Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Length of stay prediction for ICU patients using individualized single classification algorithm

Xin Ma[a], Yabin Si[a], Zifan Wang[a], Youqing Wang[a,b,*]

[a] *Beijing University of Chemical Technology, China*
[b] *Shandong University of Science and Technology, China*

ARTICLE INFO

ABSTRACT

Background and Objective: In intensive care units (ICUs), length of stay (LOS) prediction is critical to help doctors and nurses select appropriate treatment options and predict patients' condition. Considering that most hospitals use universal models to predict patients' condition, which cannot meet the individual needs of special ICU patients. Our goal is to create a personalized model for patients to determine the number of hospital stays. *Methods:* In this study, a new combination of just-in-time learning (JITL) and one-class extreme learning machine (one-class ELM) is proposed to predict the number of days a patient stays in hospital. This combination is shortened as one-class JITL-ELM, where JITL is used to search for personalized cases for a new patient and one-class ELM is used to determine whether the patient can be discharged within 10 days. *Results:* The experimental results show that the one-class JITL-ELM model has an area under the curve (AUC) index of 0.8510, lift value of 2.1390, precision of 1, and G-mean is 0.7842. Its accuracy, specificity, and sensitivity were found as 0.82, 1, and 0.6150, respectively. Moreover, a novel simple mortality risk level estimation system that can determine the mortality rate of a patient by combining LOS and age is proposed. It has an accuracy rate of 66% and the miss rate of only 6.25%. *Conclusions:* Overall, the one-class JITL-ELM can accurately predict hospitalization days and mortality using early physiological parameters. Moreover, a simple mortality risk level estimation system based on a combination of LOS and age is proposed; the system is simple, highly interpretable, and has strong application value.

© 2019 Published by Elsevier B.V.

1. Introduction

Intensive care units (ICUs) have always played an important role in hospitals [1–7]. They are well equipped and run by a large number of medical workers; therefore, accurate assessment of the physiological status of patients in ICUs is beneficial for the rational allocation of medical resources, while providing major safeguards for the treatment of patients. Moreover, although the longer patients stay in the hospital, the more their treatment, it does not necessarily favor their prognosis [8–16]. According to [17], mortality rate increases after a prolonged stay in the ICU (over 30 days), and over 70% of patients die in hospital or within one year of treatment. Hence, it should not be neglected that lengthy hospitalization time also reflects on the severity of the patients' condition.

Furthermore, hospitalization cost is a financial burden on the family [18]. The cost of intensive care is increasing yearly. When patients struggle with a disease, their family members also suffer from both mental and economic pressures. Therefore, for both doctors and patients, LOS estimation is a vital practical concern.

With the advancements in medical field, some scoring systems have been developed to judge the physiological parameters of ICU patients, such as the acute physiology and chronic health evaluation scoring system [19], simplified acute physiology score (SAPS) [20], and the mortality probability model [21]. The systemic inflammatory response score and early warning score are used to assess the severity of sepsis in ICU patients. The quick sofa score for predicting mortality and LOS has a high internal validity [22].

Following the development of medical big data processing, a large number of machine learning technologies have been applied to ICU. To better detect and predict patients' situation, artificial neural networks [23,24], support vector machine (SVM) [25–29], extreme learning machine (ELM) [30–33], and other methods can estimate the future situation only according to the patients physiological indicators.

In the current stage of intelligent medical treatment, artificial intelligence (AI) technology as a vital role is integrated into the whole process of medical health [34]. In clinical work, doctors'
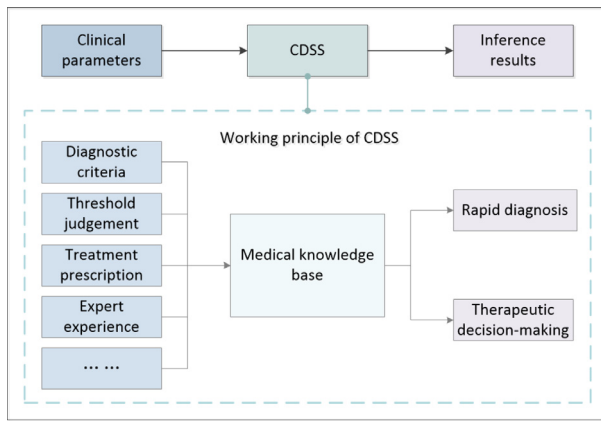
Fig. 1. Working principle of CDSS.

Table 1
Physiologicaldataset of two sample patients.

RecordID=132554				Final State: 0		LOS: 17 days
Time	HR	Temp	GCS	...		NIDiasABP
0:00	−1	−1	−1	...		−1
3:52	127	37.3	15	...		71
4:27	127	−1	−1	...		71
5:27	130	−1	15	...		69
6:37	128	−1	−1	...		71
7:27	128	−1	15	...		62
...	...	...	...	...		...
46:27	130	−1	−1	...		47
47:27	122	15	15	...		53
RecordID=132555				Final State: 1		LOS: 8 days
Time	HR	Temp	GCS	...		NIDiasABP
0:00	−1	−1	−1	...		−1
0:20	−1	−1	−1	...		−1
0:57	67	34.8	−1	...		−1
1:12	68	−1	−1	...		−1
1:16	−1	−1	−1	...		−1
1:27	72	34.9	−1	...		−1
...	...	...	...	...		...
46:57	99	−1	−1	...		57
47:57	78	−1	−1	...		49

−1 denotes the missing value.

decision-making runs through the whole process of clinical practices. When facing one patient, the doctor must make decisions successively according to the patient's condition characteristics and examination results and strive to make the most accurate diagnosis. As a virtual assistant to doctors, the clinical decision support system (CDSS) takes clinical data as input information and outputs inference results. It not only considers the prediction accuracy but also takes into account patients personalized diagnosis and treatment methods. The basic principle of the CDSS is to construct knowledge bases for various diseases; input diagnostic criteria, estimated threshold, treatment prescriptions, and expert experience of various diseases into the computer; use the superstrong and accurate information storage, extraction function, and fast computing ability of the computer; simulate doctors' diagnostic and therapeutic thinking through AI technology and computer logical reasoning operation; and help doctors make quick decisions (Fig. 1.) This study is focused on enriching the components of the CDSS system, using ICU patient's hospitalization days as an indicator to help doctors judge their condition.

Due to the specificity of patients in the ICU, this study fully considers the estimation of individual patient's LOS. Just-in-time learning is used to select from the database the case closest to that of a patient, and one-class ELM is used to personally classify the patient to accurately determine if the patient can be discharged within 10 days. In this paper, the patient's information in the test data is considered as historical data. From the historical data, the JITL idea is used to select the most similar 300 patient cases for each new patient to establish a online personalized model. This meets the needs of contemporary patients for personalized medicine. In feature selection, we use the two-step principal component analysis (TS-PCA) algorithm to extract the most essential static features from the patient's dynamic information, making the classification more accurate.

The results show that the one-class JITL-ELM model has an AUC index of 0.8510, lift value of 2.1390, precision of 1, and G-mean of 0.7842, while, its accuracy, specificity, and sensitivity are 0.82, 1, and 0.6150, respectively. Moreover, a simple innovative system for mortality risk level estimation is proposed; the system estimates the mortality rate of patients by combining their LOS and age. It has an accuracy rate of 66% and miss rate of only 6.25%.

Overall, the main contributions of this study are as follows: (1) A personalized model to predict whether ICU patients can be discharged after 10 days is established. (2) The model allows for personalized search to find similar cases, which greatly saves doctors' time, improves their efficiency, and saves hospital medical resources. (3) Regarding the binary classification problem as a single classification problem, the model can predict the patient's condi-

tion more accurately. (4) According to the discharge situation, a fast scoring method that can predict the mortality risk of patients is proposed.

The remaining part of this paper is organized as follows: Section 2 summarizes data processing, TS-PCA and relevant method; Section 3 presents illustrative experiments and results; Section 4 discusses the results of the experiment and introduces the proposed fast and effective scoring method for predicting mortality; finally, Section 5 presents the conclusions.

