

Mortality Prediction for ICU Patients Using Just-in-time Learning and Extreme Learning Machine

Yangyang Ding, Xuejian Li and Youqing Wang*

Abstract—Many types of severity or prognostic scoring systems have been developed for patients in intensive care units (ICUs). They provide evaluation of patients' status so that they can get the best distribution of the intensive care. However, the accuracy and reliability of the existing systems is still not ideal. A new combination of just-in-time learning (JITL) and extreme learning machine (ELM) was proposed, aiming at improving mortality prediction accuracy. JITL was utilized to gather the most relevant data samples for patient-specific modeling while ELM was chosen for fast model building. In this study, 4000 records of ICU patients from PhysioNet database were selected, including 554 dead and 3446 survival records in which physiological parameters values were used for mortality prediction. In terms of the area under receiver-operating curve (AUC), JITL-ELM achieved the best performance, compared with ELM, BP neural network, logistic regression model and traditional score models.

I. INTRODUCTION

INTENSIVE care occupies an important position in the modern medical system, which is also an large share of the health care budget in the last decade. The evaluation of patients' status in intensive care units (ICUs) is necessary to help medical staff prioritize resources and take the appropriate diagnostic/therapeutic measures.

Advances in technology are improving the quality and quantity of physiological measurement data available in ICUs. Based on these physiological values, prediction-of-mortality models are widely used to assess patient's status. Linear model or generalized linear models are developed by using kinds of physiological variables in the past years, such as the acute physiology and chronic health evaluation system (APACHE) [1, 2], the simplified acute physiology score (SAPS) [3], the mortality probability model (MPM) [4], and so on. Some decision rules can be easily extracted from these initial linear models, which provide a great convenience for medical staff to make fast decisions. In APACHE and SAPS systems, the worst physiological values in the first 24 hours after patients entered ICU are selected to establish the logistic regression (LR) model, whereas the other data are not used, this led to the loss of information. In addition, linear model make it inaccurate to predict the patients' situations as well.

Nonlinear models, such as artificial neural networks (ANN)

[5, 6], support vector machines (SVM) [7], decision tree (DT) [8], naive Bayesian model [9], and so on, are also explored to improve the mortality prediction accuracy. More complex models are developed to portray patients' characteristics and give more accurate results. A particle swarm optimization (PSO)-based fuzzy hyper-Rectangular composite neural network (PFHRCNN) were proposed to predict the survival rate and get a better result than linear regression models and SVM [10]. Vairavan et al. proposed an algorithm based on hidden Markov model and logistic regression model [11], which also perform well.

With the increasing volume of clinical data, the above-mentioned models developed by using the whole training set are not suitable for single patient in ICUs. In fact, one patient's behavior might deviate significantly from the average model that established offline, because most patients behave in highly individual ways [12]. According to the basic idea of diagnosis, similar symptoms characterize similar results, "just-in-time learning" (JITL) method is an attractive option for individualized modeling. Thus, a new combination of JITL and extreme learning machine (ELM), termed as JITL-ELM, was proposed to improve the mortality prediction accuracy. JITL was utilized to gather the most relevant data samples for patient-specific modeling while ELM was chosen for fast model building. What's more, JITL solves the problems that sample distribution is uneven in the neural network training process to some extent.

Real physiological data of ICU patients were selected to test the performance of the JITL-ELM model, and the experimental result shows that the classification accuracy of ELM is significantly improved thanks to the addition of JITL strategy. Compared with other conventional methods, such as ELM, BP neural network, LR and SAPS-I, JITL-ELM always performs the best.

The main findings and contributions are as follows. 1) The paper presents a new method called "JITL-ELM" for ICU patients, which performs the best for improving mortality prediction accuracy compared with other conventional methods. 2) Many synthesis parameters with physiological meanings were derived as well, which make a contribution for reaching an ideal classification effect. 3) The study offers a potential application for early warning systems in the future, which is also in step with the development trend of personalized medicine. Moreover, there is still a larger space for development about JITL modeling, such as patient-specific treatment, for which this paper lays a good foundation.

