Available online at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/burns

A novel mathematical model to predict prognosis of burnt patients based on logistic regression and support vector machine

Yinghui Huang [a,b,c], Lei Zhang [d], Guan Lian [a], Rixing Zhan [a], Rufu Xu [e], Yan Huang [c], Biswade Mitra [f], Jun Wu [a,*], Gaoxing Luo [a,*]

[a] Institute of Burn Research, Southwest Hospital, Third Military Medical University, Chongqing, China
[b] Institute of Combined Injury, State Key Laboratory of Trauma, Burns and Combined Injury, Chongqing Engineering Research Center for Nanomedicine, College of Preventive Medicine, Third Military Medical University, Chongqing, China
[c] Department of Biochemistry and Molecular Biology, Third Military Medical University, Chongqing, China
[d] College of Communication Engineering, Chongqing University, Chongqing 400044, China
[e] The Department of Epidemiology, Third Military Medical University, Chongqing, China
[f] Trauma Service Center, Alfred Hospital, 55 Commercial Road, Melbourne, VIC 3004, Australia

ARTICLE INFO

Article history:
Received 26 February 2015
Received in revised form
9 July 2015
Accepted 7 August 2015
Available online xxx

Keywords:
Burn injury
Risk factors
Support vector machine
Prediction of prognosis

ABSTRACT

Objective: To develop a mathematical model of predicting mortality based on the admission characteristics of 6220 burn cases.
Methods: Data on all the burn patients presenting to Institute of Burn Research, Southwest Hospital, Third Military Medical University from January of 1999 to December of 2008 were extracted from the departmental registry. The distributions of burn cases were scattered by principal component analysis. Univariate associations with mortality were identified and independent associations were derived from multivariate logistic regression analysis. Using variables independently and significantly associated with mortality, a mathematical model to predict mortality was developed using the support vector machine (SVM) model. The predicting ability of this model was evaluated and verified.
Results: The overall mortality in this study was 1.8%. Univariate associations with mortality were identified and independent associations were derived from multivariate logistic regression analysis. Variables at admission independently associated with mortality were gender, age, total burn area, full thickness burn area, inhalation injury, shock, period before admission and others. The sensitivity and specificity of logistic model were 99.75% and 85.84% respectively, with an area under the receiver operating curve of 0.989 (95% CI: 0.979–1.000; $p < 0.01$). The model correctly classified 99.50% of cases. The subsequently developed support vector machine (SVM) model correctly classified nearly 100% of test cases, which

* Corresponding authors. Tel.: +86 2368754173.
E-mail addresses: ikkyhuang@163.com (Y. Huang), leizhang@cqu.edu.cn (L. Zhang), lianguan@126.com (G. Lian), zhanrx@sina.com (R. Zhan), xrf@tmmu.edu.cn (R. Xu), huangy@163.com (Y. Huang), biswadev.mitra@monash.edu (B. Mitra), junwupro@cta.cq.cn (J. Wu), logxw@hotmail.com (G. Luo).

Abbreviations: SVM, support vector machine; TBA, total burn area; ROC, receiver operating characteristic; PCA, principal component analysis.

http://dx.doi.org/10.1016/j.burns.2015.08.009

could not only predict adult group but also pediatric group, with pretty high robustness (92%–100%).
Conclusion: A mathematical model based on logistic regression and SVM could be used to predict the survival prognosis according to the admission characteristics.

1. Introduction

Predicting death risk of burnt patients is one of the useful ways to reduce the mortality. Burn parameters quantized from patients provide useful information for evaluating patients' status. Comprehensive analysis of these parameters would help clinicians assess the prognosis of burn patients and guide therapy. However, predicting survival among victims of major burns trauma remains challenging.

Various injury and physiological variables may impact on mortality post burns, such as age, total burn area (TBA), depth of burn injury, presence of inhalation injury, the sites involved. Previous studies had tried to explore stable models to predict the risk of death after burn injury. However, most of them only focused on the effect of a single factor on mortality or were limited by small numbers of cases, not to mention classifying adult group and pediatric group respectively [1]. Currently, there are few practical, stable models that can predict mortality post burns injury accurately. In addition to the high mortality, major burns injury is associated with substantial morbidity and accurate prediction may enable effectiveness and palliation [2–4].

The aims of this study were to retrospectively analyze data on burns patients to develop a mathematical model of predicting mortality based on admission characteristics.

2. Methods

2.1. Clinical data collection and primary analysis

This study was approved by the ethics committee of Southwest Hospital (No.2108A0248). Data on all the burn patients presenting to Institute of Burn Research, Southwest Hospital, Third Military Medical University from January of 1999 to December of 2008 were extracted from the departmental registry. Clinical data included patient outcomes and 11 possible risk factors for mortality, including gender, age, cause, total burn area (TBA), full-thickness area, shock, inhalation injury, hours before hospitalized, combined injury and primordial condition. After the model was constructed, the accuracy, robustness and other features were determined. After the model was constructed, the accuracy, robustness and other features were determined. The evaluation of burn area and depth was based on the rule of nine and three degree four classifications. All evaluation was conducted by experienced burns surgeons in the institute. Shock status was evaluated at admission. Inhalation injury was diagnosed by bronchofibroscope and classified according to the involved range in airway when patient was suspected to have suffered inhalation injury at admission. Patients without acute burn injury, such as those for plastics or cosmetic surgery and the cases with missing data were excluded. The pediatric age group was defined by age ≤14 years.