2. Methods

2.1. Data collection

The physiological data of ICU patients needed for this study were obtained from PhysioNet website [35]. The website allows access to complex physiological signals and real biomedical signal research resources to help scholars conduct in-depth research, and the physiological data in this website can be used free of charge. Many scholars use this authoritative website because of its scientific and rigorous information.

As shown in Table 1, the dataset contained records of 4000 adult patients, including their basic information (record ID, age, height, weight, and ICU type), the physiological parameters of each ICU patient in the first 48 hours after their admission to the ICU, final survival status (0 = survivor, 1 = died), and days of hospitalization.

However, the sampling frequencies of these physiological parameters are largely different, and some parameters are not practical for predicting the hospitalization days of patients due to too many missing values, such as HCT, BUN, Creatinine, Glucose, HCO$_3$, Mg, Platelets, K, Na, WBC, PaO$_2$, SaO$_2$, Albumin, ALP, ALT, AST, Bilirubin, Lactate, Cholesterol, TroponinI and TroponinT. Finally, the 15 physiological indicators retaining the most complete information are used for modeling, as shown in Table 2.

The LOS is closely related to mortality. We use mortality as a dividing criterion, as shown in Table 4. The table shows the number of survived patients who were not discharged within the specified number of days. The data show that as the number of hospital stays increases, the risk of death also increases. As shown in Fig. 2, patient hospitalization days follow a positive skew distribution. In medical statistics, due to the asymmetry of the original data dis-

Table 2
List of selected physiological parameters and abbreviations.

Abbreviation	Name
Temp	Temperature (°C)
HR	Heart rate (bpm)
Urine	Urine output (mL)
PH	Arterial pH <0–14>
RespRate	Respiration rate (bpm)
GCS	Glasgow coma index<3–15>
FiO$_2$	Fractional inspired O$_2$ <0–1>
PaCO$_2$	Partial pressure carbon dioxide
MAP	Invasive mean arterial blood pressure (mmHg)
SysABP	Invasive systolic arterial blood pressure (mmHg)
DiasABP	Invasive diastolic arterial blood pressure (mmHg)
NIMAP	Non-invasive mean arterial blood pressure (mmHg)
NIDiasABP	Non-invasive diastolic arterial blood pressure (mmHg)
MechVent	Mechanical ventilation respiration (0: false, or 1: true)
NISysABP	Non-invasive systolic arterial blood pressure (mmHg)
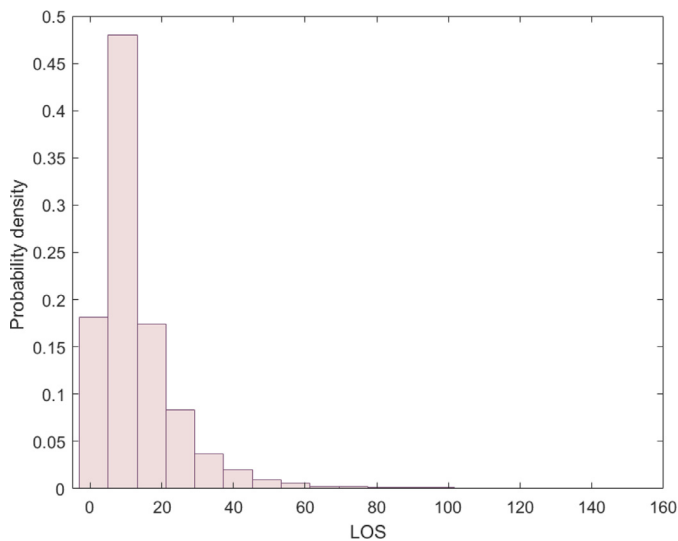
Fig. 2. The probability density of patients' LOS.

tribution, the geometric mean of the data is often selected. The formula is as follows:

$$G = \sqrt[n]{L_1 L_2 L_3 \ldots L_n} \qquad (1)$$

After calculation, the geometric mean of the LOS is 10.4238 days. At the same time, the median of this group of data is also 10 days. The median is the number in the middle of a set of data arranged in order, which is a very representative medical statistical feature [36,37]. To facilitate statistical analysis, 10 days is chosen as the criterion for this trial.

2.2. Data interpolation

In this study, the sampling frequency of the physiological data selected was different. This difference existed not only among different parameters but also for each single parameter. To interpolate the missing data while maximally retaining the time information, the irregular ICU physiological variables for each sampling period were converted into a time series of periodic variables sampled every two hours [38].

The data processing procedure was as follows: First, the physiological indicator was divided into 24 segments at a two-hour interval. In each part, if multiple numbers exist, their average was taken. If there was no record within two hours, the information was interpolated at the mean of this indicator. If the patient's indicator was completely missing, all patients values in the database
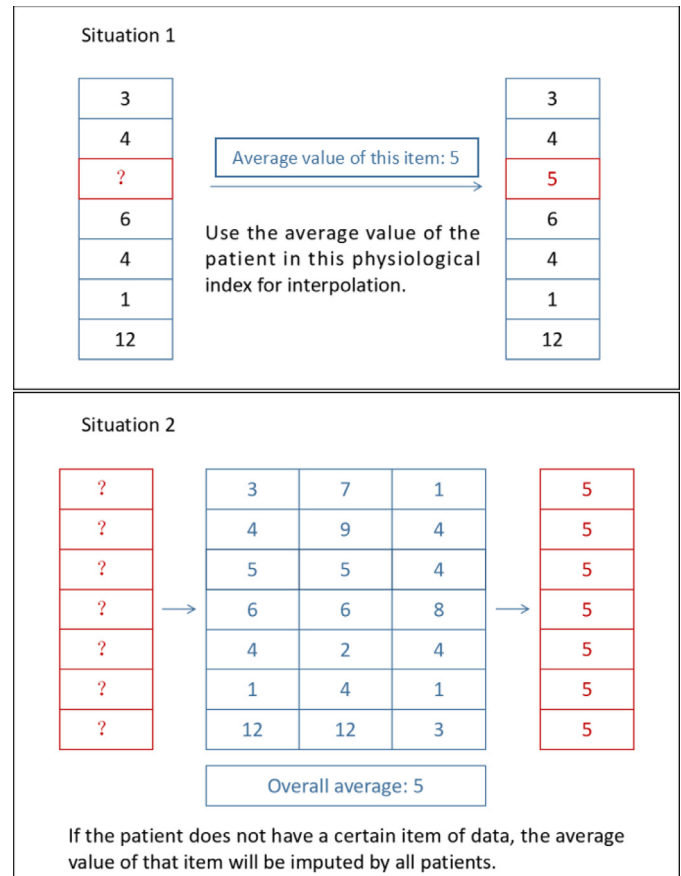
Fig. 3. Data interpolation rules.

Table 3
Survival of undischarged patients.

Day	5	7	10	14
Undischarged - death	421	346	259	195
Undischarged - survival	2806	2226	1582	1142
Mortality	0.1305	0.1345	0.1407	0.1458

Table 4
Comparison with unsupervised algorithms.

	One-class JITL-ELM	Kmeans	iForest
Sn	0.6149	0	0.1497
Sp	1	0.9953	0.8460
Pv+	1	0	0.4598
G-mean	0.7842	0	0.3553
Lift	2.1390	0	1.0084
Acc	0.82	0.53	0.5205
AUC	0.8510	0.4624	0.4777

were interpolated at the mean of the item. In order to better explain this rule, we use a simple numerical example to illustrate the detailed calculation procedure, shown in Fig. 3.

Finally, 360 medical information features were extracted. These features are derived from the 15 physiological indicators in Table 3, each of which takes 24 time periods, and the features were used for subsequent processing and modeling.

2.3. Two-step principal component analysis

Directly using the extracted 360-dimensional features for classification will not only result in a large amount of calculation but also a low accuracy. Therefore, to improve the accuracy, the dimen-

sionality of the data needs to be reduced while retaining the most critical information.

