

Reinforcement learning to help intensivists optimize mechanical ventilation settings (EZ-Vent): Derivation and validation using large databases

Siqi Liu

National University of Singapore

Zhuoyang Xu

Ping An Healthcare Technology

Zhuo Liu

Ping An Healthcare Technology

Xingzhi Sun

Ping An Healthcare Technology

Guotong Xie

Ping An Healthcare Technology

Mengling Feng (✉ ephfm@nus.edu.sg)

National University of Singapore

Kay Choong See

National University Hospital

Research Article

Keywords: mechanical ventilation, reinforcement learning, artificial intelligence

Posted Date: October 13th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2146974/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Mechanical ventilation is the cornerstone of critical care medicine. However, choosing the optimal ventilator strategy for a patient remains imprecise. Existing guidelines provide one-size-fits-all recommendations, but do not personalize treatments for different intensive care unit (ICU) patients. In this study, we aimed to design and evaluate an artificial intelligence (AI) solution that could tailor an optimal ventilator strategy for each critically ill patient who requires mechanical ventilation.

Methods

We proposed a reinforcement learning-based AI solution using observational data from multiple ICUs in the US. The primary outcome was hospital mortality. Secondary outcomes were the proportion of optimal oxygen saturation and the proportion of optimal mean arterial blood pressure. We trained our AI agent to learn each patients' treatment trajectory and thus to recommend low/medium/high levels of three ventilator settings, namely the positive end-expiratory pressure, fraction of inspired oxygen and ideal body weight-adjusted tidal volume. Off-policy evaluation metrics were applied to evaluate the AI policy.

Results

We studied 5105 and 21595 patients' ICU stays from the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC-IV) and eICU Collaborative Research (eICU) databases respectively. Observed hospital mortality rates were 18.2% (eICU) and 31.1% (MIMIC-IV). For the learnt AI policy, we estimated the hospital mortality rate (eICU $14.7 \pm 0.7\%$; MIMIC-IV $29.1 \pm 0.9\%$), proportion of optimal oxygen saturation (eICU $57.8 \pm 1.0\%$; MIMIC-IV $49.0 \pm 1.0\%$), and proportion of optimal mean arterial blood pressure (eICU $34.7 \pm 1.0\%$; MIMIC-IV $41.2 \pm 1.0\%$). Based on multiple quantitative and qualitative evaluation metrics, our proposed AI solution has potential to outperform observed clinical practice.

Conclusions

Our proposed approach has potential to be applied as a clinical decision support tool that helps intensivists make better treatment decisions and to improve the survival and prognosis of critically ill patients who require invasive respiratory support.

Background

Mechanical ventilation is the cornerstone of critical care medicine, and one of the most common interventions for patients admitted to intensive care units (ICUs). Studies showed that approximately one third of ICU patients require mechanical ventilation in the US⁽¹⁾. In the recent years, due to both the

COVID-19 pandemic and aging populations in many developed countries, the use of mechanical ventilation in ICU has been constantly increasing.

Despite decades of research, choosing the optimal ventilator strategy for a patient remains imprecise. Appropriate ventilator settings are important but complicated by significant inter-patient variability. Current clinical guidelines provide one-size-fits-all recommendations, but do not personalize treatment for different ICU patients. In particular, existing clinical guidelines do not address personalized optimal settings for mechanical ventilation, including positive end-expiratory pressure (PEEP) level, fraction of inspired oxygen (FiO_2) and ideal body weight-adjusted tidal volume(2). With the understanding that mechanical ventilation itself can cause and potentiate lung injury, it is important to choose appropriate ventilatory strategies to mitigate ventilator-induced lung injury (VILI)(3). Nonetheless, even guideline recommendations may not be adhered to, since a wide discrepancy in practice exists and evidence-based interventions are underused for the task(4).

In this study, we propose an artificial intelligence (AI) solution based on reinforcement learning (RL) to address the existing guideline-related gaps. We name the AI solution “EZ-Vent”. The framework of the proposed solution is shown in Fig. 1. We first collected mechanical ventilated patients’ health data and intensivists’ treatment actions from two large electronic health records (EHR) datasets in the US. We then trained a type of RL-based model, named the Batch Constrained deep Q-learning (BCQ), to learn from physicians’ treatment actions, and develop an optimal strategy for setting mechanical ventilation. Note that our AI solution learns from the fixed data sets without interaction with real patients. This type of learning is commonly known as batch learning in RL. Many widely used RL algorithms have failed in the batch setting, as the models learnt often suffer from overestimation and poor performance when encounter data that are not contained in the provided batch. Different from such traditional RL algorithms, the BCQ algorithm applies constraints to ensure the learnt policy is not too distant from physicians’ policy in such scenarios. As a result, we chose to implement BCQ in our solution due to its capability in developing safe policy from observational data. Understand that safe policy learning is important in healthcare applications, the proposed AI solution aims to be used as a clinical decision support (CDS) tool to help intensivists make better decisions for critically ill patients who require mechanical ventilation.

Our proposed AI solution recommends the optimal ventilator settings for levels of PEEP, FiO_2 and ideal body weight-adjusted tidal volume by taking into account the individual patients’ conditions including their demographic features, physiological status, and multiple comorbidities. Compared to the existing guidelines, the proposed solution can adjust treatment recommendations based on changes in a patient’s condition. In addition, we designed a set of flags to capture the sudden changes in patients’ health and used the flag timings to cut patients’ trajectories into small time-varying intervals, because such timings were likely the decision points for physicians to make necessary interventions in practice. If the model could be implemented at bedside in real time, it can assist intensivists to make better decisions. Therefore, our proposed AI solution has the potential to improve outcomes for patients who require

mechanical ventilation in the ICUs. Considering that there are millions of ventilated patients per year, improved ventilator settings could save thousands of lives.

Methods

Study population and datasets

The observational data for mechanically ventilated patients were extracted from two large EHR databases in the US: the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC-IV) database and the eICU Collaborative Research (eICU) database. We included patients who were 16 years old and above, and whose ventilation duration was at least 24 hours. Only the first ICU admission for each patient was considered, and we studied the first 48 hours' of ventilation data. Our exclusion criteria were as follows: 1) patients with missing outcomes; 2) age of 89 years old and above; 3) incomplete demographics; 4) missing ventilator recordings of PEEP, FiO_2 and tidal volume for the entire ventilation duration; 5) patients who had prolonged ventilation for more than 2 weeks. After those exclusions, 5105 patients in MIMIC-IV and 21595 patients in the eICU dataset remained. The flowchart for cohort selection is shown in Fig. 2. We split the eICU dataset into 70% for training and 30% for validation. The full MIMIC-IV dataset was held out as the testing set.

