Early-Warning of Peri-operative Critical Event Based on Multimodal Information Fusion

Yuwen Chen

Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu, China; Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China; University of Chinese Academy of Sciences, Beijing, China

Yujie Li

Department of Anesthesiology, Southwest Hospital, The Third Military Medical University, Chongqing, China.

Wei Huang

Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China; University of Chinese Academy of Sciences, Beijing, China

Ju Zhang

Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China

Bin Yi

Department of Anesthesiology, Southwest Hospital, the Third Military Medical University, Chongqing, China

Xiaolin QIN*

Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu, China

ABSTRACT

The occurrence of perioperative critical adverse events will affect the quality of medical services and threaten the safety of patients. Using scientific methods to assess the risk of critical illness in perioperative period is of great significance to improve the quality of medical service and ensure the safety of patients. However, the diagnosis and treatment data of perioperative patients are multi-source and irregular, and only one physiological information can not accurately reflect the patient's condition. In previous studies, it is found that a variety of physiological information can transmit the information of human health or not, which can be used to evaluate critical illness and physical condition. Therefore, this paper integrates the preoperative clinical structure data, intraoperative vital signs monitoring time series data and intraoperative anesthesia event time series data. Based on deep learning technology, the multimodal data of patients are embedded and mapped into the same recessive semantic space to realize the real-time tracking and early warning of severe events, reduce postoperative complications and improve the early diagnosis efficiency of critical adverse events. The results showed that the performance of the model based on the multi-modal data was better than that based on the real military data.

CCS CONCEPTS

• Applied computing; • Health informatics;

*Corresponding Author: XIAOLIN QIN. Email: keyanche@163.com.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICIMH 2021, August 13–15, 2021, Macau, China

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8590-9/21/08...\$15.00

<https://doi.org/10.1145/3484377.3484389>

KEYWORDS

Multi-modal, Critical event, Risk prediction

ACM Reference Format:

Yuwen Chen, Yujie Li, Wei Huang, Ju Zhang, Bin Yi, and Xiaolin QIN. 2021. Early-Warning of Peri-operative Critical Event Based on Multimodal Information Fusion. In *2021 the 3rd International Conference on Intelligent Medicine and Health (ICIMH 2021)*, August 13–15, 2021, Macau, China. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3484377.3484389>

1 INTRODUCTION

More than 310 million surgical procedures are carried out worldwide every year [1]. A prospective cohort study in 27 countries showed that about 16.8% patients developed complications and 0.5% died before hospital discharge after elective surgeries [2]. Postoperative complications increase treatment costs and reduce both life expectancy and quality of life. However, about half of post-operatively adverse events are avoidable [3]. Applying surgical safety checklist recommended by WHO to the clinical work reduces the mortality but makes no statistical significance in post-operative complication rate [4]. It may be owing to the clinical outcomes (post-operative complications and death) are associated with both patients' factors and every aspect of medical interventions. Moreover, these factors have internal relationships with each other, for example, patients' pre-operative health may affect the decision-making of surgery or anesthesia, and the quality of surgery may affect the intra- and post-operative managements. Therefore, relationships between factors and clinical outcomes have the characteristics of causality and time series which made the risk assessment of post-operative mortality and complications quite difficult.

From last century, numerous studies have used various scoring systems to perioperative predict the clinical outcome of surgical patients. Among them, the most famous and widely pre-operatively used is the American Society of Anesthesiologists (ASA) physical status classification. ASA classification uses medical history and physical status evaluation to predict the risk of 30-day mortality after surgery, with the limitation of subjectivity and low specificity[5].

The Surgical Apgar score uses intra-operative variables, namely, estimated blood loss, the lowest mean arterial pressure and heart rate during surgery to predict major postoperative complications, which achieved poor ability of discrimination due to its quantity of information [6]. Preoperative Score to Predict Postoperative Mortality (POSPOM) is developed and validated in a large population which using medical history and surgical type as variables based on Logistics Regression (LR) [7]. Despite the consideration of surgical type and comorbidities, POSPOM scoring system lacks the information of intra-operative and post-operative management which may explain the low specificity or sensitivity in predicting in-hospital mortality on another cohort[5]. Except for the scoring systems above, some researchers tried to combine pre-operative and intra-operative Surgical Apgar score systems for risk assessment of mortality but with a negative result [8]. The existing scoring systems based on conventional statistical methods cannot meet the need to dynamically predict, accurately early warning and automatically evaluate the risk of post-operative complications and mortality.

