TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 03/16 pp639–648
DOI: 10.26599/TST.2022.9010027
Volume 28, Number 4, August 2023

Use of Deep Learning for Continuous Prediction of Mortality for All Admissions in Intensive Care Units

Guangjian Zeng, Jinhu Zhuang, Haofan Huang, Mu Tian, Yi Gao, Yong Liu*, and Xiaxia Yu*

Abstract: The mortality rate in the intensive care unit (ICU) is a key metric of hospital clinical quality. To enhance hospital performance, many methods have been proposed for the stratification of patients' different risk categories, such as severity scoring systems and machine learning models. However, these methods make capturing time sequence information difficult, posing challenges to the continuous assessment of a patient's severity during their hospital stay. Therefore, we built a predictive model that can make predictions throughout the patient's stay and obtain the patient's risk of death in real time. Our proposed model performed much better than other machine learning methods, including logistic regression, random forest, and XGBoost, in a full set of performance evaluation processes. Thus, the proposed model can support physicians' decisions by allowing them to pay more attention to high-risk patients and anticipate potential complications to reduce ICU mortality.

Key words: deep learning; representation learning; mortality; risk prediction; critical care

1 Introduction

Patients in the intensive care unit (ICU) tend to have life-threatening conditions or the potential to develop one during their ICU stay. Therefore, early recognition of their illnesses' changes in severity is invaluable in helping them recover from life-threatening injuries and illnesses[1] and stabilizing their condition. Early and reliable prediction tools for sensitive medical conditions are useful caregiving aids.

Outcome prediction models are one of the prognostic tools for estimating the probability of a pre-specified outcome[2]. In-hospital mortality is the most important outcome in the ICU[3], thus making mortality prediction a crucial task[4]. Statistics indicate that about 11% of deaths are due to the failure to identify patients at risk of deterioration[5]. Mortality prediction aims to identify high-risk people, make the right decisions, and save ICU beds for patients in need[6]. An accurate mortality prediction model provides a stratification of the risk of an outcome at the population level[7]. These models generally provide a numerical estimate of that risk based on estimates from previous populations[8].

In the past, rule-based severity scoring systems were developed based on experts' experiences[9–14]. Later, the rapid development of artificial intelligence techniques and their applications in health care has resulted in the development of machine learning models to achieve the same goals[15–18]. However, previous reports of the continuous application of static scoring systems[19–25] demonstrate a long-standing desire to have a continuously updated patient assessment system. Automated, continuous assessment of the severity of a patient's illness can provide decision support and alert clinicians to a patient's changing status during intensive care. Moreover, the availability of temporal trends in the ICU offers the opportunity to apply a time sequence to state-of-the-art artificial intelligence

• Guangjian Zeng, Jinhu Zhuang, Haofan Huang, Mu Tian, Yi Gao, and Xiaxia Yu are with the School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China. E-mail: {2017222043, 2017222059, huanghaofan2016, mtian, gaoyi}@email.szu.edu.cn, xiaxiayu@szu.edu.cn.

• Yong Liu is with Shenzhen Hospital, Southern Medical University, Shenzhen 518060, China. E-mail: liuyongjoy@outlook.com.

∗ To whom correspondence should be addressed.
Manuscript received: 2022-01-09; revised: 2022-05-12; accepted: 2022-07-05

methods beyond existing conventional models[26–28]. Therefore, a critical step is to develop predictive models that can make continuously updated and accurate predictions of the mortality of all ICU patients to facilitate the administration of preventive treatments.

Deep learning techniques are commonly used in medical applications and perform well in the classification, prediction, and retrieval domains[29–32]. Recurrent neural networks (RNNs) based on long short-term memory (LSTM) were originally proposed in other literature[33]. These networks can learn sequences of observations, thus causing the model to be well suited for time series applications. The recurrent architecture allows the integration of information from previous timesteps with newly acquired data to update its risk assessment, making the model dynamic instead of static. RNNs analyze all available data with neither preconceptions about which measures may be important in determining a patient's clinical status nor the need to engineer features specific to a given clinical condition[34]. Previous work has demonstrated that RNNs are robust when using high-dimensional inputs that may include extraneous features for predicting a range of clinical outcomes[35]. The flexibility and accuracy of RNNs have made them increasingly popular for the predictive modeling of many time-based clinical tasks[35–41].

Therefore, a predictive model was developed based on RNN in this study to continuously predict all-cause mortality in ICU settings based on the framework proposed by Tomašev et al.[42], but with several improvements: (1) A dynamic window was used to generate sequential model input; (2) a resample operation was introduced to improve label balance; and (3) a set of model evaluation methods was performed to fully validate the model performance.

2 Method

2.1 Study population

This study was a cohort retrospective study based on the Medical Information Mart for Intensive Care III (MIMIC-III) databases[43]. MIMIC-III is a large single-center database that contains information about ICU admissions from major tertiary care hospitals. In this study, all ICU admissions were considered except those that belong to four exclusive criteria: (1) Laboratory measurement was not performed during the ICU stay; (2) the patients had never been in the ICU; (3) the patient's survival information was missing; and (4) the results of the laboratory measurement were not numerical. We obtained the mortality labels from the mortality information in hospitalization records. Finally, 46 467 patients were included, with a total of 334 722 encounters. The flowchart of data processing is shown in Fig. 1.

