Raising High Risk-aware in Hemodynamic Treatment with Reinforcement Learning for Septic Shock Patients

Meicheng Yang[1], Runfa Li[1], Tong Hao[2], Caiyun Ma[1], Jianqing Li[1]*, Chengyu Liu[1]*

[1]State Key Laboratory of Bioelectronics, School of Instrument Science and Engineering, Southeast University, Nanjing, China
[2]School of Medicine, Southeast University, Nanjing, China

Abstract

The resuscitation of septic shock with hemodynamic support to increase the total circulating volume and cardiac output is essential in clinical rescuing. This study aimed to raise high risk-aware in hemodynamic treatment with reinforcement learning for septic shock patients instead of finding the best treatments. Retrospective data from 7792 septic shock patients (mortality of 24.7%) were used. Data were coded as multivariate discrete-time series up to 72 h with 4-h time steps. The medical treatments of interest are the sum of intravenous fluids and a maximum dose of vasopressors. State spaces are constructed using an autoencoder network. Two separate double deep Q-Networks are trained to produce the value estimates of the embedded patient states administrated given treatments to assess the risk of transitioning to poor outcomes. Results reported that when we set a threshold of -0.16 for alarming the high-risk flag, 8.3% of treatments administered to nonsurvivors were alarmed 24 h before the outcome of death with 0.7% false alarms that misclassified the patients as near death. The global and individual trajectories of clinical variables around the first raised flag indicate the method's effectiveness. This could help warn possible high-risk treatments and help clinicians pay more attention to the alarmed patients.

1. Introduction

Septic shock is a life-threatening condition that occurs when the blood pressure drops to a dangerously low level after a body-wide infection and is a leading cause of death worldwide in the intensive care unit (ICU) [1]. The shock status should be corrected as soon as possible to prevent the subsequent negative outcome [2]. The resuscitation of septic shock with hemodynamic support involves intravenous (IV) fluid infusion and the use of vasopressors, which help increase the total circulating volume and cardiac output for maintaining blood pressure [3]. Although the Surviving Sepsis Campaign guidelines recommend several goals to guide resuscitation [4], the best treatment strategy remains uncertain due to the great clinical variability in septic shock.

Reinforcement learning (RL) aims to find an optimal policy that identifies the best action for each state, similar to clinicians' goal to make therapeutic decisions to maximize patients' good outcomes [5]. Komorowski et al. [6] first proposed an RL model with a discretized state space to suggest optimal treatment of sepsis patients in ICU. Subsequently, Nature Medicine [7] reported their further work on developing a tabular Q-learning-based RL model and validated its effectiveness on an independent database. Even though directly deploying RL into clinical decision-making systems and using the output treatments from RL for clinicians would be difficult with the limitations of safety and trust [8]. Therefore, enabling whether the treatment is secure and raising awareness of its high risk that might lead to poor outcomes is important.

To address this issue, retrospective electronic health record (EHR) data from septic shock patients admitted to the ICU of the Beth Israel Medical Center (BIMC) were used. We apply the dead-end discovery method [9] to raise high risk-aware in hemodynamic treatment with reinforcement learning. Sequential EHR data are coded as multivariate discrete-time series up to 72 h with 4-h time steps. The medical treatments of interest are the sum of IV fluids and a maximum dose of vasopressors in 4-h steps. The reward (+1/-1) is defined as surviving or not. Patient states measured by vital signs and clinical laboratory values are constructed using an autoencoder network. Two separate double deep Q-Network [10] are trained to produce the value estimates of the embedded patient states administrated given treatments to evaluate the probability of transitioning to poor outcomes.

