Predicting ICU Mortality by Supervised Bidirectional LSTM Networks

Yao Zhu[1,4], Xiaoliang Fan[1,4,*], Jinzhun Wu[2,4], Xiao Liu[3], Jia Shi[5], Cheng Wang[1]
[1] Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, Xiamen, China
[2] The First Affiliated Hospital, Xiamen University, Xiamen, China
[3] School of Information Technology, Deakin University, Melbourne, Australia
[4] Digital Fujian Institute of Healthcare & Biomedical Big Data Research, Xiamen University, Xiamen, China
[5] School of Information Science and Engineering, Lanzhou University, Lanzhou, China
csyzhu@stu.xmu.edu.cn, xfb_fxl@xm.gov.cn, 1923731201@qq.com, xiao.liu@deakin.edu.au, shij2016@lzu.edu.cn, cwang@xmu.edu.cn

Abstract. Mortality prediction in the Intensive Care Unit (ICU) is considered as one of critical steps for the treatment of patients in serious condition. It is a big challenge to model time-series variables for mortality prediction in ICU, because physiological variables such as heart rate and blood pressure are sampled with inconsistent time frequencies. In addition, it is difficult to capture the timing changes of clinical data and to interpret the prediction result of ICU mortality. To deal with these challenges, in this paper, we propose a novel ICU mortality prediction algorithm combining bidirectional LSTM (Long Short-Term Memory) model with supervised learning. First, we preprocess 37 time-series variables related to patients' signs. Second, we construct the Bidirectional LSTM model with supervision technique to accurately reflect significant changes in patients' signs. Finally, we train and evaluate our model using a real-world dataset containing 4,000 ICU patients. Experimental results show that our proposed method can significantly outperform many baseline methods.

Keywords: Deep learning, ICU, mortality prediction, LSTM.

1 Introduction

Intensive Care Unit (ICU) is a rescue center for critically ill patients, which provides the upmost service to reduce "dead in bed" events. For example, in the United States[1], 5.7 million people are admitted to the ICU each year, and 2.3 million will require a mechanical ventilator to help them breathe. Mortality prediction in ICU is considered as one of critical steps for the treatment of patients with serious condition in ICU. The major objectives of mortality prediction in ICU are: 1) assess and monitor the severity of patients' illness continuously based on their physical condition; and 2) determine those patients of the highest risk to be treated with the upmost treatments, interventions and resources. Therefore, accurate ICU mortality prediction can not only give clinicians

* Correspondence author.
[1] Critical Care Statistics, http://www.sccm.org/Communications/Pages/CriticalCareStats.aspx

a better and earlier sense of which patients are likely to get worse, but also facilitates the efficient allocation of hospital resources.

Many machine learning methods [4-7] have been employed to optimize the prediction model with the availability of high-quality ICU datasets. In fact, the mortally ill patients in ICU could generate a large amount of time-series data, such as heart rate, blood pressure, temperature, Glasgow Coma Scale (GCS), and so on. Recently, deep neural networks such as RNNs [8], LSTM [9], have been applied to process continuous data sequences such as time-series variables in ICU.

Inspired by studies on ICU mortality prediction with deep neural networks (DNNs) [10, 11], we aim to make use of the strong learning ability of DNNs to capture the fluctuations in time-series variables which could adequately reflect changes in patients' illness states. However, it is non-trivial to capture the fluctuations in time-series variables in ICU, and we need to address the following two challenges:

—To model time-series variables in ICU, which is challenging because physiological variables are usually sampled with inconsistent time patterns. For instance, heart rate variables are collected every 20 minutes, while urine data is sampled in a 5-hour interval.

—To incorporate the learning ability of deep neural networks into the interpretability of the prediction result of ICU mortality, which is challenging as deep learning approaches are known to have difficulties in inherently modeling the causation that could provide straightforward decision supports for clinicians.

To address these challenges, in this paper we propose a novel ICU mortality prediction method, named BiLSTM-ST, which combines a bidirectional LSTM (Long Short-Term Memory) model with supervised learning technique. First, based on the experience of the clinicians and the occurrence frequency of time-series variables, we choose the available general descriptors and time-series variables as the input of each patient. And the time-series variables are preprocessed via data simplification, data completion and data normalization. Second, we construct the Bidirectional LSTM model with supervision technique to accurately reflect significant changes in patients' signs. Finally, we train and evaluate the proposed model on a public ICU dataset with 4,000 ICU patients. Experimental results show that our proposed Bidirectional LSTM model with supervised learning can significantly outperform seven baseline methods.