Outcome variables

The primary outcome was hospital mortality. Secondary outcomes were the proportion of optimal oxygen saturation (SpO_2) and the proportion of optimal mean arterial blood pressure (MBP). The optimal ranges of SpO_2 and MBP were defined as follows: $94\% < SpO_2 < 98\%$ (5); $70mmHg \leq MBP \leq 80mmHg$ (6).

Reinforcement learning: a primer

RL is a goal-oriented AI method where a computer agent, acting like a decision maker, analyzes available data within its defined environment, derives a policy for taking actions, and optimizes long-term rewards. The agent is the computational model we want to develop. In general, an agent gets evaluative feedback (reward) about the performance of its action at each consecutive time step, allowing it to improve the performance of subsequent actions by trial-and-error. Mathematically, this sequential decision-making process is called the Markov Decision Process (MDP). An MDP is defined by a few major components:

- **State:** a state ($s \in \text{statespace}$) represents the environment. In this study, it represents the health status of a patient at each timestamp. We constructed the patient's state by using 40 relevant physiological features containing demographics, laboratory values and vital signs (see Table e1 for the full list).
- **Action:** an action ($a \in \text{actionspace}$) is the treatment option that the agent takes at each time step, which influences the next state. In our study, the action space was constructed as 18 possible

discrete actions from combinations of low/medium/high levels of the three ventilator settings: PEEP, FiO_2 and ideal body weight-adjusted tidal volume (Figure e1).

- Reward: the reward function is the observed feedback given a state-action (s, a) pair. The reward of our model reflected the objective of an RL agent, which was to improve survival, and to achieve oxygen saturation and mean arterial pressure within their respective optimal ranges. Hospital mortality was used as the terminal reward, whereas SpO_2 and MBP were applied as intermittent rewards.

The solution of the MDP is an optimized set of rules, i.e., the RL policy. The ability of RL to learn complex sequential decisions makes it suitable for critical care applications, and we hope to use its capability of learning to provide individualized treatment policies that could improve survival of mechanically ventilated patients.

Reinforcement learning: time-varying interval

We applied time-varying intervals to discretized state and action pairs. For each patient's treatment trajectory, we analyzed the data from the 1st hour of using mechanical ventilation till the 48th hour or till ventilator weaning, whichever came earlier. Then we cut the trajectories into 4-hourly timesteps, except for cases when flags were raised to further cut the trajectories. We designed the set of flags as follows: 1. the SpO_2 dropped under 90%; 2. the partial pressure of oxygen (PaO_2) dropped under 60 mmHg; 3. $pH < 7.25$ or $pH > 7.45$. When any flag was raised, the trajectories would be further cut into shorter timesteps. We designed these set of flags to reflect real-world conditions, where such flags would prompt changes in ventilator settings. Next, we selectively merged timesteps if they were too short, with the minimum time interval no less than one hour. For multiple values in one timestep, we computed a time-weighted average value. Missing values were imputed with the nearest value before the timestep, and if this was not available, missing values were imputed with the value from the next timestep. Binary state variables were represented using 0 or 1. Continuous state variables were normalized or log-normalized to (0, 1) as appropriate.

Reinforcement learning: model development

RL as a CDS tool has gained increasing popularity for applications in the ICU setting due to its ability to solve complex sequential decision-making problem using offline patients' data from EHR databases. However, one major drawback of using historical data to train a RL model is extrapolation error. The term extrapolation is a statistical technique for estimating values that extend beyond a particular collection of data or observations. Extrapolation error is caused by the mismatch between the data distribution in the offline dataset and future observations. To mitigate this problem, we applied the BCQ model in this study, which is an offline RL model that has the advantage over other RL algorithms in the batch setting by addressing the extrapolation error(7).

The BCQ model consists of two modules, a supervisor network, and a policy network. The supervisor network is used to mimic the physicians' policy from the observational data. The policy network fits the

optimal policy under the constraint of the supervisor network, where only actions likely to be taken in the observational data are considered and evaluated. The final optimized policy is then expected to lead to good future outcome as well as to select a safe action.

In the training stage, when BCQ receives the training sample, the supervisor network will first learn the mapping from state to action, that is, which action would be taken based on historical data. Then, the policy network will optimize its policy with the reward information and the output of the supervisor network. This training process is iterated several times until we derive the final AI policy.

We designed a clinically guided reward function which outputs a reward (penalty) when the patients' state improves (deteriorates) based on short-term and long-term health outcomes. The short-term health indicators are patients' MBP as well as SpO_2 . At each intermittent (i.e. non-terminal) timestep of a patient's trajectory, he/she would receive a positive short-term reward if his MBP fell within the range of [70mmHg, 80mmHg] or his SpO_2 fell within (94%, 98%) based on the literature of maintaining optimal level of vital signs and blood gas(6), and the patient would receive a penalty (negative reward) when MBP and SpO_2 fell out of the range. At terminal timesteps, each patient would receive a final reward (or penalty) if a patient survived (or deceased) at discharge. The reward information at each timestep would help BCQ to learn those action patterns from physicians that lead to good short-term and long-term outcomes.

Reinforcement learning: benchmark policies

We evaluated our RL based policy by comparison with three benchmark policies:

- Random policy: All the 18 discrete actions have equal probabilities to be chosen.
- One-size-fits-all policy: The action with the highest probability in the cohort is always chosen.
- Physicians' policy: The actual observed policy in the validation and testing sets.

Reinforcement learning: evaluation metrics

We used extensive quantitative and qualitative analyses to evaluate the performance of the learnt AI policy and benchmarks. Firstly, to understand the relationship between the expected return of the learnt policies and the clinical outcomes, we mapped the expected return to the estimated outcome occurrence. We sorted the expected returns of the physicians' policy into discrete bins and obtained the average empirical mortality rate from the patients in each bin. The empirical mortality estimate was used to derive a relationship between the range of computed returns of the AI policy against the observed mortality. This estimation process was performed for secondary outcomes too.

