Predicting Blood Pressure Response to Fluid Bolus Therapy Using Attention-Based Neural Networks for Clinical Interpretability

Background Excessive positive fluid balance in critically ill patients in intensive care units (ICUs) has been proposed as a risk factor for severe organ dysfunctions, prolonged mechanical ventilation, longer length of stay in ICU, and increased mortality [10, 12, 15]. Fluid bolus therapy (FBT), the rapid infusion of fluid, has been recommended as the primary-line treatment for acute hypotensive episode (AHE) that occurs in about 41% of patients in ICU [13]. However, previous studies have reported that approximately one-third of AHE cases are not successfully resolved by FBT treatment [5, 9, 11].Considering that FBT accounts for about 30-50% of the total fluid volume administrated for ICU patients [2], identifying patients for whom FBT would not likely resolve AHE may prevent an inappropriate increase of the total fluid volume. [9]. Previous studies have focused on predicting cardiac output (CO) as patients' response to FBT since fluid infusion increases CO by increasing the venous return to heart [14]. A 15% increase in CO after FBT has been used as the success benchmark [6]. However, in clinical settings, only 11% of ICU clinicians use the increased CO as the patient's positive response to FBT. 67% of them judged the patient's response as positive when the patient's blood pressure (BP) increases [3]. Of course CO is a gold-standard measure to evaluate patient's physiological response to FBT, but the increased BP might be used as a practical alternative to evaluate the patient's response to FBT in the ICU.

A previous study had constructed the non-machine learning models to predict BP response to FBT, yet the performance of the area under the curve (AUC) ranged from 0.5 to 0.6, which was not acceptable to be implemented in the clinical setting [11]. Therefore, we investigated both time-aggregated and time-series structured data for modeling. In this study, we applied regularized logistic regression, as well as the stacked long short term memory network (LSTM) and gated recurrent units network (GRU) models with and without the attention mechanism to identify hypotensive critically ill patients in the ICU who would obtain sufficient BP recovery after the FBT [1, 4, 7].

The goal of this study is to achieve high model performance and clinical interpretability for real-world implementation. Particularly, the contributions of this study include the following:

1. This is the first study that utilizes machine learning algorithms to develop models for predicting successful blood pressure response to FBT in critically ill patients in the ICU
2. The regularized regression model and LSTM/GRU models with the attention mechanism provide us with certain important features for clinical interpretability.

Methods

Dataset and Cohort Study data was collected from the MIMIC-III database [8], which contains 58,976 ICU patients admitted to the Beth Israel Deaconess Medical Center (BIDMC). For the cohort selection, we considered only (1) the first ICU stay during the hospital stay, (2) patients who were more than 18 years old on the first day of admission, (3) patients with a length of ICU stay more than 12 hours in order to include only true ICU patients, (4) patients who received their first FBT during their first 24 hours in the ICU, where FBT is defined as the crystalloid fluid infusion rate $>$248 ml/hr and volume $>$248 ml , and (5) patients who are hypotensive (mean atrial pressure (MAP) =< 65 mmHg) when the first FBT started. 17,977 patients were selected for the final patient cohort.

Clinical Features and Outcome We extracted 29 clinically meaningful features from the MIMIC-III database, which include time-static features: (1) patient demographics (age, gender, race/ethnicity, weight, height, and SOFA score at ICU admission) and (2) comorbidity condition using Elixhauser coding algorithm, and time-varying features: (1) physiological parameters (heart rate, respiratory rate, temperature, oxygen saturation, systolic blood pressure, diastolic blood pressure, mean arterial pressure, and urine output), (2) laboratory examination results (pH, PaO2, PaCO2, bicarbonate, base excess, lactate, sodium, potassium and chloride), and (3) vasopressor dosage measured (norepinephrine, epinephrine, phenylephrine, vasopressin and dopamine). For time-aggregated modeling, we collected the data at the time interval between 30 minutes before and 30 minutes after 3 important time points — 6 hours before the first FBT, 2 hours before the first FBT, and right at the start of the first FBT. All raw values of features were normalized in population level and re-scaled to obtain a value between zero and one. The missing values were imputed by median values of the features. The primary outcome of this study is the physiological response, which is reflected by the change of MAP as a success or failure. The successful FBT is defined by intensive care experts as the presence of $max(\text{MAP}_{fbt}) > 1.15 \times average(\text{MAP}_{all})$ at least once, where $max(\text{MAP}_{fbt})$ is the maximal MAP from the FBT starting time to two hours after FBT, and $average(\text{MAP}_{all})$ is the average MAP from 30 minutes before FBT until 10 minutes after FBT.

Experiment Settings We split our cohort into a 75% training and 25% testing set. For both the time-aggregated and time-series settings, we investigated the normalized raw feature and autoencoder-constructed distributed representations. For the time aggregated setting, we used LASSO and Ridge logistic regression models and the multiple layer perceptron (MLP) model. For the time series setting, we collected data every certain number of minutes based on the number of timesteps specified. For time-varying covariate variables, data collection started six hours before the first FBT was administered and ended at the time when the first FBT was administered. We used 12, 36, and 72 timesteps for sequential modeling. We adopted the stacked three layer LSTM and GRU models with the attention mechanism for prediction.

Evaluation We computed the accuracy and AUC of all models and identified certain features for human experts to qualitatively evaluate the interpretability of the models. The top five features with the highest coefficients (absolute value) in LASSO regression, and the most important timesteps with the highest weights in the neural network models with attention mechanism were extracted for interpretation.