Principal component analysis is a commonly used dimensionality reduction algorithm, but the traditional PCA algorithm is limited as it assumes that the data sampled at each time is independent; that is, the current data will not be affected by historical data. Therefore, PCA only considers the correlation among variables at the same time but does not consider the interaction of variables at different times. In other words, the traditional PCA algorithm is only a static modeling method. However, physiological processes show obvious dynamic characteristics; that is, the sampling data at different times are time-dependent. For such a process, the static PCA algorithm is not applicable.

In this study, we use a new data dimensionality reduction method called TS-PCA to process dynamic data [39]. The main idea of the TS-PCA is extracting the innovation components. The innovation components refer to the independent driving force introduced in the process. Ordinarily, innovation components are time-independent and only contain cross-correlation characteristics. Hence, the traditional PCA may be suitable for innovation components.

The specific algorithm of TS-PCA is divided into two steps: the first step is to estimate the dynamic structure matrix and extract the innovation components $\mathbf{U}^{org}(t)$ in the data; the second step is to reduce the dimension of the innovation components $\mathbf{U}^{org}(t)$ using the traditional PCA method.

In the first step, since the expectation of the innovation component $\mathbf{U}^{org}(t)$ is not 0, it is necessary to centralize it first. Since the expectations at different times $\mathbf{U}^{org}(t)$ are the same, the expected value of the difference is 0 after the $\mathbf{U}^{org}(t)$ is subtracted at two different times. To perform differential processing on two data at different times, the equation is expressed as:

$$
\begin{aligned}
\triangle \mathbf{X}^{org}(t) &= \mathbf{X}^{org}(t) - \mathbf{X}^{org}(t-D) \\
&= \left(\tilde{\mathbf{X}}^{org}(t) - \tilde{\mathbf{X}}^{org}(t-D) \right) \tilde{A} \\
&\quad + \left(\mathbf{U}^{org}(t) - \mathbf{U}^{org}(t-D) \right) \\
&= \triangle \tilde{\mathbf{X}}^{org}(t) \tilde{A} + \triangle \mathbf{U}^{org}(t)
\end{aligned}
\tag{2}
$$

where,

$$
\begin{aligned}
\mathbf{X}^{org}(t) &= \mathbf{X}^{org}(t-1)\mathbf{A}_1 + \mathbf{X}^{org}(t-2)\mathbf{A}_2 + \\
&\quad \cdots + \mathbf{X}^{org}(t-q)\mathbf{A}_q + \mathbf{U}^{org}(t) \\
&= \tilde{\mathbf{X}}^{org}(t)\tilde{A} + \mathbf{U}^{org}(t)
\end{aligned}
\tag{3}
$$

The formula can also be written as a matrix form as follows:

$$
\triangle \mathbf{X} = \triangle \tilde{\mathbf{X}} \tilde{A} + \triangle \mathbf{U}
$$
$$
\begin{cases}
\triangle \mathbf{X} = [\triangle \mathbf{X}^{org}(D+1)^{\mathbf{T}}, \triangle \mathbf{X}^{org}(D+2)^{\mathbf{T}}, \cdots, \triangle \mathbf{X}^{org}(n-1)^{\mathbf{T}}]^{\mathbf{T}} \\
\triangle \tilde{\mathbf{X}} = [\triangle \tilde{\mathbf{X}}^{org}(D+1)^{\mathbf{T}}, \triangle \tilde{\mathbf{X}}^{org}(D+2)^{\mathbf{T}}, \cdots, \triangle \tilde{\mathbf{X}}^{org}(n-1)^{\mathbf{T}}]^{\mathbf{T}} \\
\triangle \mathbf{U} = [\triangle \mathbf{U}^{org}(D+1)^{\mathbf{T}}, \triangle \mathbf{U}^{org}(D+2)^{\mathbf{T}}, \cdots, \triangle \mathbf{U}^{org}(n-1)^{\mathbf{T}}]^{\mathbf{T}}
\end{cases}
\tag{4}
$$

where $\tilde{A}$ is the dynamic matrix, q represents the order of the process, and D represents the time interval between two sampled data. Since $\tilde{\mathbf{X}}^{org}(t)$ is affected by $\mathbf{U}^{org}(t-D)$, from a strict theoretical viewpoint, $\triangle \mathbf{U}^{org}(t)$ and $\tilde{\mathbf{X}}^{org}(t)$ are not independent; therefore, the TS-PCA algorithm is not applicable. However, when the time interval D is large enough, the effect of $\mathbf{U}^{org}(t-D)$ on $\tilde{\mathbf{X}}^{org}(t)$ is very small, while $\triangle \mathbf{U}^{org}(t)$ and $\tilde{\mathbf{X}}^{org}(t)$ can almost be regarded as two independent variables. Then $\tilde{A}$ can be estimated as $\hat{\tilde{A}}$ using least squares algorithm, and $\mathbf{U}^{org}(t)$ can be obtained using (2). The second step of TS-PCA is to use the traditional PCA method to reduce dimensionality. Notice that the TS-PCA algorithm has two more parameters: q and D, when compared with PCA.

Regarding the selection of the lag order q, many articles have been published [39]; however, the proposed methods are all for the steady-state process and cannot be used for the unsteady state process; therefore, the TS-PCA requires a new order selection method.
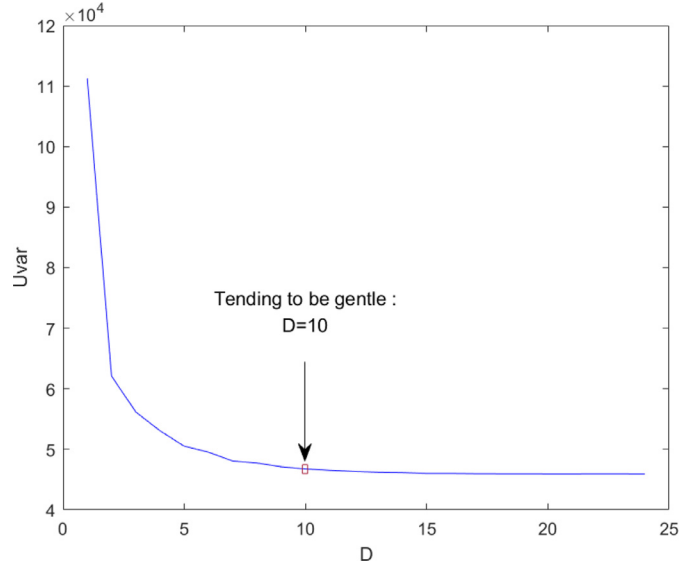

Fig. 4. Relation curve between D and Uvar.

Considering that $\hat{\tilde{A}}$ is obtained using least squares estimation, one gets:

$$
\begin{aligned}
\hat{\mathbf{U}}^{\mathbf{org}}(\mathbf{t}) &= \mathbf{X}^{org}(\mathbf{t}) - \tilde{\mathbf{X}}^{\mathbf{org}}(\mathbf{t})\hat{\tilde{A}} \\
&= \mathbf{X}^{org}(t) - \tilde{\mathbf{X}}^{org}(t)\tilde{A} + \tilde{\mathbf{X}}^{org}(t)(\hat{\tilde{A}} - \tilde{A}) \\
&= \mathbf{U}^{org}(t) + \tilde{\mathbf{X}}^{org}(t)(\hat{\tilde{A}} - \tilde{A})
\end{aligned}
\tag{5}
$$

This formula represents the innovation error of $\tilde{\mathbf{X}}^{org}(t)(\hat{\tilde{A}} - \tilde{A})$ due to the $\hat{\tilde{A}}$ deviation. Assuming $\tilde{\mathbf{X}}^{org}(t)(\hat{\tilde{A}} - \tilde{A}) \in N(\mu'(t), \Sigma'(t))$, one gets $\hat{\mathbf{U}}^{org}(t) \in N(\mu + \mu'(t), \Sigma + \Sigma'(t))$. Since the trace of the covariance matrix must be greater than 0, i.e., $tr(\Sigma'(t)) > 0$, we can obtain $tr(\Sigma + \Sigma'(t)) > tr(\Sigma)$. In other words, if the dynamic matrix $\hat{\tilde{A}}$ is estimated to be wrong, then the variance of the estimated innovation component $\hat{\mathbf{U}}^{org} = [\hat{\mathbf{U}}_1^{org}, \hat{\mathbf{U}}_2^{org}, \cdots, \hat{\mathbf{U}}_s^{org}]$ will be greater than the variance of the normal value $\mathbf{U}^{org}(t)$.