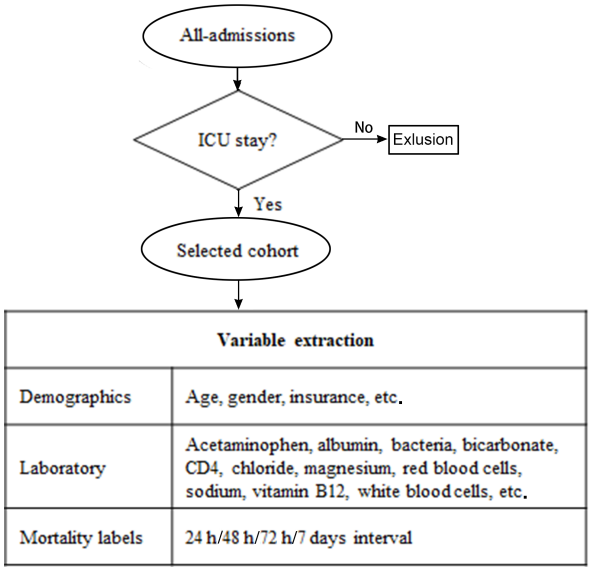
2.2 Data pre-processing

2.2.1 Data extraction

Data were extracted from the MIMIC-III database. The assessed variables included demographic information and laboratory measurements of patients; more details about the variables are shown in Fig. 1.

2.2.2 Patient sequence creation

A fixed time window was employed to generate a



Variable extraction	
Demographics	Age, gender, insurance, etc.
Laboratory	Acetaminophen, albumin, bacteria, bicarbonate, CD4, chloride, magnesium, red blood cells, sodium, vitamin B12, white blood cells, etc.
Mortality labels	24 h/48 h/72 h/7 days interval

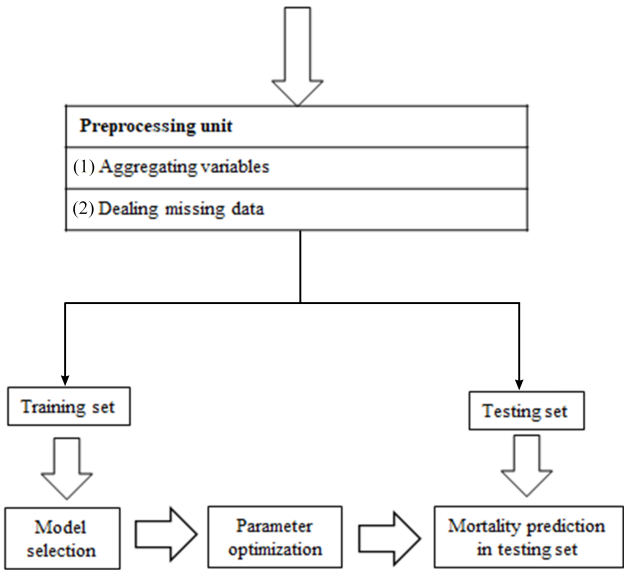


Fig. 1 Flowchart of data extraction.

sequential representation of patient features. Each data point extracted from the database had a unique timestamp. With the uneven gaps between consecutive events, we adopted a fixed time window, i.e., 24 h, to aggregate each variable within this window. A clinical indicator may have been measured multiple times during the time window, which is why we used the moving average as the aggregation method. The original data were then transformed into sequences that share the same time steps, as shown in Fig. 2. By constructing patient sequences, we can achieve continuous prediction, which is an improvement over traditional static prediction.

2.2.3 Outlier filtering

The lower quartile point (Q1), upper quartile point (Q3), and interquartile range (IQR) of each feature were computed. Features beyond the upper/lower limit (defined by Q3 + 3IQR/Q1 − 3IQR) were clipped by the corresponding boundary values.

2.2.4 Feature normalization

The z-score was utilized to normalize numerical features. The features were projected into the range of -1 to 1 using the z-score method, ensuring that all features were at the same magnitude to improve the convergence speed and accuracy of the model.

2.2.5 Missing value

Zero imputation and an additional indicator were introduced to handle null values. Not all checks were completed in each time step, which is why many null values existed in the sequential representation of patients. Thus, the sequential representation was transformed into a feature vector by special imputation. As normalization was previously performed, zero imputation was selected, though many other imputation strategies could be considered. To distinguish the original zero value from the interpolated zero value, additional binary indicator variables were used to supplement the null value's information. The indicator was set to 1 when its corresponding value was not null.

2.2.6 Sequence generation with dynamic window

We improved the sequence input problem with a dynamic window. Conventional methods input the whole patient sequence into the predictive model[42]. However, their solutions do not scale up as the system sees through longer sequences. In addition, large variations in the length of stay create challenges for algorithm robustness. To generate near real time inferences with constant complexity, we used a moving window of 7 d to cut off sequential input and feed it into our predictive model. Figure 2 shows how this "dynamic window" mechanism works.

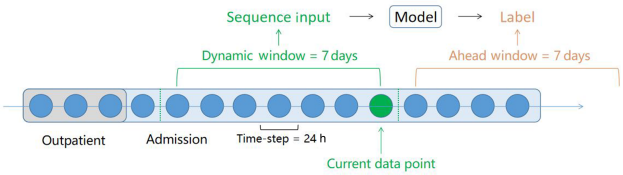


Fig. 2 Patient time sequence for continuous prediction.

