communications medicine

Article

https://doi.org/10.1038/s43856-024-00665-x

Deep reinforcement learning extracts the optimal sepsis treatment policy from treatment records

Yunho Choi[1], Songmi Oh[1], Jin Won Huh[2], Ho-Taek Joo[1], Hosu Lee[3], Wonsang You[1], Cheng-mok Bae[1], Jae-Hun Choi[4] & Kyung-Joong Kim[1]

Abstract

Background Sepsis is one of the most life-threatening medical conditions. Therefore, many clinical trials have been conducted to identify optimal treatment strategies for sepsis. However, finding reliable strategies remains challenging due to limited-scale clinical tests. Here we tried to extract the optimal sepsis treatment policy from accumulated treatment records.

Methods In this study, with our modified deep reinforcement learning algorithm, we stably generated a patient treatment artificial intelligence model. As training data, 16,744 distinct admissions in tertiary hospitals were used and tested with separate datasets. Model performance was tested by *t* test and visualization of estimated survival rates. We also analyze model behavior using the confusion matrix, important feature extraction by a random forest decision tree, and treatment behavior comparison to understand how our treatment model achieves high performance.

Results Here we show that our treatment model's policy achieves a significantly higher estimated survival rate (up to 10.03%). We also show that our models' vasopressor treatment was quite different from that of physicians. Here, we identify that blood urea nitrogen, age, sequential organ failure assessment score, and shock index are the most different factors in dealing with sepsis patients between our model and physicians.

Conclusions Our results demonstrate that the patient treatment model can extract potential optimal sepsis treatment policy. We also extract core information about sepsis treatment by analyzing its policy. These results may not apply directly in clinical settings because they were only tested on a database. However, they are expected to serve as important guidelines for further research.

Plain language summary

Sepsis is one of the most life-threatening medical conditions. It can be challenging to select the best treatment strategy for individual patients. We developed a computational model to identify optimal treatment strategies and applied it to a large amount of data obtained from patients with sepsis. We identified particular types of information about the patients that can be used to decide on the best medication and dose to treat sepsis. Further development of our treatment model could potentially enable it to be used to improve the survival of patients with sepsis. Also, the results we obtained could be used to improve the general guidance followed when treating people with sepsis.

Sepsis, one of the most life-threatening medical conditions induced by infection, is a leading cause of death that makes it difficult to decide the optimal treatment strategy[1–3]. Even though there were improvements in the treatment of sepsis, the mortality rate for sepsis still remains relatively high at 30%[4]. Furthermore, it could sharply worsen the mortality rate due to multiple organ failures if proper treatment is not provided[5,6].

Although numerous clinical tests have been conducted to determine the optimal sepsis treatment, deciding the optimal strategies for septic patients is still difficult[7–10]. For instance, the "sepsis campaign" has updated evidence-based sepsis treatment guidelines over 20 years and is still changing due to new evidence[10,11]. Because if different medical trial results have been reported for a similar treatment strategy due to the different patient group conditions, their findings tend to be less reliable[10,12–14]. In this case, large-scale experiments with varied cohorts are required to ensure the reliability of the evidence. However, applying unstable treatments to large-scale patient groups is unethical because

[1]School of Integrated Technology, Gwangju Institute of Science and Technology, Chemdan-gwa-gi-ro, Gwangju, 61005, Republic of Korea. [2]Pulmonary and Critical Care Medicine, Asan Medical Center, 88 Olympic-ro, Seoul, 05505, Republic of Korea. [3]Department of Control and Robot Engineering, Gyeongsang National University, Jinju-daero, Jinju-si, 52828, Republic of Korea. [4]Medical Information Lab, Electronics and Telecommunications Research Institute, 218 Gajeong-ro, Yuseong-gu, Daejeon, 34129, Republic of Korea. e-mail: kjkim@gist.ac.kr

clinical trials must be conducted safely only for an inevitably small group of patients[15].

The reinforcement learning (RL) algorithm extracts optimal treatment strategies from large-scale accumulated patient treatment records[16]. RL optimizes the sequence of actions by choosing the most valuable action to maximize the outcome, similar to how doctors optimize treatment to maximize patients' condition[17,18]. Recent studies have presented promising results showing that RL can optimize intravenous fluid (IV), vasopressor, and ventilation treatments for septic patients[3,19]. Furthermore, with deep RL (DRL) with deep representations, some studies have attempted to generate treatment regimens by referring to individual patient states without clustering[20,21].

However, the current RL-based treatment models are limited in determining optimal sepsis treatment because of their unstable results. For example, there are safety concerns regarding providing the same treatment to all patients in a clustered patient group. The clustering method depends on the start parameters and sometimes generates different results[22]. In addition, the clusters are easily influenced by outlier data[23], which is common in medical datasets. Similarly, DRL methods also have safety concerns. Due to the limited number of patient records (offline RL problem), DRL methods are often unable to predict the precise treatment[24,25]. Assigning incorrect values to the treatment is associated with suboptimal AI training[26], which can compromise patient safety. This problem may be more severe in the medical field because there is always uncertainty in the treatment process. Therefore, to avoid imprecise valuation, several studies have tried to train the treatment model by combining DRL with human knowledge, such as incorporating reward strategy or physicians' actions[27,28]. However, as our human understanding of sepsis is imperfect, these approaches may lead to suboptimal model training.

Considering the medical situation, we present a sepsis treatment AI algorithm that can train DRL to predict treatment values precisely using only treatment records to address this research gap. Our algorithm focuses on creating AI-driven treatments personalized for each patient instead of applying a uniform treatment to a group of patients. Note that our AI algorithm also considers the following characteristics of sepsis treatment: First, physicians may provide modified treatment (from the sepsis guidelines) by referring to the individual patient states to give them the best result. Therefore, we used deep representation-based treatment to individually manage all patients' health states (no clustering). Second, uncertainties are pervasive throughout the treatment stages. Therefore, we applied the highlight-RL algorithm, which highly refers to the treatment results rather than the middle of the treatment. Third, physicians should precisely judge treatment expectations without overestimation or underestimation. Therefore, we applied an over/underestimation-avoiding algorithm that automatically changes the parameter of highlight-RL. Finally, our understanding of sepsis is not yet perfect. Therefore, we trained the AI model to obtain an unbiased treatment model without additional medical information; instead, we only used previously collected treatment data. Using this algorithm, we demonstrate the potential of the DRL-based sepsis treatment model, which can extract an optimal treatment regime from accumulated treatment data.

The treatment strategies recommended by AI using a pre-collected patient test dataset (MIMIC-III) produce a 10.03% higher estimated survival rate. In addition, when we apply the same model to a different dataset (eICU), the estimated survival rate of patients increases by 9.81%. The model's suggested dose is associated with the highest estimated survival rate, indicating its optimality. Notably, this estimated survival rate is calculated using a pre-collected real dataset to make it more reliable compared with value-based survival rate estimation[3,21]. We also provide detailed analysis results and discussions to provide a proper understanding of the AI-treatment model's behavior. Through our analysis, we extract optimal treatment strategies and treatment-highly-related variables. Moreover, we reveal some interesting evidence on how the AI model improves treatment performance. Our results reveal that vasopressor dosing strategies should be revisited. Several variables such as age, BUN, SOFA score, and Shock index will influence the recommended medication dosages administered to sepsis patients. Even though this study is not a clinical test, it is worthwhile as it provides results from comparing a large dataset and extraction of treatment strategies from previously conducted treatments. Therefore, our data could be used as a reference for future studies. In addition, our methodology could be applied to other medical conditions requiring continuous care because our method finds the optimal treatment regime by optimizing the treatment sequence[16,29].

