This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Evidential Reasoning Rule-Based Decision Support System for Predicting ICU Admission and In-Hospital Death of Trauma

Guilan Kong, Dong-Ling Xu, Jian-Bo Yang, Tianbing Wang, and Baoguo Jiang

***Abstract*—We propose to employ evidential reasoning (ER) rule to construct a clinical decision support system (CDSS) to aid physicians to predict the probability of intensive care unit (ICU) admission and in-hospital death for trauma patients once they arrive at a hospital. A generalized Bayesian rule is used to mine evidence from historical data. Evidence is profiled using a format of belief distribution, where the belief degrees of different trauma outcomes are assigned with derived probabilities linked to the corresponding outcomes. Inputs to the CDSS are clinical data of a patient, and output from the system is predicted belief degree of severe trauma, including ICU admission and in-hospital death. The inner logic of the CDSS is that pieces of evidence that match the clinical data of a patient are identified from the evidence base first, and then the ER rule-based evidence aggregation mechanism is utilized to combine the matched evidences to arrive at a prediction. The reliability, weight, and interdependence of clinical evidence are taken into account. Moreover, an evidence weight training module is constructed. The ER rule-based prediction model has superior performance compared with logistic regression and artificial neural network models. An innovative and pragmatic ER rule-based CDSS for trauma outcome prediction is contributed by this article. In the era of big data, this CDSS helps predict patient outcomes based on historical data and helps physicians in emergency departments make proper trauma management decisions.**

***Index Terms*—Decision support systems, evidence aggregation, evidential reasoning (ER) rule, outcome prediction, severe trauma.**

Manuscript received October 31, 2018; accepted January 2, 2020. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 81771938, in part by the Peking University in China under Grant PKU2017LCX05 and in part by the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization under Grant U1709215. This article was recommended by Associate Editor F. Chiclana. *(Corresponding authors: Guilan Kong; Baoguo Jiang.)*

Guilan Kong is with the National Institute of Health Data Science and the Center for Data Science in Health and Medicine, Peking University, Beijing 100191, China (e-mail: guilan.kong@hsc.pku.edu.cn).

Dong-Ling Xu and Jian-Bo Yang are with the Decision and Cognitive Sciences Research Centre, University of Manchester, Manchester M15 6PB, U.K. (e-mail: ling.xu@manchester.ac.uk; jian-bo.yang@manchester.ac.uk).

Tianbing Wang and Baoguo Jiang are with the Trauma Center, Peking University People's Hospital, Beijing 100044, China (e-mail: wangtianbing@pkuph.edu.cn; jiangbaoguo@vip.sina.com).

Digital Object Identifier 10.1109/TSMC.2020.2967885

I. Introduction

THE OCCURRENCE rate of medical errors and adverse events in hospitals is high, and the accompanying patient safety problem cannot be ignored. Research [1] shows that nearly half of those medical errors could be prevented if appropriate measures and techniques are employed. Appropriate use of information technologies, particularly clinical decision support systems (CDSSs), may aid clinicians to make better clinical decisions and thus reduce the rate of medical errors [2]. From the early generation of knowledge-based CDSSs such as MYCIN [3] to the evolution of clinical guideline-driven decision support tool such as GLIF [4] and clinical data-based decision support tools developed using various machine learning techniques [5], [6], CDSSs have a history of nearly 50 years. A conventional CDSS contains a user interface, a knowledge component, and a reasoning or inference component. The domain knowledge stored in the knowledge base may be gained from medical textbooks, expert experiences, or clinical datasets. The inference component of these systems contains algorithms for matching clinical rules in the knowledge base to a patient's clinical data [7].

In recent years, the growth of information storage technology and the wide use of electronic medical records (EMRs) systems have helped generate large amounts of medical data. A crucial need exists for CDSS researchers to develop decision models that can mine knowledge from a large volume of historical clinical data and optimize clinical decision making, thus improving patient safety and healthcare quality [8].

Trauma causes high mortality and disability. Around 90% of the medical burden of trauma is in underdeveloped countries [9]. Studies [10]–[12] show that whether a patient may die in hospital or need to be admitted to the intensive care unit (ICU) can be predicted if clinical signs of deterioration are identified just after arrival at emergency departments (EDs). In China, no standard guidelines exist to assist physicians to assess trauma severity in EDs. The Chinese prehospital emergency system deals with trauma patients using a proximity principle, which requires a trauma patient to be sent to the nearest hospital. Therefore, some patients with high severity may be sent to the nearest hospitals having no abilities to handle severe trauma and those patients have to be retransferred. The working environments in EDs are complex, and many high-tech medical devices are utilized in this area, thus making optimal clinical decisions is difficult for ED physicians. In this case, trauma outcomes are relatively

2168-2216 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

poor in China, and trauma has become a main cause of death in young people [13]. Therefore, for optimal management of trauma patients and improved trauma outcome, ED physicians have a strong need to identify patients with severe trauma to provide appropriate trauma management.

In the literature, clinical signs, including systolic blood pressure (SBP), pulse rate, respiration rate, consciousness level, and body temperature, have been considered as contributing factors to predict ICU admission and in-hospital death [12], [14]. For trauma patients, other clinical variables, including age, comorbidities, mechanism of injury, and location of injury, are also useful predictors of in-hospital death and ICU admission [15], [16]. Various scoring systems have been developed to assess trauma severity. The main scoring tools studied in the literature include injury severity score (ISS) [17], trauma and ISS (TRISS) [11], early warning score (EWS) [12], prehospital index (PHI) [18], and Glasgow coma score (GCS) [19]. Those scoring systems would give each patient a severity score on the basis of clinical signs, and no one tool has been shown to have superior performance than the others [20]. Specifically, PHI, GCS, and EWS are mainly based on physician observations or measurements, and ISS is based on detailed diagnoses of injuries to the different parts of the body. TRISS can provide the survival probability for a trauma patient and is based on ISS, vital signs, injury pattern, and age. However, the use of ISS and TRISS in emergency response to predict trauma patient outcomes is inapplicable because detailed diagnoses of trauma cannot be made by ED physicians.

Driven by the actual need of trauma outcome prediction in clinical practice, how to combine vital signs and other clinical variables to predict the belief degree or probability of ICU admission and in-hospital death has attracted considerable research attention [11], [12], [15], [16].

From a methodology perspective, logistic regression (LR) [10], [11], [14], [16], and artificial neural network (ANN) [21]–[23] are frequently employed by researchers to construct prediction models. In this article, we propose to employ the evidential reasoning (ER) rule [24] to develop a CDSS to aid physicians in the ED to make appropriate clinical decisions by predicting the belief degree of ICU admission and in-hospital death for trauma patients. The original ER rule is for inference with independent evidence profiled in the belief distribution (BD) format, with the reliability and weight taken into account. The ER rule is developed from the ER algorithm [25] which has been employed to construct expert systems for disease diagnosis [26]–[28]. The ER rule has been employed to do online safety assessment of complex systems [29]. In clinical practice, two pieces of evidence may not be fully independent in leading to one clinical outcome, and the interdependence between two pieces of evidence may affect the joint probability distribution on different outcomes. Therefore, in addition to the weight and reliability associated with each piece of evidence, we take into account the interdependence between evidence in our research. If the belief degrees in the ER rule inference paradigm are generated by normalizing likelihoods in the Bayesian paradigm, Bayesian inference would be the ER rule [30].

Compared with Bayesian inference, if the studied dataset is complete and has no ambiguity, the ER rule has the following advantages. First, the ER rule takes the evidence reliability into account in the inference process, while Bayesian inference assumes that any observation is fully reliable. Second, the ER rule is a likelihood (or prior-free) inference process, while Bayesian inference depends on prior distribution as prerequisite [31], although the former can also take prior distribution as a piece of independent evidence if available. Third, the degree of interdependence between two pieces of evidence can be incorporated into the ER rule inference paradigm [32], while Bayesian inference assumes that all predictor variables are independent [33]. Although Bayesian net takes the interdependence between factors into account, the prior knowledge about which factor is dependent on which and which factors are conditionally independent are also needed to develop a Bayesian net [34]. While in the ER rule, no prior knowledge about the dependence relationship between evidence is required, the dependence index between two pieces of evidence can be calculated from historical data. Therefore, we choose the ER rule in this article for trauma outcome prediction.

The remainder of this article is structured as follows. In Section II, the trauma dataset is described first, followed by the ER rule model for trauma outcome prediction and the functional structure design of the ER rule-based CDSS. Then, a prediction performance comparison study is presented, where the ER rule-based model is compared with two other well-established models, namely, LR and ANN. Section III presents the evidence mined from the training dataset, the learning results of evidence weights, and the results of the comparison study. Section IV summarizes this article and provides discussions about the findings, limitations, and possible research directions in the future.

II. Materials and Methods

A. Data

The trauma dataset recorded at ED of Kailuan Hospital, China, was used for prediction model development. The dataset contains 1299 trauma patients. Demographic variables, date of ED arrival, comorbidities, body region injured, cause of injury, type of injury, blood pressure, respiration rate, body temperature, pulse rate, and consciousness level are recorded in the dataset. To identify those severe trauma patients who may be admitted to ICU or die in hospital, patient information about condition at discharge, time of ICU admission and discharge were retrieved from the EMR data.