The remainder of this paper is organized as follows. Data

This work was supported by National Natural Science Foundation of China (61374099) and the Program for New Century Excellent Talents in University (NCET-13-0652).

Yangyang Ding, Xuejian Li and Youqing Wang are with the College of Information Science and Technology, Beijing University of Chemical Technology, 15# Beisanhuan East Road, Chaoyang District, Beijing 100029, China (* corresponding author. Email: wang.youqing@ieee.org).

collection, feature creation, and principal component analysis are summarized in Section II; the proposed method is introduced in Section III; some experiment results are given in Section IV; finally, the conclusions are drawn in Section V.

II. DATA PROCESSING

A. Data Collection

The data used in this study comes from PhysioNet, a complex physiological signals and biomedical signal research resources website, which offers free web access to resources of recorded physiologic signals (physioBank) and related open-source software (PhysioToolkit) [13]. In this study, 4000 records of ICU patients from PhysioNet database were selected, including 554 dead and 3446 survival records, in which some measured physiological values were collected for mortality prediction.

TABLE I
LIST OF SELECTED PHYSIOLOGICAL PARAMETERS AND THEIR ABBREVIATIONS

ABBREVIATION	NAME	ABBREVIATION	NAME
HR	Heart rate (bpm) ^A	Glucose	Serum glucose (mg/dL)
BUN	Blood urea nitrogen (mg/dL)	K	Serum potassium (mEq/L)
GCS	Glasgow coma index<3-15> ^B	Mg	Serum magnesium (mmol/L)
Creatinine	Serum creatinine (mg/dL)	Na	Serum sodium (mEq/L)
DiasABP	Invasive diastolic arterial blood pressure (mmHg)	MAP	Invasive mean arterial blood pressure (mmHg)
NIMAP	Non-invasive mean arterial blood pressure (mmHg)	SysABP	Invasive systolic arterial blood pressure (mmHg)
FiO2	Fractional inspired O2 <0-1>	HCO3	Serum bicarbonate (mmol/L)
PaO2	Partial pressure of arterial O2 (mmHg)	PaCO2	Partial pressure carbon dioxide
NIDiasABP	Non-invasive diastolic arterial blood pressure (mmHg)	NISysABP	Non-invasive systolic arterial blood pressure (mmHg)
HCT	Hematocrit (%)	Temp	Temperature(°C)
PH	Arterial pH <0-14>	Urine	Urine output (mL)
Platelets	Blood platelet (cells/nL)	WBC	White blood cell count(cells/nL)

A: measurement unit;
B: measurement range;

The patients were all adults, whose stay were all more than 48 hours, and they were admitted for numerous reasons to cardiac, surgical, medical or trauma ICUs [13]. Six general descriptors were recorded on admission, e.g. “Record ID”, “Age”, “Height”, “Gender”, “Weight”, and “ICU Type”. Moreover, measured values of more than 30 physiological parameters were collected during the first 48 hours after admission to the ICU; however, not all physiological parameters are available for each patient, and the sampling frequencies for these parameters are inconsistent. Among all

these parameters, some are selected for modeling, whereas others are not. Unsurprisingly, parameters available for most patients are likely to be selected. Hence, 24 physiological parameters recorded for more than 75% patients are selected, which are shown in Table I.

In order to remove outliers and evident errors for each selected parameter which obeys the law of approximate normal distribution, Chauvenet criterion [14] is used. Then some statistics will be exacted from measured physiological values as input data set, while the records of death or not are used as the output data in the following modeling method.

