Journal of
Clinical Medicine

MDPI

Article

Development of a Reinforcement Learning Algorithm to Optimize Corticosteroid Therapy in Critically Ill Patients with Sepsis

Razvan Bologheanu [1,2,*,†], Lorenz Kapral [2,†], Daniel Laxar [2], Mathias Maleczek [1,2], Christoph Dibiasi [1], Sebastian Zeiner [1], Asan Agibetov [1], Ari Ercole [3], Patrick Thoral [4], Paul Elbers [4], Clemens Heitzinger [5] and Oliver Kimberger [1,2]

[1] Department of Anaesthesia, Intensive Care Medicine and Pain Medicine, Medical University of Vienna, 1090 Vienna, Austria
[2] Ludwig Boltzmann Institute for Digital Health and Patient Safety, 1090 Vienna, Austria
[3] Centre for Artificial Intelligence in Medicine, University of Cambridge, Cambridge CB2 0QQ, UK
[4] Department of Intensive Care Medicine, Laboratory for Critical Care Computational Intelligence, Amsterdam UMC, Vrije Universiteit, 1081 HV Amsterdam, The Netherlands
[5] Institute of Analysis and Scientific Computing, Department of Mathematics and Geoinformation, Technical University of Vienna, 1040 Vienna, Austria
* Correspondence: razvan.bologheanu@meduniwien.ac.at
† These authors contributed equally to the work.

Abstract: Background: The optimal indication, dose, and timing of corticosteroids in sepsis is controversial. Here, we used reinforcement learning to derive the optimal steroid policy in septic patients based on data on 3051 ICU admissions from the AmsterdamUMCdb intensive care database. Methods: We identified septic patients according to the 2016 consensus definition. An actor-critic RL algorithm using ICU mortality as a reward signal was developed to determine the optimal treatment policy from time-series data on 277 clinical parameters. We performed off-policy evaluation and testing in independent subsets to assess the algorithm's performance. Results: Agreement between the RL agent's policy and the actual documented treatment reached 59%. Our RL agent's treatment policy was more restrictive compared to the actual clinician behavior: our algorithm suggested withholding corticosteroids in 62% of the patient states, versus 52% according to the physicians' policy. The 95% lower bound of the expected reward was higher for the RL agent than clinicians' historical decisions. ICU mortality after concordant action in the testing dataset was lower both when corticosteroids had been withheld and when corticosteroids had been prescribed by the virtual agent. The most relevant variables were vital parameters and laboratory values, such as blood pressure, heart rate, leucocyte count, and glycemia. Conclusions: Individualized use of corticosteroids in sepsis may result in a mortality benefit, but optimal treatment policy may be more restrictive than the routine clinical practice. Whilst external validation is needed, our study motivates a 'precision-medicine' approach to future prospective controlled trials and practice.

Keywords: sepsis; corticosteroids; outcomes; artificial intelligence; reinforcement learning

Citation: Bologheanu, R.; Kapral, L.; Laxar, D.; Maleczek, M.; Dibiasi, C.; Zeiner, S.; Agibetov, A.; Ercole, A.; Thoral, P.; Elbers, P.; et al. Development of a Reinforcement Learning Algorithm to Optimize Corticosteroid Therapy in Critically Ill Patients with Sepsis. *J. Clin. Med.* **2023**, *12*, 1513. https://doi.org/10.3390/jcm12041513

Academic Editor: Sergio Ruiz-Santana

Received: 11 January 2023
Revised: 30 January 2023
Accepted: 6 February 2023
Published: 14 February 2023

1. Introduction

Sepsis represents a significant cause of morbidity and is responsible for 11 million deaths globally each year [1]. Defined as "life-threatening organ dysfunction caused by a dysregulated host response to infection", sepsis is an umbrella term for a heterogeneous syndrome with many distinct phenotypes and wide variation in outcomes [2,3]. As a result, clinical trials have provided conflicting evidence concerning the benefit of specific therapies beyond source control, antibiotics, and maintenance of tissue perfusion [4,5].

Corticosteroids have been extensively investigated as a therapeutic option for sepsis ever since Cook et al. first advocated their use seven decades ago, but uncertainty regarding

their optimal use nevertheless persists [6]. More recently, the case for corticosteroids in sepsis was based on the evidence of adrenal insufficiency accompanying critical illness [7]. Since diagnostic criteria for adrenal insufficiency are missing, identifying patients that should receive corticosteroids is challenging [7]. In addition, several studies have found that corticosteroids can lead to a faster resolution of shock but provided equivocal results concerning survival [8–10].

Currently, guidelines for the management of sepsis suggest using corticosteroids in septic patients with ongoing vasopressor requirement [5]. However, the optimal treatment regimen, particularly timing, duration, and dose of corticosteroids, is not known, and the clinical significance of potential adverse effects of corticosteroid therapy is unclear [5]. Identifying patients who are likely to benefit from corticosteroids is essential and attempts at personalizing corticosteroid therapy using novel approaches, such as machine learning and transcriptomics, have been reported [11,12].

Since interventional studies in sepsis are challenging due to the extreme heterogeneity of its phenotypes, machine learning could represent a complementary evaluation method for specific treatments using observational data. In essence, the aim is to construct an algorithm that can exploit clinician variances in treatment policy over a large dataset in a way that it is possible to find the effects of the treatment on similar patients at a given time. Reinforcement learning, one of the three primary machine learning branches, can be applied to this type of problem [13,14]. Reinforcement learning algorithms can serve as the foundation for decision support tools in intensive care, where decision making is based on sequential, highly granular data [15,16]. In brief, such algorithms attempt to find an 'optimal' policy that maximizes some reward function (for example survival), given a particular treatment strategy with a comprehensive description of the state of the patient at that time [13]. In the present study, we describe the development of a reinforcement learning algorithm to find the optimal approach to corticosteroid therapy in septic patients based on high-resolution clinical data from an intensive care database.

2. Materials and Methods

2.1. Data Sources and Data Processing

All data were queried from the AmsterdamUMCdb database. Approval was obtained for 3rd party re-use of AmsterdamUMCdb data for research from its steering group, and the research was conducted according to the data use agreement. Such a study of deidentified data is not subject to the need for ethical review. The ethical approvals for the AmsterdamUMCdb have been previously described [17]. AmsterdamUMCdb contains high-resolution clinical data related to 23,106 ICU admissions of 20,109 patients from 2003 to 2016 [17]. Patients with sepsis were identified based on the Sepsis-3 criteria2 Accordingly, patients with new organ dysfunction as indicated by either a SOFA score ≥ 2 at admission or an increase of 2 points or more in the SOFA score during the ICU stay, in the context of suspected infection as described in Supplemental Table S1, were included in the sepsis cohort [2,18,19]. Patients aged <18 years at the time of the ICU admission and patients who stayed in the ICU less than 24 h were excluded. The onset of the septic episode was considered the day the change in the SOFA score occurred and patients remained in the sepsis cohort until discharge or death.