Treatment recommendation is an off-policy learning problem, which aims to learn an optimal policy using trajectories from an observed behavior policy (physicians' policy). Evaluating the learnt policy with off-policy estimation (OPE) methods is crucial for healthcare applications to avoid high risk of failure of negative impact. OPE methods use examples from the behavior policy to evaluate the performance of the

learnt policy. We used a commonly applied advanced off-policy estimator, Consistent Weighted Per-Decision Important Sampling (CWPDIS)(8) for the learnt AI policy as well as for the benchmark policies. In addition, we used a random forest classification model to rank the importance of various predictors for the actions under the physicians' policy (See Figure e3 to Figure e5). This allowed us to understand physicians' behavior with regard to choosing ventilator settings.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had access to all the data in the study and was responsible for the decision to submit for publication.

Results

Patient characteristics of the selected cohorts from the MIMIC-IV and eICU datasets are provided in Table 1. There were no statistically significant differences in age, gender or body weight. However, we observed that patients in the MIMIC-IV dataset had greater illness severity compared with those in the eICU dataset. Patients in MIMIC-IV dataset had higher Elixhauser score (5.0 [0.0, 12.0] vs. 3.0 [0.0, 7.0]), higher re-intubation rate (30.7% vs. 16.7%), longer hospital stay (291.0 hours [171.0, 477.0] vs. 191.4 hours [120.1, 307.5]) and higher hospital mortality rate (31.1% vs. 18.2%) compared to patients in the eICU dataset.

We plotted the action frequency distributions of the physicians' policy and the learnt AI policy. We compared the learnt policy against the physicians' policy for low, medium, and high SOFA levels for patients in the MIMIC-IV (Fig. 3) and the eICU databases (Figure e2). For each SOFA group, we counted the number of actions for the three action categories: PEEP, FiO_2 and ideal body weight-adjusted tidal volume. The blue bars represent the actions under the physicians' policy, and the red bars represent actions under the learnt AI policy.

We derived relationships between the range of computed returns of the learnt policy against various outcomes for both the eICU validation set and the MIMIC-IV test set (Fig. 4). The proportion of mortality (red) against the returns showed inverse relationships, whereas proportion of optimal SpO_2 (green) and the proportion of optimal MBP (blue) against returns showed overall positive relationships in both datasets.

The OPE performance of the learnt policy is shown in Table 2. We estimated the hospital mortality rate (eICU $14.7 \pm 0.7\%$; MIMIC-IV $29.1 \pm 0.9\%$), proportion of optimal SpO_2 (eICU $57.8 \pm 1.0\%$; MIMIC-IV $49.0 \pm 1.0\%$), and proportion of optimal MBP (eICU $34.7 \pm 1.0\%$; MIMIC-IV $41.2 \pm 1.0\%$) for the learnt policy. We also report outcomes for the physicians' policy, including the observed mortality rate (eICU 18.2%; MIMIC-IV 31.1%), proportion of optimal SpO_2 (eICU 45.1%; MIMIC-IV 40.5%) and proportion of optimal MBP (eICU 33.4%; MIMIC-IV 37.1%) in the two datasets respectively.

To examine the effectiveness of using time-varying intervals in the action setting, we visualize a representative patient case in Fig. 6. The action plots for PEEP/ FiO_2 /ideal body weight-adjusted tidal volume for the patient is shown on left/middle/right of the figure respectively. The blue lines represent the raw data; green lines represent fixed 4-hourly time interval; red lines show the time-varying interval of the three actions. Flags to cut intervals in time-varying setting are showed as vertical dotted lines with yellow shadows.

We report the feature importance with regard choosing ventilator settings under the physicians' policy in Figures e3-e5. The top 10 important features included: PaO_2/FiO_2 ratio, $PaCO_2$, PaO_2 , creatinine, lactate, prothrombin time (PTT), base excess, age, admission weight and Richmond Agitation Sedation Scale (RASS) score.

Discussion

In this study, we used a reinforcement learning based AI model named BCQ, to learn the optimal ventilation policy for critically ill patients who require mechanical ventilation. We validated the policy using two large public datasets from the US: the eICU and MIMIC-IV datasets. In both datasets, the learnt policy had superior performance compared to observed physician policy, based on several quantitative and qualitative evaluation metrics.

We first formulated the clinical problem of choosing optimal ventilator settings in the ICU as a reinforcement learning problem. We then used relevant physiological variables to represent patients' health status as states and cut the ventilator treatment trajectories into time-varying steps to reflect the changes in patients' condition. We designed a set of flags to capture the sudden changes in patients' health and used the flag timings to further cut the trajectory, because such timings were the likely decision points for physicians to make necessary interventions. From the visualization of time-varying intervals in Fig. 6, we observed that when the flags were raised (vertical dotted line), time-varying interval setting (red lines) can better reflect the changes in raw data (blue lines) of ventilator settings in a timely manner compared to fixed 4-hourly time intervals (blue lines).

The action space was designed as 18 discrete actions comprising low/medium/high combinations of three ventilator settings: PEEP, FiO_2 and ideal body weight-adjusted tidal volume. At each timestep, the AI model took input from the patients' status and physicians' actions, received the evaluative feedback (reward) on those action, and adjusted itself to maximize survival and keep SpO_2 and MBP within their optimal ranges. Hospital mortality was used as the terminal reward, whereas SpO_2 and MBP were applied as intermittent rewards.

Although patients from the MIMIC-IV dataset were generally sicker compared to those in the eICU dataset, this provided an opportunity of us to test the extrapolation ability of the BCQ model. Given the consistently superior performance of the BCQ model-derived RL policy compared to physicians' policy in both datasets, we felt that the extrapolation ability of the BCQ model was satisfactory.

From the action frequency distribution plot (Fig. 3) for patients in MIMIC-IV, we found that the actions from physicians (red) and the actions recommended by AI policy (blue) have some discrepancies in all ventilator settings. This result is desirable, because the supervisor network in the BCQ model does not aim to copy from physicians' choices completely. On the contrary, the supervisor network was used to learn good action patterns from physicians and limit the choice of actions with constraints. In addition, we found the learnt policy recommended low-level PEEP and high-level ideal body weight-adjusted tidal volume more frequently compared to physicians' current practices for all the SOFA level groups. This finding suggests that the high PEEP-low tidal volume strategy for acute respiratory distress syndrome (ARDS) (2, 9) may not be optimal for all mechanically ventilated patients and should not be applied as a one-size-fits-all approach. For the management of FiO_2 , the learnt policy suggested more frequent use of low and medium levels and avoided high levels FiO_2 for all SOFA groups. This policy suggestion is in line with the known harm from excessive oxygenation, which has been found across different types of critical illness(5, 10, 11).

