Reinforcement Learning for Medication Dosing using Observational Data and Causal Inference

Joram F. Bakekolo[1], Supreeth P. Shashikumar[2], and Shamim Nemati[3]

[1]Msc student, African Master for Machine Intelligence, Kigali, Rwanda
[2]Division of Biomedical Informatics, University of California San Diego, La Jolla, USA
[3]Division of Biomedical Informatics, University of California San Diego, La Jolla, USA

Abstract

Over the past decade, various methods have been proposed to train artificial intelligence (AI) algorithms for treating sepsis patients with vasopressors and IV fluids. However, development of an AI algorithm that learns to optimallly administer vasopressors and fluids still remains elusive. In this study, we propose to employ a reinforcement learning (RL) based approach that utilizes recurrent neural networks (RNN) to recommend the right amount of vasopressors and fluids to administer to septic patients in order to improve their survival rates. Our RL agent is trained using the MIMIC III dataset which consists of 5,366 septic patients. Using our proposed RL agent, we observed that the net improvement in mortality was 0.190 for the median and 0.242 for the interquartile range. We also report the performance of the RL dosing agent under varying design choices of the neural networks - including various types of recurrent neural networks, various number of RNN layers and various number of training epochs.

1 Introduction

Optimal medication dosing has been an active area of research over the past decade [Organization et al., 2016, Lin et al., 2018, Nemati et al., 2016]. Indeed, it is challenging for clinicians to prescribe the optimal medication dosing to patients in critical situation due to various factors including availability of pertinent patient information, the complexity of human physiology among others. Additionally, sepsis is a critical disease which requires treatment within an optimal timeframe.

Sepsis has become one of the most dangerous and fatal disease having caused death of tremendous people through the world, 49 millions individuals was estimated affected in 2017 and approximately 11 million deaths worldwide [Organization et al., 2020]. Usually injected through vein or under the skin, different treatments have been proposed with different effects, the quantity and the type of fluid to administrate with minimum effect remain one of the main problem. Indeed, it is challenging to recommend the optimal dose since the misdosing (overdosing or under dosing) may worsen the situation, sometimes leading to death. Throughout the years different technique have been used to address the issue [Raghu, 2019].

In this study we addressed the challenge through the use of an automatic controller or techniques rooted from machine learning. A reinforcement learning based close loop controller aims to help clinicians to provide and personalize an optimal drug dosing necessary to be administered to each patient, the method has shown its effectiveness in healthcare applications [Malof and Gaweda, 2011, Gaweda et al., 2004, 2005]. Our work is organized as follows, first we describe sepsis and the various treatments used for treating sepsis. Second, we describe the reinforcement learning technique and algorithm used for optimization. Finally we evaluate the performance of the algorithm using an open-source dataset.

2 Sepsis: Description and Dataset

Defined as a life threatening organ dysfunction caused by a dysregulated host response to infection, sepsis is amongst the leading cause of deaths [Singer et al., 2016]. A new report from the World Health Organization (WHO) stated the disease affects an estimated 49 million people and causes around 11 million deaths globally every year [Organization et al., 2020].

2.1 Descriptions and Treatments of Sepsis

The human immune system and cells play essential role for the protection and defense from different invasion of pathogens and foreign substances (bacteria, virus or other microbes). A variety of pro-inflammatory substances and pro-inflammatory cytokines are released and activated throughout the pathogens associated, generating a cytokine storm [Teijaro, 2017, Chousterman et al., 2017]. In previous studies, inflammatory responses have been observed with diseases related to respiratory viral infection, this is important for the defense against pathogens.

However, it may also deteriorate the normal tissues and cells as well. Indeed, exaggerated inflammatory reactions lead to risky and unnecessary situation, in such case immune suppression from immune system is necessary to allay the system to decrease the ability of antigen [Patil et al., 2017, Mira et al., 2017]. The immune suppression is important for enhancing inflammation and pathogen loading of the body. The inflammatory response and immune suppression both have different actions and contribute to the protection, while eliminating pathogens, the body renovates to a normal state.

In the case of the inflammatory response, immune suppression is exorbitant and out of control, there is no protection, destruction starts, sepsis then develops. Sepsis patients die early from over inflammation reactions. Patients with sepsis are mostly exposed to a variety of infections attacks that has caused a rising in many deaths [Lin, 2020].

