International Journal of Biomedical and Clinical Sciences
Vol. 5, No. 4, 2020, pp. 408-417
http://www.aiscience.org/journal/ijbcs

A Machine Learning Algorithm to Classify High Risk Mortality Patients for Intensive Care Unit Admissions

Carol Anne Hargreaves[1, *], Tan Hui Ling Juliet[1], Dinesh Kumar Srinivasan[2]

[1]Department of Statistics and Applied Probability, National University of Singapore, Singapore
[2]Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Abstract

In this paper, we focus on the classification of high mortality risk patients for Intensive Care Units (ICUs). Classification algorithms for identifying ICU mortality are necessary for measuring and improving ICU performance. Mortality risk severity scores are an essential part of hospital management and clinical decision-making. Proper application of classification models can help in decision making to lower hospital costs. In fact, classification high mortality risk models have become a necessary tool to explain differences in mortality risk. Purpose of Study: The purpose of this study is to develop and evaluate a new algorithm which more accurately predicts patient mortality in ICU, using patient information of vital signs and laboratory results only in the first 24 hours of ICU admission. We convert continuous variables into categorical variables and identify optimal threshold cut points for stabilizing the coefficients of the classification mortality risk model. In this paper, an optimal set of 3 threshold values were derived, that partitioned the data into 4 groups, resulting in the patient mortality risk scores being more distinguishable across the 4 partitioned groups. The most important variables for the ICU Mortality Risk was PO2 (120 – 125), followed by Cardiac Arrest (Yes), Bilirubin (0.75 – 1), Vasopressors (Yes), SPO2 (< 66), Bilirubin (>7.75), Foley (<6), Severe COPD (Yes), WBC (> 19.5) and BUN (> 49). Our proposed optimal threshold cut point model performed substantially better (AUC=0.944) in identifying ICU patients with high mortality risk compared to the current scoring systems commonly used in hospitals, such as the SAPS 11 (AUC =0.771), APACHE 11 (AUC=0.736) and SOFA (AUC=0.699). This accuracy is at least 30% (1.35 times) better than current mortality risk scoring systems. SAPS 11, APACHE 11 and SOFA are static algorithms whereas our new optimal threshold algorithm is a data-driven algorithm which predicts mortality in ICU patients in real-time and may be useful for the timely identification of deteriorating patients. Our new binary classification algorithm will allow clinicians to accurately identify high-mortality risk patients early within 24 hours so that they can be given prompt treatment to reduce their risks of deteriorating or dying.

Keywords

Intensive Care Units (ICU), Critical Care, Mortality Risk, Classification Algorithm, Optimal Threshold Cut Points, SAPS11, APACHE11, SOFA

Received: August 21, 2020 / Accepted: September 15, 2020 / Published online: December 11, 2020

1. Introduction

Quantifying patient health status and predicting mortality outcome is a challenge in critical care research. ICU mortality prediction plays an important role in patient health care and hospital resource allocation, which contributes to improving patient survival [1]. Furthermore, they provide us with a method of assessing the performance differences between different medical facilities and services and thus help to reduce disparities in health care. For these purposes,

* Corresponding author
E-mail address: stacah@nus.edu.sg (C. A. Hargreaves)

more and more researchers have been devoted to the study of improving the accuracy in identifying ICU patient mortality, from former score-based system such as APACHE, SAPS, and SOFA to now machine learning technology [1].

Three commonly used mortality prediction models are the Acute Physiology and Chronic Health Evaluation (APACHE) [2, 3], the Simplified Acute Physiology Score (SAPS) [4, 5], and the Sequential Organ Failure Assessment (SOFA) score [6]. Over the years, efforts were made to improve the performance of these pre-existing mortality risks scoring systems. The most recent versions of the SAPS and APACHE scoring systems are SAPS III, APACHE IV [3, 5]. SOFA, on the other hand, was designed to evaluate a patient throughout the ICU stay. It assigns a score of 0 (normal) to 4 (very abnormal) for six different organ systems on each day in the ICU [6]. Unlike APACHE and SAPS, SOFA was originally intended to characterize patient morbidity as opposed to predict patient mortality; however, since its development, it has often been used for the latter purpose [7].

Extensive research and validation were only done for SAPS, SAPSII, APACHE and APACHEII on different populations of ICU patients. Also, only these versions have publicly available scoring chart to guide users on how to tabulate the mortality risk scores. Therefore, our best choice for baseline scoring systems for this paper are SAPSII, APACHEII and SOFA. In all three systems, the decision of which variables to include relied, at least in part, on clinical expertise and domain knowledge [2, 5, 6]. In the case of SOFA, the entire system was designed through clinical consensus. Our study is an improved system for predicting mortality risk in ICU as it uses a data-driven classification model which is more dynamic and proved to be more accurate than the SAPSII, APACHEII and SOFA methods.

In a clinical study [8] with 3,700 patients admitted to the ICU in a university teaching hospital between 1997 to 2003, the top risk factors for death in ICU were concluded to be cardiovascular failure and need for re-admission to ICU. There was also a significantly positive relationship between the number of failing organs and the ICU mortality rate. In a separate cohort study [9], the author stresses that the majority of patients present had at least one organ failure at the time of death. This provided evidence for multiple organ dysfunction syndrome as another crucial risk factor of mortality in ICU. In another study [10], it identified cancer, blood PH, and level of consciousness at the ICU admission as significant risk factors for mortality.