2.2. Assignment of the collected clinical factors of the valid burn patients

All the collected factors possible to impact mortality were assigned. The final outcome of patient was assigned as Y, gender as X1, age as X2, cause of burn injury as X3, TBA as X4, full thickness area as X5, shock as X6, inhalation injury as X7, hours before hospitalization after burn injury as X8, involved sites as X9, combined injury as X10, premorbid condition as X11. The detail assignments of variates were shown as in Table 1 in supplementary.

For the causes of burn injury varied too much, subvariation [5] was introduced in the analysis and model building as presented in Table 2 in supplementary.

2.3. Building a multi-factor logistic regression predictive model

All variables showing statistical correlation with mortality were entered into a logistic regression model [6] to determine independent associations with mortality. This model was developed using a stepwise selection procedure and a

Table 1 – Verification of the constructed predictive model based on multifactor logistic regression.

Observed			Predicted		
			Y		Percentage Correct
			Survivors	Non_survivors	
Step 11	Y	survivors	6092	15	99.75
		non_survivors	16	97	85.84
	Overall Percentage				99.50

Table 2 – Area under the curve.

Area	Std. Error(a)	Asymptotic Sig. (b)	Asymptotic 95% Confidence Interval	
			Lower bound	Upper bound
.9892	.0055	.0000	.9785	1.0000

backwards elimination procedure before undergoing assessment for clinical and biological plausibility. Variables, with significant independent associations with mortality were combined to develop a mathematical prediction model. A database of all the mentioned factors was recorded using Microsoft Excel. SPSS v 13.0 was used for all statistical analyses. Odds ratios with 95% CI for mortality were derived for each variable. A p-value of <0.05 was considered to be statistically significant.

Predictive accuracy of accumulative effect of these variables was assessed by measuring the area under a receiver operating characteristic (ROC) curve. The Hosmer–Lemeshow test was used for assessing the goodness of fit of the model.

2.4. *Building a novel mathematical predictive model based on support vector machine (SVM)*

SVM is a very popular method in pattern recognition. SVM is developed from optimal separating line in linearly separable condition. The linearly inseparable data were mapped to a high-dimension space by a nonlinear function, the data become linearly separable and the optimal separating line was conversed to optimal separating plane. Therefore, SVM can separate two kinds of data with maximized classification interval. A SVM classifier can be constructed by a kernel function and some parameters. Classic kernel functions include linear kernel function, polynomial kernel function, RBF kernel function and sigmoid kernel function, which are used to map the raw data into a high-dimensional space. The formulation of the SVM is shown as follows:

Given a training set of N data points $\{\mathbf{x}_i, y_i\}_{i=1}^{N}$, where the label $y_i \in \{-1, 1\}$, $i = 1, \cdots, N$. According to the structural risk minimization principle, SVM aims at solving the following risk bound minimization problem with inequality constraint.

$$\begin{aligned} &\min_{\mathbf{w},\xi_i} \frac{1}{2}\|\mathbf{w}\|^2 + \gamma \cdot \sum\nolimits_{i=1}^{N} \xi_i, \\ &\text{s.t.} \quad \xi_i \geq 0, \quad y_i\left[\mathbf{w}^T\varphi(\mathbf{x}_i) + b\right] \geq 1 - \xi_i \end{aligned} \tag{1}$$

where $\varphi(\cdot)$ is a linear/nonlinear mapping function, $\mathbf{w}$ and b are the parameters of classifier hyper-plane.

Generally, for optimization, the original problem (1) of SVM can be transformed into its dual formulation with equality constraint by using Lagrange multiplier method. One can construct the Lagrange function

$$\begin{aligned} L(\mathbf{w}, b, \xi_i; \alpha_i, \lambda_i) = &\frac{1}{2}\|\mathbf{w}\|^2 + \gamma \cdot \sum\nolimits_{i=1}^{N} \xi_i \\ &- \sum\nolimits_{i=1}^{N} \alpha_i \left(y_i\left[\mathbf{w}^T\varphi(\mathbf{x}_i) + b\right] - 1 + \xi_i\right) - \sum\nolimits_{i=1}^{N} \lambda_i \xi_i \end{aligned} \tag{2}$$

where $\alpha_i \geq 0$ and $\lambda_i \geq 0$ are Lagrange multipliers. The solution can be given by the saddle point of Lagrange function (2) by solving

$$\max_{\alpha_i,\lambda_i} \min_{\mathbf{w},b,\xi_i} L(\mathbf{w}, b, \xi_i; \alpha_i, \lambda_i) \tag{3}$$

