Predicting and simulating effects of PEEP changes with machine learning

Claas Strodthoff, Inéz Frerichs, Norbert Weiler, Björn Bergh

Abstract—Background/Objectives Choosing ventilator settings, especially positive end-expiratory pressure (PEEP), is a very common and non-trivial task in intensive care units (ICUs). Established solutions to this problem are either poorly individualised or come with high costs in terms of used material or time. We propose a novel method relying on machine learning utilising only already routinely measured data.

Methods Using the MIMIC-III (with over 60000 ICU stays) and eICU databases (with over 200000 ICU stays) we built a deep learning model that predicts relevant success parameters of ventilation (oxygenation, carbon dioxide elimination and respiratory mechanics). We compare a random forest, individual neural networks and a multi-tasking neural network. Our final model also allows to simulate the expected effects of PEEP changes.

Results The model predicts arterial partial pressures of oxygen and carbon dioxide and respiratory system compliance 30 minutes into the future with mean absolute percentage errors of about 22 %, 10 % and 11 %, respectively.

Conclusions The deep learning approach to ventilation optimisation is promising and comes with low cost compared to other approaches.

Index Terms—MIMIC-III, eICU, PEEP, Ventilation, Deep Learning, artificial neural networks

I. INTRODUCTION

Mechanical ventilation is one of the central purposes of intensive care units (ICU). Selection of ventilator parameters is being performed countless times every day worldwide. One of the most controversial settings is the positive end-expiratory pressure (PEEP). Large clinical studies showed the importance of PEEP selection for patient mortality, although not in a consistent way [1], [2]. The least one can conclude from these studies is that setting an individually adequate PEEP is an important and difficult task.

There is a range of established methods for PEEP optimisation. PEEP can be set as a function of the required oxygen fraction [3] or chosen to optimise compliance following a PEEP trial [4]. More sophisticated methods make use of additional medical equipment such as an esophageal pressure probe [5] or electrical impedance tomography [6], [7]. While these are very promising approaches which offer new insight into the topic, their use is associated with extra material cost and it requires extra effort of trained personnel. Probably the most common method to optimise PEEP relies on personal clinical experience as well as trial and error, which is hard to

evaluate objectively. All in all, there is no consensus about the best approach to optimise PEEP.

In this paper we present an alternative tool which does not have the aforementioned drawbacks of lacking individualisation or high cost. Taking advantage of the advances in machine learning, in particular deep learning, as well as the current development to record and save more and more data in the ICU setting, we developed a deep learning model to help with this decision. The aim of mechanical ventilation is to deliver sufficient gas exchange while at the same time causing as little harm as possible to the patient. Therefore, with this approach we offer the clinician real-time information and also simulations for relevant target variables of oxygenation, carbon dioxide elimination and respiratory system mechanics.

Machine learning-based clinical decision support systems become increasingly common in the ICU [8]. With respect to this work, some publications are especially important: Similar to our study, Ghazal et al. tried to predict oxygen saturation in infants after a change in ventilator settings [9]. The authors ultimately did not reach satisfactory accuracy, supposedly because of insufficient data. Relying on classical machine learning techniques, Luo et al. predicted laboratory results from previously known other laboratory results [10], however, without incorporating other measurement data. Komorowski et al. aimed to find the best treatment regime for the circulatory system (catecholamines versus intravenous fluids) in sepsis [11]. They used a reinforcement learning approach, which is considerably more complicated than the supervised learning approach used in this work but offers the possibility to consider the treatments' long-term effects like mortality more easily.

II. METHODS

A. Data Sources

1) *MIMIC-III dataset*: In the field of intensive care databases for machine learning there is one database that stands out in its size, quality and frequency of use in machine learning publications. Medical Information Mart for Intensive Care III (MIMIC-III) is an openly accessible database of over 40000 patients that were treated in the ICU of the Beth Israel Deaconess Medical Center in Boston between 2001 and 2012 [12]. The database includes demographic data, vital signs, laboratory results, interventions, medication, notes of the medical personnel, radiologic findings and mortality data of over 60000 ICU stays. We used this database as a training and test (10%) set.