Contrary to young people who present more resistance to inflammatory attacks, the elderly are more exposed to sepsis because of weakening immune system, historical chronic diseases [Martin et al., 2006, Hinojosa et al., 2009]. This can explain why the group that consists of the elderly people are more exposed to death in the case of COVID-19. In his work [Lin, 2020] proposed that the outbreak of COVID-19, which has killed one million of people to the whole world, could be viewed as a viral infection. Indeed, multiple organ dysfunction caused by inflammatory response and more have been observed in severe cases to COVID-19 patient admitted to ICU [Liu et al., 2020].

2.1.1 Treatments

In the United States, the mortality rate has decreased from 80% to 20% -30% in the early years due to improved surveillance, early treatment, and advances in support for failing organs [Semler and Rice, 2016]. Sepsis management guideline requires enormous attention and caution not only from clinical also from patients, quick and efficient recognition, on-time control of the source of infection, administration of appropriate dose of antibiotics, and hospital admission. The unusual low blood pressure is one of the first direction for diagnosing septic shock, early intervention is necessary by maintaining the blood pressure at 65 mmHg or above, and the amount of lactate upper to 2mmol/l.

However, there are no dosing rules or protocol for keeping up the mean arterial pressure (MAP) above 65 mmHg. The use of vasopressors and intravenous (IV) fluid, which is considered the first treatment [Marik et al., 2020], can vary according to different parameters, which makes it difficult to determine the ideal combination of vasopressors and IV fluids to administer. Vasopressors induce vasoconstriction and thus help to elevate the MAP, correct the vascular tone depression, improve organ perfusion pressure [Shi et al., 2020]. Despite the surviving sepsis campaign (SSC) recommendation of using norepinephrine or dopamine as first-choice of vasopressors followed by epinephrine [Shi et al., 2020], the type and dose of vasopressor remain a problem [Russell, 2019].

Likewise, fluid administration are essential in hemodynamic stabilization and resuscitation. IV fluids are recommended as major therapeutic to sepsis patient for maintaining and replacing the total body water, electrolytes, as carriers for medications [Avila et al., 2016]. Similarly, the type, dose and timing of administration remain a challenge.

During treatment, patients who respond well to early resuscitation therapy and do not present end-organ hypoperfusion, need a general hospital unit admission. However an admission to ICU is compulsory for patient in septic shock and does not respond to initial treatment. Fundamentally, the type, dose, and strategy for fluid resuscitation and vasopressor remain a great challenge, it is an ongoing debate.

2.2 Data Description

The dataset used in this work is from the MIMIC-III database. MIMIC stands for Medical Information Mart for Intensive Care. MIMIC-III is part of the large PhysioNet dataset, an openly available source collection of physiologic and clinical data from different institutes, developed and maintained by the Massachusetts Institute of Technology (MIT) Lab for computational physiology [Johnson et al., 2016]. The dataset encompasses health information for over 53 thousand patient's hospital admissions for adults aged between 16 years and above, admitted to medical, surgical, pediatric, and neonatal in Intensive Care Unit at Beth Israel Deaconness Medical Center, in Boston Massachusetts. [Johnson et al., 2016]. The MIMIC dataset was developed to support research in epidemiology, clinical decision making, and electronic tool development, enormous updates and improvement have been done through the years. It is freely accessible, open to researcher academy and industry internationally under the data use agreement [Szczepaniak et al., 2006].

Identification of each patient in the dataset, details about each hospital stay, patients movement from ICU also in the hospital are tracked and recorded. Information collected while the patient was in critical care units is stored across eight tales, caregivers, and all observations are stored in the dataset. The dataset contains also the laboratory measurement for patients both within the hospital and those seen outside the hospital and clinic, all fluids administrated to the patient or removed through a drain, and medications ordered or prescribed.

However, the full dataset is available and requires specific certifications and licensing agreements, a subset of 100-patients was released by the MIT team to the public domain, which reflects the patient population characteristics, of the full MIMIC-III dataset.