According to the above conclusion, the q estimation method is as follows: try different q values, so that the variance and minimum of the innovation component $\hat{\mathbf{U}}^{org} = [\hat{\mathbf{U}}_1^{org}, \hat{\mathbf{U}}_2^{org}, \cdots, \hat{\mathbf{U}}_s^{org}]$, as shown in the formula:

$$
q^* = \arg\min_q \sum_{i=1}^{s} var(\hat{\mathbf{U}}_i^{org})
\tag{6}
$$

In this study, the q value that is most suitable for the ICU physiological data, i.e., $q^*=1$, is selected. Because the patients' data are continuously placed in the training set, the purpose of this selection is to reduce the impact among different patients during the process of increasing the training set batch.

As with the order, the time interval parameter D also affects the identification result of the dynamic matrix $\tilde{A}$. Therefore, the parameter D can also be selected according to the variance of the innovation component $\hat{\mathbf{U}}^{org}$. The relationship between the time interval parameter and the Uvar is shown in Fig. 4. When D is large enough, $\mathbf{U}^{org}(t-D)$ and $\tilde{\mathbf{X}}^{org}(t)$ can be regarded as two variables that are independent of each other. It is of little significance to increase the value of D afterwards, and it will only increase the demand for training data in vain. When the curve change decreases, it is meaningless to choose a larger D value, and only the demand for training data is increased in vain. Here, a suitable value, $D^*=10$, is chosen.

After selecting the parameters $q^*=1$ and $D^*=10$ for TS-PCA, the rest steps are the same as those using the traditional PCA algorithm for dimensionality reduction.

At this time, the feature dimension extracted from the original data is high. If it is directly input into the model, the calculation amount is too large, so it is necessary to reduce the calculation complexity through data dimension reduction processing. Like PCA, TS-PCA has the ability to reduce data dimension and save main information, so we use TS-PCA to reduce data dimension. Therefore, 216 principal components, including 70% cumulative variance contribution rate, are retained as input data for the final model.

2.4. Just in time learning

Just-in-time learning algorithm is a multi-local modeling strategy whose basic idea is " similar input produces similar output." Compared with the traditional method of modeling with the whole data, JITL finds the database for a series of samples with the most similar status as the training dataset and builds a local personalized model to obtain a more accurate output. Patients of ICUs are highly specific and vary widely. Personalized modeling for each patient can greatly improve the treatment level. The basic idea of JITL is introduced as follows:

The similarity criterion is used for the current query sample data; then the k samples with the most similarity are collected in the stored history database [40]. Finally, a local model is established as the training dataset to obtain personalized results. The similarity D between patients is composed of two parts: Euclidean distance and angular distance. The formula can be described as:

$$D(X_q, X_i) = w\sqrt{e^{-E(X_q, X_i)^2}} + (1 - w)\cos(X_q, X_i) \qquad (7)$$

where $E(X_q, X_i) = \sqrt{(X_q - X_i)^T(X_q - X_i)}$, and $\cos(X_q, X_i) = \frac{X_q^T X_i}{\|X_i\|_2 \|X_i\|_2}$. $w \in (0, 1)$ indicates the weight coefficient, X_i represents patients in historical databases, and X_q represents the current patient. A smaller D value represents a closer similarity between patients. In this study, $w=0.6$ was chosen as a reasonable weight, and 300 similar patients were selected as individualized training sets.

To sum up, JITL considers the relationship between sequences and can effectively search for similarity. In this study, using the JITL method to search for similar cases in the database for new patients not only reduces the amount of calculation but also improves the accuracy of personalized modeling, which accords with practical application needs.

2.5. One-class extreme learning machine

Extreme learning machine is a machine learning algorithm based on feedforward neural network. As a neural network algorithm, its structure is very simple. The parameters of hidden layer nodes can be given randomly without adjustment, and the learning process only requires the calculation of output weight. The ELM algorithm has the advantages of high learning efficiency and generalization ability and is widely used in classification, regression, clustering, and feature learning.

The difference between one-class ELM and ELM lies only in the selection of training sets. When addressing binary classification problems, the ELM training set contains two types of data labeled 0 or 1; however, the training set of one-class ELM contains only one kind of data; that is, the label is only 0. In the process of establishing the model, only one kind of data is learned, and the boundary of this data can be better described. In the test set, when different classes of data are encountered, they are judged as abnormal data and given label 1. Therefore, one-class classification
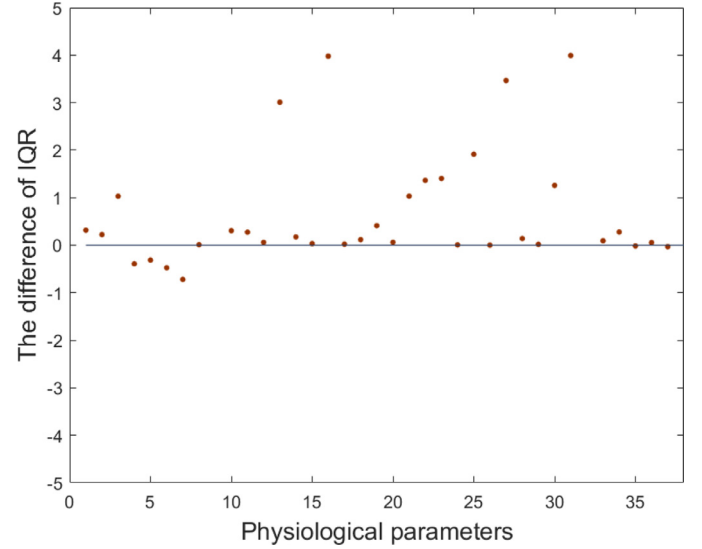

Fig. 5. The difference of IQR between discharged and undischarged patients.

problem can also be regarded as abnormal point detection problem. The principle of one-class ELM is as follows:

The dataset is $(x_j, y_j)_{j=1}^N$, where $x_j \in R^m$ indicates the input training data, and y_j is a label that represents the categories; here $y_j=0$. Then, the one-class ELM model can be described as

$$\sum_{i=1}^L \beta_i g(w_i x_j + b_i) = y_j, j = 1, 2, \cdots, N \qquad (8)$$

where L is the number of hidden nodes; w_i indicates the weight vectors between the input layer and hidden layer; β_i is the weight vector between the hidden and output layers; and $g(\cdot)$ is the activation function.

Because the weight matrix w and offset vectors b are given randomly, the output matrix H of the hidden layer can be expressed as follows:

$$H(w_1, w_2, ., w_L; b_1, b_2 \ldots b_L; x_1, x_2, \ldots, x_N) =$$
$$\begin{bmatrix} g(w_1x_1 + b_1) & g(w_2x_1 + b_2) & \cdots & g(w_Lx_1 + b_L) \\ g(w_1x_2 + b_1) & g(w_2x_2 + b_2) & \cdots & g(w_Lx_2 + b_L) \\ \vdots & \vdots & \cdots & \vdots \\ g(w_1x_N + b_1) & g(w_2x_N + b_2) & \cdots & g(w_Lx_N + b_L) \end{bmatrix}_{N \times L} \qquad (9)$$

The only unknown variable is β. According to the basic model of a single hidden feedforward networks, β satisfies

$$H\beta = Y \qquad (10)$$

Simple calculation steps, randomly given weights, and high accuracy are all advantages of the one-class ELM. In this study, the one-class ELM is chosen as a classification algorithm mainly because of the data characteristics. When one kind of data distribution is more centralized and the other is more decentralized, this algorithm is suitable to use. To clearly describe the ICU data distribution, a statistical index interquartile range (IQR) is introduced. The IQR is a quantity used to represent the degree of data dispersion in robust statistics. The larger the IQR value, the more scattered the data, and vice-versa.