Methods

Dataset

We used the Medical Information Mart for Intensive Care database (MIMIC-III)[30] and the eICU Collaborative Research Database v2.0[31], which contain treatment data of 61,293 and 200,859 patients, respectively. The MIMIC-III data is not publicly available, but access can be requested as detailed at https://physionet.org/content/mimiciii/1.4/. The eICU data is also not publicly available, but access can be requested as detailed at https://physionet.org/content/eicu-crd/2.0/. Each dataset was collected from an intensive care unit (ICU); therefore, the datasets are very similar and have over 60 patient variables like vital signs, demographic information, health scores, and prescription information. We extracted the patients' sepsis data from the whole data by following Singer et al.'s international sepsis definition[1]. Altogether, the data of 20,927 and 14,875 patients were obtained from the MIMIC-III and eICU datasets, respectively. The MIMIC-III dataset contains more patients with sepsis and includes survival data for 90 days after discharge[30]. This variable is useful when considering death from sepsis after discharge (long-term aspect training). Therefore, 80% of MIMIC-III data ($n = 16,744$) was used for training, and 20% of MIMIC-III data ($n = 4183$) was used for testing. The eICU dataset contains multiple ICU patient treatment records (208 hospitals), suitable for cross-institution testing ($n = 14,875$).

Statistics and reproducibility

To train the patient treatment model, we preprocessed the defined sepsis patients' data to make it suitable for model training. The followings are structured processes:

Initial data preparation. Because sepsis treatment is highly affected at the initial state of sepsis, the first 80 h of data were used for the strategy training sequence. To ensure that each data contains reasonable updates on patients' states, we separated the collected data by 4 h.

Data imputation. Due to various reasons such as long measurement intervals, human omission, or unmeasurable data, medical data often contains missing data. In this study, data imputation was conducted to facilitate smooth training of the treatment model. Data imputation was conducted following these steps, in the listed order: First, we filled in missing values with the sample and hold method. Here, if a patient has missing data at a certain state, but if there are measured values before that state, those values are used to fill in the missing data. Only when no previous data are available do we substitute missing values with future data by assuming that the previous state has been maintained. Second, we filled the missing values with calculated values from other measurements. If some data is still missing but can be calculated using the measured values of other variables (such as SOFA score), then the missing value is filled in through calculation.

We employed a data-driven approach to address missing values not handled by the imputation steps. In the case of the MIMIC-III dataset, we used the k-nearest neighbor imputation method, as described by ref. 32. However, owing to the extensive amount of missing data, including whole columns, within the eICU dataset collected from multiple institutions, applying the k-nearest neighbor method presented challenges. Under these circumstances, we chose median imputation as the most conservative approach. While potentially overly cautious, this method ensures the replacement of missing values with the general median, mitigating the risk

of inaccurate predictions based on incomplete data. This decision is supported by previous research, which suggests that median imputation performs comparably to more complex methods in specific applications, such as cardiovascular prediction models or metabolomics data[33,34].

Finally, 47 variables reflective of a patient's health status were selected. All variables used are listed in Supplementary Table 1. We referred to a previous study's open-source code at this stage[3].

Outlier removal. After all data were set, we removed outlier data. To utilize data from as many people as possible, medically nonsensical and only clear outliers were removed. Detailed criteria for outliers are provided in Supplementary Table 2.

Normalize. The data were normalized over three stages: z-score normalization, data clipping, and min-max normalization. First, to align the ranges of measurements by column, we converted each column's measurements to z-scores through z-score normalization. However, the normalized z-score includes extreme data; therefore, DRL training might be poor if the data distribution is not uniform. Therefore, we clipped data with a z-score exceeding the range [−3.3,3.3] (over 99.94% chance of abnormal data. We then conducted min-max normalization (0 ~ 1) for efficient DRL training. Both MIMIC-III and eICU were normalized in the same manner. However, when performing z-scoring on the eICU data, the mean and variance of the MIMIC-III data were used to ensure that the z-scoring was performed on the same basis as the MIMIC-III.

Treatment model implementation

We implemented the treatment model using preprocessed data. The detailed sequence of this implementation is shown in Fig. 1. After preprocessing, we define a state s, an action a, a next state s', and a reward r. These variables are used for DRL training to decide optimal action (a) by referring to the current state (s) for reaching a more valuable next state (s') and getting rewards (r). Each variable was defined as follows.

State (s) indicates the current patient condition. Demographic information, electronic health records, and culture data, and other information are necessary to determine treatment. Action (a) indicates the physician's treatment record. The IV and vasopressor injection levels over five stages (25 discrete actions in total, Supplementary Table 3). Next State (s') indicates the patient's condition after 4 h. When a physician treats a patient in state s, the changes are recorded 4 h later. Reward (r) indicates the reward for action a when the patient's condition changes from s to s'. Following the approach adopted in DRL models addressing other complex problems like Go and video games, we incorporated a straightforward and intuitive reward function to prevent the model from converging on suboptimal solutions[35–37]. The reward when the patient survived for 90 days was +1, and the reward when the patient died, it was −1; the reward for all actions during treatment was 0.

We trained the DRL model using the above variables with 80% of MIMIC-III data. The DRL model estimates 25 possible action values (output) from 47 patient states (input). In the training, the DRL model updates its action value estimator to reduce the error between the expected action value (Q) and the estimated action value of physicians' actions. The expected action value can be calculated by adding the actual reward of physicians' actions and the discounted (γ) estimated best action value for the next state from another estimator (Q_{target}). Q_{target} acts like a teacher for the training DRL model, and it is updated by the training DRL model in a specific period (in our case, 1 iteration). This loss system can be expressed as follows:

$$Loss = MSE(Q(t), r + \gamma \times Q_{target}(t+1)) \quad (1)$$

The training lasted until it converged, and in our case, it was 200 iterations of the whole dataset training. To confirm the robustness of the DRL model training process, we conducted 500 training sessions repeatedly. We then tested each model using the remaining 20% of the MIMIC-III test data. We calculated each model's estimated patient survival rate at the test stage based on AI suggestions. The estimated survival rate was calculated by

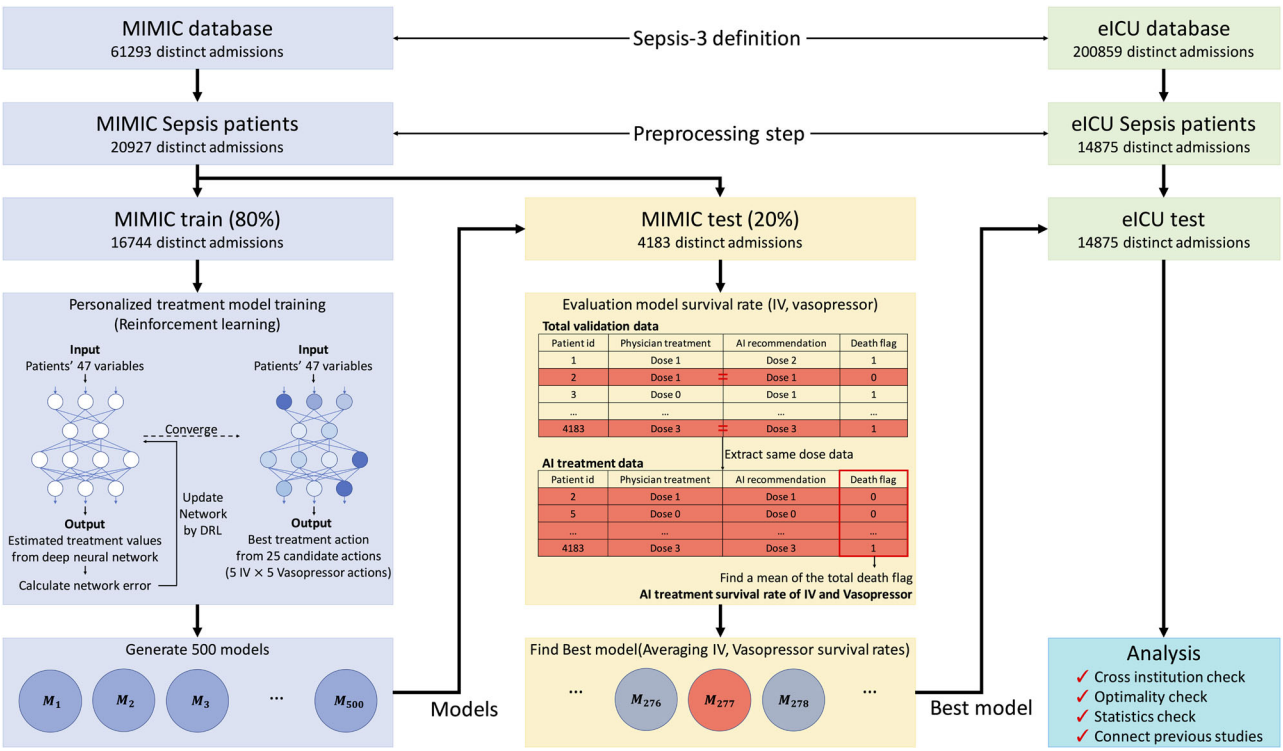

Fig. 1 | Treatment model training and testing sequences. The overall process of training, testing, and analyzing the treatment model. Sepsis patients are extracted from the MIMIC and eICU datasets, which are then used to train, select, test, and analyze the reinforcement learning-based treatment model.