3 Reinforcement Learning: Concepts and Method

3.1 Reinforcement Learning

Reinforcement learning (RL) is a computational approach whereby an agent tries to learn by trial-and-error-based knowledge acquisition used by humans and animals [Sutton et al., 1998]. The approach is based on the idea from psychology that serves for control theory and stochastic optimization [Gaweda et al., 2005]. The agent gains experience by interacting with the environment(complex and uncertain), by executing actions and observing the outcome(reward or punished). The aim of the agent is to learn a policy(actions) that maximizes a reward [Wang et al., 2018].

RL has shown it effectiveness for learning optimal policies for simulated environment using distributed training with extensive compute capacity [Andersen et al., 2020]. Its combination with deep neural network, has given powerful algorithms which has driven impressive advances in artificial intelligence in recent years, exceeding human performance in games [Botvinick et al., 2019], chemical production [Hubbs et al., 2020] and healthcare [Yu et al., 2019a].

RL algorithms can be broadly divided into online and offline approaches [Schwab and Ray, 2017]. Online approaches assume that the agent interacts directly with the domain and learns a policy, while offline approaches assumes the agent learns a policy from a fixed dataset of trajectories without interacting with the environment [Agarwal et al., 2020]. In this study our agent was trained on retrospectively collected data that included actions administered by clinicians.

3.2 Literature Review: Reinforcement learning in Healthcare

Tremendous advancement have been achieved in the last few years in healthcare using RL due to fast and efficient algorithms developed in the field. Either online or offline, the technique has shown its effectiveness in different problems from decision making to medical dosing problem. RL has been applied in problems related to chronic diseases(Cancer, Diabetes, Anemia) also to critical care(Anesthesia, sepsis) and medical diagnosis [Yu et al., 2019a, Fox et al., 2020, Raghu et al., 2017]. Recently different techniques and algorithms have been applied to tackle the sepsis problem. These algorithms have been used for early prediction [Shashikumar et al., 2021, Wardi et al., 2020, Reyna et al., 2019], progression [Wardi et al., 2021a,b] and treatment of sepsis [Yu et al., 2019b, Peng et al., 2018, Raghu, 2019].

4 Model Dosing for Sepsis Treatment

4.1 Reinforcement Learning Framework Formulation

Conventionally, reinforcement learning consist of an agent placed in a Markov Decision Process environment composed of a set of states, S, a set of actions, A, transition probabilities P and the reward function $r(s, a)$ that depends on a given state-action pair. In the case of septic patient with vasopressors and fluids dosing, we defined:

- Our state such that $S_t^{(k)} = \left[x_t^{(k)}, c^{(k)}\right]$, where t represents the time, k is the patient considered and S_t^k is the state of patient k at time t and $x_t^{(k)}$ is the observations from vital sign and clinical variable, $c^{(k)}$ is the covariates. This includes all clinical measurements done on the patient such as heart rate, blood pressure and more.

- We defined two actions that the agent can take (the pair of dose), the intravenous fluid a^{fluid} and vasopressor a^{pressor}. This is the combination of two intervention categories, IV fluid and vasopressor treatment. At each state S_i the agent takes an actions a such that $a = (a^{\text{fluid}}, a^{\text{pressor}})$.

- Policy is defined as the recommended action to take for a given state, the recommend amount of vasopressors or IV fluid to administrate to a patientat any given state.

- The reward function of our agent is defined by

$$r_t(x_t^{(k)}) = \beta.\mathbf{x}_t^{(k)} = \beta_1 x_{t,1}^{(k)} + \beta_2 x_{t,2}^{(k)} \dots \beta_{d_f} x_{t,d_f}^{(k)} \tag{1}$$

  Where x_t represents the observational data at time t and β is a vector of weights for the corresponding observations, it is an treated as an hyper parameter that can be tuned using optimization techniques.

In our study we used the Generalized Propensity Score to check the effectiveness of the treatment dose (vasopressor and fluid) administrated by our agent to patient, we used the propensity score. The Generalized Propensity Score method has been used in many studies last years to address the problems faced using observational data and multiple treatment have been used.

