OPEN ACCESS Freely available online

IJCMB
International Journal of Clinical Medicine and Bioengineering

ISSN 2737-534X
Volume 2, Issue 2
https://www.iikii.com.sg/journal/IJCMB
International Journal of Clinical Medicine and Bioengineering

Article

Time Series Multi-task Learning for Prognosis of MICU and SICU

Yen-Jung Chiu [1], Szu-Hsien Wu [2,3,4], Ping-Feng Wu [2,5], Chao-Chun Chuang [1], Ming-Liang Hsiao [6], Mei-Jung Chen [7], Pei-Ru Chen [7], and Shih-Tsang Tang [7,*]

[1] National Center for High-Performance Computing, Taiwan; zx12as3420@gmail.com
[2] Department of Surgery, Taipei Veterans General Hospital, Taiwan; shwu3@vghtpe.gov.tw
[3] Department of Surgery, School of Medicine, National Yang Ming Chiao Tung University, Taiwan
[4] Division of Plastic and Reconstructive Surgery, Department of Surgery, School of Medicine, National Defense Medical Center, Taiwan
[5] Division of Infectious Diseases, Department of Medicine, Taipei Veterans General Hospital, Taiwan; striefcloud@gmail.com
[6] Division of Experiment Surgery, Department of Surgery, Taipei Veterans General Hospital, Taiwan; mlshiao@vghtpe.gov.tw
[7] Department of Biomedical Engineering, Ming Chuan University, Taiwan; mjchen@mail.mcu.edu.tw
* Correspondence: sttang@mail.mcu.edu.tw; Tel.: +886 3 350 7001 #3658

Received: May 10, 2022; **Accepted:** Jun 10, 2022; **Published:** Jun 30, 2022

Abstract: The prognostic assessment of an ICU patient involves assessing the severity of their condition, interventions, and length of ICU stay. Over the past 30 years, researchers have proposed numerous predictive models and severity assessment scales for ICU patients in specific regions, including APACHE II and SAPS II. However, most existing methods rely heavily on curve fitting which do not account for misclassifications caused by false negatives and positives. Specificity and sensitivity must be provided as an indicator of model performance. The primary aim in this study is to develop a machine-learning model to formulate a prognosis for MICU and SICU patients by using data from the MIMIC-IV for training. The predictive models developed in this study facilitate the prediction of mortality and other outcomes across various treatment regimens.

Keywords: MICU, SICU, Prognostic assessment, Predictive model

1. Introduction

The prognostic assessment of an intensive care unit (ICU) patient involves assessing the severity of the condition, interventions, and length of ICU stay [1]. Numerous predictive models and severity assessment scales have been developed for ICU patients to predict mortality based on the baseline values of physiological parameters. Commonly used assessment scales include the Acute Physiology and Chronic Health Evaluation (APACHE) [2], APACHE II [3], Simplified Acute Physiology Score (SAPS) [4], and SAPS II [5]. SAPS II and APACHE II are currently the predictive models most widely used in clinical practice. Despite such efforts to adapt SAPS for specific regions (e.g., Southern and Mediterranean countries versus Central and Western Europe) [6–8], predicted mortality rates are generally overestimated [9–13].

Most of the methods assess the severity of ICU cases that are based on logistic regression models, which impose strict constraints on the relationship between explanatory variables and risk of death. For example, principal term logistic regression is based on a linear and additive relationship between a predetermined transformation of the mean outcome and its predictors. However, the factors related to death in ICU patients are complex, the underlying assumptions are often far from reality, and the predictive power may be greatly reduced in a parametric model with too much deviation. Conversely, if a parametric model used is realistic, it often provides excellent predictions. Therefore, the calibration of the bias of the current severity scale can largely attribute to the underlying statistical model, rather than the variables selected for the model.