In total, 281 variables were extracted, of which 277 input variables were coded as a multidimensional time series with a time resolution of 24 h. Every ICU day was considered separately, and only current measurements available at that timepoint were included in each data point. Only numeric variables represented in more than 2% of the data points were included. The imbalance resulting from missing data and the variable sampling rate were addressed by preprocessing: missing laboratory values were imputed using forward fill, while missing medication doses were set to 0. Overall, 17.93% of all input values were imputed. Numeric data were normalized to values between -1 and $+1$; for frequently sampled parameters (e.g., heart rate), the mean, the maximum, the minimum, and standard deviation were calculated, and for others (e.g., continuously administered drugs), the sum,

i.e., the 24 h cumulative dose, was used as input instead. Therefore, the final number of extracted parameters increased to 379. The complete list of input features is provided in Supplemental Table S2.

2.2. Algorithm Development

Reinforcement learning is based on modeling a virtual decision-making 'agent' interacting with its environment described by a set of continuous states; the interaction between the agent and the environment predetermined as the action space (in this case, the finite number of treatment choices). At each step, the agent chooses an action, and the environment changes its state, returning a reward. The reward signal is used to train the agent, which gradually learns an optimal policy that maximizes return [20].

We implemented a reinforcement learning algorithm, consisting of two distinct neural networks, based on the Markov Decision Process using the temporal difference actor-critic method able to suggest the optimal corticosteroid dose for each septic patients by retrospectively analyzing clinical data [20–22]. The dataset was randomly split into a training set, consisting of 70% of all patients, and two smaller datasets for validation (20%) and testing (10%) (Figure 1). The algorithm was trained on trajectories of successive patient states, where a state corresponded to a vector of all features within a 24 h period, other than mortality and the administered corticosteroid dose. The reward signal associated with each transition was related to the ICU mortality. The action space consisted of five discrete actions, defined by converting the cumulative 24 h dose of systemic corticosteroids to the equivalent dose of hydrocortisone and binning the resulting values: the null ('no corticosteroids') action and four dose ranges: 1–100 mg, 101–200 mg, 201–300 mg, and over 300 mg hydrocortisone [23]. A detailed description of the reinforcement learning model is provided in Supplemental File S1 and Supplemental Figure S1. The reinforcement learning algorithm was built using the TensorFlow 2.7 Python library [24].

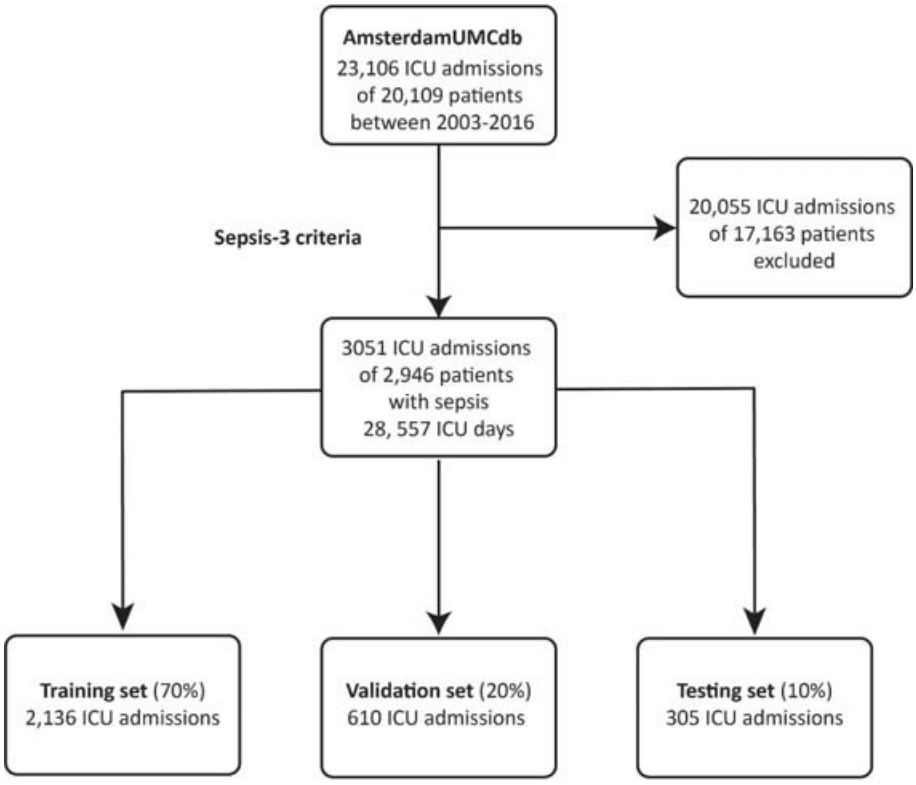

Figure 1. The Sepsis Cohort. Patients with sepsis from the AmsterdamUMC database were identified using the Sepsis-3 criteria. The sepsis cohort was randomly split in three distinct subsets used for training, evaluating, and testing the reinforcement learning algorithm.

2.3. Evaluation of the Algorithm

The reinforcement learning algorithm was initially evaluated by comparing the actual reward after concordant actions, i.e., when the actual treatment and the corticosteroid dose suggested by the agent were identical, with the reward after discordant actions in the testing set.

The performance of such reinforcement learning algorithms could not be directly evaluated by measuring the received reward of each action, since the reinforcement learning (evaluation) policy was different from the clinician (behavior) policy and the actual reward represented the performance of the clinician policy. We implemented a high-confidence off-policy evaluation (HCOPE) of the algorithm, a statistical method which compares the performance of the algorithm's policy with a baseline, the performance of the clinician policy, and computes the probability that the algorithm's policy has a performance below this baseline to select the best performing model. Using the clinician policy, a set of trajectories was generated and used to lower-bound the performance of the evaluation policy. The high-confidence off-policy evaluation (HCOPE) allowed for determining whether the 95% lower bound of the expected reward of the policy of the reinforcement learning agent exceeded the average reward of the clinician policy, i.e., the actual treatment the patients received [25,26].

Finally, we estimated the relative importance of each variable using a Layer-wise Relevance Propagation algorithm and ranked the input features of the RL algorithm according to their contribution to the agent's decision [27]. To allow for comparison between the relevance of the input features of agent's policy and the clinical practice, we developed a random forest model using the Scikit-learn Python library that predicts the clinicians' policy, simulating the clinician behavior, and we ranked the clinical variables supporting the average clinician behavior according to the parameters of the fitted model [28].

3. Results

A total of 3051 ICU admissions at the Amsterdam UMC corresponding to 2946 distinct patients were included (Figure 1).