For performance comparison, in addition to the ER rule model, two other modeling methods LR and ANN were investigated using the same dataset. As LR and ANN models cannot deal with missing data, we removed those patient data with incomplete records and obtained 1189 cases for analysis. As a result, 1189 cases were used for the derivation and test of the three models.

Two patient outcomes exist in this article. One outcome is severe trauma, which is defined as including ICU admission and in-hospital death [35]. The other outcome is nonsevere

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

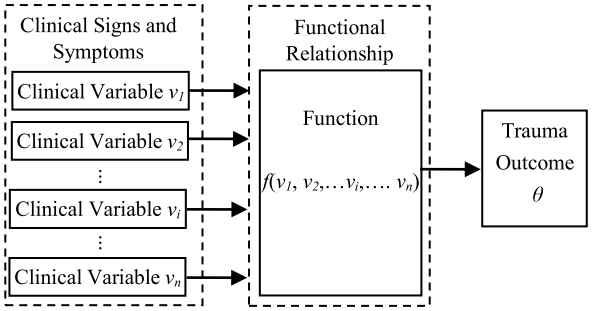

Fig. 1. Functional relationship between clinical variables and trauma outcome.

trauma, in which patients recover from their injuries without needing intensive care. We choose severe trauma as the primary outcome in this article, and the focus of this article is to identify severe trauma in ED.

B. ER Rule-Based Trauma Outcome Prediction Model

1) Problem Description: A variety of factors may affect the outcome of trauma patients. With the utilization of the EMR system in hospitals, clinicians can collect as many clinical variables as they can in the ED. Assume n clinical variables can be collected in the ED, and the outcome of a trauma patent is denoted using θ. "What is the functional relationship between these n clinical variables and the outcome θ?" and "How can n clinical variables be employed to predict patient outcome θ?" are practical clinical problems that ED physicians have to face daily. The functional relationship between clinical variables and outcome is depicted in Fig. 1.

2) Structure of the ER Rule-Based Outcome Prediction Model: The original ER rule is developed from the evidence theory [36] and the ER approach [25] and is established to conjunctively aggregate independent pieces of evidence, with the reliability and weight taken into account. In the ER rule inference paradigm, evidence is profiled using a format of weighted BD with reliability (WBDR), which is an extension of BD as introduced in the evidence theory. In the WBDR format, evidence is still profiled using BD, while the reliability and weight of evidence are considered. The use of WBDR to profile evidence is described as follows.

Assume that there are N propositions that are mutually exclusive, denoted by $\{\theta_n, n = 1, \ldots, N\}$ with $\theta_i \cap \theta_j = \varphi$ for any $i, j \in \{1, \ldots, N\}$ and $i \neq j$, and a discernment frame Θ is composed of θ_n. The power set $P(\Theta)$ contains 2^N subsets of Θ. Use $e_i (i = 1, \ldots, L)$ to denote a piece of evidence that supports one proposition θ in Θ. Profiled in the format of BD, evidence e_i is defined as follows:

$$e_i = \left\{ \left(\theta, p_{\theta,i}\right) \quad \forall \theta \subseteq \Theta, \sum_{\theta \subseteq \Theta} p_{\theta,i} = 1 \right\} \tag{1}$$

where $p_{\theta,i}$ represents the belief degree which measures how strong evidence e_i supports proposition θ. Furthermore, when evidence e_i is profiled in the WBDR format, the weight of e_i, denoted using ω_i with $0 \leq \omega_i \leq 1$, and the reliability of e_i, denoted by r_i with $0 \leq r_i \leq 1$ are introduced to describe evidence e_i in the ER rule [24].

The concept of dependence index [31], [32] was proposed to measure the degree of interdependence between any two pieces of evidence that point to the same proposition. Suppose e_i and e_j are two pieces of evidence that support proposition θ. $p(\theta|e_i, e_j)$, $p(\theta|e_i)$, and $p(\theta|e_j)$ are used to represent the belief degree in support of θ when both pieces of evidence (e_i and e_j), single evidence e_i and single evidence e_j are taken into consideration, respectively. Use $\alpha_{e_i,e_j}(\theta)$ to represent the dependence index between evidences e_i and e_j in support of proposition θ. As evidences e_i and e_j support different propositions with different belief degrees, the dependence index α_{e_i,e_j} between evidences e_i and e_j should be different over different propositions. That is to say, $\alpha_{e_i,e_j}(\theta)$ is related to proposition θ and the belief degree supported by evidences e_i and e_j, respectively. We consider $p(\theta|e_i, e_j) = \alpha_{e_i,e_j}(\theta)p(\theta|e_i)p(\theta|e_j)$ [31]. Thus, $\alpha_{e_i,e_j}(\theta)$ can be calculated by

$$\alpha_{e_i,e_j}(\theta) = p\left(\theta \middle| e_i, e_j\right) / \left(p(\theta|e_i) p\left(\theta \middle| e_j\right)\right) \tag{2}$$

where the numerator $p(\theta|e_i, e_j)$ represents the joint belief degree distributed on proposition θ after considering both evidences e_i and e_j, and the denominator is the product of degree of belief in θ provided by e_i and e_j, respectively. We set $\alpha_{e_i,e_j}(\theta) = 1$, if $p(\theta|e_ie_j) = 0$, $p(\theta|e_i) = 0$, or $p(\theta|e_j) = 0$; and $\alpha_{e_i,e_j}(\theta) = 0$, if $p(\theta|e_ie_j) = 0$, $p(\theta|e_i) \neq 0$, and $p(\theta|e_j) \neq 0$. If $\alpha_{e_i,e_j}(\theta)$ is larger than 1, then evidence e_j is considered to have a positive effect on evidence e_i in support of proposition θ. If smaller than 1, then evidence e_j is considered to have a negative effect on evidence e_i in support of θ.

Assume that there are L pieces of interdependent evidence, each of which is profiled using the WBDR format, and the dependence index, which measures the degree of interdependence between any two pieces of evidence, is calculated by using (2). $e_{(i)}$ is used to denote the combined evidence after aggregating the first i pieces of evidence. We use $c_{rw,i}$ to denote a regulative coefficient that can adjust ω_i according to reliability r_i with $c_{rw,i} = 1/(1 + \omega_i - r_i)$, and $\tilde{m}_{\theta,e_{(i)}}$ to denote the probability mass that θ is backed up by $e_{(i)}$. Aggregation of the first i pieces of interdependent evidence with weight ω_i and reliability r_i is given by the following orthogonal sum operation:

$$\tilde{m}_{\theta,e_{(i)}} = \left[\tilde{m}_1 \oplus \cdots \oplus \tilde{m}_i\right](\theta) = \begin{cases} 0 & \theta = \varphi \\ \frac{\hat{m}_{\theta,e_{(i)}}}{\sum_{D \subseteq \Theta} \hat{m}_{D,e_{(i)}} + \hat{m}_{P(\Theta),e_{(i)}}} & \theta \neq \varphi \end{cases} \tag{3}$$

$$\begin{aligned} \hat{m}_{\theta,e_{(i)}} &= \left[(1 - \tilde{\omega}_i)\tilde{m}_{\theta,e_{(i-1)}} + \tilde{m}_{P(\Theta),e_{(i-1)}}\tilde{m}_{\theta,i}\right] \\ &\quad + \sum_{B \bigcap C = \theta} \alpha_{e_{(i-1)}(B),e_i(C)}(\theta)\tilde{m}_{B,e_{(i-1)}}\tilde{m}_{C,i} (i = 2, \ldots, L) \quad \forall \theta \subseteq \Theta \end{aligned} \tag{4}$$

$$\hat{m}_{P(\Theta),e_{(i)}} = (1 - \tilde{\omega}_i)\tilde{m}_{P(\Theta),e_{(i-1)}} \tag{5}$$

where $\tilde{\omega}_i$ is the adjusted weight of e_i with $\tilde{\omega}_i = c_{rw,i}\omega_i$, $\tilde{m}_{\theta,i}$ represents the belief degree provided by e_i to support θ with the adjusted weight $\tilde{\omega}_i$, with $\tilde{m}_{\theta,i} = \tilde{\omega}_i p_{\theta,i}$ for any $\theta \subseteq \Theta$ and $\tilde{m}_{P(\Theta),i} = 1 - \tilde{\omega}_i = c_{rw,i}(1 - r_i) \cdot \tilde{m}_{\theta,e_{(1)}} = \tilde{m}_{\theta,1}$, $\tilde{m}_{\theta,1} = \tilde{\omega}_1 p_{\theta,1} = c_{rw,1}\omega_1 p_{\theta,1}$, $\tilde{m}_{P(\Theta),e_{(1)}} = \tilde{m}_{P(\Theta),1}$, $\tilde{m}_{P(\Theta),1} = 1 - \tilde{\omega}_1 = 1 - c_{rw,1}\ \omega_1 = 1 - \omega_1/(1 + \omega_1 - r_1) = c_{rw,1}(1 - r_1)$, and $\sum_{\theta \subseteq \Theta} \tilde{m}_{\theta,i} + \tilde{m}_{P(\Theta),i} = 1$ are recursively