B. Feature Creation

Simple statistical features are created, including minimum, maximum, mean and standard deviation of the time-series values for each selected parameter [15]. Moreover, many synthesis parameters with physiological meanings were derived as well, for example,

- abs(dHR) : the absolute value of the continuous changes in heart rate measurement, and it is related to many cardiovascular diseases [15].
- HR/BP : HR to BP ratio, which is known as an early indicator of cardiogenic, hypovolemic, and septic shock [16];
- HR*BP : rate pressure product, which reflects the situation of cardiac oxygen consumption [17];
- ECO : Estimated Cardiac Output. Cardiac output is an important index for circulatory system efficiency, defined as: $\text{HR}*(\text{SysABP} - \text{DiasABP})/\text{MAP}$ [18];

Here BP represents three variables, namely DiasABP, SysABP and MAP shown in Table I. In the same way, statistical features (minimum, maximum, mean and standard deviation of the time-series values) for each synthesis parameters are extracted in this procedure. The number of extracted features is 137 in total.

What is worth mentioning, in the procedure of creating the synthesis parameters, only the values that measured at the same moment are calculated, e.g. when calculating a series of HR*SysABP for a patient, the value of HR and SysABP should be measured at the same time. Otherwise the physiological meanings of synthesis parameters cannot be guaranteed.

C. Principal Component Analysis

Missing information is replaced by the mean values of the other patients with the same ICU type and age group. That means the descriptor parameters “ICU Type” and “Age” are used as a reference variable when running in the missing values. Because not all parameters are available for each patient and the parameters are measured using different sampling rate, data interpolation work in this study is conducted on “features” directly rather than on the raw data.

Principal component analysis (PCA) is most commonly used for dimension reduction. Taking into account the correlations between variables or properties, PCA can extract features without a significant loss of information [19].

In order to reduce the computational complexity and avoid the influence of useless features, PCA is adopted as

pre-processing before features are fed into the model. Moreover, in order to make full use of information as possible, 60 principal components are selected, which would be the input data of JITL-ELM model in the next step.

III. METHODS

A. ELM

In the past years, ELM has proved itself being a good candidate for classification and regression. As a new learning algorithm of single hidden layer feed forward neural network, ELM has simpler structure, less parameter adjustment, as well as the advantages of low computational complexity and fast training speed. To make this paper self-contained, a brief introduction of ELM will be presented in this section. The detailed information of ELM could be found in [20].

ELM does not tune but randomly generates the weights between the input and hidden layer. In order to estimate the output weights of hidden layer, ELM offers a simple mathematical way rather than identify the best parameters with iterative tuning like the traditional methods. Instead, ELM employed a direct solution that solving a norm least squares problem, which is converted to solving the generalized inverse of matrix. The procedure of simple ELM is as follows:

First, the training data set is given by

$$\{(x_j, t_j) | x_j \in R^d, t_j \in R^m, j = 1, L, N\};$$

Then ELM model can be described as

$$\sum_{i=1}^L \beta_i g(x_j) = \sum_{i=1}^L \beta_i g(w_i x_j + b_i) = t_j, j = 1, 2, L, N$$

where $w_i = (w_{i1}, w_{i2}, \dots, w_{iL})^T$ indicates the weight vectors between the input and hidden layer; $\beta = [\beta_1^T, \beta_2^T, \dots, \beta_L^T]^T$ is the weight vector between the hidden and output layer. L denotes the number of hidden nodes.

Step 1. Randomly generate the weight and offset vectors

$$w_i = (w_{i1}, w_{i2}, \dots, w_{iL})^T, i = 1, 2, L, N$$

$$b_i = (b_1, b_2, \dots, b_L)^T$$

Step 2. Compute the output matrix H of hidden layer

$$H(w_1, w_2, \dots, w_L; b_1, b_2, \dots, b_L; x_1, x_2, \dots, x_N) = \begin{bmatrix} g(w_1 x_1 + b_1) & g(w_2 x_1 + b_2) & \dots & g(w_L x_1 + b_L) \\ g(w_1 x_2 + b_1) & g(w_2 x_2 + b_2) & \dots & g(w_L x_2 + b_L) \\ \vdots & \vdots & \ddots & \vdots \\ g(w_1 x_N + b_1) & g(w_2 x_N + b_2) & \dots & g(w_L x_N + b_L) \end{bmatrix}_{N \times L}$$

$$Y = [Y_1^T, Y_2^T, \dots, Y_N^T]^T_{M \times N}$$

Step 3. It is clear that the only unknown variable is β , which can be calculated by