2 Related Works

High-quality ICU datasets are gradually becoming open and available for research purpose, which enables an increasing number of works on mortality prediction in ICUs. To assess the mortality risk of patients in ICUs, there are several classic scoring methods such as SAPS II [1] and APACHE IV [2]. SAPS stands for Simplified Acute Physiology Score while APACHE stands for Acute Physiology And Chronic Health Evaluation, and they are both applied within 24 hours of admission of a patient to the ICU. However, it is well known that these scores are limited to a few fixed indicators, which could make it very difficult to accurately reflect the dynamic evolution of patients' signs in ICU [3].

Machine learning technology has been widely utilized in statistical analysis. The work in [4] achieves a high score in Mortality Prediction Challenge of Physionet by using Support Vector Machine (SVM) classifiers. In addition, both general descriptors and aggregated variables are used as features for the model. Later, the authors in [6] come up with the CHISQ Classification Algorithm which is designed to address the imbalance problem in the binary classification. However, the inconsistence in sampling frequencies for time-series variables is known to cause great difficulties in statistical analysis, as the time gaps could not be easily characterized to provide valuable decision supports for clinicians.

More recently, deep neural networks are utilized to learn from the sequence data and able to grasp the long-term dependence of time series data [15-16]. Since a large part of the medical data is of time-series type, LSTM networks have been applied to learn to classify diagnoses based on patient's Electronic Health Record in pediatric intensive care unit (PICU) [12]. A bidirectional LSTM model with attention mechanism to predict mortality outcomes in ICUs is proposed showing competitive results on 2012 PhysioNet datasets [9]. The work in [13] combines LSTM and latent topic modeling for mortality prediction so that it can not only predict but also interpret the predictive results on mortality.

Furthermore, the work in [14] proposed a full-time supervision based bidirectional RNN method, called FTS-BRNN, for QA tasks. Compared with our work, the major difference is that we directly modify the LSTM neural network by adding supervision technique, and we design the particular *Loss* function method to better apply to clinical data.

3 Preliminary and Datasets

3.1 Basic Notations

In this work, we attempt to predict the mortality of patients in the ICU using deep learning methods. The data we possess includes up to 37 time-series variables of each patient recorded during the first 48 hours of their stay in the ICU, including such as heart rate, blood pressure, weight, etc. We are taking advantage of these time-series data to explore the patterns of the patient's physical condition changes, in order to achieve the high accuracy of mortality prediction.

The mortality prediction in the ICU essentially evaluates the risk of the death based on the specific patient's current physiological condition so that the doctor can take an appropriate care. Given the clinical experience of doctors, patients' physiological states during the first 48 hours in the ICU could profoundly influence his or her clinical trends within the next 30 days. Thus, as Equation (1) shows, we divided the predictions into two categories: 1) death in ICU within 30 days, and 2) death in ICU exceeding 30 days or survival in ICU. In other words, the first type of patients requires more monitoring and care services than those of the second type.

$$Prediction = \begin{cases} 1 & survival\ days \leq 30 \\ 0 & survival\ \text{days} > 30 \end{cases} \tag{1}$$

3.2 Datasets

The ICU data we employed are from the PhysioNet/Computing in Cardiology Challenge 2012[2]. There are 4,000 patients and records for each patient consist of two parts: general descriptors (as shown in italic fonts in Table 1) and time-series data. First, general descriptors contain patients' basic information, including recordID, age, gender, height, and ICU type. Second, time-series data is composed of 37 variables that reflect the patient's physiological state. Each variable has an associated time-stamp indicating the observed time of the variable. In addition, the dataset provides the number of days the patient has stayed in ICUs and the fact of death or survival for patients. Thus, we label patients who died in the ICU after more than 30 days or survival in ICU as 0, while patients who died in ICU within 30 days as 1.

