Future Generation Computer Systems 124 (2021) 351–360

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Predictive classification of ICU readmission using weight decay random forest

Bin Wang [a], Shuai Ding [a,*], Xiao Liu [b], X. Li [a], Gang Li [c]

[a] School of Management, Hefei University of Technology, China
[b] School of Information Technology, Deakin University, Melbourne, Victoria, Australia
[c] Centre for Cyber Security Research and Innovation, Deakin Unviersity, Geelong, Victoria, Australia

ARTICLE INFO

Article history:
Received 8 January 2020
Received in revised form 11 April 2021
Accepted 7 June 2021
Available online 9 June 2021

Keywords:
Predictive classification
Sparse data
Imbalanced data
Weight decay
Feature engineering

ABSTRACT

Intensive care unit (ICU) readmissions of critically ill patients result in significant increases in mortality rates and costs, but most readmissions could be avoided. Therefore, the medical management community has devoted considerable effort to developing predictive classifications for ICU readmissions. However, the existing classification methods lack effective feature engineering and are dependent on large quantity of imbalanced and sparse data. In this paper, we use an objective quantitative data set to estimate the probability of ICU readmission for patients who have been transferred from the ICU to the general ward at various risk levels. To implement valuable feature selection for imbalanced time series data, we integrate the missing value analysis and the likelihood ratio test for the distribution characteristics of time series indicators and introduce a weight decay random forest model to achieve ICU readmission classification based on sparse data. Using these approaches, we can rank the most relevant factors that affect the probability of ICU readmission and identify the missing indicators that have the greatest impact on ICU readmission classification. Comprehensive experimental results show that our proposed method can outperform other traditional methods according to seven different performance indicators.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The intensive care unit is a designated location that helps patients with the most severe conditions, the most unstable vital signs and the fastest changes in condition, that is, patients who are fighting for their lives, and it provides them with continuous, efficient and high-quality care. This unit is one of the most time consuming and costly units in the hospital. The volume of ICU admissions increased significantly between 2001 and 2009, especially in the emergency department ICU, where the volume increased by more than 50%, exceeding the number of available beds and medical personnel and resulting in the ICU operating at nearly full capacity [1–3]. Each nurse usually takes care of only one or two patients. Correspondingly, the care costs of patients needing treatment in the ICU are approximately 3.5 times those of patients not requiring ICU care during hospitalization [4,5]. Taking the United States as an example, $121 billion to $263 billion is spent per year in the ICU, accounting for 17.4%–39% of the total hospital costs or approximately 1% of the GDP [6,7].

With the increase in attention paid to ICUs, there has been a surge in research. Coming from a management perspective, this work primarily focuses on healthcare operations management and operational decision assistance with predictive modelling. The former primarily includes patient flow management [8–10], bed allocation [11] and coordination among medical teams [12], while the latter is more concerned with mortality prediction [13–15] and the effect of delays on service times [16]. However, researchers and scholars often overlook the importance of predicting ICU readmissions. First, there is a 7% average (range: 4% to 14%) ICU readmission rate in developed countries, which is far greater than the mortality rate of approximately 1% [17,18]. Moreover, the average mortality and length of stay for critically ill patients who were readmitted to the ICU are approximately 5 times those of patients during their first ICU admission [19]. In addition, ICU readmission is generally considered a key indicator of ICU medical quality in the US, Europe, Australia and China [20].

Although extensive attention has been devoted to ICU readmission, the traditional perspective is based on hospital operating costs [21,22]. Fortunately, with the widespread use of electronic health records (EHRs) and the rise of a new generation of information technology, accurate prediction research has ushered in new opportunities for development. However, there is still a lack of objective quantitative data that can be used to accurately predict the possibility of ICU readmission. For predicting ICU deaths based on these data, the improved random forest is

* Corresponding author.
E-mail address: dingshuai@hfut.edu.cn (S. Ding).

https://doi.org/10.1016/j.future.2021.06.011
0167-739X/© 2021 Elsevier B.V. All rights reserved.

better [23–25]. However, there are two major challenges with regard to predicting the possibility of ICU readmission: first, the lengths of stays for critically ill patients vary, resulting in imbalanced ICU time series data over all time periods, making effective feature engineering difficult to achieve; second, because these patients come from different departments and their conditions vary widely, different examination indexes lead to severe data sparseness.

To address these challenges, we propose a weight decay random forest model, which is helpful for predicting ICU readmission based on sparse data. Our method considers similarity constraints under textual and numerical data to construct a binary classification model with three important contributions. First, we focus on objective quantitative data, which are derived from ICU readmission records in clinical practice and are used for intelligent prediction. It is likely to become a trend to study ICU deaths and reduce hospital operating costs. Second, we use distribution characteristics of the indicators to calculate the eigenvalues of the time series data in the EHRs, which are employed to select features through missing value analysis and a likelihood ratio test based on imbalanced data. Third, a weight decay term is added to the random forest model to adjust the sparse data and predict ICU readmission.

The remainder of this paper is structured as follows: Section 2 briefly reviews some relevant literature. Section 3 introduces our research framework, feature engineering and the proposed approaches in detail. Section 4 presents our experimental results. Section 5 summarizes the conclusions and envisions future work.

2. Literature review

(1) ICU readmission predictive modelling

Within the process of predictive modelling for use in operations management, an important research area is the use of objective quantitative data to predict mortality during the first 48 hours of ICU admission. Our research is closely related to this work, from the perspective of readmission, which strongly affects ICU mortality. ICU readmission has also been analysed and discussed in many studies. For example, the effect of health information technology on ICU readmission for patients with congestive heart failure was studied [26]. Sarah Vollam et al. studied the relationship between non-working or working time spent transferring out of the ICU and ICU readmission [17]. Aseel K. AbuSara et al. studied the relationships among first mechanical ventilation, thrombocytopenia and ICU readmission [27]. However, our study perspective is different; previous studies on ICU readmission primarily addressed the influencing factors through statistical analysis and empirical research, while this paper focuses on the objective quantitative indicators associated with patients and attempts to predict the probability of ICU readmission. Thus, we focus on the new perspective of objective quantitative ICU data streams (including demography, biochemical indicators, laboratory indicators, laboratory scores, etc.) to predict the probability of ICU readmission.

(2) Feature engineering based on imbalanced data

The research on mortality primarily considers the first 48 hours of ICU admission, using the RNN and LSTM deep learning networks [28,29]. However, there are three primary drawbacks to this traditional method. First, there is no evidence to explain why quantitative data would have no impact on mortality and the readmission rate 48 h after admission. Second, the measurements of objective quantitative indicators and measurement intervals are different, making it difficult to determine the important factors that affect ICU readmission. Third, the differences in monitoring devices between critically ill patients have not been considered within previous studies, which may result in differences between the predicted and actual values. In contrast to previous work, we use the distribution characteristics of the indicators to calculate the eigenvalues of the time series data from the EHRs to capture the changes in and distributions of indicators during ICU hospitalization. Another problem with routine regression analysis is that some observations may unduly influence the results of the study based on adjustments to the pure model [11]. To address this problem, we use a combination of the missing value analysis, the variation coefficient and the likelihood ratio test to effectively select features. These means can solve the problems of feature engineering based on imbalanced data.