By calculating the partial derivatives of Lagrange function (2) with respect to $\mathbf{w}$, b and ξ_i, one can obtain

$$\begin{cases} \dfrac{\partial L(\mathbf{w}, b, \xi_i; \alpha_i, \lambda_i)}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum\nolimits_{i=1}^{N} \alpha_i y_i \varphi(\mathbf{x}_i) \\ \dfrac{\partial L(\mathbf{w}, b, \xi_i; \alpha_i, \lambda_i)}{\partial b} = 0 \rightarrow \sum\nolimits_{i=1}^{N} \alpha_i y_i = 0 \\ \dfrac{\partial L(\mathbf{w}, b, \xi_i; \alpha_i, \lambda_i)}{\partial \xi_i} = 0 \rightarrow 0 \leq \alpha_i \leq \gamma \end{cases} \tag{4}$$

Then (3) is re-written as

$$\begin{aligned} &\max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j) \\ &\text{s.t.} \quad \sum\nolimits_{i=1}^{N} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq \gamma \end{aligned} \tag{5}$$

By solving α of the dual problem (5) with a quadratic programming, the goal of SVM is to construct the following decision function (classifier),

$$f(\mathbf{x}) = sgn\left(\sum\nolimits_{i=1}^{M} \alpha_i y_i \kappa(\mathbf{x}_i, \mathbf{x}) + b\right) \tag{6}$$

where $\kappa(\bullet)$ is a kernel function. $\kappa(\mathbf{x}_i, \mathbf{x}) = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}) = \mathbf{x}_i^T \mathbf{x}$ for linear SVM and $\kappa(\mathbf{x}_i, \mathbf{x}) = \exp(-\|\mathbf{x}_i - \mathbf{x}\|^2/\sigma^2)$ for RBF-SVM. In this paper, RBF kernel function was applied to build a SVM model [7–9], and MATLAB R2009a software was used to program.

In order to construct a classification model with high robustness, adult and paediatric datasets were divided into training, test and validation samples, respectively. The training samples and test samples were distributed evenly, and the number ratio of the training samples to the test samples was near 3:1. Variables independently associated with mortality from the multivariate regression model were input to the predictive SVM model.

3. Results

During the study period, there were 8059 inpatients. There were 1825 patients excluded for presentations unrelated to acute burns injuries and 14 cases were excluded for missing data. This left 6220 cases that were included in this study. There were 113 deaths at hospital discharge (1.8%).

3.1. *Distribution of the valid cases*

The distributions of adult and pediatric cases were scattered by principal component analysis (PCA) [10], and shown in Fig. 1A and B, respectively. It could be found that the survived or died patients could be divided markedly in either adult or pediatric group. A mathematical model could be built according to the PCA distribution, and the prognosis of a new admitted patient could be predicted based on the built distribution model cursorily.

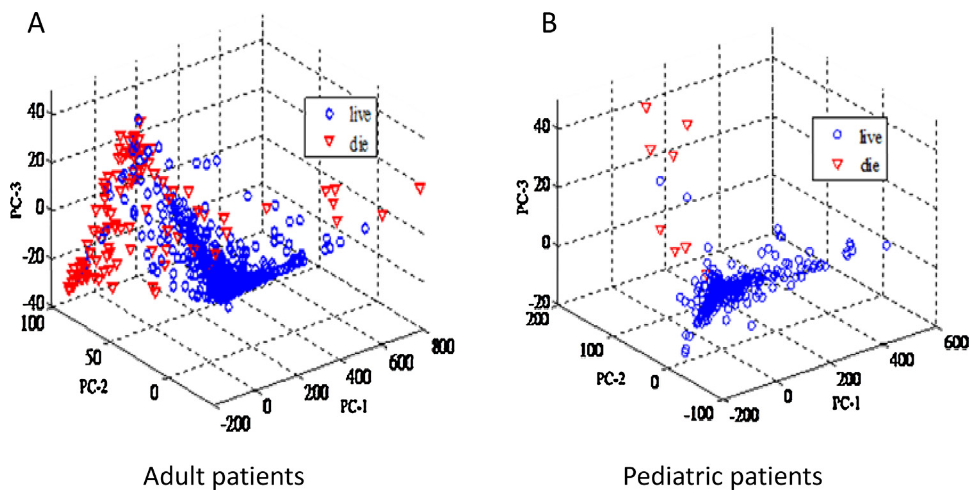

Fig. 1 – Distribution of the burn cases by principal component analysis. Blue circle: survived patients, red triangle: died patients.