Repeated admissions to the ICU, both remote and during the same hospital stay, were included if they met the sepsis definition and were analyzed as independent ICU stays. 1395 admissions were associated with vasopressor use and lactate values >2 mmol/l during the ICU stay, therefore meeting the criteria for septic shock. The cumulative length of stay from the onset of sepsis until ICU discharge was 28,557 days corresponding to as many data points. The training dataset comprised 2136 randomly selected ICU admissions, leaving a total of 610 and 305 admissions in the evaluation and testing datasets, respectively (Figure 1). Patients' characteristics are summarized in Table 1.

The relative error of the actor-critic model decreased over the training steps and converged after 250 epochs at 0.044 of the initial relative error (Figure 2a). The concordance between the virtual agent's action and the retrospective action by ICU physicians started at 22%, which was the expected value considering the dimension of the action space (five possible actions). The overall agreement between the virtual agent and the human clinicians reached 63% after convergence (Figure 2c). Similarly, the probabilities of choosing each action from the action space were equal initially. Over the training epochs, the virtual agent increasingly tended towards withholding corticosteroids. After convergence, in 65% of ICU days, the agent chose to withhold corticosteroids, and in patients where corticosteroids were prescribed, the suggested dose was low (Figure 2b). In contrast, the human clinicians prescribed corticosteroids in 45% of data points. Although the virtual agent displayed a tendency towards passive behavior, in 49% of the cases where the agent chose to administer glucocorticoids, the ICU physicians acted concordantly.

Table 1. Summary of patients' characteristics. Each ICU admission is considered separately.

Characteristics	Summary (Total)	Summary (Survivors)	Summary (Non-Survivors)
Total number of ICU admissions	3051	2336	715
Male sex, No. (%)	1758 (57.6%)	1353 (57.9%)	405 (56.6%)
Age group (years), No. (%)	–	–	–
18–39	342 (11.2%)	303 (12.9%)	39 (5.4%)
40–49	322 (10.5%)	265 (11.3%)	57 (7.9%)
50–59	518 (17.0%)	414 (17.7%)	104 (14.5%)
60–69	757 (24.8%)	591 (25.2%)	166 (23.2%)
70–79	709 (23.2%)	506 (21.6%)	203 (28.3%)
>80	403 (13.2%)	257 (11%)	146 (20.4%)
Highest SOFA score during the ICU stay, Median (IQR)	10 (6)	9 (6)	13 (6)
Sofa score at sepsis onset, Median (IQR)	9 (6)	8 (5)	11 (7)
Septic shock, No. (%)	1395 (45.7%)	845 (36.1%)	550 (76.9%)

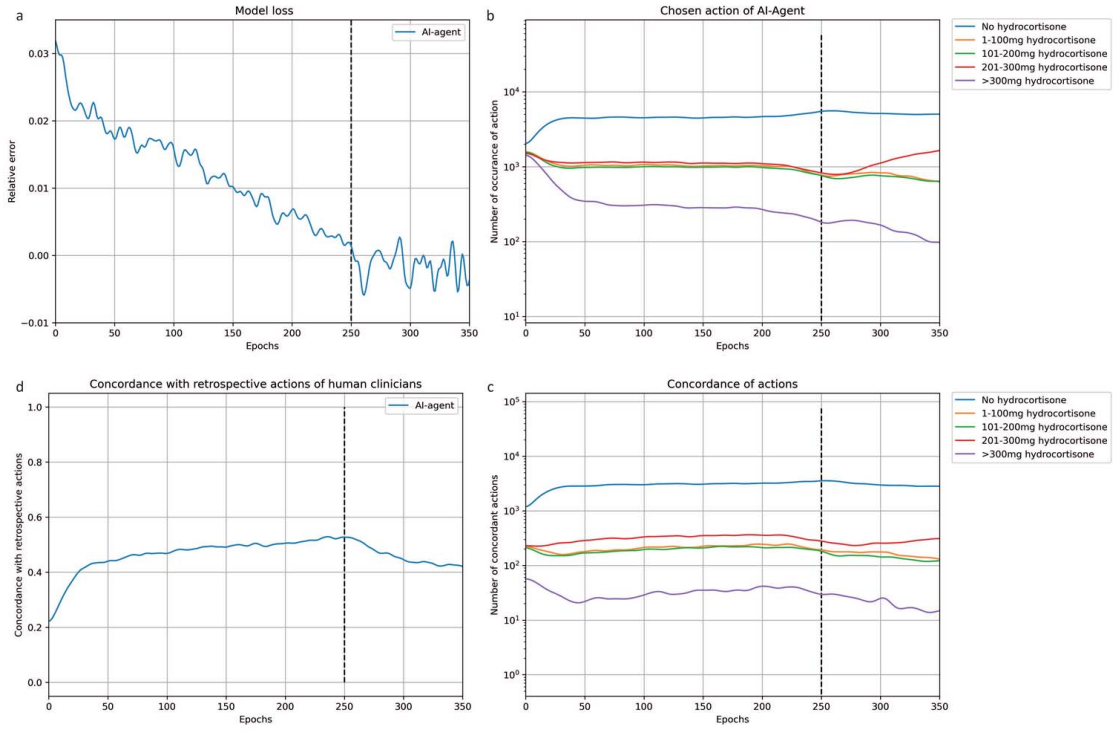

Figure 2. Training process of the virtual agent. Figure 2 shows how the performance and the behavior of the RL agent changed during the training process. On the X-axis, the number of epochs, i.e., how many times the algorithm had worked through the learning dataset, since the beginning of the training is displayed. The vertical dotted line marks the end of the training process. (**a**) The decrease in the relative error, which reflects the accuracy of the model's output, during the training process. (**b**) The number of occurrences for each action suggested by the algorithm during training is displayed in the (**b**). All five possible actions are equally represented at the beginning of the training. After 50 epochs, the algorithm's tendency to withheld corticosteroids becomes obvious. (**c**) The increasing overall agreement between the RL policy and the actual historic treatment. (**d**) The number of occurrences when agreement between the RL policy and the retrospective treatment was reached is displayed across the five possible actions in (**b**).

In the testing dataset, the treatment suggested by the virtual agent matched the retrospective action by ICU physicians in 59% of the data points. The agent's tendency to prescribe less corticosteroids was also confirmed in the testing dataset: corticosteroids were withheld in 62% of the ICU days, compared to 52% according to the ICU physicians. Accordingly, the average daily corticosteroid dose prescribed by the virtual agent was lower (Figure 3). Both ICU physicians and the RL agent tended to prescribe corticosteroids in the early phase of the septic episode and corticosteroid use dropped sharply after 10 days (Figure 3).

