

Belief rule-based inference for predicting trauma outcome



Guilan Kong^{a,*}, Dong-Ling Xu^b, Jian-Bo Yang^b, Xiaofeng Yin^c, Tianbing Wang^c, Baoguo Jiang^{c,*},
Yonghua Hu^a

^a Medical Informatics Center, Peking University, Beijing 100191, China

^b Decision and Cognitive Sciences Research Centre, The University of Manchester, Manchester M15 6PB, UK

^c Traffic Medical Center, Peking University, Beijing 100191, China

ARTICLE INFO

Article history:

Received 11 June 2015

Revised 30 November 2015

Accepted 2 December 2015

Available online 11 December 2015

Keywords:

Belief rule base

Evidential reasoning approach

Logistic regression

Support vector machine

Artificial neural network

Outcome prediction

ABSTRACT

A belief rule-based inference methodology using the evidential reasoning approach (RIMER) is employed in this study to construct a decision support tool that helps physicians predict in-hospital death and intensive care unit admission among trauma patients in emergency departments (EDs). This study contributes to the research community by developing and validating a RIMER-based decision tool for predicting trauma outcome. To compare the prediction performance of the RIMER model with those of models derived using commonly adopted methods, such as logistic regression analysis, support vector machine (SVM), and artificial neural network (ANN), several logistic regression models, SVM models, and ANN models are constructed using the same dataset. Five-fold cross-validation is employed to train and validate the prediction models constructed using four different methods. Results indicate that the RIMER model has the best prediction performance among the four models, and its performance can be improved after knowledge base training with historical data. The RIMER tool exhibits strong potential to help ED physicians to better triage trauma, optimally utilize hospital resources, and achieve better patient outcomes.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Trauma has become one of the leading causes of mortality and disability worldwide. Trauma accounts for 16% of the global burden of disease, and 16,000 people die from injury daily. Death and disability from trauma frequently occur in low- and middle-income countries where approximately 90% of the total burden of trauma is reported [1,2]. In developed countries, pre-hospital triage can help stratify trauma patients into different severity levels, and different levels of trauma centers have been established to treat trauma patients with varying degrees of severity [3–5]. Unfortunately, no nationwide initial trauma assessment guidelines or tools exist to aid physicians in pre-hospital environments or emergency departments (EDs) in many less developed countries [6]. In China, a pre-hospital “120” (the ambulance call number used in China) emergency system adopts the principle of proximity to transport a trauma patient to the nearest hospital. Consequently, some severe trauma patients have been sent to low-level hospitals, which cannot treat severe trauma patients, so these patients have to be retransferred to higher-level hospitals. Trauma patient outcomes in China are rela-

tively poor compared with those in developed countries. More than 400,000 people die from injury in China each year, and trauma is the fifth leading cause of death after malignant tumors and cardiac, cerebral, and respiratory diseases. Trauma is considered the most common cause of death of young people aged 18–40 years in China [7]. Research shows that a large proportion of in-hospital mortality can be predicted and prevented if clinical deterioration is recognized early [8,9]. Therefore, for improved patient outcomes and optimal utilization of hospital resources, physicians in EDs need to provide a rapid initial assessment of illness severity for trauma patients immediately after their arrival at the hospital, so as to make appropriate decisions regarding the treatment of patients with a high probability of in-hospital death or intensive care unit (ICU) admission [10].

Vital signs including pulse rate, systolic blood pressure, respiratory rate, body temperature, and level of consciousness are used to assign an early warning score [11–14] to assess illness severity. In trauma care, other physiological scoring tools such as the Pre-Hospital Index [15], the Trauma Index [16,17], the Glasgow Coma Score [18,19], and the Revised Trauma Score [20] have been developed to assess trauma severity before detailed diagnoses can be made for trauma patients. The nature of existing physiological trauma severity assessment tools is to assign a severity score to a patient based on physician observations and the instrumentally measured vital signs

* Corresponding authors. Tel.: +8618710098511.

E-mail addresses: Guilan.kong@hsc.pku.edu.cn (G. Kong),
Jiangbaoguo@vip.sina.com (B. Jiang).

of the patient. However, the existing trauma scoring tools to aid physicians in EDs are most often used to stratify patients into different severity levels and rarely to predict the probability of in-hospital death and ICU admission. To explore the relationship between clinical variables available for data collection in EDs and trauma outcome, such as in-hospital death and ICU admission, logistic regression (LR) [21,22], support vector machines (SVMs) [23,24], and artificial neural networks (ANNs) [25–28] are usually employed to construct prediction models. None of the LR, SVM, or ANN models require concrete knowledge about the relationship between antecedent factors and dependent outcomes, and these methods are completely data-driven, which means sufficiently large sample data are needed to learn prediction models. The performance and efficacy of data-driven prediction models are determined not only by the learning dataset, but also by the unknown dataset to which the prediction model is applied.

In the present study, vital signs are used as antecedent factors to predict in-hospital death and ICU admission. We propose the use of a generic belief rule-based inference methodology using the evidential reasoning approach (RIMER) [29] to develop a clinical decision model. This model is aimed at helping ED physicians predict the probability of in-hospital death and ICU admission for trauma patients [30]. In RIMER, an initial belief rule base (BRB) consisting of belief rules for predicting in-hospital death and ICU admission must first be constructed based on domain expert knowledge and clinical experiences. Inference with the BRB is implemented using the evidential reasoning (ER) approach [31,32], which was originally proposed for combining multiple independent assessments of one alternative on individual criteria or attributes. The ER approach can handle both quantitative and qualitative attributes or criteria under uncertainties [33,34]. In an RIMER-based prediction model, the inputs include the clinical values of vital signs, which are used to infer with or match the belief rules in the BRB. In ER-based inference, the packet antecedent of each belief rule triggered by the inputs is considered a basic attribute with an attribute weight, which is assessed using all possible consequents with belief degrees as presented in the BRB. Thus, assessments on the packet antecedents of multiple triggered belief rules can be combined by the ER approach to achieve aggregated belief degrees in all possible consequents of the BRB. The output of the model is a combined belief degree or probability linked to the trauma outcome, including in-hospital death and ICU admission for each patient. The BRB in the model can be fine-tuned by accumulated historical data. The model is transparent in that all belief rules can be checked by experienced physicians for validity and the inference process is also transparent and can be traced for better informed decision making.

In addition, LR-based, SVM-based, and ANN-based prediction models are constructed for comparison using the same dataset, whereas the antecedent variables and dependent outcome of each of these completely data-driven prediction models are the same as those of the RIMER model. To validate the prediction performance of all RIMER-based, LR-based, SVM-based, and ANN-based models, a five-fold cross-validation method is applied.

The remainder of this paper is structured as follows. The Materials and methodology section provides a brief introduction to the following: the data source, RIMER methodology, LR analysis, SVM, ANN, five-fold cross-validation method, and area under the receiver operating characteristic (ROC) curve (AUC), which we used to measure prediction performance. The Results section compares the prediction performance of the RIMER-based model before and after BRB training, the LR-based model, an optimal SVM-based model, and an optimal ANN-based model in each training round. Discussion section elaborates on the four different types of prediction models, especially the advantages and limitations of the RIMER model. Conclusions section summarizes this study and presents conclusions drawn from the study.

2. Materials and methodology

2.1. Dataset

A sample of trauma patients sent to Kailuan Hospital, North China, between 2008 and 2009 was employed for prediction model development and validation. Patients were included for analysis if they met the following criteria: (a) directly sent to the ED from an accident site; (b) with the five vital signs recorded upon their arrival at the ED; and (c) possible to retrieve corresponding in-hospital data. No further restrictions were made on the severity or characteristics of the cases. A total of 1299 trauma patients were directly sent to the ED at Kailuan Hospital within the sampling period, among which 1190 (91.61%) had both ED vital signs data and in-hospital data. The remaining 109 patients had either missing data on vital signs or missing in-hospital data, so they were excluded from data analysis.

The primary outcome of this study is a composite one, including in-hospital death and ICU admission.

2.2. RIMER methodology

In the RIMER methodology, traditional IF-THEN rules are extended to belief rules by embedding belief degrees in all possible consequents of a rule. Meanwhile, other knowledge representation parameters, including rule weights, antecedent attribute weights, and consequent belief degrees, are embedded in the belief rules. Inference with a BRB in a RIMER system is implemented using ER. The RIMER system presents the advantages of using belief rules to represent clinical domain knowledge under uncertainty and inference with uncertain clinical data using the ER approach. The knowledge representation parameters, including rule weights, antecedent attribute weights, and consequent belief degrees in the BRB, can be fine-tuned or trained using accumulated historical data [35]. The RIMER methodology has been employed to stratify patients with cardiac chest pain [36,37], diagnose lymph node metastasis in gastric cancer [38,39], and many other areas [40–42]. A brief introduction to BRB, inference with BRB, and the training of BRB follows.