(3) Predictive modelling based on sparse data

Previous literature have developed predictive models for hospitalization outcomes of the critically ill patients, such as death, readmission, and normal discharge. The primary motivation behind these efforts is to use these models to inform operational decisions and to guide medical personnel to make better use of limited healthcare resources. We found that the application of random forests has achieved good prediction results [30,31]. However, we do not directly use random forests to predict the probability of ICU readmission due to the sparse available data. Researchers did not process high-dimensional sparse data previously, but they directly selected the data features with the most complete data volume and replaced the missing values with the averages [32–34]. This choice may lead to changes in the information from the original data structure and prevent the most useful information from being extracted for more accurate predictions. To solve these problems, a weight decay term is added to the random forest model to adjust the sparse data for ICU readmission classification.

In short, traditional prediction methods cannot address problems such as large amounts of sparse data and imbalanced data. In this work, we use a combinatorial approach to select imbalanced data features and use a weight decay random forest to create predictive models based on sparse data.

3. Models

3.1. Readmission classification framework

Clearly, it is crucial to predict the probability of ICU readmission to assist the attending physician in deciding whether to transfer critically ill patients based on ICU data streams (including demography, biochemical indicators, laboratory indicators, laboratory scores, etc.). In this work, we propose an ICU readmission classification framework that integrates multiple approaches. Calculating distribution characteristics, performing missing value analysis, and applying the likelihood ratio test can result in effective feature engineering and enhanced interpretability based on the imbalanced data, and the improved random forest model can better predict the probability of ICU readmission based on sparse data. This programme consists of two steps, as shown in Fig. 1.

In the first step, we perform feature engineering on the raw data based on the imbalanced data. We use distribution characteristics to calculate the eigenvalues of the time series indicators, next exclude large numbers of sparse data through the missing analysis, then use the coefficients of variation to remove abnormal index items, and finally select the relevant factors for ICU readmission using the likelihood ratio test.

In the second step, a random forest model with supervised binary classification is established. To solve the sparse data problems, a weight decay term is added to the random forest model to adjust the sparse data and predict the probability of ICU readmission with objective quantitative data.

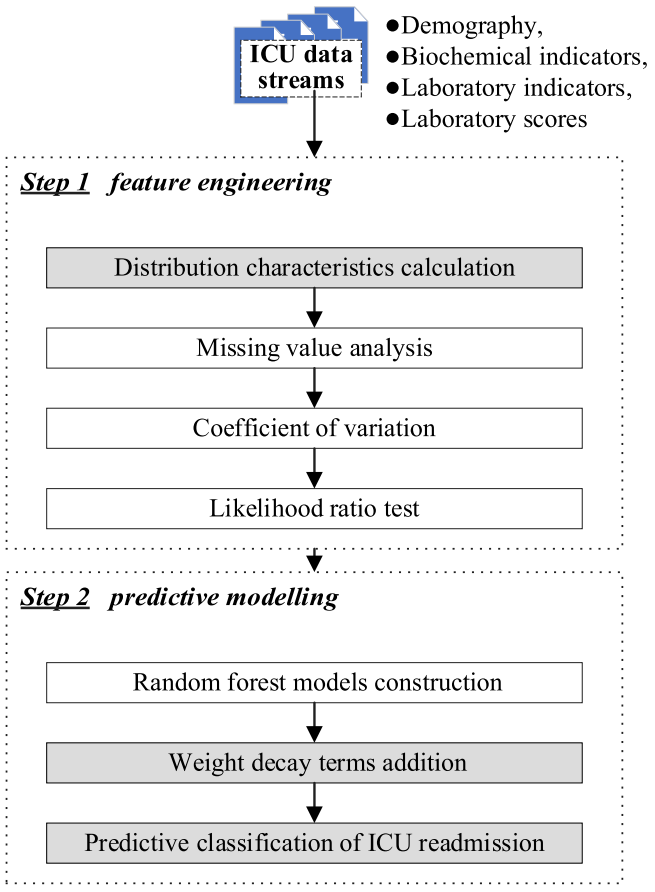

Fig. 1. Framework for readmission classification of intensive care patients.

3.2. Feature engineering

3.2.1. Data pre-processing

Due to inconsistent recording times at different levels, the residence time and measurement metrics are different. Although the time series indicators for critical patients are always monitored during ICU hospitalization, there is a serious imbalance in the data recorded in the EHRs. In this context, we use the distribution characteristics of indicators to calculate the eigenvalues of time series data in the EHRs. Specifically, we calculate the mean (mean), median (median), maximum (max), minimum (min), standard deviation (std), kurtosis (kurt), and skewness (skew). These seven commonly used statistical indicators describe the concentration trend, dispersion propensity, and distribution status of time series indicators of patients during ICU hospitalization, including the maximum and minimum values of patients used to predict mortality during ICU hospitalization. The calculation formulas for these indicators are shown in Table 1 (sorted in ascending order $z_1, z_2, \ldots z_i, \ldots, z_n$; in this work, n is odd). In addition, although some complex indicators may be better than the mentioned indicators that we commonly use in performance, they lack reasonable and effective explanatory properties [35].

Non-time series indicators of critically ill patients during ICU hospitalization primarily include textual data and numerical data. For the former, we first use text segmentation technology for text mining pre-processing. Then, the type of the corresponding field area strings is counted, and finally, the string of the type corresponding to the critical patient is extracted by string matching. For the string matching, the value given by the attending physician may involve a range of values and be entered into the database after screening.

Table 1
Calculation formulas for distribution characteristics of time series indicators.

Indicators	Calculation formulas
Minimum	$x_{min} = z_1$
Maximum	$x_{max} = z_n$
Median	$x_{median} = z_{\frac{n+1}{2}}$
Mean	$x_{mean} = \frac{1}{n}\sum_{i=0}^{n} z_i$
Standard deviation	$x_{std} = \sqrt{\left[\frac{1}{n}\sum_{i=0}^{n}(z_i - x_{mean})^2\right]}$
Kurtosis	$x_{kurt} = \frac{\frac{1}{n}\sum_{i=0}^{n}(z_i - x_{mean})^4}{\frac{1}{n^2}\sum_{i=0}^{n}\left[(z_i - x_{mean})^2\right]^2} - 3$
Skewness	$x_{skew} = \frac{\frac{1}{n}\sum_{i=0}^{n}(z_i - x_{mean})^3}{[\frac{1}{n-1}\sum_{i=0}^{n}(z_i - x_{mean})^2]^{\frac{3}{2}}}$

If the stay duration in the ICU is less than six hours, the sample information is too limited. Staying in the ICU for longer than 720 h is medically abnormal. Therefore, we select patients who were hospitalized for more than 6 h and fewer than 720 h as a study sample. We exclude cases resulting in death, and we delete data sets related to critically ill patients admitted in the past month because these samples may be subject to readmission.

3.2.2. Feature selection

This study involves various indicators, but some indicators may have an undue influence on the results of the experiment [11,26]. Fortunately, feature selection plays a crucial role in effect prediction and interpretability. This section is focused on the joint analysis methods for feature selection. We use a new likelihood ratio test for feature selection. Recent studies have shown that when the likelihood ratio test method is used for this purpose, the predictive discriminant effect is optimal [36–39].

(1) Missing value analysis. We perform a missing value analysis and the results show missing values (red parts) for some variables in Fig. 2. A total of 114 variable dimensions are shown in Fig. 2. For example, HR_max represents the maximum heart rate of critically ill patients. Through field research in the ICU and consultation with attending physicians, we know that the hospital will consider the cost of medical expenses and does not need to measure the values of variables that some attending physicians already have. Accordingly, we must perform effective feature extractions before conducting predictive modelling to reduce the interference of the model information with incomplete data.