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

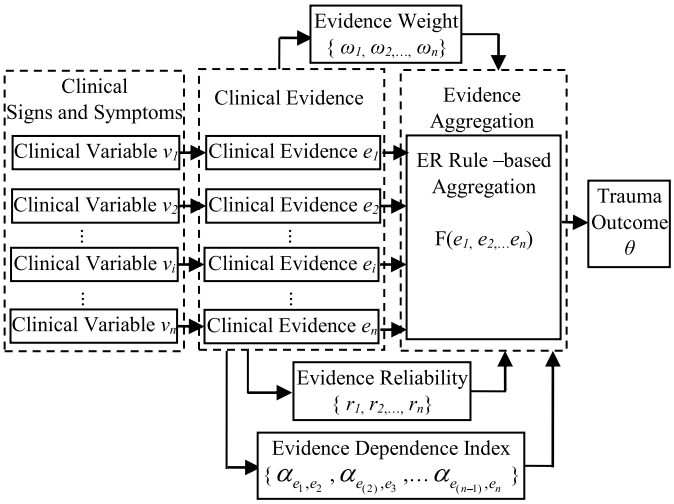

Fig. 2. ER rule-based trauma outcome prediction model.

set for $i = 1, \ldots, L$. $\alpha_{e_{(i-1)}(B),e_i(C)}(\theta)$ measures the degree of interdependence between evidence e_i and its preceding combined evidence $e_{(i-1)}$, which is generated by combining the first $(i-1)$ pieces of evidence in support of proposition θ with $B \cap C = \theta$.

The final aggregated belief degree after combining all the L pieces of interdependent evidence, with adjusted weight $\tilde{\omega}_i$ and dependence index $\alpha_{e_{(i-1)},e_i}(\theta)$, which point to proposition θ, is given by the following equation:

$$p_\theta = p_{\theta,e_{(L)}} = \begin{cases} 0 & \theta = \varphi \\ \frac{\hat{m}_{\theta,e_{(L)}}}{\sum_{D\subseteq\Theta} \hat{m}_{D,e_{(L)}}} \theta \subseteq \Theta, & \theta \neq \varphi \end{cases} \tag{6}$$

where $\hat{m}_{\theta,e_{(L)}}$ is generated by recursively applying (3)–(5) to combine i pieces of evidence for $i = 2, \ldots, L$, $0 \leq p_\theta \leq 1$ $\forall\theta \subseteq \Theta$ and $\sum_{\theta\subseteq\Theta} p_\theta = 1$.

The ER rule would turn to Dempster's rule [36] if all evidence is fully reliable and independent. Also, the ER rule would be the ER algorithm [25] if the evidence is independent, and the evidence weight and reliability are equivalent and normalized as $\sum_{i=1}^{L} \omega_i = 1$.

Fig. 2 shows the structure of the ER rule-based model for predicting trauma outcome when the ER rule is employed to model the functional relationship between clinical variables and clinical outcome.

In the model, clinical variables v_i $(i = 1, \ldots, n)$ need to be transformed to corresponding clinical evidence e_i $(i = 1, \ldots, n)$ before the ER rule-based aggregation scheme is employed to fuse evidence. Each piece of evidence has weight $\omega_i (i = 1, \ldots, n)$ and reliability $r_i (i = 1, \ldots, n)$. Each piece is correlated with its previous $(i-1)$ combined evidence and the dependence is measured using dependence index $\alpha_{e_{(i-1)},e_i} (i = 2, \ldots, n)$. For example, "if a trauma patient at the ED ages 70 (belonging to old age group), then the patient may be admitted to ICU with a probability of 70%" is a piece of evidence that corresponds to "age" variable, and it can be denoted as {"old," (admitted to ICU, 70%)}. In the ER rule-based evidence aggregation scheme, evidence weight ω_i, reliability r_i, and the dependence index $\alpha_{e_{(i-1)},e_i}$ between each piece of evidence and its preceding combined evidence are taken into account in evidence aggregation.

3) Implementation of the ER Rule-Based Trauma Outcome Prediction Model: We develop and test an ER rule-based trauma outcome prediction model through the following sequential steps.

First, we consider that each clinical variable provides a piece of evidence for predicting the outcome. The evidence that a variable can be transformed to depend on the category that the variable is classified into. A generalized Bayesian inference method [30] is employed to mine pieces of evidence from historical patient data. Evidence is profiled using the WBDR format, and the degree of belief in one specific outcome provided by each piece of evidence is calculated using the probability linked to the outcome. As such, an evidence base for predicting ICU admission and in-hospital death is constructed.

Second, we employ the ER rule as a likelihood inference scheme to analyze and combine multiple pieces of matched evidence for predicting the belief degree of ICU admission and in-hospital death. In this article, as all patient data come from the same source, we consider that the reliability of each piece of evidence equals its importance. The weight of a piece of evidence can be determined subjectively or learned from historical data. In this article, an evidence weight training module is constructed to get an optimal set of evidence weight. The degree of interdependence between any piece of evidence e_i and its preceding combined evidence, is calculated from historical data using (2).

Third, based on the constructed evidence base, the ER rule-based inference mechanism, and weight training module, an ER rule-based outcome prediction model and CDSS are constructed. The inputs of the CDSS are the clinical data of a patient, and the output of the CDSS is the belief degree of being admitted to ICU or having in-hospital death.

Finally, we conduct a performance comparison study, where the ER rule-based prediction model is compared with two other well-established prediction models: LR and ANN.

The ER rule-based prediction CDSS has the advantage of making full use of historical data by constructing an evidence base. The evidence weight and reliability together with the dependence index between each piece of evidence and its preceding combined evidence are taken into account in evidence aggregation. The ER rule-based prediction is a transparent process, as its reasoning mechanism can tell which pieces of evidence are combined.

a) Mining evidence from historical data: Use p_{ij} to denote the supporting degree that clinical sign or symptom s_j (corresponding to evidence e_j) provides for proposition θ_i for $i = 1, \ldots, N; j = 1, \ldots, L$ in the ER paradigm, where N denotes the total number of all propositions, and L denotes the total number of possible clinical signs or symptoms. Use c_{ij} to denote the likelihood that the jth sign or symptom s_j may happen if the ith proposition θ_i is correct and prior distribution is known in the Bayesian paradigm. Bayesian inference can be generalized to ER if the belief degree p_{ij} is generated by normalizing likelihoods c_{ij} [30]. The belief degree p_{ij} from

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

likelihood c_{ij} is calculated as follows:

$$p_{ij} = c_{ij} \Big/ \sum_{n=1}^{N} c_{nj} \text{ for } i = 1, \ldots, N \text{ and } j = 1, \ldots, L \quad (7)$$

where c_{ij} represents the likelihood of evidence e_j (transformed from the jth clinical sign or symptom) occurs if the proposition, which here means one clinical outcome θ_i happens.

This finding means that from the historical dataset, we can mine clinical evidence profiled in the BD format by normalizing likelihoods via (7). A piece of evidence e_j means the fact that clinical outcome θ_i would occur with the probability p_{ij} when the clinical sign or symptom s_j happens. We use this generalized Bayesian method to mine evidence from the collected dataset as follows.

i) Categorization of clinical variables in the dataset: To mine clinical evidence from historical data, the first step is to clarify the number of categories that a clinical variable can be classified into, because a different piece of clinical evidence will be produced if the clinical variable falls into a different category. There are ten clinical variables, namely, *age* (denoted by v_1), *gender* (v_2), *comorbidities* (v_3), *body temperature* (v_4), *SBP* (v_5), *pulse rate* (v_6), *respiration rate* (v_7), *consciousness level* (v_8), *injury mechanism* (v_9), and *injury region* (v_{10}), that can be collected in ED once a patient arrives at a hospital. In this article, each clinical variable can be transformed to a piece of clinical evidence according to its value. The details of transforming clinical variables to clinical evidence are provided as follows.

The number of evidence that a variable can provide depends on the number of categories that the variable can be classified into. For example, for variable *age* (v_1), no universal standard for age grouping exists, while 18 and 64 are typically used as threshold values to categorize patients into "adult or child" and "adult or old" groups, respectively. Similarly, in this article, *age* (v_1) is divided into three categories, including "old" (older than 64), "adult" (18–64), and "child" (younger than 18), and the belief degree distributed on a particular outcome provided by each age group is calculated by normalizing likelihoods of each age group linked to the outcome.

Generally, variable *gender* (v_2) is categorized into two groups: 1) male and 2) female. The categorization of *comorbidities* (v_3) depends on those pre-existing diseases that appeared in historical patient data. In this article, high-frequency comorbidities include "diabetes," "hypertension," and "heart diseases," and some patients may have coexisting diseases. Thus, the categories of *comorbidities* (v_3) include "multiple diseases," diabetes, hypertension, heart diseases, "other disease," and "no diseases," where multiple diseases represents a coexistence of all or any two of diabetes, hypertension and heart diseases; other diseases represents some other diseases except for the above three.