2.2.7 Labeling

Fundamentally, our model takes the inputs from the "past" to predict the "future". During training, an "ahead window" at each time step is adopted to generate the label. If the patient died within the "ahead window", we mark the current time step as positive. The model trained in this manner would then predict the likelihood of mortality within the "ahead window" from each time step. In this paper, we used 48 h, 72 h, and 7 d as three options for the "ahead window", as shown in Fig. 2.

2.2.8 Dataset splitting

We use 80% of the entire dataset for training, 10% for internal validation and the remaining 10% for testing.

2.2.9 Resampling

To mitigate the imbalance between positive and negative labels, we perform downsampling with bootstrap to balance the dataset while preserving original distributions. The validation and test sets were kept unchanged.

2.3 Prediction algorithm

Our RNN-based architecture is shown in Fig. 3. The whole model consists of three components: Signal encoding and fusion layers, deep sequence network, and classifier layer for prediction.

2.3.1 Encoding and fusion layer

The input of the whole model consists of two parts: the numerical features $N \in \mathbb{R}^{k\times l} = \{n_i|i = 1, 2, ..., l\}$, where $\mathbb{R}$ is the set of real numbers, and the categorical features $C \in \mathbb{R}^{m\times l} = \{c_i|i = 1, 2, ..., l\}$, where n_i is a numerical vector of size k in time T_i on the sequence, c_i is a binary vector of size m in time T_i on the sequence, and l is the length of sequence. Many zero values exist in the features because of the zero-interpolation operation, thereby making proper encoding of the sparsity necessary. After N and C pass through their independent encoding layers, the encoded features $E_1 \in \mathbb{R}^{e\times l}$ and $E_2 \in \mathbb{R}^{e\times l}$, where e is the encoding size will

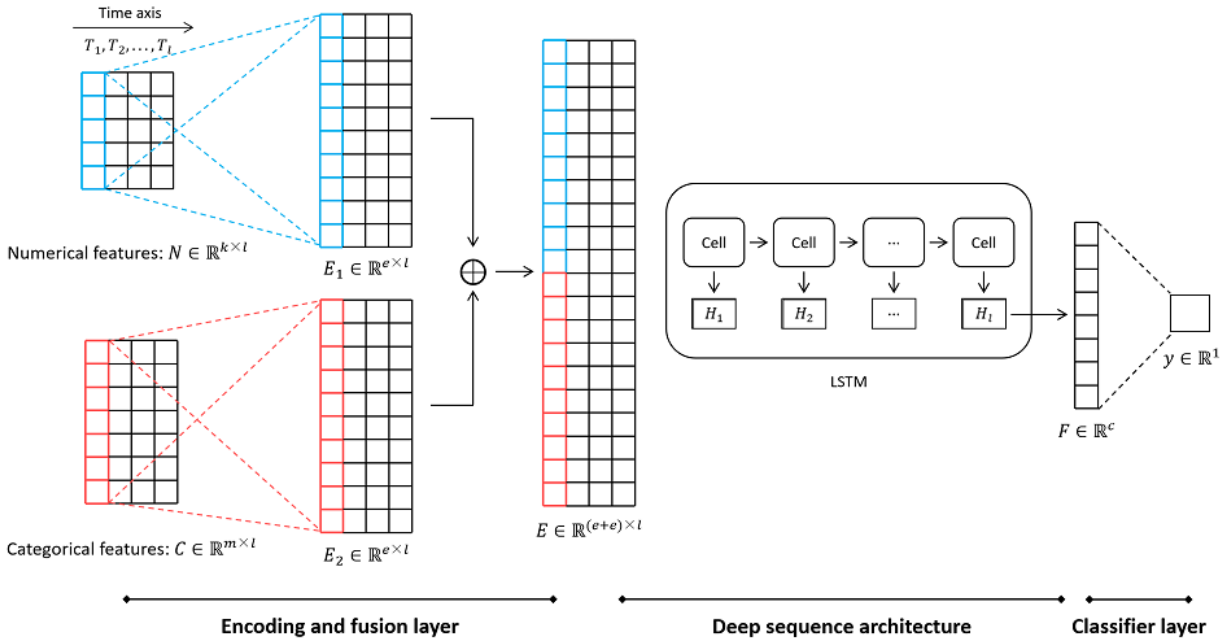


Fig. 3 Model structure for continuous prediction.

be fused by concatenation into $E \in \mathbb{R}^{(e+e)\times l}$ to the following deep recurrent architecture.

2.3.2 Deep sequence architecture

RNN is used to extract sequential information from E. For sequence data, RNN can capture the time information that exists between the data. At the same time, it can handle variable-length sequences and build an internal memory that keeps track of relevant information seen up to that point. Many different RNN structures are considered; LSTM is used in this work as shown in Fig. 3. The encoded vector $e_i \in E$, where $1 \leqslant i \leqslant l$ is sent into the LSTM cell to generate hidden state H_i in chronological order while H_i will be reentered into the cell to help generate H_{i+1}. Finally, the hidden layer vector H_l becomes the input of the last layer; $F \in \mathbb{R}^c$, where c is the size of the LSTM cell.

2.3.3 Classifier layer

Fully connected layers are used to achieve the final prediction $y \in \mathbb{R}^1$. To map the output of multiple neurons to the range of 0 to 1, which can be understood as a probability for multi-classification, the softmax activation function is introduced.