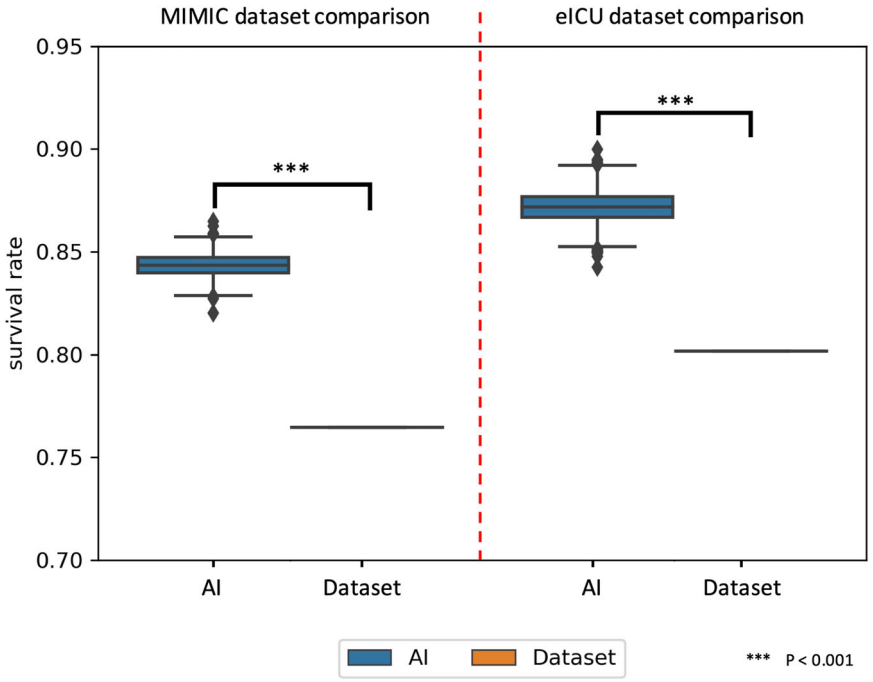

Fig. 2 | Patients' estimated survival rates from AI and whole dataset treatment records. AI test results at each MIMIC-III and eICU dataset and overall survival rate of MIMIC-III and eICU data records. Error bars indicate outlier bounds of each model testing. 500 samples of the treatment model have been compared with the total MIMIC dataset patients' survival rate and 2000 samples of treatment model results have been compared with the total eICU sepsis patients' survival rate.

averaging the survival rates of patients who are prescribed the same dose of medication as suggested by AI from the total test dataset. Note that this estimated survival rate is derived from a data-driven approach, which does not accurately represent the survival rate verified through a human-in-the-loop approach. Therefore, it does not reflect the "true" survival rate that might be observed if our treatment model were implemented in a clinical trial. However, consistent with previous studies, this estimated survival rate serves as a valuable indicator for indirectly assessing the potential performance of artificial intelligence treatment models in patient care[3,19,21]. Finally, the best model with the highest estimated survival rate among the 500 candidates was tested with another hospital dataset (eICU). Here, we examined whether it could maintain its optimal treatment performance despite using data from different institutions'. Furthermore, we analyzed (with statistics and literature) the training stages and results to understand AI behavior. In addition, with in-depth discussion, we investigated how the AI model works effectively and which variables are highly related to decision-making when treating patients with sepsis.

Algorithm

In this study, we adopted the Double-Dueling-Deep-Q-Network (DDDQN) proposed by ref. 38 as the base training model owing to its exceptional suitability for the patient care model. DDDQN is an advanced version of the Deep-Q-Network introduced by ref. 39. Its incorporation of dueling and double structures distinguishes it, making RL training more similar to real clinical scenarios (Detailed explanations of DDDQN described in Supplementary Note 1). The dueling structure enables the simultaneous estimation of state values and action advantages, a departure from traditional RL algorithms that primarily focus on action values. This capability allows DDDQN to evaluate the patient's current health state and the potential impact of various treatment actions.

Moreover, DDDQN addresses the common issue of action overestimation encountered in standard Q-learning techniques. Typically, when a single network uses the q-value to select the best action and subsequently updates the q-value based on it, there is a risk of reinforcing a specific behavior, potentially leading to overestimating its value. To mitigate this, the double structure of DDDQN employs one network to select the best action and another to update the q-value, as outlined by ref. 26. This separation is critical for ensuring a fair assessment of treatment options without overvaluing any particular action.

However, DDDQN is mainly used in online RL. However, in the healthcare domain, it differs from the existing online RL in that the agent must be trained from the collected data sets (offline RL). Therefore, we added two additional algorithms to make DDDQN work well in offline RL training.

First, we applied the Highlight DDDQN. Highlight DDDQN, add a variable termed "Highlight coefficient (h)" and apply it to the DDDQN loss function. Q and Q_{target} are both estimated values and each estimator is updated during the training process. However, because uncertainties are always present in the treatment stages, estimators can be trained in unwanted directions. To prevent this unwanted direction training, we intentionally suppressed estimator updates to make it slowly improve. At the same time, it strongly updates as much as h when the actual reward income. Therefore, the reward propagates more effectively from the final stage of treatment to the uncertain treatment stages. In conclusion, our estimator learns from reliable data quickly, while it learns from less reliable data slowly in the middle of treatment when the reward is 0 (Detailed experiment results of highlight coefficient described in Supplementary Fig. 1). This can be expressed as follows:

$$Loss = MSE\left(\frac{1}{h}Q(t), r + \frac{\gamma}{h} \times Q_{target}(t+1)\right) \quad (2)$$

Second, we applied an over/under estimation monitoring algorithm. Highlight coefficients can suppress unstable updates by as much as h to make them never greater than the expected maximum reward (hr). If $\frac{Q}{h}$ is getting out of the final reward r range ($+1 \sim -1$) in some cases, which is a clear overestimation signal, we immediately stop training and update h as a greater value. Then, we start training again with updated h. Greater h suppresses unstable network updates harder, so overestimation disappears because h getting bigger. In our case, we set h as 100.