2.2.1. BRB

A belief rule can be described as R_k :

$$\text{If } A_1^k \wedge A_2^k \wedge \dots \wedge A_{T_k}^k,$$

$$\text{then } \{(D_1, \beta_{1k}), (D_2, \beta_{2k}), \dots, (D_N, \beta_{Nk})\} \left(\beta_{jk} \geq 0, \sum_{j=1}^N \beta_{jk} \leq 1 \right),$$

with a rule weight θ_k and attribute weights $\delta_1, \delta_2, \dots, \delta_{T_k}$,

$$k \in \{1, \dots, L\}, \quad (1)$$

where $A_i^k (i = 1, \dots, T_k)$ is the referential category or grade of the i th antecedent attribute used in the k th rule; $\beta_{jk} (j = 1, \dots, N; k = 1, \dots, L)$ is the belief degree assigned to consequent D_j , and it can initially be given by experts; $\delta_i (i = 1, \dots, T_k)$ is the antecedent attribute weight representing the relative importance of the i th attribute; and θ_k is the rule weight representing the relative importance of the k th rule. L represents the number of all belief rules in the rule base. T_k is the number of all antecedent attributes used in the k th belief rule. N is the number of all possible consequents in the BRB. Traditional IF-THEN rule can be represented as a special case of belief rule with only one consequent, and the consequent belief degree is always 100%.

Initial belief rules in this study were provided by domain experts, and the five vital signs, namely, body temperature, respiratory rate, systolic blood pressure, pulse rate, and level of consciousness, are used as antecedent factors in the rule base. The possible consequents of the rule base include “occurrence of in-hospital death or ICU

Table 1
Initial belief rules provided by experts [30].

No.	Antecedent					Consequent
	A ¹	A ²	A ³	A ⁴	A ⁵	
1	AN	AN	AN	AN	AN	{{(occurrence, 70%), (nonoccurrence, 30%)}}
2	AN	AN	AN	AN	N	{{(occurrence, 60%), (nonoccurrence, 40%)}}
3	AN	AN	AN	N	AN	{{(occurrence, 60%), (nonoccurrence, 40%)}}
4	AN	AN	AN	N	N	{{(occurrence, 50%), (nonoccurrence, 50%)}}
5	AN	AN	N	AN	AN	{{(occurrence, 60%), (nonoccurrence, 40%)}}
6	AN	AN	N	AN	N	{{(occurrence, 50%), (nonoccurrence, 50%)}}
7	AN	AN	N	N	AN	{{(occurrence, 50%), (nonoccurrence, 50%)}}
8	AN	AN	N	N	N	{{(occurrence, 30%), (nonoccurrence, 70%)}}
9	AN	N	AN	AN	AN	{{(occurrence, 60%), (nonoccurrence, 40%)}}
10	AN	N	AN	AN	N	{{(occurrence, 50%), (nonoccurrence, 50%)}}
11	AN	N	AN	N	AN	{{(occurrence, 50%), (nonoccurrence, 50%)}}
12	AN	N	AN	N	N	{{(occurrence, 30%), (nonoccurrence, 70%)}}
13	AN	N	N	AN	AN	{{(occurrence, 50%), (nonoccurrence, 50%)}}
14	AN	N	N	AN	N	{{(occurrence, 30%), (nonoccurrence, 70%)}}
15	AN	N	N	N	AN	{{(occurrence, 30%), (nonoccurrence, 70%)}}
16	AN	N	N	N	N	{{(occurrence, 10%), (nonoccurrence, 90%)}}
17	N	AN	AN	AN	AN	{{(occurrence, 60%), (nonoccurrence, 40%)}}
18	N	AN	AN	AN	N	{{(occurrence, 50%), (nonoccurrence, 50%)}}
19	N	AN	AN	N	AN	{{(occurrence, 50%), (nonoccurrence, 50%)}}
20	N	AN	AN	N	N	{{(occurrence, 30%), (nonoccurrence, 70%)}}
21	N	AN	N	AN	AN	{{(occurrence, 50%), (nonoccurrence, 50%)}}
22	N	AN	N	AN	N	{{(occurrence, 30%), (nonoccurrence, 70%)}}
23	N	AN	N	N	AN	{{(occurrence, 30%), (nonoccurrence, 70%)}}
24	N	AN	N	N	N	{{(occurrence, 10%), (nonoccurrence, 90%)}}
25	N	N	AN	AN	AN	{{(occurrence, 50%), (nonoccurrence, 50%)}}
26	N	N	AN	AN	N	{{(occurrence, 30%), (nonoccurrence, 70%)}}
27	N	N	AN	N	AN	{{(occurrence, 30%), (nonoccurrence, 70%)}}
28	N	N	AN	N	N	{{(occurrence, 10%), (nonoccurrence, 90%)}}
29	N	N	N	AN	AN	{{(occurrence, 30%), (nonoccurrence, 70%)}}
30	N	N	N	AN	N	{{(occurrence, 10%), (nonoccurrence, 90%)}}
31	N	N	N	N	AN	{{(occurrence, 10%), (nonoccurrence, 90%)}}
32	N	N	N	N	N	{{(occurrence, 5%), (nonoccurrence, 95%)}}

admission” and “nonoccurrence of in-hospital death or ICU admission.” The referential grades of the five vital signs include “normal (N)” and “abnormal (AN).” One example of a belief rule is “IF body temperature is AN, respiratory rate is AN, systolic blood pressure is AN, pulse rate is AN, and level of consciousness is AN, THEN {(occurrence of in-hospital death or ICU admission, 70%), (nonoccurrence of in-hospital death or ICU admission, 30%)},” where {(occurrence of in-hospital death or ICU admission, 70%), (nonoccurrence of in-hospital death or ICU admission, 30%)}

means that if the clinical data of one patient satisfy the compact antecedent of the rule, the patient will have in-hospital death or ICU admission with 70% certainty and 30% certainty of eliminating the two adverse events. In total, 32 belief rules in the BRB are listed in Table 1, where A¹ represents body temperature, A² is the respiration rate, A³ is the pulse rate, A⁴ is the systolic blood pressure, and A⁵ represents the level of consciousness.

2.2.2. Inference with BRB using the ER approach

Inference with the BRB in the RIMER system is implemented using the ER algorithm [31,43].

First, transform numerical values or subjective judgments associated with antecedent clinical variables to belief distributions on cor-

responding reference categories used in the BRB. This transformation can be implemented by rule- or utility-based equivalence transformation techniques [32]. For example, set the input value for the *i*th antecedent attribute as *U_i* along with a belief degree ϵ_i . The input value can be transformed as

$$S(U_i, \epsilon_i) = \{ (A_{ij}, \alpha_{ij}); j = 1, \dots, J_i \}, i = 1, \dots, T \tag{2}$$

where *A_{ij}* is the *j*th referential category of the *i*th antecedent attribute, α_{ij} is the degree to which the input *U_i* with belief degree ϵ_i belongs to the referential category *A_{ij}* with $\alpha_{ij} \geq 0$ and $\sum_{j=1}^{J_i} \alpha_{ij} \leq 1$ (*i* = 1, ..., *T*), and *J_i* is the number of all referential categories of the *i*th antecedent attribute.

In this study, the degree of “N” or “AN” associated with each vital sign variable is calculated from the severity score assigned to the corresponding vital sign on the basis of physician observation or instrumental measurement. The criteria for the severity score assignment are based on the early warning score, as introduced by Smith et al. [12]. Details of the scoring criteria are shown in Table 2.

In Step One of this input transformation, a severity score equal to 0, 1, 2, or 3 is assigned to a patient. This score pertains to his or her vital signs on the basis of physician observations or the numerical

Table 2
Early warning scores for trauma patients [12].

Score	3	2	1	0	1	2	3
Pulse rate (bpm)		<40	40–50	51–100	101–110	111–130	>130
Systolic blood pressure (mmHg)	<70	70–80	81–100	101–200		>200	
Respiration rate (breaths/min)		<9		9–14	15–20	21–30	>30
Temperature (°C)		<35.1	35.1–36.5	36.6–37.5	>37.5		
Consciousness				A	V	P	U

A=Alert; V=Reaction when verbally addressed; P=Response to pain; U=No reaction.

values (exact numerical values for body temperature, respiration rate, pulse rate, and systolic blood pressure, and physician observations for level of consciousness). The assignment is based on the early warning score criteria, where a score of 0 means normal status and the scores 1, 2, and 3 mean abnormal status. The degree of abnormality increases with the score. Step Two involves transforming a numerical severity score to a distribution on the referential grades “N” and “AN” based on the following transformation rule. With pulse rate as an example, the pulse rate of a patient is set as “N” with 100% certainty if its associated severity score is 0, “AN” with 100% certainty if its associated severity score is 3, {(N, 66.67%), (AN, 33.33%)} if its associated severity score is 1, and {(N, 33.33%), (AN, 66.67%)} if its associated severity score is 2. For the other four vital signs, that is, body temperature, respiration rate, systolic blood pressure, and level of consciousness, we used the same rules as discussed above for input transformation.