As the development of medical informatic systems and machine learning methods, some groups tried to use high dimensional clinical variables based on different machine learning algorithms to predict the clinical outcomes of surgical patients. Bihorac et al [9] constructed an automated predictive analytics tool based on a generalized additive model (GAM) using only preoperative predictors such as demographic, laboratory results, to assess the risk of surgical complications and death at different time point post-operatively. Deep neural networks can also predict in-hospital mortality based on pre-operative and intra-operative data from electronic medical records, but are not superior to existing methods [5]. Furthermore, the automatically extractable intraoperative data are statistical features of the whole surgery not time series data, such as the minimum or average of blood pressure during operation [5]. The above two methods to assess post-operative outcome make up the problems of automation, however, none of the two researches can realize dynamically assessment at different stage of hospital stay for they did not deal with time series data. For the high dimension, complicated medical data, and rapidly changing status of patients, methods dealing with time series data from various sources or with internal associated are badly needed. Therefore, we aimed to present a multimodal deep learning algorithm using medical history, intra-operative anesthesia events and monitoring data to dynamically assess the risk of perioperative critical complications and death, which may realize early warning of critical events and help doctors make decisions.

The main contributions of this paper can be summarized as follows: We propose a method of implicit semantic mapping of medical multimodal data, and implement the prediction model of critical illness based on the temporal deep learning network of attention mechanism, which is verified in real medical practice.

2 METHODOLOGY

2.1 Multi-modal critical event prediction

Let there be I represent the perioperative patient data set. Each patient i is associated with a set of static variables $s_i \in R^{ms}$, such as preoperative examination and demographic information. As well

as, it also includes the monitoring temporal sequence data of vital signs $v_{i,t}$ and the operation data of anesthesia $e_{i,t}$. As shown in Figure 1, critical event risk prediction is based on patients' past observable diagnosis and treatment information $[s_i, v_{i,t}, e_{i,t}]$, to predict the risk of critical illness in the future. The definition is as follows:

$$Y(t) = F(y_{i,t-1}, s_i, v_{i,t-1}, e_{i,t-1})$$

Where $Y(t)$ is the patient's predicted risk at time t , and $F(\cdot)$ is a prediction model.

In line with other methods, we not only consider the static variables of patients' preoperative diagnosis and treatment, but also consider the intraoperative monitoring situation of patients and the doctor's operation events, and integrate all the information of patients into a limited time window, and make prediction based on temporal data.

2.2 Model Architecture

Patients' health and doctors' operation are the main factors leading to critical illness. In view of this, we integrate the diagnosis and treatment data of patients and doctors' operation data, and predict the risk of critically ill patients based on deep learning technology. Figure 2 shows the overall framework of our approach for critical event prediction (CEP). It consists of six part:1) Gated Residual Network (GRN),2) Variable selection networks(VSN),3) Static variable encoder(SVE),4) Temporal processing(TP),5) Multi-Head Attention(MHA),6)Risk Prediction(RP). GRN enable efficient information flow with skip connections and gating layers. Variable Selection is used for judicious selection of the most salient features based on the input. SVE integrates static features into the network through context vector coding. Temporal processing learns the long-term and short-term temporal relationships from the observed preoperative and intraoperative data. Local relationships are learned by recursive networks, while long-term dependence is captured by multi attention mechanism. RP predicts the patient's risk through forward network. Individual components described in detail in the subsequent sections.

In our model, we use entity embeddings [10] for anesthetic event variables as feature representations, and linear transformations were used to encode preoperative static variables and intraoperative monitoring data variables- transforming each input variable into a d_{model} -dimensional vector which matches the dimensions in subsequent layers for skip connections. s_i is the static variable vector of patient i . $s_i \in R^{NX1}$, N is the number of static variables of patients. v_i is the vital sign vector monitored during operation of patient i . $v_i \in R^{MXT}$, M is the number of intraoperative monitoring attributes, and T is the length of monitoring time. e_i is the anesthesia operation event of patient i during the operation. $e_i \in R^{KX1}$, K is the number of anesthesia events.

$$s_{i_emb} = Linear(s_i)$$

$$v_{i_emb} = Linear(v_i)$$

$$e_{i_emb} = Embedding(e_i)$$

Where Linear is linear transformations [11], Embedding is Word embedding [12]. $s_{i_emb} \in R^{NX d_{model}}$, $v_{i_emb} \in R^{MXTX d_{model}}$, $e_{i_emb} \in R^{KX d_{model}}$.