Ethics

MIMIC-III data collection and release were approved by the institutional review boards of Beth Israel Deaconess Medical Center (Boston, MA) and the Massachusetts Institute of Technology (Cambridge, MA) [Johnson et al (2016)]. For the eICU data, institutional review board approval was exempted due to the retrospective design, lack of direct patient intervention, and the security schema, for which the re-identification risk was certified as

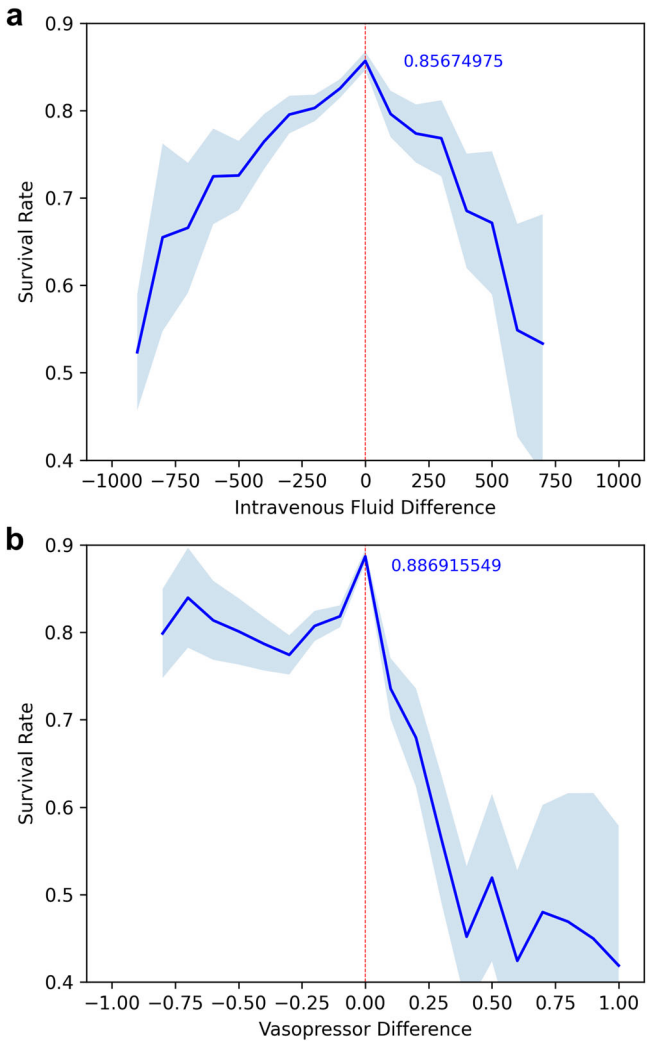

Fig. 3 | Estimated changes to survival rates at different doses. In each figure panel, the red line shows the AI model's recommended IV (**a**) and vasopressor (**b**) doses. The red line's left side indicates when the prescribed dose was lower than the AI model's recommended dose. The right side of the red line denotes when the prescribed dose was greater than that suggested by the AI model. The blue line indicates the average survival rate value from the total experiment results. The blue shade indicates the standard deviation of 2000 resampling experiment results. A total of 14875 patient treatment records have been used for testing.

meeting safe harbor standards by an independent privacy expert (Privacert, Cambridge, MA) (Health Insurance Portability and Accountability Act Certification no. 1031219-2) [Pollard et al (2018b)]. In both projects, the requirement for individual patient consent was waived because the projects did not impact clinical care, and all protected health information was unidentified [Johnson et al. (2016); Pollard et al. (2018a)]. All authors of the present study completed registration at PhysioNet (data provider; https://physionet.org), including completion of the 'research with human subjects' course, and signed a data use agreement mandating responsible use of the data. This study has been waived from IRB approval by the Ethical Committee of the Gwangju Institute of Science and Technology with the following statements. No issues related to safety, protection of personal information, or compliance with bioethics were identified, and it was determined that the exemption from review was possible because the academic and ethical legitimacy was sufficient.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Model performance

We used a pre-collected clinical dataset (MIMIC-III) containing various prescribed doses and their results to evaluate the AI model results. When the model generated prescription suggestions, we collected data on patients who received the same drug dosage from the physicians in the dataset. We classified the collected patient dataset as the AI-treatment group and calculated their estimated survival rate. To confirm the robustness of the model performance, we conducted 500 individual model training with 80% MIMIC-III data and tested them with 20% unseen MIMIC-III data. The test result illustrated in the MIMIC test is shown in Fig. 2. The AI-MIMIC test denotes a box plot of the MIMIC-III data test result of the 500 trained models. The median estimated survival rate was 84.35%, whereas the highest and lowest estimated survival rates were 86.48% and 82.03%, respectively, with an average of 84.32 ± 0.59%. The average estimated survival rate of patients with sepsis in the MIMIC-III dataset was 76.45%. The t-tests revealed statistically significant differences (P value < 0.001, t = 296.251, Supplementary Note 2). Here, we observed that patients who took the same amount of medication as the AI recommendation survived by up to 10.03%. Of note, every 500-test result improved the current treatment regime's result in the MIMIC-III dataset, ensuring the safety of the AI-treatment model.

Additionally, cross-institution tests were conducted using the eICU dataset. For these tests, we selected the best model that performed in the MIMIC test. Using this model, we executed 2000 evaluations using the same survival rate estimation process as the MIMIC test. At each test, 50% of the randomly resampled patient data in the eICU dataset was used for evaluation. The cross-institution test results of the eICU test are illustrated in Fig. 2. The median estimated survival rate was 87.18%, whereas the highest and lowest estimated survival rates were 89.98% and 84.25%, respectively, with an average value of 87.18 ± 0.75%. The average estimated survival rate of the patients with sepsis included in the eICU dataset was 80.17%. The t-tests revealed statistically significant differences (P value < 0.001, t = 416.257, Supplementary Note 2) even when using other hospital datasets. Also we confirmed that these increased survival rates were also observed for underrepresented races in the data. Detailed treatment results in underrepresented groups can be found in Supplementary Table 4.

Furthermore, we conducted an optimality check on the eICU dataset to determine whether the AI-treatment model could provide an optimal treatment regime for cross-institution patients. In this test, we investigated patients' estimated survival rates using the difference between the actual prescribed doses and the AI-suggested prescribed dose. We conducted 2000 evaluations with 50% resampled data and then averaged it. Since there are not enough patient samples to continuously investigate different survival rates associated with the dosages, we grouped patients on similar mean dosages (split with discrete value; IV range: 100ml/4 hour, vasopressor range: 0.1 mcg/kg/min). We then calculated each patient group's survival rate. Interestingly, the highest estimated survival rate was noted in patients prescribed the same dose as the AI model's suggestion in both IV (85.68%) and vasopressor (88.70%) treatment scenarios (Fig. 3). In both cases, the estimated survival rate decreased when the difference between the AI-treatment model's suggested dose and actually prescribed dose increases. These results are not biased on the small-sized patient groups, as each group comprised sufficient patients except IV over 400ml/4 hours and vasopressor over 0.5mcg/kg/min. Detailed patient distribution can be found in Supplementary Fig. 2.

Model analysis

To confirm whether the AI model learns the treatment strategy properly, we visualized the AI model's treatment valuation. We investigated all estimated treatment values for all patients and every sequence of their treatments. Figure 4 shows how AI improves treatment valuation during training. Each dot represents the estimated treatment value (q value in the DRL aspect) for every treatment on patients. The blue dots denote the surviving patient samples, and the red dots denote the dead patient samples. We assume that the treatments for surviving patients are more valuable than those for dead