Numerous studies have concluded that the use of nonparametric methods to predict mortality in ICU patients is at least as good as, and possibly better than traditional standard logistic regression [14–19]. Note that many of these methods are based on neural networks or data mining. In 2019, Shillan et al. reviewed the literature on the use of machine learning to analyze ICU data, including 258 papers (24% used MIMIC-II/III). They determined the two most common AI methods that were used for neural network models with support vector machine (SVM). Roughly 27% of the articles focused on mortality prediction with sample cohorts of 100–1,000 patients. Only 7% of the articles used independent test data. Most of them used k-fold cross-validation, followed by randomly sampled test data subsets. Most of the studies used the area under the curve (AUC) to present prediction results [20], although AUC depends heavily on curve fitting. The receiver operating characteristic curve AUC cannot account for misclassifications caused by false negatives and positives. Specificity and sensitivity must also be provided as an indicator of model performance.

Therefore, the primary aim of this study was to develop a machine learning-based model by which to perform prognostic assessments of ICU patients based on the data from the Medical Information Mart for Intensive Care IV (MIMIC-IV) [21,22]. Previous studies combined data from MICU (medical ICU) and SICU (surgical ICU) despite fundamental differences in the causes of death in the two facilities (i.e., organ failure in MICU versus infection in SICU). The predictive models developed in this study can be used as a reference by ICU physicians in assessing mortality outcomes for a range of treatments.

2. Materials and Methods

2.1. MIMIC-IV

The MIMIC database is an open-source electronic health record (EHR) dataset (e.g., clinical data, vital signs, and biochemistry) from a large number of ICUs. The MIMIC database is created and managed by the Massachusetts Institute of Technology (MIT) Laboratory. The current version is MIMIC-IV 1.0, covering the period from 2008 to 2019 [21]. The MIMIC-IV v1 includes 43 data tables covering six category modules: CORE, HOSP, ICU, ED, CXR, and NOTE. The CORE module contains patient tracking data (e.g., demographics, hospital admissions, and in-hospital ward transfers). The HOSP module contains data from a variety of hospital departments (e.g., laboratory, pharmaceuticals, and diagnostics). The ED module contains data from emergency departments (e.g., reasons for admission, triage assessment, vital signs, and medicine reconciliation). The CXR module contains imaging data (e.g., chest X-rays). The Note module contains deidentified free-text clinical notes. The ICU module contains ICU-specific data from the Clinical Information System (CIS) (e.g., intravenous medications and ventilator settings). Note that the ICU data are from seven ICU branches, including the CVICU (Cardiac Vascular Intensive Care Unit), CCU (Coronary Care Unit), MICU (Medical Intensive Care Unit), MICU/SICU (Medical/Surgical Intensive Care Unit), Neuro SICU (Neuro Surgical Intensive Care Unit), SICU (Surgical Intensive Care Unit), and TSICU (Trauma SICU). MICU and SICU account for more than half of the total number of patients.

2.2. Target Data Sifting

Target data for model development were collected by using patient identifiers from MICU and SICU facilities. Three MIMIC-IV patient identifiers were used: subject_id, hadm_id, and stay_id. Every patient was assigned a unique subject_id. Patients were assigned one hadm_id for each hospitalization. Patients were also assigned one stay_id for each ICU admission. Thus, any given patient could be tracked via one subject_id and at least one hadm_id or stay_id. Fig. 1 illustrates the process of target data collection, including data related to the patient and data related to physiological measurements. Patient data collection focused on patients who visited the MICU or SICU, which resulted in the following: subject_id (53,150), hadm_id (69,211), and stay-id (76,540). Retaining only the most recent records of hospitalization and ICU admission resulted in an equal number of hadm_id and stay_id (53,150). Hans-Christian and Thorsen-Meyer reported that predictions based on data obtained over 72 h resulted in an AUROC accuracy of 0.7 [20]. Exclusion criteria included patients under the age of 18 and patients whose ICU stay was less than 3 days. Following exclusions, the study group included 16,380 subject_ids.