We computed the learnt policy's expected return, and we plotted it against mortality risk in Fig. 4. We observed inverse relationships between expected return and mortality (red) in both validation and testing datasets. This indicates that the optimal policy (high return) results in lower mortality for patients. For the secondary outcomes related to maintaining SpO_2 and MBP within their respective optimal ranges, the expected return showed positive relationships (green for SpO_2 , blue for MBP). This indicates that the optimal policy (high return) leads to higher proportions of SpO_2 and MBP within their respective optimal ranges.

We also illustrated the observed mortality in terms of the differences between AI and physicians' ventilator setting (Fig. 5). An effective policy has the lowest mortality when the recommended and administered ventilator settings coincide (x-axis value is zero), indicating when the practice strictly following the AI policy, it has the lowest mortality. At the same time, for an effective policy, the observed mortality should increase as the administered ventilator settings deviate from recommended settings from the AI policy. Accordingly, an effective policy should have a V-shaped curve with the minimum at 0, which we observed for the AI policy under all the three action groups (PEEP, FiO_2 and tidal volume).

From the quantitative evaluation using CWPDIS, we found the learnt policy had the lowest estimated mortality compared with all three benchmark policies. At the same time, the learnt policy achieved the highest proportion of optimal SpO_2 and MBP in both datasets. Intuitively and as expected, the random policy had the worst outcome among all the policies.

Although our study harnessed two large databases for derivation and external validation of an RL model, several limitations remain. Firstly, our study is retrospective, and the results require prospective validation. Secondly, patients were treated in the US, which is a high-income country with advanced medical care. Whether the RL model would perform similarly in a lower resourced country is unknown and requires further study. Thirdly, our model is a standalone AI model, which may be made more effective by combining AI with human input (i.e. collaborative intelligence(12)).

Despite the above limitations, our study highlights the potential of AI and RL to personalize medical care by accounting for the myriad variations in patients' clinical features and tailoring treatment recommendations according to those variations. Our method may also be applied to complex clinical decision making beyond mechanical ventilation, such as sepsis management(13) and drug dosing(14).

Conclusions

In conclusion, management of the critically ill patient is particularly challenging due to the complexity and dynamic changes in patients' conditions. To assist physicians, we introduced an RL-based AI solution (EZ-Vent) that aimed to learn an optimal ventilator strategy for mechanical ventilated patients. We applied extensive quantitative and qualitative evaluations and showed that our solution outperformed current clinical practice in two large observational datasets. Prospective validation of RL as a clinical decision support tool will be performed as future work.

Abbreviations

AI: artificial intelligence; ARDS : acute respiratory distress syndrome; BCQ: Batch Constrained deep Q-learning; CDS: clinical decision support; CWPDIS: Consistent Weighted Per-Decision Important Sampling; EHR: electronic health records; eICU: eICU Collaborative Research Database; FiO2: fraction of inspired oxygen; ICU: intensive care unit; MBP: mean arterial blood pressure; MDP: Markov Decision Process; MIMIC-IV: Multiparameter Intelligent Monitoring in Intensive Care Database; OPE: off-policy estimation; PaO2: partial pressure of oxygen; PEEP: positive end-expiratory pressure; PTT: prothrombin time; RASS: Richmond Agitation Sedation Scale; RL: reinforcement learning; SpO2: optimal oxygen saturation; VILI: ventilator-induced lung injury.

Declarations

Author contributions: All authors contributed to the study design. KS contributed to literature search. SL collected the data. SL, ZX and ZL contributed to data analysis, figures and tables. All authors contributed to the writing and approval of the manuscript.

Acknowledgement: None

Funding: This research is supported by the National Research Foundation Singapore under its AI Singapore Programme (Award Number: AISG-100E-2020-055 and AISG-GC-2019-001-2A).

Availability of data and materials: The data supporting the findings of the current study are available from the MIMIC-IV and eICU databases. However, these data were used under license, and restrictions apply to the availability of the data. Data are thus not publicly available but available from the corresponding author (Dr. Mengling Feng) upon reasonable request and with permission of the holder of the database.

Consent for publication: All authors have reviewed and approved the manuscript for publication.

Ethics approval and consent to participate: Consent was obtained when the databases were established.

Competing interests: See KC declares receipt of honoraria from GE Healthcare and Medtronic.

References

1. Wunsch H, Wagner J, Herlim M, Chong D, Kramer A, Halpern SD. ICU occupancy and mechanical ventilator use in the United States. *Critical care medicine*. 2013;41(12).
2. Fan E, Del Sorbo L, Goligher EC, Hodgson CL, Munshi L, Walkey AJ, et al. An official American Thoracic Society/European Society of Intensive Care Medicine/Society of Critical Care Medicine clinical practice guideline: mechanical ventilation in adult patients with acute respiratory distress syndrome. *American journal of respiratory and critical care medicine*. 2017;195(9):1253-63.
3. Slutsky AS, Ranieri VM. Ventilator-induced lung injury. *New England Journal of Medicine*. 2013;369(22):2126-36.
4. Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, et al. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *Jama*. 2016;315(8):788-800.
5. van den Boom W, Hoy M, Sankaran J, Liu M, Chahed H, Feng M, et al. The search for optimal oxygen saturation targets in critically ill patients: observational data from large ICU databases. *Chest*. 2020;157(3):566-73.
6. Vincent J-L, Nielsen ND, Shapiro NI, Gerbasi ME, Grossman A, Doroff R, et al. Mean arterial pressure and mortality in patients with distributive shock: a retrospective analysis of the MIMIC-III database. *Annals of intensive care*. 2018;8(1):1-10.
7. Fujimoto S, Meger D, Precup D, editors. Off-policy deep reinforcement learning without exploration. *International Conference on Machine Learning*; 2019: PMLR.
8. Thomas PS. Safe reinforcement learning. 2015.
9. Briel M, Meade M, Mercat A, Brower RG, Talmor D, Walter SD, et al. Higher vs lower positive end-expiratory pressure in patients with acute lung injury and acute respiratory distress syndrome: systematic review and meta-analysis. *Jama*. 2010;303(9):865-73.
10. Asfar P, Schortgen F, Boisramé-Helms J, Charpentier J, Guérot E, Megarbane B, et al. Hyperoxia and hypertonic saline in patients with septic shock (HYPER2S): a two-by-two factorial, multicentre, randomised, clinical trial. *The Lancet Respiratory Medicine*. 2017;5(3):180-90.
11. Girardis M, Busani S, Damiani E, Donati A, Rinaldi L, Marudi A, et al. Effect of conservative vs conventional oxygen therapy on mortality among patients in an intensive care unit: the oxygen-ICU randomized clinical trial. *Jama*. 2016;316(15):1583-9.
12. See KC. Collaborative intelligence for intensive care units. *Critical Care*. 2021;25(1):1-2.

13. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nature medicine*. 2018;24(11):1716-20.
14. Liu S, See KC, Ngiam KY, Celi LA, Sun X, Feng M. Reinforcement learning for clinical decision support in critical care: comprehensive review. *Journal of medical Internet research*. 2020;22(7):e18477.

Tables

Table 1: Patient characteristics

Variables	MIMIC-IV (n=5,105)	eICU (n= 21,595)
Female (%)	2,154 (42.2)	9,244 (42.8)
Age, years (median [IQR])	65.0 [53.0, 76.0]	64.0 [53.0, 74.0]
Body weight, kg (median [IQR])	80.9 [67.6, 97.2]	81.9 [68.0, 99.7]
Reintubation (%)	1,565 (30.7)	3,661 (16.7)
Elixhauser score (median [IQR])	5.0 [0.0, 12.0]	3.0 [0.0, 7.0]
First SOFA score (median [IQR])	2.0 [0.0, 4.0]	2.0 [0.0, 4.0]
Hospital LOS, hours (median [IQR])	291.0 [171.0, 477.0]	191.4 [120.1, 307.5]
Hospital mortality (%)	1,590 (31.1)	3,934 (18.2)
PEEP, cmH2O (median [IQR])	5.3 [5.0, 10.0]	5.0 [5.0, 5.0]
FIO2, % (median [IQR])	50.0 [40.0, 60.0]	49.2 [40.0, 61.1]
Tidal volume, ml/kg IBW (median [IQR])	7.2 [6.4, 8.3]	7.8 [6.9, 8.8]

IBW = Ideal body weight, Tidal volume = ideal weight-adjusted tidal volume, LOS = length of stay.

Table 2: Outcomes on test set of the datasets

Dataset	Policy	Proportion of time within SpO_2 target range (%)	Proportion of time within MBP target range (%)	Estimated mortality (%)	Observed mortality (%)
eICU	Physician	45.1	33.4	16.2 ± 0.8	18.2
	Random	35.2 ± 1.1	28.0 ± 1.1	17.6 ± 0.9	
	One-size-fit-all	39.7 ± 1.1	30.7 ± 1.1	17.4 ± 0.8	
	AI	57.8 ± 1.0	34.7 ± 1.0	14.7 ± 0.7	
MIMIC-IV	Physician	40.5	37.1	30.6 ± 0.9	31.1
	Random	34.6 ± 1.0	36.2 ± 1.0	32.3 ± 0.9	
	One-size-fit-all	40.9 ± 1.0	38.5 ± 1.0	32.0 ± 0.9	
	AI	49.0 ± 1.0	41.2 ± 1.0	29.1 ± 0.9	

Physician = The actual observed policy in the validation and testing set, **Random** = All the 18 discrete actions have the equal probabilities to be chosen, **One-size-fit-all** = The action with the highest probability in the cohort is always chosen. AI = AI policy from BCQ model, SpO_2 target range = $94\% < SpO_2 < 98\%$, MBP target range = $70\text{mmHg} \leq MBP \leq 80\text{mmHg}$.