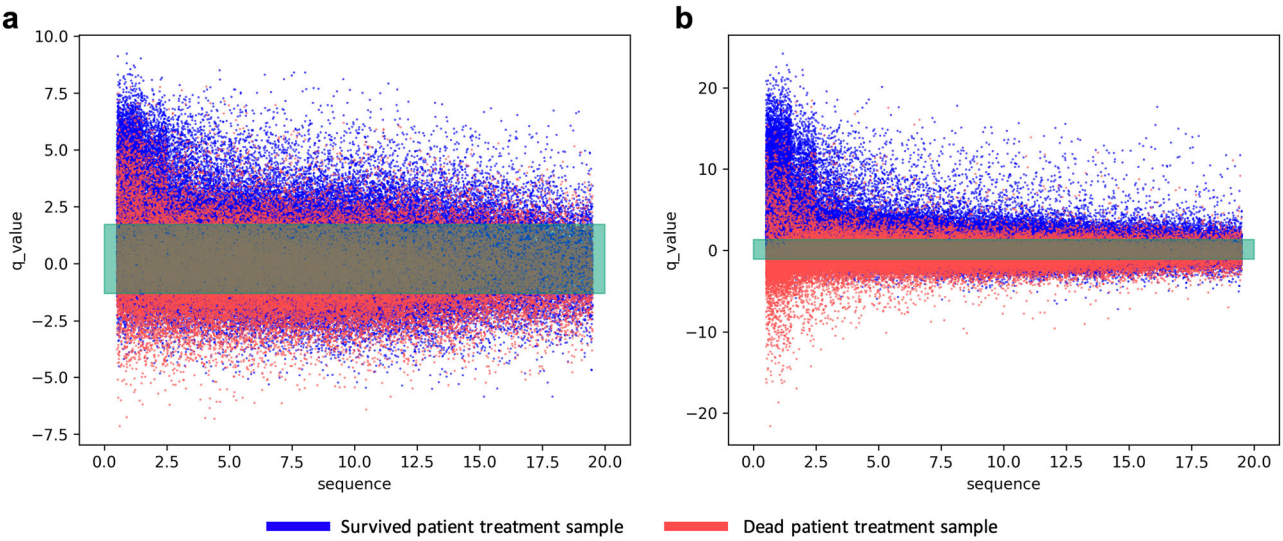

Fig. 4 | Individual treatment value visualization when model trained. Visualized dot-graph of individual treatment value by AI estimation, with the last treatment (1) to the first treatment (19). As there was so much data in each discrete sequence, we uniformly spread the data in each sequence (1~19) in a range *sequence* ± 0.5 to clearly see all data. **a** treatment value visualization when the treatment model is trained in one cycle. **b** treatment value visualization when the treatment model is trained in 200 cycles. Blue points denote the surviving patients' treatment records, and red points denote the dead patient's treatment records. The green area is the not separable (valuable or not valuable treatment) area where the individual estimated treatment values of survival/dead patients data are strongly mixed. A total of 204,800 individual treatment records have been used to compare; those data contain 153,785 survival patients' treatment records and 51,015 dead patients' treatment records.

patients. According to this assumption, well-trained AI models should clearly separate the blue and red dots to specify the valuable and nonvaluable individual treatment values. Therefore, we set the 'not separable area' as the area where the values are lower than the top 10% of the dead patient treatment value and higher than the bottom 10% of the survived patient treatment value. In this area, the treatment values of the surviving and deceased patients are strongly mixed.

In contrast, outside this area, there is more separated data (surviving or dead patients). As a result, different from the initial step of the training, each dot color is separated more clearly when the training is completed. At the initial stage, 70.58% of the data were found in the 'not separable area,' which means that the AI model could specify only 29.42% of the clear valuable/nonvaluable individual treatments. However, at the end of the AI model training, only 57.63% of the data were found in the green area. Notably, the AI model could separate 12.95% of the more valuable/nonvaluable data.

We used a heatmap graph to investigate how AI changed treatment behavior to improve patient treatment outcomes. Figure 5 shows the results of the individual treatment analysis for each IV and vasopressor dose. The AI's treatment agreed to some extent with the physicians' prescription regarding IV dosage (36% of individual treatments agreed); a little more or less fluid was administered (34% has 1 level of difference). The activated diagonal line indicates a high level of agreement between the AI model and the physician's response to the same patient condition in individual treatment decisions. In contrast, regarding the vasopressor dosage, the AI model showed a high agreement (67% of data) when physicians decided on minimum prescriptions. At the same time, they are less likely to agree on the other levels of prescriptions (20.5% agree, 39.5% 1 level difference on average). Notably, when the physician prescribed level 1 or 2 vasopressors, the agreements were lower than those for other levels of treatments.

Conversely, the AI model suggested a high-dose vasopressor to be prescribed to many patients who have been prescribed level 0 vasopressors by a physician. Even though their agreement rates were low (6% ~ 14%), considering the total number of treatments (217,093 cases), many cases exist. The AI-treatment model prescribes a level 0 vasopressor dose in only 59.87% of the treatment cases. This means that the AI-treatment model more actively prescribes vasopressors, compared with treatment cases in MIMIC-III (84.42% of prescriptions were level 0 in the MIMIC-III dataset).

Figure 6 shows the variables that affect the AI model's decision-making. A decision tree classifier has been used to calculate the importance of each feature of patients' states. Interestingly, although age and blood urea nitrogen (BUN, the indicator for renal function) levels are less referred to in clinical practice (compared to other vital signs), the AI model highly refers to these values.

We investigate which important variables, as shown in Fig. 6, have different tendencies than the physician's. Figure 7 shows several examples of variables with different tendencies. We investigated the average dosage of each medication (IV, vasopressor) according to the value of each variable. Age and BUN levels do not meaningful affect the average dosage in IV cases. However, AI more actively suggested vasopressor dosing according to age or BUN level increase in the vasopressor case. On the other hand, physicians tend to refer more highly to the SOFA or shock index scores than the AI model. As these two values increase, physicians quickly increase the vasopressor and IV dosage compared with AI.

Discussion

We created an optimal AI sepsis treatment model by extracting optimal treatment regimens for IV and vasopressor dosages from the accumulated treatment data. Our study is important because large treatment records were used to obtain optimal treatment strategies; thus, a more reliable treatment regime was extracted.

Remarkably, our analysis revealed that the estimated survival rate for patients who received the same dosage as recommended by the AI exceeded the overall survival rate by up to 10.03% and 9.81% in the MIMIC-III and eICU test data, respectively (Fig. 2). This result indicates that the AI-treatment model could support physicians' decision making through high-quality patient analysis and dosing suggestions. Furthermore, the AI-treatment model shows their prescription optimality in the other hospital dataset (Fig. 3). This is also a remarkable result because the AI-treatment model could maintain its suggestion quality, irrespective of the institution. In other words, these results indicate that our AI model could be used in several institutions even if their data were not included in the training dataset. Note that these results are reliable because all estimated survival rates are calculated based on actual patient survival rates.

Moreover, our AI treatment model could provide individualized treatment. As shown in Fig. 4, our model was trained to evaluate every

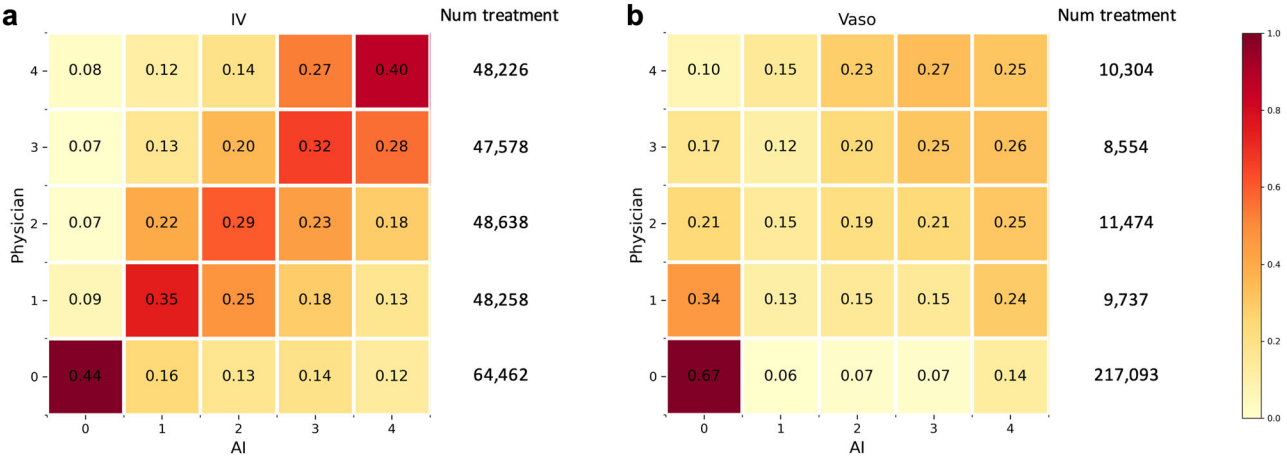

Fig. 5 | AI and physician treatment tendency comparison. AI-treatment Model Action (horizontal) Agreement Rate (sum = 1 for horizontal) per Physician Treatment Action (vertical). Each heatmap square color and number indicate the action agreement rate. A darker color denotes a high agreement rate. The num treatment denotes the number of all treatments that received level X (0 ~ 4) from the physicians. **a** shows action comparison between AI and physicians in IV dosage. **b** shows action comparison between AI and physicians in vasopressor dosage.