With the increasing popularity and successes of machine learning in the data science field, many researchers started to adopt these models to do clinical predictions. Earlier [11] and recent works [12] showed that machine learning models obtained good predictive performance, and significantly outperformed SAPS, APACHE and SOFA. This was further validated by a recent study [13] which uses the same MIMIC III database for their clinical analysis.

In one published medical paper [14], it argued that most of the traditional scoring systems like SAPS II and SOFA "assume that risk factors are independent from one another, and hence are not sensitive to underlying complex homeostatic physiologies of patients". In response to this limitation, it introduced a new method called the Auto-Triage scoring system. The basic idea was to assign weights to combinations of variables based on the corresponding correlations between individual, pairs and triplets of variables with in-hospital death. A logistics regression model was then used to scale and combine the weights to form the Auto-Triage scores. The Auto-Triage method provided a baseline prediction score 12 hours in advance for ICU patients, and it showed significant improvement in the accuracy and specificity as compared to the SAPSII and SOFA severity scores.

In another research paper [15], the author presented a new novel approach called the Univariate Flagging Algorithm to predict mortality outcome for patients in ICU. Our main methodology, which is to be discussed in a later section, is inspired by this research paper. The basic idea behind this approach was to identify optimal thresholds for each input variable from a list of candidate threshold values. The Z-statistic was used to test for statistical significance. This algorithm first finds the optimal threshold below the median, and the procedure was repeated to find the optimal threshold above the median, to eventually give 2 most significant thresholds as the optimal threshold values for each input variable. After which, the usual model fitting and cross validations were done, and all the machine learning methods once again outperformed the pre-existing scoring systems by significant margins. In this paper, machine learning models were also shown to consistently outperform the SAPSII, APACHEII and SOFA scoring systems.

This paper is structured into 5 sections. While Section 1 is the introduction, Section 2 gives a brief overview of the methods used, Section 3 the results, and Section 4 the discussion, after which Section 5 presents the conclusion.

2. Methods

2.1. Data Source

Our data came from the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) III database [16]. The MIMIC-III Clinical Database is a large, publicly available database which contains de-identified health related data of

about 60,000 admissions, including readmissions, of patients who stayed in the ICU (critical care units) of the Beth Israel Deaconess Medical Care Center between 2001 and 2012.

The database includes information such as demographics, vital sign measurements made at the bedside, laboratory test results, procedures, medications, nurse and physician notes, imaging reports and out-of-hospital mortality. Only patients with at least 25 hours of length of stay or were alive for at least 25 hours were included in our cohort sample. As a result, our sample only has 11,105 patients, filtered from a total of 13,651 patients from the MIMIC III Metavision database record system. A 5-fold cross validation dataset was used where, the data was first split into 5 mutually exclusive datasets of approximately equal size. Each algorithm was run iteratively on the 5 training sets and used to predict mortality risk for patients using the remaining 4 validation datasets.

2.2. Feature Selection

To build a classification model for mortality risk, we first need to know the variables that should be included in the model. Table 1. shows the list of all candidate variables.

Table 1. Candidate Variables for Classification.

Organ System	SAPS II, APACHE II, SOFA Variables
Respiratory	PO2, FiO2, Heart rate, Respiratory rate, A-aPO2
Nervous	Glasgow Coma Scale – eye opening, verbal response, motor response
Cardio-vascular	Blood Pressures – Systolic, Diastolic, Mean, Vasopressors
Liver	Bilirubin
Blood	Platelets, White blood Cells, Bicarbonate, Sodium, Potassium, Blood Urea Nitrogen
Kidney	Creatinine, Hematocrit
Others	Temperature, Admission Age, Admission Type, Hematologic Malignancy*
ORGAN SYSTEM	NEW VARIABLES FROM LITERATURE REVIEWS
Respiratory	SPO2, O2 flow, Mechanical Ventilation
Cardio-vascular	Past History of Cardiac Arrest
Liver	Past History of Cirrhosis
Endocrine	Glucose
Kidney	Foley
Others	Past History of COPD*, Immunocompromised*, Metastatic Nonblood Cancer*, Admission Location

*These variables are specific diagnosed conditions upon admission.

With these candidate variables, we will first discuss how to select the significant numeric/temporal variables, followed by the steps taken for the categorical variables.

2.2.1. Steps for Numerical Variable Selection

Since we are constructing a baseline mortality risk prediction model, we set the window frame for measurements for each input variable to be within 24h from admission time.