Second, calculate the activation weight of each belief rule in the BRB based on the weight associated with each rule and the degrees of belief in various reference categories after input transformation. Suppose ω_k ($k = 1, \dots, L$) is the rule activation weight, which measures the degree to which the packet antecedent A^k in the k th rule is activated by the inputs. It can be calculated by

$$\omega_k = \frac{\theta_k \alpha_k}{\sum_{j=1}^L \theta_j \alpha_j} = \frac{\theta_k \prod_{i=1}^{T_k} (\alpha_i^k)^{\delta_{ki}}}{\sum_{j=1}^L \left[\theta_j \prod_{i=1}^{T_k} (\alpha_i^j)^{\delta_{ji}} \right]}, \quad k = 1, \dots, L \quad (3)$$

where $\delta_{ki} = \frac{\delta_{ki}}{\max_{i=1, \dots, T_k} \{\delta_{ki}\}}$ ($0 \leq \delta_{ki} \leq 1$) is transformed from the antecedent attribute weight δ_{ki} ($i = 1, \dots, T_k; k = 1, \dots, L$) representing the relative importance of the i th antecedent attribute in the k th rule [29]. α_i^k ($i = 1, \dots, T_k$) is the individual matching degree to which the input U_i ($i = 1, \dots, T_k$) belongs to A_i^k ($i = 1, \dots, T_k; k = 1, \dots, L$), which is the referential category of the i th antecedent attribute in the k th rule, and it is generated from the input transformation as described by (2), with $\alpha_i^k \geq 0$ and $\sum_{i=1}^{T_k} \alpha_i^k \leq 1$. $\alpha_k = \prod_{i=1}^{T_k} (\alpha_i^k)^{\delta_{ki}}$ ($k = 1, \dots, L$) is the combined matching degree to which the input vector matches the packet antecedent A^k in the k th rule. T_k is the total number of antecedent attributes in the k th rule. L is the total number of belief rules in the BRB.

Third, use the ER algorithm to aggregate all the activated belief rules in the BRB, thereby generating a distributed trauma outcome prediction with combined belief degrees distributed on all possible consequents or outcomes. The analytic ER algorithm [44] is given as follows:

$$\beta_j = \frac{\mu \times \left[\prod_{k=1}^L (\omega_k \beta_{jk} + 1 - \omega_k \sum_{j=1}^N \beta_{jk}) - \prod_{k=1}^L (1 - \omega_k \sum_{j=1}^N \beta_{jk}) \right]}{1 - \mu \times \left[\prod_{k=1}^L (1 - \omega_k) \right]}, \quad j = 1, \dots, N \quad (4)$$

where $\mu = \left[\sum_{j=1}^N \prod_{k=1}^L (\omega_k \beta_{jk} + 1 - \omega_k \sum_{j=1}^N \beta_{jk}) - (N-1) \times \prod_{k=1}^L (1 - \omega_k \sum_{j=1}^N \beta_{jk}) \right]^{-1}$, ω_k is the rule activation weight calculated by (3), and β_j ($j = 1, \dots, N$) is the final combined belief degree associated with the corresponding consequent D_j ($j = 1, \dots, N$).

In this study, all rule weights and antecedent attribute weights are assumed to be equal, and a final combined outcome prediction with belief degrees distributed on “occurrence of in-hospital death or ICU admission” and “nonoccurrence of in-hospital death or ICU admission” was generated for each case that was included for analysis.

Note that in this study, all recorded clinical values or clinician judgments were certain values and could be transformed to early warning scores without uncertainty. Therefore, those early warning scores could be further transformed into precise distributed assessments on “N” and “AN” using the previously discussed transformation rules. In other disease diagnoses, there may be ambiguous judgments or missing values of some clinical variables that cannot be measurable at a certain time point.

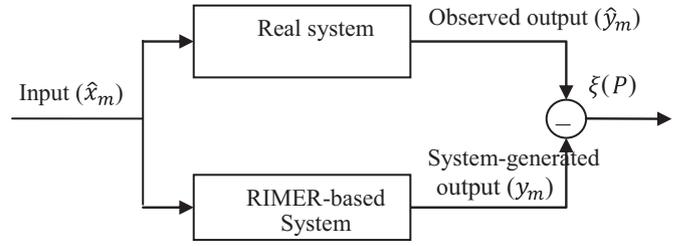


Fig. 1. BRB training process.

If clinician judgements contain ambiguity and we cannot decide whether the patient is normal or abnormal with 100% certainty, we can convert the uncertain judgments to some degree (α) of normal and some degree (β) of abnormal. If $\alpha + \beta = 1$, then the inputted information is complete and the transformed inputs are complete distributed assessments on “N” and “AN”, which is the same as the transformation of those early warning scores. Thus the inference is the same as what were discussed above. If $\alpha + \beta < 1$, it brings ignorance to the inputs and the incomplete inputs would cause incomplete assessments on different consequents. Although the procedures of inference with incomplete inputs are the same as that for the complete inputs, the final distributed assessments on different consequents after aggregating all triggered belief rules should not be the inferred certain belief degrees, because there is also ignorance in the consequents. In the latter situation, the final distributed assessments on different consequents would be interval belief degrees.

If missing values for the input clinical variables exist, the BRB with certain consequent belief degrees (Table 1) can be extended to a BRB with interval belief degrees. A belief rule with interval belief degrees can be something like “IF body temperature is AN, respiratory rate is AN, systolic blood pressure is AN, pulse rate is AN, and level of consciousness is unknown, THEN {(occurrence of in-hospital death or ICU admission, [60–70%]), (nonoccurrence of in-hospital death or ICU admission, [30–40%])}.” Inference with this type of BRB with interval belief degrees needs to be implemented using the ER approach for combining multiple independent assessments with interval belief degrees, as proposed by Wang et al. [33]. The inference steps of the BRB with interval belief degrees are similar to the aforementioned steps. All sample data included in this study are certain data; thus, the details of the uncertainty handling capability of the RIMER model will not be discussed. Readers can refer to [33] for details of using the ER approach to deal with interval belief degrees.

2.2.3. Training BRB using historical data

The belief rules in the initial BRB are provided by domain experts and may not represent the real situation with 100% accuracy. A change in the knowledge representation parameters can lead to changes in the performance of a RIMER-based prediction model. Thus, the knowledge representation parameters need to be fine-tuned or trained using accumulated historical data. Suppose (\hat{x}_m, \hat{y}_m) ($m = 1, \dots, M$) is used to denote input–output pairs of M clinical cases in the training dataset. The process of training BRB from these M datasets can be depicted as in Fig. 1, where \hat{x}_m is the clinical inputs of the m th case, y_m is the combined degree of belief generated by the RIMER-based model (implemented using Eqs. (2), (3), and (4)) in “occurrence of in-hospital death or ICU admission” with the inputs \hat{x}_m , \hat{y}_m is the actual or observed output of the m th case (a binary value), and $\xi(P)$ represents the difference between the observed output and the system-generated output.

The objective of BRB training is to determine an optimal set of knowledge representation parameters ($\beta_{jk}, \delta_i, \theta_k$) for the BRB by minimizing the discrepancies between system-generated outputs and the observed outcomes for all cases in the training dataset. In this study, all rule weights and antecedent attribute weights were set to be equal, and only consequent belief degrees β_{jk} were set as training parameters in the BRB training process. The constraints of the training model are given as $0 \leq \beta_{jk} \leq 1 (j = 1, \dots, N; k = 1, \dots, L)$ and $\sum_{j=1}^N \beta_{jk} = 1$, where L is the number of belief rules, and N is the number of all possible consequents in the BRB. The *fmincon* function in MATLAB was used to solve the single-objective optimization model. Except for *MaxIter* set at 60 and *TolX* at 0.00001, all other options of the function were set to the default values in MATLAB. The fine-tuned consequent belief degrees of each belief rule in the BRB after each training round are presented in the Results section.

2.3. LR analysis

LR measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities. LR is a maximum-likelihood method that has been used in many trauma outcome studies [21,22,25,45,46]. The binary LR model is used in this study. The binary logistic model is used to predict a binary outcome based on one or more predictor variables. In the model, the logit transformation of the probability p_i linked to a particular outcome of the i th case is expressed as a linear prediction function of the independent or predictor variables

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_j x_{j,i} + \dots + \beta_m x_{m,i} \tag{5}$$

where $j = 1, \dots, m$, and m is the number of predictor variables. The predictors can be categorical dummy variables or continuous variables. The coefficient β_j represents the change in the logit for each unit change in the associated predictor.