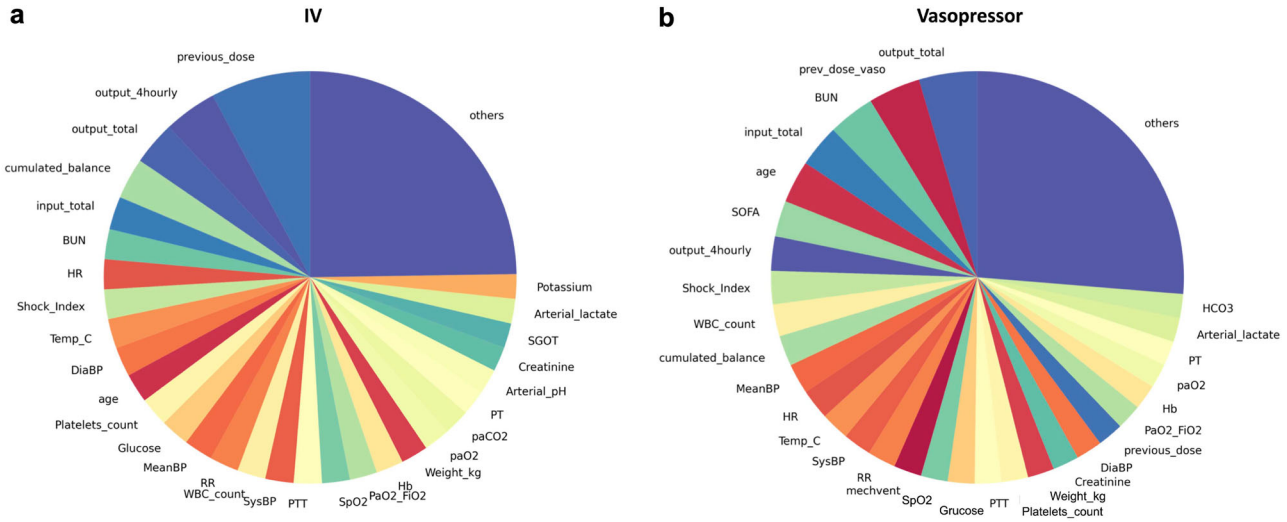

Fig. 6 | Important feature pie chart. This chart indicates the importance of features when AI makes decisions. **a** shows important features when AI decides on IV dosage. **b** shows important features when AI decides on vasopressor dosage. Each importance has been extracted using a decision tree when classifying the AI model's suggested actions by features. Each abbreviation means as following: previous IV dose(previous_dose); previous urine output(output_4hourly); total urine(output_total); total IV input dose(input_total); blood urea nitrogen(BUN); heart rate(HR); body temperature(temp_c); diastolic blood pressure(diaBP); mean blood pressure(meanBP); respiratory rate(RR); white blood cell count(WBC_count); systolic blood pressure(sysBP); partial thromboplastic time test(PTT); oxygen saturation(SpO2); PaO2/FiO2 rate(pao2_fio2); hemoglobin(Hb); oxygen pressure in arterial(paO2); carbon dioxide pressure in arterial(paCO2); prothrombin time test(PT); serum glutamic oxaloacetic transaminase(SGOT); previous vasopressor dose(pre_dose_vaso); Sequential Organ Failure Assessment Score(SOFA); mechanical ventilation flag(mechvent); bicarbonate(HCO3).

treatment value individually by referring to each patient's state. Therefore, without using any additional machine learning technique, such as clustering, it could evaluate every possible treatment for individual patients. Therefore, the AI-treatment model could immediately provide the most effective dosage by being placed at the patient's side while continuously referring to the patient's health status to increase patients survival rate[40]. This result is remarkable because one of the essential things to managing sepsis is continuous monitoring and providing immediate response to changes in the patient's status[10]. A previous study also showed that the AI model could provide more immediate and frequent dosing changes to reduce physicians' heavy workload and decrease patient mortality[19].

The AI-treatment model's notable results could be explained by AI's changed policy, as illustrated in Fig. 5. Figure IV shows a strong relationship between the AI model and physicians' prescriptions. Approximately 70% of the AI model's IV prescription doses matched or were within one level higher or lower than the doses prescribed by physicians. This suggests that the physician and AI model may have a high level of agreement in their individual treatment responses to the same patient condition. Furthermore, in the feature importance test, the AI model considered input/output, cumulated balance, and diaBP as important features for IV dose decision-making (Fig. 6). This is also interesting because even though we did not provide any medical information, the AI-treatment model extracted the core criteria from the patient's treatment records.

However, there was a notable difference in the vasopressor treatment dose between AI and physicians' actions. When physicians prescribe high or low doses of vasopressors, the AI-treatment model tends to agree with them. However, when they prescribe medium doses of a drug, the AI-treatment model recommends high or low doses, not medium doses. This finding is

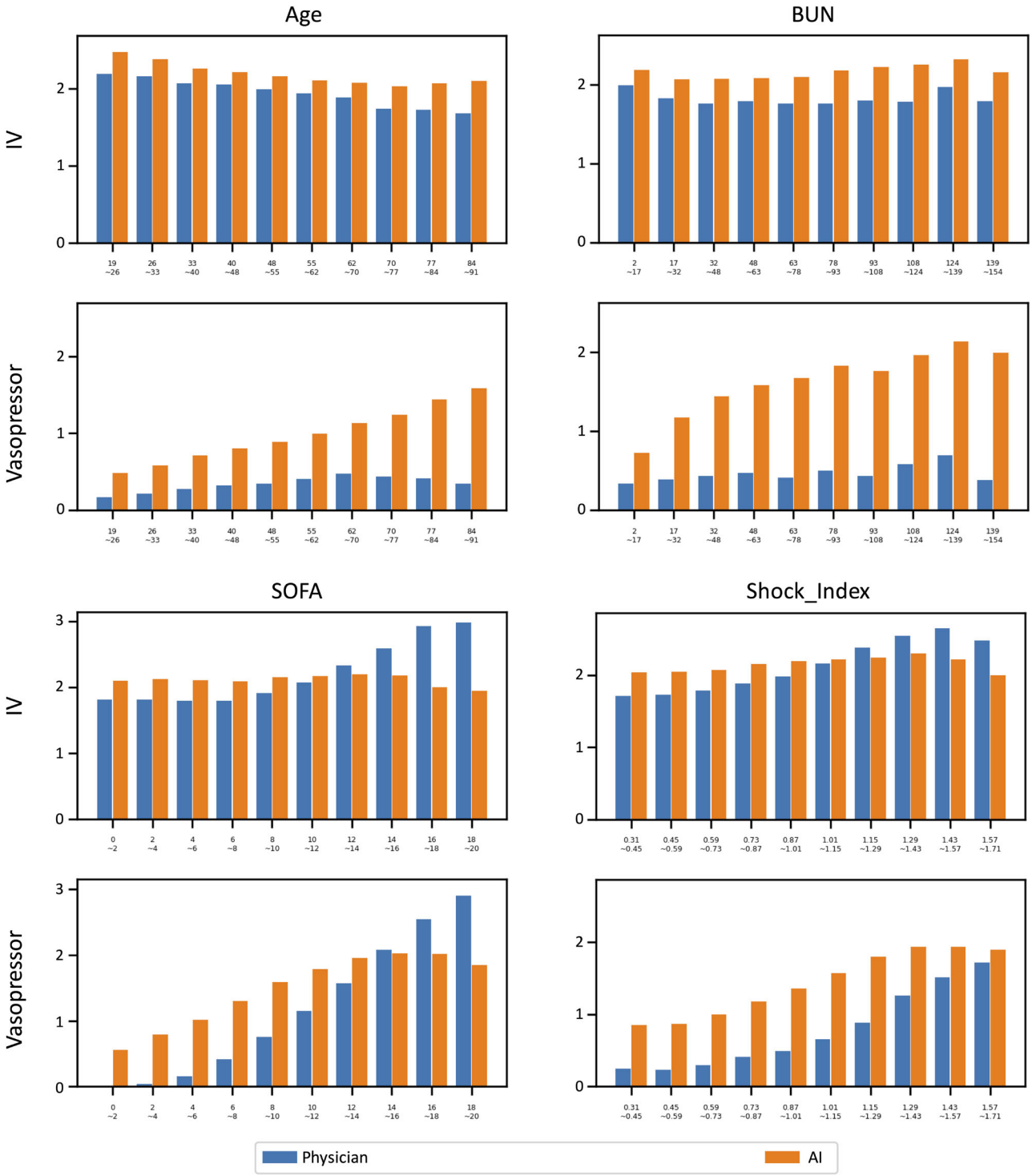

Fig. 7 | AI and physician treatment tendency comparison on specific features. The horizontal line denotes the column values, whereas the Vertical line denotes the averaged AI model's prescription level (0 ~ 4;5 levels). Age, BUN, SOFA score, and Shock index were selected to compare. Each abbreviation means the following: blood urea nitrogen(BUN) and sequential Organ Failure Assessment Score (SOFA).