2.3.4 Hyperparameter sweep

We improved the time-consuming and labor-intensive problem of manual parameter adjustment in the past through the hyperparameter sweeps. Hyperparameter sweeps are defined based on domain knowledge and previous literature after model-building to confirm the best model. Table 1 shows the hyperparameter and corresponding value considered in this study. Different hyperparametric values are combined as a group, and the best group is selected.

Table 1 Hyperparameter sweep used in prediction.

Hyperparameter	Values considered
Embedding size	200, 250, 300, 400, 500
Embedding dropout	0.3, 0.5, 0.7
Embedding residual connect	True, false
Embedding activatione	Relu, tanh, leaky relu, hardtanh, elu, sigmoid
RNN cell size	100, 150, 200, 250, 300, 400, 500
Number of RNN layers	1, 2, 3
Batch size	32, 64, 128, 256, 512
Number of epochs	150, 200, 250
Learning rate	0.01, 0.001, 0.0001, 0.000 01
Number of schedule epochs	10, 20, 30, 40
Schedule gamma	0.7, 0.8, 0.85, 0.9, 0.95

2.3.5 Optimization

The loss between the prediction and the target is calculated by the cross-entropy loss function. The Adam optimizer performs gradient descent based on this loss. Variable training steps can be set in the training schedule to ensure the best result. A suitable training schedule is also defined in hyperparameter sweep.

2.4 Performance evaluation

2.4.1 Cross-validation and performance matrix

Fivefold cross-validation was performed. The area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), accuracy, and F-measure (F1) were calculated as the items in the performance matrix.

2.4.2 Discrimination

The corresponding discrimination slopes were demonstrated to present the differences between the predicted risks in survivors and non-survivors.

2.4.3 Reclassification

The categorical net reclassification index and reclassification table were used to measure reclassification for the various methods.

2.4.4 Clinical usefulness

Decision curve analysis (DCA) was performed to estimate the clinical usefulness and net benefit of the proposed prediction models.

3 Result

3.1 Model building and validation

The newly proposed models performed much better than the baseline algorithms did, with an accuracy of 0.92, an F1-score of 0.45, an AUROC of 0.90, and an AUPRC of 0.45 in the 24-h mortality task; an accuracy of 0.91, an F1-score of 0.52, an AUORC of 0.93, and AUPRC of 0.52 in the 48-h mortality task; and an accuracy of 0.93, an F1-score of 0.67, an AUROC of 0.96, and an AUPRC of 0.67 in the 7-day mortality task (Table 2).

3.2 Discrimination

The whole population was grouped into five bins according to the predicted risk scores to determine the discrimination ability of our predicted results. The observed mortality rate was calculated. As demonstrated in Table 3, a low relative rate of patient mortality corresponds to a low predicted risk score, especially in patients with risk scores greater than 80% or lower than 20%. These results suggest that our model can successfully identify survivors and non-survivors, especially among patients with high (0.8–1.0) and low (<0.2) risk scores.

Table 2 Model performance for mortality prediction in validation.

Task	Method	Accuracy	F1-score	AUROC	AUPRC
48-h mortality	LR	0.75	0.23	0.81	0.20
	RF	0.79	0.25	0.84	0.18
	XGBoost	0.80	0.27	0.86	0.28
	Our method	**0.92**	**0.45**	**0.90**	**0.45**
72-h mortality	LR	0.71	0.25	0.81	0.23
	RF	0.77	0.27	0.83	0.20
	XGBoost	0.80	0.30	0.86	0.31
	Our method	**0.91**	**0.52**	**0.93**	**0.52**
7-day mortality	LR	0.70	0.31	0.82	0.31
	RF	0.78	0.36	0.83	0.27
	XGBoost	0.81	0.40	0.87	0.41
	Our method	**0.93**	**0.67**	**0.96**	**0.67**

Table 3 Relative risk ratios for various risk groups with our method.

Risk	Number of patients	48-h mortality		72-h mortality		7-day mortality	
		Mortality (%)	Relative ratio	Mortality (%)	Relative ratio	Mortality (%)	Relative ratio
>80%	6360	45.78	4.9	27.85	4.2	23.6	4.2
60%–79%	6694	3.38	0.4	5.62	0.9	4.27	0.8
40%–59%	6694	0.28	0.0	0.75	0.1	1.14	0.2
20%–39%	6695	0.09	0.0	0.13	0.0	0.19	0.0
<19%	7029	0.00	0.0	0.03	0.0	0.04	0.0
Total	33 472	9.45	–	6.60	–	5.61	–

3.3 Reclassification

The net reclassification index (NRI), which was devised to overcome the limitations of usual discrimination measures, was computed to compare our proposed algorithm with XGBoost,which had the best performance among the traditional machine learning methods. The reclassification tables that involve our method's score and the XGBoost scores are provided in Table 4. Our method's score proposals resulted in the reclassification of a large proportion of patients.

