systems, Chest. 141 (1) (2012) 245–252.

[4] H. Lee, Y.J. Shon, H. Kim, H. Paik, H.P. Park, Validation of the APACHE IV model and its comparison with the APACHE II, SAPS 3, and Korean SAPS 3 models for the prediction of hospital mortality in a Korean surgical intensive care unit, Korean J. Anesthesiol. 67 (2) (2014) 115–122.

[5] P. Ferrando-Vivas, A. Jones, K.M. Rowan, D.A. Harrison, Development and validation of the new ICNARC model for prediction of acute hospital mortality in adult critical care, J. Crit. Care 38 (2017) 335–339.

[6] A. Pérez, R.J. Dennis, J.F.A. Gil, M.A. Rondón, A. López, Use of the mean, hot deck and multiple imputation techniques to predict outcome in intensive care unit patients in Colombia, Stat. Med. 21 (24) (2002) 3885–3896.

[7] W.A. Knaus, J.E. Zimmerman, D.P. Wagner, E.A. Draper, D.E. Lawrence, APACHE-acute physiology and chronic health evaluation: a physiologically based classification system, Crit. Care Med. 9 (8) (1981) 591–597.

[8] J.R. Le Gall, P. Loirat, A. Alperovitch, et al., A simplified acute physiology score for ICU patients, Crit. Care Med. 12 (11) (1984) 975–977.

[9] S. Lemeshow, D. Teres, H. Pastides, J.S. Avrunin, J.S. Steingrub, A method for predicting survival and mortality of ICU patients using objectively derived weights, Crit. Care Med. 13 (7) (1985) 519–525.

[10] R.P. Moreno, P.G. Metnitz, E. Almeida, et al., SAPS 3–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, Intensive Care Med. 31 (10) (2005) 1345–1355.

[11] T.L. Higgins, D. Teres, W.S. Copes, B.H. Nathanson, M. Stark, A.A. Kramer, Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III), Crit. Care Med. 35 (3) (2007) 827–835.

[12] E. Paul, M. Bailey, D. Pilcher, Risk prediction of hospital mortality for adult patients admitted to Australian and New Zealand intensive care units: development and validation of the Australian and New Zealand Risk of Death model, J. Crit. Care 28 (6) (2013) 935–941.

[13] D.A. Harrison, G.J. Parry, J.R. Carpenter, A. Short, K. Rowan, A new risk prediction model for critical care: the Intensive Care National Audit & Research Centre (ICNARC) model, Crit. Care Med. 35 (4) (2007) 1091–1098.

[14] W.A. Knaus, E.A. Draper, D.P. Wagner, J.E. Zimmerman, APACHE II: a severity of disease classification system, Crit. Care Med. 13 (10) (1985) 818–829.

[15] J.R. Le Gall, S. Lemeshow, F. Saulnier, A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study, JAMA 270 (24) (1993) 2957–2963.

[16] W.A. Knaus, D.P. Wagner, E.A. Draper, et al., The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults, Chest 100 (6) (1991) 1619–1636.

[17] S. Lemeshow, D. Teres, J. Klar, J.S. Avrunin, S.H. Gehlbach, J. Rapoport, Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients, JAMA. 270 (20) (1993) 2478–2486.

[18] G.S. Doig, K.J. Inman, W.J. Sibbald, C.M. Martin, J.M. Robertson, Modeling mortality in the intensive care unit: comparing the performance of a back-propagation, associative-learning neural network with multivariate logistic regression, Proc. Symp. Comput. Appl. Med. Care (1993) 361–365.

[19] R. Dybowski, P. Weller, R. Chang, V. Gant, Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm, Lancet 347 (9009) (1996) 1146–1150.

[20] L.S. Wong, J.D. Young, A comparison of ICU mortality prediction using the APACHE II scoring system and artificial neural networks, Anaesthesia 54 (11) (1999) 1048–1054.

[21] G. Clermont, D.C. Angus, S.M. DiRusso, M. Griffin, W.T. Linde-Zwirble, Predicting hospital mortality for patients in the intensive care unit: a comparison of artificial neural networks with logistic regression models, Crit. Care Med. 29 (2) (2001) 291–296.

[22] A. Nimgaonkar, D.R. Karnad, S. Sudarshan, L. Ohno-Machado, I. Kohane, Prediction of mortality in an Indian intensive care unit. Comparison between APACHE II and artificial neural networks, Intensive Care Med. 30 (2) (2004) 248–253.

[23] M. Suka, S. Oeda, T. Ichimura, K. Yoshida, J. Takezawa, Comparison of proportional hazard model and neural network models in a real data set of intensive care unit patients, Stud. Health Technol. Inform. 107 (Pt 1) (2004) 741–745.

[24] L.G. Gortzis, F. Sakellaropoulos, I. Ilias, K. Stamoulis, I. Dimopoulou, Predicting ICU survival: a meta-level approach, BMC Health Serv. Res. 8 (2008) 157.

[25] S. Kim, W. Kim, R.W. Park, A comparison of intensive care unit mortality prediction models through the use of data mining techniques, Healthc. Inform. Res. 17 (4) (2011) 232–243.

[26] R. Pirracchio, M.L. Petersen, M. Carone, M.R. Rigon, S. Chevret, M.J. van der Laan, Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study, Lancet Respir. Med. 3 (1) (2015) 42–52.

[27] A. Awad, M. Bader-El-Den, J. McNicholas, J. Briggs, Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach, Int. J. Med. Inform. 108 (2017) 185–195.

[28] P.J. Stow, G.K. Hart, T. Higlett, et al., Development and implementation of a high-quality clinical database: the australian and New Zealand intensive care society adult patient database, J. Crit. Care 21 (2) (2006) 133–141.

[29] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (3) (1988) 837–845.

[30] N.E. Breslow, N.E. Day, Statistical methods in cancer research. Volume II–the design and analysis of cohort studies, IARC Sci. Publ. (82) (1987) 1–406.