$$H\beta = T \quad (1)$$

$$\|H\hat{\beta} - T\| = \min_{\beta} \|H\beta - T\| \quad (2)$$

$$\text{when } N > L, \hat{\beta} = \left(\frac{1}{C} + HH^T \right)^{-1} H^T Y \quad (3)$$

$$\text{when } N < L, \hat{\beta} = H^T \left(\frac{1}{C} + HH^T \right)^{-1} Y \quad (4)$$

where $\hat{\beta}$ is the least square norm solution [21, 22] refer to (1). As can be seen in the two formulas above, a positive value $1/C$ is added to the diagonal of HH^T or $H^T H$ in order for better generalization performance. Appropriate formula can be selected according to the size of training data.

In general, N is far larger than L so that formula (3) is often used in most cases, so it is with this paper. What's more, if the weight of each training sample is considered in the ELM modeling process [23], then $\hat{\beta}$ can be described as

$$\hat{\beta} = \left(\frac{1}{C} + H^T W H \right)^{-1} H^T W Y \quad (5)$$

ELM offers a new idea for the development of neural network technology. With slightly improved performance but a significantly faster learning speed, ELM have certain advantages compared with conventional learning techniques for feed forward neural networks [24]. Furthermore, unlike the results of general back-propagation learning, ELM have a better generalization ability, and its structure is easy to determine only by setting the number of hidden layer nodes, thus ELM is adopted when establishing the mortality prediction model.

B. JITL-ELM

The idea of JITL is to obtain the prediction result of the current testing patient with the k most relevant/similar samples collected from the database. The forecast is operated for patients one by one, namely the model changes for each patient. The idea of JITL is introduced as follows:

Given the data set $(X_i, y_i)_{i=1}^N$, and their relationship is

$$y_i = f(X_i, \theta) + \varepsilon_i \quad (6)$$

where $f(\cdot, \cdot)$ indicates the functional relationship between the input and output data, θ represents the coefficient vector of X_i ; ε_i is the modeling error which satisfies a Gaussian distribution.

Based on the current testing sample data, the similar samples will be gathered from the history database, which will be used to establish ELM model. Thus the prediction for the current testing patient will be obtained according to the established ELM model. This process will be repeated when new testing samples is read in, and it is converted into solving the following optimization problem

$$\theta^* = \arg \min_{\theta} \sum_{(X_i, y_i) \in \Omega} (y_i - f(X_i, \theta))^2 w_i \quad (7)$$

$$w_i = \exp \left(\frac{-\|X_i - X_q\|^2}{2\sigma^2} \right) \quad (8)$$

where Ω indicates the data set or domain space composed of k similar samples; w_i denotes the corresponding weight according to Gaussian Kernel Function refer to (8), which reflects the influence on the output value that similar samples exert; σ reflects the width parameters of the kernel function.

According to the idea "similar inputs produce similar output", the more similar to the current input data, the more

information can be reflected about output characters, and the corresponding w_i is larger relatively.

The domain space Ω can be described as

$$\Omega = (X_1, X_2, \dots, X_k) = \{X_i \mid d(X_i, X_q) < h\} \quad (9)$$

where h indicates the radius of domain space, determining the size of the similar data set; $d(\cdot, \cdot)$ indicates the distance function, which is used as a measurement index of the degree of similarity, X_1, X_2, \dots, X_k indicates the most similar k samples inside Ω .

Similar samples obtained by neighborhood searching have a great influence on the prediction accuracy of JITL models. It can be accomplished by setting a proper value for h , which is hard to realize. If h is too large, it will increase the computational difficulty of the algorithm; but if h is too small, the generalization ability of the model will decline.

The basic idea of JITL shows the key problems need to be solved.

- Selection of effective sample information when modeling.
- Determine the size of Ω or the value of h .
- Determination of w_i .
- Choice of modeling method.