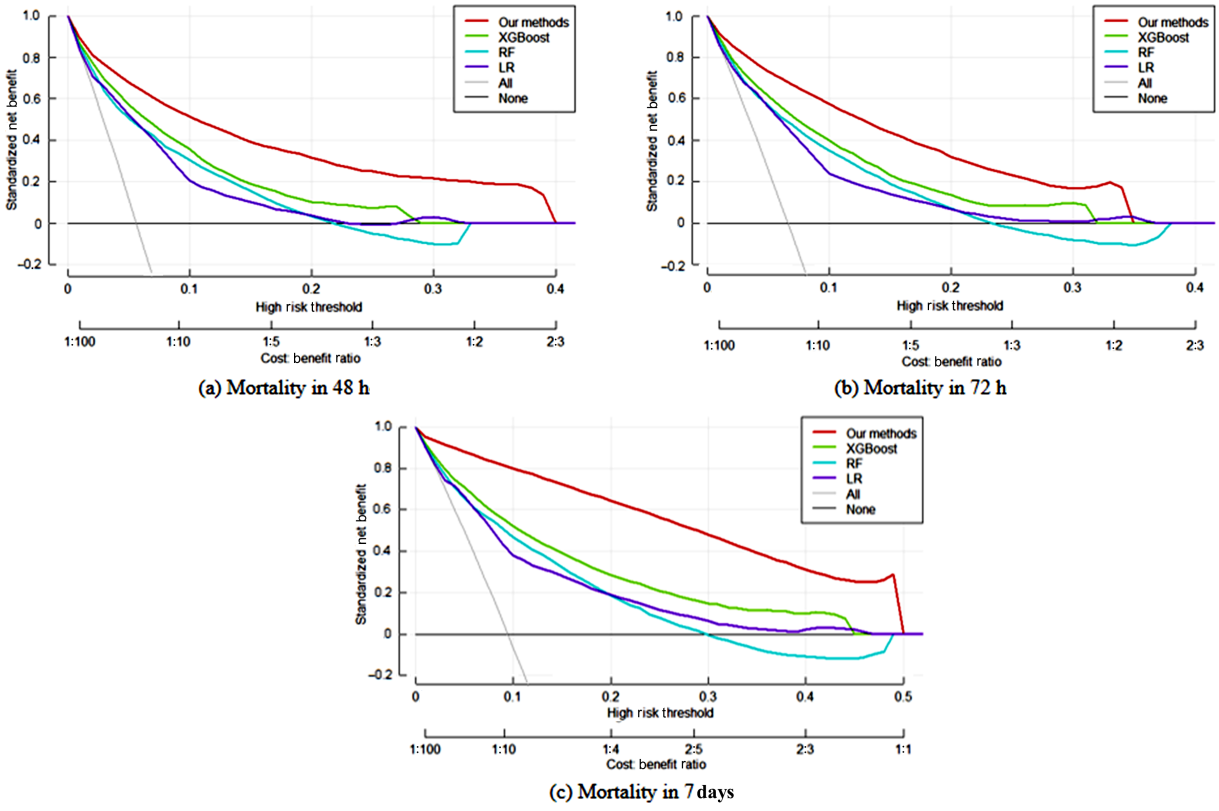
3.4 Decision curve analysis

DCA was performed to estimate the clinical usefulness and net benefit of the intervention[44]. The decision curve[45] was grounded in a decision-theoretical framework that accounted for both the benefits and the costs of intervention to a patient who could not benefit from it. As shown in Fig. 4, the standardized net benefit yielded by the model developed in this study is larger across the major high-risk ranges compared with other models.

Table 4 Reclassification tables between our method and XGBoost.

Task	Updated risk	Reclassification of initial risk				Reclassified ratio (%)	Statistics	
		<25%	25%–50%	50%–75%	>75%		NRI (95 CI)	*P*
48-h mortality	<25%	15 347	1409	696	391	14	0.206 (0.178, 0.235)	<0.001
	25%–50%	2503	909	668	517	80		
	50%–75%	1993	814	787	923	83		
	>75%	1691	838	1055	2931	55		
72-h mortality	<25%	15 439	588	418	786	10	0.328 (0.302, 0.354)	<0.001
	25%–50%	3085	439	369	891	91		
	50%–75%	2668	468	467	1387	91		
	>75%	1942	460	624	3442	47		
7-day mortality	<25%	15 118	300	206	678	7	0.566 (0.546, 0.587)	<0.001
	25%–50%	3989	274	231	922	95		
	50%–75%	3620	311	261	1659	96		
	>75%	2065	275	277	1286	44		

Note: Updated risk, predicted probability according to our method. Initial risk, predicted probability according to XGBoost.



(a) Mortality in 48 h

(b) Mortality in 72 h

(c) Mortality in 7 days

Fig. 4 DCA of our methods and traditional models in different tasks.

4 Discussion

In this study, we developed a framework to continuously predict the mortality of all admissions in the ICU by using RNN. The strengths of this study are as follows. First, unlike the static prediction method, our proposed method could use real-time data to continuously predict patients' mortality during their hospitalization by constructing the patient sequence; that is, it could capture the time information in the sequence to improve the performance. In addition, the problem of the large difference in the lengths of the traditional method's sequence input could be addressed by introducing dynamic window sliding for each patient. Under the framework of our proposed method, three models were constructed with different lengths of ahead windows, thus enabling different predictive horizons to be developed and validated in large, heterogeneous populations of ICU patients. The outcomes of these three predictive models are given in terms of mortality in 48 h, 72 h, and 7 d. Moreover, label resampling was introduced to mitigate the imbalance between positive and negative labels, and a hyperparameter sweep was used to obtain the best model and save time on experimental validation.