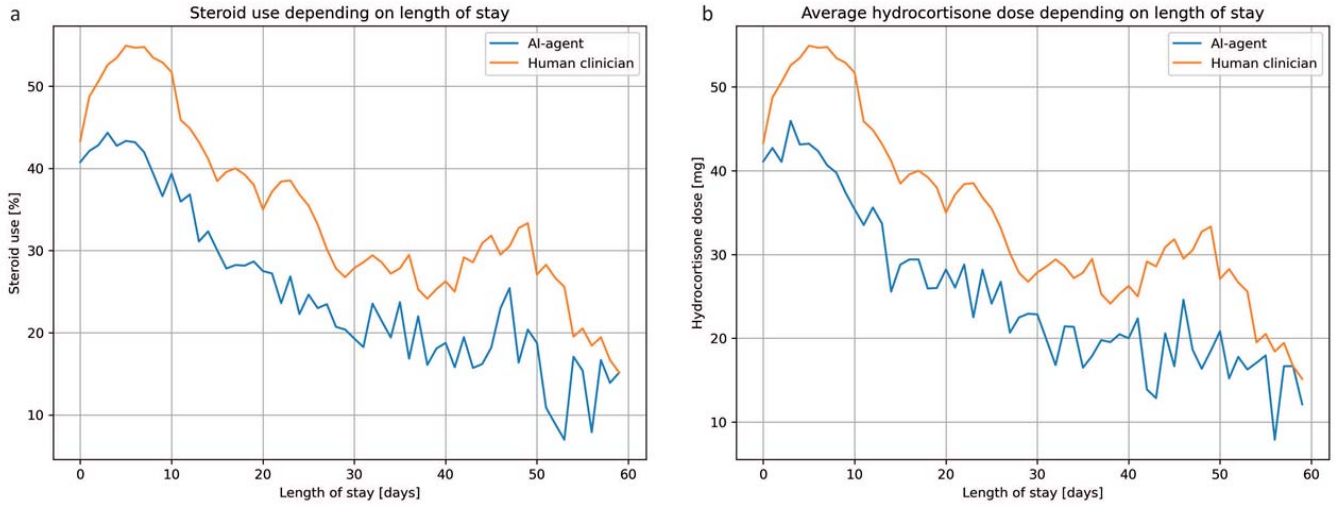

Figure 3. Comparison of corticosteroid use between ICU physicians and the RL agent. Use of corticosteroids as percentage of patients receiving corticosteroids (**a**) and average cortisone dose (**b**) is compared between the historic treatment in the ICU and the RL policy after adjusting for the ICU length of stay. Both ICU physicians and the RL agent tend to prescribe corticosteroids during the early phase of the septic episode. Notably, the RL policy is more restrictive compared to the actual treatment the patients received.

The ratio between the reward of the agent's policy and the clinicians' policy increased over the training process and high-confidence off-policy evaluation (HCOPE) demonstrated that the 95% lower bound of the expected average reward for the agent's policy was higher compared to the average reward for the historical decisions by clinicians after 200 epochs (Figure 4). Accordingly, the normalized expected mortality rate decreased and was lower than 0.7. Overall, when patients from the testing set received the same glucocorticoid therapy as suggested by the RL agent, mortality was lower: the mortality across all ICU days, i.e., the ICU days that eventually result in patient's death, when the decisions made by the RL agent and the ICU physician were identical was 22.38% compared to 28.33% in case the actions were different. This finding was consistent both when the RL agent withheld corticosteroids (25.85% of the data points compared to 32.22%) and when the RL agent suggested using corticosteroids (33.02% of the data points compared to 34.27%).

We modeled the retrospective treatment policy by the ICU physicians using a random forest model that predicted the clinicians' treatment decisions. The micro-average multiclass Area under the Receiver Operator Characteristic Curve for the random forest model was 0.8 (Supplemental Figure S2). The most relevant input features underlying the decisions of the reinforcement learning algorithm and the random forest model, respectively, are presented in Supplemental Tables S3 and S4 and Supplemental Figure S3. Both algorithms relied on vital parameters and laboratory values to determine the optimal treatment policy. However, vasopressor use and PEEP were distinctly more relevant for the clinician policy. Accordingly, although the reinforcement learning agent was consistently more restrictive

compared to human clinicians, the difference is more obvious in patients who met the criteria for septic shock (Figure 5).