4.2 Model Description

In our study, we used an algorithm constituted of different type of networks. Each network used plays crucial role, we initialized the algorithm with random parameters to feed into our networks. Value network which is the first network parameterized, predicts the expected value of a given state of the patient. Advantage network gives the advantage of choosing a particular action by the agent same as the policy network which recommends action to choose [Wang et al., 2018]. The GPS network measures the difference for fluid between the recommendation of the agent and the clinician actions. The Average Dose Response Function (ADRF) networks predicts patient's likelihood of survival given their state and treatment level, the ADRF is calibrated. The counterfactual network iterates through the different treatment levels and estimates the survival improvement of each treatment level to assess the potential benefit of the new treatment method figure 1.

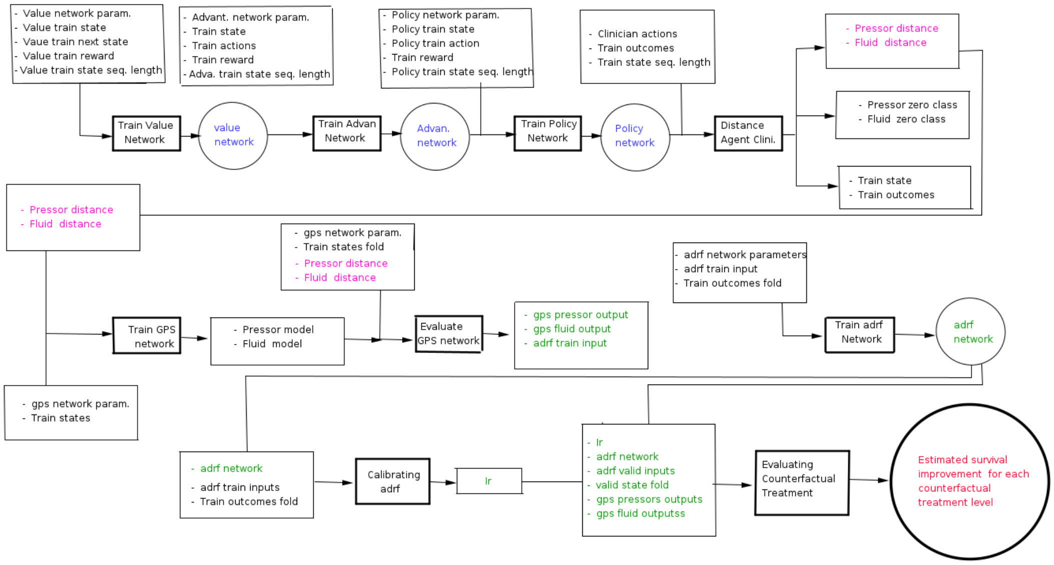

Figure 1: Block Diagram of the overall model

Algorithm 1 Train Agent for Dosing treatment

```
procedure TRAIN_REWARD_AND_AGENT
    Randomly initialize parameter $\beta^{(0)}$
    for $k = 1 : K$ do
        $R_t^{(k)} = \beta_1^{(0)} x_{1,t}^{(k)} + \beta_2^{(0)} x_{2,t}^{(k)} + \cdots + \beta_{d_f}^{(0)} x_{d_f,t}^{(k)}$
    for $n = 0 : N_b - 1$ do
        Train Value Network.
        Train Advantage Network.
        Train Policy Network
        Calculate distance between Agent and Clinician
        Train GPS network.
        Evaluate GPS network.
        Train ADRF network.
        Calibrating ADRF network.
        Evaluating treatment.
        Propose the next sampling point $\beta^{(n+1)}$.
        for $k = 1 : K$ do
            $R_t^{(k)} = \beta_1^{(n+1)} x_{1,t}^{(k)} + \beta_2^{(n+1)} x_{2,t}^{(k)} + \beta_{d_f}^{(n+1)} x_{d_f,t}^{(k)}$

    $\beta \leftarrow \arg\min_{\beta \in \{\beta^0, \ldots \beta^{N_b}\}} [cost(\pi^a; \beta)]$
```

5 Experiments and Results

To check the performance of the algorithm, we trained and tested our agent on a subset of the MIMIC dataset, containing various treatment recommendations from expert clinicians for sepsis and septic shock. The task of our agent was to recommend the right amount of fluid and vasopressor to patient at the right time, for early management of septic shock. Vasopressors are used to restore a patient's blood pressure when it is abnormal or below the average of 65 mm Hg. The best performance, in terms of the expected improvement in survival rate, is achieved when the clinician dosing policy closely follows the agent's recommended dosing of fluids and pressors. As expected, dramatic deviations from the RL policy are not associated with a positive change in survival rate, and may even hurt the patients.