1. Check the patient coverage for each of the 27 numeric variables and filter out variables that do not have adequate patient coverage.
2. Replace missing measurements with the median value of the respective mortality status group that the patient belongs to, for numeric variables with adequate patient coverage. For example, the survivors group has a median eye opening (minimum) score of 3. Hence a patient who survived with missing eye-opening measurement will be assigned an eye-opening (minimum) score of 3.
3. For each numeric variable, perform a 2-Sample T-test, to test whether the mean of the survivor group is significantly different from the mean of the non-survivor group. In other words, we test the following hypothesis:

$H_0: \mu_{survivor} = \mu_{nonsurvivor}$

$H_1: \mu_{survivor} \neq \mu_{nonsurvivor}$

For each variable, we test for significant differences for at least one of the following descriptive statistics: minimum, maximum and median. For example, for the minimum, we extract the minimum measurement value within the specified window frame for each patient and use these values to construct our t-statistic to test for significant differences in the means between the survivors and non-survivor groups. We will then compare the t-test results for all the three descriptive statistics to select the most significant one. In short, for each numeric variable, with at least 1 descriptive (minimum/maximum/median) t-test statistic result with a significant p-value (<0.05), choosing the descriptive statistic with the most significant p-value for a candidate selection variable for our classification algorithm. Otherwise, we drop the variable entirely (i.e. p-value > 0.05 for all descriptive statistics)

Table 2 below gives us the finalized list of shortlisted numeric variables as well as their corresponding descriptive statistics used.

Table 2. Significant Numeric Variables.

GCS eye opening (min)
GCS verbal response (min)
GCS motor response (min)
Systolic Blood pressure (min)
Diastolic Blood pressure (min)
Mean Blood pressure (min)
Temperature (min)
Heart rate (max)
Respiratory rate (median)
PO2 (median)
SpO2 (min)
Bicarbonate (min)
Bilirubin (max)
White blood cell (min)
Creatinine (max)
Hematocrit (max)
Platelets (min)
Sodium (min)
Potassium (median)
Blood urea nitrogen (min)
Foley (min)
Glucose (max)
Admission Age

It was interesting to discover that the Mean Blood Pressure (min) was more significant for classifying ICU patient mortality risk than Mean Blood Pressure (max).

2.2.2. Steps for Categorical Variable Selection

To select the categorical variables, the chi-square test is used to test for significant differences in the mortality rates between the factor levels of each categorical variable. Table 3 shows the finalized list of selected categorical variables and their corresponding factor levels

Table 3. Significant Categorical Variables.

ORGAN SYSTEM	VARIABLE	FACTOR LEVELS
Respiratory	Mechanical Ventilation	Yes
		No
Cardio-vascular	Vasopressors	Yes
		No
	H/O Cardiac Arrest	Yes
		No
Liver	H/O Cirrhosis	Yes
		No
Others	Admission Location	Emergency Room/ICU
		Ward
		Outside ICU
	Admission Type	Planned
		Unplanned
	Hematologic Malignancy	Yes
		No
	H/O Immuno-Compromised	Yes
		No
	H/O COPD	Yes
		No
	H/O Metastatic Cancer	Yes
		No
	Co-Morbidities	None
		1
		2
		3-4

2.3. Logit Regression Using Threshold Cut points for Continuous Variables

Instead of keeping the numeric variables as continuous variables, we convert them into categorical variables by setting partition values. The basic idea is to find a set of threshold cut points that can maximize the differences in mortality risk for survivor and non-survivor patients.

In Table 4. below, we have the binary outcome variable, Y = {0,1} (0: survivor patient; 1: non-survivor patient), the predictor variable, X, and c, the candidate threshold value.

Table 5. is the scenario with 2 threshold values, c1 and c2.

Table 4. Partitioning of a Continuous Variable with 1 threshold value, c1.

	$X \leq c_1$	$X > c_1$
Y = 0	n_{11}	n_{12}
Y = 1	n_{21}	n_{22}

*Note: n11, n12, n21, n22 are the cell counts.

Table 5. Partitioning of a Continuous Variable with 2 threshold values, c1 & c2.

	$X < c_1$	$c_1 \leq X \leq c_2$	$X > c_2$
Y = 0	n_{11}	n_{12}	n_{13}
Y = 1	n_{21}	n_{22}	n_{23}

*Note: n11, n12, n13, n21, n22, n23 are the cell counts.

For selected variables with very high variances, to avoid having small numbers in one of the groups following the dichotomization, thereby prevent substantial losses in statistical power, we will exclude the outer 10% of the continuous covariate distribution, and only use the inner 90% of the distribution as our selection interval to choose the threshold cut points.

For each variable, let the number of candidate threshold values be k. For each candidate threshold value, obtain the chi-square test statistic and p-value and arrange the k candidate threshold values in descending order of the magnitude of chi-square statistic value. To ensure the reliability of the chi-square test statistic, we only keep threshold values such that n11+ n21> 50 and n12+ n22> 50 (Refer to Table 4). In other words, the partitioned cell counts must be sufficiently large. Candidate threshold values with insignificant chi-square test statistics < critical value x_1^2 (0.05) = 3.84 are also discarded.

Next, we select the best candidate threshold value c1 with the highest chi-square test statistic v1 and rerun the chi-square test on the (k-1) pairs of candidate threshold values – (c1, c2), (c1, c3), …, (c1, ck). Similarly, to ensure the reliability of the chi-square test statistic, we only keep pairs of threshold values such that n11+ n21 > 50, n12+ n22 > 50 and n13+ n23 > 50 (Refer to Table A1 in the appendix). Threshold pairs with insignificant chi-square test statistics <

critical value x_2^2 (0.05) = 5.99 are also discarded.

Arrange the pairs of threshold values in descending order of the magnitude of their corresponding chi-square test statistic values. Let the best performing pair of threshold values be c (1,2) and its corresponding p-value be v (1,2). If v (1,2) < v1, we will choose c (1,2) as our optimal threshold cut points. Otherwise, we keep c1 as our optimal threshold cut point for the variable.