2) *eICU dataset*: The eICU Collaborative Research Database holds data associated with over 200000 patient

Claas Strodthoff, Inéz Frerichs and Norbert Weiler are with Department of Anaesthesiology and Intensive Care Medicine, University Medical Centre Schleswig-Holstein, Campus Kiel, Kiel, Germany, e-mail: claas.strodthoff@uksh.de.

Björn Bergh is with Department of Medical Informatics, University Medical Centre Schleswig-Holstein, Campus Kiel, Kiel, Germany.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

stays recorded in numerous ICUs in the United States, made available by Philips Healthcare in partnership with the MIT Laboratory for Computational Physiology [13]. As far as the contents are concerned, this database is very similar to the MIMIC-III database with the advantage of being more diverse because of its multiple origins. In this work we used this second database as an additional independent test set which the model never sees during training.

For both databases we only used data from ventilated patients at least 16 years old.

B. Data Preprocessing

We started with all time points where the desired target variable has been measured (ground truth). For each of these events, we collected the desired input variables according to a predefined template from the minutes and hours before (see table I). The used items have different characteristics: items can either be set (F_iO_2 , PEEP, norepinephrine rate) or measured (S_pO_2 , temperature, mean arterial pressure); they can be timed (S_pO_2 , F_iO_2) or constant over one ICU stay (age, height); they can be continuous (S_pO_2 , temperature) or categorical (ventilator mode, sex). From these items we always took the latest measurement that was available before the time of measurement of the target variable. In order to include older measurements for some items, we discarded the latest n minutes before selecting a measurement (denoted as "item name [n]"). Whenever columns contained missing values, an extra boolean column indicating the presence/absence of values was added and the missing values were imputed as the mean. We performed only minimal data curation: removal of impossible values and conversion of units or representations to be identical for one variable in case they varied considerably between but also within one database(s) (temperature, height, weight, fractions vs. percentages, ventilator mode names).

In this paper, we present two similar setups, representing two clinical use cases. In the *all data* setup, the model predicts one or more target variables at a certain point in time, using all data that is available up to that time. This aims to help clinicians with real-time information instead of them having to wait for the results of a blood gas analysis or other measurement. Figure 1 illustrates how we selected and preprocessed measurements for the *all data* setup. The *30 min blinded* setup is identical to *all data* except that all measured variables (i.e. variables that are not set) in the last 30 minutes are removed. This corresponds to a prediction 30 to 60 minutes into the future. This model is used in the simulation setting which is explained below in section II-E.

From the MIMIC-III/eICU dataset, we included a total of 205736/133234, 205737/131800 and 207956/170851 individual observations for P_aO_2 , P_aCO_2 and respiratory system compliance C_{rs} , respectively.

C. Model selection, training, and testing

The deep learning model uses a fully connected artificial neural network with 2 layers with 1024 and 512 nodes. Categorical variables were represented using embeddings of sizes 3 to 7. Settings for the training were as follows: Learning rate