Figures

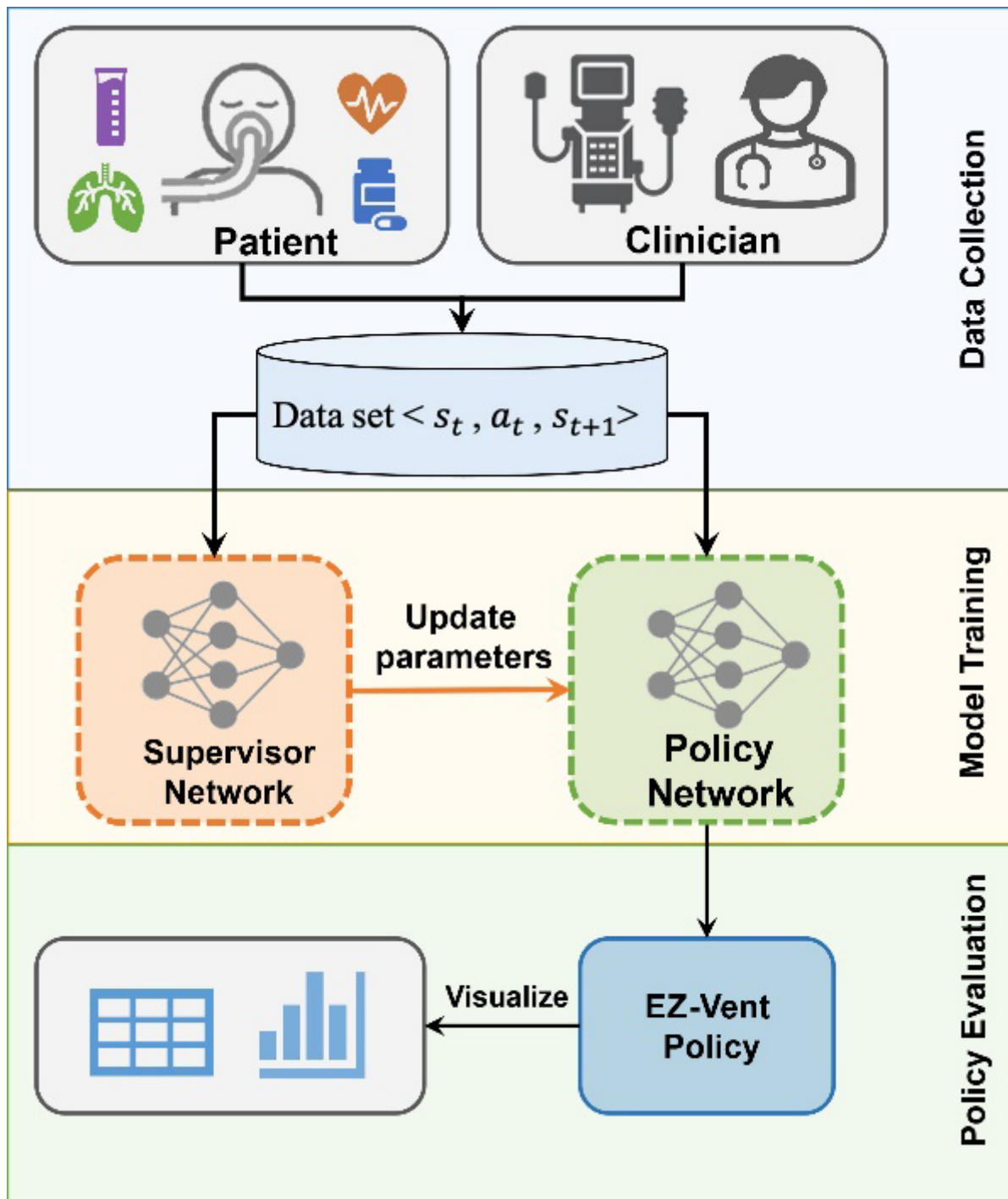


Figure 1

EZ-Vent framework

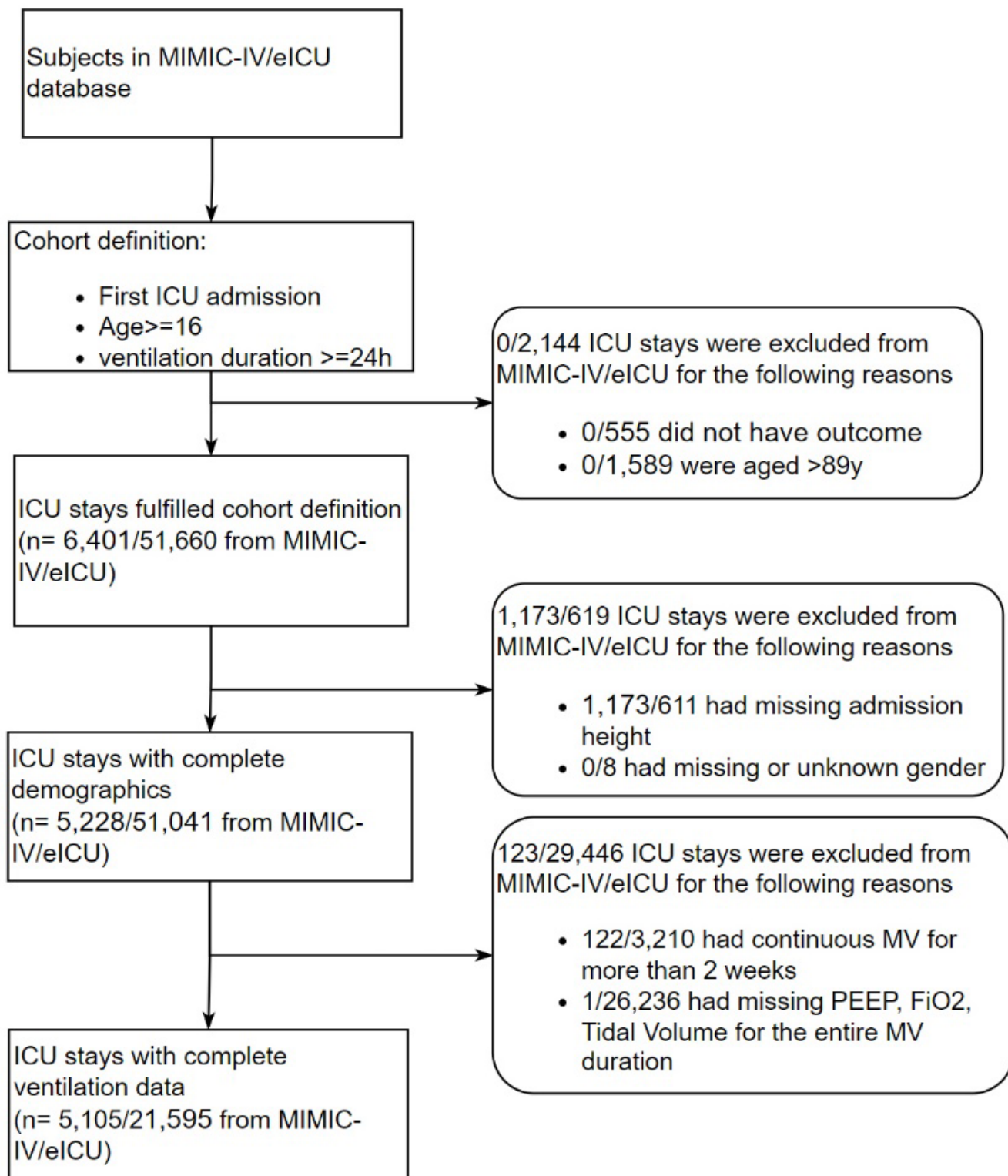


Figure 2

Flowchart of cohort

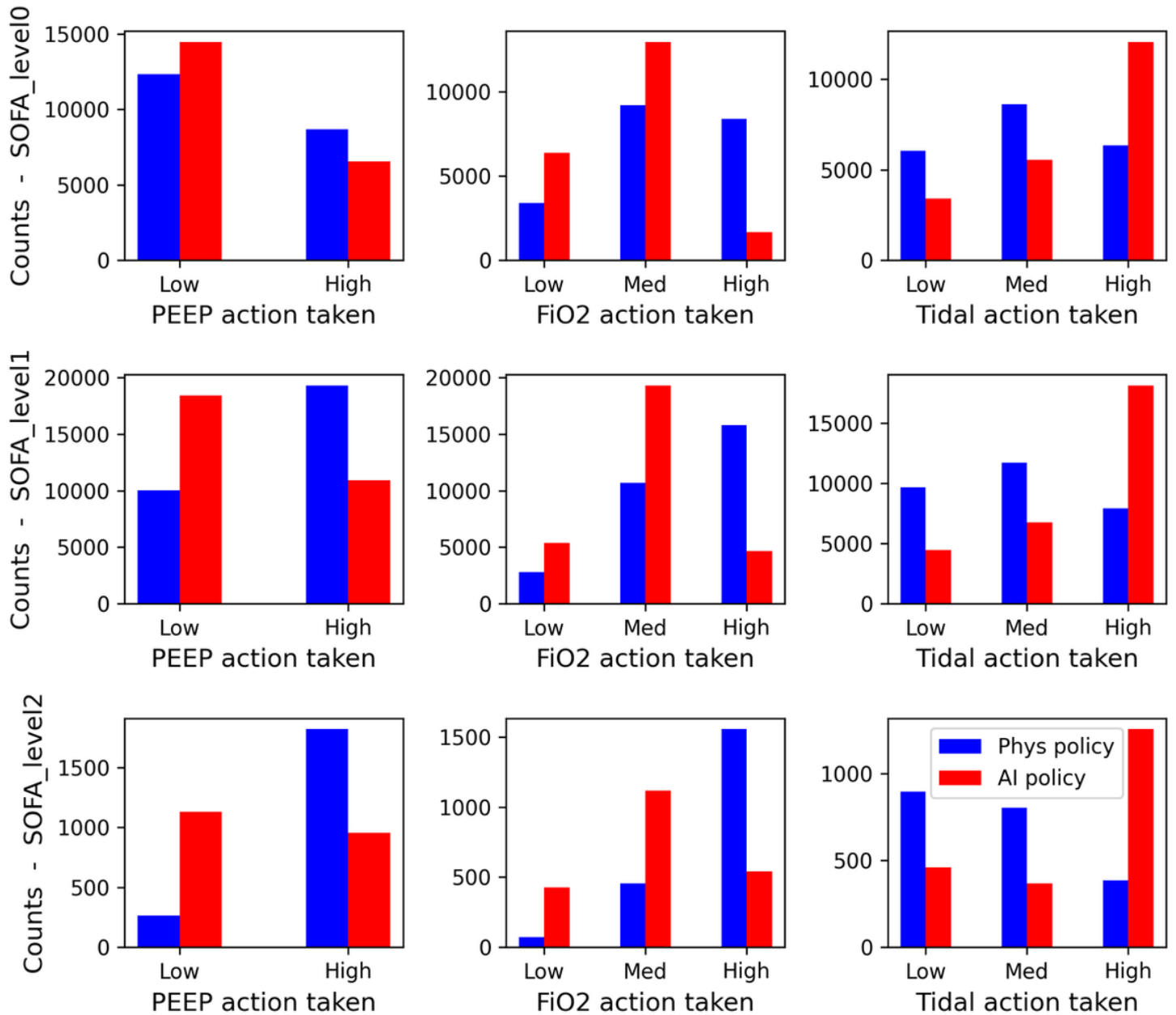
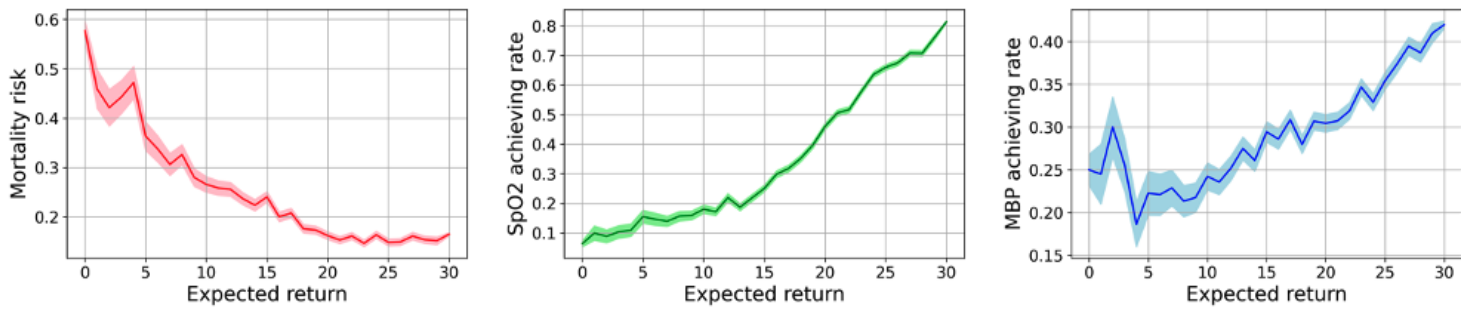


Figure 3

Action distributions for physicians (blue) and learnt policy (red) for patients with low/mid/high levels of SOFA score in MIMIC-IV test set

