Fall 2024

Importance Sampling to Learn Vasopressor Dosage to Optimize Patient Mortality in an Interpretable Manner

Anshul Rastogi

IMPORTANCE SAMPLING TO LEARN VASOPRESSOR DOSAGE TO OPTIMIZE PATIENT MORTALITY IN AN INTERPRETABLE MANNER

Anshul Rastogi
College of Engineering
University of Connecticut
Storrs, CT 06269, USA
anshul.rastogi@uconn.edu

Dr. Dongjin Song
(Mentor)
Department of Computer Science & Engineering
University of Connecticut
Connecticut, United States of America
dongjin.song@uconn.edu

ABSTRACT

Sepsis is a life-threatening organ dysfunction resulting from an improperly compensated bodily response to infection. There is a high urgency among clinicians to develop a set of real-time explainable treatment guidelines and tools to address the high mortality rate of sepsis patients. We present a reinforcement learning approach for vasopressor drug dosage in intensive care unit sepsis patients to achieve a better-than-expert treatment policy. We preserved interpretability with a prototype learning layer and learn actions in an off-policy manner with importance sampling. We evaluated our design on the MIMIC-IV deidentified electronic health record dataset with an 80%-20% training-validation split for 300 epochs. For patient trajectories where expert actions agreed with over 80% of model policy actions, we achieved a mortality rate of approximately 35%, outperforming the expert policy mortality rate of 38%. Where expert actions agreed with the model policy for less than 80% of states, patient mortality reached 92%. This work indicates the model policy outperformed the expert policy by a significant margin while preserving interpretability. Future directions include improved data cleaning and extraction techniques, more tailored patient feature selection, exploring solutions to address limited trajectory exploration in training data, and improved evaluation metrics.

1 INTRODUCTION

The advent of artificial intelligence (AI) and other computational techniques displays great potential to transform medical treatment and diagnoses as a result of AI's ability to glean patterns from complex, high-dimensional data (Younis et al., 2024; Gulshan et al., 2016). AI has demonstrated success in various clinical decision-making tasks, including in patient trials (Zhang et al., 2017; Wang et al., 2020; Liu et al., 2020a; Lam et al., 2022).

One disease of concern is sepsis. Sepsis is defined as life-threatening organ dysfunction resulting from improper bodily response to infection (Singer et al., 2016). Sepsis is a leading cause of death; sepsis patients suffer a high mortality rate estimated to range from 20-50% (Nasir et al., 2015; Fleischmann et al., 2016; Fleischmann-Struzek et al., 2020; Rhee et al., 2017), and sepsis-related deaths account for approximately 20% of global mortality as of 2020 (Rudd et al., 2020). Septic shock, the most severe escalation of sepsis, persistently maintains a high mortality rate at over 50% (Daniels, 2011; Vincent et al., 2006). Despite medical advances, mortality for leading causes of

sepsis has not declined for two decades now (Prest et al., 2022). In light of this, general guidelines for sepsis management have been established, such as the Surviving Sepsis Campaign (Evans et al., 2021). However, effective tools for real-time sepsis patient management are lacking in various aspects of treatment (Marik, 2015), such as fluid rescuscitation (Marik & Bellomo, 2016; Byrne & Van Haren, 2017; Marik et al., 2017) and vasopressor dosage (Bai et al., 2014; Waechter et al., 2014), where immediate intervention is urgent.

Sepsis management lends itself to the development of computational tools as the typical intensive care unit (ICU) environment is data-rich due to close patient monitoring (Johnson et al., 2023a). Computational techniques, such as predictive analyses, have been successful at improving real patient outcomes for sepsis (Cull et al., 2023; Komorowski et al., 2018). However, deep learning approaches are generally black-box (Molnar, 2022). In the high-stakes patient care environment, faithful interpretability is crucial to both clinicians and patients. This is a major roadblock to the integration of such techniques into real clinical settings; understandably, clinicians trust explainable models better (doi, 2020). Methods in literature often attempt explanation of model policy decisions in a manner disjoint from the model's functioning, such as with retrospective model-agnostic feature importance analysis or with a separate model dedicated to policy explanation Molnar (2022); Komorowski et al. (2018). However, literature suggests the use of inherently interpretable models; "explainable" models risk being unfaithful or improperly portraying model decisions (Rudin, 2019).

One approach for inherent explainability is the use of a prototype layer (Molnar, 2022). The incorporation of prototypes mimics clinical case study-based decision-making in a meaningful manner, and thus can avoid compromising model performance (Rudin, 2019); indeed, prototype learning has demonstrated success in various real-world decision-making tasks (Biehl et al., 2016). By learning exemplar states and using comparisons with these prototypes to guide model decisions, faithful interpretability arises in a human-comprehensible manner (Ming et al., 2019a).

Previously, our lab co-developed an interpretable skill learning (ISL) framework, an imitation learning approach to mimic expert total intravenous fluid dosages for sepsis patients in the intensive care unit (ICU) (Jiang et al., 2023). Reliant on prototype learning and behavior cloning, the ISL achieved state-of-the-art performance—particularly for accuracy and AUC metrics—while providing inherent interpretability via the prototypes learning layer.

However, literature persistently finds that expert treatments for sepsis patients are frequently suboptimal (Marik & Bellomo, 2016; Byrne & Van Haren, 2017; Marik et al., 2017; Bai et al., 2014; Waechter et al., 2014), with various aspects of sepsis patient treatment debated in medical literature. Rather than imitation learning, an off-policy reinforcement learning-based approach has potential to perform better than the expert policy (Prasad et al., 2017; Fujimoto et al., 2019; Sutton & Barto, 2018), especially for long-term patient outcomes such as mortality. Moreover, reinforcement learning has demonstrably performed well in other clinical contexts(Wang et al., 2018; Yu et al., 2019; Liu et al., 2020b).

In this paper, we present an expansion of the ISL's methods via the introduction of reinforcement learning to achieve a better-than-expert policy. We preserve the prototype learning layers in our design while utilizing off-policy importance sampling to determine a better policy from expert data to optimize sepsis patient mortality in an innately interpretable manner. Rather than intravenous fluid intake, we seek a policy to assign vasopressor dosage. Vasopressors are crucial for the management of circulatory regulation in ICU patients, ensuring appropriate blood flow to organ systems (VanValkinburgh et al., 2024). Clinical literature remains controversial on the optimal policy for vasopressor dosage for sepsis patients (Bai et al., 2014; Waechter et al., 2014); accordingly, this is a medically compelling action objective to determine an optimal computational policy for. We demonstrate that this usage of reinforcement learning succeeds in outperforming the expert policy while the skill embedding of the prototype learning layer provides reasonable interpretability.