Solving for p_i results in the individual probability of the particular outcome as

$$p_i = \frac{\exp[\beta_0 + \beta_1 x_{1,i} + \dots + \beta_j x_{j,i} + \dots + \beta_m x_{m,i}]}{1 + \exp[\beta_0 + \beta_1 x_{1,i} + \dots + \beta_j x_{j,i} + \dots + \beta_m x_{m,i}]} \tag{6}$$

Models derived using LR analysis typically employ only variables that are statistically significant predictors of a particular outcome. In LR analysis, a backward stepwise method can be used to eliminate variables without statistical significance from the model in an iterative process, whereas the fit of the model is tested after each variable is eliminated to ensure that the model still adequately fits the training data.

For deriving LR models, we have attempted to transform continuous variables, such as body temperature, pulse rate, respiration rate, and systolic blood pressure, to categorical variables in accordance with the early warning score criteria used in the RIMER methodology for input transformation. However, the LR model derived using transformed categorical predictor values obtained a worse prediction performance than the LR model that was derived using original continuous predictor values. Thus, the continuous values of the five vital signs, except level of consciousness, were used in the LR model derivation; only the value pertaining to level of consciousness was categorical like that in the RIMER model. In terms of the predictor variables in the LR model, we used both the backward stepwise method and the simple entering method. The latter method used all available predictors without considering corresponding statistical significance in model derivation. We found that the two LR models constructed with different independent predictor variables showed similar prediction performance. As the RIMER model employed all the five vital signs as antecedent attributes in its BRB, we also used

a saturated LR model with the five vital signs as predictors in the LR model to compare the prediction performance of both models in the same prediction environment.

In this study, the same dataset used for RIMER training and testing was used for LR analysis. SPSS was employed for LR analysis.

2.4. SVM

SVM [47] is a supervised learning method used to perform dichotomy classification of multidimensional feature vectors. The basic idea under the SVM method is to transform the input features into a higher-dimensional space where the two classes can be linearly separated by a high-dimensional surface known as hyperplane. Given a training dataset $\{(x_i, y_i) | x_i \in R^L, y_i \in \{-1, +1\}\}_{i=1}^N$ with N samples, where y_i is either 1 or -1, indicating the class to which the data point x_i belongs. Each x_i is one L -dimensional real vector. The decision function of SVM is described as

$$h(x) = \mathbf{w}^T \varphi(x) + b \tag{7}$$

where $\varphi(x)$ represents a mapping of sample x from the input space to a high-dimensional feature space, \mathbf{w} denotes the normal vector to the learned hyperplane, and b is the model bias.

An SVM classifier needs to satisfy the following conditions:

$$\begin{cases} \mathbf{w}^T \varphi(x_i) + b \geq 1 \text{ if } y_i = +1 \\ \mathbf{w}^T \varphi(x_i) + b \leq -1 \text{ if } y_i = -1 \end{cases} \tag{8}$$

The equation $\mathbf{w}^T \varphi(x_i) + b = 0$ constructs a hyperplane that discriminates between the two classes, where $|b|/|\mathbf{w}|$ is the perpendicular distance from the hyperplane to the origin, and $1/|\mathbf{w}|$ is the shortest distance from the hyperplane to the closest positive (negative) sample. A typical two-class problem classified using SVM is shown in Fig. 2. Line H represents the separating hyperplane, and each of the two half spaces separated by H corresponds to one class, H_1 for $y_i = +1$, and H_2 for $y_i = -1$. The SVM determines a hyperplane with the maximal margin. The margin of an SVM classifier is the minimal distance of any training point to the hyperplane, which is the distance between the dotted lines H_1 and H_2 and the solid line H , as shown in Fig. 2. Therefore, the margin of a hyperplane will be $1/|\mathbf{w}| + 1/|\mathbf{w}|$. Calculating the optimal separating hyperplane is equivalent to maximizing the separation margin or distance between the two dotted lines H_1 and H_2 [48]. The training points that lie on the dotted lines H_1 and H_2 are referred to as support vectors and are colored in gray in Fig. 2.

In practice, a separating hyperplane may not exist. To allow for the possibility of training samples violating the edges of the margin, slack variables $\xi_i (i = 1, 2, \dots, N)$ with $\xi_i \geq 0$ are introduced. We can optimize the values of \mathbf{w} and b by solving the following optimization problem:

$$\begin{aligned} &\underset{\mathbf{w}, b, \xi_i}{\text{minimize}} && \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \\ &\text{subject to} && y_i (\mathbf{w}^T \varphi(x_i) + b) \geq 1 - \xi_i, \\ &&& \xi_i \geq 0, \quad i = 1, 2, \dots, N \end{aligned} \tag{9}$$

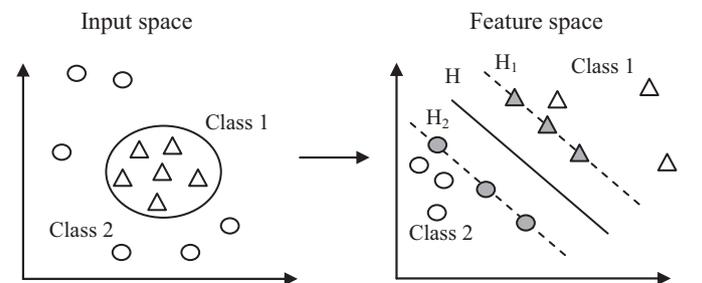


Fig. 2. Mapping between input space and feature space in a two-class problem classified with the SVM. Class 1: triangle; Class 2: circle. [48].

where $C (>0)$ is the regularization parameter (penalty factor) that regulates the relationship between training accuracy and model generalization, slack variables ξ_i are used to achieve a soft margin, and φ is a non-linear mapping from an input space into a feature space. By introducing the Lagrange multiplier α_i , a corresponding dual problem can be derived as the following quadratic programming problem:

$$\begin{aligned} \text{maximize}_{\alpha_i} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\ \text{subject to} \quad & \sum_{i=1}^N \alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, i = 1, 2, \dots, N \end{aligned} \quad (10)$$

where k is a kernel function with $k(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$, which maps the input vectors into a suitable feature space. Commonly used kernel functions in SVM include linear kernel, polynomial kernel, and radial basis function kernel. Once the dual quadratic programming problem is solved, the resulting decision function at any test data point x is as follows:

$$h(x) = \mathbf{w}^T \varphi(x) + b = \sum_{i=1}^N \alpha_i y_i k(x_i, x) + b = \sum_{i \in SV} \alpha_i y_i k(x_i, x) + b \quad (11)$$

Only those data points for which α_i is nonzero are referred to as support vectors, and they define the decision function.

In the test phase, to obtain a binary outcome, we estimate the class of the test data point x based on $\text{sign}(h(x))$. To obtain a binary outcome together with probability information, we need to extend the SVM to provide probabilities. Details of probability estimates obtained using SVM can be found in [49].

The present study employed the LIBSVM toolbox (for MATLAB environment) developed by Chang and Lin [50]. The *svmtrain* and *svmpredict* functions in the LIBSVM toolbox were used for SVM model training and testing. The procedures of employing *svmtrain* and *svmpredict* to obtain an optimal SVM model are as follows. First, we used the general and default kernel function (radial basis function) of LIBSVM to build SVM. Second, we established training samples and called *svmtrain* in LIBSVM to train SVM. Third, we called *svmpredict* to predict the samples in the training set using well-trained SVM models. To find an optimal set of parameters C and γ for the radial basis function kernel SVM, we tested different sets of C and γ using a five-round model training method. In this method, the whole dataset was split into five folds (roughly similar data distribution in each fold) and used any four folds for model training in each training round. The ranges of C and γ were set as $[1, 2^4]$ and $[2^{-5}, 2^{-1}]$, respectively, and their initial values were set to be 1 and 2^{-5} , respectively. The value assignments of C and γ in the parameter searching process were set as $(2^0, 2^{0+step}, 2^{0+step+step}, \dots, 2^4)$ and $(2^{-5}, 2^{-5+step}, 2^{-5+step+step}, \dots, 2^{-1})$ ($step=1$) respectively. Finally, we selected the pair of parameters C and γ with the best average *mean squared error* (*MSE*) in the training dataset after five rounds of model training. After the parameters C and γ were determined, the same dataset for LR model training and testing was used for SVM model analysis. Instead of using the originally recorded values of body temperature, respiration rate, pulse rate, systolic blood pressure, and the numerical score of the level of consciousness, we normalized the input values to $[0, 1]$ in the SVM model analysis, as recommended by Hsu et al. [51]. The outcome of the SVM model included two values ranging from 0 to 1, which indicate the probabilities linked to the occurrence and nonoccurrence of ICU admission or in-hospital death, respectively.