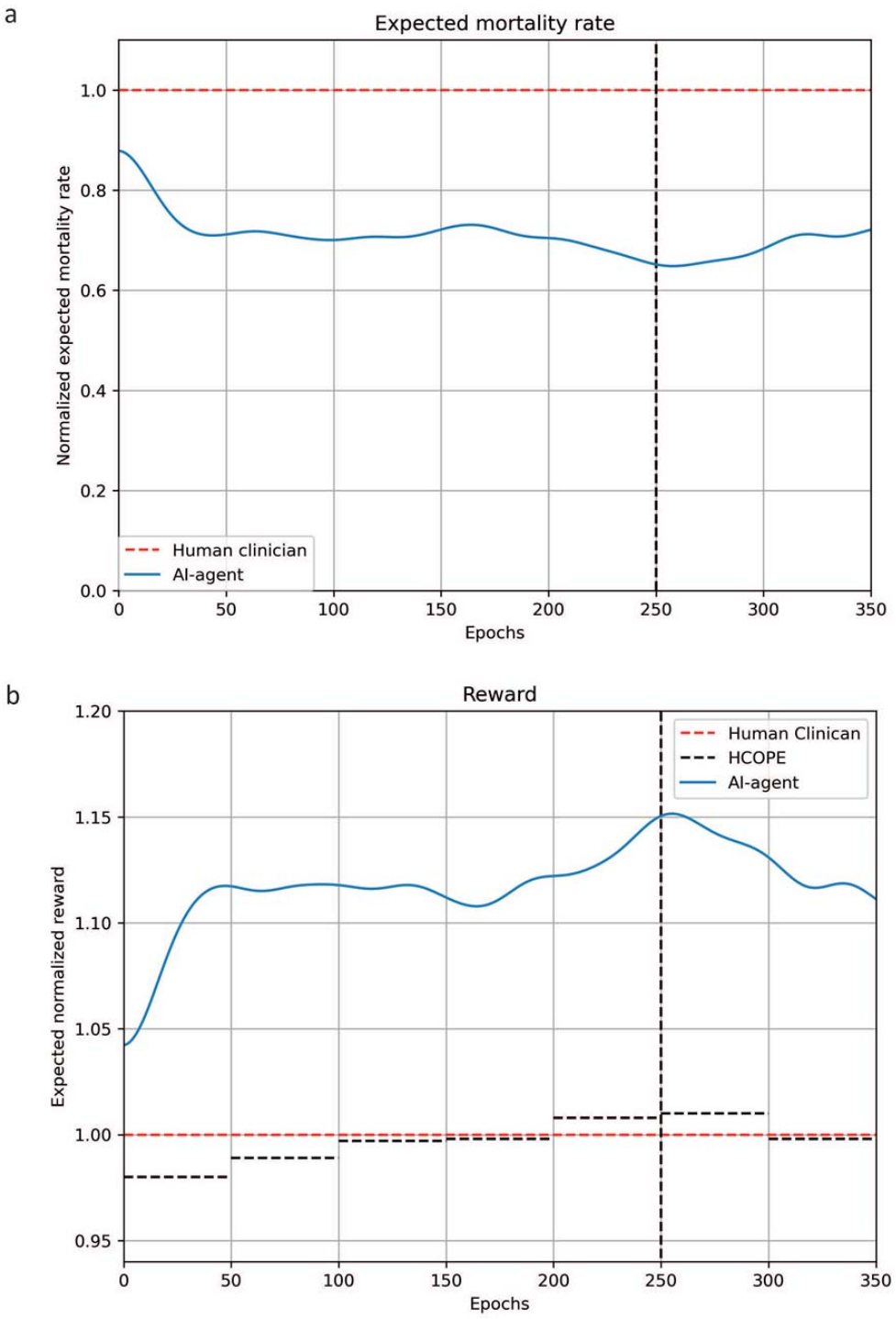

Figure 4. Comparison between the evaluation (RL) policy and the behavior policy (the actual treatment). (**a**) The change in the normalized expected mortality rate across training epochs, (i.e., the number of iterations or how many times the algorithm had worked through the learning dataset, since the beginning of the training) is represented in Figure 5a. (**b**) The 95% lower bound of the normalized expected reward of the RL policy (black dotted line) determined by high-confidence off-policy evaluation compared to the estimated reward of the clinician policy (red dotted line) is shown in (**b**).

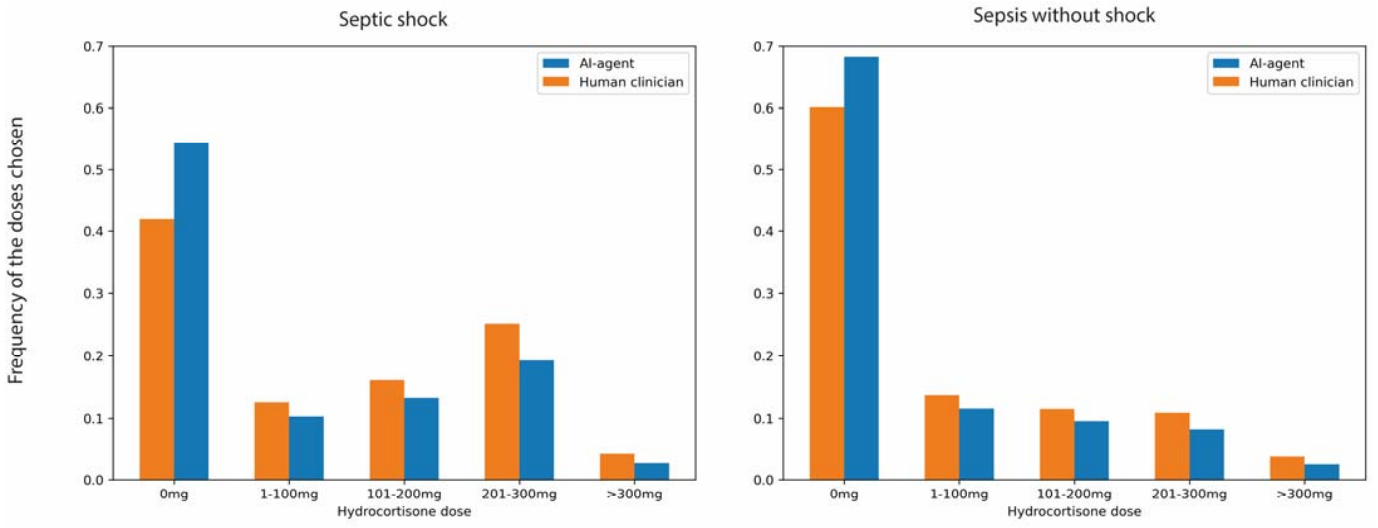

Figure 5. Comparison between the RL and physician policy in patient states grouped by septic shock criteria.

4. Discussion

We present a reinforcement learning algorithm trained to optimize the corticosteroid treatment strategy for a specific patient state in critically ill patients with sepsis. The novelty of our approach is that it potentially enables an individualized therapy to improve a highly relevant outcome based on clinical parameters routinely collected in the ICU. The goal of our reinforcement learning algorithm, determined by the reward signal, was to minimize mortality. Indeed, in the testing dataset, ICU mortality was the lowest in patients who received a treatment identical to the action suggested by the algorithm. Off-policy evaluation confirmed that the algorithm performed well within the given environment and even outperformed the clinician policy in the validation dataset.

Currently, the rationale for corticosteroids in sepsis is based on several studies suggesting faster resolution of shock in septic patients who require vasopressors despite adequate fluid resuscitation [5]. While earlier studies showed a mortality benefit, this was not consistently confirmed in subsequent trials [8,9,29–32]. This led to frequent changes in the clinical practice to accommodate new, often conflicting evidence, which have been likened to a "swinging pendulum" situation [30]. The most recent guidelines for the treatment of sepsis suggest corticosteroids as early as 4 h after the initiation of treatment in patients who require vasopressors. In the testing subset of our sepsis cohort, where 45.7% of patients met the criteria for septic shock, corticosteroids were suggested by the virtual agent in 38% of the ICU days. Conversely, ICU physicians used corticosteroids in 48% of the data points, yet only in 49% of the cases where the reinforcement learning agent suggested using corticosteroids, the actual treatment prescribed in the ICU was concordant. This difference may be a result of at least two factors. First, the reward signal used for training was related to the mortality and the reinforcement learning agent aimed to maximize survival. Second, corticosteroids have been historically reserved for patients who require more vasopressors and have higher severity of disease and, therefore, worse outcomes. Indeed, the random forest model we developed to simulate decision making by the ICU physicians showed that blood pressure and vasopressor use were most consistently associated with corticosteroid use. Furthermore, due to the retrospective nature of our study, we expected that the association between higher severity scores and corticosteroid use in the database would translate in a bias of the RL policy towards the null action.

We identified patients from the database with sepsis algorithmically, and this required a pragmatic operationalization of the Sepsis-3 criteria, using a data-driven approach, instead of relying on coding data to be defined [2,19]. This method has been used before and has the advantage of being more reliable; more reproducible; and therefore, appropriate for

epidemiological or database studies [33]. These operational criteria can provide consistent estimates of the sepsis incidence over longer periods, despite its inherent limitations, such as the assumptions about suspected infections being confirmed, pre-admission organ function, and the impact of the caregivers' decisions on the SOFA score [33–35].