2 Related Work

Effective interpretability techniques to address the black-box problem remains challenging; various approaches exist, ranging from model-agnostic analyses such as Shapley values to feature visualization (Molnar, 2022). Prototype learning has emerged in research as a technique to meet the intrinsic interpretability demands of higher-stakes decision settings (Ming et al., 2019b; Ni et al., 2021).

By presenting exemplar environment states as learned model weights, models employing prototype learning are capable of faithfully rationalizing their decision-making process.

Prototype learning allows not only for inherent interpretability of the model but additionally provides a form of contextual hierarchy, which models clinician decisions more closely; moreover, hierarchical policies have demonstrably performed well in healthcare contexts (Sharma et al., 2019; An et al., 2021). In the ISL framework, prototypes meaningfully represented short-term patient state trends and thus improved performance while introducing a higher-level hierarchy preceding the lower-level policy dictated by behavior cloning (Jiang et al., 2023). Moreover, these prototypes are further learned from training data, superseding the otherwise labor-intensive need to establish hierarchical policies with expert-directed subgoals.

Reinforcement learning as an approach for treatments in the clinical setting is an avid area of current research (Liu et al., 2020b) and has shown success in various medical sectors (Prasad et al., 2017; Melanie K Bothe & Faisal, 2013; Lowery & Faisal, 2013; Wang et al., 2018). However, numerous obstacles exist. The complexity of human biology and the resultant data suggests the use of AI but concomitantly represents a near-infinite state space (Khezeli et al., 2023). As such, assigning rewards, too, is often challenging. Many clinical datasets in particular suffer from a narrow exploration of potential trajectories as deviation from typical expert actions may be hazardous to the patient.

One project that informed our reinforcement learning approach strongly is the AI Clinician (Komorowski et al., 2018). The AI Clinician was developed for a similar objective—to optimize long-term patient outcomes for sepsis patients via management of vasopressor and intravenous fluid dosage. This is achieved via policy iteration. However, the AI Clinician seeks interpretability not inherently but rather with retrospective feature importance analysis, revealing the role of clinically meaningful values in its decision-making.

Our particular approach—coupling off-policy reinforcement and prototype learning architectures to allow intrinsic interpretability in the context of medical treatment across patient trajectories while seeking to achieve better-than-expert performance—appears to be a unique contribution in literature.

3 METHODOLOGY

3.1 PROBLEM FORMULATION

To guide treatment, a pertinent framework for clinical decision-making is the dynamic treatment regime (DTR). DTRs are an emerging methodology within the paradigm of personalized medicine that allow for individualized treatments for patients adaptive to time-varying settings. DTRs are composed of multi-stage interventions with discretized patient states (Chakraborty & Murphy, 2014); thus, DTR patient trajectories can be described as a sequence of state-action pairs (s_t, a_t), where s_t and a_t denote the patient state and corresponding expert action, respectively, at timestep t. We denote the length (in timesteps) of the patient stay until termination as T; as such, a corresponding DTR trajectory is denoted $[(s_1, a_1), (s_2, a_2), \ldots, (s_T, a_T)]$. Evidence-based DTRs have garnered significant interest for management of chronic diseases and are keenly amenable to computational and statistical techniques (Sun et al., 2023). Furthermore, DTRs may be modeled as Markov decision processes, rendering them ideal for predictive tasks. Here, the DTR model for patient trajectories permits discretization of expert actions and comparison to policy behavior.

However, clinician decisions are not determined solely by singular patient states. Rather, trends in patient conditions are observed. Therefore, as per the ISL framework design, our design relies on patient segments rather than individual states. At timestep $t = t_0$, the corresponding patient segment S_{t_0} is a series of m contiguous patients states ending at s_{t_0} such that $S_{t_0} = [s_t]_{t=t-m+1}^{t_0} = [s_{t_0-m+1}, \ldots, s_{t_0}] \in \mathbb{R}^{m \times d}$, where d denotes the dimension of any given patient state s_t. A patient trajectory $[(s_t, a_t]_{t=1}^{T}$ can be decomposed into $n = T - m + 1$ overlapping segments and can be written as $[S_j]_{j=m}^{T} = [[s_t]_{t=j-m+1}^{j}]_{j=m}^{T}$. Segments mimic clinical decision-making more closely by including some medical history of the treatment, displaying a brief trend in patient status. Moreover, segments somewhat mitigate the typical independent-and-identically-distributed assumption that otherwise promotes compounding errors in many model designs, such as in a flat vanilla behav-

ior cloning policy. Finally, we assume that the segment granularity of examining patient trajectories is more adaptable to varying DTR task contexts.

Our model policy objective, then, is to suggest a vasopressor dosage that improves long-term ICU sepsis patient outcomes with the greatest likelihood. Similar to previous work (Komorowski et al., 2018), our long-term optimization objective is the minimization of 90-day patient mortality.

However, no objective optimal policy is known for for vasopressor dosage. As such, determining an appropriate metric to evaluate the model is challenging. One more straightforward approach that may gauge the model's performance in comparison to the expert policy may be the measurement of mortality rates where expert decisions align with model decisions and mortality rates where expert decisions differ. In particular, we refer to the 90-day mortality rate of patient trajectories where over 80% of expert and model decisions are in agreement as *aligned trajectory mortality*; 90-day mortality rate for trajectories where expert and model agreement is lower than 80% of segments is referred to as the *differing trajectory mortality*. We hope to minimize aligned trajectory mortality while having higher-than-expert mortality rates for differing trajectory mortality.

3.2 Patient Inclusion-Exclusion Criteria

We first selected all sepsis patients by identifying any patients billed for sepsis-related ICD9 or ICD10 codes. We then filtered out patients meeting any of the following criteria:

- **Withdrawal of treatment** As per prior literature (Komorowski et al., 2018), patients whose treatment is withdrawn were removed from the dataset as clinical decisions are not made to optimize patient survival in such cases. Withdrawal of treatment was inferred from patients having receiving a positive dosage of vasopressor at any point during their ICU stay and suffering mortality within 24 hours of discharge from the ICU.
- **Too-brief ICU stay** Patient trajectories consisting of less than m timesteps were removed due to being unable to be decomposed into segments.
- **Pediatric patients** Patients < 18 years of age at any point were excluded as treatment for pediatric patients differs from that of adult patients and is medically distinct (Fraser & Campbell, 2007; King et al., 2022).

3.3 Design

The model architecture, visualized in Figure 1, consists of three layers, all following a data preprocessing pipeline. The first two—segment embedding and prototype learning—are preserved in design from the ISL (Jiang et al., 2023). We briefly discuss them here.