Repeat this partitioning procedure to obtain the best performing set of 3 threshold values (Let the best performing set of 3 threshold values be c (1,2,3)), and do the p-value test again to determine whether to update the optimal threshold set from c (1,2) to c (1,2,3). The number of threshold values should be capped at 3 (i.e. maximum number of corresponding partitioned groups is 4), as it is not ideal to have too many partitions as it can introduce problems such as overfitting.

Table A1 and Table A2 in the appendix show the optimal set threshold values for 2 and 3 threshold values (i.e. c (1,2) and c (1,2,3)) for all continuous variables respectively. It is clear that the optimal set of 3 threshold values separate the data better and the mortality rates are more distinguishable across the 4 partitioned groups. We will use the 3-partition version for all variables. An exception is the temperature variable where its optimal set of 2 threshold values will be used instead of the optimal set of 3 threshold values, even though the p-value for the set of 3 threshold values is significantly smaller (Refer to Table 4 and Table 5) This is because it is not sensible to include an additional threshold value (c2) of 36.1 that is so close to the pre-existing threshold value of 36.0 (c1) (Refer to Table 5).

Note that for the 3 Glasgow Coma Scale variables (i.e. eye opening, verbal response and motor response), the original values are retained and are not subject to partitioning. Even though these variables are numeric, they are discrete (not continuous) variables, with integer values ranging from 1 to 4, 1 to 5 and 1 to 6 respectively.

With the exception of the Glasgow Coma Scale variables, all the other partitioned continuous variables and categorical variables are converted into dummy variables (0 and 1).

The stepwise variable selection algorithm was implemented in R using the My.stepwise.glm function from the My.stepwise R package.

3. Results

The stepwise variable selection procedure (with iterations between 'forward' and 'backward' steps) was used to obtain the best candidate reduced logistics regression model. The significance levels for entry (SLE) and for stay (SLS) were set at 0.15, and the best reduced regression model was identified manually by adding or dropping variables one at a time until all regression coefficients were significantly different from 0 at the chosen alpha level of 0.05.

For the stepwise variable selection procedure, we set the reference partition to the factor level with lowest mortality rate among all the factor levels for the variable. This was to allow easier analysis of the fitted coefficients of the logistics regression model to derive the scoring algorithm.

Using the scoring chart in Table A3 in the appendix, we obtain a Total Score for each patient, by summing up the relevant scores (fitted coefficients). The formulae below convert the Total Score into the Predicted Mortality Risk (%).

$$\text{Logit} = -6.29014 + \text{Total score}$$

$$Mortality\ Risk = \frac{\exp(Logit)}{1 + \exp(Logit)}$$

$$Mortality\ Outcome = \begin{cases} Survived, if\ Mortality\ Risk < 0.5 \\ Died, if\ Mortality\ Risk \geq 0.5 \end{cases}$$

The performance of our Logistic Optimal Threshold Classification Model was evaluated using the Area under the Curve (AUC), Sensitivity and Specificity. All three metrics produced excellent performance results. See Table 6 below.

Table 6. Performance of Logistic Optimal Threshold Model.

Area under the Curve (AUC)	**0.925**
Sensitivity	0.871
Specificity	0.834

The binary optimal threshold cut point model results were considered good.

From Figure 1, the top 10 factors contribute to 40% of the variability for mortality risk scoring. Further, PO2 (120-125), Age (>=73) and Eye Opening are the most important factors for classifying high mortality risk ICU patients.

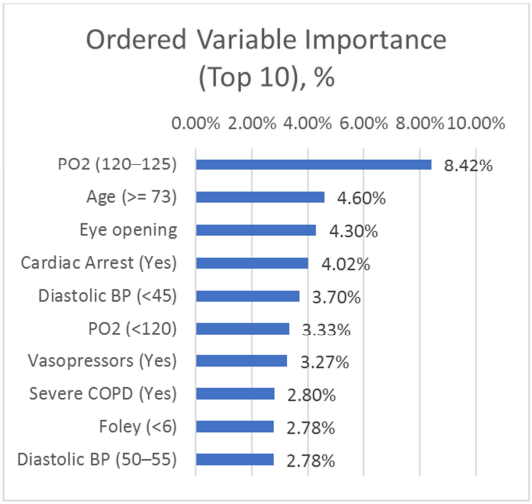

Figure 1. Variable Importance for Classifying ICU Mortality Risk.

We tested the performance of our model on unseen data. The binary optimal threshold cut point model approach performed extremely well with an AUC of 0.944, sensitivity of 0.867 and specificity of 0.876.

4. Discussion

Logistics regression models are sensitive to outliers. Extreme outliers have a significant impact on the fitted model, and this will in turn result in unreliable fitted coefficients. To address this problem, we partitioned the continuous variables into optimal threshold sub groups, to maximize the mortality differences across the sub groups. Hence, instead of using the raw continuous variables to fit the model, we use the partitioned threshold groups as the input variables.

The most crucial step to obtain a good model is the choice of threshold values, as the stepwise variable selection relies heavily on it to obtain the best reduced logistics regression model. As seen in the previous section, the selected variables can change drastically when the threshold values for just 1 variable (bilirubin) are modified. However, a consistent pattern in the order of the top few most important variables was observed. For instance, vasopressors, eye-opening and PO2 levels were consistently the top 3 most important variables in almost all the fitted models.