Although traditionally, artificial intelligence algorithms have been often compared to a black box, several methods are available to provide insight into which variables contributed most to the algorithmic decisions [36]. We ranked the input features based on relevance, showing that our model was explainable and valid from a clinical standpoint and that the agent relied on plausible clinical variables to make its decisions. If the random forest model accurately simulates the decision-making process by ICU physicians, comparing the relative relevance of the input features between the reinforcement learning algorithm and the random forest model can reveal how a treatment policy can be developed to maximize ICU survival contrasts with actual care. Unlike current clinical practice, where refractory shock is the single most important factor considered to prescribe corticosteroids, vasopressor requirements and lactate only had a limited influence on the reinforcement learning policy while being highly relevant for the clinicians' policy. Similarly, the time elapsed since the onset of sepsis ranked distinctly higher amongst input features for the historical treatment by ICU physicians compared to the reinforcement learning treatment. These findings confirm the usual practice of prescribing corticosteroids early for patients in septic shock [5]. Interestingly, the machine learning policy resulted in a similar corticosteroid use pattern, characterized by an abrupt fall in steroid use after the 10th day since onset without explicitly relying as much on the time elapsed from the onset of sepsis. Conversely, total protein in cerebrospinal fluid (CSF) and the standard deviation of the heart rate ranked higher among the input parameters of the virtual agent only. It might seem surprising that a parameter that is rarely sampled is highly relevant for the output of the algorithm. Although corticosteroids are recommended for prevention of neurological sequelae in patients with bacterial meningitis, they have no effect on mortality [37]. Alternatively, lumbar puncture might be performed as a part of the work-up in patients with fever of unknown origin and subtle neurological symptoms [38]. In either case, since non-missing values are highly suggestive of a neurological diagnosis, informative missingness might explain its relevance for the reinforcement learning policy.

Arterial blood pressure, leucocyte count, serum sodium, and blood glucose levels were similarly influential in both algorithms. These findings seem biologically plausible, given the essential role of corticosteroids in regulating glucose metabolism and electrolyte homeostasis [39]. Corticosteroids also potentiate the effects of catecholamines and mobilize neutrophils, leading to leukocytosis and neutrophilia [40,41]. It is reasonable that clinical variables related to the physiological effects of corticosteroids could help guide therapy in septic patients by accurately predicting their effects in specific patient states. However, these results must be interpreted cautiously. The method we used to rank input variables estimates the overall contribution of all variables to the output of the model. Furthermore, unlike traditional statistical modeling, neural networks are less suitable for determining relationships between variables. Finally, all input variables were normalized between -1 and +1, and the relation between the normalized values, the actual values, and the reference range for each variable was determined by the variable's distribution and is not obvious or readily interpretable for clinicians.

We acknowledge several limitations of our study. First, we used a single database to develop our algorithm and our findings have not been externally validated, which considerably limits the clinical applicability of the model. Like most of the artificial intelligence research in the intensive care, our study is in the prototype phase, and broad implementation remains a distant goal [42] Although machine learning models could be transferred across ICUs, moving these models to the bedside proves challenging [42]. Artificial intelligence holds great promise to enhance the practice of intensive care and the management of sepsis in the ICU; however, the current state of AI in intensive care does not support its routine use due to regulatory reasons, but also because uncertainty

surrounds how these models could be included in daily practice, and good prospective studies still need to be included. Second, data used to train and test the model originate from a single medical center over several years. Changes in the best care practices over time and differences between local policies concerning ICU admission and sepsis management might result in relevant heterogeneity of the sepsis cohort and the outcomes. However, the aim of this study was to create an algorithm that can exploit these differences to derive an optimal treatment policy by analyzing several different suboptimal policies. Third, data were anonymized, and in the process, all notes were removed. Consequently, we could not account for the withdrawal of life-sustaining therapies. Fourth, by using a 24 h step to model the patients' trajectories, our model artificially creates data points that encompass more data than are available to the clinician at any given time. We considered the time resolution of 24 h and the action space defined as the cumulative 24 h dose of corticosteroids due to several reasons, since this approach allowed us to compare different treatment regimens, using different substances, doses, and intervals. Furthermore, in our experience, therapy goals and some therapeutic measures for the next 24 h are defined during the ICU rounds, once daily. Therefore, modelling clinical data as time-series data with a resolution of 24 h resembles, to some extent, clinical practice.

Decision making in the ICU typically takes place during the once-daily rounds and the cumulative 24 h dose allows for different treatment regimens to be compared regardless of substance and timing. Finally, we analyzed all clinical data from onset of sepsis until discharge from the ICU, which most likely covers a significantly longer period than the duration of the septic shock. However, clearly delineating between the acute critical illness, and subsequent organ dysfunction and persistent critical illness does not seem feasible in the context of the present study.

5. Conclusions

We developed and evaluated a reinforcement learning algorithm that used clinical data to derive the optimal corticosteroid therapy aimed at improving mortality in patients with sepsis. The algorithm performed well in the testing dataset, and the reinforcement learning policy was associated with a lower mortality than the clinician's policy. Due to the exploratory nature of our work, future research focusing on external validation of the model is required before prospective evaluation at the bedside. Our model suggests that a more targeted and individualized, reinforcement learning-driven approach to corticosteroids is possible and motivates prospective evaluation of treatment scenarios beyond refractory shock.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/jcm12041513/s1, Supplemental Table S1: Diagnosis of sepsis. Supplemental Table S2: Input features included in development of the algorithm. Supplemental Figure S1: Development of the RL Algorithm. Supplemental Figure S2: Micro-average ROC curve of the SVM algorithm. Supplemental Table S3: The most relevant predictors of the clinicians' policy according to the SVM algorithm ordered from the lowest to highest rank. Supplemental File S1: The 20 most relevant input features for the RL and random forest models [43–45].

Author Contributions: Conceptualization, R.B., M.M. and O.K.; methodology, L.K.; software, D.L, M.M. and L.K.; validation, L.K. and. A.A.; formal analysis, O.K.; investigation, R.B.; resources, A.E., P.E. and P.T.; data curation, S.Z., D.L. and C.D.; writing—original draft preparation, R.B.; writing—review and editing, A.E., O.K. and P.T.; visualization, L.K. and R.B.; supervision, O.K. and C.H.; project administration, R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Approval was obtained for third party re-use of AmsterdamUMCdb data for research from its steering group, and the research was conducted according to the data use agreement. Such a study of deidentified data is not subject to the need for ethical review. The ethical approvals for the AmsterdamUMCdb have been previously described.

Data Availability Statement: Access to the dataset used in this manuscript may be requested from Amsterdam Medical Data Science (https:/amsterdammedicaldatascience.nl/ accessed on 1 January 2021).