2.5. ANN

ANN has different types of structures [52]. One of the most common ANN structures is the multi-layer perceptron network, which is

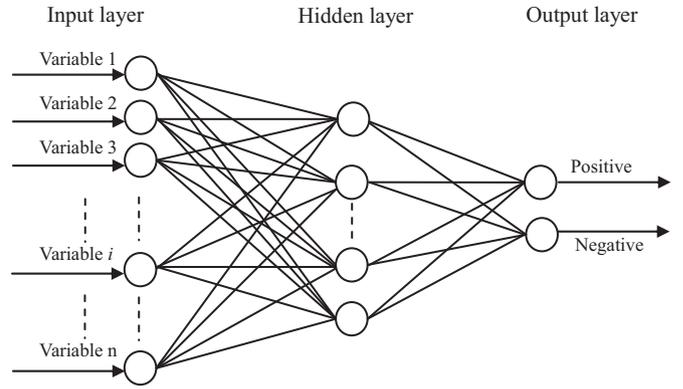


Fig. 3. Architecture of a typical multi-layer perceptron network.

a feedforward type of the neural network [48,53]. A multi-layer perceptron with a back-propagation algorithm was used in the current study. A multi-layer perceptron network consists of multiple layers of neurons: one input layer receiving external inputs, one output layer generating the classification results, and one or more hidden layers. Each layer is fully connected to the next one. A typical multi-layer perceptron network structure with an input layer, a hidden layer, and an output layer is shown in Fig. 3.

Except for the input layer, each neuron in the network is a computational element that receives information from neurons of the previous layer, produces an output signal by an activation or transfer function, and passes it on to neurons of the next layer. The principle of the network is that when data are presented at the input layer, the neurons in the consecutive layers run computations until an output value is obtained at each of the output neurons. This output will indicate the appropriate outcome for the input data.

Each neuron in the hidden layer receives an activation signal, which is the weighted sum of all inputs that enter into the neuron. The resulting sum is used to generate an output of the neuron through an activation function. This process is defined as follows:

$$v_j = \sum_{i=1}^p x_i w_{ij} \quad (12)$$

$$y_j = f(v_j) \quad (13)$$

where v_j is the linear combination of inputs x_1, x_2, \dots, x_n , w_{ij} is the connection weight between the input x_i and the j th neuron, f is the activation function used for the neuron, and y_j is the output of the neuron. The most commonly used activation function is the *sigmoid* function, and its general form is expressed as

$$f(t) = \frac{1}{1 + e^{-t}} \quad (14)$$

The neurons in the output layer receive the following activation signals from the hidden neurons:

$$h_k = \sum_{j=1}^n y_j w_{jk} \quad (15)$$

where w_{jk} is the weight of the connection between neuron j in the hidden layer and neuron k in the output layer. These activation signals are then transformed again to generate the outputs of the network as follows:

$$o_k = f(h_k) = \frac{1}{1 + e^{-h_k}} \quad (16)$$

where o_k is the output generated by the network.

Different algorithms have been developed to learn a multi-layer perceptron network using a set of training samples. The most popular algorithm is the backpropagation algorithm, a gradient descent

method that adjusts the weights of interconnections to minimize output error. The error function of the output neuron is defined as

$$E = \frac{1}{2} \sum_k (d_k - o_k)^2 \quad (17)$$

where d_k and o_k are the observed and network-generated values of the outputs, respectively. The objective of network learning is to minimize the error E for the network to achieve the best performance.

All weight vectors w are initialized with small random values from a pseudorandom sequence generator. The following steps are repeated until the convergence (i.e., when the error E is below a pre-set value).

The weights between the hidden layer and the output layer w_{jk} are updated by

$$w_{jk}(t+1) = w_{jk}(t) + \Delta w_{jk}(t) \quad (18)$$

where

$$\Delta w_{jk}(t) = -\eta \frac{\partial E(t)}{\partial w_{jk}} \quad (19)$$

$$\Delta w_{jk}(t) = -\eta \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial h_k} \frac{\partial h_k}{\partial w_{jk}} = \eta (d_k - o_k) f'(h_k) y_j = \eta \delta_{ok} y_j \quad (20)$$

where η is the learning rate, and δ_{ok} is the error signal of the neuron k in the output layer given by

$$\delta_{ok} = (d_k - o_k) f'(h_k) \quad (21)$$

To avoid oscillation at large η , the changes in the weights are made to depend on the past changes by adding a momentum term α :

$$\Delta w_{jk}(t+1) = \eta \delta_{ok} y_j + \alpha \Delta w_{jk}(t) \quad (22)$$

Similarly, the changes in the weights between the neurons in the input and hidden layer, $\Delta w_{ij}(t)$, are given by

$$\Delta w_{ij}(t) = -\eta \frac{\partial E}{\partial w_{ij}} = -\eta \sum_k \left(\frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial h_k} \frac{\partial h_k}{\partial y_j} \frac{\partial y_j}{\partial v_j} \frac{\partial v_j}{\partial w_{ij}} \right) = \eta \delta_{yj} x_i \quad (23)$$

where δ_{yj} is the error signal of the neuron j in the hidden layer and is given by

$$\delta_{yj} = f'(v_j) \sum_k \delta_{ok} w_{jk} \quad (24)$$

Similarly, a momentum term can be added to prevent oscillation.

Through the learning process, a set of optimal weights can ensure that the resulting vector generated by the network is the same or sufficiently close to the observed output vector for each input vector.

A previous study [52] showed that an ANN with a single hidden layer with a sufficiently large number of neurons can relate any given set of inputs to outputs. Therefore, a three-layer ANN was constructed in this study.

The input variables of the ANN were the same as the inputs of the LR models. The only output of the ANN is a value ranging from 0 to 1, which may represent the probability of the primary outcome for each studied sample. Meanwhile, numerous thumb rules are used for determining the number of neurons in the hidden layer. An empirical equation used in this study is as follows.

$$h = \sqrt{m + n} + a \quad (25)$$

where h is the target number, m is the number of neurons in the output layer, n is the number of variables in the input layer, and a is an integer constant with $a \in [1, 10]$. According to the above formula, the number of neurons in the hidden layer may be between 3 and 12.

The *feedforwardnet*, *train*, and *sim* functions in MATLAB were employed to create, train, and test ANN models. In creating and training

the ANN, the training function was set as *traingdm*, the transfer function was set as *logsig*, and all other parameters in the ANN models were set to their default values in MATLAB. To find an optimal network structure, we used the trial-and-error method to train and test different ANN models with various numbers (ranging from 3 to 12) of neurons in the hidden layer. Similar to optimal parameter searching in SVM model analysis, the five-round model training method was used to train different ANN models. An ANN model with the best average MSE in the training dataset was selected after five rounds of ANN training. After the ANN model structure was determined, the same dataset used for SVM model analysis was employed for ANN model training and testing.

2.6. Five-fold cross-validation

The five-fold cross-validation method was employed in this study to train and validate all constructed prediction models. The studied dataset was split into five folds with equal cases (with each fold containing roughly the same proportions of the two types of outcome). A single fold out of the five folds is retained as the validation set for model testing, and the remaining four folds are used as the training set for model derivation. Thus, the four constructed models were trained and tested five times using any four folds as the training set and the remaining fold as the test set.

2.7. Performance comparison of prediction models

Plotting the ROC curve is a commonly used method of illustrating the discriminatory accuracy of a diagnostic tool for detecting or predicting whether a patient has a disease or outcome [54]. In medicine, a person is assessed as diseased (positive) or healthy (negative) depending on whether the corresponding test value is greater than, less than, or equal to a given threshold value. Associated with any threshold value are the probability of a true positive (*sensitivity*) and the probability of a true negative (*specificity*). The theoretical ROC curve is a plot of *sensitivity* versus *1-specificity* for all possible threshold values. The most commonly used global index of diagnostic accuracy is the AUC [54]. The AUC has also been employed to measure the performance of prediction models in many trauma studies [25,45,46,55,56].

With the observed outcome as the gold standard, the following were used to plot the ROC curves: the final combined belief degree distributed on “occurrence of in-hospital death or ICU admission” generated by the RIMER model, the predicted probabilities of in-hospital death or ICU admission generated by the LR model, the SVM model, and the ANN model in each test fold. The average AUC value was used as a measure to compare the prediction performance of the four different models.

3. Results

The consequent belief degree in the “occurrence of in-hospital death or ICU admission” of each belief rule in the RIMER model before and after each training round is shown in Table 3, where R^1 , R^2 , R^3 , R^4 , and R^5 represent the first, second, third, fourth, and fifth training rounds, respectively. The sum of consequent belief degrees in “occurrence” and “nonoccurrence” of in-hospital death or ICU admission in each belief rule is always equal to 1 in the RIMER model constructed in this study.