Physiological measurements included 2,216 entries for the 16,380 subject_ids in the MIMIC-IV table labeled “chartevents”. Removal of non-numerical measurement data resulted in 358 measurements. Our focus on time-series data also led to the exclusion of non-timed data, as shown in Table 1. This resulted in six measurements on MICU and nine measurements on SICU as shown in Table 2. The proposed model generated a prognosis for a period of 7, 14, or 28 days. Thus, we excluded patients whose ICU stay exceeded 30 days. This resulted in a total of 5,713 patients, including 3,087 MICU and 2,626 SICU patients.

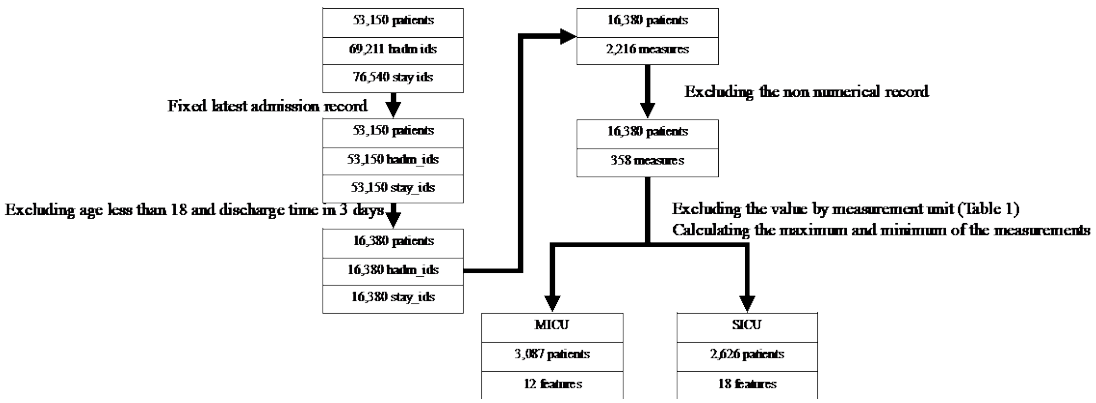

Fig. 1. Process used in sifting through target data.

Table 1. Physiological data discarded according to measurement units.

Unit	Number of Discarded Measurements
cm	1
kg	1
/min	1
kcal/kg	71
g/kg	1
ml/kg	1
hour	1
degree	1
ears	3
inch	4
kcal/day	1
°C	2
°F	9
lbs	3

Table 2. Measurements from the MICU and SICU.

MICU Measurements	SICU Measurements
Heart Rate	Heart Rate
Non-Invasive Blood Pressure, systolic	Non-Invasive Blood Pressure, systolic
Non-Invasive Blood Pressure, diastolic	Non-Invasive Blood Pressure, diastolic
Non-Invasive Blood Pressure, mean	Non-Invasive Blood Pressure, mean
Respiratory Rate	Respiratory Rate
O_2 saturation pulse oximetry	O_2 saturation pulse oximetry
	Arterial Blood Pressure, systolic
	Arterial Blood Pressure, diastolic
	Arterial Blood Pressure, mean

2.3. Development of Time-Series Data

The first step in developing time-series data was to screen out measurements obtained during the previous ICU stay using the timestamps in the tables "chartevents" and "icustays". Note that the above-mentioned data-sifting process resulted in six measurements for MICU and nine measurements for SICU, both of which were compiled over 72 h. The measurements were then listed as hourly observations, which were compiled in 72 tables. The maximum and minimum measurement values were used as features to generate a prognosis, resulting in 12 features for MICU and 18 features for SICU. There was a serious problem concerning missing data for three of the SICU measurements (systolic blood pressure, diastolic blood pressure, and mean arterial blood pressure). Specifically, these measurements were linked to only 36 records with 2 cases of mortality. Thus, these three SICU measurements were excluded, leaving 12 SICU and 12 MICU features.