Acknowledgments: The authors acknowledge the European Society of Intensive Care Medicine (ESICM) for support as part of the 2021 ESICM Datathon project.

Conflicts of Interest: The authors declare no conflict of interest regarding the contents of this submission.

References

1. Rudd, K.E.; Johnson, S.C.; Agesa, K.M.; Shackelford, K.A.; Tsoi, D.; Kievlan, D.R.; Colombara, D.V.; Ikuta, K.S.; Kissoon, N.; Finfer, S.; et al. Global, regional, and national sepsis incidence and mortality, 1990–2017: Analysis for the Global Burden of Disease Study. *Lancet* **2020**, *395*, 200–211. [CrossRef]
2. Singer, M.; Deutschman, C.S.; Seymour, C.W.; Shankar-Hari, M.; Annane, D.; Bauer, M.; Bellomo, R.; Bernard, G.R.; Chiche, J.-D.; Coopersmith, C.M.; et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* **2016**, *315*, 801–810. [CrossRef] [PubMed]
3. Seymour, C.W.; Kennedy, J.N.; Wang, S.; Chang, C.-C.H.; Elliott, C.; Xu, Z.; Berry, S.; Clermont, G.; Cooper, G.; Gomez, H.; et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA* **2019**, *321*, 2003–2017. [CrossRef]
4. Iwashyna, T.J.; Burke, J.F.; Sussman, J.B.; Prescott, H.C.; Hayward, R.A.; Angus, D.C. Implications of Heterogeneity of Treatment Effect for Reporting and Analysis of Randomized Trials in Critical Care. *Am. J. Respir. Crit. Care Med.* **2015**, *192*, 1045–1051. [CrossRef]
5. Evans, L.; Rhodes, A.; Alhazzani, W.; Antonelli, M.; Coopersmith, C.M.; French, C.; Machado, F.R.; Mcintyre, L.; Ostermann, M.; Prescott, H.C.; et al. Surviving sepsis campaign: International guidelines for management of sepsis and septic shock 2021. *Intensive Care Med.* **2021**, *47*, 1181–1247. [CrossRef]
6. Cook, C.; Smith, C. Sepsis and cortisone. *Nature* **1952**, *170*, 980. [CrossRef]
7. Annane, D.; Pastores, S.M.; Arlt, W.; Balk, R.A.; Beishuizen, A.; Briegel, J.; Carcillo, J.; Christ-Crain, M.; Cooper, M.S.; Marik, P.E.; et al. Critical illness-related corticosteroid insufficiency (CIRCI): A narrative review from a Multispecialty Task Force of the Society of Critical Care Medicine (SCCM) and the European Society of Intensive Care Medicine (ESICM). *Intensiv. Care Med.* **2017**, *43*, 1781–1792. [CrossRef] [PubMed]
8. Annane, D.; Bellissant, E.; Bollaert, P.E.; Briegel, J.; Confalonieri, M.; De Gaudio, R.; Keh, D.; Kupfer, Y.; Oppert, M.; Meduri, G.U. Corticosteroids in the treatment of severe sepsis and septic shock in adults: A systematic review. *JAMA* **2009**, *301*, 2362–2375. [CrossRef]
9. Annane, D.; Bellissant, E.; Bollaert, P.E.; Briegel, J.; Keh, D.; Kupfer, Y. Corticosteroids for treating sepsis. *Cochrane Database Syst. Rev.* **2015**, *12*, CD002243. [CrossRef]
10. Rygård, S.L.; Butler, E.; Granholm, A.; Møller, M.H.; Cohen, J.; Finfer, S.; Perner, A.; Myburgh, J.; Venkatesh, B.; Delaney, A. Low-dose corticosteroids for adult patients with septic shock: A systematic review with meta-analysis and trial sequential analysis. *Intensiv. Care Med.* **2018**, *44*, 1003–1016. [CrossRef] [PubMed]
11. Pirracchio, R.; Hubbard, A.; Sprung, C.L.; Chevret, S.; Annane, D.; for the Rapid Recognition of Corticosteroid Resistant or Sensitive Sepsis (RECORDS) Collaborators. Assessment of Machine Learning to Estimate the Individual Treatment Effect of Corticosteroids in Septic Shock. *JAMA Netw. Open* **2020**, *3*, e2029050. [CrossRef]
12. Antcliffe, D.B.; Burnham, K.L.; Al-Beidh, F.; Santhakumaran, S.; Brett, S.J.; Hinds, C.J.; Ashby, D.; Knight, J.C.; Gordon, A.C. Transcriptomic Signatures in Sepsis and a Differential Response to Steroids. From the VANISH Randomized Trial. *Am. J. Respir. Crit. Care Med.* **2019**, *199*, 980–986. [CrossRef] [PubMed]
13. Doya, K. Reinforcement learning: Computational theory and biological mechanisms. *HFSP J.* **2007**, *1*, 30–40. [CrossRef]
14. 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.* **2018**, *24*, 1716–1720. [CrossRef]
15. Liu, S.; See, K.C.; Ngiam, K.Y.; Celi, L.A.; Sun, X.; Feng, M. Reinforcement Learning for Clinical Decision Support in Critical Care: Comprehensive Review. *J. Med. Internet Res.* **2020**, *22*, e18477. [CrossRef] [PubMed]
16. Liu, S.; Ngiam, K.Y.; Feng, M. Deep Reinforcement Learning for Clinical Decision Support: A Brief Survey. *arXiv* **2019**, arXiv:1907.09475.
17. Thoral, P.J.; Peppink, J.M.; Driessen, R.H.; Sijbrands, E.J.; Kompanje, E.J.; Kaplan, L.; Bailey, H.; Kesecioglu, J.; Cecconi, M.; Churpek, M.; et al. Sharing ICU Patient Data Responsibly Under the Society of Critical Care Medicine/European Society of Intensive Care Medicine Joint Data Science Collaboration: The Amsterdam University Medical Centers Database (AmsterdamUMCdb) Example. *Crit. Care Med.* **2021**, *49*, e563–e577. [CrossRef] [PubMed]