(2) Coefficient of variation. Inevitably, the dimensions between the data or their values are different, and the degree of data dispersion cannot be selected and processed. We eliminate this effect by using the coefficient of variation, which is defined as ratio of the standard deviation to the average of the original data, reflecting the absolute value of the degree of data dispersion. Relevant theories show that when the coefficient of variation is greater than 15%, there may be data anomalies that must be considered [40]. In this work, indicators with a coefficient of variation of greater than 15% are removed.

(3) Likelihood ratio test. It is assumed that n random samples of the characteristics of critically ill patients (X_1, X_2… X_n) come from the density function h (X; θ), where θ is an unknown parameter. The null hypothesis H_0 is $\theta = \theta_0$ and the alternative hypothesis H_1 is $\theta \neq \theta_0$. The test standard is β. The ratio of the value of the likelihood function at $\theta = \theta_0$ to that at $\theta = \theta_1$

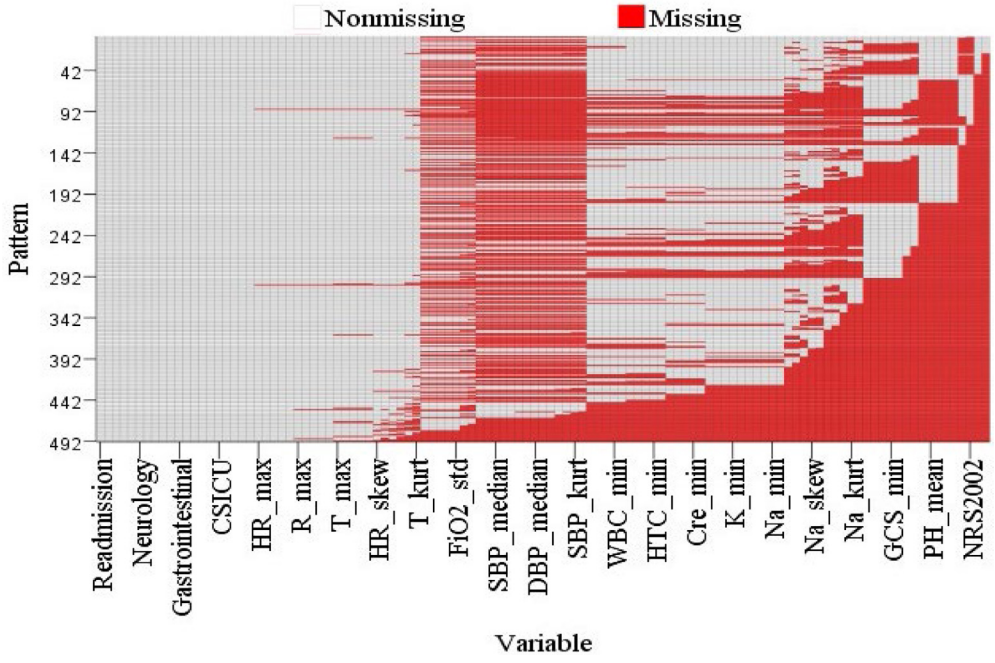

Fig. 2. Missing value analysis of characteristic distribution variables.

(maximum point) is recorded as Λ(x).

$$\Lambda(x) = \frac{sup_{\theta=\theta_0} h(x;\theta)}{sup_{\theta=\theta_1} h(x;\theta)} = \frac{h\left(x;\widehat{\theta}_0\right)}{h\left(x;\widehat{\theta}_1\right)} \tag{1}$$

Formula (1) is used to find the ratio of likelihood functions $\widehat{\theta}_0$ and $\widehat{\theta}_1$ at parameters $\theta = \theta_0$ and $\theta = \theta_1$ (maximum value), respectively. The closer the value of θ_0 is to θ_1, the greater the value of Λ(x); in contrast, the greater the difference between θ_0 and θ_1 is, the smaller the value of Λ(x). The significance threshold of this research is $\beta = 0.9$, which can be inferred according to the following rules: when $\Lambda \leq 0.9$, H_0 is rejected and H_1 is accepted, and the feature is not selected. When Λ is >0.9, we do not reject H_0, and this feature is selected.

This section presents the core content necessary to achieve imbalanced data feature selection. In the first part, data preprocessing is performed to enhance the differentiation between features and eigenvalues. The second part focuses on feature selection to reduce the dimensions of the features in order to make the model more versatile and interpretable. Feature selection could be utilized to increase the performance of our model and help researchers to understand the fundamental structure and components of the data. For example, using a missing value analysis is a convenient way for us to understand the real data; the coefficient of variation filters data with a relatively large degree of dispersion to eliminate the influence of extreme data; the likelihood ratio test calculates the correlation between the explanatory variable and the explained variable through statistical theory, and selects some of the most effective features are explanatory variables used to reduce the dimensionality of the data set.

3.3. Predictive modelling

3.3.1. Random forest model construction

In this study, we attempt to predict ICU readmissions using a combination of machine learning and traditional statistical methods. There are some objective quantitative data on critically ill patients during ICU hospitalization, including their vital signs, laboratory indicators, dynamic scoring indicators such as the heart rate (HR), oxygen concentration in inhaled air (FiO_2), Glasgow Coma Scale (GCS), etc. We attempt to use this information to predict whether patients would need to be readmitted after transfer. The ICU readmission classification is based on the physiological state during a past period so the attending physician can take appropriate measures to avoid the outcome. Therefore, we need to make a binary classifier to measure ICU readmissions. Our prediction has only two outcomes as shown in formula (2): readmission and normal discharge.

$$\text{objective classification function} = \begin{cases} 1, & \text{readmission} \\ 0, & \text{normal discharge} \end{cases} \tag{2}$$

We assume that the objective classification function is f(X), where the objective variable is ICU readmission, and X represents many factors affecting ICU readmission. $g(X)$ denotes the implicit function of the model (e.g., random forest, support vector, etc.), b denotes the constant term of the objective classification function, and μ denotes the random perturbation term, indicating the influences of random factors in the model and other factors outside the consideration of the model. The original random model is

$$f(X) = g(X) + b + \mu \tag{3}$$

where $X = (x_1, x_2, \ldots, x_i)$,with x_i denoting the influencing factors of ICU readmission after feature engineering treatment.

3.3.2. Weight decay term addition

Although we initially screened the data through feature selection, there are still sparse data in the indicator items. For example, the distribution characteristics of GCS scores for patients in non-cardiac macrovascular departments are missing in large numbers. In clinical practice, attending physicians typically do not need to measure the patient sparsity index when determining whether a critically ill patient meets the criteria for transfer out of the ICU. To achieve ICU readmission classification without amending the sparse data, we solve the method of reference weight decay for sparse data problems [41–45].

Since data sparsity remains a problem in our data after feature selection, we add the weight decay term $C(\omega) = \frac{\alpha}{2n}\sum_{\omega}\omega^2$. When the explanatory variable still has missing values, the process of calculating the term is as follows: take the sum of the squares of the weights for the ownership parameter ω divided by the sample size n of the training set. α is the weight decay coefficient and weighs the ratio of the attenuation term to the original $f(X)$. In addition, the coefficient 1/2 is primarily applied for the convenience of taking the derivative. The improved model is as follows:

$$f(X,\omega) = g(X,\omega) + C(\omega) + b + \mu \tag{4}$$

where X represents many factors affecting ICU readmission, and ω represents the weight decay variable.