Categorization of the five vital sign variables, including body temperature (v_4), SBP (v_5), pulse rate (v_6), respiration rate (v_7), and consciousness level (v_8), is based on the scoring criteria used in the EWS [12], where a score between 0 and 3 is given to a patient according to physician observation or instrumental measurement of each vital sign. The categorization of the five vital signs in this article does not completely follow the EWS criteria, because the number of cases in some category is too small and we need to merge those adjacent categories with small number of cases. The category merge would not affect the reliability of evidence provided by the merged category, because the belief degrees in support of possible propositions are calculated from real-world data using (7), while the consistency or inconsistency among case outcomes in the merged category would be reflected in the distribution of belief degrees on different propositions. Moreover, in medical statistics, merging adjacent categories with small number of cases is a recommended way to deal with category variables with sparse data [37].

We categorized the variables *injury mechanism* (v_9) and *injury region* (v_{10}) according to the categorization in related trauma study [38]. The final categorization of the two variables is dependent on the actual injuries that occurred to the involved patients. Similar to [38], *injury mechanism* (v_9) is categorized into "blunt injury," "penetrating injury," "burn injury," "destructive injury," "crush injury," "fall," "traffic accident," and "others." We merged the categories of "burn injury," destructive injury, and "others" into one nonspecific category "others" because the number of cases in any of these three categories is small. By reference to anatomic location used in the ISS [17], *location of injury* (v_{10}) is categorized into "multiple injuries," "head, face, or neck," "chest," "abdomen," "back and spine," and "extremities" in this article. With a slight difference from location categorization in the ISS, we merged "injuries to face" and "injuries to head or neck" into one category because the number of patients who have only face injuries is small and injuries to face is always together with injuries to head or neck in the collected dataset. Also, we merged the categories of chest, abdomen, and back and spine into the "main body" category because of the small number of cases in any one of these categories.

ii) Evidence mining: Based on the categorization of clinical variables as discussed above, the evidence mining process is illustrated as follows. Take variable *age* (v_1) for example. *age* is divided into three categories: 1) "old" (older than 64); 2) "adult" (18–64); and 3) "child" (younger than 18). The degree of belief in a particular outcome provided by each age group is calculated by normalizing likelihoods of each group linked to the outcome. If we use the whole dataset without missing data (1189 cases) to mine evidence, the characteristics of *age* (v_1) in the historical dataset are shown in Table I, where θ_1 represents the occurrence of the primary outcome which contains ICU admission and in-hospital death, and θ_2 represents the nonoccurrence of the primary outcome.

Use e_0 to denote prior distribution, e_1 to represent a single evidence when a patient is in the child category, e_2 to represent evidence when the patient is in the adult category, and e_3 to represent evidence when the patient is in the old category. A piece of evidence tells how strong (measured using belief degree) it supports the primary outcome. According to the data characteristics as shown in Table I, the *prior* distribution of the primary outcome in the studied population is $p_{10} = p(\theta_1|e_0) = 28/1189 = 0.0235$ and $p_{20} = p(\theta_2|e_0) =$

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
CHARACTERISTICS OF *Age* VARIABLE IN THE STUDIED DATASET

Data		Age (v_1)			Total outcomes
		Child	Adult	Old	
Patient outcomes	In-hospital death/ICU admission (θ_1)	0	25	3	28
	Neither one (θ_2)	34	1020	107	1161
	Total patients	34	1045	110	1189

TABLE II
LIKELIHOODS OF *Age* VARIABLE (v_1)

Patient outcome	Prior Probability (e_0)	Likelihoods		
		Child (e_1)	Adult (e_2)	Old (e_3)
In-hospital death /ICU admission (θ_1)	0.0235	$c_{11}(0)$	$c_{12}(0.8929)$	$c_{13}(0.1071)$
Neither one (θ_2)	0.9765	$c_{21}(0.0293)$	$c_{22}(0.8786)$	$c_{23}(0.0922)$

TABLE III
EVIDENCE PROVIDED BY *Age* VARIABLE (v_1) IN THE BD FORMAT

Child (e_1)	Adult (e_2)	Old (e_3)
$\{(\theta_1,0), (\theta_2,1)\}$	$\{(\theta_1,0.5040), (\theta_2,0.4960)\}$	$\{(\theta_1,0.5374), (\theta_2,0.4626)\}$

$1161/1189 = 0.9765$. The variable v_1 can be transformed to three different pieces of evidence: e_1, e_2, and e_3, when its clinical value matches child, adult, or old group, respectively. The likelihoods c_{11}, c_{21}, c_{12}, c_{22}, c_{13}, and c_{23} of the evidences e_1 (child), e_2 (adult), and e_3 (old), can be calculated from the collected dataset as follows: $c_{11} = p(e_1|\theta_1, e_0) = 0/28 = 0$, $c_{21} = p(e_1|\theta_2, e_0) = 34/1161 = 0.0293$, $c_{12} = p(e_2|\theta_1, e_0) = 25/28 = 0.8929$, $c_{22} = p(e_2|\theta_2, e_0) = 1020/1161 = 0.8786$, $c_{13} = p(e_3|\theta_1, e_0) = 3/28 = 0.1071$, and $c_{23} = p(e_3|\theta_2, e_0) = 107/1161 = 0.0922$. The likelihoods of *age* variable (v_1) linked to different outcomes are shown in Table II.

Based on the likelihoods calculated from the collected dataset as shown in Table II, degrees of belief in support of different outcomes provided by the three pieces of evidence, which are transformed from v_1, are shown in Table III.

Take evidence e_2 for example. The degree of belief in the outcome θ_1 is calculated from likelihoods as $0.8929/(0.8929+0.8786) = 0.5054$. Similar calculations are made for evidences e_1 and e_3. Similar to *age* variable (v_1), the other nine clinical variables are categorized according to their data characteristics, and pieces of evidence that correspond to those variables can be mined from the training dataset.

One thing cannot be neglected here is the imbalance of samples as shown in Table I, which tells us that the proportion of patients having sever trauma is small compared with the patients with nonsevere trauma. Generally, the data imbalance problem can be addressed from three different perspectives: 1) data itself; 2) learning algorithms; and 3) model performance measures [39], [40]. From data aspect, sampling techniques including oversampling and undersampling are usually employed to balance the dataset for model training. From algorithms aspect, various approaches such as cost-sensitive learning are utilized in data mining to eliminate the imbalance caused by data. From model performance measures aspect, to eliminate the effect of disease prevalence, the receiver operating characteristic (ROC) curve instead of accuracy is employed to measure model prediction performance. In this article, we tried to eliminate the effect caused by data imbalance on the ER rule-based model from both the algorithms and performance measures aspects. In performance comparison, we employed the ROC curve to measure prediction performance. In evidence mining, we normalized the likelihoods of each variable category linked to different outcomes to generate the belief degrees provided by the evidence (transformed from a clinical variable category). Thus, the degrees supported by each piece of evidence for different outcomes can be balanced. The logic of the normalization is that cases with different outcomes have the same proportion in the dataset, and thus the effect of data imbalance on prior distribution can be removed.

b) Evidence dependence index calculation: In this article, to reduce computation and take full advantages of the recursive inference algorithm as shown in (3)–(5), the dependence index between any evidence e_i and its preceding joint evidence $e_{(i-1)}$ instead of the dependence index between any two pieces of evidences e_i and e_j is calculated from historical data for later evidence weight training and ER-rule based inference.

Let $\alpha_{e_{(i-1)},e_i}(\theta)$ represent the dependence index between evidence e_i and its preceding joint evidence $e_{(i-1)}$ in support of proposition θ. $\alpha_{e_{(i-1)},e_i}(\theta)$ is calculated from historical data using (2). Similar to $p(\theta|e_i)$, the degree of support for proposition θ provided by evidence e_i, the supporting degree provided by its preceding joint evidence $e_{(i-1)}$ for θ, $p(\theta|e_{(i-1)})$, is calculated from the historical dataset using the same method as we introduced before for evidence mining. In detail, the likelihoods of each category of the joint variable [considering all the preceding $(i-1)$ variables] linked to possible outcomes are calculated; then, the degrees of belief in different outcomes provided by each piece of joint evidence are calculated by normalizing the likelihoods. Take variables *age* (v_1), *gender* (v_2), and *comorbidities* (v_3) for example. Based on the evidence base constructed in the above evidence mining step, *age* (v_1) can be transformed to three different pieces of evidence e_{1_1} (child), e_{1_2} (adult), and e_{1_3} (old), *gender* (v_2) to two different pieces of evidence e_{2_1} (male) and e_{2_2} (female), and *comorbidities* (v_3) to six pieces of evidence e_{3_1} (multiple diseases), e_{3_2} (diabetes), e_{3_3} (hypertension), e_{3_4} (heart diseases), e_{3_5} (other diseases), and e_{3_6} (no disease). Assume that through evidence mining from historical data, among other obtained evidence, three pieces of evidence e_{1_2}, e_{2_1}, and e_{3_3} are $\{(\theta_1|e_{1_2}, 0.5384), (\theta_2|e_{1_2}, 0.4616)\}$, $\{(\theta_1|e_{2_1}, 0.4828), (\theta_2|e_{2_1}, 0.5172)\}$, and $\{(\theta_1|e_{3_3}, 0.6999), (\theta_2|e_{3_3}, 0.3001)\}$, respectively, where θ_1 and θ_2 represent the severe and nonsevere trauma, respectively. If the basic probability distribution over outcomes θ_1 and θ_2 after considering both evidences e_{1_2} and e_{2_1} is $\{(\theta_1|e_{1_2}, e_{2_1}, 0.6999), (\theta_2|e_{1_2}, e_{2_1}, 0.3001)\}$, and the joint distribution after considering evidences e_{1_2}, e_{2_1}, and e_{3_3} is $\{(\theta_1|e_{1_2}, e_{2_1}, e_{3_3}, 0.6999), (\theta_2|e_{1_2}, e_{2_1}, e_{3_3}, 0.3001)\}$, then the dependence index between e_{2_1} and e_{1_2} in supporting θ_1 is calculated as: $\alpha_{e_{1_2}, e_{2_1}}(\theta_1) = p(\theta_1|e_{1_2}e_{2_1})/(p(\theta_1|e_{1_2})p(\theta_1|e_{2_1})) = 0.6999/(0.5384 * 0.4828) = 2.6925$ according to (2), and