The value network, advantage network and policy network presented in the algorithm play crucial role by predicting the state of the patient and helping to decide the dose to administrate and advantages of the dose administrated. Sepsis is a condition that progresses over time, and to help capture this temporal trend we employed recurrent neural networks. We compared the performances of the RL agent across the Gated recurrent unit network (GRU), Long short-term memory network (LSTM), and the simple recurrent network (RNN). Note that we trained all networks with a batch size of 1000 and for different epochs and layers.

Figure 1 shows the result of average improvement obtain by administrating vasopressor and fluid to sepsis and septic shock patients. The networks were trained on 100, 200, and 300 epochs with 1, 2, 3 and 5 layers. For the GRU network, we have GRU_{100}, GRU_{200} GRU_{300}, the simple RNN network, RNN_{100}, RNN_{200} RNN_{300}, the same LSTM network, $LSTM_{100}$, $LSTM_{200}$ $LSTM_{300}$.

Values (improvement) greater than zero corresponds to reduction in mortality when the RL's optimal policy is followed vs clinician policy. When values is equal to zero, that mean the dosing has not brought any change to the patient and the situation remain the same. However, negative values show the administration has not ameliorated or improved the patient situation, on the contrary it has degraded the situation, the agent has made a worse recommendation vs clinician recommendation.

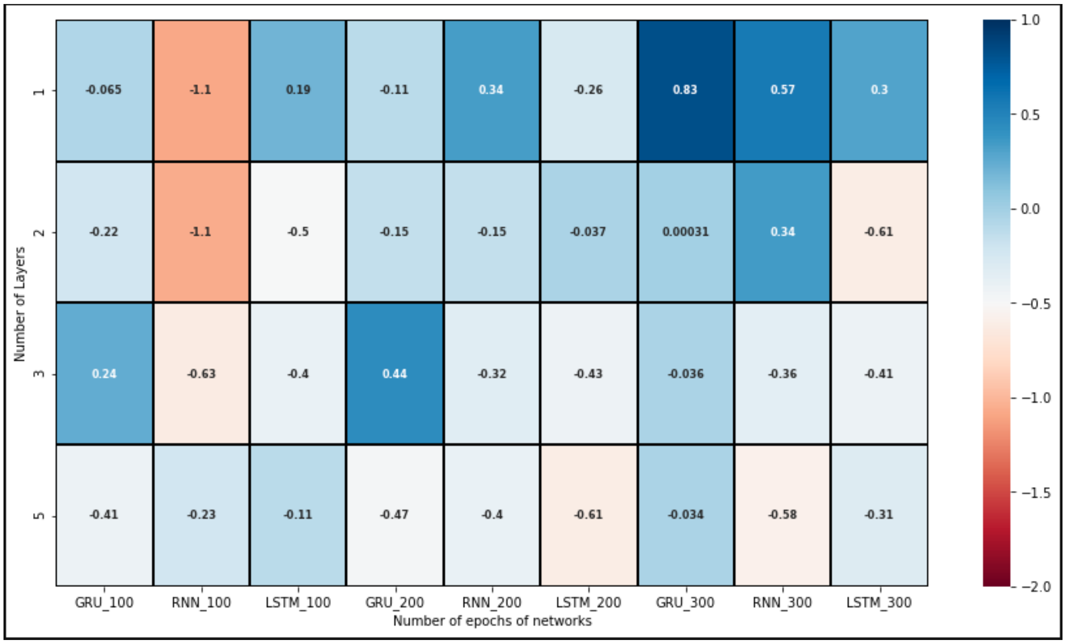

Figure 2: Improvement sepsis treatment using recurrent neural network

We can observe on the figure, for instance by training the agent (value network, advantage network and policy network) using the Gated recurrent unit network for 300 epoch and 1 layer, the average improvement is 0.831026 which is a good recommendation, whereas the situation is worst, the agent starts gave bad recommendations -0.035629 and -0.033629 using the same network for 3 and 5 layers respectively.

We can observe from figure 2 that by training our agent on 300 epochs and 1 layer, the Gated recurrent unit network has outperformed the simple recurrent network and the Long short-term memory network. Distribution obtained by training with the Gated recurrent unit network, the simple recurrent network and the Long short-term memory network for different epochs and different layers can be summarized such that figure 3.