Table 3 shows that the difference between fine-tuned consequent belief degrees and initial values is not large in each training round. This information indicates that the trained BRB has not significantly departed from the initial BRB provided by domain experts. The small variation also reveals that the fine-tuned consequent belief degrees are reasonable and reflect the real situation of patient outcomes in the corresponding training dataset.

Table 3

Consequent belief degrees in in-hospital death or ICU admission before and after training in the RIMER model.

No.	Consequent belief degree in occurrence of in-hospital death or ICU admission					
	Initial	R ¹	R ²	R ³	R ⁴	R ⁵
1	0.700	0.747	0.714	0.731	0.700	0.743
2	0.600	0.582	0.579	0.564	0.538	0.530
3	0.600	0.843	0.694	0.797	0.652	0.863
4	0.500	0.545	0.494	0.527	0.533	0.564
5	0.600	0.717	0.714	0.845	0.886	0.885
6	0.500	0.528	0.482	0.476	0.488	0.406
7	0.500	0.680	0.594	0.630	0.535	0.610
8	0.300	0.132	0.103	0.000	0.042	0.002
9	0.600	0.623	0.607	0.615	0.600	0.621
10	0.500	0.491	0.498	0.500	0.494	0.493
11	0.500	0.658	0.552	0.589	0.536	0.661
12	0.300	0.575	0.357	0.345	0.469	0.473
13	0.500	0.554	0.515	0.534	0.530	0.507
14	0.300	0.498	0.321	0.357	0.426	0.190
15	0.300	0.103	0.347	0.387	0.250	0.177
16	0.100	0.000	0.004	0.070	0.048	0.044
17	0.600	0.623	0.607	0.615	0.600	0.621
18	0.500	0.520	0.449	0.429	0.392	0.381
19	0.500	0.850	0.600	0.978	0.999	0.979
20	0.300	0.179	0.252	0.282	0.231	0.203
21	0.500	0.718	0.590	0.696	0.731	0.608
22	0.300	0.032	0.195	0.061	0.033	0.012
23	0.300	0.643	0.600	0.659	0.640	0.534
24	0.100	0.000	0.000	0.008	0.000	0.003
25	0.500	0.511	0.503	0.507	0.500	0.510
26	0.300	0.329	0.293	0.317	0.315	0.319
27	0.300	1.000	0.392	0.752	0.863	0.999
28	0.100	0.498	0.169	0.068	0.223	0.205
29	0.300	0.395	0.326	0.359	0.363	0.299
30	0.100	0.037	0.009	0.000	0.004	0.000
31	0.100	0.358	0.534	0.340	0.530	0.448
32	0.050	0.000	0.000	0.000	0.000	0.000

Table 4

Coefficients in the LR models in different training rounds.

Coefficient	R ¹	R ²	R ³	R ⁴	R ⁵
β_0	0.256	9.639	6.241	2.453	1.170
β_1	-0.308	-0.502	-0.472	-0.374	-0.313
β_2	0.043	0.016	0.018	0.012	0.025
β_3	0.055	0.033	0.067	0.073	0.048
β_4	0.006	0.007	0.003	0.001	0.007
β_5	4.806	4.968	4.892	5.544	5.075
β_6	3.740	3.564	2.301	3.341	3.456
β_7	-15.889	3.853	3.209	2.930	3.271
β_8	3.579	3.607	3.270	2.243	2.925

model with $C = 8$ and $\gamma = 0.25$ to perform SVM model training and testing.

As discussed in Section 2.5, we employed an ANN model with one hidden layer and tested different ANN structures with 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12 neurons in the hidden layer. To find an optimal ANN structure, we used both originally recorded data and normalized data as in SVM analysis to perform ANN model training and testing. We found that ANN models performed much better with originally recorded values. The ANN model that could achieve the best average MSE in the training dataset was an ANN model with five neurons in the hidden layer.

The AUCs that represent the prediction performance of the RIMER model, the LR model, the optimal SVM model, and the optimal ANN model in different training and testing rounds are shown in Table 5.

As shown in Table 5, the prediction performance of the RIMER model before BRB training is stable, which shows that the BRB model can represent medical knowledge and mimic medical reasoning as an open-box simulator in a robust way. Its performance can be further improved after fine-tuning its parameters. This learning ability of the BRB model differentiates itself from a traditional rule-based system that is also an open-box simulator. Among the other three data-driven models, the models showing performances from the best to the worst are as follows: the LR model, SVM model, and ANN model. The prediction performance of the three data-driven models (LR, SVM, and ANN) are clearly not as stable as the RIMER model and may differ significantly when they are learned from different training datasets. For instance, the LR model may provide a perfect prediction performance such as that in the fourth training round but the prediction performance of the learned LR model may not be good enough in the first training round. This finding is expected because the LR model is a simulator with an assumed model structure; its parameters are completely determined by the training dataset. The same reasons apply to the SVM and ANN models.

4. Discussion

This study is the first to employ the RIMER methodology to develop a trauma outcome prediction tool. It is the first comparison

Coefficients in the derived LR model in different training rounds are shown in Table 4. β_0 represents the constant; β_1 , β_2 , β_3 , and β_4 represent coefficients associated with the continuous variables body temperature, respiration rate, pulse rate, and systolic blood pressure, respectively; and β_5 , β_6 , β_7 , and β_8 represent coefficients associated with four different levels of consciousness, which is a categorical predictor in different LR models derived in this study.

Table 4 indicates that coefficients in LR models after training with different datasets vary significantly, and the values of coefficients in each model are completely determined by data characteristics in the corresponding training dataset.

Regarding SVM models, we found through five rounds of training that the parameter pair [$C = 8$, $\gamma = 0.25$] for the radial basis function kernel SVM could help achieve the best average MSE in the training dataset. Therefore, we chose the radial basis function kernel SVM

Table 5

AUCs of the RIMER, LR, SVM, and ANN prediction models.

Models/Folds	1st fold	2nd fold	3rd fold	4th fold	5th fold	Average	
RIMER	Before training	0.821 (95%CI: 0.767–0.868)	0.993 (95%CI: 0.972–0.999)	0.912 (95%CI: 0.869–0.945)	0.973 (95%CI: 0.944–0.990)	0.981 (95%CI: 0.955–0.994)	0.936
	After training	0.850 (95%CI: 0.798–0.893)	0.995 (95%CI: 0.975–1.000)	0.940 (95%CI: 0.902–0.967)	0.988 (95%CI: 0.964–0.998)	0.985 (95%CI: 0.961–0.996)	0.952
LR	0.590 (95%CI: 0.524–0.653)	0.984 (95%CI: 0.959–0.996)	0.890 (95%CI: 0.842–0.927)	0.991 (95%CI: 0.970–0.999)	0.972 (95%CI: 0.942–0.989)	0.885	
SVM	0.712 (95%CI: 0.380–1.000)	0.983 (95%CI: 0.957–1.000)	0.820 (95%CI: 0.619–1.000)	0.735 (95%CI: 0.451–1.000)	0.857 (95%CI: 0.615–1.000)	0.821	
ANN	0.722 (95%CI: 0.403–1.000)	0.978 (95%CI: 0.945–1.000)	0.544 (95%CI: 0.273–0.815)	0.984 (95%CI: 0.967–0.995)	0.723 (95%CI: 0.392–1.000)	0.790	

CI: Confidence Interval.

study in which the prediction performance of the RIMER methodology is compared with the three most commonly used data-driven modeling methods LR, SVM, and ANN. Based on the study results, the prediction performance of the RIMER model is stable and can be improved after training. The initial BRB in the RIMER-based prediction tool is provided by domain experts, and its accuracy directly affects the prediction performance of the model.

LR analysis, SVM, and ANN, the three most commonly employed tools to derive outcome prediction tools in trauma studies, are completely data-driven. Data-driven methods have advantages in model derivation because they require no prior knowledge about a concrete relationship between independent predictors and a particular dependent outcome. However, the prediction performance of the derived LR, SVM, and ANN models may widely vary in different training rounds despite having the same model structure. This variation is attributed to the fact that the prediction performance of a model completely derived from historical data is determined by the training dataset.

Among the three data-driven models, LR analysis exhibits its own characteristics. Through LR analysis, we cannot only obtain an outcome prediction model, but also know which predictor shows a statistically significant correlation with the particular outcome. Only predictors with statistical significance are typically used to construct LR models. However, to compare the four models using the same predictive variables, we employed the same variables that were used as antecedent attributes in the RIMER methodology to construct LR models. We found that the LR model constructed with all five vital signs as predictors and the LR model with only statistical significant variables as predictors showed similar prediction performance.

Meanwhile, SVM and ANN models exhibit similar features. The final structure of either model is determined by the training dataset. Nevertheless, the optimal parameters and internal functions of both models need to be determined using a trial-and-error method, and the initial values of those parameters need to be set by system designers based on their domain knowledge and experiences. Therefore, different studies may yield different SVM or ANN models even though the same modelling methodology and dataset are used. Moreover, to system users, the LR, SVM, and ANN models are completely black boxes, and clinicians may be reluctant to use these prediction tools in their practice because they do not know what clinical rules have been applied in these tools.