supported by a recent animal study reporting that vasopressors may suppress immune responses[41]. Therefore, we considered that the AI-treatment model tried to avoid a moderate and lasting range of dosages to maintain the functioning of the human immune system. The AI-treatment model recommended short-term, intensive prescriptions when a vasopressor is needed. This interpretation aligns with a previous study reporting that norepinephrine may aggravate sepsis-induced immunoparalysis[42]. Although it is not strictly proven in a clinical test, this result found an interesting possible guideline to further study vasopressor dosages. Note that the AI model's vasopressor strategies show higher estimated patient survival rates (Fig. 3).

It is also interesting to note that the AI-treatment model considers BUN and age as the most important criteria (Fig. 6). When the patients' BUN and age increase, the AI model increases the vasopressor dose even when the physician does not increase the dose. Since BUN is strongly related to renal function, this result is reasonable because the kidney is

the second most frequently failing organ (15%) in patients with sepsis[6]. Additionally, recent research reveals that a high BUN level is related to 30-day mortality in patients with sepsis[43]. Since proper vasopressor treatment also affects renal function, proper vasopressor prescription may have beneficial effects on obtaining renal function recovery[44]. Several articles reported that treatment for sepsis should be applied according to patients' age[45,46]. More specifically, older patients more frequently develop acute renal failure than younger people[47]. Thus, some studies have reported that a high mean arterial pressure (MAP) goal would be useful in improving renal function[48,49].

Moreover, other researchers argue that a higher MAP goal is potentially beneficial in older patients because of their other comorbidities[50], which is consistent with our study findings. Although sepsis treatment protocols remain consistent, clinical practice advises a more cautious approach to IV fluid administration in elderly patients with elevated BUN, owing to the increased risk of pulmonary edema resulting from volume overload. The use of ventilators owing to pulmonary edema in these patients may exacerbate the risk of increased morbidity and mortality, requiring a conservative approach to IV fluid management. Similarly, vasopressors, while crucial for managing hypotension, need to be administered with caution to mitigate potential cardiovascular complications and avoid triggering arrhythmias. However, our AI model seems to judge that these patients are in more dangerous conditions and recommend a more aggressive management strategy involving IV fluids and vasopressors for these patient groups (Fig. 7).

Interestingly, AI demonstrates more sensitivity to severity scores, such as the SOFA score and shock index, compared to physicians. Instead, AI tends to initiate IV fluid and vasopressor administration before serious increases in SOFA and shock index scores, and less so thereafter. In clinical practice, SOFA score and shock index, which require manual calculation, are essential tools for assessing disease severity and monitoring patient progress, as indicated by refs. 51–53. However, integrating these metrics into real-time treatment decisions poses challenges. Interventions such as IV fluids and vasopressors are typically initiated in response to hypotension, often solely based on blood pressure readings. Yet, as depicted in Fig. 7, initiating treatment based on the SOFA score from the outset, rather than reacting solely to changes in blood pressure, appears to reduce the required doses of IV fluids and vasopressors in patients with elevated SOFA scores. Furthermore, in cases of an elevated shock index, a more aggressive vasopressor regimen may be warranted. Conversely, considering previous clinical studies indicating that intensive vasopressor treatment in patients with multiple organ failure may reduce survival rates, reducing vasopressor dosing in patients with elevated SOFA scores may be imperative, as highlighted by ref. 54.

However, our study has several limitations. First, this research version primarily focused on expanding patient states to facilitate individualized treatment in limited aspects, specifically dosage decisions. Consequently, our model requires further expansion of actions, encompassing considerations such as medication type, concentration, injection speed, or continuous dosage adjustments. Because our model does not consider the side effects on patients, taking these into account is also a part that physicians must handle. In addition, because our study is a data-driven research, our model is not yet ready for direct clinical use. Clinical work is far more complex and diverse, and always unexpected situations exist. Since medical decision-making is closely related to patient survival, to apply research in the current phase of data-driven approaches to real-world situations, additional in-depth research will be necessary regarding the conclusions drawn from this study.

Although there are several limitations, this study is still meaningful. It provides insight into the sepsis treatment strategy by analyzing the AI-treatment model's treatment regime. We believe it could be a relatively safe reference for future clinical trials because all these findings extracted large-sized treatment records. It is more remarkable that our methodology could be applied to other conditions that require continuous treatment. In this study, we only enrolled patients with sepsis. However, because DRL is a sequence-optimizing algorithm, our methodology could be applied to other conditions, such as diabetes and cancer, all of which necessitate continuous care records[16,29]. These findings could be extracted by accumulating treatment records without other medical information input (using naturally generated patients' treatment data). Therefore, we can extract the potential optimal treatment using only data without excessive human effort.

Data availability

The MIMIC-III data is not publicly available, but access can be requested as detailed at https://physionet.org/content/mimiciii/1.4/. The eICU data is also not publicly available, but access can be requested as detailed at https://physionet.org/content/eicu-crd/2.0/. The source data for Figs. 2–7 can be found in https://zenodo.org/records/13842300(https://doi.org/10.5281/zenodo.13842300) Sepsis_treatment_model/experimental_results folder[55].

Code availability

The underlying code for this study is available in the Zenodo can be accessed via the following link: https://zenodo.org/records/13842300(https://doi.org/10.5281/zenodo.13842300)[55].

Received: 2 June 2023; Accepted: 5 November 2024;
Published online: 22 November 2024

References

1. Singer, M. et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA* **315**, 801–810 (2016).
2. Vincent, J.-L. et al. Assessment of the worldwide burden of critical illness: the intensive care over nations (icon) audit. *lancet Respir. Med.* **2**, 380–386 (2014).
3. Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C. & Faisal, A. A. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat. Med.* **24**, 1716–1720 (2018).
4. Bösch, F., Angele, M. K. & Chaudry, I. H. Gender differences in trauma, shock and sepsis. *Mil. Med. Res.* **5**, 1–10 (2018).
5. Russell, J. A. Management of sepsis. *N. Engl. J. Med.* **355**, 1699–1713 (2006).
6. Martin, G. S., Mannino, D. M., Eaton, S. & Moss, M. The epidemiology of sepsis in the united states from 1979 through 2000. *N. Engl. J. Med.* **348**, 1546–1554 (2003).
7. Dellinger, R. P. et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Intensive Care Med.* **34**, 17–60 (2008).
8. Dellinger, R. P. et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock, 2012. *Intensive Care Med.* **39**, 165–228 (2013).
9. Rhodes, A. et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive Care Med.* **43**, 304–377 (2017).
10. Evans, L. et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med.* **47**, 1181–1247 (2021).
11. Marshall, J. C., Dellinger, R. P. & Levy, M. The surviving sepsis campaign: a history and a perspective. *Surg. Infect.* **11**, 275–281 (2010).
12. Cohen, J. et al. Sepsis: a roadmap for future research. *Lancet Infect. Dis.* **15**, 581–614 (2015).
13. Rivers, E. et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N. Engl. J. Med.* **345**, 1368–1377 (2001).
14. Investigators, P. A randomized trial of protocol-based care for early septic shock. *N. Engl. J. Med.* **370**, 1683–1693 (2014).
15. Freedman, B. *Equipoise and the ethics of clinical research* (Routledge, 2017).