Solutions in this study are given as follows:

- Combine Euclidean distance with angle distance, which are shown below. Only samples satisfying constraint conditions can be selected, and the conditions will be introduced in the following part.

$$d(X_q, X_i) = \sqrt{(X_q - X_i)^T (X_q - X_i)}$$

$$\cos(\theta_i) = \frac{X_q^T X_i}{\|X_q\|_2 \|X_i\|_2}$$

- It is difficult to determine the size of domain space Ω or h , whereas the number of similar samples k can be tuned artificially based on the distance function.
- All weights are set to 1 in this study, namely all similar samples searched in the previous step have the same influence on the output of the model.
- ELM was selected as a modeling method because of its good generalization ability and fast training speed.

The procedure of selecting similar samples using the combination of Euclidean distance and angle distance is as follows: k similar samples, which have the k smallest Euclidean distance, would be selected first. However, samples obtained here will also be deleted if they do not satisfy the angle constraint, namely the angle distance should be larger than a certain value which is set by 0.7 in this study. Then the rest samples obtained can be used as training set for ELM modeling. Moreover, if the size of the similar sample data set satisfying the conditions is too small to yield reliable results, 80 samples with the smallest Euclidean distance will be chosen, and meanwhile its corresponding angle distance should be larger than zero. As shown in Figure 1, JITL-ELM modeling steps are described as follows.

Step 1. Divide the 4000 records into two parts, that is, the training sample library and the testing samples, in a ratio of 4:1 chosen in this study.

Step 2. Read in one testing sample every time, and then its

k most similar samples will be collected in the training sample library, serving as the modeling data in the next step.

Step 3. For the current testing patient, ELM model will be established using its similar samples.

Step 4. Input the current testing data to the established model to obtain the corresponding predicted value. Repeat Step2 if there is new testing sample data.

Fig. 1 shows the procedure of JITL-ELM model. Testing samples are read in one by one, and ELM model will be established for each patient trained by its corresponding relevant samples. That is to say, the old ELM model will be discarded when new sample is read in.

JITL strategy has a similar idea to the process of diagnosing disease. Doctors often use their professional knowledge and experiences to diagnose diseases or to do a general judgment by observing the patient's symptoms, which is the same to the procedure of searching for the similar cases. Moreover, the "symptoms" here can be understood as "features" that introduced in section 2.

JITL belongs to the local modeling method which means that the training samples data is partly utilized for modeling. Only the similar data is collected for use according to the distance function. Compared with the traditional modeling method, JITL makes the model simpler, and the simpler model will be discarded after giving the corresponding forecast for the current testing sample. When new testing sample came in, a new similar data set would be gathered again. This idea is similar to doctors' diagnosis process, and it also reflects the idea "similar illness produces similar symptoms most likely".

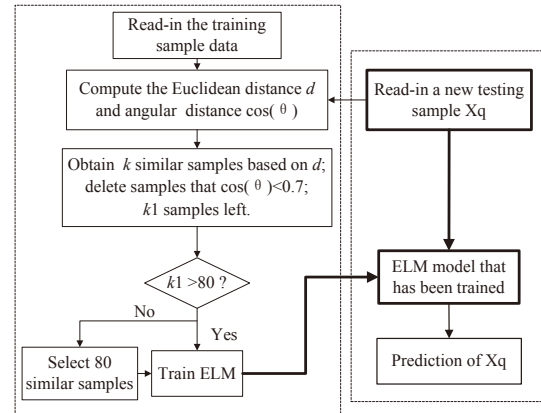


Fig.1. JITL-ELM model procedure

JITL can solve the low prediction accuracy problem caused by the distribution imbalance of training samples. For the artificial neural network, if the number of survival/dead records is 5 times or more than the other one, both of the training and testing accuracy will be affected unsurprisingly. However, JITL provides a method by establishing a local model. For survival patients, which are the majority group in this study, only a few dead samples are contained in its similar data set, which do very little contribution to the prediction result. For the testing patients who are dead, similar data set contains two classes, and the imbalance of which is significantly reduced. It will make little difference if the number of similar samples k is determined, appropriately.