The segment embedding layer receives from the data preprocessing pipeline patient segments consisting of m patient states. Each patient state is denoted as described by d features and is therefore of dimension d. Therefore, a segment S_t can be represented as $[s_{t-m+1}, \ldots, s_t] \in \mathbb{R}^{m \times d}$. The segment embedding layer first receives the segment input to a step-wise multilayer perceptron (MLP), generating embeddings $f_t = \textbf{MLP}(s_t)$, where $f_t \in \mathbb{R}^{d'}$, for each constituent state of the segment. As such, $\textbf{MLP}(S_t) = [f_{t-m+1}, \ldots, f_t] \in \mathbb{R}^{m \times d'}$. Via a 1D convolution layer, the tensor of state embeddings is transformed and concatenated into a singular segment embedding $z_t \in \mathbb{R}^h$. Generally, we seek to select a small m such that the convolutional neural network design of the segment embedding layer performs well in extracting a representative segment embedding without being impeded by not accounting for long-term dependencies as an LSTM/GRU or Transformer design may have offered (Hochreiter & Schmidhuber, 1997; Cho et al., 2014; Vaswani et al., 2023).

The segment embedding z_t is then passed to the prototype learning layer. We constrain the prototype layer to k prototypes $p = [p_i]_{i=1}^k = [p_1, p_2, \ldots, p_{k-1}, p_k] \in \mathbb{R}^{k \times h}$, which are treated as learnable model parameters and ultimately represent exemplar segment embeddings. As detailed in our ISL work, the Euclidean distance between z_t and each prototype is used to calculated similarity scores corresponding each p_i (Jiang et al., 2023). These scores are used to transform z_t into o_t, a weighted sum of all p_i.

We preserve the jointly-constraint learning objectives for the segment embedding and prototype learning layers from the ISL framework. These learning objectives are summarized as follows. Note

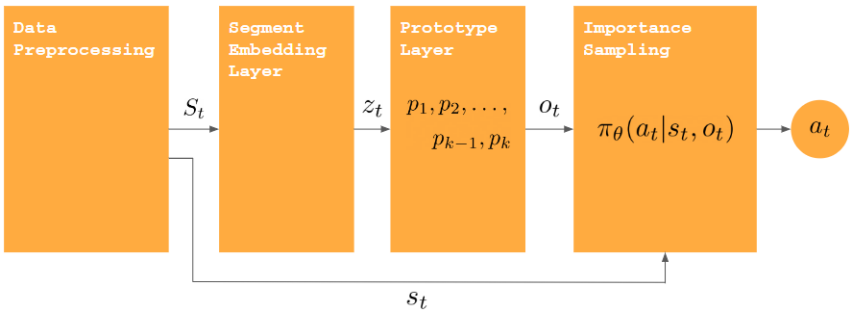

Figure 1: Summary of data pipeline and the model architecture, consisting of 1) the data preprocessing pipeline, 2) the segment embedding layer, 3) the prototype learning layer, and 4) the importance sampling layer.

that we denote the L_2-norm via the notation $\|\cdot\|_2$ and that $[\![x \ldots y]\!] = [z|x \leq z \leq y, z \in \mathbb{N}]$ for given $x, y \in \mathbb{N}$.

- **Clustering structure regularization** To ensure that segment embeddings are closeby to nearby learned prototypes so as to ensure prototypes are representative of nearby segment clusters, we minimize the L_2 distance:

$$\mathcal{L}_{\text{cluster}} = \sum_{j=m}^{T} \min_{i \in [\![1 \ldots k]\!]} \|z_j - p_i\|_2^2$$

- **Prototype-segment evidence regularization** To ensure prototype-segment association—thus improving interpretability of the model—and that prototypes do accurately resemble a closeby batch of segments, we minimize the L_2 distance between the prototype vector and the nearest segment embedding:

$$\mathcal{L}_{\text{evidence}} = \sum_{i=1}^{k} \min_{j \in [\![m \ldots T]\!]} \|p - z_j\|_2^2$$

Additionally, to reduce redundancy in prototypes and ensure their meaningfulness, we impose the following diversity regularization term on the prototype learning layer that encourages L_2 distance between distinct prototypes to match a given proximity threshold d_{min}:

$$\mathcal{L}_{\text{diversity}} = \sum_{i=1}^{k} \sum_{i' \neq i}^{k} \max\left(0, d_{\text{min}} - \|p_i - p_{i'}\|_2^2\right)$$

As a result of these regularization terms, following training loss convergence, each prototype p_i is often closeby to a nearby segment. To enhance interpretability by ensuring that prototypes correspond to real patient states, we assigned prototypes to their nearest segment embedding in latent space by L_2-norm from the training set.

The weighted sum of prototypes o_t is concatenated to the most recent state in the segment, s_t, and passed from the prototype layer to the importance sampling layer, which operates a model policy $\pi_\theta(a_t|s_t, o_t)$ parametrized by θ. We denote the expert policy as π_E. Given rewards $R = [R_j]_{j=n}^{T}$ corresponding to respective segments in a given trajectory, our importance sampling loss is defined as follows:

$$\mathcal{L}_{\text{importance}} = \frac{1}{T} \sum_{j=n}^{T} \frac{\log \pi_\theta(a_j|s_j, o_j)}{\log \pi_E(a_j|s_j)} R_j$$

However, as in many clinical datasets, π_E provides a very limited visibility of potential patient trajectories. To allow for exploration, we "softened" the policy such that, with a frequency of 5%, a_j

is replaced a randomly selected action excluding the expert treatment decision (Komorowski et al., 2018). If the original a_j was part of a trajectory with a positive 90-day mortality outcome—that is, the patient survives—the corresponding R_j for the replaced action is reduced by 90%. Otherwise, R_j is amplified by a factor of 2.

The total objective function for the model is ultimately

$$\mathcal{L} = \mathcal{L}_{\text{importance}} + \lambda_1 \mathcal{L}_{\text{cluster}} + \lambda_2 \mathcal{L}_{\text{evidence}} + \lambda_3 \mathcal{L}_{\text{diversity}},$$

where λ_1, λ_2, and λ_3 are weights to balance the various regularization term constituents.

4 EVALUATION

4.1 DATASET

To evaluate our design, we selected the MIMIC-IV dataset (Johnson et al., 2023a). MIMIC-IV is the most recent iteration of a series of electronic health record datasets sourced from the data of the Beth Israel Deaconess Medical Center and contains detailed information on over 65,000 distinct ICU patient stays. The MIMIC-IV dataset is openly available, and the modular organization lends itself to convenient data extraction for machine learning tasks.