Table 1. Data samples of the PhysioNet dataset

Time	**Parameter**	**Value**	**Time**	**Parameter**	**Value**
00:00	*RecordID*	*132548*	01:39	GCS	15
00:00	*Age*	*68*	01:39	NIDiasABP	80
00:00	*Gender*	*0*	01:39	NIMAP	112.7
00:00	*Height*	*162.6*	01:39	NISysABP	178
00:00	*ICUType*	*3*	02:14	MAP	144
00:09	GCS	15	19:09	Urine	18
00:09	NIDiasABP	79	20:09	Weight	87
00:09	NIMAP	112	25:09	Temp	36.7
00:09	NISysABP	178	47:09	HR	60

3.3 Observations

3.3.1 General Descriptors

We perform the basic statistics on general descriptors including patients' basic information. The preliminary observations are described as follows: 1) 12.88% of patients survived for less than 30 days; 2) male patients are slightly more than female patients; 3) there are four different types of ICU: coronary care unit (14.43%), cardiac surgery recovery unit (21.85%), medical ICU (37.02%), and surgical ICU (26.70%); and 4) the elderly (age over 60) take the highest proportion (61.98%).

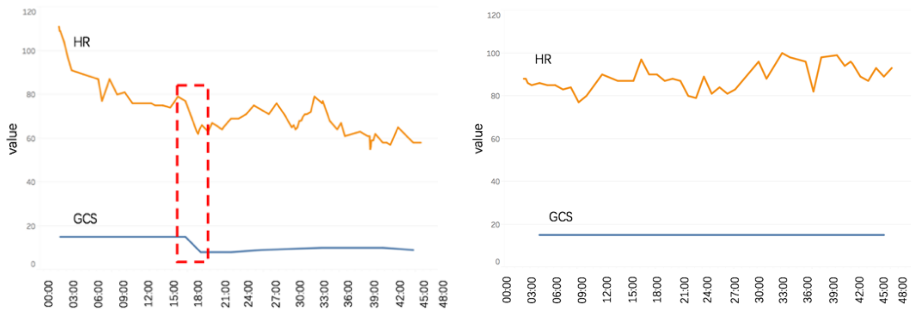
3.3.2 Time-series Data

In Fig. 1, we further illustrate the 48-hour time-series data between two patients, patient_a was dead in ICU within 30 days, while patient_b survived in the ICU. We selected two high-frequency variables, HR and GCS, for comparison. heart rate (HR) is the most basic sign of patients and GCS reflects the patient's degree of coma.

---

[2] PhysioNet/Computing in Cardiology Challenge 2012, https://www.physionet.org/challenge/2012/

For example, the fluctuating range of patient_a's heart rate (HR) is relatively larger, showing an obvious downward trend at the same time, while patient_b's heart rate has fluctuated within the normal range. In addition, GCS of patient_a has a big drop to below 10 at 18:00, whereas GCS of patient_b remains stable within 48 hours. The dashed box marks the period during which both HR and GCS indicators of patient_a change drastically.



(a) patient_a (dead in ICU within 30 days) (b) patient_b (survived in ICU)

Figure 1. The comparison between two patients.

4 The Proposed Method

4.1 Data Preprocessing

There are some general descriptors for each patient and 37 time-series variables in the dataset. For the general descriptors, we chose the *Age* and *ICUType* as part of input for the individual patient. It's worth mentioning that the *ICUType* is represented by the one-hot code. For the 37 time-series variables, *Cholesterol*, *TropI* and *TropT* even have no record in the dataset. Thus, we utilized 34 time-series variables for patients ultimately. To sum up, the input of each patient consists of two general descriptors and thirty-four time-series variables.

We recorded 34 time-series variables about the vital signs of the patients with a one-hour time interval, that is to say, there are at most forty-eight records in 48 hours for the individual variable. Specifically, three data preprocessing steps are conducted as follow: **1) data simplification.** For multiple records of the same variable in an hour, we use the average value as there are very small changes; **2) data completion.** For missing records in an hour, there are two circumstances. For the first circumstance, if there is no such data for a period of 48 hours, the variable is assigned with the average value of the variable in the same type of ICU the patient belongs to. For the second circumstance, if the record is only missing occasionally within 48 hours, the missing values are replaced by the neighboring records; **3) data normalization.** The mean and standard deviation are used for normalization, so that the time-series information for each patient is represented by a 34×48 matrix.

Furthermore, we labeled all patients as {0, 1}, according to Equation (1). Since the information provided in the dataset includes the patient's survival days, we could screen cases by the 30-day threshold. Thus, we obtained the label for each patient.