TABLE II
CLASSIFICATION PERFORMANCE RESULTS USING DIFFERENT MODELING METHODS

	JITL-ELM								ELM	LR	BP	SAPS-I
	80 ^A	100	200	300	400	500	800	1000				
AUC	0.8000	0.8067	0.8123	0.8139	0.8151	0.8169	0.8174	0.8177	0.7434	0.7847	0.7190	0.6668
SD	0.0172	0.0170	0.0146	0.0172	0.0149	0.0165	0.0162	0.0166	0.0258	0.0216	0.0151	\

A: the similar sample number k in JITL-ELM model; AUC: area under receiver-operating curve; SD: standard deviation.

IV. RESULTS

Cross validation (CV) is widely used to evaluate the performance of models, which can eliminate consequences that rely on the dataset. The implementation procedures are as follows. First, the whole dataset is divided into m partitions arbitrarily without overlapping. Then one group is selected from the m partitions as the testing dataset and the other $m-1$ groups are the training datasets. The procedure will be repeated until all m partitions have been served as the testing dataset to evaluate the pertinent performance [25]. In this study, samples are divided into five parts equally and arbitrarily.

In this section, the performance of JITL-ELM is compared with some traditional algorithms, i.e., ELM, LR, BP neural network, LR. In addition, SAPS-I is also selected as a comparative method, because it is a kind of indicator to patients' status in hospital. The validation was carried out in MATLAB 2014, where the toolboxes of BP network and logistic regression are used.

A. Receiver Operating Curve

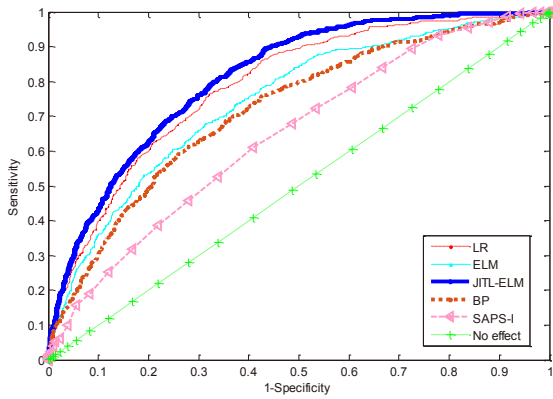


Fig. 2. ROC curves for ELM, BP, LR, JITL-ELM, and SAPS-I scoring system. 400 similar data was selected for the current testing data when using JITL-ELM model here.

Figure 2 shows the receiver-operating curves (ROC) when using different methods, i.e., ELM, BP, LR, JITL-ELM algorithm and SAPS-I Scoring system. When using the JITL-ELM model, 400 similar samples data was selected for the current testing sample data. It is intuitive that the performance became better after adding JITL strategy compared with the pure ELM model, and JITL-ELM performs best among all these methods. Over-fitting occurred when using BP algorithm, which decreased its accuracy.

B. AUC Index

AUC index, referring to the area under the ROC curve, is often used to evaluate the performance of a binary classifier quantitatively. Results of the classification using the above-mentioned algorithms are shown in Table II, in which lists the mean values and standard deviation (SD) values of the AUC index. The tiny standard deviation demonstrates that the results of five parts in cross validation are all reliable. It suggests that the performance of JITL-ELM became better when more similar samples were selected, but did not improve much.

To explain, when the similar sample number k is larger than about 500, the increasing of AUC basically depends on the improvement of prediction accuracy for survival patients, but the distribution unbalance will occur easily in the similar sample data set, which may decrease the training performance. Among the samples of 4000 patients, only 554 dead records are provided, accounting for only 13.85%. After selecting similar samples according to Euclidean distance, samples whose corresponding angle constraint distances smaller than 0.7 would be deleted. That means the number of similar samples for testing data especially the dead testing data actually is smaller than k . Furthermore, the prediction accuracy for dead samples will decrease when k becomes larger, which has less effect on the AUC index than survival testing records. Due to the prediction accuracy for survival samples was increased mainly, AUC index was improved, but the result of increase is tiny which is shown in Table II.