16. Yu, C., Liu, J. & Nemati, S. Reinforcement learning in healthcare: a survey. *ACM Computing Surveys* (Association for Computing Machinery, 2021).
17. Gottesman, O. et al. Guidelines for reinforcement learning in healthcare. *Nat. Med.* **25**, 16–18 (2019).
18. Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction* (MIT press, 2018).
19. Peine, A. et al. Development and validation of a reinforcement learning algorithm to dynamically optimize mechanical ventilation in critical care. *NPJ Digit. Med.* **4**, 1–12 (2021).
20. Yu, C., Ren, G. & Liu, J. Deep inverse reinforcement learning for sepsis treatment. In: *2019 IEEE ICHI*, 1–3 (2019).
21. Wu, X., Li, R., He, Z., Yu, T. & Cheng, C. A value-based deep reinforcement learning model with human expertise in optimal treatment of sepsis. *NPJ Digit. Med.* **6**, 15 (2023).
22. Ayed, A. B., Halima, M. B. & Alimi, A. M. Survey on clustering methods: towards fuzzy clustering for big data. In: *2014 6th International conference of soft computing and pattern recognition (SoCPaR)*, 331–336 (IEEE, 2014).
23. Rai, P. & Singh, S. A survey of clustering techniques. *Int. J. Comput. Appl.* **7**, 1–5 (2010).
24. Fujimoto, S., Meger, D. & Precup, D. Off-policy deep reinforcement learning without exploration. In: *International Conference on Machine Learning*, 2052–2062 (PMLR, 2019).
25. Levine, S., Kumar, A., Tucker, G. & Fu, J. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv* https://arxiv.org/abs/2005.01643 (2020).
26. Van Hasselt, H., Guez, A. & Silver, D. Deep reinforcement learning with double q-learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, **30** (2016).
27. Raghu, A. et al. Deep reinforcement learning for sepsis treatment. *arXiv* https://arxiv.org/abs/1711.09602 (2017).
28. Peng, X. et al. Improving sepsis treatment strategies by combining deep and kernel-based reinforcement learning. In: *AMIA Annual Symposium Proceedings*, 887 (American Medical Informatics Association, 2018).
29. Coronato, A., Naeem, M., De Pietro, G. & Paragliola, G. Reinforcement learning for intelligent healthcare applications: a survey. *Artif. Intell. Med.* **109**, 101964 (2020).
30. Johnson, A. E. et al. Mimic-iii, a freely accessible critical care database. *Sci. Data* **3**, 160035 (2016).
31. Pollard, T. J. et al. The eicu collaborative research database, a freely available multi-center database for critical care research. *Sci. Data* **5**, 180178 (2018).
32. Batista, G. E. et al. A study of k-nearest neighbour as an imputation method. *His* **87**, 48 (2002).
33. Berkelmans, G. F. et al. Population median imputation was noninferior to complex approaches for imputing missing values in cardiovascular prediction models in clinical practice. *J. Clin. Epidemiol.* **145**, 70–80 (2022).
34. Armitage, E. G., Godzien, J., Alonso-Herranz, V., López-Gonzálvez, Á. & Barbas, C. Missing value imputation strategies for metabolomics data. *Electrophoresis* **36**, 3050–3060 (2015).
35. Fu, J., Luo, K. & Levine, S. Learning robust rewards with adversarial inverse reinforcement learning. In: *International Conference on Learning Representations* (ICLR, 2018).
36. Silver, D. et al. Mastering the game of go without human knowledge. *Nature* **550**, 354–359 (2017).
37. Vinyals, O. et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature* **575**, 350–354 (2019).
38. Wang, Z. et al. Dueling network architectures for deep reinforcement learning. In: *International conference on machine learning*, 1995–2003 (PMLR, 2016).
39. Mnih, V. et al. Human-level control through deep reinforcement learning. *Nature* **518**, 529 (2015).
40. Neuraz, A. et al. Patient mortality is associated with staff resources and workload in the icu: a multicenter observational study. *Crit. Care Med.* **43**, 1587–1594 (2015).
41. Devi, S. et al. Adrenergic regulation of the vasculature impairs leukocyte interstitial migration and suppresses immune responses. *Immunity* **54**, 1219–1230.e7 (2021).
42. Stolk, R. F. et al. Potentially inadvertent immunomodulation: norepinephrine use in sepsis. *Am. J. Respir. Crit. Care Med.* **194**, 550–558 (2016).
43. Li, X. et al. Association between blood urea nitrogen and 30-day mortality in patients with sepsis: a retrospective analysis. *Ann. Palliat. Med.* **10**, 11653–11663 (2021).
44. Schetz, M. Vasopressors and the kidney. *Blood Purif.* **20**, 243–251 (2002).
45. Martin, G. S., Mannino, D. M. & Moss, M. The effect of age on the development and outcome of adult sepsis. *Crit. Care Med.* **34**, 15–21 (2006).
46. Starr, M. E. & Saito, H. Sepsis in old age: review of human and animal studies. *Aging Dis.* **5**, 126 (2014).
47. Yegenaga, I. et al. Clinical characteristics of patients developing arf due to sepsis/systemic inflammatory response syndrome: results of a prospective study. *Am. J. Kidney Dis.* **43**, 817–824 (2004).
48. Asfar, P. et al. High versus low blood-pressure target in patients with septic shock. *N. Engl. J. Med.* **370**, 1583–1593 (2014).
49. Hu, J.-R. et al. Risk-standardized sepsis mortality map of the united states. *Digit. Health* **8**, 20552076211072400 (2022).
50. Clifford, K. M. et al. Challenges with diagnosing and managing sepsis in older adults. *Expert Rev. Anti-Infect. Ther.* **14**, 231–241 (2016).
51. Lambden, S., Laterre, P. F., Levy, M. M. & Francois, B. The sofa score—development, utility and challenges of accurate assessment in clinical trials. *Crit. Care* **23**, 1–9 (2019).
52. Berger, T. et al. Shock index and early recognition of sepsis in the emergency department: pilot study. *West. J. Emerg. Med.* **14**, 168 (2013).
53. Tseng, J. & Nugent, K. Utility of the shock index in patients with sepsis. *Am. J. Med. Sci.* **349**, 531–535 (2015).
54. Abid, O., Akça, S., Haji-Michael, P. & Vincent, J.-L. Strong vasopressor support may be futile in the intensive care unit patient with multiple organ failure. *Crit. Care Med.* **28**, 947–949 (2000).
55. Choi, Y. Sepsis treatment model code. *Zenodo* https://doi.org/10.5281/zenodo.13842300 (2024).

Acknowledgements

This work was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-00982, Development of the multi-center distributed intelligence reinforcement federation technology for cooperating optimal treatments) This work was supported by the GIST-MIT Research Collaboration grant funded by the GIST in 2024. This work was supported in part by the Glocal University 30 Project Fund of Gyeongsang National University in 2024, and in part by "Leaders in INdustry-university Cooperation 3.0" Project funded by the Ministry of Education and National Research Foundation of Korea.

Author contributions

Y.C., K.K., and J.C. built the original concept and direction of the study. Y.C., S.O., and K.K. have developed experimental codes, analyzed methods, and led the whole project. J.H. was involved in the discussion writing to provide medical expertise. W.Y. conducted data preprocessing. C.B. conducted data visualization. H.J. and H.L. have been involved in the result writing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43856-024-00665-x.

Correspondence and requests for materials should be addressed to Kyung-Joong Kim.

Peer review information *Communications Medicine* thanks Joo Heung Yoon, Longxiang Su and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. [Peer reviewer reports are available].