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

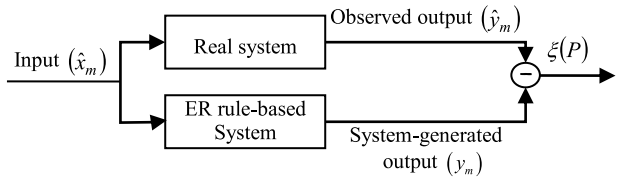

Fig. 3. Weight training process.

the dependence index between e_{3_3} and $e_{(2)_1}$ (considering both evidences e_{1_2} and e_{2_1}) for θ_1 is calculated as: $\alpha_{e_{3_3},e_{(2)_1}}(\theta_1) = p(\theta_1|e_{1_2}, e_{2_1}, e_{3_3})/(p(\theta_1|e_{1_2}e_{2_1})p(\theta_1|e_{3_3})) = 0.6999/(0.6999 * 0.6999) = 1.4288$.

In this article, all dependence indexes between any piece of evidence and its preceding joint evidence were calculated using the same method as described above.

c) Evidence weight training: Through evidence matching, clinical signs and symptoms of a new trauma patient can be transformed to pieces of clinical evidence. To predict the clinical outcome of one patient, matched pieces of evidence need to be combined to predict the probability of the primary outcome. Prior to this step, the reliability and weight of each piece of evidence need to be determined. In this article, as all patient data come from the same data source, we consider that the evidence reliability and weight share the same definition and have the same value. However, we cannot tell the exact weight of each piece of evidence. A training model is constructed to learn weight ω_i of evidence e_i from historical data, and construction of the training model is briefly introduced as follows.

The purpose of the training model is to obtain an optimal set of evidence weights that can minimize the discrepancy between the actual outcomes and the system-generated outcomes in the training dataset. In this article, the ER rule-based CDSS generated output is an aggregated degree of belief in the primary outcome, and it is numerical. The observed outcome of each patient is a binary one, that is, having the primary outcome or not, and quantitative values 1 and 0 are used to represent the observed outcomes, namely, severe trauma and nonsevere trauma, respectively.

Given that both the system-generated output and the observed output are numerical, the training process can be depicted as in Fig. 3, where $\hat{x}_m(m = 1, \ldots, M)$ is the clinical input of the mth patient, and M represents the total number of patients in the training dataset. $y_m(m = 1, \ldots, M)$ is the combined degree of belief in the primary outcome, which is the output of the ER rule-based prediction CDSS with the input $\hat{x}_m$. $\hat{y}_m(m = 1, \ldots, M)$ is the observed output with a binary 0 or 1, which is the actual outcome of the mth patient with the input $\hat{x}_m$. $\xi(P)$ denotes the discrepancy between the system-generated output and the actual outcome.

This article investigates the situation in which the weights of different pieces of evidence provided by the same clinical variable may be different. For example, *gender* (v_2) is the second variable in evidence aggregation for outcome prediction, and the weight of "male" evidence may be different from that of "female" evidence. Therefore, the weight associated with v_2 may have two different values according to its value. After discussing with collaborating physicians, we set reliability equal to weight for each piece of evidence, and the initial weights were set as 0.90 for all pieces of evidence in the training model. Function *fmincon* in the MATLAB Optimization Toolbox was used to develop the training model to fine-tune weight $\omega_i(i = 1, \ldots, L)$. After model training, an optimal weight set $\omega_i(i = 1, \ldots, L)$ associated with evidence can be obtained. In other situations, if datasets from different sources are used for evidence mining, the reliability of evidence extracted from different data sources would be different, and then both the evidence weight and reliability would need training by the same training process and logic as shown in Fig. 3. In that case, the trained parameters would include both evidence weights and reliabilities.

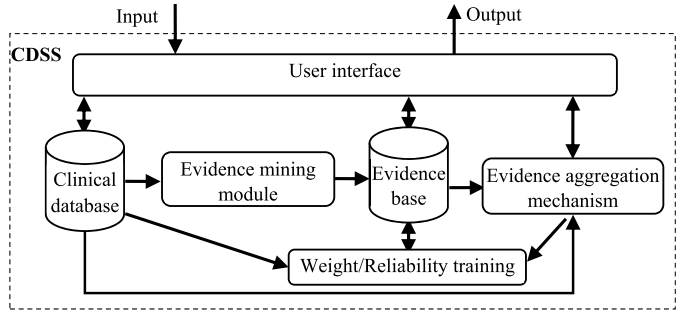

Fig. 4. System functional structure.

C. System Functional Structure Design

As shown in Fig. 4, the core system components of the prediction CDSS are designed to include a clinical database, an evidence mining module, an evidence base, an evidence weight and reliability training module, an evidence combination mechanism, and a user interface.

The clinical database in the CDSS is used to store clinical data of patients in a structural manner. The evidence mining module is designed to mine evidence from the historical dataset, which is stored in the clinical database. The mining module is implemented using the generalized Bayesian inference method as discussed in Section II-B3a. The evidence mining module does not need to run at all times and can be scheduled to run monthly or quarterly. The evidence base stores the mined evidence together with corresponding dependence indexes, weights, and reliabilities. The weight training module is designed to train weights periodically, and the trained weights should be stored in the evidence base for later expert validation. The evidence aggregation mechanism is implemented using the ER rule. First, it identifies matched evidence from the evidence base for each patient, and then it combines multiple pieces of matched evidence to produce an aggregated degree of belief in severe trauma. In this article, the clinical database and evidence base is implemented using the SQL server. The evidence aggregation mechanism and weight training module are implemented in the MATLAB development environment.

D. Prediction Performance Comparison

In the literature, frequently used trauma prediction models are LR [10], [11], [14], [16], and ANN [21]–[23], whose

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

outputs are numerical and are the probabilities linked to the primary outcome. To evaluate the prediction model developed in this article, the ER rule-based model was compared with the ANN and LR models using the trauma dataset collected at Kailuan Hospital.

According to the characteristics of patient data, we tried to divide patient data with similar characteristics into both training and test datasets, and finally around 65% patient data were used for model training and the remaining around 35% were used as test dataset.

For the ER rule model construction, pieces of evidence, as shown in Table IV, were mined from the training dataset. The trained evidence weights are shown in Table V, where reliabilities of those evidence are equal to their corresponding weights. For model test, the ER rule-based evidence aggregation mechanism was utilized to generate a combined degree of belief in the primary outcome for each patient in the test dataset.

For the LR model development, similar to categorizing variables to mine evidence in the ER rule model, we transformed those clinical variables, including age, body temperature, pulse rate, respiration rate, and SBP, which have continuous values to categorical ones by using the same categorizing rules as used in the ER rule model. We used SPSS to perform LR analysis for the training dataset. A total of 22 coefficients were learned from the training dataset. We then used the learned LR function to predict the probability of severe trauma for patients in the test dataset. It is worth noting that an LR model with original continuous values as inputs performs inferior than the LR model as discussed above with transformed categorical inputs.

For the ANN model development, the same set of predicting variables used for the ER rule and LR model development were employed, and those categorical inputs were transformed to several binary ones with 1 or 0 as inputs. In total, there were 25 binary input variables after the variable transformation for the ANN model. The output generated by the ANN model is a numerical value between 0 and 1, and it represents the belief degree of severe trauma. A multilayer perceptron (MLP) network is selected for the ANN model construction. An MLP consists of multiple layers. Generally, an ANN with a layer for inputting, a layer for outputting, and a layer between with a sufficient number of hidden neurons is enough for mapping any relation between inputs and outputs [41]. For this reason, we constructed a three-layer ANN.

We employed the *feedforwardnet*, *train*, and *sim* functions in MATLAB for the ANN model creation, training, and testing. The trial-and-error approach was used to obtain an optimal ANN model. We compared ANN models with different numbers of neurons in the hidden layer, while the number of neurons was calculated using the empirical equation as used in [42]. The number of hidden neurons is between 6 and 15 in this article. The total mean squared error (MSE) of each ANN model was used as a performance criterion to choose an optimal ANN structure. The MSEs in the test dataset generated by ANN models with different structures are shown in Table VI. Based on the MSE of each ANN model, we selected an ANN model with the least MSE, which has 11 hidden neurons.