Results and Discussion The results of the binary classification task of predicting whether the FBT yielded successful blood pressure improvement using the different machine learning settings are shown in Table 1. In general, the attention-based neural network models which considered time sequence with higher temporal granularity yielded higher performance for prediction. The result indicates that the autoencoder-derived representation is compact but still informative even if it is in a lower dimension.

Algorithm	Timesteps	Accuracy		AUC	
		Raw features	Distributed	Raw features	Distributed
L1-regularized logistic regression	-	0.706	0.688	0.680	0.659
L2-regularized logistic regression	-	0.699	0.695	0.679	0.664
Multiple layer perceptron	-	**0.712**	0.688	**0.690**	0.642
LSTM	12	0.751	0.705	0.818	0.718
GRU	12	0.748	0.721	0.813	0.770
LSTM + Attention	12	0.747	0.727	0.822	0.795
GRU + Attention	12	0.747	0.716	0.818	0.786
LSTM	36	0.814	0.827	0.899	0.899
GRU	36	0.820	0.794	0.898	0.869
LSTM + Attention	36	0.819	0.820	0.902	0.895
GRU + Attention	36	0.812	0.777	0.893	0.858
LSTM	72	0.843	0.834	**0.926**	0.915
GRU	72	0.848	0.836	0.922	0.904
LSTM + Attention	72	**0.852**	0.831	0.925	0.920
GRU + Attention	72	0.841	0.803	0.917	0.882

Table 1: Model performance in accuracy and AUC between different experimental settings. **Boldface** denotes the best performance in each group.

The top features learned from LASSO regression should undergo further investigation to understand their clinical value. We are also able to identify the key timesteps using the RNN models with the attention mechanism by extracting the attention weights. The timesteps closer to FBT have higher impact to the prediction. The result is clinically meaningful since the time points closer to the time of FBT are the most important, which is explainable from the clinical perspective. For future work, we will investigate the optimal strategy to determine when and how much fluid bolus a patient should receive using reinforcement learning. On the clinical side, we will include the parameters of mechanical ventilation settings as covariates to make the model more robust. The study results may support intensive care clinicians to identify whether the hypotensive episode in ICU patients will resolve with FBT.

References

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *ICLR*, 2015.

[2] Shailesh Bihari, Shivesh Prakash, and Andrew D Bersten. Post resusicitation fluid boluses in severe sepsis or septic shock: prevalence and efficacy (price study). *Shock*, 40(1):28–34, 2013.

[3] Maurizio Cecconi, Christoph Hofer, Jean-Louis Teboul, Ville Pettila, Erika Wilkman, Zsolt Molnar, Giorgio Della Rocca, Cesar Aldecoa, Antonio Artigas, Sameer Jog, et al. Fluid challenges in intensive care: the fenice study. *Intensive care medicine*, 41(9):1529–1537, 2015.

[4] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8)*, 2014.

[5] Manuel Ignacio Monge García, Pedro Guijo González, Manuel Gracia Romero, Anselmo Gil Cano, Chris Oscier, Andrew Rhodes, Robert Michael Grounds, and Maurizio Cecconi. Effects of fluid administration on arterial load in septic shock patients. *Intensive care medicine*, 41(7):1247–1255, 2015.

[6] Neil J Glassford, Glenn M Eastwood, and Rinaldo Bellomo. Physiological changes after fluid bolus therapy in sepsis: a systematic review of contemporary data. *Critical care*, 18(6):696, 2014.

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

[8] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3:160035, 2016.

[9] Matthew R Lammi, Brianne Aiello, Gregory T Burg, Tayyab Rehman, Ivor S Douglas, Arthur P Wheeler, National Institutes of Health, ARDS Network Investigators, et al. Response to fluid boluses in the fluid and catheter treatment trial. *Chest*, 148(4):919–926, 2015.

[10] Joon Lee, Emma de Louw, Matthew Niemi, Rachel Nelson, Roger G Mark, Leo Anthony Celi, Kenneth J Mukamal, and John Danziger. Association between fluid balance and survival in critically ill patients. *Journal of internal medicine*, 277(4):468–477, 2015.

[11] Giuseppe Natalini, Antonio Rosano, Carmine Rocco Militano, Antonella Di Maio, Pierluigi Ferretti, Michele Bertelli, Federica de Giuli, and Achille Bernardini. Prediction of arterial pressure increase after fluid challenge. *BMC anesthesiology*, 12(1):3, 2012.

[12] Didier Payen, Anne Cornélie de Pont, Yasser Sakr, Claudia Spies, Konrad Reinhart, and Jean Louis Vincent. A positive fluid balance is associated with a worse outcome in patients with acute renal failure. *Critical care*, 12(3):R74, 2008.

[13] Mohammed Saeed, Christine Lieu, Greg Raber, and Roger G Mark. MIMIC II: a massive temporal icu patient database to support research in intelligent patient monitoring. In *Computers in Cardiology, 2002*, pages 641–644. IEEE, 2002.

[14] Laura Toscani, Hollmann D Aya, Dimitra Antonakaki, Davide Bastoni, Ximena Watson, Nish Arulkumaran, Andrew Rhodes, and Maurizio Cecconi. What is the impact of the fluid challenge technique on diagnosis of fluid responsiveness? a systematic review and meta-analysis. *Critical Care*, 21(1):207, 2017.

[15] Jean-Louis Vincent, Yasser Sakr, Charles L Sprung, V Marco Ranieri, Konrad Reinhart, Herwig Gerlach, Rui Moreno, Jean Carlet, Jean-Roger Le Gall, and Didier Payen. Sepsis in european intensive care units: results of the SOAP study. *Critical care medicine*, 34(2):344–353, 2006.