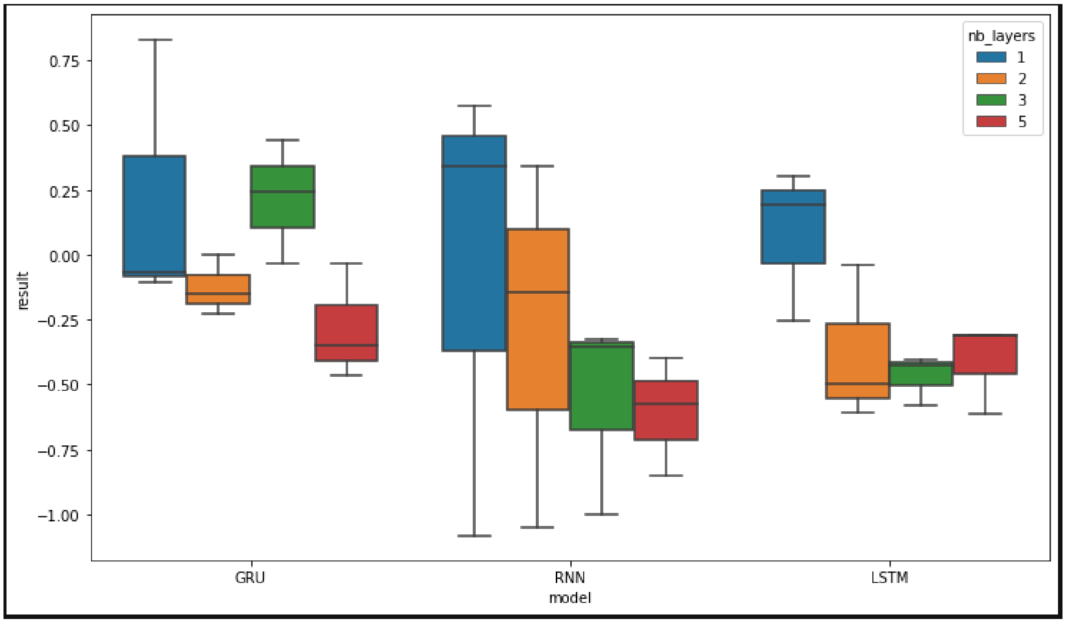

Figure 3: Distributions of improvement using different networks

6 Conclusion and Perspectives

In conclusion, our method (algorithm) presents significant novel contributions, a good step forward for getting a time series agent able to recommend in time the appropriate amount of fluid and vasopressor for sepsis management, compared to previous studies done on sepsis and septic shock treatment. Our agent used time-series recurrent neural network to assign a state to a patient from clinical information and recommendation in time as described. Positive or good recommendations have been done by the agent using time-series recurrent neural network, this is significant, it can improve health situation of patients. However, amelioration needs to be done to overcome it limitations. At this stage our agent cannot be used in real life for treating sepsis patients accurately. In this work the policy evaluation has been done using standard counterfactual reasoning, in future training the agent using a better offline reinforcement learning policy would be better. An agent able to learn a better policy from observing clinician, rewarded by patient improvement over time will perform better. We also need to optimize our network parameters for better improvement using optimization technique (Bayesian optimization).

Regardless of these limitations which will be explored in future work, our reinforcement learning approach marks an interesting first step towards providing real-time, individualized computer-assisted treatment of sepsis and septic shock management.

Acknowledgments

I would like to express my deepest feeling of gratitude toward Dr Shamim Nemati, my supervisor and Dr Supreeth P. Shashikumar, my wonderful co-supervisor, you are the perfect guides. Each moment I discussed with you was a piece of new knowledge, you were available, open to discussion and very serious in the project.

References

Rishabh Agarwal, Dale Schuurmans, and Mohammad Norouzi. An optimistic perspective on offline reinforcement learning. In *International Conference on Machine Learning (ICML)*, 2020.

Per-Arne Andersen, Morten Goodwin, and Ole-Christoffer Granmo. Towards safe reinforcement-learning in industrial grid-warehousing. *Information Sciences*, 2020.