variable	value	variable	value
P_aO_2 [dep]	87.2 mmHg	RR_{meas} [30]	10 /min
P_aCO_2 [dep]	36.5 mmHg	RR_{meas} [60]	10 /min
C_{rs} [dep]	NA	RR_{meas} [120]	5 /min
respiratory		VT_{set} [0]	700 mL
P_aO_2 [30]	88.8 mmHg	VT_{set} [30]	700 mL
P_aO_2 [60]	88.8 mmHg	VT_{set} [60]	700 mL
P_aO_2 [120]	88.8 mmHg	VT_{set} [120]	700 mL
P_aO_2 [240]	88.8 mmHg	VT_{meas} [30]	NA
P_aO_2 [480]	215 mmHg	VT_{meas} [60]	NA
P_aCO_2 [30]	31.2 mmHg	VT_{meas} [120]	NA
P_aCO_2 [60]	31.2 mmHg	$VentMode$ [0]	NA
P_aCO_2 [120]	31.2 mmHg	circulatory	
P_aCO_2 [240]	31.2 mmHg	HR [30]	91 /min
P_aCO_2 [480]	37 mmHg	HR [120]	92 /min
C_{rs} [30]	NA	HR [480]	97 /min
C_{rs} [60]	NA	MAP [30]	72 mmHg
C_{rs} [120]	NA	MAP [120]	69 mmHg
C_{rs} [240]	NA	MAP [480]	73 mmHg
C_{rs} [480]	NA	NE [0]	0 $\mu g/(kg \cdot min)$
S_pO_2 [30]	0.98	laboratory	
S_pO_2 [60]	0.98	albumin [30]	NA
S_pO_2 [120]	1.00	bilirubin [30]	NA
$PEEP_{set}$ [30]	5 cmH ₂ O	creatinine [30]	1.00 mg/dL
$PEEP_{set}$ [60]	5 cmH ₂ O	lactate [30]	NA
$PEEP_{set}$ [120]	5 cmH ₂ O	temperature [30]	38 °C
$PEEP_{set}$ [240]	5 cmH ₂ O	platelets [30]	73 /nL
$PEEP_{set}$ [480]	5 cmH ₂ O	leukocytes [30]	7.22 /nL
$P_{peakinsp}$ [30]	NA	Hb [30]	8.7 g/dL
$P_{peakinsp}$ [60]	NA	pH [30]	7.513
$P_{peakinsp}$ [120]	NA	INR [30]	1.49
F_iO_2 [0]	0.4	demographics	
F_iO_2 [30]	0.4	age [-]	73 a
F_iO_2 [60]	0.4	gender [-]	male
F_iO_2 [120]	0.4	height [-]	157 cm
F_iO_2 [240]	0.4	weight [-]	119 kg
F_iO_2 [480]	1.0		
RR_{set} [0]	11 /min		
RR_{set} [30]	11 /min		
RR_{set} [60]	11 /min		
RR_{set} [120]	11 /min		

TABLE I: Example of a single sample for the *30 min blinded* setup, totalling 71 items per event. P_aO_2 : arterial oxygen partial pressure, P_aCO_2 : arterial carbon dioxide partial pressure, C_{rs} : respiratory system compliance, S_pO_2 : peripheral oxygen saturation (oximetry), PEEP: positive end-expiratory pressure, $P_{peakinsp}$: peak inspiratory pressure, F_iO_2 : inspired oxygen fraction, RR: respiratory rate, VT: tidal volume, HR: heart rate, MAP: mean arterial pressure, NE: norepinephrine rate, Hb: haemoglobin, INR: International Normalised Ratio; subscript *set/meas*: set/measured, numbers in square brackets indicate the time in minutes up to which the variable was blinded, [dep]: dependent variable, NA: not available.

$2 \cdot 10^{-3}$ with Adam optimiser, 20 epochs. For regularisation we used a weight decay of 0.1 and a dropout of 0.04, 0.05 and 0.5 for the embeddings and the first and second fully connected layers, respectively. We did not perform any hyperparameter tuning on test data.

With the used target variables, we care for relative rather than absolute errors. Therefore we opted for mean squared logarithmic error (MSLE) as our loss function. We trained the model for every target variable individually as well as together in the sense of multi-task learning [14]. In the latter case, the mean of the individual loss functions for each task was used as loss function. We used the MIMIC-III dataset as training and

	random forest				neural network, individual models				neural network, multi-task			
	MIMIC-III test set		eICU test set		MIMIC-III test set		eICU test set		MIMIC-III test set		eICU test set	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
- all data -												
P _a O ₂ [mmHg]	27.15	21.76%	43.80	32.16%	28.66	22.49%	46.98	42.56%	27.16	21.11%	41.81	38.74%
P _a CO ₂ [mmHg]	4.496	10.66%	6.688	15.61%	4.498	10.55%	7.644	18.66%	4.264	10.00%	5.492	13.27%
C _{rs} [mL/cmH ₂ O]	5.029	16.82%	10.48	31.67%	5.155	17.09%	12.14	28.26%	3.201	11.20%	9.195	22.54%
- 30 min blinded -												
P _a O ₂ [mmHg]	28.16	22.67%	44.68	33.12%	29.44	23.39%	46.60	42.28%	28.09	22.20%	43.04	41.17%
P _a CO ₂ [mmHg]	4.547	10.79%	6.804	15.83%	4.557	10.68%	7.509	18.26%	4.338	10.15%	5.452	12.77%
C _{rs} [mL/cmH ₂ O]	5.059	17.07%	10.17	29.71%	5.169	17.21%	11.92	28.25%	3.240	11.33%	8.957	22.47%