4.2 Algorithm

In order to realize this classification model for mortality prediction, we try to utilize deep learning method to construct the model. Specifically, LSTM recurrent neural network is employed in this paper as it is more suitable for time-series data. Therefore, we start from the basic LSTM classification model and then propose the Bidirectional LSTM model with continuous improvement. At the same time, supervised learning technique is applied to these neural network structures to effectively improve the accuracy of the prediction results. Finally, during the model training phase, we find that the imbalance between positive and negative samples significantly affects the classification result. To deal with this problem, we use the up-sampling method to balance the training data and obtain a more accurate model.

4.2.1 Bidirectional LSTM

In recurrent neural networks, Long Short-Term Memory (LSTM) is a relatively more efficient structure. Compared with the basic RNN, LSTM are able to make better use of long-term dependence among the data. The following Equation (2) shows how the LSTM structure works in a looping module. Specifically, s_c^t represents the state of Cell at time t, which is related to a_c^t (the input at time t) and s_c^{t-1} (the state at time t-1). For each patient, his or her physical condition at time t is affected by the previous illness state and the specific vital signs of the subject at this time. Of course, in order to deal with long-term dependencies, b_f^t means the forget gates' reservations about the state of the Cell at the last moment, but b_i^t decides to what extent the input information is received. The absolute value of the disease index and its changes reflect the patient's physiological status. Clearly, the abnormally fluctuating time-series data in the model require more attention than the normal stable values. From this point of view, memory and forgetting mechanisms are very important to the prediction model.

$$s_c^t = b_f^t * s_c^{t-1} + b_i^t * g(a_c^t) \tag{2}$$

For our scenario, each patient has 48 hours of time-series variables, which can be captured by the LSTM network. Furthermore, the introduction of Bidirectional LSTM can help us make better use of both past and future time-series data at each time step. In fact, BiLSTM consists of forward and backward LSTMs, which allows for a more complete understanding of the characteristics of all time-series data. In another word, BiLSTM could avoid the blindness that propagation in unidirectional may cause. As a result, BiLSTM is employed to process the input data to reflect the fluctuation of patients' illness condition as it is more capable of sensitively capturing the changes of patient's physical signs than unidirectional LSTM models.

4.2.2 Supervision Technique

In a simplest Bidirectional LSTM recurrent neural network, there is only one output per recurrent module of the hidden layer. When the gradients back propagate, the entire module calculates the loss value only at the last step. In contrast, a recurrent module can actually be unfolded into multiple modules, each part of which corresponds to a certain moment of the input. Thus, every step can generate output respectively. As for the calculation of the loss value, we apply a supervision technique. For each generated output, as shown in Equation (3) and (4), we compute the loss with the label value, and the $\mathcal{L}$ (loss at last) is weighted by all *N* losses through all *N* steps. *N* is the total number of steps, and g_i are linear interpolation between 0 and 1. In this way, during the training period of the model, the output of all steps will affect the final parameters so that the model can detect subtle changes of the vital signs more sensitively.

$$g_i = \frac{i}{N} \qquad i = 1,2 \dots, N \tag{3}$$

$$\mathcal{L} = \sum_{i=1}^{N} g_i \times Loss_i \tag{4}$$

The supervision technique here has some advantages over the attention mechanism. First, the attention mechanism only concerned with global dependencies, but in fact, it is the local state that affects the outcome for the mortality prediction. And grabbing characteristics of local state is exactly what the supervision technique is good at. Second, the supervision technique shows the ordered incremental importance for all time steps, which is consistent with the trend of patients' states. However, the attention on time steps is unordered according to the attention mechanism. Third, attention mechanism tends to converge slower and need a larger amount of computation when dealing with superabundant steps, but supervision technique improves the loss value in recurrent module to make the model effectively handle the long sequence problem. For example, in our experiment, model with attention mechanism converge slower than the supervised LSTM networks when the number of total steps equals to 48.