3.2. *Independent-sample t test of the collected clinical data*

Each variate in the database was described and analyzed. The data either in survival group or died group were analyzed with independent sample t test firstly. It was found that the average age in died group was older than the survival group (37.06 VS 24.33, $p < 0.0001$). For the adult patients, the average ages were 38.40 and 40.00 years old in survival and died group respectively ($p < 0.05$). On the other hand, for the pediatric patients, they were 3.51 and 3.11 years old in survival and died group, respectively ($p < 0.05$). It is also found that female accounted 29.41% in survival group, which was much higher than that of died group (13.27%, $p < 0.0001$). However, the gender did not impact the mortality in the pediatric patients ($p = 0.076$). There were no differences of the other variates such as X31, X32, X33, X36, X37, X11 between survival and died group by independent-samples T test. The detail can be found in Table 3 in supplementary.

3.3. *Univariate logistic analysis of the collected data*

All the variates were analyzed with wald method of binary logistic in SPSS 13.0 software [11]. Standard of p value was set as 0.05, and it was considered statistical significance when OR is more than 1. The variates of age, period before hospitalization, TBA, third degree burn area, involved sites, complicated shock, inhalation, combined injury, basic health condition were found as possible death risk factors. All the p values were lower than 0.001. The accuracies to predict outcomes of survival and died by third degree burn area alone were 98.3% and 23.9% respectively. The outcome of death could be predicted by none of other possible risk factors alone, but they could predict survival, the accuracies were more than 97.8%. It was suggested that mortality should not be predicted by a single risk factor, a multifactor logistic regression model should be built to predict mortality of a new admitted burn patients. The detail can be found in Table 4 in supplementary.

3.4. *Construction and verification of a predictive model based on multifactor logistic regression*

3.4.1. *Construction of a multifactor logistic regression model to predict the outcome of burn patients*

All the variates in the database were analyzed and calculated by 11 steps with the forward: wald gradually move forward method in SPSS13.0 software [11]. When the p value more than 0.05 was set as introducing standard, all the eleven variates, i.e., gender, age, TBA, TBA of Third degree, inhalation injury, complicated with shock, period before hospitalization, involved sites, causes of burn

Table 3 – Prognosis prediction of total burnt patients based on SVM model.

	Total samples	Training samples	Testing samples	Validating samples
Number of samples	6220	1266	549	4405
Sampled cases	6107(113)[a]	1200 (66)	502 (47)	4405 (0)
Verified results	5967(108)	1198 (61)	502 (42)	4147 (0)
Accurate rate	97.71%(95.58%)	99.83% (92.42%)	100% (89.36%)	94.14%

[a] :the numbers out of the brackets were for survival cases, in the brackets were died cases.

Table 4 – Prognosis prediction of adult burnt patients based on SVM model.

	Total samples	Training samples	Testing samples	Validating samples
Number of samples	3808	761	244	2803
Sampled cases	3704(104)[a]	700 (61)	201 (43)	2803 (0)
Verified results	3567 (97)	699 (58)	201 (43)	2599 (0)
Accurate rate	96.30%(93.27%)	99.86% (95.08%)	100% (100%)	92.72%

[a] :the numbers out of the brackets were for survival cases, in the brackets were died cases.

injury could be introduced into the multifactor logistic regression model. Finally, the death risk could be demonstrated as:

$$\mathrm{Logit}(P) = \ln[P/(1-P)] = -2.114X_1 + 0.031X_2 + 3.7X_4 + 1.5X_5 + 0.734X_7 + 2.059X_{34} - 2.376X_{35} + 0.59X_6 + 0.008X_8 + 2.104X_9 + 2.958X_{10} - 19.474 \quad (7)$$

It was found by verification the total accuracy to predict burn patients outcome was 99.50%, the sensitiveness was 99.75%(6092/(6092 + 15)), and the specificity was 85.84%(97/(97 + 16)) (Table 1).

3.4.2. *Evaluating the accuracy of the built model based on multifactor logistic regression by Receiver Operating Characteristic Curve.*

The confidence interval was set as 95%, the areas under the ROC curves [11] of the above eleven possible risk factors were calculated and shown in Table 5 in supplementary, respectively. Except gender, the areas under the ROC curves of other ten variates were more than 0.5 ($p < 0.05$). The areas of TBA, full thickness area and involved sites were 0.967, 0.920, and 0.972 respectively. The factors of inhalation injury, shock, involved sites, combined injury were more than 7.60. It was also found that age and hours before hospitalization impacted the mortality significantly (Table 5 in supplementary).

Using all the eleven factors together, the area under the ROC curve was as high as 0.9892, $p = 0.0000$. The detail could be found in Table 2. It suggested the specificity be very high.

3.4.3. *Chi-square test of the coefficients of the built model*

By 11 step by step calculation, the coefficients of the built model based on multifactor logistic regression were analyzed by Chi-square test, it was found the p values were less than 0.0001, which suggested the coefficients in the model were statistical significant.

3.4.4. *Fit test of the goodness of the built model*

The built model based on multifactor logistic regression was applied goodness of fit test by Hosmer and Lemeshow Test of SPSS13.0 software [11,12]. It was found that the Chi-square value was 0.016, p value was 1.000, respectively, which suggested the theoretical frequencies be as same as the actual numbers.