TABLE II: Performance of models for both setups. Best (lowest) metrics highlighted. P_aO₂: arterial oxygen partial pressure, P_aCO₂: arterial carbon dioxide partial pressure, C_{rs}: respiratory system compliance, MAE: mean absolute error, MAPE: mean absolute percentage error.

Top	P _a O ₂	P _a CO ₂	C _{rs}
1	F _i O ₂ [0]	P _a CO ₂ [30]	C _{rs} [30]
2	P _a O ₂ [30]	VT _{set} [0]	VT _{set} [0]
3	F _i O ₂ [30]	VentMode [0]	P _{peakinsp} [30]
4	S _p O ₂ [30]	RR _{set} [30]	C _{rs} [480]
5	F _i O ₂ [120]	RR _{set} [120]	RR _{set} [0]
6	F _i O ₂ [480]	P _a CO ₂ [480]	VentMode [0]
7	PEEP _{set} [30]	F _i O ₂ [0]	PEEP _{set} [30]
8	RR _{set} [0]	pH [30]	C _{rs} [60]
9	VentMode [0]	P _a CO ₂ [60]	C _{rs} [240]
10	S _p O ₂ [120]	VT _{meas} [30]	RR _{set} [120]
...			
26		PEEP _{set} [30]	

TABLE III: Top 10 most important features ranked by permutation importance for different target variables (multi-task model for the 30 min blinded setup); P_aO₂: arterial oxygen partial pressure, P_aCO₂: arterial carbon dioxide partial pressure, C_{rs}: respiratory system compliance, F_iO₂: inspired oxygen fraction, PEEP: positive end-expiratory pressure, RR: respiratory rate, VentMode: ventilator mode, S_pO₂: peripheral oxygen saturation (oximetry), VT: tidal volume, P_{peakinsp}: peak inspiratory pressure; subscript *set/meas*: set/measured, numbers in square brackets indicate the time in minutes up to which the variable was blinded.

and possibly a low transpulmonary pressure which might mitigate ventilator associated lung injury [19]. The time it takes for different variables to reach steady states after a PEEP change ranges from 5 to over 60 minutes and depends on the variables in question and on the direction of the PEEP step [20]. We chose 30 minutes as our prediction horizon as a compromise between the quantity of expected changes and the uncertainty and thus difficulty of the prediction task. Because of the binning of the variables, this choice is equivalent to a prediction 30 to 60 minutes into the future.

Methods for clinical decision support become more and more common. The strength of this tool is that it offers a very high degree of individualisation while at the same time being relatively easy to apply. Once the digital infrastructure is set up, the model takes a small number of routine measurements and outputs results - there is no need for additional material, measurements, interventions or manoeuvres. This simplicity is also true for the model itself, which not only makes it easier to understand and explain to clinical users, it also makes it possible to train the model in under 5 minutes on a single consumer graphics processing unit (GPU). This means that

the model can easily be trained or fine-tuned regularly or when circumstances change, like during the beginning of a pandemic. Sometimes clinical decision support systems seem to take the power out of the clinician's hand because they learn to rely on a proposed decision. The advantage of the solution presented here is that the clinicians still can and have to balance the expected effects on the different target variables and make their own decisions. We therefore refrain from offering concrete rules as to how one should react to certain model or simulation outputs.