The below pointers summarize the important results of our methodology.

1. Out of the 3 Glasgow coma scale scoring variables, eye opening is the most important indicator for mortality risk.
2. There is an exponential increase in mortality risk as age increases.
3. For patients with hypertension (i.e. high blood
4. pressure), the maximum blood pressure recorded in the first 24h (baseline) is not correlated with mortality risk. Instead, the minimum blood pressure measurements provide more useful information in the prediction of ICU mortality risks.
5. While the partitioned numeric temporal variables were able to produce more accurate mortality prediction results in comparison to using the raw numeric variables, the partitions were heavily dependent on the data. It is most applicable when the clinical data outliers are a result of measurement recording error.
6. Most of the non-temporal variables that were used in this report are critical factors of ICU mortality, as they provide more concrete information about the severity of the patients' conditions upon admission or during the first 24h of admission.
7. The one factor that stood out from the rest is whether the patient has cardiac arrest upon admission. Among the 277 patients who had cardiac arrest upon admission, 49.5% of them did not survive.
8. As a risk factor indicator on its own, the number of co-morbidities that a patient has is clearly positively correlated with mortality risk. However, when we fit the model to include all the variables, the number of co-morbidities does not provide new information and hence was dropped from the model.

5. Implementation of R Shiny Tool

Our end goal is to be able to come up with an efficient tool for clinicians to predict the mortality risk of ICU patients within 24 hours from admission. Also, we will want the tool to be able to help pinpoint the crucial factors that are driving the high or low mortality risk so that the clinicians can know what to pay attention to in order to reduce the patient's risk of dying as early as possible. The Mortality Risk Calculator R Shiny App acts as an interactive platform to serve this purpose. The link for this App is at [17].

This R Shiny App (see Figure A1 and Figure A2 in the Appendix) is interactive as it allows clinicians to choose the patient information they want to view, by selecting the patient ID from the sidebar panel. The main panel has 3 main tabs. The first one is named as "Admission/ICU data" and it displays all the values/ raw measurements of the input variables. The second tab is "Mortality Outcomes" (see figure A1 in the Appendix) and it shows the predicted mortality status, mortality risk group as well as the mortality risk probability.

For comparison purposes, we also included the actual mortality status as one of the columns. The last tab is the "Notes for Clinicians", and it shows the summary of the main drivers of the patient's low or high mortality risk. These summaries help the doctors to quickly understand a patient's health status at first glance.

6. Conclusion

Improved mortality prediction for patients in intensive care units (ICU) remains an important challenge. The purpose of this study was to develop and evaluate a new algorithm which more accurately predicts patient mortality in ICU, using patient information of vital signs and laboratory results only in the first 24 hours of ICU admission.

The Area under the Curve (AUC) is a good metric for determining model performance. Our new binary optimal

threshold cut point algorithm performed extremely well with an AUC of 0.944, sensitivity of 0.867 and specificity of 0.876. In fact, our proposed optimal threshold cut point model performed substantially better in identifying ICU patients with high mortality risk compared to the current scoring systems commonly used in hospitals, such as the SAPS 11 (AUC =0.771), APACHE 11 (AUC=0.736) and SOFA (AUC=0.699).

The current mortality risk scoring systems do not take into the account history of co-morbidities. Our new binary optimal threshold algorithm included co-morbidities such as

1) History of COPD

2) History of Cardiac Arrest

3) History of Cirrhosis

Our study predicts patient mortality based on vitals, demographics, and lab results during the first 24 hours, with new identified key predictors for ICU mortality risk. See Table 7 below.

Table 7. New Key Predictors for ICU Mortality Risk.

Organ System	New Variables	Contribution (%)
Cardiovascualr	History of Cardiac Arrest	4.02%
Others	History of COPD	2.8%
Kidney	Foley	2.78%

While PO_2(120-125), Age (>=73) and Eye opening are the top three variables of importance for identifying ICU mortality risk, H/O Cardiac Arrest contribute, H/O COPD and Foley are in the top ten most important variables (see figure 1).

Our new algorithm has made substantial improvements for predicting ICU mortality risk of patients, an increased accuracy of at least 17% when compared to current scoring systems. Further, SAPS 11, APACHE 11 and SOFA are static algorithms whereas our new optimal threshold algorithm is a data-driven algorithm which predicts mortality in ICU patients in real-time and may be useful for the timely identification of deteriorating patients.

Our algorithm developed machine learning methods to identify patients at high mortality risk in real time. The results suggest better accuracy compared to established scoring systems. The optimal threshold cut-point technique provides clinicians with thresholds that can be interpreted better in ICU environments. It makes it easier for doctors to apply the results from the classification algorithm to alleviate mortality risk.

Appendix

Table A1. Optimal set of 2 Threshold Values for Continuous Variables.