Examining ICU stays, we extracted all patient trajectories as per our inclusion-exclusion criteria (see 3.2). We received 10,877 sepsis patients. This cohort suffered a 30-day mortality rate of 25.0% and a 90-day mortality rate of 34.5%, consistent with the high mortality rate suggested for sepsis patients in literature.

We discretized patient trajectories into states with a timestep of $\Delta t = 4$ hours. Across each 4-hour interval described by a patient state, we extracted the corresponding comprehensive vasopressor dosage in norepinephrine equivalent with an equation adapted from (Goradia et al., 2020). We further referenced the well-iterated Minimal Sepsis Data Model (MSDM) developed by the creators of the OpenSep pipeline, an open-source data processing pipeline to identify sepsis patients in EHR datasets (Hofford et al., 2022). As per the MSDM, dobutamine was treated equivalent to dopamine in dosage and we thus incorporated dobutamine dosage into the total vasopressor calculation accordingly:

$$\begin{aligned}\text{Total vasopressor} = \text{norepinephrine} &+ \text{epinephrine} \\ &+ \frac{\text{phenylephrine}}{10} + \frac{\text{dopamine + dobutamine}}{100} \\ &+ \frac{\text{metaraminol}}{8} + \text{vasopressin} \times 2.5 \\ &+ \text{angiotensin II} \times 10\end{aligned}$$

We then discretized vasopressor dosage into 5 intervals. An action of 0 signified no vasopressors provided; similarly, an action of 1 meant total vasopressor dosage was within the interval (0,0.08], and action of 2 corresponded to the interval (0.08,0.22], and action of 3 to (0.22, 0.45], and and action of 4 to (0.45,∞). These bounds were based on common vasopressor dosages in the dataset. Notably, approximately 68% of patient states received a vasopressor dosage of 0.

To describe patient states, we compiled a selection of features based on relevance in previous literature (Jiang et al., 2023; Komorowski et al., 2018; Sharma et al., 2019). We further included features with less than 70% missing data for sepsis patients and bearing a Spearman correlation coefficient of greater than 0.1 to total vasopressor dosage. In total, we compiled 96 different features, including demographics (language, gender, marital status, etc.), lab measurements (chloride, partial oxygen pressure, etc.), vitals (mean arterial pressure, Glasgow Coma Scale scores, etc.), procedures (mechanical ventilation, Foley catheter, etc.), and pertinent comorbidities (diabetes, hypertension, congestive heart failure, etc.). We performed data cleaning by winsorizing outliers identified via Tukey's method. For missing values, we utilized simple imputation and replaced them with median values. We then took the logarithm of features exhibiting a log-normal distribution.

Additionally, for the calculation of $\mathcal{L}_{\text{importance}}$, we must extract $\pi_E(a_t|s_t)$ in some form. Specifically, we require a method to estimate the state-transition probabilities of the expert policy. To achieve this, we utilized K-means with 25 clusters to group patient states. Probability for various actions provided the state s_t were calculated to estimate $\pi_E(a_t|s_t)$.

4.2 Reward Function

The task objective for our model is to optimize long-term patient outcomes. Due to the high patient mortality rates encountered in sepsis, we hope to increase patient survival. Accordingly, our reward function must correspond to lower patient mortality.

One metric of consideration is the Sequential Organ Failure Assessment (SOFA) score at a given patient state, which quantifies the amount and severity of dysfunction in organ systems for patients with acute ICU morbidity and has been found to bear a positive correlation to mortality (Jones et al., 2009). This score is typically a significant factor in the clinical assessment of sepsis patients as a value ≥ 2 is required to diagnose a patient with sepsis (Lambden et al., 2019a; Singer et al., 2016). Moreover, a higher SOFA score indicates a higher likelihood of mortality.

Reliability of the SOFA score as a predictor of mortality may vary due to challenges in clinical measurements and consistency (Lambden et al., 2019b). Our SOFA calculation method mimics that of the MSDM (Hofford et al., 2022). This method is summarized in Table 1 and follows established metrics (Vincent et al., 1996; Hofford et al., 2022).

Table 1: SOFA score calculation subscore assignment by organ system metrics. Adapted from (Vincent et al., 1996).

Organ system subscore component	**0**	**1**	**2**	**3**	**4**
Respiratory PaO_2/FiO_2; mmHg	> 400	≤ 400	≤ 300	≤ 200	≤ 100
Nervous Glasgow Coma Scale (GCS) Score	15	13-14	10-12	6-9	<6
Cardiovascular Mean arterial pressure (MAP); mmHg Vasopressor administration; μg/kg/min	MAP > 70	MAP ≤ 70	Dopamine ≤ 5 or Dobutamine > 0	Dopamine > 5 or Epinephrine ≤ 0.1 or Norepinephrine ≤ 0.1	Dopamine > 15 or Epinephrine > 0.1 or Norepinephrine > 0.1
Liver Bilirubin; mg/dl	< 1.2	1.2-1.9	2.0-5.9	6.0-11.9 ≥ 12	
Coagulation Platelets $\times 10^3$/ml	> 150	< 150	< 100	< 50	< 20
Kidneys Creatinine; mg/dl Urine output; ml/day	< 1.2	1.2-1.9	2.0-3.4	3.5-4.9 Urine output < 500	≥ 5.0 Urine output < 200

The total SOFA score is the sum of the subscores and can range from 0 to 24. Similar to the MSDM Hofford et al. (2022), any SOFA subscore for which data is missing is assumed to be 0. The same practice is extended to the Glasgow Coma Scale (GCS) subscores when determining the total GCS score. Following this, we calculated SOFA scores for each patient state and their correspondence to 30- and 90-day mortality. This correspondence is displayed in Figure 2. SOFA scores displayed a high positive linear correlation to these metrics, with an r-value of 0.970 for 30-day mortality and 0.992 for 90-day mortality. This suggests that SOFA scores are an excellent metric to predict mortality for this dataset.

Crucially for the determination of an appropriate reward function, changes in SOFA score may serve as an indicator for better or poorer actions; on average, patients that survived beyond 90 days following discharge typically had a SOFA score approximately 1.8 points lower at termination state than at the start of their trajectory whereas patients that suffered 90-day mortality often terminated their ICU stays with relatively unchanged SOFA scores.