With mean absolute percentage errors of 22%, 10% and 11% for the three tasks on the MIMIC-III test set, the model appears to reach a clinically useful performance level, especially for P_aCO₂ and C_{rs}. To determine the caretaker acceptance and the perceived and actual benefit for patient care, a clinical study would be needed. When comparing the results of the *all data* and *30 min blinded* setup, the differences in prediction error are surprisingly small. Discarding 30 minutes of measurements does not seem to increase the difficulty of the task to a relevant degree. This may be in part explained by the fact that all target variables tend to change slowly, leaving the last known value as a very good guess. Another reason might be the high variability of the measurement of the ground truth, which for P_aO₂ and P_aCO₂ is subject to numerous preanalytical and (allowable) analytical errors [21], [22]. The fact that models were tested on a separate dataset from a different source and still performed well shows its generalisation capacity. The performance gap between the MIMIC-III test set (held separate from the training set) and the eICU test set is probably due to structural differences between the databases, their collection and the underlying treatments. We are confident that the real-world performance after fine-tuning with local data will be closer to the MIMIC-III test set performance. We reached our best result using the multi-tasking neural network. It outperforms both the random forest and the individual neural networks except for the MAPE metrics for P_aO₂ in the eICU test set in both setups and the MAE metric for P_aO₂ in the MIMIC test set in the *all data* setup, where the random forest performed better. The results clearly show the advantage of the multi-tasking model over the individual models. This is a known phenomenon for related tasks where the model may profit from the information gained from one task in another task [14]. Additionally, multi-tasking models simplify training and inference because one has to

train, evaluate and store only one model. We chose the mean of the individual losses as our multi-tasking loss because of simplicity and because the individual losses were sufficiently similar in size as not to overpower one another.

Our analysis of feature importance shows that for all tasks high importance is attributed to features that seem clinically reasonable. Specifically, most important features naturally include the last known measurement of the variable to be predicted and variables that are closely correlated. For P_{aO_2} these include inspired oxygen fraction and oximetric peripheral oxygen saturation, for P_{aCO_2} these are the set tidal volume and set respiratory rate as well as pH, and for C_{rs} the set tidal volume and peak inspiratory pressure. For all target variables, the currently set ventilator mode occupies a top 10 position, showing that the model gathers valuable information from this categorical variable's embeddings. The last known set PEEP occurs at positions 7, 26 and 7, respectively. Only the reason for the occurrence of the set respiratory rate (for P_{aO_2} and C_{rs}) and inspired oxygen fraction (for P_{aCO_2}) are not immediately obvious. It is important to note that this permutation based approach to determine feature importance should only be seen as a very rough peek under the model's hood as it is ignorant of the complex interactions of multiple input variables because it considers every variable independently [23]. All in all, these results can be seen as a reason to have confidence in the decision process of the proposed model. This kind of interpretability is of major importance for the acceptance of machine learning solutions by clinicians and the population in general. Similar to the simulation approach used in this work, it would also be possible to better understand the model's decision for a single case by manipulating or obscuring individual input variables.

From the simulation example (figure 2) we get an impression of how this tool might be used practically in clinical care. In the two shown cases, one has an expected increase in P_{aO_2} in response to an increased PEEP while the other does not. This kind of graph might be useful for a clinician who is trying to optimise oxygenation or carbon dioxide elimination, especially in the short term.

A possible shortcoming of the simulation design we applied here is the risk of information leakage in the training and testing phases. Even if the model is blinded for all measured variables within the prediction horizon, it has still access to the set variables from this time period, which may be a clinician's reaction to a measurement. However, this situation will not affect the model performance as no wrong information is introduced. As already mentioned above, the model succeeded to transfer knowledge from one US database to another. It could be more challenging, however, to transfer to the situation in Europe, for example because pressure controlled ventilation is much more common there. Additionally, medical knowledge has a short half-time for example as patient demographics change or new treatment options arise. What has been accurate for the MIMIC-III dataset recorded between 2001 and 2012 may be inaccurate in 2021. This makes fine-tuning on recent local data even more important. One big problem we encountered, as one does in many machine learning applications, was the quality of the data. While the MIMIC-III database is

often praised for its quality, these large databases still suffer from missing and corrupt data or incomplete documentation. This becomes especially evident when moving across different databases. Finally, due to the chosen approach in this work, it is not easily possible to include long-term clinical endpoints like mortality into the prediction because we use a fixed prediction horizon. For these endpoints the much more complicated reinforcement learning approach would be necessary [11].