2. Methodology

2.1. Patient cohorts

Data is sourced from Medical Information Mart for Intensive Care database-IV (MIMIC-IV) [11], which

captured de-identified health information for 76,540 ICU stays admitted to the ICUs at BIMC between 2008 and 2019 in Boston, MA, USA. We include septic shock patients fulfilling the sepsis-3 criteria [1]. Sepsis is defined as suspected or documented infection plus an increase in the Sequential Organ Failure Assessment (SOFA) score of 2 points or more. The time event defined the sepsis onset is followed as [7]. After that, the time of septic shock onset is defined as the timestamp of requiring vasopressors to maintain a mean arterial pressure of 65 mmHg or greater after sepsis onset and serum lactate level greater than two mmol/L during the one day around the time.

2.2. Data extraction and pre-processing

After determining the time onset of septic shock, we extract data started 24 h prior to this timestamp up to 48 h after, resulting in a total of 72 h data as [7] did. We use 46 variables as described in Table 1 to represent the physiological state of patients. The outcome of interest is in-hospital mortality. We then pre-process the raw EHR data in the following four steps.

(1) Step 1: Aggregate the sequential data into 4-h time steps by averaging or summing when there are multiple measurements.

(2) Step 2: Remove the outliers of all the numerical features that are not clinically plausible values according to a frequency histogram method.

(3) Step 3: Impute missing data with forward-filling strategy, linear interpolation, and KNN imputation.

(4) Step 4: Standardize normally distributed data. Log-normal distributed variables are log-transformed before standardization. Binary data is centered on zero.

Table 1. Patient variables.

Category	Variables
Demographics (4)	Age, Gender, Weight, Re-admission
Vital Signs (11)	SOFA, SIRS, GCS, Heart rate, Systolic, Mean and Diastolic blood pressure, Shock index, Respiratory rate, SpO2, Temperature
Lab Values (25)	Potassium, Sodium, Chloride, Glucose, BUN, Creatinine, Magnesium, Calcium, SGOT, SGPT, Total bilirubin, Hemoglobin, WBC, Platelets, PTT, PT, INR, pH, PaO2, PaCO2, Base excess, Bicarbonate, Lactate, FiO2, PaO2/FiO2
Intake and Output Events (6)	Fluid intake and output (4 h), Total output, Cumulated Balance, Mechanical ventilation, Max vasopressor dose

2.3. Markov decision process formulation

The Markov Decision Process (MDP) provides a mathematical framework for modeling the sequential decision-making process [12]. An MDP could be used to model the patient environment and trajectories. The agent (clinician) iteratively interacts with a State (patient physiological state) by performing an Action (medical treatments), coming to the following State, and receiving a reward (survival). The MDP is defined by tuple $\{S, A, T, R, \gamma\}$, with S and A being the finite set of states and actions. $T(s', s, a)$ is the transition function defining the probability of state s_t transiting to s_{t+1}, i.e., s', when taking action a. R is the reward function and γ is the discount factor. The patient trajectory could be simulated by $\{S_t, A_t, R_t, S_{t+1}\}$. Policy $\pi(a|s)$ represents how an action is given at state s.

Given a policy $\pi(a|s)$, the cumulative reward received by the trajectory τ with a length of L is return $G(\tau) = \sum_{t=0}^{L-1} \gamma^t r(s_t, a_t, s_{t+1})$. The goal is to maximize the expected return. A state-action value function $Q^{\pi}(s, a)$ is defined as the expected total return obtained from the initial state s with the action a and then executing the strategy π. The optimal state-treatment value function is defined as $Q^*(s, a) = \max_{\pi} Q^{\pi}(s, a)$, which is the maximum expected return of all trajectories starting from (s, a).

2.3.1. State Space

We perform an autoencoder architecture with a recurrent neural network to form sequential latent state representations of patient physiological state vectors using 4-h time steps observations from septic shock patients.

2.3.2. Action space

IV fluids include bonuses and background infusions of crystalloids, colloids, and blood products, normalized by tonicity. The vasopressors including norepinephrine, epinephrine, vasopressin, dopamine and phenylephrine are converted to equivalent dose of norepinephrine by the following equation.