Thus, given a patient trajectory $\{(s_t, a_t)\}_{t=1}^{T}$, where $s_t^{(\text{SOFA})}$ is the patient's SOFA score corresponding to timestep t, we calculated a SOFA score-based reward component $r_t^{(\text{SOFA})} = e^{s_t^{(\text{SOFA})} - s_{t-1}^{(\text{SOFA})}}$. As

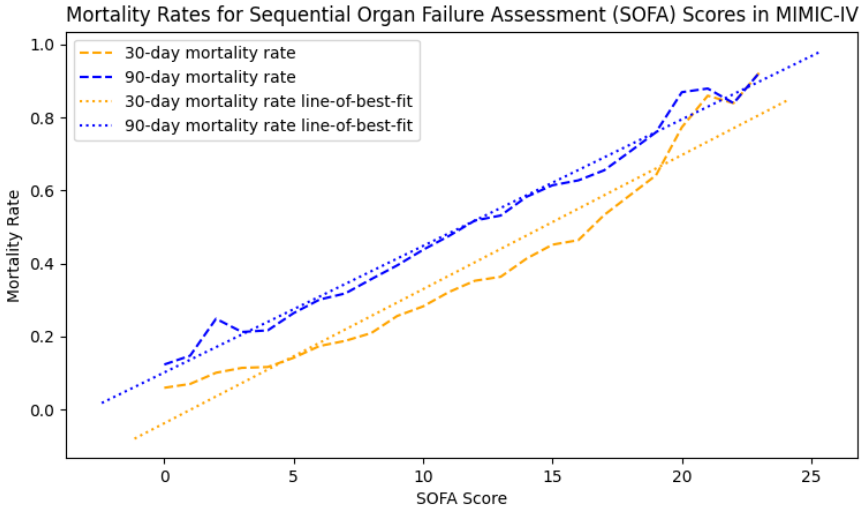

Figure 2: 30- and 90-day mortality versus SOFA scores

such, lower reward is awarded to an action when the subsequent patient state has a higher SOFA score.

In addition to SOFA scores, we incorporated 30- and 90-day mortality with reward components $r^{(30)}$ and $r^{(90)}$, respectively. For positive mortality outcomes in which the patient survives, the corresponding reward component was set to 1 while negative mortality outcomes resulted in 0.

Similar reinforcement learning designs, such as the AI Clinician Komorowski et al. (2018), rely on a γ discount factor to account for long-term rewards; moreover, the discount factor optimizes for trajectory length, with higher short-term rewards resulting from longer trajectories. This can be relevant when trajectory length corresponds to the task objective. In our dataset, however, the lengths of trajectories did not appear to differ significantly by mortality status. For instance, the mean stay length for patients with negative 90-day outcomes (death) is 37.7 ± 50.4 timesteps whereas the mean stay length for patients with positive 90-day outcomes (survival) is 32.1 ± 46.8 timesteps. Since trajectory length did not appear to bear any correlation to patient survival, we used a γ discount factor of 0.

As such, the final reward R_t at a given timestep t is described by the expression

$$R_t = c^{(\text{SOFA})} r_t^{(\text{SOFA})} + c^{(30)} r_t^{(30)} + c^{(90)} r_t^{(90)},$$

where $c^{(\text{SOFA})}$, $c^{(30)}$, and $c^{(90)}$ are arbitrary constants.

4.3 Results

We trained our model for 300 epochs with an 80%-20% training-validation split. The model trained on batches of $2^{10} = 1024$ trajectories during each epoch. We set number of prototypes k to 25 as per the stable prototype amount determined in the ISL previously (Jiang et al., 2023). The importance sampling layer was a 3-layer MLP. Finally, we weighted the objective function $\mathcal{L}$ with $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, and $\lambda_3 = 0.1$. Figure 3 displays aligned and differing trajectory mortality outcomes on the validation dataset for the model. As the majority of patient states received zero total vasopressors, we measured aligned and differing trajectory mortality for a zero-drug policy as well. At baseline, the validation dataset 90-day mortality rate was 38.1%.

The aligned trajectory mortality stabilised at approximately 35.2%, outperforming the expert policy. More notably, the differing trajectory mortality reached as high as 92.3%, suggesting that disagree-

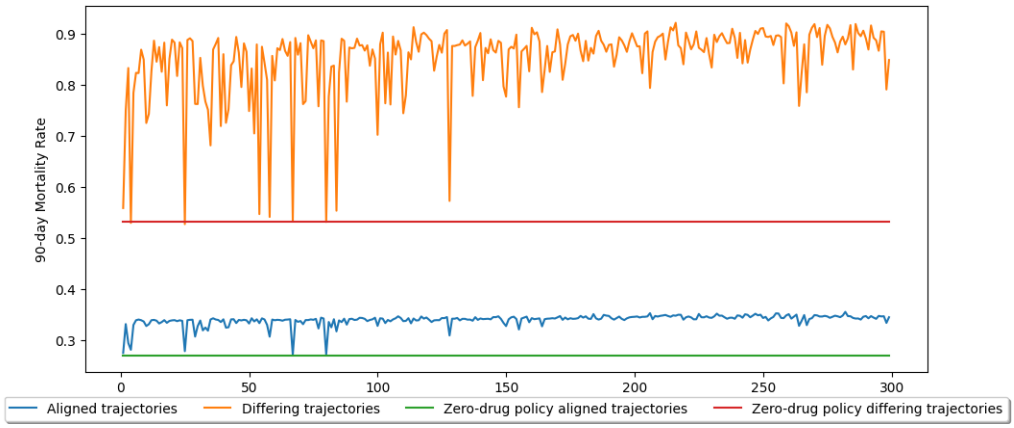

Figure 3: 90-day mortality rates for patient trajectories by alignment to model and to zero-drug policy

ment with the model policy risked a high patient mortality. On average, the model suggested a higher vasopressor dosage than the expert policy.

However, interestingly, the zero-drug policy appears to have outperformed the model on aligned trajectory mortality, although it did perform much worse on differing trajectory mortality. A similar result was noted also in the AI Clinician (Komorowski et al., 2018), where a zero-drug policy yielded a higher median return than most model policies generated. One possible explanation is that zero drugs were most often given to patients that already did not have severe sepsis—those with a positive prognosis. On average, the mean SOFA score for zero-drug patient states 5.9 whereas the mean SOFA score for positive-drug patients was 10.8, mirroring this.

Additionally, although the differing trajectory mortality continues to increase as expected, the aligned trajectory mortality also increases for several epochs before stabilizing at 35%. The cause of this is uncertain. Finally, the tendency of performance to dip notably at times likely reflects a catastrophic forgetting problem.

5 Conclusion

5.1 Discussion

In this paper, we presented the use of reinforcement learning in conjunction with inherent interpretability techniques to explore a novel approach to sepsis patient treatment. Ultimately, the model's policy outperformed the expert policy on a DTR dataset, suggesting the importance of AI integration into healthcare and its ability to improve and advise treatment decisions in high-stakes contexts in an interpretable manner.