As to the RIMER methodology, the results indicate that the initial BRB is robust, and fine-tuning the initial BRB using a historical dataset can help to make consequent belief degrees in the initial BRB more accurately reflect the real situation in the training dataset. The performance improvement of the RIMER model through training is not that significant because the initial BRB is already very close to reality and can already provide very good prediction performance. Alternatively, if the initial BRB is not provided by domain experts and deviates far from reality, the initial prediction performance could be poor, and the differences of consequent belief degrees in the BRB before and after training can be large. Thus, in such situations, the performance improvement may be obvious and significant, depending on the initial BRB and the training and testing datasets.

Compared with the LR, SVM, and ANN models, the RIMER model uses both historical datasets and domain expert knowledge. Due to the fact that the learned model performs less well on a new dataset than on the dataset used for model training [57], the performance of a completely data-driven model, such as the LR, SVM, and ANN models derived in this study, is not ensured. In a model developed by combining the knowledge-based method and the data training method, such as the RIMER-based model in this study, the prediction performance can be secured to a certain degree. In detail, the initial knowledge base in the RIMER model is provided by domain experts, and it is based on expert knowledge and clinical experiences that have been proven correct by numerous clinical cases, thereby assuring its clin-

ical robustness. Furthermore, fine-tuning the initial knowledge base (BRB) using a historical dataset improves the accuracy of the knowledge representation parameters to fit the specific data characteristics in the training dataset. The method also enhances the knowledge base to better represent the relationship between independent predictors and the particular dependent outcome. It may explain why the RIMER model provided a more robust and better prediction performance than the LR, SVM, and ANN models.

However, the RIMER methodology has its limitations. In other disease areas in which a concrete relationship between predictive variables and outcome variable is unavailable, an initial knowledge base will be difficult to construct. In such cases, the performance of the RIMER methodology will depend on data only, similar to the LR, SVM, and ANN modeling methodologies. Under the circumstances, machine learning methods such as decision tree learning can be employed to learn traditional "IF-THEN" rules from historical data first, and then those learned traditional rules can be extended to belief rules according to the distribution of real medical outcome in the training dataset. In such cases, it is necessary to train the initial belief rules using accumulated historical data before the knowledge base is put into practice.

In real implementation, the fine-tuned or trained consequent belief degrees should not be directly applied to infer with input clinical data immediately after training. All trained belief rules need to be approved by domain experts prior to their real application. This procedure helps ensure that only rational and approved knowledge is used in the knowledge-based prediction model. This method also guarantees the accuracy of the embedded knowledge base.

5. Conclusions

This study contributes to the research by developing and validating a RIMER-based outcome prediction model to assist ED physicians to predict the probability of in-hospital death and ICU admission among trauma patients. The RIMER model is compared to the most commonly used methods in deriving trauma outcome prediction models, which include LR analysis, SVM, and ANN. The five-fold cross-validation method is employed in this study to validate the RIMER-based, LR-based, SVM-based, and ANN-based prediction models. To compare the prediction performance of the newly developed RIMER model with those of the other three commonly used data-driven models, the AUC values that represent the prediction performance of the four models were obtained. The results show that the RIMER model performs the best. Meanwhile, its initial BRB provided by trauma experts exhibits robust prediction performance, which can be further improved after fine-tuning through historical trauma data. The LR model performs better than the SVM and ANN models. Among all models, the ANN model shows the least AUC.

To conclude, the RIMER methodology provides an innovative way for clinical decision science researchers to fully benefit from both expert knowledge and historical patient data to develop a decision support tool to assess illness severity and predict patient outcome. Compared with LR analysis, SVM, and ANN, the RIMER-based model provides the best prediction performance. The RIMER tool exhibits strong potential to help ED physicians to better triage trauma, optimally utilize hospital resources, and achieve better patient outcomes. In our future work, we plan to implement the RIMER-based outcome prediction model in real practice and generalize the RIMER-based outcome prediction model to other diseases and departments.

Acknowledgments

This study was supported by a grant from the [National Natural Science Foundation of China \(NSFC\)](#) under Grant No.: 81301296. The study was also supported by a humanities and social sciences project funded by the Ministry of Education of China under Grant

No.: 13YJC630066. This study was also supported by a grant from the National Health and Family Planning Commission of China under Grant No. 201002014.