The performance of our method was compared with that of other machine learning algorithms, including logistic regression (LR), random forest (RF), and XGBoost. The predicted performance had superior accuracy, F1-score, AUROC, and AUPRC for all tasks. Overall, our method performed better than the baseline algorithms.

In addition to the performance matrix, discrimination, and reclassification, DCA was introduced to evaluate the performance of the validation. Findings show that our model could successfully identify survivors and non survivors, and its score proposals resulted in the reclassification of a large proportion of patients. At the same time, it could yield a larger standardized net benefit compared with the baseline models.

On the basis of the above findings, the newly developed models can be potentially applied to the risk stratification of ICU patients right after their admission and can be monitored continuously because our proposed framework provides the predicted mortality risk in real time as the new lab results are updated in the EHR database. Under these circumstances, with the help of such a predictive model, more appropriate care can be given to patients with a high risk of mortality. Therefore, we can better allocate the limited medical resources in ICU settings.

The limitations and the future work of this study should be mentioned. First, because of the nature of real-world medical data, missing values and imbalanced positive-negative sets were major problems during the development and validation of the predictive model. Although missing values and imbalance were properly dealt with during data pre-processing, more advanced methods should be tested to improve the data quality. Second, the predictive models were built to predict the mortality risk of all patients admitted to the ICU by using demographics and lab data. Patients with different medical histories and comorbidities vary greatly in the risk of mortality; thus, including medical history and comorbidity information in the model construction would be valuable.

5 Conclusion

Three continuous predictive models with different time windows were developed in this study. These models exhibit significantly improved performance compared with current machine learning methods, namely, LR, RF, and XGBoost. We innovated several ideas, including constructing patient sequences to achieve continuous prediction, using appropriate dynamic windows to ensure more accurate results, and adopting a series of result evaluation methods to provide more reliable results. Findings indicate that our method achieves the best performance. Thus, our new models are promising tools for building mortality prediction models in clinical and research settings. This model was developed to predict the risk of in-hospital mortality for all patients admitted to the ICU, thus helping physicians pay more attention to high-risk patients and anticipate potential complications.

Acknowledgment

This research was supported by the Key Discipline Fund of Shenzhen Hospital of Southern Medical University (No. 2021-2023ICU), the New-Generation Information Technology by the Scientific Research Platform of Institutions of Higher Education of the Education Department of Guangdong Province (No. 2021ZDZX1014), and the Shenzhen University (SZU) Top Ranking Project (No. 86000000210). We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

References

[1] G. Rouleau, M. P. Gagnon, and J. Côté, Impacts of information and communication technologies on nursing care: An overview of systematic reviews (protocol), *Syst. Rev.*, vol. 4, no. 1, p. 75, 2015.

[2] K. G. M. Moons, R. F. Wolff, R. D. Riley, P. F. Whiting, M. Westwood, G. S. Collins, J. B. Reitsma, J. Kleijnen, and S. Mallett, PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration, *Ann. Intern. Med.*, vol. 170, no.1, pp. W1–W33, 2019.

[3] R. Lilford and P. Pronovost, Using hospital mortality rates to judge hospital performance: A bad idea that just won't go away, *BMJ*, vol. 340, p. c2016, 2010.

[4] K. M. D. M. Karunarathna, Predicting ICU death with summarized patient data, presented at the 2018 IEEE 8th Annu. Computing and Communication Workshop and Conf. (CCWC), Las Vegas, NV, USA, 2018, pp. 238–247.

[5] R. T. Thomson, D. Leuttel, F. Healey, and S. Scobie, *Safer Care for the Acutely Ill Patient: Learning from Serious Incidents*. London, UK: National Patient Safety Agency, 2007.

[6] R. Sadeghi, T. Banerjee, and W. Romine, Early hospital mortality prediction using vital signals, *Smart Health*, vols. 9–10, pp. 265–274, 2018.

[7] B. E. Keuning, T. Kaufmann, R. Wiersema, A. Granholm, V. Pettilä, M. H. Møller, C. F. Christiansen, J. C. Forte, H. Snieder, F. Keus, et al., Mortality prediction models in the adult critically ill: A scoping review, *Acta Anaesthesiol. S*cand., vol. 64, no. 4, pp. 424–442, 2020.

[8] S. Lemeshow and J. R. Le Gall, Modeling the severity of illness of ICU patients: A systems update, *JAMA*, vol. 272, no. 13, pp. 1049–1055, 1994.

[9] J. I. F. Salluh and M. Soares, ICU severity of illness scores: APACHE, SAPS and MPM, *Curr. Opin. Crit. Care*, vol. 20, no. 5, pp. 557–565, 2014.

[10] J. L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. K. Reinhart, P. M. Suter, and L. G. Thijs, The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure, *Intensive Care Med.*, vol. 22, no. 7, pp. 707–710, 1996.

[11] J. E. Zimmerman, A. A. Kramer, D. S. Mcnair, and F. M. Malila, Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients, *Crit. Care Med.*, vol. 34, no. 5, pp. 1297–1310, 2006.

[12] W. A. Knaus, E. A. Draper, D. P. Wagner, and J. E. Zimmerman, APACHE II: A severity of disease classification system, *Crit. Care Med.*, vol. 13, no. 10, pp. 818–829, 1985.

[13] T. L. Higgins, D. Teres, W. Copes, B. Nathanson, M. Stark, and A. Kramer, Updated mortality probability model (MPM-III), *Chest*, vol. 128, no. 4, p. 348S, 2005.