3.4.5. *Determination of the stability of the built model based on multifactor logistic regression*

The stability of the built model was determined by jackknife method [13]. It was found that the changes of coefficient in the model were less than 0.1%, or even there were no changes at all when some cases were removed from the database randomly. The results demonstrated the model was stable. The mortality of a new admitted burn patient could be predicted by the built model based on multifactor logistic regression.

3.4.6. *Instance verification of the built model based on multifactor logistic regression*

The mortalities of the patients in the database were calculated based on the built mathematical model. The schemas of survival group and died group were made using quantity of cases as x axis, possibility of mortality as ordinate, respectively. It was found that there were 19 cases in survival group (6107), 16 cases in died group (113) out of the predictions(Fig. 2), the accuracies were 99.85% and 85.84%,respectively.

3.4.7. *Limitations of the predictive logistic model*

Predicting either pediatric or adult group with logistic model, it was found some important factors such as age, burn area did

Table 5 – Prognosis prediction of burnt children based on SVM model.

	Total samples	Training samples	Testing samples	Validating samples
Number of samples	2412	505	305	1602
Sampled cases	2403(9)[a]	500 (5)	301 (4)	1602 (0)
Verified results	2395(8)	500 (4)	301 (4)	1583 (0)
Accurate rate	99.67%(88.89%)	100% (80.00%)	100% (100.00%)	98.81%

[a] :the numbers out of the brackets were for survival cases, in the brackets were died cases.

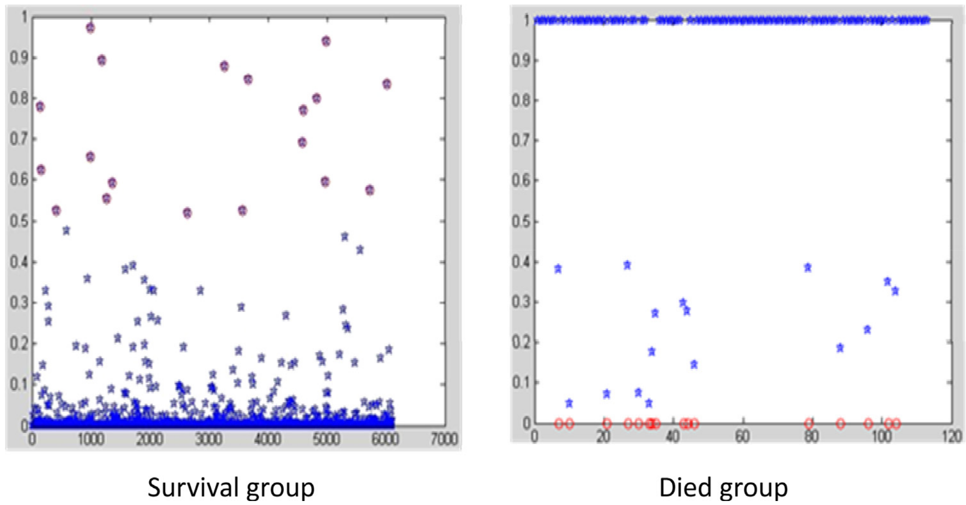

Fig. 2 – Accuracy of the logistic model.
Blue star: survived patients, red circle: died patients.

not play a vital role in the model to predict the mortality, though these factors in deed impact the mortality. Therefore, a more precise model should be built.

3.5. *Construction of a novel model to predict mortality of burnt patients based on SVM*

3.5.1. *Building a predictive model based on SVM*

Firstly, in order to construct a classification model with high robustness, all the cases, adult cases or pediatric cases were divided into training samples, test samples and validation samples, respectively. The training samples and test samples were distributed evenly, and the number ratio of the training samples to the test samples was near 3:1. From our logistic regression model, it was found that most causes of burns (variable of X31, X32, X33, X36 and X37) and basic conditions (variable of X11) impacted the prognosis insignificantly. Therefore, the other ten variates, i.e., gender, age, total burn area, full thickness burn area, inhalation injury, combined shock, period of admission after injury, involved sites, combined injury, other reasons of burns including explosive and electronic burns were constructed into the predictive SVM model.

In this paper, Gaussian RBF kernel function is used in SVM. Therefore, two model parameters such as regularization coefficient γ and kernel parameter σ are involved. The best parameters are determined by using the popular grid-search and 10-fold cross-validation on the training set. The model parameter of adult classification model were calculated as $\gamma_1 = 1200.6655$, $\sigma_1^2 = 1259.4127$; in the pediatric group, $\gamma_2 = 226.8305$; $\sigma_2^2 = 1328.0958$. The classifier parameters $\boldsymbol{\alpha}$ in Eq. (6) including are solved by using a SVM toolbox based on the training set. The SVM training and testing process are illustrated as follows

First, the SVM training is to solve the maximization model in Eq. (5) based on the given training set. The outputs of SVM training are the classifier parameters $\boldsymbol{\alpha}$. The model solving (training) process can be implemented by an existing toolbox (e.g. libsvm).