Variable	c_1	c_2	% death ($X < c_1$)	% death ($c_1 \leq X < c_2$)	% death ($X \geq c_2$)	p-value	X^2 statistic
Temperature	36.0	37.1	12.8%	7.9%	12.7%	$1.2*10^{-12}$	55
Systolic BP	80.0	90.0	18.8%	14.6%	6.9%	$1.8*10^{-42}$	192
Diastolic BP	45.0	50.0	15.5%	5.4%	8.6%	$6.9*10^{-37}$	167
Mean BP	50.0	55.0	16.3%	11.4%	8.4%	$7.8*10^{-18}$	79
Heart rate	105	135	7.8%	12.7%	18.8%	$1.3*10^{-20}$	92
Respir rate	20	24	7.0%	14.0%	23.8%	$3.1*10^{-51}$	233
PO2	120	155	17.5%	29.1%	4.3%	$2.2*10^{-153}$	703
SPO2	66	86	41.4%	21.3%	8.9%	$1.8*10^{-35}$	160
Bicarbonate	16	18	29.7%	17.5%	7.8%	$1.3*10^{-68}$	313
Bilirubin	1.5	7.0	8.9%	16.3%	27.2%	$1.4*10^{-23}$	105
WBC	12.5	19.5	8.1%	15.0%	24.3%	$1.8*10^{-32}$	146
Creatinine	1.3	2.0	7.2%	12.5%	20.0%	$6.6*10^{-40}$	180
Hematocrit	31	49	12.7%	9.3%	18.6%	$9.7*10^{-7}$	28
Platelets	20	34	15.7%	8.9%	14.8%	$8.9*10^{-12}$	51
Sodium	129	143	19.0%	9.3%	18.4%	$6.4*10^{-14}$	61
Potassium	3.4	5.1	12.8%	9.2%	21.7%	$1.6*10^{-16}$	73
BUN	23	49	7.2%	14.0%	22.7%	$5.0*10^{-45}$	204
Foley	6	20	27.7%	15.3%	6.2%	$6.8*10^{-88}$	401
Glucose	82	241	20.9%	8.6%	19.2%	$6.9*10^{-29}$	130
Age	61	73	6.2%	8.4%	14.8%	$2.0*10^{-29}$	132

Table A2. Optimal set of 3 Threshold Values for Continuous Variables.

Variable	c_1	c_2	c_3	%death ($X < c_1$)	%death ($c_1 \leq X < c_2$)	%death ($c_2 \leq X < c_3$)	%death ($X \geq c_3$)	p-value	X^2 statistic
Temperature	36.0	36.1	37.1	12.8%	12.2%	7.4%	12.7%	$2.3*10^{-14}$	67
Systolic BP	80.0	90.0	95.0	18.8%	11.8%	15.1%	6.3%	$5.8*10^{-44}$	204
Diastolic BP	45.0	50.0	55.0	15.5%	4.8%	9.4%	8.5%	$4.7*10^{-38}$	176
Mean BP	40.0	50.0	55.0	20.3%	14.8%	11.4%	8.4%	$6.6*10^{-19}$	88
Heart rate	105	120	135	7.8%	11.7%	14.8%	18.8%	$3.8*10^{-21}$	98
Respir rate	18	20	24	6.1%	9.1%	14.0%	23.8%	$1.1*10^{-52}$	244
PO2	120	125	155	17.5%	61.7%	15.1%	4.3%	$2*10^{-256}$	1184
SPO2	66	86	96	41.4%	21.7%	8.3%	13.1%	$1.1*10^{-37}$	175
Bicarbonate	16	18	20	29.7%	18.1%	12.6%	7.2%	$2.2*10^{-69}$	322
Bilirubin	0.5	1.5	7.0	11.7%	8.3%	16.3%	27.2%	$1.3*10^{-25}$	119
WBC	3.0	12.5	19.5	17.6%	7.8%	15.0%	24.3%	$8.0*10^{-37}$	171
Creatinine	1.3	2.0	8.5	7.2%	12.5%	20.3%	1.1%	$3.4*10^{-45}$	210
Hematocrit	31	37	49	12.7%	10.2%	8.7%	18.6%	$4.5*10^{-7}$	33
Platelets	20	28	34	15.7%	8.4%	10.0%	14.8%	$5.5*10^{-12}$	55
Sodium	129	134	143	19.0%	12.6%	8.7%	18.4%	$4.1*10^{-17}$	79
Potassium	3.4	4.6	5.1	12.8%	8.8%	11.1%	21.7%	$6.7*10^{-17}$	78
BUN	19	23	49	6.5%	10.4%	14.0%	22.7%	$5.1*10^{-47}$	218
Foley	6	20	25	27.7%	14.1%	14.5%	5.6%	$5.6*10^{-90}$	417
Glucose	82	204	241	20.9%	8.2%	11.5%	19.2%	$9.4*10^{-30}$	138
Age	47	61	73	4.9%	6.8%	8.4%	14.8%	$6.1*10^{-29}$	134

Table A3. The Logistic Regression Model Results.