To be able to use the developed model in Europe, we plan to continue training the model on data from local ICUs. A possible next step would be to include more ventilator settings into the simulation part, eventually making it possible to try out different ventilator modes, pressures and timings without even having to bother (and possibly harm) the patient. The relatively simple process and model design we applied to a ventilation problem here would also be suited for a wide range of applications outside of ventilation therapy, for example the prediction of drug concentrations, regulation of coagulation, fluid balance or circulation. For all these applications, clinicians currently rely mostly on experience or rules of thumb and would definitely benefit from a reliable and objective prediction.

V. CONCLUSION

Deep learning for tabular data is a useful approach for this clinically relevant problem of ventilation optimisation. It delivers reliable results and comes with no extra cost once the digital infrastructure is in place. There is a large potential for this tabular approach for numerous other applications in the ICU.

REFERENCES

- [1] M. Briel, M. Meade, A. Mercat, R. G. Brower, D. Talmor, S. D. Walter, A. S. Slutsky, E. Pullenayegum, Q. Zhou, D. Cook, L. Brochard, J.-C. M. Richard, F. Lamontagne, N. Bhatnagar, T. E. Stewart, and G. Guyatt, "Higher vs lower positive end-expiratory pressure in patients with acute lung injury and acute respiratory distress syndrome," *JAMA*, vol. 303, p. 865, Mar. 2010.
- [2] A. B. Cavalcanti, É. A. Suzumura, L. N. Laranjeira, D. de Moraes Paisani, L. P. Damiani, H. P. Guimarães, E. R. Romano, M. de Moraes Regenga, L. N. T. Taniguchi, C. Teixeira, R. P. de Oliveira, F. R. Machado, F. A. Diaz-Quijano, M. S. de Alencar Filho, I. S. Maia, E. B. Caser, W. de Oliveira Filho, M. de Carvalho Borges, P. de Aquino Martins, M. Matsui, G. A. Ospina-Tascón, T. S. Giancursi, N. D. Giraldo-Ramirez, S. R. R. Vieira, M. da Graça Pasquotto de Lima Assef, M. S. Hasan, W. Szczeklik, F. Rios, M. B. P. Amato, O. Berwanger, and C. R. R. de Carvalho, "Effect of lung recruitment and titrated positive end-expiratory pressure (PEEP) vs low PEEP on mortality in patients with acute respiratory distress syndrome," *JAMA*, vol. 318, p. 1335, Oct. 2017.
- [3] "Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome," *New England Journal of Medicine*, vol. 342, pp. 1301–1308, May 2000.
- [4] R. M. Kacmarek, J. Villar, D. Sulemanji, R. Montiel, C. Ferrando, J. Blanco, Y. Koh, J. A. Soler, D. Martínez, M. Hernández, M. Tucci, J. B. Borges, S. Lubillo, A. Santos, J. B. Araujo, M. B. P. Amato, and F. Suárez-Sipmann, "Open lung approach for the acute respiratory distress syndrome," *Critical Care Medicine*, vol. 44, pp. 32–42, Jan. 2016.
- [5] D. Talmor, T. Sarge, A. Malhotra, C. R. O'Donnell, R. Ritz, A. Lisbon, V. Novack, and S. H. Loring, "Mechanical ventilation guided by esophageal pressure in acute lung injury," *New England Journal of Medicine*, vol. 359, pp. 2095–2104, Nov. 2008.