Note: SOFA level 0 (low): SOFA<5, SOFA level 1 (mid): 5<SOFA<15, SOFA level 2 (high): SOFA>15

eICU validation set



MIMIC-IV test set

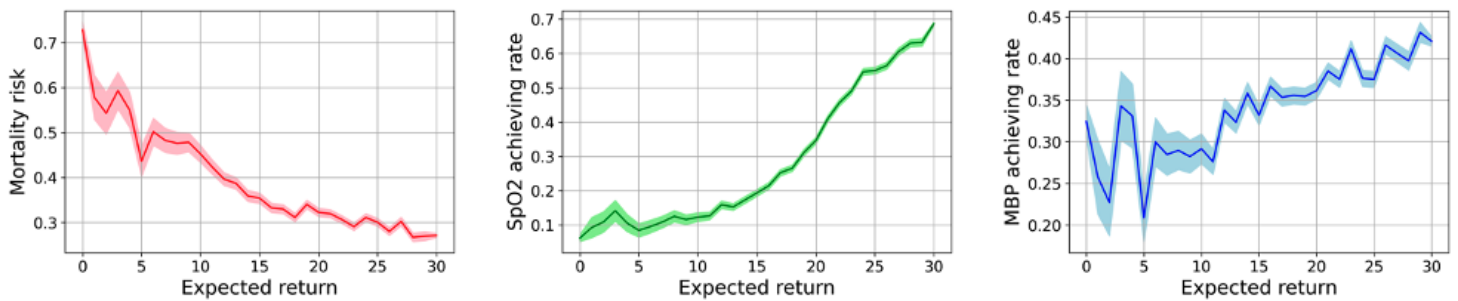


Figure 4

Changes in observed mortality (red), SpO2 optimal rate (green) and MBP optimal rate (blue) vs. the expected return curves for learnt policies in eICU validation set and MIMIC-IV test set.