5.2 Future Work

Various further steps remain to substantiate the robustness of this methodology. More accurate metrics for policy value—especially off-policy policy evaluation methods—such as boostrapping may provide greater insight into model performance (Komorowski et al., 2018; Jiang & Li, 2016). Evaluation of prototype accuracy as well as an ablation analysis on the prototype layer are imminent.

In regards to data preprocessing, several major refinements remain to be made. Foremost, determining sepsis state less naively, such as via the methods developed for the MSDM in OpenSep (Hofford et al., 2022), is crucial to ensuring the appropriate patient population is being considered, especially in light of challenges in sepsis identification presented by the recent International Sepsis-3 guidelines (Johnson et al., 2018). In doing so, as well as for feature extraction, the utilization of the

mimic-derived dataset, contributed to by various users to develop various derived datasets from MIMIC-IV (Johnson et al., 2023b).

Hand-in-hand with this step is more careful feature selection, as our amount of features (96) greatly exceeds that used in other literature (often 40-50). This will allow for more tailored data cleaning of more important patient features. More sophisticated imputation techniques for missing data (Shortreed et al., 2011), such as K-means imputation, will also become more feasible.

Lastly, exploring memory replay or other techniques to mitigate any potential catastrophic forgetting problem may be critical to ensuring consistent improvements in model performance during training.

ACKNOWLEDGMENTS

I would like to acknowledge Dr. Dongjin Song for his mentorship, guidance, and support throughout the research process. I would also like to thank the Holster Scholar Program at the University of Connecticut for funding, support, and the opportunity.

6 CONFLICTS OF INTEREST

This research was funded by the Holster Scholar Program at the University of Connecticut.

REFERENCES

Should health care demand interpretable artificial intelligence or accept "black box" medicine? *Annals of Internal Medicine*, 172(1):59–60, 2020. doi: 10.7326/M19-2548. URL https://www.acpjournals.org/doi/abs/10.7326/M19-2548. PMID: 31842204.

Guangzhou An, Masahiro Akiba, Kazuko Omodaka, Toru Nakazawa, and Hideo Yokota. Hierarchical deep learning models using transfer learning for disease detection and classification based on small number of medical images. *Sci Rep*, 11(1):4250, March 2021.

Xiaowu Bai, Wenkui Yu, Wu Ji, Zhiliang Lin, Shanjun Tan, Kaipeng Duan, Yi Dong, Lin Xu, and Ning Li. Early versus delayed administration of norepinephrine in patients with septic shock. *Crit Care*, 18(5):532, October 2014.

Michael Biehl, Barbara Hammer, and Thomas Villmann. Prototype-based models in machine learning. *Wiley Interdiscip Rev Cogn Sci*, 7(2):92–111, January 2016.

Liam Byrne and Frank Van Haren. Fluid resuscitation in human sepsis: Time to rewrite history? *Ann Intensive Care*, 7(1):4, January 2017.

Bibhas Chakraborty and Susan A Murphy. Dynamic treatment regimes. *Annu Rev Stat Appl*, 1: 447–464, 2014.

Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches, 2014. URL https://arxiv.org/abs/1409.1259.

John Cull, Robert Brevetta, Jeff Gerac, Shanu Kothari, and Dawn Blackhurst. Epic sepsis model inpatient predictive analytic tool: A validation study. *Crit Care Explor*, 5(7):e0941, June 2023.

Ron Daniels. Surviving the first hours in sepsis: getting the basics right (an intensivist's perspective). *J Antimicrob Chemother*, 66 Suppl 2:ii11–23, April 2011.

Laura Evans, Andrew Rhodes, Waleed Alhazzani, Massimo Antonelli, Craig M Coopersmith, Craig French, Flávia R Machado, Lauralyn Mcintyre, Marlies Ostermann, Hallie C Prescott, Christa Schorr, Steven Simpson, W Joost Wiersinga, Fayez Alshamsi, Derek C Angus, Yaseen Arabi, Luciano Azevedo, Richard Beale, Gregory Beilman, Emilie Belley-Cote, Lisa Burry, Maurizio Cecconi, John Centofanti, Angel Coz Yataco, Jan De Waele, R Phillip Dellinger, Kent Doi, Bin Du, Elisa Estenssoro, Ricard Ferrer, Charles Gomersall, Carol Hodgson, Morten Hylander Møller, Theodore Iwashyna, Shevin Jacob, Ruth Kleinpell, Michael Klompas, Younsuck Koh, Anand Kumar, Arthur Kwizera, Suzana Lobo, Henry Masur, Steven McGloughlin, Sangeeta Mehta, Yatin

Mehta, Mervyn Mer, Mark Nunnally, Simon Oczkowski, Tiffany Osborn, Elizabeth Papathanassoglou, Anders Perner, Michael Puskarich, Jason Roberts, William Schweickert, Maureen Seckel, Jonathan Sevransky, Charles L Sprung, Tobias Welte, Janice Zimmerman, and Mitchell Levy. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med*, 47(11):1181–1247, October 2021.

Carolin Fleischmann, André Scherag, Neill K J Adhikari, Christiane S Hartog, Thomas Tsaganos, Peter Schlattmann, Derek C Angus, Konrad Reinhart, and International Forum of Acute Care Trialists. Assessment of global incidence and mortality of hospital-treated sepsis. current estimates and limitations. *Am J Respir Crit Care Med*, 193(3):259–272, February 2016.

C Fleischmann-Struzek, L Mellhammar, N Rose, A Cassini, K E Rudd, P Schlattmann, B Allegranzi, and K Reinhart. Incidence and mortality of hospital- and ICU-treated sepsis: results from an updated and expanded systematic review and meta-analysis. *Intensive Care Med*, 46(8):1552–1562, June 2020.

James Fraser and Mark Campbell. Teenagers in intensive care: adult or paediatric icu? *Paediatrics and Child Health*, 17(11):454–459, 2007. ISSN 1751-7222. doi: https://doi.org/10.1016/j.paed.2007.09.004. URL https://www.sciencedirect.com/science/article/pii/S1751722207002399.

Scott Fujimoto, David Meger, and Doina Precup. Off-policy deep reinforcement learning without exploration, 2019. URL https://arxiv.org/abs/1812.02900.

Shruti Goradia, Arwa Abu Sardaneh, Sujita W Narayan, Jonathan Penm, and Asad E Patanwala. Vasopressor dose equivalence: A scoping review and suggested formula. *J Crit Care*, 61:233–240, November 2020.

Varun Gulshan, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, Ramasamy Kim, Rajiv Raman, Philip C. Nelson, Jessica L. Mega, and Dale R. Webster. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA*, 316(22):2402–2410, 12 2016. ISSN 0098-7484. doi: 10.1001/jama.2016.17216. URL https://doi.org/10.1001/jama.2016.17216.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

Mackenzie R Hofford, Sean C Yu, Alistair E W Johnson, Albert M Lai, Philip R O Payne, and Andrew P Michelson. OpenSep: a generalizable open source pipeline for SOFA score calculation and sepsis-3 classification. *JAMIA Open*, 5(4):ooac105, December 2022.