[14] K. Strand and H. Flaatten, Severity scoring in the ICU: A review, *Acta Anaesthesiol. Scand.*, vol. 52, no. 4, pp. 467–478, 2008.

[15] S. Vairavan, L. Eshelman, S. Haider, A. Flower, and A. Seiver, Prediction of mortality in an intensive care unit using logistic regression and a hidden Markov model, presented at the 2012 Computing in Cardiology, Krakow, Poland, 2012, pp. 393–396.

[16] S. Barnes, E. Hamrock, M. Toerper, S. Siddiqui, and S. Levin, Real-time prediction of inpatient length of stay for discharge prioritization, *J. Am. Med. Inform. Assoc.*, vol. 23, no. e1, pp. e2–e10, 2016.

[17] N. El-Rashidy, S. El-Sappagh, T. Abuhmed, S. Abdelrazek, and H. M. El-Bakry, Intensive care unit mortality prediction: An improved patient-specific stacking ensemble model, *IEEE Access*, vol. 8, pp. 133541–133564, 2020.

[18] G. Gutierrez, Artificial intelligence in the intensive care unit, *Crit. Care*, vol. 24, no. 1, p. 101, 2020.

[19] S. Leteurtre, A. Duhamel, B. Grandbastien, F. Proulx, J. Cotting, R. Gottesman, A. Joffe, B. Wagner, P. Hubert, A. Martinot, et al., Daily estimation of the severity of multiple organ dysfunction syndrome in critically ill children, *CMAJ*, vol. 182, no. 11, pp. 1181–1187, 2010.

[20] S. Leteurtre, A. Duhamel, V. Deken, J. Lacroix, F. Leclerc, and Groupe Francophone de Réanimation et Urgences Pédiatriques, Daily estimation of the severity of organ dysfunctions in critically ill children by using the PELOD-2 score, *Crit. Care*, vol. 19, no. 1, p. 324, 2015.

[21] O. Badawi, X. G. Liu, E. Hassan, P. J. Amelung, and S. Swami, Evaluation of ICU risk models adapted for use as continuous markers of severity of illness throughout the ICU stay, *Crit. Care Med.*, vol. 46, no. 3, pp. 361–367, 2018.

[22] M. J. Rothman, S. I. Rothman, and J. Beals IV, Development and validation of a continuous measure of patient condition using the electronic medical record, *J. Biomed. Inform.*, vol. 46, no. 5, pp. 837–848, 2013.

[23] M. J. Rothman, .J. J. Tepas, A. J. Nowalk, J. E. Levin, J. M. Rimar, A. Marchetti, and A. L. Hsiao, Development and validation of a continuously age-adjusted measure of patient condition for hospitalized children using the electronic medical record, *J. Biomed. Inform.*, vol. 66, pp. 180–193, 2017.

[24] C. W. Hug and P. Szolovits, ICU acuity: Real-time models versus daily models, presented at the AMIA 2009 Symp. Proc., American Medical Informatics Association, San Francisco, CA, USA, 2009, pp. 260–264.

[25] B. Shickel, T. J. Loftus, L. Adhikari, T. Ozrazgat-Baslanti, A. Bihorac, and P. Rashidi, DeepSOFA: A continuous acuity score for critically ill patients using clinically interpretable deep learning, *Sci. Rep.*, vol. 9, no. 1, p. 1879, 2019.

[26] G. Clermont, D. C. Angus, S. M. Dirusso, M. Griffin, and 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.*, vol. 29, no. 2, pp. 291–296, 2001.

[27] G. Meyfroidt, F. Güiza, J. Ramon, and M. Bruynooghe, Machine learning techniques to examine large patient databases, *Best Pract. Res. Clin. Anaesthesiol.*, vol. 23, no. 1, pp. 127–143, 2009.

[28] S. Kim, W. Kim, and R. W. Park, A comparison of intensive care unit mortality prediction models through the use of

data mining techniques, *Healthc. Inform. Res.*, vol. 17, no. 4, pp. 232–243, 2011.

[29] M. Babaie, S. Kalra, A. Sriram, C. Mitcheltree, S. J. Zhu, A. Khatami, S. Rahnamayan, and H. R. Tizhoosh, Classification and retrieval of digital pathology scans: A new dataset, presented at the 2017 IEEE Conf. Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 2017, pp. 760–768.

[30] A. Khatami, M. Babaie, A. Khosravi, H. R. Tizhoosh, S. M. Salaken, and S. Nahavandi, A deep-structural medical image classification for a Radon-based image retrieval, presented at the 2017 IEEE 30th Canadian Conf. Electrical and Computer Engineering (CCECE), Windsor, Canada, 2017, pp. 1–4.

[31] A. Khatami, A. Khosravi, C. P. Lim, and S. Nahavandi, A wavelet deep belief network-based classifier for medical images, in *Proc. 23rd Int. Conf. Neural Information Processing*, Kyoto, Japan, 2016, pp. 467–474.

[32] A. Khatami, A. Khosravi, T. Nguyen, C. P. Lim, and S. Nahavandi, Medical image analysis using wavelet transform and deep belief networks, *Expert Syst. Appl.*, vol. 86, pp. 190–198, 2017.