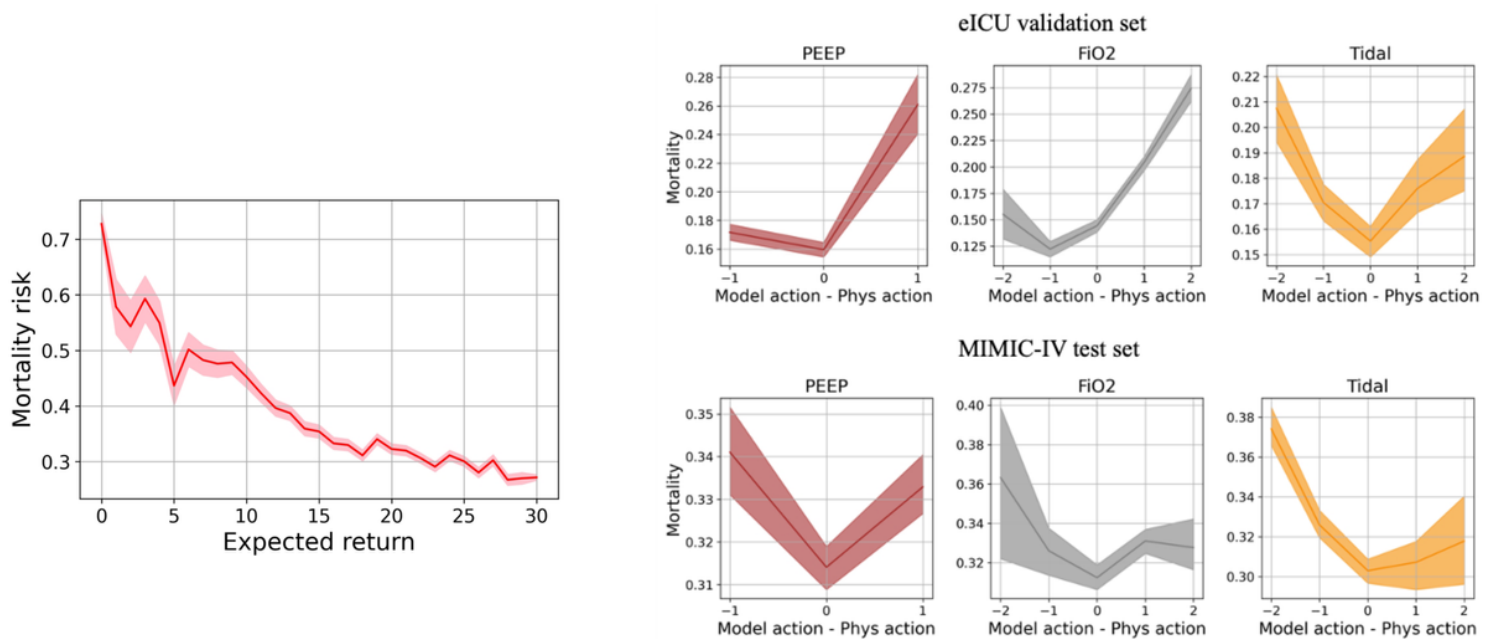


Figure 5

Changes in observed mortality (y-axis) versus the difference between the mechanical ventilation settings recommended by the optimal policy and the settings administered by clinicians (x-axis) on eICU validation set and MIMIC-IV test set

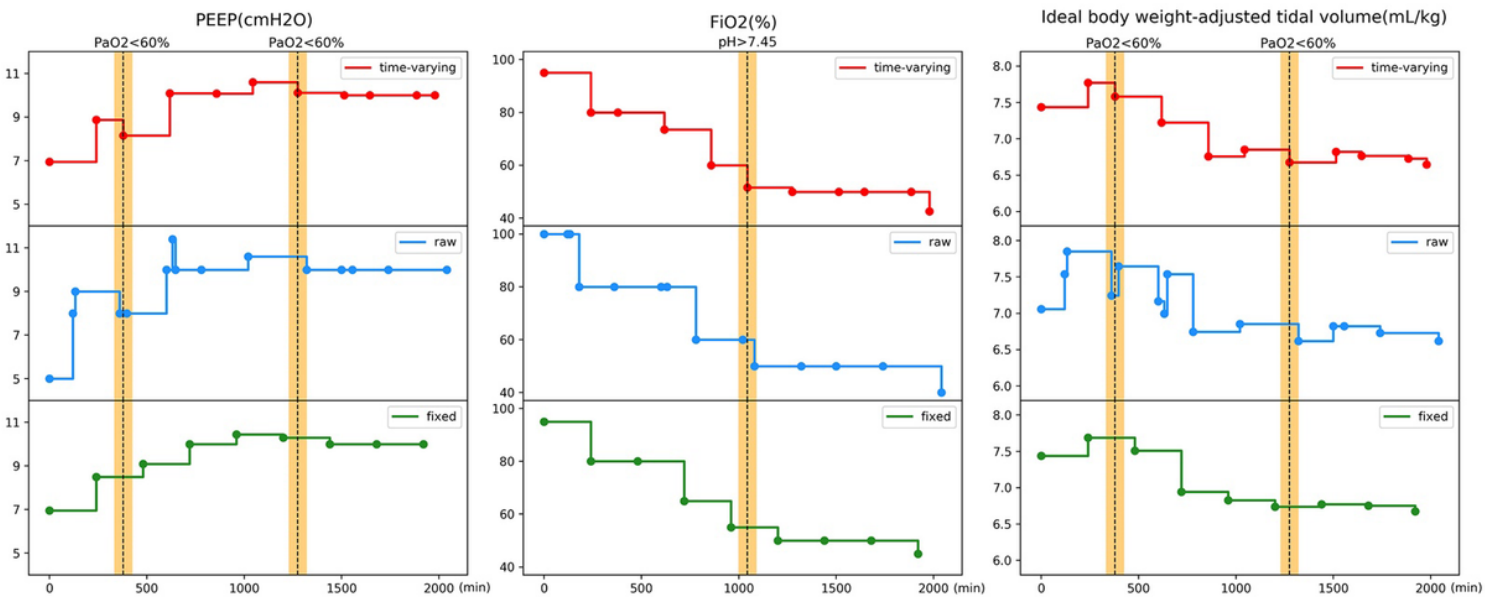


Figure 6

Visualization of representative patient case in raw/fix/time-varying interval for mechanical ventilator action setting. Visualization of representative case study for mechanical ventilator settings of PEEP (left), FiO2 (middle), Ideal body weighted-adjusted tidal volume (right) using time-varying interval (red), raw data (blue), fixed 4-hourly time interval (green). Flags to cut intervals in time-varying setting are showed as vertical dotted lines with yellow shadows.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementary.docx](#)