The problem of missing data, which is commonly encountered in machine learning applications, can be dealt with via data imputation, wherein a synthetic value is assigned for each missing value [23]. For example, missing body height values could be filled in as 170 cm, or missing body temperature values as 36.6°C. However, vital signs can be strongly correlated with each other. Therefore, missing values cannot be filled in arbitrarily (i.e., the correlation between measurements must be taken into account). In the current study, we employed the imputation method KNNImputer in the Python package Scikit-learn [24]. To minimize the skewing effects of the imputed data, we applied an additional criterion wherein only patients with measurement data available for at least 64 hours were included. This reduced the number of MICU patients to 726 and the number of SICU patients to 364. The prognostic outcomes in this study were mortality and length of stay (LOS) in the ICU. In-hospital predictions of mortality were binary (alive or dead). The models proposed in the current study estimated the probability of mortality in 7, 14, or 28 days. LOS values were obtained by using the following attributions from the "admissions" table: "admittime" (admission time), "dischtime" (discharge time), and "deathtime" (time of death).

2.4. Model Development

The proposed model was based on the extended least short-term memory model (LSTM) and a recurrent neural network (RNN) applicable to the analysis of time-series data. We adopted the bidirectional-LSTM model in assessing the prognosis of ICU patients, wherein each output was connected to the two previous vital sign measurements (obtained at intervals of one hour). As for

hyperparameters, the output number of patients matched the input number using batches of 8, 16, and 32 (based on the number of patients). The model developed in this study is shown in Fig. 2, where "t" indicates the length of the ICU stay (measured in hours), the green block is the classification model used for mortality (two outputs), and the orange block is the model for LOS (one output). To reduce the risk of overfitting, the loss function for LOS was multiplied by three during training, and early stop criteria (5 epochs) were imposed. The fully connected layer is connected to outputs in a backpropagation manner. There were two fully connected models from the same LSTM output. The fully connected neural network used the cross-entropy loss function for the mortality model and the mean square error loss function for LOS. The prediction performance of the model was assessed in terms of accuracy, sensitivity, and specificity.

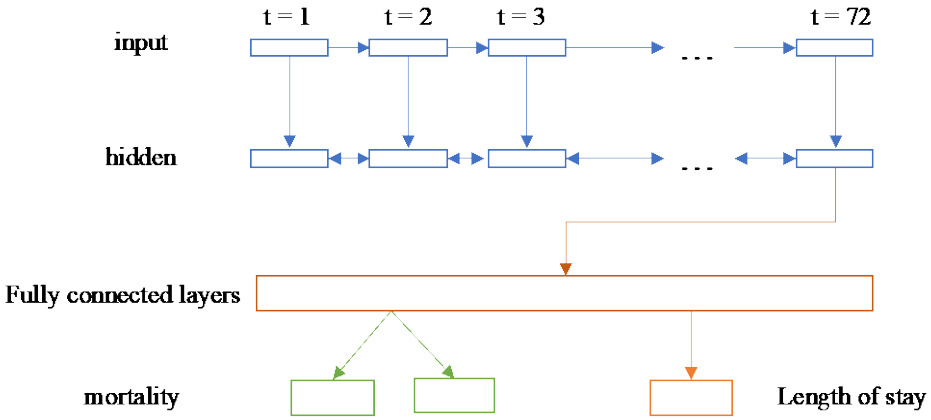

Fig. 2. Proposed prediction model.

3. Results

Model performance was assessed by using 1,000 random samples. Training data included measurements from 90% of the MICU patients and 50% of the SICU patients. The remaining data (10% MICU and 50% SICU) were used for testing. Using an NVIDIA Tesla V100 GPU, roughly 2 min was required for the completion of each epoch.