As the ROC curve is frequently used as one performance measure in trauma studies [16], [21], the area under the ROC curve (AUC) is employed in this article as a performance measure to compare the prediction performance of the three models. Usually, we assess a person as diseased or not according to his or her clinical test value. A threshold value exists for each clinical test, and based on the threshold value, we can calculate *sensitivity*, which is the probability that an assessed diseased is a true one, and *specificity*, which is the probability that a person who is assessed to be healthy is truly healthy. The ROC curve plots *sensitivity* (y-coordinate) versus 1-*specificity* (x-coordinate) for all possible threshold values. The Youden index [43], defined as *sensitivity* + *specificity* − 1, is used as a summary measure of the ROC curve. An optimal cut-off value can be selected by maximizing the Youden index [44]. The AUC of each model and the corresponding *sensitivity* and *specificity* according to the optimal cut-off value of each model are shown in Table VII.

III. Results

The evidence mined from the training dataset are shown in Table IV.

The fine-tuned weights and corresponding evidence are shown in Table V. We can see that most of the weights have been slightly fine-tuned compared with the initial weights assigned by the experts. However, the weights of some pieces of evidence changed substantially after training. For example, the weight of the evidence transformed from clinical variable *Pulse rate* (v_6) when it is labeled as "EWS:1&2&3" changed from 0.9000 to 0.0001 after training. This weight change may be caused by close correlation between the corresponding evidence and its preceding combined evidence, or there are a small number of training cases with variable v_6 labeled as "EWS:1&2&3" having in-hospital death or ICU admission. These trained weights are determined by the real probability distribution on the occurrence of severe trauma in the training dataset.

Table VI shows different MSEs obtained from the test dataset when different ANN models were employed. An optimal three-layer ANN model with 11 hidden neurons was selected because it has the lowest MSE of 0.2521.

The performances of the three prediction models are shown in Table VII. We can find that the best-performing prediction model is the ER rule-based model according to the AUC. In addition, when an optimal cut-off is selected for each model, the three models have the same sensitivity and the ER rule model has the best specificity. Therefore, from both perspectives of AUC and optimal sensitivity and specificity, the ER rule-based prediction model performs best, and the LR model performs better than the ANN model.

IV. Discussion and Conclusion

To optimally manage trauma patients and help them achieve ideal outcomes, trauma patients with a high probability of being admitted to the ICU or dying in hospital need to be

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
EVIDENCE GENERATED FROM THE TRAINING DATASET (Θ_1/Θ_2: SEVERE/NONSEVERE TRAUMA)

Variable		Evidence in BD format
Age (v_1)	Child	$\{(\theta_1,0), (\theta_2,1)\}$
	Adult	$\{(\theta_1,0.5096), (\theta_2,0.4904)\}$
	Old	$\{(\theta_1,0.5206), (\theta_2,0.4794)\}$
Gender (v_2)	Male	$\{(\theta_1,0.5149), (\theta_2,0.4851)\}$
	Female	$\{(\theta_1,0.4341), (\theta_2,0.5659)\}$
Comorbidities (v_3)	Multiple diseases	$\{(\theta_1,0.7835), (\theta_2,0.2165)\}$
	Diabetes	$\{(\theta_1,0), (\theta_2,1)\}$
	Hypertension	$\{(\theta_1,0.3969), (\theta_2,0.6031)\}$
	Heart diseases	$\{(\theta_1,0), (\theta_2,1)\}$
	Other diseases	$\{(\theta_1,0.8167), (\theta_2,0.1833)\}$
	No disease	$\{(\theta_1,0.4666), (\theta_2,0.5334)\}$
Body temperature (v_4)	EWS:0	$\{(\theta_1,0.4466), (\theta_2,0.5534)\}$
	EWS:1&2	$\{(\theta_1,0.5361), (\theta_2,0.4639)\}$
SBP (v_5)	EWS:0	$\{(\theta_1,0.4693), (\theta_2,0.5307)\}$
	EWS:1&2&3	$\{(\theta_1,0.6773), (\theta_2,0.3227)\}$
Pulse rate (v_6)	EWS:0	$\{(\theta_1,0.3171), (\theta_2,0.6829)\}$
	EWS:1&2&3	$\{(\theta_1,0.9143), (\theta_2,0.0857)\}$
Respiratory rate (v_7)	EWS:0&1	$\{(\theta_1,0.3672), (\theta_2,0.6328)\}$
	EWS:2&3	$\{(\theta_1,0.7507), (\theta_2,0.2493)\}$
Level of consciousness (v_8)	EWS:0	$\{(\theta_1,0.2476), (\theta_2,0.7524)\}$
	EWS:1&2&3	$\{(\theta_1,0.9610), (\theta_2,0.0390)\}$
Mechanism of Injury (v_9)	Blunt injury	$\{(\theta_1,0), (\theta_2,1)\}$
	Penetrating injury	$\{(\theta_1,0), (\theta_2,1)\}$
	Crush injury	$\{(\theta_1,0.5175), (\theta_2,0.4825)\}$
	Fall	$\{(\theta_1,0.4712.), (\theta_2,0.5288)\}$
	Traffic accident	$\{(\theta_1,0.7147), (\theta_2,0.2835)\}$
	Others	$\{(\theta_1,0.4325), (\theta_2,0.5675)\}$
Location of injury (v_{10})	Multiple injuries	$\{(\theta_1,0.6979), (\theta_2,0.3021)\}$
	Head, face or neck	$\{(\theta_1,0.6286), (\theta_2,0.3714)\}$
	Main body	$\{(\theta_1,3863), (\theta_2,0.6137)\}$
	Extremities	$\{(\theta_1,0), (\theta_2,1)\}$

identified quickly and accurately upon their arrival at a hospital. The main contribution of this article is an ER rule-based CDSS. The ER rule is further developed and employed in this CDSS for predicting whether a trauma patient needs to be admitted to the ICU or may die in hospital after arriving at the ED. In the CDSS, core components include a clinical database, an evidence mining module, an evidence base, a weight and reliability training module, an ER rule-based evidence aggregation mechanism, and a user interface. Historical patient data are used to mine pieces of clinical evidence that are profiled in the BD format. The evidence mining module is implemented using a generalized Bayesian inference method, which would turn to the ER rule if the probability or belief degree is calculated by normalizing likelihoods. In the inference mechanism, the evidence reliability and weight, together with the dependence index between one piece of evidence and its preceding combined evidence, are taken into consideration in evidence aggregation for outcome prediction. Moreover, an evidence weight training module is developed in this CDSS to extend the ER rule-based inference mechanism, and the training module would help obtain an optimal set of evidence weights that are close to reality.

TABLE V
FINE-TUNED WEIGHTS OF PIECES OF EVIDENCE

Clinical variable	Evidence category	Initial weight	Trained weight
v_1	Child	0.9000	0.8076
	Adult	0.9000	0.9000
	Old	0.9000	0.9000
v_2	Male	0.9000	0.8578
	Female	0.9000	0.8717
v_3	Multiple diseases	0.9000	0.8165
	Diabetes	0.9000	0.9001
	Hypertension	0.9000	0.9003
	Heart diseases	0.9000	0.9000
	Other diseases	0.9000	0.8968
	No disease	0.9000	0.8382
v_4	EWS:0	0.9000	0.9858
	EWS:1&2	0.9000	0.7986
v_5	EWS:0	0.9000	0.9214
	EWS:1&2&3	0.9000	0.3698
v_6	EWS:0	0.9000	0.9992
	EWS:1&2&3	0.9000	0.0001
v_7	EWS:0&1	0.9000	1.0000
	EWS:2&3	0.9000	0.8860
v_8	EWS:0	0.9000	0.9775
	EWS:1&2&3	0.9000	0.7204
v_9	Blunt injury	0.9000	0.9063
	Penetrating injury	0.9000	0.9978
	Crush injury	0.9000	0.9028
	Fall	0.9000	0.9141
	Traffic accident	0.9000	0.5809
	Others	0.9000	0.9363
v_{10}	Multiple injuries	0.9000	0.8080
	Head, face or neck	0.9000	1.0000
	Main body	0.9000	0.9127
	Extremities	0.9000	1.0000

TABLE VI
MSES IN ANN MODELS WITH DIFFERENT ARCHITECTURE

Number of neurons in hidden layer	MSE (test dataset)	Number of neurons in hidden layer	MSE (test dataset)
6	0.2535	11	0.2521
7	0.2637	12	0.2551
8	0.2569	13	0.2528
9	0.2533	14	0.2539
10	0.2528	15	0.2533

TABLE VII
PREDICTION PERFORMANCES OF THE ER RULE, LR, AND ANN MODELS

		ER rule model	LR model	ANN model
AUC		0.9490	0.8440	0.8150
[95% confidence interval]		[0.8970,1.0000]	[0.7390,0.9500]	[0.6600,0.9700]
Maximised Youden Index	Sensitivity	1.0000	1.0000	1.0000
	Specificity	0.8950	0.6890	0.6360

The comparison study shows that the ER rule-based model has superior prediction performance compared with the LR and ANN models. The first possible reason for the results

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

is that the ER rule model makes no assumption about the predicting variables and its evidence base is totally dependent on historical data, while the LR model assumes that the logit transformation of the probability linked to outcome can be expressed using a linear function of predicting variables and the ANN model assumes that the relationship between the predicting variables and the outcome variable can be simulated using layers of nonlinear functions such as sigmoid function. The second possible reason is that the ER rule model takes into consideration the degree of interdependence between one piece of evidence and its preceding combined evidence, while neither the LR nor the ANN model considers the interdependence so explicitly and precisely. Therefore, the ER rule-based prediction model can help model the relationship between predictors and outcomes closer to reality. Consequently, it has better prediction performance than the ANN and LR models, which make assumptions about predictors.