As shown in Fig. 5, the X-axis corresponds to each physiological parameter, and the Y-axis represents the difference of the IQR between patients who can be discharged from hospital within 10 days and patients who cannot be discharged considering this physiological parameter. The blue line indicates that the difference is 0 and the IQR is equal. If the red dot is higher than the blue line, it means that the IQR of the undischarged patient is larger than that of the discharged patient. In Fig. 5, the points above the blue line
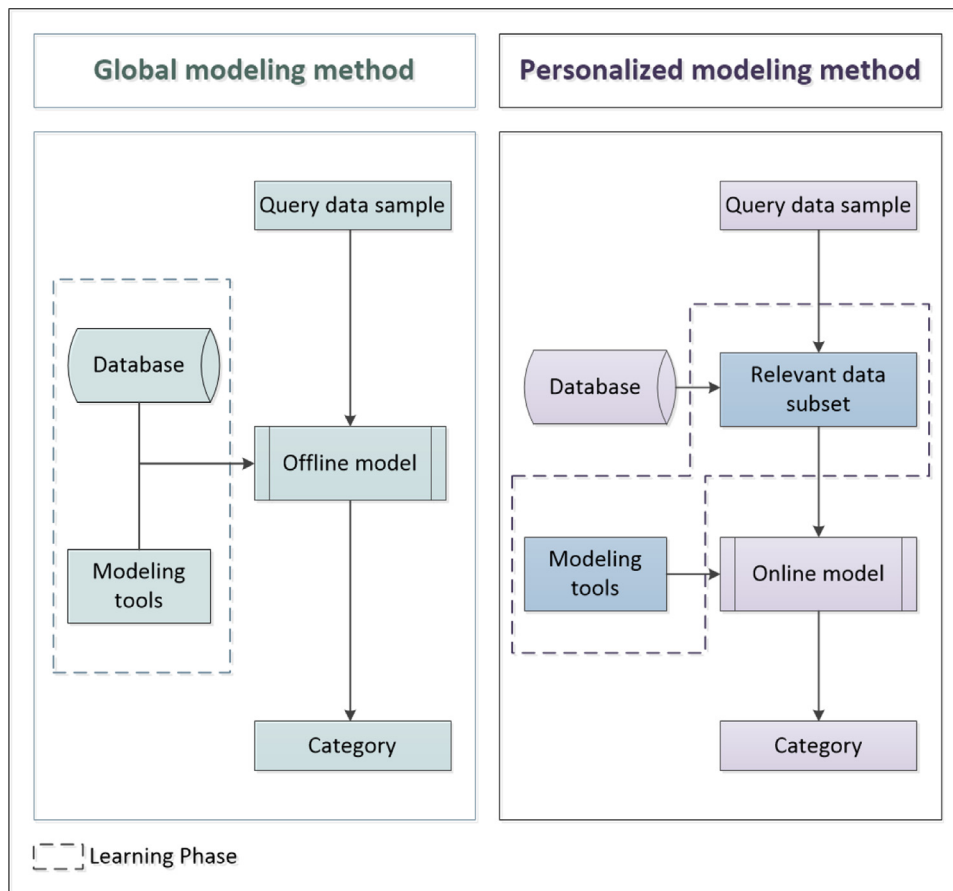
Fig. 6. Comparison between traditional model and personalized model.

are much more than points below the line. This indicates that most IQR values are lower in discharged patients than in undischarged patients. This shows that the distribution of physiological parameters of discharged patients is more centralized and in line with certain rules, while the distribution of physiological parameters of undischarged patients is more scattered and more disordered.

Hence, we chose a more centralized discharged data as the training set and use a one-class classification method to model. If the classifier learns two kinds of data, the result will become inaccurate.

2.6. One-class JITL-ELM

Patients of ICU have a strong specificity and critical condition; therefore, the one-class JITL-ELM, which combines personalized search and rapid classification, is appropriate to predict whether patients can be discharged within 10 days. The traditional prediction methods use all the data to model, while the one-class JITL-ELM only uses the closest cases to model, which meets the need of modern medical individualization and improves the judgment accuracy, as shown in Fig. 6.

One-class JITL-ELM modeling can be described as follows:

Step 1. A total of 400 records are randomly selected as the testing samples, and 1946 records for cases of discharge within 10 days are selected from the remaining data as the whole training set.

Step 2. The most similar 300 cases for each new patient in the test set are found. These records become personalized training sets.

Step 3. For the current testing patient, a one-class ELM model is established using the 300 most similar samples.

Step 4. Predictions are obtained from this personalized model.

Step 2 is repeated if new testing sample data are available.

Fig. 7 illustrates the working process of the one-class JITL-ELM model. With the addition of new patients, the new model was built in real time and the old model was disregarded. This process is a simulation of the doctor's consultation. When a new patient is encountered, the doctor searches the brain for similar cases based on the patient's symptoms and gives a judgment.

Compared with the traditional classification methods, the one-class JITL-ELM model shows better performance. Meanwhile, such a design not only meets the individual needs but also highlights the development of AI technology in medicine.

3. Results

In this section, we demonstrate the superiority of the one-class JITL-ELM in estimating patients' discharge time through a series of experiments. To comprehensively evaluate the results, the selection of parameters not only considers the global accuracy but also the internal situation of two types of data. In the medical field, negative and positive represent two types of data. In this paper, they respectively represent patients who can and cannot be discharged within 10 days.

Confusion matrix is often used to evaluate the classification result accuracy, in which TP, FP, FN, and TN respectively indicate the number of samples of true positive, false positive, false negative, and true negative in the experiment. In the experiment, we introduce other evaluation indices consisting of confusion matrix, such as