3.1 Model Prediction

The number of survival cases greatly exceeded the number of deaths, so we based the number of training samples on the number of deaths. Therefore, in the training, the numbers of images were the same from the classes of dead and alive patients. Table 3 lists the sizes of MICU and SICU samples. Using the MICU 7-day sample for training, the number of deaths was calculated as $82 \times 0.9 \cong 73$ (73 deaths and 73 survival cases), resulting in a training set with 146 samples. The testing dataset included 9 deaths ($82 \times 0.1 \cong 9$) and 571 survival cases ($644 - 73 = 571$). Using the SICU 7-day sample for training resulted in 17 deaths ($34 \times 0.5 = 17$) matched with 17 survival cases. The testing dataset included 17 deaths ($34 \times 0.5 = 17$) and 313 survival cases ($330 - 17 = 313$).

Model performance was evaluated based on accuracy, sensitivity, and specificity. Accuracy indicates the degree of correspondence between the actual numbers of death and survival cases versus the predicted values. Sensitivity indicates the proportion of actual deaths that were predicted. Specificity indicates the proportion of actual survival cases that were predicted. Optimization was performed using three batch sizes (8, 16, and 32) and two optimization schemes, including Adam (Adaptive Moment Estimation) and SGD (Stochastic Gradient Descent), with a dropout rate of 0.5. The SGD optimizer results were better suited to SICU whereas the Adam optimizer results were better suited to MICU.

Table 3. Sample sizes for MICU and SICU.

MICU	Death	Alive	Training	Testing-death	Testing-alive
7 days	82	644	146	9	571
14 days	162	564	290	17	419
28 days	212	514	380	22	323
SICU	**Death**	**Alive**	**Training**	**Testing-death**	**Testing-alive**
7 days	34	330	34	17	313
14 days	53	311	52	27	285
28 days	70	294	70	35	259

3.2 Performance Assessment

Tables 4 and 5 list the model training performance for MICU and SICU. Note that the values listed for accuracy, sensitivity, and specificity are medians. In the parenthesis, the first value indicates the 75% quantile and the second value indicates the 25% quantile. For both MICU and SICU model training, a batch size of 32 samples resulted in the best performance. Decreasing the batch size was shown to undermine performance and led to overfitting. Even if the model's predictions of LOS were not significantly different between batch sizes and death identified (Tables 6 and 7), which also outperformed the model without using LOS loss. Both models achieved the best accuracy, sensitivity, and specificity and the lowest interquartile ranges when using data of 28 days.

Table 4. Performance of LSTM model for MICU.

Batch size	Death Identified	Accuracy	Sensitivity	Specificity
32	7 days	0.62 (0.79, 0.33)	0.63 (0.80, 0.32)	0.67 (0.89, 0.33)
32	14 days	0.71 (0.79, 0.59)	0.72 (0.81, 0.59)	0.52 (0.71, 0.24)
32	28 days	0.70 (0.74, 0.62)	0.71 (0.75, 0.62)	0.63 (0.68, 0.50)
16	7 days	0.56 (0.84, 0.45)	0.55 (0.85, 0.44)	0.67 (0.78, 0.31)
16	14 days	0.53 (0.74, 0.33)	0.52 (0.76, 0.31)	0.7 (0.88, 0.45)
16	28 days	0.64 (0.74, 0.6)	0.64 (0.75, 0.6)	0.59 (0.68, 0.53)
8	7 days	0.02 (0.84, 0.02)	0 (0.85, 0)	1 (1, 0)
8	14 days	0.63 (0.92, 0.37)	0.63 (0.95, 0.35)	0.58 (0.82, 0.16)
8	28 days	0.57 (0.72, 0.43)	0.56 (0.72, 0.41)	0.72 (0.78, 0.53)

Table 5. Performance of LSTM model for SICU.