The training of the LR model consumed the least time and is the second best-performing model. The training of the evidence weight in the ER rule-based model took nearly 3 min, and the training of the ANN model consumed more time, because the trial-and-error approach was used to search for an optimal ANN model. In clinical practice, the computation time of model training would not affect the running efficiency of a CDSS, because the model training would be arranged to run on schedule at backend.

The ER rule-based prediction CDSS has the advantages of a reliable and transparent evidence base constructed by mining pieces of evidence from historical data, and a rational and understandable reasoning process which can tell which evidence have been used in evidence aggregation for outcome prediction. However, for the derivation of an ANN model, users with different background knowledge may produce different ANN models even if the same training dataset is used, because they may use different transfer functions and different network structures. The ER rule-based CDSS is innovative in evidence mining, dependence index calculation, and evidence aggregation for outcome prediction. Compared with the LR and ANN models, which are black boxes for clinicians, the evidence base and the evidence aggregation process in the ER rule model are transparent and trackable.

The ER rule-based model is especially useful when prior expert knowledge or clinical experiences are nonexistent, and only historical patient data are available for decision making or prediction. If applying the ER rule-based CDSS to other disease areas, the CDSS may play a significant role in predicting patient outcomes when a new type of infectious disease (e.g., severe acute respiratory syndromes) is spread and professional knowledge about the disease is not available.

However, the accuracy of clinical evidence mined from historical data may be affected by the size and quality of the dataset. The evidence base needs to be updated periodically at the backend.

In conclusion, the design and development of a novel ER rule-based CDSS is presented in this article for predicting trauma patient outcomes. In the era of big data, the CDSS provides an innovative and pragmatic way to predict patient outcomes based on available historical data. The CDSS aims to aid physicians in predicting patient outcomes and thus may suggest appropriate trauma management strategies for ED physicians. One of our next research directions is to generalize the ER rule-based prediction CDSS to other clinical areas and other hospital departments.

Acknowledgment

The authors would like to thank Dr. X. Yin at Peking University People's Hospital and Dr. J. Wang at Peking University Medical Informatics Center.

References

[1] L. T. Kohn, J. M. Corrigan, and M. S. Donaldson, *To Err Is Human: Building a Safer Health System*. Washington, DC, USA: Nat. Acad. Press, 2000.

[2] B. Sheehan *et al.*, "Informing the design of clinical decision support services for evaluation of children with minor blunt head trauma in the emergency department: A sociotechnical analysis," *J. Biomed. Inf.*, vol. 46, no. 5, pp. 905–913, 2013.

[3] E. H. Shortliffe, *Computer-Based Medical Consultations: MYCIN*. New York, NY, USA: Elsevier, 1976.

[4] M. Peleg *et al.*, "GLIF3: The evolution of a guideline representation format," in *Proc. AMIA Annu. Symp.*, 2000, pp. 645–649.

[5] G. Guidi, M. C. Pettenati, P. Melillo, and E. Iadanza, "A machine learning system to improve heart failure patient assistance," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 6, pp. 1750–1756, Nov. 2014.

[6] D. Tay, C. L. Poh, E. Van Reeth, and R. I. Kitney, "The effect of sample age and prediction resolution on myocardial infarction risk prediction," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 3, pp. 1178–1185, May 2015.

[7] S. A. Spooner, "Mathematical foundations of decision support systems," in *Clinical Decision Support Systems: Theory and Practice*, 2nd ed. E. S. Berner, Ed. New York, NY, USA: Springer-Verlag, 2007.

[8] A. Gupta and R. Sharda, "Improving the science of healthcare delivery and informatics using modeling approaches," *Decis. Support Syst.*, vol. 55, no. 2, pp. 423–427, 2013.

[9] M. N. Chawda, F. Hildebrand, H. C. Pape, and P. V. Giannoudis, "Predicting outcome after multiple trauma: Which scoring system?" *Injury*, vol. 35, no. 4, pp. 347–358, 2004.

[10] S. Alghnam, M. Palta, A. Hamedani, M. Alkelya, P. L. Remington, and M. S. Durkin, "Predicting in-hospital death among patients injured in traffic crashes in Saudi Arabia," *Injury*, vol. 45, no. 11, pp. 1693–1699, 2014.

[11] D. A. Kuhls, D. L. Malone, R. J. McCarter, and L. M. Napolitano, "Predictors of mortality in adult trauma patients: The physiologic trauma score is equivalent to the trauma and injury severity score," *J. Amer. Coll. Surg.*, vol. 194, no. 6, pp. 695–704, 2002.

[12] T. Smith, D. D. Hartog, T. Moerman, P. Patka, E. M. M. Van Lieshout, and N. W. L. Schep, "Accuracy of an expanded early warning score for patients in general and trauma surgery wards," *Brit. J. Surg.*, vol. 99, no. 2, pp. 192–197, 2012.

[13] B. Jiang *et al.*, "Transport and public health in China: The road to a healthy future," *Lancet*, vol. 390, pp. 1781–1791, Oct. 2017.

[14] W. K. Utomo, B. J. Gabbe, P. M. Simpson, and P. A. Cameron, "Predictors of in-hospital mortality and 6-month functional outcomes in older adults after moderate to severe traumatic brain injury," *Injury*, vol. 40, no. 9, pp. 973–977, 2009.

[15] M. Ottochian, A. Salim, J. Dubose, P. G. R. Teixeira, L. S. Chan, and D. R. Margulies, "Does age matter? The relationship between age and mortality in penetrating trauma," *Injury*, vol. 40, no. 4, pp. 354–357, 2009.

[16] S. E. Brooks *et al.*, "Do models incorporating comorbidities outperform those incorporating vital signs and injury pattern for predicting mortality in geriatric trauma?" *J. Amer. Coll. Surg.*, vol. 219, no. 5, pp. 1020–1027, 2014.

[17] S. P. Baker, B. O'Neill, W. J. Haddon, and W. B. Long, "The injury severity score: A method for describing patients with multiple injuries and evaluating emergency care," *J. Trauma*, vol. 14, no. 3, pp. 187–196, 1974.

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[18] J. J. Koehler, L. J. Baer, S. A. Malafa, M. S. Meindertsma, N. R. Navitskas, and J. E. Huizenga, "Prehospital Index: A scoring system for field triage of trauma victims," *Ann. Emerg. Med.*, vol. 15, no. 2, pp. 178–182, 1986.

[19] G. Teasdale and B. Jennett, "Assessment of coma and impaired consciousness," *Lancet*, vol. 304, pp. 81–84, Jul. 1974.

[20] J. P. Salomone, "Prehospital triage of trauma patients: A trauma surgeon's perspective," *Prehospital Emerg. Care*, vol. 10, no. 2, pp. 311–313, 2006.

[21] S. M. Dirusso *et al.*, "Development of a model for prediction of survival in pediatric trauma patients: Comparison of artificial neural networks and logistic regression," *J. Pediatr. Surg.*, vol. 37, no. 7, pp. 1098–1104, 2002.

[22] B. Eftekhar, K. Mohammad, H. E. Ardebili, M. Ghodsi, and E. Ketabchi, "Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data," *BMC Med. Inf. Decis. Making*, vol. 5, p. 3, Feb. 2005.

[23] S. D. Izenberg, M. D. Williams, and A. Luterman, "Prediction of trauma mortality using a neural network," *Amer. Surg.*, vol. 63, pp. 275–281, Mar. 1997.

[24] J.-B. Yang and D.-L. Xu, "Evidential reasoning rule for evidence combination," *Artif. Intell.*, vol. 205, pp. 1–29, Sep. 2013.

[25] J.-B. Yang and D.-L. Xu, "On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 32, no. 3, pp. 289–304, May 2002.

[26] J.-B. Yang, J. Liu, J. Wang, H.-S. Sii, and H.-W. Wang, "Belief rule-base inference methodology using the evidential reasoning approach—RIMER," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 36, no. 2, pp. 266–285, Mar. 2006.

[27] M. S. Hossain, F. Ahmed, J. F. Tuj, and K. Andersson, "A belief rule based expert system to assess tuberculosis under uncertainty," *J. Med. Syst.*, vol. 41, p. 43, Mar. 2017.