Audrey A Avila, Eliezer C Kinberg, Nomi K Sherwin, and Robinson D Taylor. The use of fluids in sepsis. *Cureus*, 8(3), 2016.

Matthew Botvinick, Sam Ritter, Jane X Wang, Zeb Kurth-Nelson, Charles Blundell, and Demis Hassabis. Reinforcement learning, fast and slow. *Trends in cognitive sciences*, 23(5):408–422, 2019.

Benjamin G Chousterman, Filip K Swirski, and Georg F Weber. Cytokine storm and sepsis disease pathogenesis. In *Seminars in immunopathology*, volume 39, pages 517–528. Springer, 2017.

Ian Fox, Joyce Lee, Rodica Pop-Busui, and Jenna Wiens. Deep reinforcement learning for closed-loop blood glucose control. In *Machine Learning for Healthcare Conference*, pages 508–536. PMLR, 2020.

Adam E Gaweda, Alfred A Jacobs, George R Aronoff, and Michael E Brier. Intelligent control for drug delivery in management of renal anemia. In *2004 International Conference on Machine Learning and Applications, 2004. Proceedings.*, pages 355–359. IEEE, 2004.

Adam E Gaweda, Mehmet K Muezzinoglu, George R Aronoff, Alfred A Jacobs, Jacek M Zurada, and Michael E Brier. Individualization of pharmacological anemia management using reinforcement learning. *Neural Networks*, 18(5-6):826–834, 2005.

Ernesto Hinojosa, Angela R Boyd, and Carlos J Orihuela. Age-associated inflammation and toll-like receptor dysfunction prime the lungs for pneumococcal pneumonia. *The Journal of infectious diseases*, 200(4):546–554, 2009.

Christian D Hubbs, Can Li, Nikolaos V Sahinidis, Ignacio E Grossmann, and John M Wassick. A deep reinforcement learning approach for chemical production scheduling. *Computers & Chemical Engineering*, page 106982, 2020.

Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.

Hong-Yuan Lin. The severe covid-19: A sepsis induced by viral infection? and its immunomodulatory therapy. *Chinese Journal of Traumatology*, 2020.

Rongmei Lin, Matthew D Stanley, Mohammad M Ghassemi, and Shamim Nemati. A deep deterministic policy gradient approach to medication dosing and surveillance in the icu. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4927–4931. IEEE, 2018.

Di Liu, Qiang Wang, Huacai Zhang, Li Cui, Feng Shen, Yong Chen, Jiali Sun, Lebin Gan, Jianhui Sun, Jun Wang, et al. Viral sepsis is a complication in patients with novel corona virus disease (covid-19). *Medicine in Drug Discovery*, 8:100057, 2020.

Jordan M Malof and Adam E Gaweda. Optimizing drug therapy with reinforcement learning: The case of anemia management. In *The 2011 International Joint Conference on Neural Networks*, pages 2088–2092. IEEE, 2011.

Paul E Marik, Liam Byrne, and Frank van Haren. Fluid resuscitation in sepsis: the great 30 ml per kg hoax. *Journal of Thoracic Disease*, 12(Suppl 1):S37, 2020.

Greg S Martin, David M Mannino, and Marc Moss. The effect of age on the development and outcome of adult sepsis. *Critical care medicine*, 34(1):15–21, 2006.

Juan C Mira, Lori F Gentile, Brittany J Mathias, Philip A Efron, Scott C Brakenridge, Alicia M Mohr, Fredrick A Moore, and Lyle L Moldawer. Sepsis pathophysiology, chronic critical illness and pics. *Critical care medicine*, 45(2):253, 2017.

Shamim Nemati, Mohammad M Ghassemi, and Gari D Clifford. Optimal medication dosing from suboptimal clinical examples: A deep reinforcement learning approach. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2978–2981. IEEE, 2016.

World Health Organization et al. *Medication errors*. World Health Organization, 2016.

World Health Organization et al. Global report on the epidemiology and burden of sepsis: current evidence, identifying gaps and future directions. 2020.

Naeem K Patil, Yin Guo, Liming Luan, and Edward R Sherwood. Targeting immune cell checkpoints during sepsis. *International Journal of Molecular Sciences*, 18(11):2413, 2017.