Theorem 1. *Weight decay does not affect b, but it affects the updating of ω.*

Proof. Partial derivative for $f(X,\omega)$

$$\nabla_{\omega} f(X,\omega) = \frac{\alpha}{n}\omega + \nabla_{\omega}(X,\omega) \tag{5}$$

$$\nabla_{b} f(X,\omega) = \nabla_{b} f(X) \tag{6}$$

Update weights using single-step gradient descent

$$\omega \to \omega - \varepsilon\left[\frac{\alpha}{n}\omega + \nabla_{\omega} g(X,\omega)\right] = \left(1 - \varepsilon\frac{\alpha}{n}\right)\omega - \varepsilon\nabla_{\omega} g(X,\omega) \tag{7}$$

$$w = \left(1 - \varepsilon\frac{\alpha}{n}\right)\omega - \varepsilon\nabla_{\omega} g(X,\omega) \tag{8}$$

Theorem 2. *Weight decay can optimize the solutions for models and implement local optimal solutions.*

Proof. Assuming that Q(X) is the loss function of $f(X)$, the parametric of the model can be further expressed as shown in formula (9).

$$\omega_0 = \frac{\lambda_i}{\lambda_i + \alpha}\omega_i \tag{9}$$

The improved model adds a control factor to the original parameter model, and λ_i is the eigenvalue of the Hessian matrix.

(1) When λ_i is much larger than α, that is, when the patient data characteristics are complete, the weight decay is barely effective.

(2) When λ_i is much smaller than α, that is, when the patient data characteristics are sparse, the corresponding parameter is reduced to 0, and the parameter weight decay is implemented.

For a class with many stagnation points, increasing the weight decay term means adding prior knowledge to the region in which the original derivative is zero to distinguish the values. Geometrically, this addition means that the area of the original platform is tilted in the 0 direction. This approach can help the optimization algorithm converge to at least a local optimal solution instead of staying at a saddle point.

By limiting parameter ω to approximately 0, the convergence can be sped up, and the optimization difficulty can be reduced, and according to the monotone convergence theorem, we can find a large enough area to make its derivative close to 0, which means that the improvement of the variable by the gradient method becomes extremely slow, even when affected by the floating-point precision and other factors. Then, by using the weight attenuation term, the size of the control variable is approximately 0, which can help prevent the above situation; hence, the optimization algorithm can be accelerated to a large extent.

In summary, the advantages of the improved random forest model are as follows: first, it can process high-dimensional sparse data. Second, the generalization capability of the model is strong due to the use of unbiased estimates. However, our model also has some shortcomings. First, random forest models will result in over-fitting in some categories. Second, variables with too many values will have an adverse effect on the model results. For the first drawback, we employ the joint analysis method before using the model and apply weight decay in predictive modelling to reduce the overfitting problem. To reduce over-fitting problems, we adopted the method of likelihood ratio test, which was used for the effective feature selection of the existing explanatory variables before modelling and prevented collinearity between variables (information overlap between variables). The comparative experiment in Section 4.4.2 shows that our joint analysis method improves our accuracy. When $\beta = 0.9$, the effect is the best. For the second shortcoming, we remove the unrealistic values during the data filtering process.

4. Experiments

4.1. Data description

All the data we used in the experiment come from the EHRs of critically ill patients during ICU hospitalization. These data were collected from a large tertiary hospital in Anhui from March 2017 to December 2018, with total ICU inpatient cases reaching 4697, including 244 cases of mortality in the ICU (this part of the sample has been filtered out during feature selection) and 618 cases of ICU readmission. The predictors used in the ICU readmission classification framework include time series indicators and non-time series indicators. The time series indicators include patient demographic information, the department of origin, the ICU inpatient department, the disease level, the nursing level, the nutrition risk screening (NRS), the length of stay in ICU hospitalization, etc. The non-time series indicators include vital sign information, laboratory test indicators and dynamic score indexes. The raw data and the name of the data variable after the feature engineering (input variables) are detailed in the Appendix. It is worth noting that all the above data are exported by SQL statements from multiple databases in the hospital in the same batch, and the unique indexed patient ID is used as the key–value pair of all the tables. In addition, we use Python according to the patient's name and ID number. The last six items and whether there is different admission time are used to judge whether the critically ill patient is readmitted to the hospital to ensure the objectivity and validity of the data.

Through consultation with attending physicians and field investigations, the Acute Physiology and Chronic Health Enquiry (APACHE-II) score is identified as the most effective way for hospitals to determine whether critically ill patients need to be readmitted to the ICU. In addition, we found that there were relatively few sparse items in APACHE II scoring indicators through the missing value analysis. Therefore, we selected 13 APACHE II scoring items, including the oxygen concentration (FiO_2), body temperature (T), diastolic blood pressure (SBP), systolic blood pressure (SDP), PH artery (PH), heart rate (HR), respiratory rate (R), serum potassium (K), serum sodium (Na), serum creatinine (Cre), haematocrit (HTC), white blood cell count (WBC), and Glasgow Coma Scale (GCS) score as predictors of time series indicators. The details can be found in Table 7 and Table 8.

Table 2
Calculation formulas for evaluation metrics.

Indicators	Calculation formulas
Accuracy	$\mathrm{Acc} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}$
Sensitivity	$\mathrm{Sen} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$
Specificity	$\mathrm{Spe} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$
Misdiagnosis rate	$\mathrm{Mdr} = 1 - \mathrm{Sen}$
Missed diagnosis rate	$\mathrm{Mr} = 1 - \mathrm{Spe}$
$\mathrm{F1}_{\mathrm{score}}$	$\mathrm{F1}_{\mathrm{score}} = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}}$

4.2. Evaluation metrics

To evaluate the proposed method, we used the following indicators: accuracy (Acc), sensitivity (Sen), specificity (Spe), misdiagnosis rate (Mr) and missed diagnosis rate (Mdr). In addition, we used the $\mathrm{F1}_{\mathrm{score}}$ and receiver operating characteristic (ROC) curve indexes to further our evaluation of the purposed model. The formulas for the calculation of indicators above are summarized in Table 2. TP (true positives) means the number of samples of ICU readmissions properly classified by the model; TN (true negatives) means the number of normal discharges correctly classified by the model; FN (false negatives) denotes samples classified as normal discharges when they are actually ICU readmissions; and FP (false positives) denotes samples classified as ICU readmissions when they belong to the normal discharge category.

It is notable that during the process of assigning sample labels, we found that the positive and negative samples of our data set are of 1:9 ratio, which indicates heavy imbalance. Therefore, we primarily rely on the comprehensive performance indicators AUC (area under ROC curve) and $\mathrm{F1}_{\mathrm{score}}$ to measure our model. The higher the $\mathrm{F1}_{\mathrm{score}}$ and AUC values are, the better our classification results.

4.3. Baseline models

To demonstrate the validity of our proposed approaches, we chose baseline models from three types: random forest, feature selection and weight decay. Therefore, our comparative experiment was divided into three parts. In the first part, we chose the most commonly used artificial intelligence algorithms as a comparison of random forests. The artificial intelligence algorithms included the Support Vector Machine (SVM), k-nearest neighbour (KNN), Bayesian network (BN), decision tree algorithms (C5), chi-squared automatic interaction detector (CHAID), logistic regression (LR) and least absolute shrinkage and selection operator (LASSO). In the second part, the likelihood ratio test levels were 0, 0.1, 0.2...1 (where 0 means no feature selection and 1 is replaced with 0.999). In the third part, we compared the abilities of the filling technologies to fill in the missing values and the weight decay method. The filling technologies included zero-filling, average-filling, median-filling, and linear interpolation. We compared the weight decay random forest model with the following seven weight decay baseline models in the first part (Table 3).

4.4. Results

In this section, the proposed method is compared with the baseline methods shown in Table 3. The primary content of the experimental comparative analysis is the indicators selected in Section 4.2.