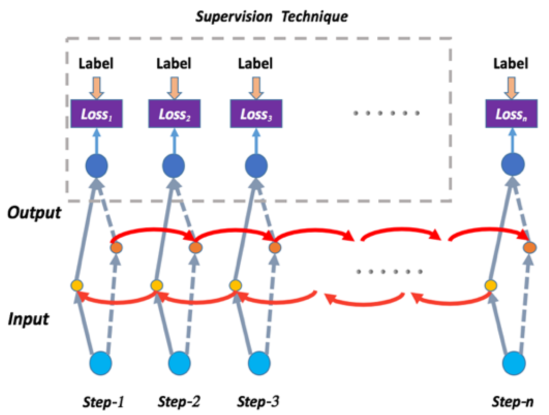


Figure 2. BiLSTM model with supervised learning

Fig. 2 depicts the structure of the supervision technique applied to the Bidirectional LSTM network. The figure shows the LSTM structure in both forward and backward directions after the hidden layer is expanded, where Label is the expected output of the module. In addition, dotted arrows represent inputs to the forward LSTM process, while solid arrows represent inputs to the backward LSTM process. The small orange circle is the cell of the forward LSTM network and the small yellow circle is the cell of the backward LSTM network.

4.2.3 Samples Balance

In the process of scanning the datasets with 4,000 patients, we find that samples with positive and negative labels are uneven. For example, there are 3,485 patients with the label 0, while only 515 patients are labeled as 1. If trained with the unbalanced data directly, it will be difficult for the model to capture features of patients with a label of 1. Thus, we augment the sample with the up-sampling method to reduce the gap between the positive and negative samples. Finally, with a total of 4,200 patients for training, there are 2,786 negative samples and 1,414 positive samples so that a reasonable balance is achieved.

5 Experiments and Evaluations

5.1 Experimental Settings

We perform the experiment on 4,000 cases of the Physionet dataset, with 3,200 patients as the training set and the remaining 800 as the test set. After data preprocessing, for the training set, we up-sampled the records for balance, expanding the training set to 4,200. We apply the supervision technique to three different types of recurrent neural networks including GRU, LSTM and Bidirectional LSTM, so as to produce a variety of models for mortality prediction for comparison purpose. There are 128 cells per hidden layer in our final model, and the softmax function is utilized to classify the output into two categories. We make parameters optimization with 5-fold cross-validation by taking 20% of the training set for parameter validation.

The algorithms are developed using Python 3.6.1, TensorFlow 1.2.1 and Keras 2.1.1. Meanwhile, the experiment is conducted on 14 CPU cores (Intel(R) Xeon(R) CPU E5-2683 v3 @ 2.00 GHz), with two GPUs (GeForce GTX TITAN X).

5.2 Evaluation Metrics

As the prediction of mortality is a binary classification problem, we choose *Precision, Recall, F1,* and *AUC* to evaluate our model and compare with baselines. For a binary classification problem, we usually take the class of interest as the positive class and the others as negative class. Thus, there are four cases where the classifier is predicted correctly or incorrectly on the dataset, as shown in Table 2.

TABLE 2. Confusion Matrix.

Predicted \ *Outcome*	**1**	**0**
1	***TP***	***FP***
0	***FN***	***TN***

Precision equals *TP* divided by *TP* plus *FP*, and *Recall* equals *TP* divided by *TP* plus *FN*. Besides, *F1* is the harmonic mean of Precision and Recall, which is defined as:

$$\frac{2}{F1} = \frac{1}{Precision} + \frac{1}{Recall} \tag{5}$$

In addition, *AUC* characterizes the ability of the classifier to rank positive samples in front of negative samples. The greater the value of *AUC* is, the better the effect of classification achieves.

5.3 Baselines

We compare our BiLSTM-ST model with the following six baselines:

- **CNN** [17]**.** Convolutional Neural Network, one of the most popular models of deep learning but it lacks handling of time series data.
- **LSTM** [18]**.** Long short-term memory (LSTM) network is a special kind of RNN. It is explicitly designed to avoid the long-term dependency problem and it is capable of remembering information for long periods of time.
- **BiLSTM** [19]. Bidirectional LSTM network has both forward and backward LSTM structure and tt has a better overall understanding of time series data.
- **GRU** [20]. Gated Recurrent Unit, is a variant of LSTM that maintains the effect of LSTM while making the structure simpler.
- **GRU-ST.** Gated Recurrent Unit (GRU) network with supervision technique, improve the prediction of the model to a certain extent.
- **LSTM-ST.** Long short-term memory (LSTM) network with supervision technique.
- **BiLSTM (attention)** [9]. Bidirectional LSTM network with attention mechanism. The attention mechanism is to weight importance of each time step and it might capture the human decision making implicitly.