Second, after obtaining the classifier parameter $\boldsymbol{\alpha}$, the testing of a new sample can be implemented by computing decision function Eq. (6). Then the predicted class label of the sample is obtained.

3.5.2. *Accuracy of predicting prognosis of burn patients with the SVM model*

3.5.2.1. *Predicting prognosis of the burn patients with the SVM model*. There were 6220 cases totally in this study. In the classification model, 1266 cases were chosen as training samples randomly, 549 cases as test samples, and 4405 cases as validation samples. Programming with Matlab R2009a, the prognosis was predicted. The predictions of different groups with the SVM model were shown in Table 3. The predicting accurate rates for total samples were 97.71% and 95.58% for survival and died patients respectively, for training samples 99.83% and 92.42%, for testing samples 100% and 89.36%, for validation samples 94.14% (no died case), as shown in Table 3.

The 3808 adult cases were divided into 761 training samples, 244 testing samples and 2803 validation samples. The verification of the samples was shown in Table 4. The predicting accurate rates for total adult samples were 96.30% and 93.27% for survival and died cases respectively, for training samples 99.86% and 95.08%, for testing samples 100% and 100%, for validation samples 92.72% for survival patients (no dead patients), as shown in Table 4.

There were 2412 pediatric cases, which were divided 505 training samples, 305 testing samples and 1602 validation samples. The verified results were shown in Table 5. The predicting accurate rates for total pediatric samples were 99.67% and 88.89% for survival and died cases respectively, for training samples 100% and 80.00%, for testing samples100% and 100%, for validation samples 98.81% for survival patient-s(no died patients), as shown in Table 5.

3.5.3. Determining robustness of the SVM model by sampling 100 times randomly

The total cases including adult and pediatric patients were sampled randomly 100 times, 3000 samples were extracted each time. The prognoses of the samples were verified with the SVM model. It was found the accuracy rates were between 92% and 99% for the total samples as shown in Fig. 3A. For the adult samples, 2000 samples were extracted every time, the accuracy rates were between 90% and 97% (Fig. 3B). For the pediatric patients, 1000 cases were extracted each time, and it was found that the accuracy rates were 97% to 100% (Fig. 3C). Predictive accuracy rates were more than 92% in all the random extracts, which suggested the robustness of the built SVM model is high enough to apply for all burn patients.

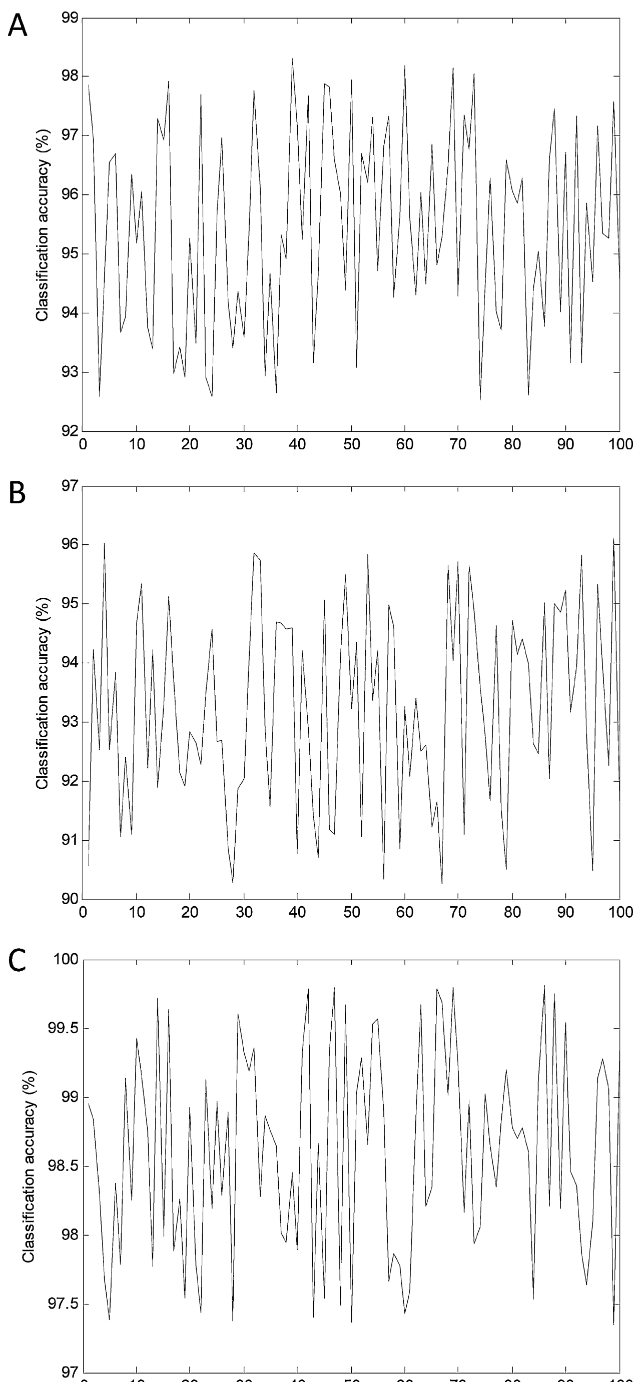

Fig. 3 – Detecting robustness of the SVM model by sampling 100 times randomly.
A. total samples B. adult samples C. pediatric samples