References

- [1] C. Mock, J.D. Lormand, J. Goosen, M. Josphipura, M. Peden, Guidelines For Essential Trauma Care, World Health Organization, Geneva, 2004.
- [2] S. Sasser, M. Varghese, A. Kellermann, J.D. Lormand, Prehospital Trauma Care Systems, World Health Organizations, Geneva, 2005.
- [3] M.-A. Purtil, K. Benedict, T. Hernandez-Boussard, S.I. Brundage, K. Kritayakirana, J.P. Sherck, A. Garland, D.A. Spain, Validation of a prehospital trauma triage tool: a 10-year perspective, *J. Trauma Acute Care Surg.* 65 (6) (2008) 1253–1257.
- [4] G. Davies, M. Wilson, Pre-hospital care and triage in the UK, *Surgery (Oxford)* 24 (6) (2006) 190–193.
- [5] S. Cox, A. Currell, L. Harriss, B. Barger, P. Cameron, K. Smith, Evaluation of the Victorian state adult pre-hospital trauma triage criteria, *Injury* 43 (5) (2012) 573–581.
- [6] C.N. Mock, M. Tiska, M. Adu-Ampofo, G. Boakye, Improvements in prehospital trauma care in an African country with no formal emergency medical services, *J. Trauma-Injury, Infect. Crit. Care* 53 (1) (2002) 90–97.
- [7] Z.-W. Zong, N. Li, T.-M. Cheng, X.-Z. Ran, Y. Shen, Y.-F. Zhao, Q.-S. Guo, L.-Y. Zhang, Current state and future perspectives of trauma care system in mainland China, *Injury* 42 (9) (2011) 874–878.
- [8] D.R. Goldhill, S.A. White, A. Sumner, Physiological values and procedures in the 24 h before ICU admission from the ward, *Anaesthesia* 54 (6) (1999) 529–534.
- [9] H. McGloin, S.K. Adam, M. Singer, Unexpected deaths and referrals to intensive care of patients on general wards. Are some cases potentially avoidable? *J. R. Coll. Phys. Lond.* 33 (3) (1999) 255–259.
- [10] G.L. Kong, X. Yin, T. Wang, R. Body, Y.-W. Chen, J. Wang, L. Cao, S. Wu, J. Gao, G. Wang, et al., Current state of trauma care in China, tools to predict death and ICU admission after arrival to hospital, *Injury* 46 (9) (2015) 1784–1789.
- [11] C.P. Subbe, M. Kruger, P. Rutherford, L. Germmel, Validation of a modified early warning score in medical admissions, *Q. J. Med.* 94 (10) (2001) 521–526.
- [12] T. Smith, D.D. Hartog, T. Moerman, P. Patka, E.M.M. Van Lieshout, N.W.L. Schep, Accuracy of an expanded early warning score for patients in general and trauma surgery wards, *Br. J. Surg.* 99 (2012) 192–197.
- [13] Royal College of Physicians, National Early Warning Score (NEWS): standardising the assessment of acute-illness severity in the NHS, Report of a working party, RCP, London, 2012.
- [14] N. Alam, I.L. Vegting, E. Houben, B. van Berkel, L. Vaughan, M.H. Kramer, P.W. Nanayakkara, Exploring the performance of the national early warning score (NEWS) in a European emergency department, *Resuscitation* 90 (2015) 111–115.
- [15] J.J. Koehler, L.J. Baer, S.A. Malafa, M.S. Meindertsmas, N.R. Navitskas, J.E. Huizenga, Prehospital Index: a scoring system for field triage of trauma victims, *Ann. Emerg. Med.* 15 (2) (1986) 178–182.
- [16] J.R. Kirkpatrick, R.L. Youmans, Trauma Index, an aide in the evaluation of injury victims, *J. Trauma* 11 (8) (1971) 711–714.
- [17] M. Ogawa, T. Sugimoto, Rating severity of the injured by ambulance attendants: field research of trauma index, *J. Trauma* 14 (11) (1974) 934–937.
- [18] G. Teasdale, B. Jennett, Assessment of coma and impaired consciousness, *Lancet* 304 (7872) (1974) 81–84.
- [19] M.C. Wang, T.A. Gennarelli, Classification systems, in: J. Jallo, C.M. Loftus (Eds.), *Neurotrauma and Critical Care of the Brain*, Thieme Medical Publishers, New York, 2009.
- [20] H.R. Champion, W.J. Sacco, W.S. Copes, D.S. Gann, T.A. Gennarelli, M.E. Flanagan, A revision of trauma score, *J. Trauma* 29 (5) (1989) 623–629.
- [21] S. Alghnam, M. Palta, A. Hamedani, M. Alkelya, P.L. Remington, M.S. Durkin, Predicting in-hospital death among patients injured in traffic crashes in Saudi Arabia, *Injury* 45 (2014) 1693–1699.
- [22] W.K. Utomo, B.J. Gabbe, P.M. Simpson, P.A. Cameron, Predictors of in-hospital mortality and 6-month functional outcomes in older adults after moderate to severe traumatic brain injury, *Injury* 40 (2009) 973–977.
- [23] B. Patil, R. Joshi, D. Toshniwal, S. Biradar, A new approach: role of data mining in prediction of survival of burn patients, *J. Med. Syst.* 35 (6) (2011) 1531–1542.
- [24] I.R. Galatzer-Levy, K.-I. Karstoft, A. Statnikov, A.Y. Shalev, Quantitative forecasting of PTSD from early trauma responses: a machine learning application, *J. Psychiatr. Res.* 59 (2014) 68–76.
- [25] S.M. DiRusso, A.A. Chahine, T. Sullivan, D. Risucci, P. Nealson, S. Cuff, J. Savino, M. Slim, Development of a model for prediction of survival in pediatric trauma patients: comparison of artificial neural networks and logistic regression, *J. Pediatr. Surg.* 37 (7) (2002) 1098–1104.
- [26] İ. Güler, Z. Gökçil, E. Güllbandilar, Evaluating of traumatic brain injuries using artificial neural networks, *Expert Syst. Appl.* 36 (7) (2009) 10424–10427.
- [27] L. Chen, T.M. McKenna, A.T. Reinsner, A. Gribok, J. Reifman, Decision tool for the early diagnosis of trauma patient hypovolemia, *J. Biomed. Inform.* 41 (3) (2008) 469–478.
- [28] M.F. Abbod, K.-Y. Cheng, X.-R. Cui, S.-J. Huang, Y.-Y. Han, J.-S. Shieh, Ensembled neural networks for brain death prediction for patients with severe head injury, *Biomed. Signal Process. Control* 6 (4) (2011) 414–421.
- [29] J.B. Yang, J. Liu, J. Wang, H.S. Sii, H.W. Wang, Belief rule-base inference methodology using the evidential reasoning approach - RIMER, *IEEE Trans. Syst. Man Cybern. Part A-Syst. Hum.* 36 (2) (2006) 266–285.
- [30] G.L. Kong, J.-B. Yang, D.-L. Xu, R. Body, X. Yin, T. Wang, B. Jiang, J. Wang, Y. Hu, Identifying severe trauma in the ED with a RIMER methodology, in: *Proceedings of the International Conference on Health Informatics and Medical Systems*, Las Vegas, USA, 2015.
- [31] J.B. Yang, M.G. Singh, An evidential reasoning approach for multiple-attribute decision making with uncertainty, *IEEE Trans. Syst., Man Cybern.* 24 (1) (1994) 1–18.
- [32] J.B. Yang, Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties, *Eur. J. Oper. Res.* 131 (1) (2001) 31–61.
- [33] Y.M. Wang, J.B. Yang, D.L. Xu, K.-S. Chin, The evidential reasoning approach for multiple attribute decision analysis using interval belief degrees, *Eur. J. Oper. Res.* 175 (1) (2006) 35–66.
- [34] C. Fu, K.-S. Chin, Robust evidential reasoning approach with unknown attribute weights, *Knowl.-Based Syst.* 59 (2014) 9–20.
- [35] J.B. Yang, J. Liu, D.L. Xu, J. Wang, H. Wang, Optimization models for training belief-rule-based systems, *IEEE Trans. Syst., Man Cybern., Part A-Syst. Hum.* 37 (4) (2007) 569–585.
- [36] G.L. Kong, D.L. Xu, R. Body, J.B. Yang, K.R.H. Mackway-Jones, S. Carley, A belief rule-based decision support system for clincial risk assessment of cardiac chest pain, *Eur. J. Oper. Res.* 219 (3) (2012) 564–573.
- [37] G.L. Kong, D.L. Xu, X.B. Liu, J.B. Yang, Applying a belief rule-base inference methodology to a guideline-based clinical decision support system, *Expert Syst.* 26 (5) (2009) 391–408.
- [38] Z.-G. Zhou, F. Liu, L.-L. Li, L.-C. Jiao, Z.-J. Zhou, J.-B. Yang, Z.-L. Wang, A cooperative belief rule based decision support system for lymph node metastasis diagnosis in gastric cancer, *Knowl.-Based Syst.* 85 (2015) 62–70.
- [39] Z.-G. Zhou, F. Liu, L.-C. Jiao, Z.-J. Zhou, J.-B. Yang, M.-G. Gong, X.-P. Zhang, A bi-level belief rule based decision support system for diagnosis of lymph node metastasis in gastric cancer, *Knowl.-Based Syst.* 54 (2013) 128–136.
- [40] Z.-J. Zhou, C.-H. Hu, X.-X. Han, H.-F. He, X.-D. Ling, B.-C. Zhang, A model for online failure prognosis subject to two failure modes based on belief rule base and semi-quantitative information, *Knowl.-Based Syst.* 70 (2014) 221–230.
- [41] L. Jin, J. Liu, Y. Xu, X. Fang, A novel rule base representation and its inference method using the evidential reasoning approach, *Knowl.-Based Syst.* 87 (2015) 80–91.
- [42] Y. Chen, Y.-W. Chen, X.-B. Xu, C.-C. Pan, J.-B. Yang, G.-K. Yang, A data-driven approximate causal inference model using the evidential reasoning rule, *Knowl.-Based Syst.* 88 (2015) 264–272.
- [43] J.B. Yang, D.L. Xu, On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty, *IEEE Trans. Syst. Man Cybern. Part A-Syst. Hum.* 32 (3) (2002) 289–304.
- [44] Y.M. Wang, J.B. Yang, D.L. Xu, Environmental impact assessment using the evidential reasoning approach, *Eur. J. Oper. Res.* 174 (3) (2006) 1885–1913.
- [45] B. Eftekhari, M.R. Zarei, M. Moezardalan, M. Zargar, E. Ketabchi, Comparing logistic models based on modified GCS motor component with other prognostic tools in prediction of mortality: results of study in 7226 trauma patients, *Injury* 36 (8) (2005) 900–904.
- [46] J.L. Guzzo, G.V. Bochicchio, L.M. Napolitano, D.L. Malone, W. Meyer, T.M. Scalea, Prediction of outcomes in trauma: anatomic or physiologic parameters? *J. Am. Coll. Surg.* 201 (6) (2005) 891–897.
- [47] V.N. Vapnik, *The Nature of Statistical Theory*, Springer-Verlag, New York, 1995.
- [48] D. Gil, J.L. Girela, J. De Juan, M.J. Gomez-Torres, M. Johnsson, Predicting seminal quality with artificial intelligence methods, *Expert Syst. Appl.* 39 (16) (2012) 12564–12573.
- [49] T.-F. Wu, C.-J. Lin, R.C. Weng, Probability estimates for multi-class classification by pairwise coupling, *J. Mach. Learn. Res.* 5 (2004) 975–1005.
- [50] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, *ACM Trans. Intell. Syst. Technol.* 2 (3) (2011) 27:21–27:27.
- [51] C.-W. Hsu, C.-C. Chang, C.-J. Lin, *A practical guide to support vector classification*, 2010.
- [52] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing, New York, 1994.
- [53] Q. Liu, X. Cui, M.F. Abbod, S.-J. Huang, Y.-Y. Han, J.-S. Shieh, Brain death prediction based on ensembled artificial neural networks in neurosurgical intensive care unit, *J. Taiwan Inst. Chem. Eng.* 42 (1) (2011) 97–107.
- [54] D. Faraggi, B. Reiser, Estimation of the area under the ROC curve, *Stat. Med.* 21 (2002) 3093–3106.
- [55] S.E. Brooks, K. Mukherjee, O.L. Gunter, O.D. Guillaumondegui, J.M. Jenkins, R.S. Miller, A.K. May, Do models incorporating comorbidities outperform those incorporating vital signs and injury pattern for predicting mortality in geriatric trauma? *J. Am. Coll. Surg.* 219 (5) (2014) 1020–1027.
- [56] O. Karam, O. Sanchez, C. Chardot, G. La Scala, Blunt abdominal trauma in children: a score to predict the absence of organ injury, *J. Pediatr.* 154 (6) (2009) 912–917.
- [57] B.S. Everitt, *Cambridge Dictionary of Statistics*, Second edn., Cambridge University Press, Cambridge, 2002.