[28] M. S. Hossain, S. Rahaman, R. Mustafa, and K. Andersson, "A belief rule-based expert system to assess suspicion of acute coronary syndrome (ACS) under uncertainty," *Soft Comput.*, vol. 22, no. 22, pp. 7571–7586, 2018.

[29] F.-J. Zhao, Z.-J. Zhou, C.-H. Hu, L.-L. Chang, Z.-G. Zhou, and G.-L. Li, "A new evidential reasoning-based method for online safety assessment of complex systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 6, pp. 954–966, Jun. 2018.

[30] J.-B. Yang and D.-L. Xu, "A study on generalising Bayesian inference to evidential reasoning," in *Belief Function: Theory and Applications, Lecture Notes in Artificial Intelligence*, F. Cuzzolin, Ed. Cham, Switzerland: Springer Int., 2014, pp. 180–189.

[31] J.-B. Yang and D.-L. Xu, "Inferential modelling and decision making with data," in *Proc. 23rd Int. Conf. Autom. Comput. (ICAC)*, Huddersfield, U.K., 2017, pp. 1–6.

[32] J.-B. Yang, D.-L. Xu, P. Stachow, and X. B. Xu, "Evidential reasoning for risk analysis," in *Proc. 2nd GOSS Private Equity Forum 9th RMI Annu. Risk Manage. Conf.*, Singapore, 2015, pp. 1–9.

[33] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. Boca Raton, FL, USA: Chapman & Hall/CRC, 2004.

[34] R. E. Neapolitan, *Learning Bayesian Networks*. Upper Saddle River, NJ, USA: Pearson Prentice-Hall, 2004.

[35] G. L. Kong *et al.*, "Current state of trauma care in China, tools to predict death and ICU admission after arrival to hospital," *Injury*, vol. 46, no. 9, pp. 1784–1789, 2015.

[36] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[37] K. J. Rothman, S. Greenland, and T. L. Lash, *Modern Epidemiology*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2008.

[38] J. Tashiro *et al.*, "Mechanism and mortality of pediatric aortic injuries," *J. Surg. Res.*, vol. 198, no. 2, pp. 456–461, 2015.

[39] N. V. Chawla, "Data mining for imbalanced datasets: An overview," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Boston, MA, USA: Springer, 2005, pp. 853–867.

[40] R. Longadge, S. S. Dongre, and L. Malik, "Class imbalance problem in data mining: Review," *Int. J. Comput. Sci. Netw.*, vol. 2, no. 1, p. 83, 2013.

[41] Y.-M. Wang and T. M. S. Elhag, "A comparison of neural network, evidential reasoning and multiple regression analysis in modelling bridge risks," *Expert Syst. Appl.*, vol. 32, no. 2, pp. 336–348, 2007.

[42] G. L. Kong *et al.*, "Belief rule-based inference for predicting trauma outcome," *Knowl. Syst.*, vol. 95, pp. 35–44, Mar. 2016.

[43] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.

[44] R. Fluss, D. Faraggi, and B. Reiser, "Estimation of the Youden index and its associated cutoff point," *Biometrical J.*, vol. 47, no. 4, pp. 458–472, 2005.

Guilan Kong received the B.Eng. and M.Eng. degrees in computer science and technology from the National University of Defense Technology, Changsha, China, in 1997 and 2000 respectively, and the Ph.D. degree in decision and system sciences from the University of Manchester, Manchester, U.K., in 2011.

She is currently an Associate Research Professor with the National Institute of Health Data Science, Peking University, Beijing. Her research interests include clinical decision support systems, artificial intelligence in medicine, medical data mining, and medical knowledge management.

Dr. Kong is an Associate Chair of Smart Health and Senior Care Committee associated to China Association of Gerontology and Geriatrics and a member of the National Traffic Trauma and Trauma Database Professional Committee and the Expert Committee of Trauma Medicine Centre at Peking University.

Dong-Ling Xu received the B.Eng. degree in electrical engineering from Hunan University, Changsha, China, in 1983, and the M.Eng. and Ph.D. degrees in system control engineering from Shanghai Jiao Tong University, Shanghai, China, in 1986 and 1988, respectively.

She is a Professor of Decision Sciences and Systems with the Allaince Manchester Business School, University of Manchester, Manchester, U.K. She has developed several interactive Web-based decision support tools, windows-based intelligent decision systems (IDS), and fault detection systems. Those systems are used in a wide range of assessment and decision-making activities by organizations, such as General Motors, Tesco, NHS, Ford, Shell, BP, and CNOOC in areas, such as safety and risk assessment, organizational self-assessment in quality management, policy assessment, and pipeline leak detection. The IDS software is used by practitioners and researchers from over 50 countries. She has published over 150 peer-reviewed journal papers, book chapters, and conference papers. Her research interests include multiple criteria and multiple objective decision modeling, analysis, and support under uncertainty; complex system and process modeling; and statistical fault detection and system development.

Jian-Bo Yang received the B.Eng. and M.Eng. degrees in system control engineering from North Western Polytechnic University, Xi'an, China, in 1981 and 1984, respectively, and the Ph.D. degree in decision sciences and system engineering from Shanghai Jiao Tong University, Shanghai, China, in 1987.

He was a Faculty Member of the University of Manchester Institute of Science and Technology, Manchester, U.K., in 1990 and from 1998 to 2004, the University of Birmingham, Birmingham, U.K., from 1995 to 1997, the University of Newcastle-upon-Tyne, Newcastle upon Tyne, U.K., from 1991 to 1995, and Shanghai Jiao Tong University from 1987 to 1989. He is a Professor of Decision and System Sciences and the Director of the Decision and Cognitive Sciences Research Center with the Alliance Manchester Business School, University of Manchester, Manchester. In the past three decades, he has been conducting research in multiple criteria decision analysis under uncertainty, multiple objective optimization, intelligent decision support systems, hybrid quantitative and qualitative decision modeling using techniques from operational research, artificial intelligence, and systems engineering, and dynamic system modeling, simulation, and control for engineering and management systems. His current applied research is in design decision making, risk modeling and analysis, production planning and scheduling, quality modeling and evaluation, supply chain modeling and supplier assessment, and evaluation of products, systems, projects, and policies. His research has been supported by EPSRC, DEFRA and SERC, HKRGC in U.K. and in Hong Kong, NSFC in China, and EC in Europe and industry. He has published 3 books, over 140 papers in international journals/book chapters, and a similar number of papers in conferences and developed several software packages in these areas, including the windows-based intelligent decision system via evidential reasoning approach.

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Tianbing Wang received the B.Med. degree in clinical medicine from Xi'an Medical University, Xi'an, China, in 1993, the M.Med. degree in hand surgery from the Shanghai Medical College, Fudan University, Shanghai, China, in 2001, and the M.D. degree in orthopedics from Peking University, Beijing, China, in 2009.

He is a Professor of Trauma and Emergency Medicine with the Trauma Center of Peking University People's Hospital, Beijing. He has published more than 30 peer-reviewed journal papers, book chapters, and conference papers. His research interests include limb fracture, hand surgery, peripheral nerve injury, and big data in trauma.

Prof. Wang currently serves as a member of the Standing Committee and the Secretary-General of China Trauma Rescue and Treatment Association, the Chairman of the Pharmaceutical Information Professional Committee of Chinese Pharmaceutical Association, the Deputy Director of the Trauma Treatment and Nerve Regeneration Laboratory (Key Laboratory of the Ministry of Education, China), and the Deputy Director of the Engineering Research Center for Mobile Digital Hospital System of the Education Ministry of China.

Baoguo Jiang received the B.Med. and M.Med. degrees in clinical medicine from China Medical University, Shenyang, China, in 1984 and 1987, respectively, and the M.D. degree in orthopedics from Beijing Medical University (later renamed as Peking University Health Science Center), Beijing, China, in 1992.

He is a Professor and an Academic Leader of Orthopedics with Peking University People's Hospital, Beijing. He has authored more than 370 scientific articles, book chapters and conference papers, and his work has been published in high-impact, peer-reviewed journals, including *Lancet*. His research interests include peripheral nerve injury and repair, spinal cord injury, biomechanics of articular fracture healing, trauma treatment, and orthopedic surgical methods.

Prof. Jiang has received several distinguished awards, including the National Science Fund for Outstanding Youth in 2006, the Chinese Medical Science and Technology Award in 2006, the Technical Invention Award from the Ministry of Education in 2010 and 2012, and the Gusi Peace Prize from Gusi Foundation in 2016. He is the President of Peking University People's Hospital and the Director of the National Center for Trauma Medicine and the Orthopedics Department, Peking University. He is the Editor-in-Chief of the *Chinese Journal of Shoulder and Elbow*, an Editorial Board Member of *Artificial Cells Blood Substitutes and Immobilization Biotechnology*, the Deputy Editor-in-Chief of the *Chinese Journal of Traumatology*, *Journal of Chinese Microsurgery*, and *Journal of Chinese Orthopaedic Trauma*. He is the Chief Scientist of China National "973" Project. He served as the Chairman of the Society of Traumatology of Chinese Medical Association (CMA).

Authorized licensed use limited to: University of Manchester. Downloaded on October 20,2020 at 22:12:51 UTC from IEEE Xplore. Restrictions apply.