$$accuracy : Acc = \frac{TP + TN}{ALL} \tag{11}$$

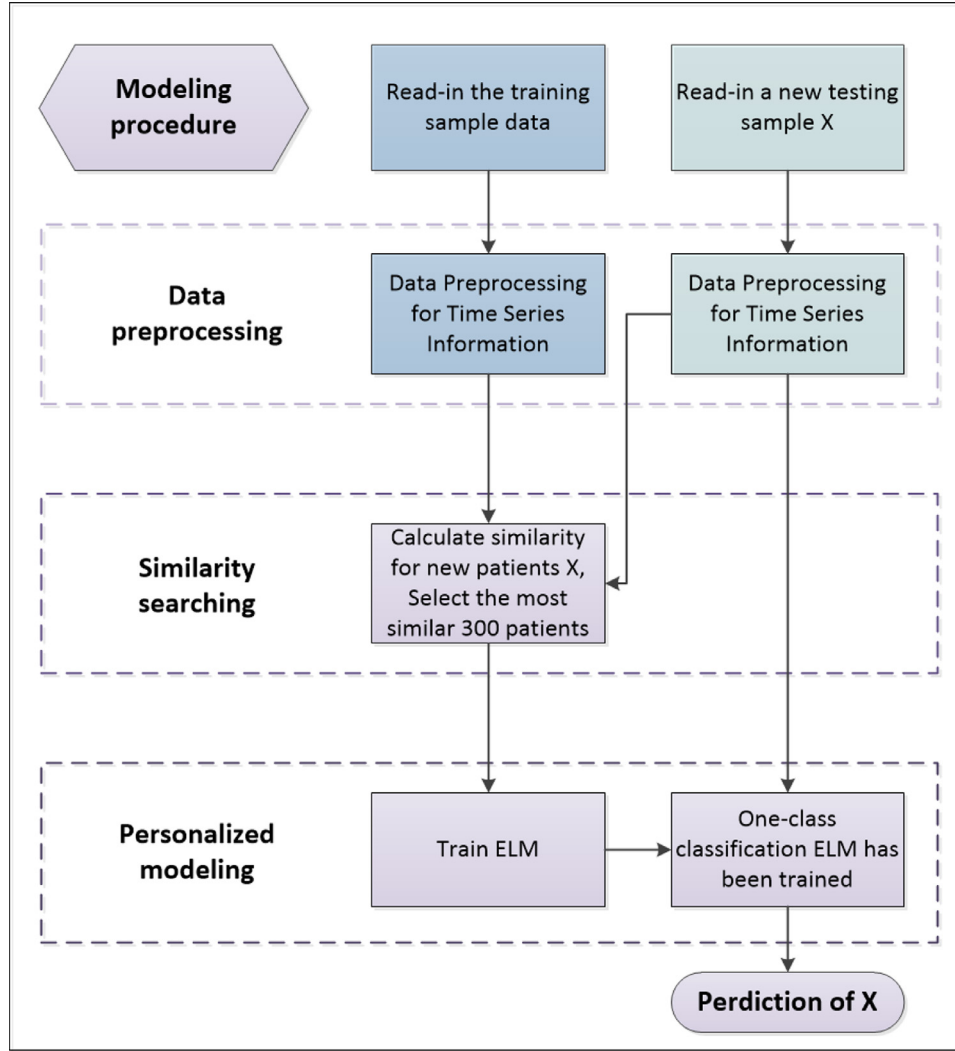**Fig. 7.** The working procedure of one-class JITL-ELM model.

$$sensitivity : Sn = \frac{TP}{TP + FN} \qquad (12)$$

$$specificity : Sp = \frac{TN}{TN + FP} \qquad (13)$$

$$precision : Pv+ = \frac{TP}{TP + FP} \qquad (14)$$

$$G-mean = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}} \qquad (15)$$

$$Lift = \frac{TP}{TP + FP} / \frac{TP + FN}{all} \qquad (16)$$

where accuracy represents the judgment of the whole data. Sensitivity and specificity are the internal accuracy metrics of the two types of data. Precision is the proportion of the positive number of correct predictions to the total number of positive predictions. G-mean considers the overall accuracy when data imbalance is considered. Lift describes the degree of improvement compared to the non-use model, with a standard of 1. Meanwhile, receiver operating characteristics (ROC) graph and AUC are also introduced as a comparison in our experiment.

To ensure fairness of the experiment, the same training set and test set were used in the experiment. The same 400 patient records were selected for each test set. The training set of the one-class classification experiment was different from that of the binary experiment. In the one-class classification experiment, 1946 records for cases of discharge within 10 days were used; in the binary classification experiment, 1654 records for cases in which patients cannot be discharged within 10 days were used, totaling 3600 training sets; in the unsupervised experiment, no training set was used.

3.1. Experiment 1: Comparison with unsupervised algorithm

In this unsupervised experiment, the classical K-means clustering algorithm and the novel isolation forest algorithm (iForest) [41] are selected for comparison.

K-means is the simplest and most efficient clustering algorithm. Its core idea is to use k initial centroids as the clustering category and repeat iteration until the algorithm converges. According to the experimental needs, k=2.

The iForest algorithm does not describe the normal point but is dedicated to isolating the abnormal point. The iForest has a very efficient strategy for judging outliers. The anomaly point here is defined as an outlier that is easily isolated. The algorithm does not require training to recursively segment the dataset until all sam-

Table 8
The relationship between state and mortality.

State	Survival	Death	Total	Mortality
Discharge within ten days	1864	295	2159	0.1366
Ten days without discharge	1582	259	1841	0.1407
Number	3446	554	4000	
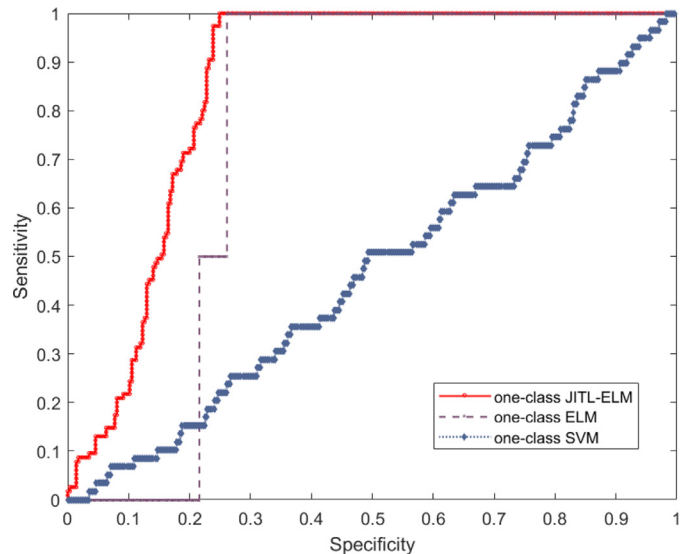
4. Discussion

4.1. Discussion on Experiment 1

The results of Experiment 1 reveal that each indicator of the one-class JITL-ELM showed that the algorithm is superior to the other two algorithms. Especially Sp=1, which means that after using the one-class JITL-ELM algorithm, all these types of data can be judged accurately. The ROC curves of one-class JITL-ELM also show great advantages.

This set of experiments validates that the one-class JITL-ELM as a supervised algorithm is better than other unsupervised algorithms.

4.2. Discussion on Experiment 2

The experimental 2 results are shown in Table 7. The one-class JITL-ELM was compared with ELM and JITL-ELM, and the ROC graph is shown in Fig. 8. The ELM performed better than JITL-ELM to a certain extent, and all the JITL-ELM's indicators show that the JITL-ELM outperformed the other two algorithms.

However, the four algorithms used for comparative experiments did not perform as well as the one-class JITL-ELM. This shows that the one-class JITL-ELM is better than the traditional binary classification algorithm in supervised learning.

### 4.3. Discussion on Experiment 3

In Experiment 3, One-class ELM achieved the same effect as the one-class JITL-ELM in the three values: Sp, Pv+, and lift. Although the AUC value of one-class ELM was slightly lower than that of one-class JITL-ELM, it was higher than that of other methods. One-class SVM had a higher G-mean value than one-class ELM, but lower than that of the one-class JITL-ELM. The results show the importance of personalized search, and it can improve the values of evaluation indices. Overall, the one-class JITL-ELM is still better.

Through three sets of experiments, the comparison of the one-class JITL-ELM with unsupervised and supervised algorithms shows the comprehensive advantages of the one-class JITL-ELM on various indicators. This fully proves that the one-class JITL-ELM can perfectly estimate whether the ICU patient can be discharged within 10 days.

4.4. Simple judgment of mortality risk level

The LOS estimation not only helps in evaluating the patient's condition laterally but also assists doctors in predicting the mortality of patients. Hence, this section presents a simple mortality risk level estimation method by analyzing the situation of ICU patients.