5.4 Results Summary

As shown in Table 3, we employ four indicators Precision, Recall, F1 and AUC to compare different models for mortality prediction.

Comparison with Baselines. We can see clearly from Table 3 that the Bidirectional LSTM model with supervised learning technique (BiLSTM-ST) outperforms baselines in general. Specifically, the prediction effect of recurrent neural networks (e.g., GRU, LSTM, BiLSTM) is obviously better than that of convolution neural network (CNN) because recurrent neural networks can better deal with the time series data. Furthermore, the effectiveness of both supervised learning technique and attention mechanism indicates that it is helpful to combine such techniques in recurrent neural networks for ICU mortality prediction.

Comparison with BiLSTM (attention). BiLSTM-ST performs better than BiLSTM (attention). The reason is that in contrast to the attention mechanism, the supervision technique could supervise each step in reasonable order so that the model can learn features and predict more effectively. Meanwhile, the attention mechanism lacks of attention to local features during the time steps. In addition, the supervision technique is more suitable for mortality prediction because it can better deal with the long sequence problem like the time series data of 48 hours.

TABLE 3. Comparison among different methods.

Model	*Precision*	*Recall*	*F1*	*AUC*
CNN	0.573	0.529	0.550	0.717
GRU	0.691	0.650	0.670	0.804
LSTM	0.638	0.733	0.682	0.809
BiLSTM	0.712	0.723	0.717	0.825
GRU-ST	0.789	0.670	0.724	0.822
LSTM-ST	0.826	0.685	0.749	0.832
BiLSTM (attention)	0.798	0.738	0.767	0.838
BiLSTM-ST	**0.848**	**0.745**	**0.793**	**0.869**
BiLSTM-ST(24-hour)	0.839	0.731	0.781	0.831
BiLSTM-ST(36-hour)	0.845	0.747	0.793	0.868

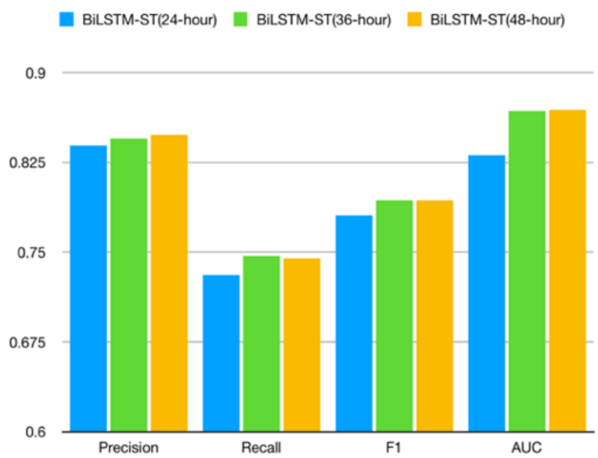


Figure 3. Comparison of BiLSTM-ST with data in varying length

Variants of BiLSTM-ST. In addition to utilizing all 48 hours of data (BiLSTM-ST), we also selected the sequence data of the first 24-hour and first 36-hour after the patient was admitted into the ICU, and the mortality prediction is carried out with the BiLSTM-ST model. As shown in Fig. 3, BiLSTM-ST(24-hour) is relatively weak compared to BiLSTM-ST(36-hour) and BiLSTM-ST(48-hour), which indicates that the data within 24 hours is not enough to accurately reflect the severity of the patient's illness condition for the following 30 days. Nevertheless, the prediction accuracy of BiLSTM-ST(36-hour) and BiLSTM-ST(48-hour) is almost identical. Specifically, BiLSTM-ST(36-hour) even exceeds BiLSTM-ST(48-hour) on *Recall*. As a result, we could use 36-hour of time-series data instead of 48-hour so as to give clinicians an earlier sense of which patients will require critical targeted treatments.

6 Conclusion

In this paper, we proposed a novel mortality prediction algorithm in Intensive Care Unit (ICU). We trained and tested our model with a real-world dataset of 4,000 patients to accurately reflect significant changes in patients' vital signs. Our experimental results demonstrated that the bidirectional LSTM model with supervised learning mechanism outperforms all other baseline methods. The significance of our novel prediction model includes: 1) by effectively capturing fluctuations in time-series variables, it could give clinicians an earlier sense of the patient's mortality status; and 2) it could be used to help hospitals to allocate ICU resources more efficiently.