Nan Jiang and Lihong Li. Doubly robust off-policy value evaluation for reinforcement learning, 2016. URL https://arxiv.org/abs/1511.03722.

Yushan Jiang, Wenchao Yu, Dongjin Song, Wei Cheng, and Haifeng Chen. Interpretable skill learning for dynamic treatment regimes through imitation. In *2023 57th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6, 2023. doi: 10.1109/CISS56502.2023.10089648.

Alistair Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, Li-wei Lehman, Leo Celi, and Roger Mark. Mimic-iv, a freely accessible electronic health record dataset. *Scientific Data*, 10:1, 01 2023a. doi: 10.1038/s41597-022-01899-x.

Alistair Johnson, Tom Pollard, Jim Blundell, A-Chahin, Brian Gow, Erinhong, Michael Schubert, Kien Dang, Nicolas Paris, Shu98, JackieMe, Angus Zhang, Eric Carlson, Qinyu Zhao, Etheleon, Andrew Barros, Lucas Bulgarelli, Alexmbennett2, Hans0124SG, Peter Szolovits, Sicheng Hao, Christian Porschen, C. V. Cosgriff, Sravan R, Thomas M Ward, Armando Fandango, Bibo Hao, Cx1111, Christopher Yarnell, and Emilyva. Mit-lcp/mimic-code: Mimic code v2.4.0, 2023b. URL https://zenodo.org/record/7604234.

Alistair E W Johnson, Jerome Aboab, Jesse D Raffa, Tom J Pollard, Rodrigo O Deliberato, Leo A Celi, and David J Stone. A comparative analysis of sepsis identification methods in an electronic database. *Crit Care Med*, 46(4):494–499, April 2018.

Alan E Jones, Stephen Trzeciak, and Jeffrey A Kline. The sequential organ failure assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation. *Crit Care Med*, 37(5):1649–1654, May 2009.

Kia Khezeli, Scott Siegel, Benjamin Shickel, Tezcan Ozrazgat-Baslanti, Azra Bihorac, and Parisa Rashidi. Reinforcement learning for clinical applications. *Clin J Am Soc Nephrol*, 18(4):521–523, February 2023.

Mary A. King, Renee I. Matos, Mitchell T. Hamele, Matthew A. Borgman, Luke A. Zabrocki, Samir K. Gadepalli, and Ryan C. Maves. Picu in the micu: How adult icus can support pediatric care in public health emergencies. *Chest*, 161(5):1297–1305, 2022. ISSN 0012-3692. doi: https://doi.org/10.1016/j.chest.2021.12.648. URL https://www.sciencedirect.com/science/article/pii/S0012369222000058.

Matthieu Komorowski, Leo A Celi, Omar Badawi, Anthony C Gordon, and A Aldo Faisal. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med*, 24(11):1716–1720, October 2018.

Thomas Y T Lam, Max F K Cheung, Yasmin L Munro, Kong Meng Lim, Dennis Shung, and Joseph J Y Sung. Randomized controlled trials of artificial intelligence in clinical practice: Systematic review. *J Med Internet Res*, 24(8):e37188, August 2022.

Simon Lambden, Pierre Francois Laterre, Mitchell M Levy, and Bruno Francois. The SOFA score-development, utility and challenges of accurate assessment in clinical trials. *Crit Care*, 23(1):374, November 2019a.

Simon Lambden, Pierre Francois Laterre, Mitchell M Levy, and Bruno Francois. The SOFA score-development, utility and challenges of accurate assessment in clinical trials. *Crit Care*, 23(1):374, November 2019b.

Deyin Liu, Yuanbo Lin Wu, Xue Li, and Lin Qi. Medi-Care AI: Predicting medications from billing codes via robust recurrent neural networks. *Neural Networks*, 124:109–116, 2020a.

Siqi Liu, Kay Choong See, Kee Yuan Ngiam, Leo Anthony Celi, Xingzhi Sun, and Mengling Feng. Reinforcement learning for clinical decision support in critical care: Comprehensive review. *J Med Internet Res*, 22(7):e18477, Jul 2020b. ISSN 1438-8871. doi: 10.2196/18477. URL https://www.jmir.org/2020/7/e18477.

Cristobal Lowery and A. Aldo Faisal. Towards efficient, personalized anesthesia using continuous reinforcement learning for propofol infusion control. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 1414–1417, 2013. doi: 10.1109/NER.2013.6696208.

P. Marik and R. Bellomo. A rational approach to fluid therapy in sepsis. *British Journal of Anaesthesia*, 116(3):339–349, 2016. ISSN 0007-0912. doi: https://doi.org/10.1093/bja/aev349. URL https://www.sciencedirect.com/science/article/pii/S0007091217304427.

P E Marik. The demise of early goal-directed therapy for severe sepsis and septic shock. *Acta Anaesthesiol Scand*, 59(5):561–567, February 2015.

Paul E Marik, Walter T Linde-Zwirble, Edward A Bittner, Jennifer Sahatjian, and Douglas Hansell. Fluid administration in severe sepsis and septic shock, patterns and outcomes: an analysis of a large national database. *Intensive Care Med*, 43(5):625–632, January 2017.

Katrin Reichel Arn Tellmann Björn Ellger Martin Westphal Melanie K Bothe, Luke Dickens and Ahmed A Faisal. The use of reinforcement learning algorithms to meet the challenges of an artificial pancreas. *Expert Review of Medical Devices*, 10(5):661–673, 2013. doi: 10.1586/17434440.2013.827515. URL https://doi.org/10.1586/17434440.2013.827515. PMID: 23972072.

Yao Ming, Panpan Xu, Huamin Qu, and Liu Ren. Interpretable and steerable sequence learning via prototypes. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pp. 903–913, New York, NY, USA, 2019a. Association for Computing Machinery. ISBN 9781450362016. doi: 10.1145/3292500.3330908. URL https://doi.org/10.1145/3292500.3330908.

Yao Ming, Panpan Xu, Huamin Qu, and Liu Ren. Interpretable and steerable sequence learning via prototypes. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining*, KDD '19. ACM, July 2019b. doi: 10.1145/3292500.3330908. URL http://dx.doi.org/10.1145/3292500.3330908.

Christoph Molnar. *Interpretable Machine Learning*. 2 edition, 2022. URL https://christophm.github.io/interpretable-ml-book.