As far as the AUC index is concerned, the performance of ELM shows a 9.64% percent increase thanks to the combination of JITL, and JITL-ELM raises 3.87%, 13.37% and 22.2% than LR, BP and SAPS-I, respectively.

C. Sensitivity and Specificity

Sensitivity and specificity are indicators of classification accuracy. Sensitivity indicates what proportion can be correctly detected in the real survival sample data set; specificity indicates what percentage has been figured out for the real dead samples. In addition, the abscissa and ordinate of ROC are *1-specificity* and *sensitivity*, respectively.

Table III shows the results of classification, when the target of sensitivity is set to be 0.90 and compute the corresponding specificity of different methods, the specificity of JITL-ELM performs the best with the value 0.5554, which performs better than the others. An optimal tradeoff was considered for JITL-ELM with a specificity and sensitivity of 0.7002 and

0.7617 respectively, and JITL-ELM performs best among these method list on Tables II and III.

TABLE III
THE COMPARISON OF SENSITIVITY AND SPECIFICITY USING DIFFERENT METHODS

	JITL_ELM	ELM	Logistic	BP	SAPS-I
Sp1	0.7002	0.7002	0.7002	0.6245	0.6002
Sn1	0.7617	0.6245	0.7202	0.7209	0.6119
Sp2	0.5554	0.3340	0.4920	0.3342	0.2522
Sn2	0.9000	0.9000	0.9000	0.9000	0.9000

Sp: specificity;
Sn: sensitivity;

V. DISCUSSION AND CONCLUSIONS

In this study, a new method combining just-in-time learning (JITL) and extreme learning machine (ELM) was proposed, which can improve the mortality prediction accuracy. JITL was used to gather the most relevant samples for patient-specific modeling while ELM was chosen for fast model building. 4000 records of ICU patients from PhysioNet database were selected, including 554 dead and 3446 survival records in which physiological parameters measured values were used for mortality prediction. In terms of the area under receiver-operating curve (AUC), JITL-ELM achieved the best performance, compared with ELM, BP neural network, logistic regression model, and traditional score models.

For an optimal tradeoff of specificity and sensitivity, the AUC index of JITL-ELM model is 0.8151, with a specificity and sensitivity of 0.7002 and 0.7617, respectively. Compared with the commonly used scoring system in hospitals, JITL-ELM model presented in this study significantly improved the classification accuracy.

The proposed JITL-ELM model offers a general framework in which JITL can search for a better domain space for testing samples, while ELM provides an advanced machine learning method to get better prediction results for ICU patients. With high adaptability and specificity, this kind of method can better achieve the aim to monitor individual patients. Furthermore, it has a potential application value for early warning systems in the future, which is also in step with the development trend of personalized medicine.

REFERENCES

- [1] Knaus WA, Draper EA, Wagner DP, Zimmerman JE. "APACHE II: a severity of disease classification system". *Critical Care Medicine*. vol.13, no. 10. pp. 818-829. 1985.
- [2] Knaus WA. "APACHE 1978-2001: the development of a quality assurance system based on prognosis: milestones and personal reflections". *Archives of Surgery*. vol.137, no. 1. pp. 37-41. 2002.
- [3] Le GJ, Lemeshow S, Saulnier F. "A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study". *JAMA the Journal of the American Medical Association*. vol.270, no. 24. pp. 2957-2963. 1993.
- [4] Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. "Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients". *JAMA the Journal of the American Medical Association*. vol.270, no. 20. pp. 2478-2486. 1993.