Table 3
Comparison with weight decay baseline models.

Models	Introductions
SVM	A linear classifier which performs binary classifications on input data.
KNN	Each sample are represented by the closest k neighbors of the category.
BN	Conditional probabilities are used to represent the relationship strength.
C5	Analyse and summarize large sample attributes by using information theory.
CHAID	Optimally segmented according to the given variable and the significance of the chi-square test.
LR	A generalized linear regression model, which is often used in data mining and disease diagnosis.
LASSO	Able to filter variables and reduce the complexity of the model.

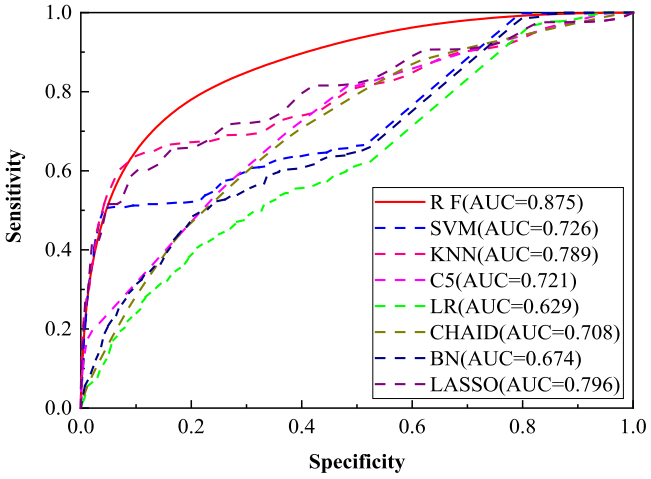

Fig. 3. ROC curves of comparison with weight decay baseline methods.

4.4.1. Comparison with weight decay baseline models

First, to confirm the advantage of the random forest method based on sparse data for predicting ICU readmission, we use a weight decay random forest model and machine learning methods for the comparative experiments. Specifically, we conduct a comparative experiment with a weight decay random forest model and other machine learning methods in Section 4.3. Table 4 lists the experimental results of the improved weight decay RF model, SVM model, KNN model, C5 model, LR model, CHAID model, BN model, and the refined regression LASSO model based on a penalty function. Specifically, the experimental results of Acc, Spe, Sen, Mr, Mdr and $\mathrm{F1}_{\mathrm{score}}$ for each method were included. The ROC curve and its AUC value are shown in Fig. 3. In particular, the feature selection and weight decay parameters are the same as those in the experimental comparison of the first part.

It is clear that the improved random forest is not the best in terms of the Acc, Spe, and Mdr. However, regarding the Sen index, the random forest achieves a value that is approximately 50% higher than those of the rest. Regarding the Mr, the random forest achieves a value approximately 50% lower than that of the others. With respect to the comprehensive performance indicators, the $\mathrm{F1}_{\mathrm{score}}$ and AUC, the performance of the random forest method is far superior to that of the other methods. Therefore, in general, the result fully demonstrates the superiority of the improved random forest method over the other methods.

4.4.2. Comparison with different likelihood ratio test levels

Different likelihood ratio test criteria may have different values for the primary factors affecting ICU readmission in feature selection. To explore the optimal likelihood ratio test criteria for

Table 4
Performance comparison with weight decay baseline methods.

Methods	Acc (%)	Sen (%)	Spe (%)	Mdr (%)	Mr (%)	$F1_{score}$
RF	87.434	65.549	90.662	34.451	9.338	0.573
SVM	87.985	17.318	98.407	82.682	1.593	0.270
KNN	87.721	13.408	98.682	86.592	1.318	0.219
C5	88.344	15.642	99.066	84.358	0.934	0.256
LR	87.075	7.449	98.819	92.551	1.181	0.129
CHAID	87.027	5.028	99.121	94.972	0.879	0.091
BN	86.955	13.594	97.775	86.406	2.225	0.211
LASSO	79.895	13.966	89.618	86.034	10.382	0.152

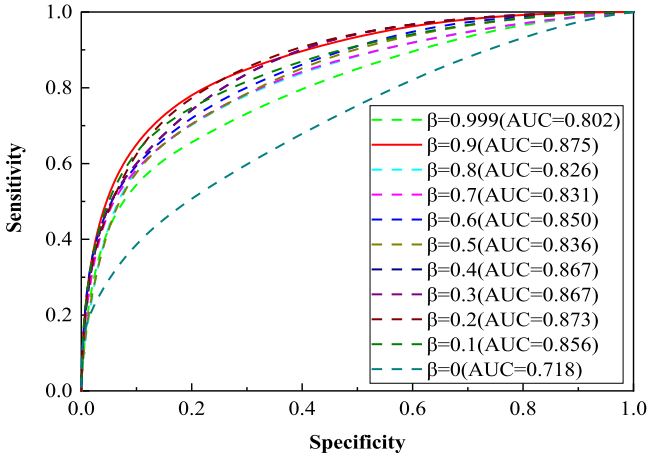

Fig. 4. ROC curves of comparison with different likelihood ratio test levels.

Table 5
Performance comparison with different likelihood ratio test levels.

Test standards	Acc (%)	Sen (%)	Spe (%)	Mdr (%)	Mr (%)	$F1_{score}$
$\beta = 0.999$	83.174	60.335	86.542	39.665	13.458	0.480
$\beta = 0.9$	87.434	65.549	90.662	34.451	9.338	0.573
$\beta = 0.8$	85.711	60.708	89.399	39.292	10.601	0.522
$\beta = 0.7$	85.184	61.453	88.684	38.547	11.316	0.516
$\beta = 0.6$	85.017	63.501	88.190	36.499	11.810	0.521
$\beta = 0.5$	83.940	64.618	86.789	35.382	13.211	0.508
$\beta = 0.4$	85.424	62.942	88.739	37.058	11.261	0.526
$\beta = 0.3$	85.424	62.942	88.739	37.058	11.261	0.526
$\beta = 0.2$	85.854	67.970	88.492	32.030	11.508	0.553
$\beta = 0.1$	86.070	66.667	88.932	33.333	11.068	0.552
$\beta = 0$	87.650	26.257	96.704	73.743	3.296	0.353

the objective classification function, we adopted different test criteria for the likelihood ratio test, which are 0 to 1 (where 1 is approximately replaced with 0.999) and verified them with the improved random forest model. The results are shown in Table 5 and Fig. 4.

Comparison experiments show that both the $F1_{score}$ value and AUC value after the likelihood ratio test are much higher than 0.353 and 0.718 without feature selection (the test standard is 0). When exploring the optimal test level suitable for this research, we found that when the test standard $\beta = 0.9$, the comprehensive performance indexes $F1_{score}$ and AUC are 0.572 and 0.875 respectively, with the best effect. We found that $\beta = 0.3$ and $\beta = 0.4$ have the same effect. By checking the data, we found that the likelihood ratio test value does not exist between 0.3 and 0.4.

4.4.3. Comparison with different sparse data processing methods

To demonstrate that weight decay is superior to the method of filling the sparse data normally, we compare the weight decay random forest model based on sparse data with several basic filling methods (zero-filling, mean-filling, median-filling and linear interpolation) after completing the likelihood ratio test. The results are shown in Table 6 and Fig. 5. In the experimental comparison of the baseline filling methods, the feature selection and the parameter design of the random forest are also the same.