Nosheen Nasir, Bushra Jamil, Shahla Siddiqui, Najeeha Talat, Fauzia A Khan, and Rabia Hussain. Mortality in sepsis and its relationship with gender. *Pak J Med Sci*, 31(5):1201–1206, September 2015.

Jingchao Ni, Zhengzhang Chen, Wei Cheng, Bo Zong, Dongjin Song, Yanchi Liu, Xuchao Zhang, and Haifeng Chen. Interpreting convolutional sequence model by learning local prototypes with adaptation regularization. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021. URL https://api.semanticscholar.org/CorpusID:239993479.

Niranjani Prasad, Li-Fang Cheng, Corey Chivers, Michael Draugelis, and Barbara E Engelhardt. A reinforcement learning approach to weaning of mechanical ventilation in intensive care units, 2017. URL https://arxiv.org/abs/1704.06300.

Jonathan Prest, Thai Nguyen, Tiffany Rajah, Alayna B Prest, Matheni Sathananthan, and Niranjan Jeganathan. Sepsis-Related mortality rates and trends based on site of infection. *Crit Care Explor*, 4(10):e0775, October 2022.

Chanu Rhee, Raymund Dantes, Lauren Epstein, David J Murphy, Christopher W Seymour, Theodore J Iwashyna, Sameer S Kadri, Derek C Angus, Robert L Danner, Anthony E Fiore, John A Jernigan, Greg S Martin, Edward Septimus, David K Warren, Anita Karcz, Christina Chan, John T Menchaca, Rui Wang, Susan Gruber, Michael Klompas, and CDC Prevention Epicenter Program. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. *JAMA*, 318(13):1241–1249, October 2017.

Kristina E Rudd, Sarah Charlotte Johnson, Kareha M Agesa, Katya Anne Shackelford, Derrick Tsoi, Daniel Rhodes Kievlan, Danny V Colombara, Kevin S Ikuta, Niranjan Kissoon, Simon Finfer, Carolin Fleischmann-Struzek, Flavia R Machado, Konrad K Reinhart, Kathryn Rowan, Christopher W Seymour, R Scott Watson, T Eoin West, Fatima Marinho, Simon I Hay, Rafael Lozano, Alan D Lopez, Derek C Angus, Christopher J L Murray, and Mohsen Naghavi. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the global burden of disease study. *Lancet*, 395(10219):200–211, January 2020.

Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, 2019. URL https://arxiv.org/abs/1811.10154.

Arjun Sharma, Mohit Sharma, Nicholas Rhinehart, and Kris M. Kitani. Directed-info gail: Learning hierarchical policies from unsegmented demonstrations using directed information, 2019. URL https://arxiv.org/abs/1810.01266.

Susan M Shortreed, Eric Laber, Daniel J Lizotte, T Scott Stroup, Joelle Pineau, and Susan A Murphy. Informing sequential clinical decision-making through reinforcement learning: an empirical study. *Mach Learn*, 84(1-2):109–136, July 2011.

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, Richard S Hotchkiss, Mitchell M Levy, John C Marshall, Greg S Martin, Steven M Opal, Gordon D Rubenfeld, Tom van der Poll, Jean-Louis Vincent, and Derek C Angus. The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA*, 315(8):801–810, February 2016.

Zhaohong Sun, Wei Dong, Haomin Li, and Zhengxing Huang. Adversarial reinforcement learning for dynamic treatment regimes. *Journal of Biomedical Informatics*, 137:104244, 2023. ISSN 1532-0464. doi: https://doi.org/10.1016/j.jbi.2022.104244. URL https://www.sciencedirect.com/science/article/pii/S1532046422002490.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, second edition, 2018. URL http://incompleteideas.net/book/the-book-2nd.html.

Danny VanValkinburgh, Connor C Kerndt, and Muhammad F Hashmi. Inotropes and vasopressors. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), January 2024.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023. URL https://arxiv.org/abs/1706.03762.

J L Vincent, R Moreno, J Takala, S Willatts, A De Mendonça, H Bruining, C K Reinhart, P M Suter, and L G Thijs. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. on behalf of the working group on Sepsis-Related problems of the european society of intensive care medicine. *Intensive Care Med*, 22(7):707–710, July 1996.

Jean-Louis Vincent, Yasser Sakr, Charles L Sprung, V Marco Ranieri, Konrad Reinhart, Herwig Gerlach, Rui Moreno, Jean Carlet, Jean-Roger Le Gall, Didier Payen, and Sepsis Occurrence in Acutely Ill Patients Investigators. Sepsis in european intensive care units: results of the SOAP study. *Crit Care Med*, 34(2):344–353, February 2006.

Jason Waechter, Anand Kumar, Stephen E Lapinsky, John Marshall, Peter Dodek, Yaseen Arabi, Joseph E Parrillo, R Phillip Dellinger, Allan Garland, and Cooperative Antimicrobial Therapy of Septic Shock Database Research Group. Interaction between fluids and vasoactive agents on mortality in septic shock: a multicenter, observational study. *Crit Care Med*, 42(10):2158–2168, October 2014.

Lu Wang, Wei Zhang, Xiaofeng He, and Hongyuan Zha. Supervised reinforcement learning with recurrent neural network for dynamic treatment recommendation, 2018. URL https://arxiv.org/abs/1807.01473.

Lu Wang, Wenchao Yu, Xiaofeng He, Wei Cheng, Martin Renqiang Ren, Wei Wang, Bo Zong, Haifeng Chen, and Hongyuan Zha. Adversarial cooperative imitation learning for dynamic treatment regimes. *Proceedings of The Web Conference 2020*, 2020. URL https://api.semanticscholar.org/CorpusID:215829171.

Hussain A Younis, Taiseer Abdalla Elfadil Eisa, Maged Nasser, Thaeer Mueen Sahib, Ameen A Noor, Osamah Mohammed Alyasiri, Sani Salisu, Israa M Hayder, and Hameed Abdulkareem Younis. A systematic review and Meta-Analysis of artificial intelligence tools in medicine and healthcare: Applications, considerations, limitations, motivation and challenges. *Diagnostics (Basel)*, 14(1), January 2024.

Chao Yu, Jiming Liu, and Hongyi Zhao. Inverse reinforcement learning for intelligent mechanical ventilation and sedative dosing in intensive care units. *BMC Med Inform Decis Mak*, 19(Suppl 2): 57, April 2019.

Yutao Zhang, Robert Chen, Jie Tang, Walter F. Stewart, and Jimeng Sun. Leap: Learning to prescribe effective and safe treatment combinations for multimorbidity. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pp. 1315–1324, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450348874. doi: 10.1145/3097983.3098109. URL https://doi.org/10.1145/3097983.3098109.