[33] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[34] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[35] E. Laksana, M. Aczon, L. Ho, C. Carlin, D. Ledbetter, and R. Wetzel, The impact of extraneous features on the performance of recurrent neural network models in clinical tasks, *J. Biomed. Inform.*, vol. 102, p. 103351, 2020.

[36] C. S. Carlin, L. V. Ho, D. R. Ledbetter, M. D. Aczon, and R. C. Wetzel, Predicting individual physiologically acceptable states at discharge from a pediatric intensive care unit, *J. Am. Med. Inform. Assoc.*, vol. 25, no. 12, pp. 1600–1607, 2018.

[37] M. C. Winter, T. E. Day, D. R. Ledbetter, M. D. Aczon, C. J. L. Newth, R. C. Wetzel, and P. A. Ross, Machine learning to predict cardiac death within 1 hour after terminal extubation, *Pediatr. Crit. Care Med.*, vol. 22, no. 2, pp. 161–171, 2021.

[38] E. Choi, A. Schuetz, W. F. Stewart, and J. M. Sun, Using recurrent neural network models for early detection of heart failure onset, *J. Am. Med. Inform. Assoc.*, vol. 24, no. 2, pp. 361–370, 2017.

[39] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. B. Liu, J. Marcus, M. M. Sun, et al., Scalable and accurate deep learning with electronic health records, *NPJ Digit. Med.*, vol. 1, p. 18, 2018.

[40] M. Saqib, Y. Sha, and M. D. Wang, Early prediction of sepsis in EMR records using traditional ML techniques and deep learning LSTM networks, presented at the 2018 40th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 2018, pp. 4038–4041.

[41] S. Kannan, G. Yengera, D. Mutter, J. Marescaux, and N. Padoy, Future-state predicting LSTM for early surgery type recognition, *IEEE Trans. Med. Imaging*, vol. 39, no. 3, pp. 556–566, 2020.

[42] N. Tomašev, N. Harris, S. Baur, A. Mottram, X. Glorot, J. W. Rae, M. Zielinski, H. Askham, A. Saraiva, V. Magliulo, et al., Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records, *Nat. Protoc.*, vol. 16, no. 6, pp. 2765–2787, 2021.

[43] A. E. W. Johnson, T. J. Pollard, L. Shen, L. W. H. Lehman, M. L. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, MIMIC-III, a freely accessible critical care database, *Sci. Data*, vol. 3, p. 160035, 2016.

[44] A. J. Vickers, A. M. Cronin, E. B. Elkin, and M. Gonen, Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers, *BMC Med. Inform. Decis. Mak.*, vol. 8, no. 1, p. 53, 2008.

[45] K. F. Kerr, M. D. Brown, K. H. Zhu, and H. Janes, Assessing the clinical impact of risk prediction models with decision curves: Guidance for correct interpretation and appropriate use, *J. Clin. Oncol.*, vol. 34, no. 21, pp. 2534–2540, 2016.

Guangjian Zeng received the BS degree in biomedical engineering from Shenzhen University, Shenzhen, China, in 2021. He is currently pursuing the MS degree in biomedical engineering at Shenzhen University, Shenzhen, China. His research interests include artificial intelligence, medical data analysis, and medical image processing.

Jinhu Zhuang received the BS degree in biomedical engineering from Shenzhen University, Shenzhen, China, in 2021. He is currently pursuing the MS degree in biomedical engineering at Shenzhen University, Shenzhen, China. His research interests include clinical informatics and medical data analysis.

Haofan Huang received the BS degree in biomedical engineering from Shenzhen University, Shenzhen, China, in 2020. He is currently working with Dr. Yi Gao as a research assistant with Shenzhen University, Shenzhen, China. His research interests include medical data analysis and medical image processing.

Xiaxia Yu received the PhD degree in computer science from Georgia State University, Atlanta, GA, USA, in 2015. She joined the School of Biomedical Engineering, Shenzhen University, Shenzhen, China, as an assistant professor, in 2017. Her research interests include clinical informatics, medical statistics, and molecular bioinformatics.

Mu Tian received the PhD degree in applied mathematics and statistics from the State University of New York at Stony Brook, New York, NY, USA, in 2017. After graduation, he joined Facebook, Inc. as a research scientist and then joined Shenzhen University, Shenzhen, China, in 2020 as an associate research fellow. His research interests include artificial intelligence, computer vision, and medical image analysis.

Yi Gao received the BS and MS degrees from Tsinghua University, Beijing, China, in 2003 and 2005, respectively, and the MS degree in mathematics and the PhD degree in biomedical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2008 and 2011, respectively. He was a postdoctoral researcher with the Harvard Medical School and an assistant professor with the Department of Biomedical Informatics, State University of New York at Stony Brook, New York, NY, USA. He is currently a professor at the School of Biomedical Engineering, Shenzhen University, Shenzhen, China. His research interests include medical image analysis.

Yong Liu received the BS degree from Tongji Medical University, Wuhan, China, in 1997, the MS degree in Fudan University, Shanghai, China, in 2004, and the PhD degree in general surgery from the Southern Medical University, Guangzhou, China, in 2009. He is a director of the Department of Critical Care Medicine, Shenzhen Hospital, Southern Medical University, Shenzhen, China. He is also an associate director of the Critical Care Medicine Branch of the Shenzhen Chinese Medical Association and Shenzhen Physicians Association and chairman of the Shenzhen ECMO Association. His research interests include artificial intelligence in the intensive care unit.