Age is one of the most important physiological indices. In this dataset, all patients are adults. According to the World Health Organization's criteria for age classification, adults of 18 to 44 years-old are young people, 45 to 59 are middle-aged people, 60 to 74 are young old people, and over 75 years are old people. According to this criterion, the mortality rate of 4000 patients in the database was found to be divided and increase significantly with age (Table 8). In the first three age groups, mortality increased by about 27% at a time. For the last age group, mortality increased by
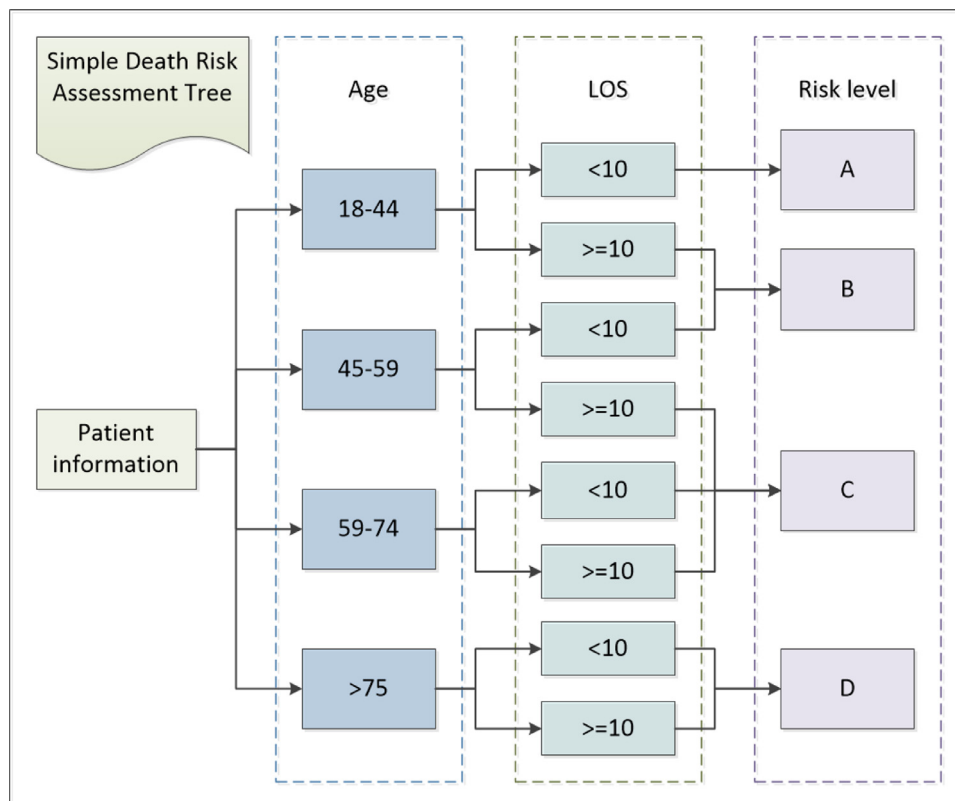
Fig. 11. Simple judgment of mortality risk level.

54%. Such an increase in mortality can clearly show that mortality is closely related to age.

Moreover, the possibility of discharge from hospital within 10 days is also related to the mortality rate, as shown in Table 8. The mortality rate of those who could not be discharged was 0.41% higher than that of those who could be discharged. Of the survivors, 1846 were discharged in 10 days, and 1582 were not discharged in 10 days. The survival rate of the former was 17.83% higher than that of the latter.

Considering age and LOS, we have developed a simple scoring system, as shown in Fig. 11. This simple judgment of mortality risk level is divided into four grades: A, B, C, and D, of which D is the most serious.

The scoring rules are as follows: Grade A is for 18–44 years-old patients that can be discharged within 10 days. Grade B is for 18–44 years-old patients that cannot be discharged within 10 days or 45–59 years-old patients that can be discharged within 10 days. Grade C is for 45–59 years-old that cannot be discharged within 10 days or 59–74 years-old patients. Grade D is for over 75 years-old patients.

To verify the reasonableness of the score, we selected 400 data from the test set to compare with the actual death situation. Here, death was predicted when a patient belongs to Grade D, and patients of other lower grades were judged to have survived. Of the total, 264 cases were predicted accurately, and the overall accuracy rate was as high as 0.6667. The accuracy of the SAPS score commonly used in hospitals on this data is only 0.6471. Among our experiments, 111 cases were false positive (determining a survived patient as dead) and 25 cases were false neglect (determining a dead patient to have survived). The miss rate was only 6.25%. In the clinical practice, the scoring system tends to consider the condition more seriously so as to pay attention to the treatment of patients.

This simple estimation method of mortality risk level is fast, highly interpretable, and highly applicable. It can simultaneously assess patients' LOS and mortality and has an extremely high clinical value.

5. Conclusion

In this study, JITL was combined with one-class ELM algorithm to predict ICU patient's LOS and yielded great results. The clinical data used in the study was sourced from the PhysioNet database. The one-class JITL-ELM demonstrated superior performance compared with unsupervised algorithms, traditional binary classification algorithms, and other one-class classification algorithms. The one-class JITL-ELM achieved the best results based on all the evaluation indices.

In summary, the one-class JITL-ELM can use early physiological parameters to predict the number of hospital days for ICU patients, which can enable the effective classification of medical resources and enable patients to receive better treatments that optimally meet their individual needs. Moreover, a simple mortality risk level estimation system based on a combination of LOS and age is proposed; the system is simple, highly interpretable, and has strong application value. This research accords with the development trend of AI in the medical field, improving medical efficiency and substantially benefiting patients. With the continuous development of deep learning, we will capture more reasonable patient data features in the future, and further improve the accuracy and judgment speed.

Statements of ethical approval

This article does not contain any studies with human participants or animals performs by any of the authors.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.inffus.2018.01.011.

References

[1] G. Valenza, H. Wendt, K. Kiyono, J. Hayano, E. Watanabe, Mortality prediction in severe congestive heart failure patients with multifractal point-process modeling of heartbeat dynamics, IEEE Trans. Biomed. Eng. 65 (2018) 2345–2354, doi:10.1109/TBME.2018.2797158.

[2] S.B. Nagaraj, L.M. Mcclain, E.J. Boyle, D.W. Zhou, S.M. Ramawamy, S. Biswal, O. Akeju, P.L. Purdon, M.B. Westover, Electroencephalogram based detection of deep sedation in ICU patients using atomic decomposition, IEEE Trans. Biomed. Eng. 65 (2018) 2684–2691, doi:10.1109/TBME.2018.2813265.

[3] J.L. Dickson, K.W. Stewart, C.G. Pretty, M. Flechet, T. Desaive, S. Penning, B.C. Lambermont, B. Benyo, G.M. Shaw, J.G. Chase, Generalisability of a virtual trials method for glycaemic control in intensive care, IEEE Trans. Biomed. Eng. 65 (2018) 1543–1553, doi:10.1109/TBME.2017.2686432.

[4] O. Sayadi, M.B. Shamsollahi, Life-threatening arrhythmia verification in ICU patients using the joint cardiovascular dynamical model and a bayesian filter, IEEE Trans. Biomed. Eng. 58 (2011) 2748–2757, doi:10.1109/TBME.2010.2093898.

[5] Y. Wang, H. Xie, X. Jiang, B. Liu, Intelligent closed-loop insulin delivery systems for ICU patients, IEEE Journal of Biomedical and Health Informatics 18 (2014), doi:10.1109/JBHI.2013.2269699. 290–209

[6] A. Michel, A. Junger, M. Benson, D.G. Brammen, G. Hempelmann, J. Dudeck, K. Marquardt, A data model for managing drug therapy within a patient data management system for intensive care units, Comput. MethodsProgr. Biomed. 70 (2003) 71–79, doi:10.1016/S0169-2607(01)00197-3.

[7] J.M. King, C.A. Bergeron, C.E. Taylor, Development of an adaptive pulmonary simulator for in vitro analysis of patient populations and patient-specific data, Comput. Methods Progr. Biomed. 161 (2018) 93–102, doi:10.1016/j.cmpb.2018.04.007.

[8] S. Siddiqui, M. Chua, V. Kumaresh, R. Choo, A comparison of pre ICU admission sirs, EWS and q SOFA scores for predicting mortality and length of stay in ICU, J. Crit. Care 41 (2017) 191–193, doi:10.1016/j.jcrc.2017.05.017.

[9] A.A. Kramer, Are ICU length of stay predictions worthwhile, Crit. Care Med. 45 (2017) 379–380, doi:10.1097/CCM.0000000000002111.

[10] U. Qazi, M. Haq, N. Rashad, K. Rashid, S. Ullah, U. Raza, Availability and use of in-patient electronic health records in low resource setting, Comput. Methods Progr. Biomed. 164 (2018) 23–29, doi:10.1016/j.cmpb.2018.06.005.