Batch size	Death Identified	Accuracy	Sensitivity	Specificity
32	7 days	0.38 (0.96, 0.02)	0.37 (0.98, 0.01)	0.75 (1.0, 0)
32	14 days	0.37 (0.74, 0.04)	0.36 (0.75, 0.02)	0.75 (1, 0.3)
32	28 days	0.69 (0.88, 0.27)	0.71 (0.9, 0.25)	0.43 (0.86, 0.14)
16	7 days	0.53 (0.92, 0.13)	0.53 (0.94, 0.11)	0.75 (1.0, 0)
16	14 days	0.31 (0.98, 0.05)	0.3 (1.0, 0.04)	0.59 (1.0, 0)
16	28 days	0.62 (0.96, 0.06)	0.64 (0.99, 0.04)	0.21 (0.90, 0.21)
8	7 days	0.59 (0.99, 0.12)	0.59 (1.0, 0.11)	0.25 (1.0, 0)
8	14 days	0.36 (0.8, 0.02)	0.36 (0.8, 0)	0.59 (1, 0.17)
8	28 days	0.34 (0.94, 0.03)	0.34 (0.97, 0)	0.64 (1, 0.14)

Table 6. Mean square error between ground truth values and predicted length of stay in MICU.

Batch size	Death Identified	Adam	SGD
32	7 days	0.11 (0.09, 0.11)	0.17 (0.15, 0.17)
32	14 days	0.09 (0.08, 0.11)	0.15 (0.13, 0.16)
32	28 days	0.06 (0.05, 0.08)	0.13 (0.1, 0.14)
16	7 days	0.14 (0.08, 0.14)	0.15 (0.13, 0.17)
16	14 days	0.08 (0.05, 0.12)	0.15 (0.11, 0.18)
16	28 days	0.08 (0.08, 0.09)	0.14 (0.13, 0.15)
8	7 days	0.12 (0.08, 0.15)	0.16 (0.12, 0.17)
8	14 days	0.09 (0.05, 0.16)	0.17 (0.11, 0.17)
8	28 days	0.05 (0.04, 0.06)	0.16 (0.13, 0.17)

Table 7. Mean square error between ground truth values and predicted length of stay in SICU

Batch size	Death Identified	Adam	SGD
32	7 days	0.14 (0.1, 0.15)	0.17 (0.14, 0.18)
32	14 days	0.1 (0.1, 0.12)	0.17 (0.14, 0.18)
32	28 days	0.1 (0.09, 0.11)	0.16 (0.12, 0.16)
16	7 days	0.13 (0.8, 0.15)	0.17 (0.16, 0.18)
16	14 days	0.12 (0.11, 0.14)	0.15 (0.11, 0.17)
16	28 days	0.1 (0.08, 0.11)	0.15 (0.14, 0.17)
8	7 days	0.12 (0.09, 0.16)	0.15 (0.13, 0.17)
8	14 days	0.1 (0.07, 0.13)	0.14 (0.1, 0.17)
8	28 days	0.06 (0.04, 0.14)	0.17 (0.12, 0.17)

4. Discussion and Conclusions

We developed LSTM models for generating prognoses in the MICU or SICU in terms of mortality and LOS. Experiment data were extracted from the MIMIC-IV datasets. Model training was performed by using data obtained in the ICU over 72 h. A serious problem was encountered related to missing data (hourly measurements of most vital signs), which resulted in the need for data imputation and the exclusion of most patients (only patients for whom data were available for at least 64 hours were included). Thus, the main factor affecting the performance of the proposed model was the issue of missing data. We found that increasing the batch size and the length of the prediction period increased accuracy, wherein the optimal mortality prediction performance was achieved with a batch size of 32 patients and a prediction window of 28 days. This appears reasonable considering that many patients will die within 28 days despite surviving for 7 days.

The source of training data in this study was MIMIC-IV version 1.0, published in 2021. Data quality in this version has been greatly improved over previous versions. In previous research based on the MIMIC III dataset, Sheikhalishahi (2020) used 48-hour data to predict ICU mortality. Their models achieved high median sensitivity (~0.9) but low median specificity (~0.5), indicating underfitting [23]. Barbieri (2020) compared neural networks and RNN models (single or mixed) [25], attention models, ordinary differential equations, and medical concept embeddings. Using ICD codes, prescription data, and vital signs as prediction features, RNN with an attention model provided the best prediction results for 30-day ICU mortality with a mean sensitivity of 0.7 and a mean specificity of 0.66. The proposed model proved to be more effective than previous methods in terms of overfitting or underfitting, resulting in superior median specificity (e.g., ~0.63 in MICU). The prediction results were on par with those obtained using categorical data analysis and RNN models.