Norepinephrine equivalent = Norepinephrine + Epinephrine + Phenylephrine/10 + Dopamine/100 + Vasopressin × 2.5

The sum of IV fluids and maximum vasopressor in 4-h steps is divided into per-drug quartiles and includes a special case of 0 represents no drug given. This result in the following 5×5 action space (Table 2).

2.3.3. Reward formulation

To identify the high-risk treatments that might lead to unavoid negative outcomes in the following ICU stays, we use the Dead-end discovery method proposed by Fatemi [9]. This method flags bad treatments rather than finding

the best ones through estimating an optimal policy π^*. Two MDPs $\mathcal{M}_D$ and $\mathcal{M}_R$ are constructed for modeling the dead and survival processes, respectively. Discount factor γ is set as 1. Reward functions are defined as $\mathcal{M}_D$ returns -1 with any transition to a negative terminal state (and zero with all other transitions). $\mathcal{M}_R$ returns +1 with any transition to a positive terminal state (zero otherwise).

Table 2. Dose for intravenous fluids and vasopressors that comprise the action space.

Acti-on	IV fluids (mL/4 hours)		Vasopressors (mcg/kg/min)	
	Range	Median	Range	Median
1	0	0	0	0
2	0-81	35	0-0.067	0.04
3	81-253	146	0.067-0.140	0.10
4	253-664	424	0.140-0.296	0.20
5	> 664	1180	> 0.296	0.47

2.4. Model training

For the two constructed MDPs $\mathcal{M}_D$ and $\mathcal{M}_R$, Q_D^* and Q_R^* are the corresponding optimal state-action value functions respectively. As been proved in [9], Q_D^* and Q_R^* could represent how a patient transitioning to a dead state or to a survival state as a result of administrating treatment a at s. As the reward limits, $Q_D^*(s,\ a) \in [-1, 0]$ and $Q_R^*(s,\ a) \in [0, 1]$ for all states and actions. In this case, we could set a threshold λ_D to identify treatments that lead immediately to dead-ends. The dead-end could also be confirmed by assessing if Q_R^* is smaller than the threshold λ_R, i.e., an alarm would be raised if $Q_D \leq \lambda_D$ and $Q_R \leq \lambda_R$.

We divide 80% of the cohorts for training and the remaining 20% for testing. The autoencoder is trained to form the sequential observations into the 64 dimensions of the state representations. After that, two separate double-DQN neural networks [10] are used to compute Q_D and Q_R for all treatments at the output states. The double-DQN network consists of two linear layers with 64 nodes. The first linear layer is followed by ReLU activation. The second layer outputs the relevant 25 actions. Adam optimizer was used for training with a mini-batch size of 64 and a learning rate of 0.001. The training runs for 200 epochs with the final model being the best one during the optimization process.

3. Results and discussions

A total of 7792 septic shock patients are included in this study, with a mortality rate of 24.7%. Nonsurvivors have a higher SOFA score and longer ICU length of stay compared with survivors, as shown in Table 3. In addition, nonsurvivors are administered higher doses of IV fluids and vasopressors obviously by the clinicians due to their complex conditions (Figure 1).

Table 3. Cohort statistics, variables count in the median.

Patients	Num-ber	SOFA	Age (years)	ICU-stay (days)
Survivors	5869	6	68	3.6
Nonsurvivors	1923	9	69	4.5

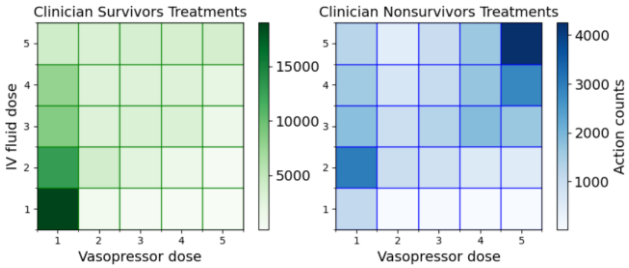

Figure 1. Clinician treatment action counts for survivors and nonsurvivors.