Figure 1: Illustration of multi-modal critical event prediction with static variables and observable patient temporal data inputs.

Figure 2: Critical event prediction (CEP) architecture.

2.3 Gated Residual Network

The gating mechanisms controlling the path through which information flows to the subsequent layers are widely used in the famous sequence models like LSTM [13], GRU [14], and ConvS2S [15]. Such multiplicative gates are adept at dealing with gradient explosion

and vanishing, which enable the information to propagate unimpededly through long time-steps or deep layers. In order to extract the precursory information of critically ill patients effectively, we introduced the gate residual network (GRN) as the basic building block of the model. As shown on the top left of Figure 2, GRN will

input a and a context vector c into the network and output:

$$GRN_w(a, c) = \text{LayerNorm}(a + GLU_w(\eta_1))$$

$$\eta_1 = W_1\eta_2 + b_1$$

$$\eta_2 = \text{ELU}(W_2a + W_3c + b_2)$$

Where ELU is the Exponential Linear Unit activation function [16]. $\eta_1 \in R^{d_{model}}$, $\eta_2 \in R^{d_{model}}$ are intermediate layers, Layer-Norm is standard layer normalization of [17]. GLU is Gated Linear Units [18]. Letting $\gamma \in R^{d_{model}}$ be input, the GLU then takes the form:

$$GLU_w(\gamma) = \sigma(W_3\gamma + b_3) \odot (W_4\gamma + b_4)$$

Where $\sigma(\cdot)$ is the sigmoid activation function, W and B are weight parameters and bias parameters respectively, $\odot$ is the element-wise Hadamard product[19]. During training, dropout [20] is applied before the gating layer and layer normalization.

2.4 Variable selection network

While multiple patient variables may be available, their relevance and specific contribution to the risk are typically unknown. Therefore, we design a variable selection network for patients with pre-operative static variables and intraoperative dynamic variables to achieve the selection of variables and delete unnecessary noise input to reduce the impact on the stability of the model. The variable selection network is constructed based on the gate residual network. As shown in the bottom left of Figure 2. Preoperative static variables, intraoperative vital signs monitoring variables and intraoperative anesthesia event operation variables were separately processed by different variable selection network.

Here we take the intraoperative vital signs monitoring variables as an illustration. Let $\xi_{j,t} \in R^{d_{model}}$ denote the transformed input of the j -th vital signs monitoring variable at time t , with $\varphi_t = \xi_{1,t}, \xi_{2,t}, \dots, \xi_{M,t}$, being the flattened vector of all the vital signs monitoring variable at time t . Variable selection weights are generated by feeding both φ_t and an external context vector c_s through a GRN, followed by a Softmax layer:

$$v_t = \text{softmax}(GRN(\varphi_t, c_s))$$

Where c_s is the preoperative static variables coding vector.

Each monitoring variable is transformed by a separate GRN and then combined with the weight to get the output,

$$\overline{\xi_{j,t}} = GRN(\xi_{j,t})$$

$$\widetilde{\xi}_t = \sum_{j=1}^m v_t \overline{\xi_{j,t}}$$

The external context variables are empty, in the selection of pre-operative static variables. When selecting variables from intraoperative vital signs monitoring data and anesthesia event operation data, context variables are the coding vectors of preoperative static variables.

2.5 Preoperative Attribute Variable encoder

Preoperative static attributes play an important role in prediction. In order to construct feature representation, three separate GRN encoders are used to generate different static variable context vectors and fuse them into different positions of the model.

$$c_s = GRN(s_{i_emb})$$

$$c_f = GRN(s_{i_emb})$$

$$c_h = GRN(s_{i_emb})$$

When encoding static attributes, the external context variable is empty. c_s is used as an external context vector in the selection of intraoperative monitoring data and anesthesia operation events. c_f is used as an external context vector to enhance temporal features. c_h is used as an external context vector for the hidden state of GRU.