In the future, our work can be extended, for example: 1) more sophisticated data preprocessing steps will be conducted to capture the characteristics of the time-series data; and 2) more extensive ICU datasets will be employed to evaluate and improve our model.

Acknowledgments

The work is supported by grants from the Natural Science Foundation of China (61300232); the China Postdoc Foundation (2015M580564); and Fundamental Research Funds for the Central Universities (lzujbky-2016-br04).

References

1. Legall, J.R., Lemeshow, S., Saulnier, F.: A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. JAMA 270, 2957--2963 (1993)
2. 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 34, 1297--1310 (2006)

3. Silva, I., Moody, G., Scott, D.J., Celi, L.A., Mark, R.G.: Predicting in-hospital mortality of icu patients: The physionet/computing in cardiology challenge 2012. Computing in Cardiology 39(20), 245--248 (2012)
4. Citi, L., Barbieri, R.: PhysioNet 2012 Challenge: Predicting mortality of ICU patients using a cascaded SVM-GLM paradigm. Computing in Cardiology 25(1), 257--260 (2012)
5. Luo, Y., Xin, Y., Joshi, R., Celi, L., Szolovits, P.: Predicting ICU Mortality Risk by Grouping Temporal Trends from a Multivariate Panel of Physiologic Measurements. In: 30th AAAI Conference on Artificial Intelligence, pp. 42--50. AAAI press, Phoenix (2016)
6. Bhattacharya, S., Rajan, V., Shrivastava, H.: ICU Mortality Prediction: A Classification Algorithm for Imbalanced Datasets. In: 31th AAAI Conference on Artificial Intelligence, pp. 1288--1294, AAAI press, San Francisco (2017)
7. Ghassemi, M., Naumann, T., Doshi-Velez, F., Brimmer, N., Joshi, R., Rumshisky, A., Szolovits, P.: Unfolding physiological state: Mortality modelling in intensive care units. In: 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 75--84, ACM press, New York (2014)
8. Carlin, C., Van Ho, L., Ledbetter, D., Aczon, M., Wetzel, R.: Predicting Individual Physiologically Acceptable States for Discharge from a Pediatric Intensive Care Unit. arXiv, preprint arXiv:1712.06214. (2017)
9. Nguyen, P., Tran, T., Venkatesh, S.: Deep learning to attend to risk in ICU. In: 2nd International Workshop on Knowledge Discovery in Healthcare Data, pp. 25--29, Morgan Kaufmann Press, Melbourne (2017)
10. Ding, J., Kang, X., Hu, X.H., Gudivada, V.: Building A Deep Learning Classifier for Enhancing a Biomedical Big Data Service. In: 2017 IEEE International Conference on Services Computing, pp. 140--147, IEEE Press, Honolulu (2017)
11. Avati, A., Jung, K., Harman, S., Downing, L., Ng, A., Shah, N.H.: Improving palliative care with deep learning. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine, pp. 311--316, IEEE Press, Kansas City (2017)
12. Lipton, Z.C., Kale, D.C., Elkan, C., Wetzel, R.: Learning to diagnose with LSTM recurrent neural networks. arXiv, preprint arXiv:1511.03677. (2015)
13. Jo, Y., Lee, L., Palaskar, S.: Combining LSTM and Latent Topic Modeling for Mortality Prediction. arXiv, preprint arXiv:1709.02842. (2017)
14. Xu, D., Li, W.J.: Full-Time Supervision Based Bidirectional RNN for Factoid Question Answering. arXiv, preprint arXiv: 1606.05854. (2016)
15. Dai, A.M., Le, Q.V.: Semi-supervised sequence learning. In: 29th International Conference on Neural Information Processing Systems, pp. 3079--3087, MIT Press, Montreal (2015)
16. Ng, Y.H., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., Toderici, G.: Beyond short snippets: deep networks for video classification. Computer Vision and Pattern Recognition, 16, 4694--4702 (2015)
17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: 26th International Conference on Neural Information Processing Systems, pp. 1097--1105, MIT Press, Lake Tahoe (2012)
18. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation, 9, 1735--1780 (1997)
19. Graves, A., Fernández, S., Schmidhuber, J.: Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition. In: 15th International Conference on Artificial Neural Networks, pp. 799--804, Springer Press, Warsaw (2005)
20. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv, preprint arXiv:1412.3555. (2014)