Xuefeng Peng, Yi Ding, David Wihl, Omer Gottesman, Matthieu Komorowski, Li-wei H Lehman, Andrew Ross, Aldo Faisal, and Finale Doshi-Velez. Improving sepsis treatment strategies by combining deep and kernel-based reinforcement learning. In *AMIA Annual Symposium Proceedings*, volume 2018, page 887. American Medical Informatics Association, 2018.

Aniruddh Raghu. Reinforcement learning for sepsis treatment: Baselines and analysis. 2019.

Aniruddh Raghu, Matthieu Komorowski, Imran Ahmed, Leo Celi, Peter Szolovits, and Marzyeh Ghassemi. Deep reinforcement learning for sepsis treatment. *arXiv preprint arXiv:1711.09602*, 2017.

Matthew A Reyna, Chris Josef, Salman Seyedi, Russell Jeter, Supreeth P Shashikumar, M Brandon Westover, Ashish Sharma, Shamim Nemati, and Gari D Clifford. Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. In *2019 Computing in Cardiology (CinC)*, pages Page–1. IEEE, 2019.

James A Russell. Vasopressor therapy in critically ill patients with shock. *Intensive care medicine*, 45(11):1503–1517, 2019.

Devin Schwab and Soumya Ray. Offline reinforcement learning with task hierarchies. *Machine Learning*, 106(9-10):1569–1598, 2017.

Matthew W Semler and Todd W Rice. Sepsis resuscitation: fluid choice and dose. *Clinics in chest medicine*, 37(2):241–250, 2016.

Supreeth P Shashikumar, Christopher Josef, Ashish Sharma, and Shamim Nemati. Deepaise–an interpretable and recurrent neural survival model for early prediction of sepsis. *Artificial Intelligence in Medicine*, page 102036, 2021.

Rui Shi, Olfa Hamzaoui, Nello De Vita, Xavier Monnet, and Jean-Louis Teboul. Vasopressors in septic shock: which, when, and how much? *Annals of Translational Medicine*, 8(12), 2020.

Mervyn Singer, Clifford S Deutschman, Christopher Warren Seymour, Manu Shankar-Hari, Djillali Annane, Michael Bauer, Rinaldo Bellomo, Gordon R Bernard, Jean-Daniel Chiche, Craig M Coopersmith, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). *Jama*, 315(8):801–810, 2016.

Richard S Sutton, Andrew G Barto, et al. *Introduction to reinforcement learning*, volume 135. MIT press Cambridge, 1998.

M Cleat Szczepaniak, KW Goodman, MW Wagner, J Hutman, S Daswani, MM Wagner, AW Moore, and RM Ayrel. Advancing organizational integration: negotiation, data use agreements, law, and ethics. *Handbook of Biosurveillance*, pages 465–480, 2006.

John R Teijaro. Cytokine storms in infectious diseases. In *Seminars in immunopathology*, volume 39, pages 501–503. Springer, 2017.

Qing Wang, Jiechao Xiong, Lei Han, Han Liu, Tong Zhang, et al. Exponentially weighted imitation learning for batched historical data. In *Advances in Neural Information Processing Systems*, pages 6288–6297, 2018.

G Wardi, S Shashikumar, M Carlile, M Krak, S Hayden, A Holder, and S Nemati. 301 use of transfer learning to improve external validity of a machine-learning algorithm to predict septic shock in the emergency department. *Annals of Emergency Medicine*, 76(4):S116, 2020.

Gabriel Wardi, Morgan Carlile, Andre Holder, Supreeth Shashikumar, Stephen R Hayden, and Shamim Nemati. Predicting progression to septic shock in the emergency department using an externally generalizable machine-learning algorithm. *Annals of Emergency Medicine*, 2021a.

Gabriel Wardi, Supreeth Shashikumar, Thomas Allen, and Shamim Nemati. 1233: Development and validation of a novel machine learning algorithm to predict sepsis readmissions. *Critical Care Medicine*, 49(1):620, 2021b.

Chao Yu, Jiming Liu, and Shamim Nemati. Reinforcement learning in healthcare: A survey. *arXiv preprint arXiv:1908.08796*, 2019a.

Chao Yu, Guoqi Ren, and Jiming Liu. Deep inverse reinforcement learning for sepsis treatment. In *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 1–3. IEEE, 2019b.