2.6 Temporal processing

The patient's intraoperative monitoring sign data and anesthesia operation event data respectively generate corresponding temporal features after passing through the variable selection network, and use two independent GRU networks to encode the time series data. The first hidden state of the GRU anesthesia event encoder is the last hidden state of the GRU encoder in the monitoring sign. As shown in Figure 2, Gate mechanism and regularization are applied to GRU model, which is convenient for model training.

$$v_{\xi_t}, h_{v_{\xi_t}} = GRU(\widetilde{\xi_{vt}}, c_h)$$

$$e_{\xi_t}, h_{e_{\xi_t}} = GRU(\widetilde{\xi_{et}}, h_{v_{\xi_t}})$$

$\widetilde{\xi_{vt}}$ and $\widetilde{\xi_{et}}$ are the monitoring time data and time series data of anesthesia event after variable selection respectively. v_{ξ_t} and e_{ξ_t} are the outputs of GRU model based on gate mechanism, respectively.

In order to fuse the static information of patients before operation, we fused the preoperative coding features with the post-processing temporal features after GRU transformation.

$$v_{\xi_{ct}} = GRN(e_{\xi_t}, c_f)$$

$$e_{\xi_{ct}} = GRN(v_{\xi_t}, c_f)$$

$v_{\xi_{ct}}$ and $e_{\xi_{ct}}$ are the local temporal features after fusion of preoperative static data. We concatenate them into a vector which contains the feature of vital signs and anesthesia events. The formula is:

$$ve_{\xi_c}(t) = [v_{\xi_{ct}}; e_{\xi_{ct}}]$$

In order to learn the long-distance dependence of temporal data, we use the attention mechanism to process the fused patient feature. Here we start with the scaled dot product attention [21], in which the inner product is applied to calculate the attention weights. In practice,

$$\text{Attention}(Q, K, V) = A(Q, K) V$$

Where $A(\cdot)$ is a normalization function. A common choice is scaled dot-product attention.

$$A(Q, K) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_{attn}}}\right)$$

Figure 3: Distribution of patients

To extend the capacity of exploring subspaces, we employ the multi-head attention [22] which consists of h parallel scaled dot-product attentions named head. The inputs including queries, keys, and values are projected into h subspaces, and the attention performs in the subspaces separately:

$$\text{MultHead}(Q, K, V) = [H_1, \dots, H_m] W_H$$

$$H_h = \text{Attention}\left(QW_Q^h, KW_K^h, VW_V^h\right)$$

$$W_K^h, W_Q^h \in R^{d_{model} \times d_{atten}}, W_V^h \in R^{d_{model} \times d_v}$$

For the training of the model, we use the gate control layer after the attention layer. The formula is:

$$B(t) = \text{MultHead}\left(ve_{\xi_c}(t), ve_{\xi_c}(t), ve_{\xi_c}(t)\right)$$

$$\delta(t) = \text{LayerNorm}\left(ve_{\xi_c}(t) + \text{GLU}(B(t))\right)$$

Finally, we apply an additional non-linear processing to the outputs of the self attention layer. this makes use of GRN:

$$\Psi(t) = \text{GRN}(\delta(t))$$

Where the weights of GRN are shared across the entire layer. We also apply a gated residual connection which skips over the entire block, providing a direct path to the sequence-to-sequence layer, as shown below:

$$\tilde{\Psi}(t) = \text{LayerNorm}(\delta(t) + \text{GLU}(\Psi(t)))$$

2.7 Risk Prediction

Finally, we feed the vectors $\Psi(t)$ into classifier to predict patient risk:

$$R_t = \text{SoftMax}(\Psi(t))$$

The loss function is written as follows:

$$L(\Omega, W) = \sum_{i=1}^M y_i \log(\tilde{y}_i)$$

Where Ω is the training data including M samples and W is the learning weight of the model

3 EXPERIMENTS

3.1 Datasets

The data set of this study is from the clinical data of surgical patients in the First Affiliated Hospital of Military Medical University of the Army from 2014 to 2018, including demographic data, laboratory data, intraoperative monitoring vital signs, intraoperative anesthetic events and other diagnostic information. The patient data is divided into static data and dynamic data: intraoperative vital signs monitoring data and anesthesia event time series data, the rest of the clinical data are static data. In this paper, the patients with critical illness were selected as positive data set, including 405 cases of heart failure, 129 cases of liver failure, 81 cases of renal failure, 115 cases of respiratory failure, 82 cases of sepsis and 1456 cases of negative patients. The distribution is shown in Figure 3

3.2 Train Details and Experimental Setup

This section mainly shows the training details and the parameter setting. The performance of the model was tested by accuracy, sensitivity / specificity and ROC value. The data were divided into training set and test set. K-fold cross validation was used to evaluate the model performance with $k = 5$. As shown in Figure 4. The parameter setting is shown in the following subsections.