- [5] Xia H, Daley BJ, Petrie A, Zhao X. "A neural network model for mortality prediction in ICU". *Computing in Cardiology (CinC)*. pp. 261 - 264. 2012;
- [6] Silva Á, Cortez P, Santos MF, Gomes L, Neves J. "Mortality assessment in intensive care units via adverse events using artificial neural networks". *Artificial Intelligence in Medicine*. vol.36, no. 3. pp. 223-234. 2006.
- [7] Chan CL, Chen CL, Ting HW. "An excellent mortality prediction model based on support vector machine (SVM)-a pilot study". *International Symposium on Computer Communication Control & Automation*. vol.1. pp. 134 - 137. 2010.
- [8] Lee S, Lee S, Park Y. "A prediction model for success of services in e-commerce using decision tree: E-customer's attitude towards online service". *Expert Systems with Applications*. vol.33, no. 3. pp. 572-581. 2007.
- [9] Rynänen O-P, Soini EJ, Lindqvist A, Kilpeläinen M, Laitinen T. "Bayesian predictors of very poor health related quality of life and mortality in patients with COPD". *BMC Medical Informatics & Decision Making*. vol.13, no. 1. pp. 1-10. 2013.
- [10] Hsieh YZ, Su MC, Wang CH, Wang PC. "Prediction of survival of ICU patients using computational intelligence". *Computers in Biology & Medicine*. vol.47, no. 1. pp. 13-19. 2014.
- [11] Vairavan S, Eshelman L, Haider S, Flower A, Seiver A. "Prediction of mortality in an intensive care unit using logistic regression and a hidden Markov model". *Computing in Cardiology (CinC)*; vol.39. pp. 393-396. 2012;
- [12] Ying ZA, Peter SA. "Patient-specific learning in real time for adaptive monitoring in critical care". *Journal of Biomedical Informatics*. vol.41, no. 3. pp. 452-460. 2008.
- [13] Mohammed S, Mauricio V, Reisner AT, et al. "Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database". *Critical Care Medicine*. vol.39, no. 5. pp. 952-960. 2011.
- [14] Pagurova VI. "On Chauvenet's test for finding several outliers". *Theory of Probability & Its Applications*. vol.30, no. 3. pp. 558-561. 1985.
- [15] Cao H, Eshelman L, Chbat N, Nielsen L, Gross B, Saeed M. "Predicting ICU hemodynamic instability using continuous multiparameter trends". in *Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society*; vol.2008. pp. 3803-3806. 2008;
- [16] Rady MY, Smithline HA, Blake H, Nowak R, Rivers E. "A comparison of the shock index and conventional vital signs to identify acute, critical illness in the emergency department". *Annals of Emergency Medicine*. vol.24, no. 4. pp. 685-690. 1994.
- [17] Gobel FL, Norstrom LA, Nelson RR, Jorgensen CR, Wang Y. "The rate-pressure product as index of myocardial oxygen consumption during exercise in patients with angina pectoris". *Circulation*. vol.57, no. 3. pp. 549-556. 1978.
- [18] Saeed M. "Temporal pattern recognition in multiparameter ICU data". *Massachusetts Institute of Technology*. 2007.
- [19] Yoon H, Park CS, Kim JS, Baek JG. "Algorithm learning based neural network integrating feature selection and classification". *Expert Systems with Applications*. vol.40, no. 1. pp. 231-241. 2013.
- [20] Huang GB, Zhu QY, Siew CK. "Extreme Learning Machine: Theory and Applications". *Neurocomputing*. vol.70, no. 1-3. pp. 489-501. 2006.
- [21] Healy MJR. "Generalized Inverse of Matrices and Its Applications by C. R. Rao; S. K. Mitra". *Operational Research Quarterly (1970-1977)*. vol.23, no. 4. 1972.
- [22] Serre D. "Matrices : theory and applications". *Mathematics*. no. 32. pp. xvi,221. 2010.
- [23] Zong W, Huang GB, Chen Y. "Weighted extreme learning machine for imbalance learning". *Neurocomputing*. vol.101, no. 3. pp. 229-242. 2013.
- [24] Huang GB, Wang D. "Advances in extreme learning machines (ELM2010)". *Neurocomputing*. vol.74, no. 16. pp. 2411-2412. 2011.
- [25] Golub GH, Heath M, Wahba G. "Generalized cross-validation as a method for choosing a good ridge parameter". *Technometrics*. vol.21, no. 21. pp. 215-223. 1979.