Author Contributions: The authors confirm contributions to the paper as follows: conceptualization, S.-H. Wu and P.-F. Wu; methodology, Y.-J. Chiu; software, Y.-J. Chiu; validation, S.-H. Wu, P.-F Wu, and M.-L. Hsiao; formal analysis, Y.-J. Chiu; investigation, Y.-J. Chiu; resources, C.-C. Chuang; data curation, Y.-J. Chiu; writing—original draft preparation, Y.-J. Chiu; writing—review and editing, M.-J. Chen and P.-R. Chen; visualization, Y.-J. Chiu; supervision, S.-T. Tang. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported in part by the Taiwan National Science and Technology Council under MOST 111-2221-E-130-004.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fernando, S.M.; Tran, A.; Taljaard, M.; Cheng, W.; Rochwerg, B.; Seely, A. J. E.; Perry, J. J. Prognostic accuracy of the quick sequential organ failure assessment for mortality in patients with suspected infection: a systematic review and meta-analysis. *Annals of Internal Medicine* **2018**, *168*(4), 266–275.
2. Knaus, W.A.; Zimmerman, J.E.; Wagner, D.P.; Draper, E.A.; Lawrence, D.E. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Critical Care Medicine* **1981**, *9*(8), 591–597.
3. Knaus, W.A.; Draper, E.A.; Wagner, D.P.; Zimmerman, J.E. APACHE II: a severity of disease classification system. *Critical Care Medicine* **1985**, *13*(10), 818–829.
4. Le Gall, J.R.; Loirat, P.; Alperovitch, A.; Glaser, P.; Granthil, C.; Mathieu, D.; Mercier, P.; Thomas, R.; Villers, D. A simplified acute physiology score for ICU patients. *Critical Care Medicine* **1984**, *12*(11), 975–977.