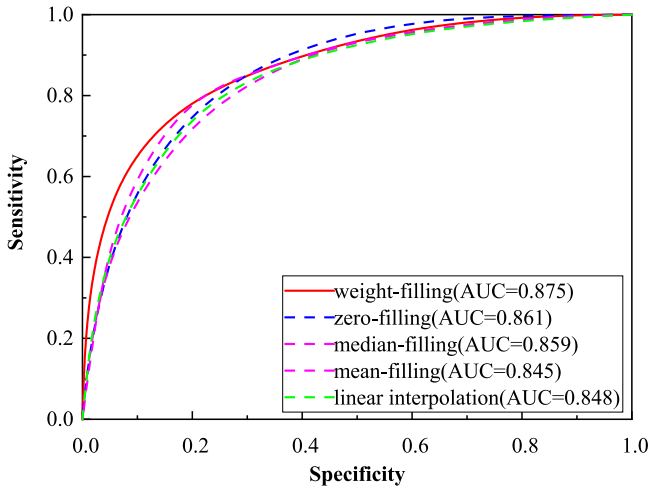

Fig. 5. ROC curves of comparisons with different sparse data processing methods.

Table 6
Performance comparison with different sparse data processing methods.

Methods	Acc (%)	Sen (%)	Spe (%)	Mdr (%)	Mr (%)	F1
Weight decay	87.434	65.549	90.662	34.451	9.338	0.573
Zero-filling	78.411	77.654	78.522	22.346	21.478	0.480
Median-filling	79.919	80.074	79.896	19.926	20.104	0.506
Mean-filling	78.554	73.743	79.264	26.257	20.736	0.469
Linear interpolation	79.105	75.605	79.621	24.395	20.379	0.482

Based on the comparison between the weight decay method and the other models, it is clear that this method is better than the others in terms of the Acc, Spe, Mr, and comprehensive $F1_{score}$ and AUC indicators. Regarding the $F1_{score}$, the weight decay performance is 0.06–0.10 higher than those of the other methods, and regarding the AUC, the weight decay performance is 0.01–0.03 higher than those of the other methods. These results fully demonstrate the superiority of the weight decay method for processing sparse data compared to the other filling methods.

4.5. Analysis and discussion

In this study, a weight decay random forest model based on sparse data was proposed, and it helps predict the probability of ICU readmission with objective quantitative data. We used a likelihood ratio test model to generate the feature scores of different predictors for feature engineering. This model produced the following 20 most important predictor lists: Stay Times, Age, R_skew, R_max, FiO_2_skew, SCICU, FiO_2_max, R_min, FiO_2_std, HR_max, R_kurt, Obstetrics, T_skew, Obstetrics, Bedridden, R_mean, R_std, Hepatobiliary, HR_kurt, HR_std, and Gastrointestinal. However, we also found it interesting that the variable data sparseness of FiO_2 is more serious than the HR and T variables according to the missing values analysis, while the feature score is higher than the HR and T values. Therefore, we reason that the missing FiO_2 indicators that have the greatest impact on ICU readmission classification.

Finally, we conclude this section by stating the limitations of our study. First, although our model can predict some of

Table 7
Sample raw data section.

Times	Names	Values	Times	Names	Values
0:00:00	ID	1 489 718	0:00:00	HR	106
0:00:00	Age	53	0:00:00	FiO_2	99
0:00:00	Gender	Male	0:00:00	FiO_2	50
0:00:00	Source	EICU	0:30:00	T	38.4
0:00:00	Stay times	79.5 h	0:30:00	R	18
0:00:00	Degree of illness	0	0:30:00	HR	113
0:00:00	Level of care	1	0:30:00	FiO_2	99
0:00:00	T	38.7	0:30:00	FiO_2	50
0:00:00	R	15	1:00:00	R	17

the factors that influence the probability of ICU readmissions, we have not clearly grasped the ways in which these factors correlate and influence one another. Second, the source of the data set we studied was a hospital, which prevented the study from expanding into other clinical settings. However, this study has demonstrated that methods such as weight decay and feature selection can improve the accuracy of models and provide a reference for attending physicians to determine whether patients should be transferred out of the ICU.

5. Conclusions and future work

There is a problem regarding the imbalanced and sparse nature of data among the objective quantitative clinical data on critically ill patients during ICU hospitalization, and it must be addressed to predict the probability of ICU readmission. We use missing value analysis and the likelihood ratio test to achieve effective feature selection based on global distribution eigenvalues of imbalanced data and then use the weighted decay random forest model to solve the data sparsity problem of predictive modelling. We draw the following preliminary conclusions: (1) The most important non-time series indicators of ICU readmission are hospitalization duration and patient age. The most important time series indicators are R, HR and T. (2) The lack of FiO_2 indicators has a greater impact on ICU readmission classification than does the lack of other indicators. (3) The correlation between ICU readmission and the concentration trend in time series data is weak. To some extent, this weakness shows that the mean-filling method of time series indicators has certain drawbacks. (4) ICU readmission is closely related to the maximum, minimum, kurtosis and skewness of time series data, indicating that the probability of ICU readmission is closely related to the extreme values and the distribution of all data during ICU hospitalization.

With the deepening of investigations into ICU readmission, this research will be further improved by extracting higher-dimensional effective predictors. Due to the limited sample size, we can improve our model by subdividing more samples into typical departments to gain more effective management ideas. In this study, we found that the time series index of EHRs is highly imbalanced and sparse. We hope that the next indexes to be measured are standard time interval data, and we will retain the original measurement data as much as possible during the prediction process. In the next step, we intend to add a large amount of textual, image and video information regarding critically ill patients to predict and further improve the robustness of our prediction framework. The research in this paper identified some influencing factors that affect ICU readmission. In future studies, we will examine how those influencing factors affect the probability of ICU readmission and how to adjust the impact factors to reduce the probability. In addition, we intend to establish a multi-parameter time series model to predict readmission time so that medical personnel can monitor and care for patients in advance.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is fully supported by the National Natural Science Foundation of China [Nos. 91846107 and 61903115], and the Fundamental Research Funds for the Central Universities [No. PA2019GDQT0021, No. PA2020GDGP0057, and No. PA2020GDKC0020].

Appendix

Table 8
The name of the data variable after feature engineering.

Readmission	Severity	GCS_max	R_std	Na_min	DBP_mean
ID	Level	GCS_mean	Cre_kurt	Na_skew	DBP_median
Age	APACHE II	GCS_median	Cre_max	Na_std	DBP_min
Gender	NRS2002	GCS_min	Cre_mean	PH_kurt	DBP_skew
Bedridden	WBC_kurt	GCS_skew	Cre_median	PH_max	DBP_std
Cardiovascular	WBC_max	GCS_std	Cre_min	PH_mean	T_kurt
Neurology	WBC_mean	HTC_kurt	Cre_skew	PH_median	T_max
Neurosurgery	WBC_median	HTC_max	Cre_std	PH_min	T_mean
Obstetrics	WBC_min	HTC_mean	K_kurt	PH_skew	T_median
Urinary	WBC_skew	HTC_median	K_max	PH_std	T_min
Hepatobiliary	WBC_std	HTC_min	K_mean	SBP_kurt	T_skew
Gastrointestinal	FiO2_kurt	HTC_skew	K_median	SBP_max	T_std
Spinal	FiO2_max	HTC_std	K_min	SBP_mean	HR_kurt
Others	FiO2_mean	R_kurt	K_skew	SBP_median	HR_max
EICU	FiO2_median	R_max	K_std	SBP_min	HR_mean
SCICU	FiO2_min	R_mean	Na_kurt	SBP_skew	HR_median
CSICU	FiO2_skew	R_median	Na_max	SBP_std	HR_min
CICU	FiO2_std	R_min	Na_mean	DBP_kurt	HR_skew
Stay times	GCS_kurt	R_skew	Na_median	DBP_max	HR_std

References

[1] S.-H. Kim, C. Chan, M. Olivares, G.J. Escobar, ICU Admission control: An empirical study of capacity allocation and its implication on patient outcomes, Manage. Sci. 61 (2015) 19–38, http://dx.doi.org/10.2139/ssrn.2062518.