We counted the intraoperative monitoring attributes of all patients, and selected the 5 attributes [heart rate, systolic blood pressure, diastolic blood pressure, spO2, respiration] with the highest monitoring frequency as the monitoring data input. Embedding coding vector dimension d_{model} is 24. We chose 48 types of anesthesia events, See Table 1 for details. Embedding coding vector dimension d_{model} is 24. We selected 20 index variables according to the doctor's experience and the preoperative test index of all critically ill patients. See table 2 for details. The vector dimension of GRU hidden layer is set to 24, which overlaps two layers. And $h = 8$ parallel attention layers are employed in multi-head attention. The variants of our models are trained with stochastic gradient descent [23] within adaptive learning rates. The batch size is 32. The learning rate is 0.001, the dropout [24] regularization ($P = 0.3$) is applied to the embedded layer and hidden layer, and the weight attenuation coefficient [25] is 0.001.

Figure 4: Schematic diagram of data division

Table 1: Preoperative static variables

Preoperative patient static variable attribute name
Age, 5'-Nucleotidase, Serum Chloride, Glomerular filtration rate, D-lactic-pyruvic transaminase, Creatinine, glucocholic acid, serum potassium, serum sodium, hyaline cylindruria, α -L-fucosidase, urea, fibrinogen, glutamic-oxaloacetic transaminase, cystatin-C, fibrinogen degradation product (FDP), D-dimer, osmotic pressure, platelet count

Table 2: Intraoperative anesthetic variables

Intraoperative anesthetic variables attribute name
monitoring depth of anesthesia, intraoperative warming, Sputum Aspiration, relaxation monitor (TOF Watch SX), arterial catheterization, Internal jugular vein catheterization, Subclavian vein catheterization, arterial blood gas analysis, thrombelastogram monitoring, ultrasound-guided Transversus Abdominis Plane Block, Patient-controlled femoral nerve block analgesia (patient-controlled epidural analgesia) PCEA, patient controlled intravenous analgesia, nasogastric intubation, video laryngoscope-guided trachea intubation, Barometric Pressure Therapeutic Apparatus, Monitoring of cardiac output, doctor shifts, deliberate hypotension, manual hyperinflation, monitoring of Cerebral oxygen saturation, positive end expiratory pressure, low Tidal Volume Mechanical Ventilation, Roll over and pat back, crystalloid-colloid ratio, albumin, erythrocyte suspension, Cisatracurium Besilate, Sufentanil, remifentanyl, etomidate, fentanyl, dexmetomidine, flumazenil, Nalmefene hydrochloride injection, Neostigmine Methylsulfate, atropine sulfate, sevoflurane, nitrous oxide, Ephedrine hydrochloride, dopamine hydrochloride, Esmolol Hydrochloride, lansoprazole, ulinastatin, heamocoagulase agkistrodon, tranexamic acid, urbason/hydrocortisone/dexamethasone, Pentethylglycoline hydrochloride, Ambroxol Hydrochloride, normal saline (antibiotic)

3.3 Experimental Results

Table 3 and Figure 5 show the performance of our model and several other methods. The contrast models can be roughly divided into three categories. The first category only uses preoperative static data of patients. The second category only uses intraoperative monitoring of vital signs data. The third category only uses intraoperative anesthetic event data.

For patient static data, we applied logistic regression, SVM, decision tree, and multilayer perceptron to conduct experiments. We used TCN [26] model and GRU model to monitor vital signs and anesthesia operation data. In addition, experiments were also carried out based on the statistical characteristics of intraoperative monitoring signs [27]. Based on the single-mode preoperative static

data, the best effect can be achieved ROC value (0.82), sensitivity (0.81), specificity (0.90). Based on the single-modality intraoperative sign monitoring data, the roc value is 0.82, the sensitivity is 0.72, and the specificity is 0.72. Based on the single-modality anesthesia event data, the roc value is 0.69, the sensitivity is 0.52, and the specificity is 0.64. Our model can get the ROC value (0.94), sensitivity (0.88), specificity (0.94). The sensitivity and specificity of the model were increased by 13%, 28%, 36%, 7%, 11% and 33% respectively. From the experimental results, we can see that the multi-mode data time series fusion can improve the prediction of critical illness, and also verify the effectiveness of the proposed model. The experimental results show that the single-mode data have different discriminative power. Combined with the complementary information of the two