After optimizing the threshold λ_D and λ_R for balancing the true alerts and the false alarms on the test cohorts (1169 survivors, 390 nonsurvivors), we set λ_D= -0.16, λ_R= 0.84 for raising high-risk flags. In this case, 8.3% of treatments administered to nonsurvivors are identified as high risk 24 h prior to the outcome of death with 0.7% false alarms that misclassify the patients as near death (Table 4).

Table 4. The proportion of the identified high-risk administered treatments X hours before the terminal.

Time (h)	-72	-48	-24	-12	-8	-4
Survivors (%)	0.3	1.0	0.7	1.9	1.4	2.1
Nonsurvivor-s (%)	0.0	4.5	8.3	13.6	19.5	29.7

The trajectories (32 h, 8 steps) of clinical variables and Q values are shown around the first raised flag (Figure 3). A clear turning point of deterioration can be seen at the timestamp of the raised flag for most variables, such as Calcium, Glucose, SpO2, and Shock_Index. For nonsurvivors, the Q values got more decreased after the flag compared with survivors meaning the possible higher risks resulting from the administrated treatments, which should catch the attention.

The individual trajectory of a nonsurvivor is shown in Figure 3. A high dose of vasopressor is given at time step 9 to avoid the blood pressure dropping to a dangerously low level. However, heart rate, Calcium, SGOT, SOFA, and Shock_Index then start to deteriorate rapidly. Meanwhile, the risk begins to trigger the threshold for alarm. This result indicates the effectiveness of the method that could warn possible high-risk treatment.

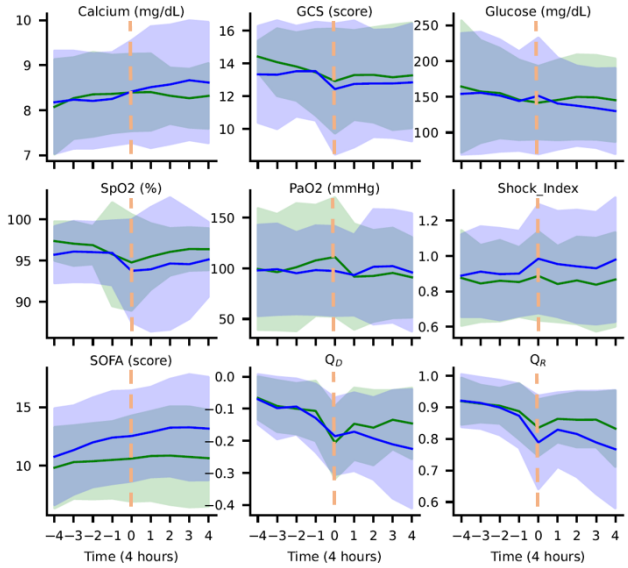

Figure 2. Average trajectories of clinical variables and Q values around the first raised high-risk flag. Blue lines represent nonsurvivors, while green lines indicate survivors. Shaded areas indicate the standard deviation.

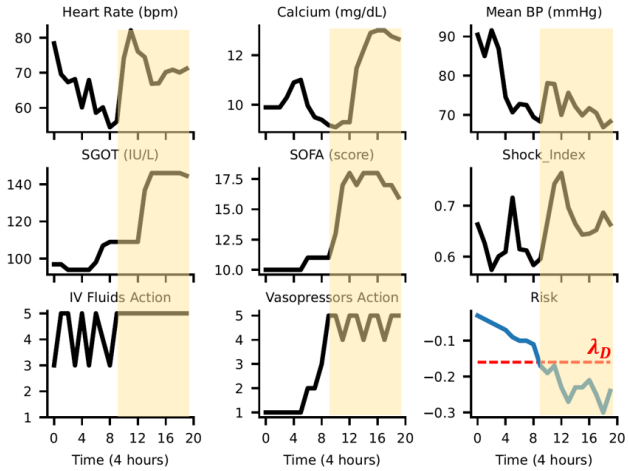

Figure 3. The individual trajectory of a nonsurvivor. Yellow shaded areas indicate the alarming high-risk periods of time.