[11] H.A. Hawkins, C.M. Lilly, D.A. Kaster, R.H. Groves, H. Khurana, ICU telemedicine comanagement methods and length of stay, Chest 150 (2016) 314–319, doi:10.1016/j.chest.2016.03.030.

[12] F. Rotolo, X. Paoletti, S. Michiels, Surrosurv: an r package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials, Comput. Methods Progr. Biomed. 155 (2018) 189–198, doi:10.1016/j.cmpb.2017.12.005.

[13] A.S. Okasha, A.M. Fayed, A.S. Saleh, The four score predicts mortality, endotracheal intubation and ICU length of stay after traumatic brain injury, Neurocrit. Care 21 (2014) 496–504, doi:10.1007/s12028-014-9995-6.

[14] Z. Zhou, M. Zhong, Y. Wang, Fault diagnosis observer and fault-tolerant control design for unmanned surface vehicles in network environments, IEEE Access (2020), doi:10.1109/ACCESS.2019.2954352.

[15] N. Li, J. Tuo, Y. Wang, M. Wang, Prediction of blood glucose concentration for type 1 diabetes based on echo state networks embedded with incremental learning, Neurocomputing (2019), doi:10.1016/j.neucom.2019.10.003.

[16] X. Feng, Y. Wang, Fault estimation based on sliding mode observer for takagi-sugeno fuzzy systems with digital communication constraints, J. Franklin Inst. (2019), doi:10.1016/j.jfranklin.2019.09.020.

[17] E. Eggenberger, S. Marquez, T. Doan, J. Chipman, K. Banton, D. Radosevich, G. Beilman, Stuck in the unit: three-year outcomes following prolonged stay in the surgical intensive care unit, Surg. Sci. 40 (2014) U227–U228.

[18] X. Ye, K. Diego, L. Yuan, Predicting ICU readmission using grouped physiological and medication trends, Artif. Intell. Med. 95 (2019) 27–37, doi:10.1016/j.artmed.2018.08.004.

[19] W.A. Knaus, E.A. Draper, D.P. Wagner, J.E. Zimmerman, APACHE II: A severity of disease classification system, Crit. Care Med. 13 (1985) 818–829, doi:10.1097/00003246-198510000-00009.

[20] J.R.L. Gall, A new simplified acute physiology score (SAPS II) based on a european/north american multicenter study, J. Am. Med. Assoc. 270 (1993) 2957–2963, doi:10.1097/00003246-198510000-00009.

[21] S. Lemeshow, D. Teres, J. Klar, J.S. Avrunin, S.H. Gehlbach, J. Rapoport, Mortality probability models (MPM II) based on an international cohort of intensive care unit patients, J. Am. Med. Assoc. 270 (1993) 2478–2486, doi:10.1001/jama.270.20.2478.

[22] A. Granholm, C.F. Christiansen, S. Christiansen, A. Perner, M.H. Moller, Performance of SAPS II according to ICU length of stay:, ProtocolObserv. Study 63 (2019) 122–127, doi:10.1111/aas.13233.

[23] A.B. Rad, T. Eftestol, K. Engan, U. Irusta, J.T. Kvalo, J. Kramer-Johansen, et al., ECG-Based classification of resuscitation cardiac rhythms for retrospective data analysis, IEEE Trans. Biomed. Eng. 64 (2017) 2411–2418, doi:10.1109/TBME.2017.2688380.

[24] Q. Li, C. Rajagopalan, G.D. Clifford, Ventricular fibrillation and tachycardia classification using a machine learning approach, IEEE Trans. Biomed. Eng. 211 (2014) 202–211, doi:10.1109/TBME.2013.2275000.

[25] P. Chen, L. Yuan, Y. He, S. Luo, An improved SVM classifier based on double chains quantum genetic algorithm and its application in analogue circuit diagnosis, Neurocomputing 211 (2016) 202–211, doi:10.1016/j.neucom.2015.12.131.

[26] V.J.R. Ripoll, A. Vellido, E. Romero, J. Carlos, Ruiz-Rodr-guez, Sepsis mortality prediction with the quotient basis kernel, Artif. Intell. Med. 61 (2014) 45–52, doi:10.1016/j.artmed.2014.03.004.

[27] A.T. Azar, S.A. EI-Said, Performance analysis of support vector machines classifiers in breast cancer mammography recognition, Neural Comput. Appl. 24 (2014) 1163–1177, doi:10.1007/s00521-012-1324-4.

[28] H.T. Shiao, V. Cherkassky, J. Lee, B. Veber, G.A. Worrell, SVM-Based system for prediction of epileptic seizures from iEEG signal, IEEE Trans. Biomed. Eng. 64 (2017) 1011–1022, doi:10.1109/TBME.2016.2586475.

[29] T.N.D. Cruz, T.M.D. Cruz, P.D.S. Wellington, Detection and classification of mammary lesions using artificial neural networks and morphological wavelets, IEEE Latin America Transactions 16 (2018), doi:10.1109/TLA.2018.8358675. 926–923.

[30] Y. Ding, X. Li, Y. Wang, Mortality prediction for ICU patients combining just-in-time learning and extreme learning machine, Neurocomputing 156 (2018) 61–71, doi:10.1109/WCICA.2016.7578592.

[31] Y. Ding, X. Ma, Y. Wang, Health status monitoring for ICU patients based on locally weighted principal component analysis, Comput. Methods Progr. Biomed. 281 (2018) 12–19, doi:10.1016/j.cmpb.2017.12.019.

[32] J. Cao, K. Zhang, H. Yong, X. Lai, B. Chen, Z. Lin, Extreme learning machine with affine transformation inputs in an activation function, IEEE Trans. Neural Netw. Learn.Syst. 99 (2018), doi:10.1109/TNNLS.2018.2877468.

[33] V.M. Janakiraman, X.L. Nguyen, J. Sterniak, D. Assanis, B. Chen, Z. Lin, Identification of the dynamic operating envelope of HCCI engines using class imbalance learning, IEEE Trans. Neural Netw. Learn.Syst. 26 (2015) 98–112, doi:10.1109/TNNLS.2014.2311466.

[34] D. Min, What can AI + healthcare do, in AI + healthcare:, Appl. FutureIntell. Med.(in Chinese) (2018) 9–16.

[35] A.E.W. Johnson, T.J. Pollard, L. Shen, L.W.H. Lehman, R.G. Mark, MIMIC-III, A freely accessible critical care database, Scientif. Data 3 (2016) 160035, doi:10.1038/sdata.2016.35.

[36] G.J. Privitera, Statistics for the Behavioral S, 2018.

[37] A. Alahmar, E. Mohammed, R. Benlamri, Application of data mining techniques to predict the length of stay of hospitalized patients with diabetes, in: Proc.of 4th International Conference on Big Data Innovations and Applications (Innovate-Data, 2018, pp. 38–43, doi:10.1109/Innovate-Data.2018.00013.

[38] S. McMillan, C.C. Chia, A.V. Esbroeck, I. Rubinfeld, Z. Syed, ICU Mortality prediction using time series motifs, Comput. Cardiol. 2012 (2012) 265–268.

[39] Z. Lou, Y. Wang, Multimode continuous processes monitoring based on hidden semi-markov model and principal component analysis, Ind. Eng. Chem. Res. 56 (2017) 13800–13811, doi:10.1021/acs.iecr.7b01721.

[40] A. Gogolou, T. Tasandilas, T. Palpanas, A. Bezerianos, Comparing similarity perception in time series visualizations, IEEE Trans. Visual. Comput. Graphics 25 (2019) 523–533, doi:10.1109/TVCG.2018.2865077.

[41] F. Liu, K. Ting, Z. Zhou, Isolation forest, IEEE Int. Conf. Data Min. (2009) 413–422, doi:10.1109/ICDM.2008.17.