18. 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* **2019**, *23*, 374. [CrossRef]
19. Thoral, P.J.; Driessen, R.H.; Peppink, J.M. AmsterdamUMCdb Github Repository. 2020. Available online: https://github.com/AmsterdamUMC/AmsterdamUMCdb (accessed on 15 September 2021).
20. Shin, J.; Badgwell, T.A.; Liu, K.-H.; Lee, J.H. Reinforcement Learning—Overview of recent progress and implications for process control. *Comput. Chem. Eng.* **2019**, *127*, 282–294. [CrossRef]
21. Li, L.; Komorowski, M.; Faisal, A.A. The Actor Search Tree Critic (ASTC) for Off-Policy POMDP Learning in Medical Decision Making. *arXiv* **2018**, arXiv:1805.11548.
22. Haarnoja, T.; Zhou, A.; Hartikainen, K.; Tucker, G.; Ha, S.; Tan, J.; Kumar, V.; Zhu, H.; Gupta, A.; Abbeel, P.; et al. Soft Actor-Critic Algorithms and Applications. *arXiv* **2018**, arXiv:181205905.
23. Liu, D.; Ahmet, A.; Ward, L.; Krishnamoorthy, P.; Mandelcorn, E.D.; Leigh, R.; Brown, J.P.; Cohen, A.; Kim, H. A practical guide to the monitoring and management of the complications of systemic corticosteroid therapy. *Allergy Asthma Clin. Immunol.* **2013**, *9*, 30. [CrossRef] [PubMed]
24. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: Large-scale machine learning on heterogeneous systems. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2015.
25. Thomas, P.; Theocharous, G.; Ghavamzadeh, M. High-Confidence Off-Policy Evaluation. In Proceedings of the AAAI Conference on Artificial Intelligence, Hollywood, FL, USA, 18–20 May 2015. [CrossRef]
26. Thomas, P.; Theocharous, G.; Ghavamzadeh, M. High Confidence Policy Improvement. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; Proceedings of Machine Learning Research, PMLR. Francis, B., David, B., Eds.; MLResearch Press: San Francisco, CA, USA, 2015; Volume 37, pp. 2380–2388.
27. Montavon, G.; Binder, A.; Lapuschkin, S.; Samek, W.; Müller, K.-R. Layer-Wise Relevance Propagation: An Overview. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2019; pp. 193–209. [CrossRef]
28. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
29. Sligl, W.I.; Milner, J.D.A.; Sundar, S.; Mphatswe, W.; Majumdar, S.R. Safety and Efficacy of Corticosteroids for the Treatment of Septic Shock: A Systematic Review and Meta-Analysis. *Clin. Infect. Dis.* **2009**, *49*, 93–101. [CrossRef] [PubMed]
30. Vincent, J.-L. Steroids in sepsis: Another swing of the pendulum in our clinical trials. *Crit. Care* **2008**, *12*, 141. [CrossRef]
31. Sprung, C.L.; Annane, D.; Keh, D.; Moreno, R.; Singer, M.; Freivogel, K.; Weiss, Y.G.; Benbenishty, J.; Kalenka, A.; Forst, H.; et al. Hydrocortisone Therapy for Patients with Septic Shock. *New Engl. J. Med.* **2008**, *358*, 111–124. [CrossRef] [PubMed]
32. Venkatesh, B.; Finfer, S.; Cohen, J.; Rajbhandari, D.; Arabi, Y.; Bellomo, R.; Billot, L.; Correa, M.; Glass, P.; Harward, M.; et al. Adjunctive Glucocorticoid Therapy in Patients with Septic Shock. *New Engl. J. Med.* **2018**, *378*, 797–808. [CrossRef] [PubMed]
33. Shah, A.D.; MacCallum, N.S.; Harris, S.; Brealey, D.A.; Palmer, E.; Hetherington, J.; Shi, S.; Perez-Suarez, D.; Ercole, A.; Watkinson, P.J.; et al. Descriptors of Sepsis Using the Sepsis-3 Criteria: A Cohort Study in Critical Care Units Within the U.K. National Institute for Health Research Critical Care Health Informatics Collaborative. *Crit. Care Med.* **2021**, *49*, 1883. [CrossRef]
34. Rhee, C.; Murphy, M.V.; Li, L.; Platt, R.; Klompas, M.; for the Centers for Disease Control and Prevention Epicenters Program. Comparison of Trends in Sepsis Incidence and Coding Using Administrative Claims Versus Objective Clinical Data. *Clin. Infect. Dis.* **2015**, *60*, 88–95. [CrossRef]
35. Valik, J.K.; Ward, L.; Tanushi, H.; Müllersdorf, K.; Ternhag, A.; Aufwerber, E.; Färnert, A.; Johansson, A.F.; Mogensen, M.L.; Pickering, B.; et al. Validation of automated sepsis surveillance based on the Sepsis-3 clinical criteria against physician record review in a general hospital population: Observational study using electronic health records data. *BMJ Qual. Saf.* **2020**, *29*, 735–745. [CrossRef]
36. Wang, F.; Kaushal, R.; Khullar, D. Should Health Care Demand Interpretable Artificial Intelligence or Accept "Black Box" Medicine? *Ann. Intern. Med.* **2020**, *172*, 59–60. [CrossRef] [PubMed]
37. Brouwer, M.C.; McIntyre, P.; Prasad, K.; van de Beek, D. Corticosteroids for acute bacterial meningitis. *Cochrane Database Syst. Rev.* **2015**, *2015*, CD004405. [CrossRef] [PubMed]
38. Cunha, B.A.; Lortholary, O.; Cunha, C.B. Fever of unknown origin: A clinical approach. *Am. J. Med.* **2015**, *128*, 1138.e1–1138.e15. [CrossRef] [PubMed]
39. Teblick, A.; Peeters, B.; Langouche, L.; Van den Berghe, G. Adrenal function and dysfunction in critically ill patients. *Nat. Rev. Endocrinol.* **2019**, *15*, 417–427. [CrossRef]
40. Walker, B.R.; Yau, J.L.; Brett, L.P.; Seckl, J.R.; Monder, C.; Williams, B.C.; Edwards, C.R. 11 beta-hydroxysteroid dehydrogenase in vascular smooth muscle and heart: Implications for cardiovascular responses to glucocorticoids. *Endocrinology* **1991**, *129*, 3305–3312. [CrossRef]
41. Shoenfeld, Y.; Gurewich, Y.; Gallant, L.A.; Pinkhas, J. Prednisone-induced leukocytosis. *Am. J. Med.* **1981**, *71*, 773–778. [CrossRef]
42. Van de Sande, D.; van Genderen, M.E.; Huiskens, J.; Gommers, D.; van Bommel, J. Moving from bytes to bedside: A systematic review on the use of artificial intelligence in the intensive care unit. *Intensive Care Med.* **2021**, *47*, 750–760. [CrossRef]
43. Richard, S. Sutton and Andrew G. Barto, Reinforcement Learning: An Introduction. *IEEE Trans. Neural Netw* **2015**, *9*, 1054.

44. Brockman, G.; Cheung, V.; Pettersson, L.; Schneider, J.; Schulman, J.; Tang, J.; Zaremba, W. Openai gym. *arXiv* **2016**, arXiv:1606.01540.
45. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In Proceedings of the International Conference on Machine Learning, PMLR 2022, Virtual Event, 7–8 April 2022. [CrossRef]