4. Discussion and conclusions

For the severely burned patient, early determination of prognosis may guide therapy. This study has developed a mathematical model which predicts prognosis following major burns. 12 items including gender, age, burn mechanism, total burn area, full thickness burn area, complicated shock and combined inhalation injury were examined to construct a mathematic model to predict prognosis of a certain burn patient. The clinical data of 6220 cases admitted to our unit during ten years were collected and analyzed. Finally, a mathematic model was constructed based on logistic regression and SVM. All the collected data were analyzed by *t*-test, chi-square test, single factor and multifactor logistic regression analysis. Eleven factors were found significantly to affect the prognosis of burn patients. These were gender, age, total burn area, full thickness burn area, combined inhalation injury, shock, the period before admission after burn injury, the sites involved, multiple injury, and injury mechanism (explosion or electricity).

In our study, it was found that the constructed mathematical model based on multifactor regression analysis was simple and practical to some extent. However, it missed age and burn area, which were very important factors to predict the risk of severely burned patients when predicting pediatric or adult patients respectively. Therefore, a novel predictive model of burn patients based on SVM was built. Based on SVM, we constructed a mathematical model to predict the prognosis of burn patients, and it was found that it had accuracy, specificity, stability and was robust. From the constructed model, items such as gender, age, total burn area, inhalation injury, compounded with shock and explosive injury were correlated with the prognosis of burn patients.

Previous research has demonstrated that age impacts prognosis significantly [4]. Lionelli found with each 10 year increase in age the possibility of death increased 10%[14]. Our study demonstrated that OR value of age was 1.031, which suggested that in older patients the prognosis was worse. However, it was shown with independent *t* test that this conclusion only fits to adult patients. For the pediatric patients, this situation was completely reversed, with the youngest having a higher risk to death.

Previous studies have demonstrated that gender is an important prognostic factor [15–17]. Skin thickness, muscle mass, IL-6 production, cellular and hormonal immune responses result in females having a higher risk of death for the same burn injury [18,19]. This was true for adult patients in our study. But there was no difference between pediatric male and female victims ($p = 0.076$). One explanation is that the gender differences described previously are unobvious between prepubertal boys and girls. Sharma and his colleagues had similar findings [20].

Inhalation injury has been shown to increase the likelihood of death post burn [4,21,22]. Our study found that inhalation injury increased the likelihood of death as high as 2.082 fold. Explosion as an injury mechanism increased the risk of death by 2.059 fold. Almost all of these patients had an inhalation injury. Moreover, our study demonstrated that the total burn or full thickness burn areas affected survival (43.28% and 79.11% vs 2.567% and 14.83%).

In our model, the score of the injured sites in the group that died was as high as 5.85. The head, face, neck and perineum were always involved, whereas the score in the survival group was 1.969, which mainly involved extremities and trunk. When the head, face and neck are injured, inhalation injury is coming and, therefore the possibility of death is high. Also perineal burns usually suggest the total burn area be large and so the risk of death increases.

Burn patients with complications at admission have a high risk of death [4,23,24]. In our study that the presence of complication was significant between those who died and the survival group (72.57% vs 3.32%, $p < 0.0001$).

It was also found that the period from injury to admission of survivors was much shorter than that of those who died (47.98 h vs 90.55 h, $p < 0.001$), which suggest that early resuscitation is critical, especially for the severely burned patients [25,26].

From our mathematical model, there were at least seven factors increasing the death risk of burn patients, which needs to be aware of by clinicians. They were age, total burn area, full thickness burn area, period before admission, involvement of the head, face, neck and perineum, combination of inhalation or other injuries, and presence of shock or other complications at the time of admission.

Because of its convenience and accuracy, our model is especially useful in receiving large amounts of burn patients at a time. It can also be used in prognosis prediction of many other kinds of disease such as tumor.

Competing interests

The authors declare that they have no competing interests relevant to this article.

Authors' contributions

HYH and LG carried out clinical data collection, statistic analysis and manuscript preparation. ZL and ZRX participated in the collection of clinical records and data analysis. ZL and XRF instructed and performed the data statistic analysis. HY helped to collect the clinical data. BM participated in the discussion of the study and helped to draft the manuscript. WJ and LGX conceived of the idea, designed and coordinated the whole study, and finished the manuscript preparation. All the authors read and approved the final manuscript.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (30672174, 61401048, 30973116, 81171809), ''863'' project in China (2012AA020504).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.burns.2015.08.009.