- [6] E. L. V. Costa, J. B. Borges, A. Melo, F. Suarez-Sipmann, C. Toufen, S. H. Bohm, and M. B. P. Amato, "Bedside estimation of recruitable alveolar collapse and hyperdistension by electrical impedance tomography," *Intensive Care Medicine*, vol. 35, pp. 1132–1137, Mar. 2009.
- [7] G. Zick, G. Elke, T. Becher, D. Schädler, S. Pulletz, S. Freitag-Wolf, N. Weiler, and I. Frerichs, "Effect of PEEP and tidal volume on ventilation distribution and end-expiratory lung volume: A prospective experimental animal and pilot clinical study," *PLoS ONE*, vol. 8, p. e72675, Aug. 2013.
- [8] D. Shillan, J. A. C. Sterne, A. Champneys, and B. Gibbison, "Use of machine learning to analyse routinely collected intensive care unit data: a systematic review," *Critical Care*, vol. 23, Aug. 2019.
- [9] S. Ghazal, M. Sauthier, D. Brossier, W. Bouachir, P. A. Jouviet, and R. Noumeir, "Using machine learning models to predict oxygen saturation following ventilator support adjustment in critically ill children: A single center pilot study," *PLOS ONE*, vol. 14, p. e0198921, Feb. 2019.
- [10] Y. Luo, P. Szolovits, A. S. Dighe, and J. M. Baron, "Using machine learning to predict laboratory test results," *American Journal of Clinical Pathology*, vol. 145, pp. 778–788, June 2016.
- [11] M. Komorowski, L. A. Celi, O. Badawi, A. C. Gordon, and A. A. Faisal, "The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care," *Nature Medicine*, vol. 24, pp. 1716–1720, Oct. 2018.
- [12] A. E. Johnson, T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, R. G. Mark, and et al., "Mimic-iii, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 1, 2016.
- [13] T. J. Pollard, A. E. W. Johnson, J. D. Raffa, L. A. Celi, R. G. Mark, and O. Badawi, "The eICU collaborative research database, a freely available multi-center database for critical care research," *Scientific Data*, vol. 5, Sept. 2018.
- [14] S. Ruder, "An overview of multi-task learning in deep neural networks." <https://arxiv.org/abs/1706.05098>, 2017.
- [15] J. Howard et al., "fast.ai." <http://fast.ai>, 2018.
- [16] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [17] E. C. Goligher, B. P. Kavanagh, G. D. Rubenfeld, N. K. J. Adhikari, R. Pinto, E. Fan, L. J. Brochard, J. T. Granton, A. Mercat, J.-C. M. Richard, J.-M. Chretien, G. L. Jones, D. J. Cook, T. E. Stewart, A. S. Slutsky, M. O. Meade, and N. D. Ferguson, "Oxygenation response to positive end-expiratory pressure predicts mortality in acute respiratory distress syndrome: a secondary analysis of the LOVS and ExPress trials," *American Journal of Respiratory and Critical Care Medicine*, vol. 190, pp. 70–76, July 2014.
- [18] P. M. Suter, H. B. Fairley, and M. D. Isenberg, "Optimum end-expiratory airway pressure in patients with acute pulmonary failure," *New England Journal of Medicine*, vol. 292, pp. 284–289, Feb. 1975.
- [19] L. Gattinoni, J. J. Marini, F. Collino, G. Maiolo, F. Rapetti, T. Tonetti, F. Vasques, and M. Quintel, "The future of mechanical ventilation: lessons from the present and the past," *Critical Care*, vol. 21, July 2017.
- [20] D. Chiumello, S. Coppola, S. Froio, C. Mietto, L. Brazzi, E. Carlesso, and L. Gattinoni, "Time to reach a new steady state after changes of positive end expiratory pressure," *Intensive Care Medicine*, vol. 39, no. 8, p. 1377–1385, 2013.
- [21] G. Baird, "Preanalytical considerations in blood gas analysis," *Bio-chemia Medica*, pp. 19–27, 2013.
- [22] N. authors listed, "Medicare, Medicaid and CLIA programs; regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA) and Clinical Laboratory Improvement Act Program fee collection–HCFA. Final rule with comment period," *Fed Regist*, vol. 58, pp. 5215–5237, Jan 1993.
- [23] G. Hooker and L. Mentch, "Please stop permuting features: An explanation and alternatives." <https://arxiv.org/abs/1905.03151>, 2019.