4. Conclusions

This study aimed to raise high risk-aware in hemodynamic treatment with reinforcement learning for septic shock patients. Quantitative and qualitative results show it could help warn possible high-risk treatments and help clinicians pay more attention to the alarmed patients.

Acknowledgments

The study was partly supported by the National Key Research and Development Program of China (2019YFE0113800), the National Natural Science Foundation of China (62171123, 62071241 and 81871444), the Natural Science Foundation of Jiangsu Province (BK20190014 and BK20192004), and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX21_0088).

References

[1] M. Singer et al., "The third international consensus definitions for sepsis and septic Shock (Sepsis-3)," *JAMA*, vol. 315, no. 8, pp. 801-810, Feb. 2016.

[2] G. Hernández et al., "Effect of a resuscitation strategy targeting peripheral perfusion status vs serum lactate levels on 28-Day mortality among patients with septic shock: The ANDROMEDA-SHOCK randomized clinical trial," *JAMA*, vol. 321, no. 7, pp. 654-664, Feb. 2019.

[3] P. Ma et al., "Individualized resuscitation strategy for septic shock formalized by finite mixture modeling and dynamic treatment regimen," *Crit. Care.*, vol. 25, no. 1, p. 243, Jul. 2021.

[4] L. Evans et al., "Surviving sepsis campaign: International guidelines for management of sepsis and septic shock 2021," *Crit. Care. Med.*, vol. 49, no. 11, pp. e1063-e1143, Nov. 2021.

[5] S. Liu, K. C. See, K. Y. Ngiam, L. A. Celi, X. Sun, and M. Feng, "Reinforcement learning for clinical decision support in critical care: Comprehensive review," *J. Med. Internet. Res.*, vol. 22, no. 7, p. e18477, Jul. 2020.

[6] M. Komorowski, A. Gordon, L. A. Celi, and A. Faisal., "A Markov Decision Process to suggest optimal treatment of severe infections in intensive care," in *NeurIPS*, Dec. 2016.

[7] M. Komorowski, L. A. Celi, O. Badawi, A. C. Gordon, and A. A. Faisal, "The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care," *Nat. Med.*, vol. 24, no. 11, pp. 1716-1720, Nov. 2018.

[8] M. Lu, Z. Shahn, D. Sow, F. Velez, and L. H. Lehman, "Is deep reinforcement learning ready for practical applications in healthcare? A sensitivity analysis of Duel-DDQN for hemodynamic management in sepsis patients," *AMIA. Annu. Symp. Proc.*, vol. 2020, pp. 773-782, 2020.

[9] M. Fatemi, T. W. Killian, J. Subramanian, and M. Ghassemi, "Medical Dead-ends and learning to identify high-risk states and treatments," in *NeurIPS*, vol. 35, pp. 1-15, Dec. 2021.

[10] H. Hasselt, A. Guez, and D Silver, "Deep reinforcement larning with Double Q-learning," in *AAAI*, vol. 16, pp. 2094-2100, Feb. 2016.

[11] A. E. Johnson et al., "MIMIC-III, a freely accessible critical care database," *Sci. Data.*, vol. 3, p. 160035, May. 2016.

[12] C. C. Bennetta and K. Hauser, "Artificial intelligence framework for simulating clinical decision-making: A markov decision process approach," *Artif. Intell. Med.*, vol. 57, no. 1, pp. 9-19, Jan. 2013.

Address for correspondence:

Jianqing Li and Chengyu Liu.
Sipailou Road 2, Southeast University, Nanjing, China.
Email: ljq@seu.edu.cn and chengyu@seu.edu.cn