Table 3: Experimental Results

Model	p/precision	sensitivity	n/precision	specificity	ROUAAU
Pre- logistic	0.71	0.73	0.74	0.70	0.73
Pre- SVM	0.81	0.74	0.78	0.89	0.71
Pre- DT	0.81	0.75	0.81	0.90	0.82
Pre- MLP	0.78	0.69	0.72	0.85	0.70
Intra-V-TCN	0.69	0.53	0.74	0.85	0.81
Intra-V-GRU	0.72	0.60	0.72	0.86	0.82
Intra-V-stat	0.61	0.60	0.64	0.74	0.73
Intra-E-TCN	0.58	0.51	0.71	0.62	0.67
Intra-E-GRU	0.59	0.52	0.72	0.64	0.69
Ours	0.93	0.88	0.94	0.97	0.94

Figure 5: Comparison of experimental results

modes, modal fusion can significantly improve the classification accuracy (0.93%). By analyzing the correlation and complementarity of each modal data in the expression ability or information tendency, multimodal fusion achieves a deep understanding of the high-level semantics of the data, so as to better realize the diagnosis and prediction of the patient's condition.

4 DISCUSSION AND CONCLUSION

It is a complex process to predict critical events in perioperative patients. Whether serious adverse events occur during and after operation depends entirely on the experience and judgment of doctors. Doctors with long working time and rich experience have higher prediction accuracy, but doctors with short working time or lack of experience have lower prediction accuracy. It is a serious problem that the judgment based on doctors' experience is lagged and the evaluation results can't be directly applied effectively. Aiming at the intelligent application of critical illness prediction, this paper explores the learning content of multimodal medical data feature representation. The data from different modalities characterize the patient from different aspects, and the complementary features from the different modalities can be fused to build an early identification of critical illness that is more robust than the single-modality method. Therefore, this paper studies the use of

multi-modal data features for Critical illness prediction based on deep learning technology to pass preoperative patient clinical structure data, intraoperative patient vital sign monitoring data, and intraoperative anesthesia event operation data through gating and self-attention mechanisms. Multi-modal numbers are embedded in the feature representation, and multi-stage, multi-source data is fused to predict risks. Therefore, this paper studies the use of multimodal data for critical illness risk prediction. Based on the deep learning technology, we embedded the clinical structure data, vital signs monitoring data and operation data of anesthesia events by the gating mechanism. Based on recurrent network learning temporal feature of patient data, and finally risk prediction based on attention mechanism. The model was experimentally verified on the data of heart failure, liver failure, renal failure, respiratory failure, and sepsis in the patient data set of the anesthesiology department of Southwest Hospital. The experimental results show that the multi-modal model proposed in this paper can effectively predict critical illness, and has higher accuracy than single-mode prediction. In addition, the model proposed in this paper can be applied to the prediction of other diseases according to different critical illness event data sets. The future research directions of this paper include the following aspects: the video data of patients'

operation includes surgeons' relevant operations on patients. Surgeons' methodical, clear anatomy and meticulous operation can greatly reduce the trauma of patients and promote the rapid recovery of patients, and reduce the occurrence of critical adverse events during and after the operation. On the contrary, those operations with confused anatomy, unclear organization and muddled operation will increase the probability of trauma and occurrence of critical adverse events, and also prolong the length of hospital stay. Increase medical expenses. That is, surgeons' operation has an indirect effect on the occurrence of serious and adverse complications. In the later stage of this study, the intraoperative surgeon video operation information is fused with multimodal prediction model, and the comprehensive information of patients is extracted from more comprehensive data to predict the perioperative critical illness. In addition, the model proposed in this paper is still a black box for doctors. If we can generate explanatory text to explain the model at the same time, we can make the conclusion of the model more accurate and accurate. This is also a key direction of the later research in this paper.