Variable	Metric[1]	Categorical Partition	Score (fitted coefficient)
SAPS11, APACHE 11, SOFA VARIABLES			
ORGAN SYSTEM: NERVOUS			
Eye opening	Worst score (Minimum)	Variable treated as numeric (integers range from 1-4)	-0.41461* (eye opening score)
Motor response		integer values range from 1 to 6	-0.1855 * (motor response score)
Verbal response		integer values range from 1 to 5	-0.1081 * (verbal response score)
ORGAN SYSTEM: RESPIRATORY			
Respiratory rate	Median	*< 18 inspir/min*	0
		18 – 20 inspir/min	+ 0.53167
		20 – 24 inspir/min	+ 0.45007
		≥ 24 inspir/min	+ 0.89472
Heart rate	Maximum	*< 105 or ≥ 135 beats/min*	0
		105 – 120 beats/min	+ 0.11356
		120 – 135 beats/min	+ 0.172
SPO2 levels	Minimum	< 66 %	+ 1.28253
		66 – 86 %	+ 0.55164
		86 – 96 %	0
		≥ 96 %	+ 0.53805
PO2 levels	Median	<120 mmHg	+ 0.93580
		120–125 mmHg	+ 4.23804
		125–155 mmHg	+ 0.58886
		≥155 mmHg	0
ORGAN SYSTEM: LIVER			
Bilirubin	Maximum	*< 0.75 mg/dl*	0
		0.75 – 1 mg/dl	+ 2.89520
		1 – 7.75 mg/dl	+ 0.69242
		≥ 7.75 mg/dl	+ 2.03523
ORGAN SYSTEM: KIDNEY			
Hematocrit	Maximum	*< 49%*	0
		≥ 49%	-0.42810
Foley	Minimum	< 6 mL	+ 0.86566
		6 – 20 mL	+ 0.28029
		20 – 25 mL	+ 0.58092
		≥ 25 mL	0
ORGAN SYSTEM: CARDIOVASCULAR			

1 Baseline window frame is set to be within 24h from admission for temporal variables. Diagnosed diseases variables are historical or diagnosed conditions upon admission. Discharged diagnoses are not included since this is a baseline mortality scoring system.

Variable	Metric[1]	Categorical Partition	Score (fitted coefficient)
Vasopressors		*No*	0
		Yes	+ 0.74001
Systolic BP	Minimum	< 80 mmHg	+ 0.38861
		80 – 90 mmHg	+ 0.32199
		90 – 95 mmHg	+ 0.63199
		≥ 95 mmHg	0
Diastolic BP	Minimum	< 45 mmHg	+ 1.29462
		45 – 50 mmHg	0
		50 – 55 mmHg	+ 0.82218
		≥ 55 mmHg	+ 0.66162
Mean BP	Minimum	< 40	-0.55085
		40 – 50 mmHg	-0.60798
		50 – 55 mmHg	-0.72236
		≥ 55 mmHg	0
ORGAN SYSTEM: BLOOD			
White blood cells	Minimum	< 3.0 ($2.5x10^3/mm^3$)	+ 0.60184
		3.0 – 12.5 ($2.5x10^3/mm^3$)	0
		12.5 – 19.5 ($2.5x10^3/mm^3$)	+ 0.41386
		≥ 19.5 ($2.5x10^3/mm^3$)	+ 0.81632
Sodium	Minimum	< 129 mEq/L	-0.72771
		129 – 134 mEq/L	+ 0.14478
		≥ 134 mEq/L	0
Blood Urea Nitrogen	Minimum	*< 49 mg/dL*	0
		≥ 49 mg/dL	+ 0.49198
Potassium	Median	< 3.4 mEq/L	+ 0.48884
		3.4 – 4.6 mEq/L	-0.15013
		4.6 – 5.1 mEq/L	0
		≥ 5.1 mEq/L	+ 0.40614
Platelets	Minimum	< 20 ($10^3/\mu l$)	+ 0.49081
		20 – 28 ($10^3/\mu l$)	0
		28 – 34 ($10^3/\mu l$)	-0.33701
		≥ 34 ($10^3/\mu l$)	-0.27353
ORGAN SYSTEM: ENDOCRINE			
Glucose	Maximum	*< 204 mg/dL*	0
		204 – 241 mg/dL	+ 0.21752
		≥ 241 mg/dL	+ 0.17071
ORGAN SYSTEM: OTHERS			
Temperature	Minimum	< 36.0 ºC	+ 0.37293
		36.0 – 37.1 ºC	0
		≥ 37.1 ºC	+ 0.45264
Age	Age upon admission	*< 47 years old*	0
		47 – 61 years old	+ 0.65465
		61 – 73 years old	+ 0.98743
		≥ 73 years old	+ 1.82696
Metastatic Cancer		*No*	0
		Yes	+ 0.48827
Admission Location		*Outside ICU*	0
		Emergency Room/ ICU ward	+ 0.61225
Hematologic Malignancy		*No*	0
		Yes	+ 0.79106
Admission Type		*Planned*	0
		Unplanned	+ 1.37918
NEW VARIABLES FROM LITERATURE REVIEW			
ORGAN SYSTEM: OTHERS			
H/O COPD		*No*	0
		Yes	+ 0.85747
ORGAN SYSTEM: CARDIOVASCULAR			
H/O Cardiac Arrest		*No*	0
		Yes	+ 2.13783
ORGAN SYSTEM: RESPIRATORY			
Mechanical Ventilation		*No*	0
		Yes	+ 0.54973
ORGAN SYSTEM: LIVER			
H/O Cirrhosis		*No*	0
		Yes	+ 0.76393