[2] D.J. Wallace, D.C. Angus, C.W. Seymour, A.E. Barnato, J.M. Kahn, Critical care bed growth in the United States: A comparison of regional and national trends, Am. J. Respir. Crit. Care Med. 191 (2015) 410–416, http://dx.doi.org/10.1164/rccm.201409-1746OC.

[3] A. Jebali, A. Diabat, A stochastic model for operating room planning under capacity constraints, Int. J. Prod. Res. 53 (2015) 7252–7270, http://dx.doi.org/10.1080/00207543.2015.1033500.

[4] N. Dellaert, E. Cayiroglu, J. Jeunet, Assessing and controlling the impact of hospital capacity planning on the waiting time, Int. J. Prod. Res. 54 (2016) 2203–2214, http://dx.doi.org/10.1080/00207543.2015.1051668.

[5] A.A. Kramer, J.F. Dasta, S.L. Kane-Gill, The impact of mortality on total costs within the ICU, Crit. Care Med. 45 (2017) 1457–1463, http://dx.doi.org/10.1097/CCM.0000000000002563.

[6] L. Lu, C.W. Chan, S. Lekwijit, G. Escobar, L.V. Green, Assessing the impact of service level when customer needs are uncertain: An empirical investigation of hospital step-down units, Manage. Sci. 65 (2018) 751–775, http://dx.doi.org/10.1287/mnsc.2017.2974.

[7] S. Yeung, F. Rinaldo, J. Jopling, B. Liu, R. Mehra, N.L. Downing, M. Guo, G.M. Bianconi, A. Alahi, J. Lee, B. Campbell, K. Deru, W. Beninati, L. Fei-fei, A. Milstein, OPEN A computer vision system for deep learning-based detection of patient mobilization activities in the ICU, Npj Digit. Med. 11 (2019) http://dx.doi.org/10.1038/s41746-019-0087-z.

[8] C.W. Chan, V.F. Farias, N. Bambos, G.J. Escobar, Optimizing intensive care unit discharge decisions with patient readmissions, Oper. Res. 60 (2012) 1323–1341, http://dx.doi.org/10.1287/opre.1120.1105.

[9] D.S. Kc, C. Terwiesch, An econometric analysis of patient flows in the cardiac ICU, Manuf. Serv. Oper. Manag. 14 (2012) 50–65, http://dx.doi.org/10.2139/ssrn.1815527.

[10] J. Li, M. Dong, W. Zhao, Admissions optimisation and premature discharge decisions in intensive care units, Int. J. Prod. Res. 53 (2015) 7329–7342, http://dx.doi.org/10.1080/00207543.2015.1059520.

[11] W. Hu, C.W. Chan, J.R. Zubizarreta, G.J. Escobar, An examination of early transfers to the ICU based on a physiologic risk score, Manuf. Serv. Oper. Manag. 20 (2018) 531–549, http://dx.doi.org/10.1287/msom.2017.0658.

[12] C. LeBaron, M.K. Christianson, L. Garrett, R. Ilan, Coordinating flexible performance during everyday work: An ethnomethodological study of handoff routines, Organ. Sci. 27 (2016) 514–534, http://dx.doi.org/10.1287/orsc.2015.1043.

[13] Q. Lin, S. Pan, H. Wang, A. Fei, J. Liu, F. Wang, The relationship between coagulation abnormality and mortality in ICU patients: a prospective, observational study, Sci. Rep. 5 (2015) 1–7, http://dx.doi.org/10.1038/srep09391.

[14] C. Trautwein, S. Loosen, F. Tacke, M. Luedde, S. Roy, N. Frey, H.-J. Hippe, C. Roderburg, A. Koch, D.V. Cardenas, M. Spehlmann, T. Luedde, P. Hoening, M. Vucur, Elevated serum levels of bone sialoprotein during ICU treatment predict long-term mortality in critically ill patients, Sci. Rep. 8 (2018) 1–10, http://dx.doi.org/10.1038/s41598-018-28201-7.

[15] S.H. Ardehali, S. Dehghan, A.R. Baghestani, A. Velayati, Z. Vahdat Shariatpanahi, Association of admission serum levels of vitamin d, calcium, phosphate, magnesium and parathormone with clinical outcomes in neurosurgical icu patients, Sci. Rep. 8 (2018) 1–8, http://dx.doi.org/10.1038/s41598-018-21177-4.

[16] C.W. Chan, V.F. Farias, G.J. Escobar, The impact of delays on service times in the intensive care unit, Manage. Sci. 63 (2016) 2049–2072, http://dx.doi.org/10.1287/mnsc.2016.2441.

[17] S. Vollam, S. Dutton, S. Lamb, T. Petrinic, J.D. Young, P. Watkinson, Out-of-hours discharge from intensive care, in-hospital mortality and intensive care readmission rates: a systematic review and meta-analysis, Intensive Care Med. 44 (2018) 1115–1129, http://dx.doi.org/10.1007/s00134-018-5245-2.

[18] M.J. Al-Jaghbeer, S.S. Tekwani, S.R. Gunn, J.M. Kahn, Incidence and etiology of potentially preventable ICU readmissions ∗, Crit. Care Med. 44 (2016) 1704–1709, http://dx.doi.org/10.1097/CCM.0000000000001746.

[19] A.A. Kramer, T.L. Higgins, J.E. Zimmerman, The association between ICU readmission rate and patient outcomes, Crit. Care Med. 41 (2013) 24–33, http://dx.doi.org/10.1097/CCM.0b013e3182657b8a.

[20] A.A. Kramer, T.L. Higgins, J.E. Zimmerman, Can this patient be safely discharged from the icu?, Intensive Care Med. 42 (2016) 580–582, http://dx.doi.org/10.1007/s00134-015-4148-8.

[21] C. Senot, A. Chandrasekaran, P.T. Ward, A.L. Tucker, S.D. Moffatt-Bruce, The impact of combining conformance and experiential quality on hospitals' readmissions and cost performance, Manage. Sci. 62 (2016) 829–848, http://dx.doi.org/10.1287/mnsc.2014.2141.

[22] D. Zhang, I. Gurvich, J.A. Van Mieghem, E. Park, R. Young, M. Williams, Hospital readmissions reduction program: An economic and operational analysis, Manage. Sci. 62 (2016) 3351–3371, http://dx.doi.org/10.2139/ssrn.2366493.

[23] L.W.H. Lehman, R.P. Adams, L. Mayaud, G.B. Moody, A. Malhotra, R.G. Mark, S. Nemati, A physiological time series dynamics-based approach to patient monitoring and outcome prediction, IEEE J. Biomed. Health Inform. 19 (2015) 1068–1076, http://dx.doi.org/10.1109/JBHI.2014.2330827.

[24] M. Rouzbahman, A. Jovicic, M. Chignell, Can cluster-boosted regression improve prediction of death and length of stay in the icu?, IEEE J. Biomed. Health Inform. 21 (2017) 851–858, http://dx.doi.org/10.1109/JBHI.2016.2525731.

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

[26] I. Bardhan, J. (Cath) Oh, Z. (Eric) Zheng, K. Kirksey, Predictive analytics for readmission of patients with congestive heart failure predictive analytics for readmission of patients with congestive heart failure, Inf. Syst. Res. Publ. 26 (2015) 19–39, http://dx.doi.org/10.1287/isre.2014.0553.