ACKNOWLEDGMENTS

This work is supported by the National Key Research & Development Plan of China (2018YFC0116704) in data collection. In addition, it is supported by Chongqing Technology Innovation and application research and development project (cstc2019jscx-msxmx0237) in the design of the study.

REFERENCES

- [1] Weiser, T.G., *et al.*, Estimate of the global volume of surgery in 2012: an assessment supporting improved health outcomes. *Lancet*, 2015. 385 Suppl 2((Suppl 2): S11): p. S11.
- [2] International Surgical Outcomes Study, g., Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. *Br J Anaesth*, 2016. 117(5): p. 601-609.
- [3] Gawande, A.A., *et al.*, The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery*, 1999. 126(1): p. 66-75.
- [4] Abbott, T.E.F., *et al.*, The surgical safety checklist and patient outcomes after surgery: a prospective observational cohort study, systematic review and meta-analysis. *Br J Anaesth*, 2018. 120(1): p. 146-155.
- [5] Lee, C.K., *et al.*, Development and Validation of a Deep Neural Network Model for Prediction of Postoperative In-hospital Mortality. *Anesthesiology*, 2018. 129(4): p. 649-662.
- [6] Haynes AB, R.S., Weiser TG, Lipsitz SR, Dziedan G, Berry WR, Gawande AA. , Surgical outcome measurement for a global patient population: Validation of the Surgical Apgar Score in 8 countries. *Surgery*, 2011. 149: p. 519-24.
- [7] Le Manach, Y., *et al.*, Preoperative Score to Predict Postoperative Mortality (POSPOM): Derivation and Validation. *Anesthesiology*, 2016. 124(3): p. 570-9.
- [8] Terekhov MA, E.J., Wanderer JP Preoperative surgical risk predictions are not meaningfully improved by including the Surgical Apgar score: An analysis of the Risk Quantification Index and present-on-admission risk models. *ANESTHESIOLOGY*, 2015. 123: p. 1059-66.
- [9] Bihorac, A., *et al.*, MySurgeryRisk: Development and Validation of a Machine-learning Risk Algorithm for Major Complications and Death After Surgery. *Ann Surg*, 2019. 269(4): p. 652-662.
- [10] Lai, S., *et al.*, How to Generate a Good Word Embedding. 2016. 31(6): p. 5-14.
- [11] Desplas, M., Generalized Inverses of Linear Transformations. 1991.
- [12] Ruder, S., I. Vuli, and A. Sgaard, A Survey Of Cross-lingual Word Embedding Models. 2017.
- [13] Gers, F.A., J. Schmidhuber, and F.J.N.C. Cummins, Learning to Forget: Continual Prediction with LSTM. 2000. 12(10): p. 2451-2471.
- [14] Chen, S., J. Wen, and R. Zhang, GRU-RNN Based Question Answering Over Knowledge Base. 2016.
- [15] Kameoka, H., *et al.*, ConvS2S-VC: Fully convolutional sequence-to-sequence voice conversion. 2018.
- [16] Clevert, D., T. Unterthiner, and S.J.a.L. Hochreiter, Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). 2015.
- [17] Xu, J., *et al.*, Understanding and Improving Layer Normalization. in neural information processing systems. 2019.
- [18] Dauphin, Y.N., *et al.*, Language Modeling with Gated Convolutional Networks. 2016.
- [19] Kim, J.H., *et al.*, Hadamard Product for Low-rank Bilinear Pooling. 2016.
- [20] Srivastava, N., *et al.*, Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 2014. 15: p. 1929-1958.
- [21] Vaswani, A., *et al.*, Attention is All you Need. in neural information processing systems. 2017.
- [22] Ambartsumian, A. and F. Popowich. Self-Attention: A Better Building Block for Sentiment Analysis Neural Network Classifiers. in empirical methods in natural language processing. 2018.
- [23] Kingma, D.P. and J.J.C.S. Ba, Adam: A Method for Stochastic Optimization. 2014.
- [24] Gal, Y. and Z. Ghahramani. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. in international conference on machine learning. 2016.
- [25] Zhang, G., *et al.*, Three Mechanisms of Weight Decay Regularization. 2018.
- [26] Bai, S., J.Z. Kolter, and V.J.a.L. Koltun, An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. 2018.
- [27] Chen, Y., B.J.B.M.I. Qi, and D. Making, Representation learning in intraoperative vital signs for heart failure risk prediction. 2019. 19(1): p. 1-15.