REFERENCES

[1] Mahar P, Wasiak J, Bailey M, Cleland H. Clinical factors affecting mortality in elderly burn patients admitted to a burns service. Burns 2008;34(5):629–36.
[2] Ryan CM, Schoenfeld DA, Thorpe WP, Sheridan RL, Cassem EH, Tompkins RG. Objective estimates of the probability of death from burn injuries. N Engl J Med 1998;338(6):362–6.
[3] Saffle JR. Predicting outcomes of burns. N Engl J Med 1998;338(6):387–8.
[4] McGwin G, George RL, Cross JM, Rue LW. Improving the ability to predict mortality among burn patients. Burns 2008;34(3):320–7.
[5] Li BS, Wang XY, Ma FL, Jiang B, Song XX, Xu AG. Is high resolution melting analysis (HRMA) accurate for detection of human disease-associated mutations? A meta analysis. PloS One 2011;6(12):e28078.
[6] Pagano M, Gauvreau K. Principles of Biostatistics. 2nd ed, New York: Duxbury Press; 2000.
[7] Hua S, Sun Z. Support vector machine approach for protein subcellular localization prediction. Bioinformatics 2001;17(8):721–8.
[8] Roshan U, Chikkagoudar S, Wei Z, Wang K, Hakonarson H. Ranking causal variants and associated regions in genome-wide association studies by the support vector machine and random forest. Nucleic Acids Res 2011;39(9):e62.
[9] Yang Z, Lin Y, Wu J, Tang N, Lin H, Li Y. Ranking support vector machine for multiple kernels output combination in protein-protein interaction extraction from biomedical literature. Proteomics 2011;11(19):3811–7.
[10] Tang Z, Li T, Liu R, Xiong W, Sun J, Zhu Y, et al. Improving the performance of beta-turn prediction using predicted shape strings and a two-layer support vector machine model. BMC Bioinformatics 2011;12:283.
[11] Coakes SJ. SPSS 13. 0 for Windows. New York: John Wiley & Sons; 2007.
[12] Andersen PK, Ekstrom CT, Klein JP, Shu Y, Zhang MJ. A class of goodness of fit tests for a copula based on bivariate right-censored data. Biom J 2005;47(6):815–24.
[13] Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis. Generalization to the population of readers and patients with the jackknife method. Invest Radiol 1992;27(9):723–31.
[14] Lionelli GT, Pickus EJ, Beckum OK, Decoursey RL, Korentager RA. A three decade analysis of factors affecting burn mortality in the elderly. Burns 2005;31(8):958–63.
[15] O'Keefe GE, Hunt JL, Purdue GF. An evaluation of risk factors for mortality after burn trauma and the identification of gender-dependent differences in outcomes. J Am Coll Surg 2001;192(2):153–60.
[16] Chang EJ, Edelman LS, Morris SE, Saffle JR. Gender influences on burn outcomes in the elderly. Burns 2005;31(1):31–5.
[17] Muller MJ, Pegg SP, Rule MR. Determinants of death following burn injury. Br J Surg 2001;88(4):583–7.
[18] Gregory MS, Faunce DE, Duffner LA, Kovacs EJ. Gender difference in cell-mediated immunity after thermal injury is mediated, in part, by elevated levels of interleukin-6. J Leukoc Biol 2000;67(3):319–26.

[19] Gregory MS, Duffner LA, Faunce DE, Kovacs EJ. Estrogen mediates the sex difference in post-burn immunosuppression. J Endocrinol 2000;164(2):129–38.
[20] Sharma PN, Bang RL, Al-Fadhli AN, Sharma P, Bang S, Ghoneim IE. Paediatric burns in Kuwait: incidence, causes and mortality. Burns 2006;32(1):104–11.
[21] Thompson PB, Herndon DN, Traber DL, Abston S. Effect on mortality of inhalation injury. J Trauma 1986;26(2):163–5.
[22] Tanizaki S, Suzuki K. No influence of burn size on ventilator-associated pneumonia in burn patients with inhalation injury. Burns 2012;38(8):1109–13.
[23] Palmieri T, Lavrentieva A, Greenhalgh DG. Acute kidney injury in critically ill burn patients. Risk factors, progression and impact on mortality. Burns 2010;36(2): 205–11.
[24] Germann G, Barthold U, Lefering R, Raff T, Hartmann B. The impact of risk factors and pre-existing conditions on the mortality of burn patients and the precision of predictive admission-scoring systems. Burns 1997;23(3):195–203.
[25] Ford HR. Mechanism of nitric oxide-mediated intestinal barrier failure: insight into the pathogenesis of necrotizing enterocolitis. J Pediatr Surg 2006;41(2):294–9.
[26] Kirkpatrick AW, Ball CG, Nickerson D, D'Amours SK. Intraabdominal hypertension and the abdominal compartment syndrome in burn patients. World J Surg 2009;33(6):1142–9.

Please cite this article in press as: Huang Y, et al. A novel mathematical model to predict prognosis of burnt patients based on logistic regression and support vector machine. Burns (2016), http://dx.doi.org/10.1016/j.burns.2015.08.009