5. Le Gall, J.R.; Lemeshow, S.; Saulnier, F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA* **1993**, *270*(24), 2957–2963.
6. Le Gall, J.R.; Neumann, A.; Hemery, F.; Bleriot, J. P.; Fulgencio, J.P.; Garrigues, B.; Gouzes, C.; Lepage, E.; Moine, P.; Villers, D. Mortality prediction using SAPS II: an update for French intensive care units. *Critical Care* **2005**, *9*(6), R645–R652.
7. Metnitz, P.G.; Valentin, A.; Vesely, H.; Alberti, C.; Lang, T.; Lenz, K.; Steltzer, H.; Hiesmayr, M. Prognostic performance and customization of the SAPS II: results of a multicenter Austrian study. Simplified acute physiology score. *Intensive Care Medicine* **1999**, *25*(2), 192–197.
8. Moreno, R.P.; Metnitz, P.G.; Metnitz, B.; Bauer, P.; Afonso de Carvalho, S.; Hoechtl, A.; Investigators, S. Modeling in-hospital patient survival during the first 28 days after intensive care unit admission: a prognostic model for clinical trials in general critically ill patients. *Journal of Critical Care* **2008**, *23*(3), 339–348.
9. Arabi, Y.; Al Shirawi, N.; Memish, Z.; Venkatesh, S.; Al-Shimemeri, A. Assessment of six mortality prediction models in patients admitted with severe sepsis and septic shock to the intensive care unit: a prospective cohort study. *Critical Care* **2003**, *7*(5), R116–122.
10. Bertolini, G.; D'Amico, R.; Apolone, G.; Cattaneo, A.; Ravizza, A.; Iapichino, G.; Brazzi, L.; Melotti, R.M. Predicting outcome in the intensive care unit using scoring systems: is new better? A comparison of SAPS and SAPS II in a cohort of 1,393 patients. GiViTi Investigators (Gruppo Italiano per la Valutazione degli interventi in Terapia Intensiva). Simplified Acute Physiology Score. *Medical Care* **1998**, *36*(9), 1371–1382.
11. Metnitz, P.G.; Lang, T.; Vesely, H.; Valentin, A.; Le Gall, J.R. Ratios of observed to expected mortality are affected by differences in case mix and quality of care. *Intensive Care Medicine* **2000**, *26*(10), 1466–1472.
12. Soares, M.; Salluh, J.I. Validation of the SAPS 3 admission prognostic model in patients with cancer in need of intensive care. *Intensive Care Medicine* **2006**, *32*(11), 1839–1844.
13. Sakr, Y.; Marques, J.; Mortsch, S.; Gonsalves, M.D.; Hekmat, K.; Kabisch, B.; Kohl, M.; Reinhart, K. Is the SAPS II score valid in surgical intensive care unit patients? *Journal of Evaluation in Clinical Practice* **2012**, *18*(2), 231–237.
14. Wise, E.S.; Hocking, K.M.; Brophy, C.M. Prediction of in-hospital mortality after ruptured abdominal aortic aneurysm repair using an artificial neural network. *Journal of Vascular Surgery* **2015**, *62*(1), 8–15.
15. Zimmerman, J.E.; Kramer, A.A.; McNair, D.S.; Malila, F.M. Acute physiology and chronic health evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Critical Care Medicine* **2006**, *34*(5), 1297–1310.
16. Ribas Ripoll, V.J.; Vellido, A.; Romero, E.; Ruiz-Rodriguez, J.C. Sepsis mortality prediction with the quotient basis kernel. *Artificial Intelligence in Medicine* **2014**, *61*(1), 45–52.
17. Hur, S.; Min, J.Y.; Yoo, J.; Kim, K.; Chung, C.R.; Dykes, P.C.; Cha, W.C. Development and validation of unplanned extubation prediction models using intensive care unit data: retrospective, comparative, machine learning study. *Journal of Medical Internet Research* **2021**, *23*(8), e23508.
18. Foltran, F.; Baldi, I.; Bertolini, G.; Merletti, F.; Gregori, D. Monitoring the performance of intensive care units using the variable life-adjusted display: a simulation study to explore its applicability and efficiency. *Journal of Evaluation in Clinical Practice* **2009**, *15*(3), 506–513.
19. Abosaif, N.Y.; Tolba, Y.A.; Heap, M.; Russell, J.; El Nahas, A.M. The outcome of acute renal failure in the intensive care unit according to RIFLE: model application, sensitivity, and predictability. *American Journal of Kidney Diseases* **2005**, *46*(6), 1038–1048.
20. Thorsen-Meyer, H.C.; Nielsen, A.B.; Nielsen, A.P.; Kaas-Hansen, B.S.; Toft, P.; Schierbeck, J.; Strom, T.; Chmura, P.J.; Heimann, M.; Dybdahl, L.; Spangsege, L.; Hulsen, P.; Belling, K.; Brunak, S.; Perner, A. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. *Lancet Digital Health* **2020**, *2*(4), e179–e191.
21. MIMIC-IV. Available online: https://physionet.org/content/mimiciv/2.0/. (Accessed on October 26, 2022)
22. Goldberger, A.; Amaral, L.; Glass, L.; Hausdorff, J.; Ivanov, P.C.; Mark, R.; Mietus, J.E.; Moody G.B.; Peng C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*(23), e215–e220.
23. Sheikhalishahi, S.; Balaraman, V.; Osmani, V. Benchmarking machine learning models on multi-centre eICU critical care dataset. *PLoS ONE* **2020**, *15*(7), e0235424.
24. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
25. Barbieri, S.; Kemp, J.; Perez-Concha, O.; Kotwal, S.; Gallagher, M.; Ritchie, A.; Jorm, L. Benchmarking Deep Learning Architectures for Predicting Readmission to the ICU and Describing Patients-at-Risk. *Scientific Reports* **2020**, *10*(1), 1111.