[27] A.K. AbuSara, L.H. Nazer, F.I. Hawari, ICU Readmission of patients with cancer: Incidence, risk factors and mortality, J. Crit. Care 51 (2019) 84–87, http://dx.doi.org/10.1016/j.jcrc.2019.02.008.

[28] S. Purushotham, D. Sontag, Y. Liu, Z. Che, K. Cho, Recurrent neural networks for multivariate time series with missing values, Sci. Rep. 8 (2018) 1–12, http://dx.doi.org/10.1038/s41598-018-24271-9.

[29] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nat. Med. 25 (2019) 65–69, http://dx.doi.org/10.1038/s41591-018-0268-3.

[30] P. Mistry, D. Neagu, P.R. Trundle, J.D. Vessey, Using random forest and decision tree models for a new vehicle prediction approach in computational toxicology, Soft Comput. 20 (2016) 2967–2979, http://dx.doi.org/10.1007/s00500-015-1925-9.

[31] H. González-Díaz, C. Fernandez-Lozano, C.R. Munteanu, A. Pazos, Experimental study and random forest prediction model of microbiome cell surface hydrophobicity, Expert Syst. Appl. 72 (2017) 306–316, http://dx.doi.org/10.1016/j.eswa.2016.10.058.

[32] S. Ghosh, M. Feng, H. Nguyen, J. Li, Hypotension risk prediction via sequential contrast patterns of ICU blood pressure, IEEE J. Biomed. Health Inform. 20 (2016) 1416–1426, http://dx.doi.org/10.1109/JBHI.2015.2453478.

[33] Z. Wei, Y. Feng, Z. Hong, R. Qu, J. Tan, Product quality improvement method in manufacturing process based on kernel optimisation algorithm, Int. J. Prod. Res. 55 (2017) 5597–5608, http://dx.doi.org/10.1080/00207543.2017.1324223.

[34] S. Kang, E. Kim, J. Shim, W. Chang, S. Cho, Product failure prediction with missing data, Int. J. Prod. Res. 56 (2018) 4849–4859, http://dx.doi.org/10.1080/00207543.2017.1407883.

[35] J. Wang, C. Li, C. Xia, Improved centrality indicators to characterize the nodal spreading capability in complex networks, Appl. Math. Comput. 334 (2018) 388–400, http://dx.doi.org/10.1016/j.amc.2018.04.028.

[36] C. Schlereth, B. Skiera, Two new features in discrete choice experiments to improve willingness-to-pay estimation that result in SDR and SADR: Separated (adaptive) dual response, Manage. Sci. 63 (2016) 829–842, http://dx.doi.org/10.1287/mnsc.2015.2367.

[37] T. Post, V. Potì, Portfolio analysis using stochastic dominance, relative entropy, and empirical likelihood, Manage. Sci. 63 (2016) 153–165, http://dx.doi.org/10.1287/mnsc.2015.2325.

[38] R.O. Murphy, R.H.W. ten Brincke, Hierarchical maximum likelihood parameter estimation for cumulative prospect theory: Improving the reliability of individual risk parameter estimates, Manage. Sci. 64 (2017) 308–326, http://dx.doi.org/10.1287/mnsc.2016.2591.

[39] G. van Ryzin, G. Vulcano, Technical note—An expectation-maximization method to estimate a rank-based choice model of demand, Oper. Res. 65 (2017) 396–407, http://dx.doi.org/10.1287/opre.2016.1559.

[40] A. Haq, M.B.C. Khoo, New adaptive EWMA control charts for monitoring univariate and multivariate coefficient of variation, Comput. Ind. Eng. 131 (2019) 28–40, http://dx.doi.org/10.1016/j.cie.2019.03.027.

[41] P. Glasserman, Q. Wu, Persistence and procyclicality in margin requirements, Manag. Sci. Publ. 64 (2017) 5705–5724, http://dx.doi.org/10.2139/ssrn.2938515.

[42] M.G. Markakis, E. Modiano, J.N. Tsitsiklis, Delay analysis of the max-weight policy under heavy-tailed traffic via fluid approximations, Math. Oper. Res. 43 (2017) 460–493, http://dx.doi.org/10.1109/Allerton.2013.6736557.

[43] R. Govind, R. Chatterjee, V. Mittal, Segmentation of spatially dependent geographical units: Model and application, Manage. Sci. 64 (2017) 1941–1956, http://dx.doi.org/10.1287/mnsc.2016.2699.

[44] P.B. Hsieh, R.A. Jarrow, Volatility uncertainty, time decay, and option bid–ask spreads in an incomplete market, Manage. Sci. 65 (2017) 1833–1854, http://dx.doi.org/10.2139/ssrn.2263877.
[45] W.H. Sandholm, M. Staudigl, Sample path large deviations for stochastic evolutionary game dynamics, Math. Oper. Res. 43 (2018) 1348–1377, http://dx.doi.org/10.1287/moor.2017.0908.

Bin Wang received his BS degree in Applied Statistics in 2017 from Anhui University of Finance and Economics. Currently, he is working towards a Ph.D. in Management Science and Engineering at the School of Management, Hefei University of Technology, China. His research focuses on the field of ICU clinical decision support systems and big data analysis and mining.

Shuai Ding is a professor of information systems at the School of Management, Hefei University of Technology, China. He has been a visiting scholar at the University of Pittsburgh. His research interests include data mining, knowledge discovery and their applications in clinical decision support. He has published papers in refereed journals and conference proceedings, such as IEEE Transactions on Knowledge and Data Engineering, Transactions on Fuzzy Systems, Decision Support Systems, and Physics of Life Reviews. He is a member of the China Medical Equipment Journal Life Support Editorial Board, the China Systems Engineering Society Service System Engineering Branch and the Intelligent Manufacturing Professional Committee. He is a reviewer of international journals such as Decision Support Systems, IEEE Transactions on Cybernetics, IEEE Transactions on Industrial Informatics, IEEE Systems Journal, Future Generation Computer Systems, Applied Mathematics and Computation, and Knowledge-Based Systems.

Xiao Liu (Senior Member, IEEE) received Bachelor's and Master's degrees in information management and information system from the School of Management, Hefei University of Technology, Hefei, China, in 2004 and 2007, respectively, and a Ph.D. in computer science and software engineering from the Faculty of Information and Communication Technologies, Swinburne University of Technology, Melbourne, Australia, in 2011. He has taught at the Software Engineering Institute, East China Normal University, Shanghai, China. He is currently a Senior Lecturer with the School of Information Technology, Deakin University, Melbourne. His current research interests include software engineering, distributed computing, and data mining, with special interests in workflow systems, cloud/fog computing, and social networks.

Xiaojian Li is currently a professor of the School of Management, Hefei University of Technology, China. He received his B.S. in Automation from Southeast University in 2011 and his Ph.D. jointly from the Department of Automation, the University of Science and Technology of China, and the Department of Mechanical and Biomedical Engineering at the City University of Hong Kong in 2017. His research interests include robot-aided in vivo cell manipulation, surgical robotics, and in vivo SLAM. He has published many papers in top journals, such as Science Robotics, Automatica, and IEEE Transactions on Robotics. He has served as the chairman of many conferences, such as the 2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, and the 2019 International Conference on Cybernetics.

Gang Li, IEEE senior member, is an associate professor in the centre for cyber security research and innovation, Deakin University (Australia). His research interests include data mining, data privacy, causal discovery, and business intelligence.