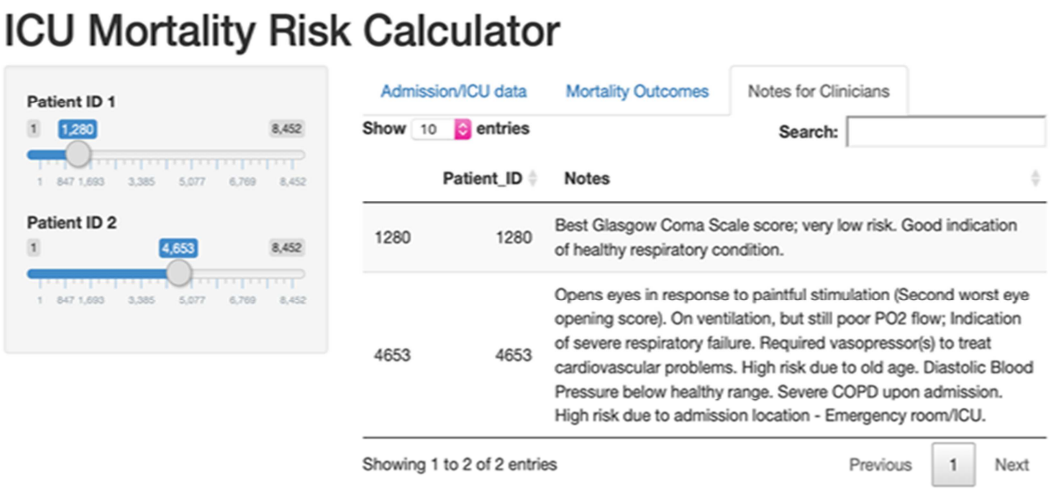

Figure A1. R Shiny ICU Mortality Risk Calculator.

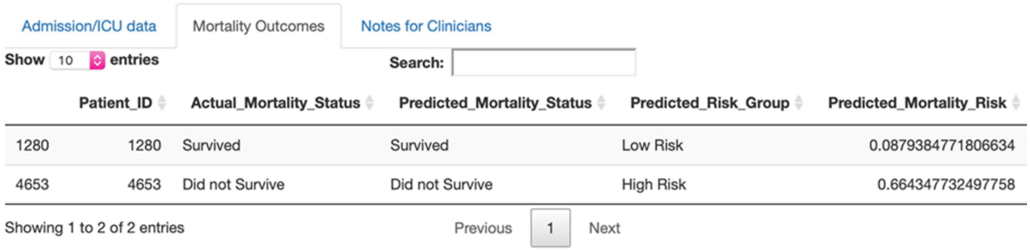

Figure A2. R Shiny Mortality Outcomes.

References

[1] Jiankang Liu, Xian Xiang Chen, Lipeng Fang, et. al. Mortality prediction based on imbalanced high-dimensional ICU big data. Science Direct, Computers in Industry, (2018), 98, pp: 218-225.

[2] Knaus WA, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. Chest 1991; 100(6): 1619-1636.

[3] Zimmeman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluations (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med 2006; 34(5): 1297-1310.

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

[5] Moreno RP, Metnitz PGH, Almeida E, et al. SAPS 3 – from evaluation of the patient to evaluation of the intensive care unit. Intensive Care Med 2005; 31: 1336-1355.

[6] Vincent JL, et al. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. Intensive Care Med 1996; 22:707-710.

[7] Minne L, Abu-Hanna A, de Jonge E. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. Crit Care 2008; 12: R161.

[8] Mayr VD, Dunser MW, Greil V, et. al. Causes of death and determinants of outcome in critically ill patients. BioMed Central Ltd, Critical Care (2006), 10(6): R154.

[9] Orban JC, Walrave Y, Mongardon N, et. al. Causes and Characteristics of Death in Intensive Care Units: A Prospective Multicenter Study. Pubmed (2017), 126(5), pp 882-889.

[10] Nasser Abdullah Alghamdi, Munni Begum. Identification of the Risk Factors Associated with ICU Mortality. Biom Biostat Int J 2017, 6(1): pp. 278-287. DOI: 10.15406/bbij.2017.06.00157.

[11] Martina, M. L. V. Todini, E. A. Libralaon. A Bayesian decision approach to rainfall threshold-based flood warning. Hydrology and Earth System Sciences, 2006, 10:413-426.

[12] Meng Hsuen Hsieh, Meng Ju Hsieh, Chin-Ming Chen, et. al. Scientific Reports 8, Comparison of machine learning models for the prediction of mortality of patients with unplanned extubation in intensive care units, 2018. Article Number:17116.

[13] Meng, C., Sanjay, P., Che, Z., & Liu, Y. Benchmark of Deep Learning Models on Large Healthcare MIMIC Datasets. Journal of Biomedical Informatics (2017).

[14] Jacob, C., Qingqing, M., Jana, L. H., et. al. Ritankar, D. Using electronic health record collected clinical variables to predict medical intensive care unit mortality. Annals of Medicine and Surgery, 2016, Volume 11, Pages 52-57.

[15] Sheth, M. Predicting Mortality for Patients in Critical Care: A Univariate Flagging Approach. MASSACHUSETTS INSTITUTE OF TECHNOLOGY (2015).

[16] Johnson, A. E., Pollard, T. J., Shen, L., et. al. MIMIC-III, a freely accessible critical care database, Sci. Data. 3 (2016).

[17] http://julietthl.shinyapps.io/rshiny