Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Conf Proc IEEE Eng Med Biol Soc. Author manuscript; available in PMC 2019 October 03.

Published in final edited form as:

Conf Proc IEEE Eng Med Biol Soc. 2018 July ; 2018: 1–4. doi:10.1109/EMBC.2018.8513185.

Brain Monitoring of Sedation in the Intensive Care Unit Using a Recurrent Neural Network

Haoqi Sun,
Department of Neurology, Massachusetts General Hospital, Boston, MA 02114 USA

Sunil B. Nagaraj,
Biomedical Signals and Systems Group, University of Twente, Enschede, Netherlands

Oluwaseun Akeju,
Department of Anesthesiology, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA 02114 USA

Patrick L. Purdon,
Department of Anesthesiology, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA 02114 USA

M. Brandon Westover
Department of Neurology, Massachusetts General Hospital, Boston, MA 02114 USA

Abstract

Over- and under-sedation are common in critically ill patients admitted to the Intensive Care Unit. Clinical assessments provide limited time resolution and are based on behavior rather than the brain itself. Existing brain monitors have been developed primarily for non-ICU settings. Here, we use a clinical dataset from 154 ICU patients in whom the Richmond Agitation-Sedation Score is assessed about every 2 hours. We develop a recurrent neural network (RNN) model to discriminate between deep vs. no sedation, trained end-to-end from raw EEG spectrograms without any feature extraction. We obtain an average area under the ROC of 0.8 on 10-fold cross validation across patients. Our RNN is able to provide reliable estimates of sedation levels consistently better compared to a feed-forward model with simple smoothing. Decomposing the prediction error in terms of sedatives reveals that patient-specific calibration for sedatives is expected to further improve sedation monitoring.

I. Introduction

Critically ill patients in intensive care unit (ICU) are broadly affected by the “ICU triad”: pain, agitation and delirium, due to many distressing interventions such as invasive mechanical ventilation [1]. Continuous sedation is one of the techniques to alleviate the negative effects of the ICU triad [2]. Unfortunately, both over- and under-sedation are common [3]. Research has shown that inappropriate levels of sedation, as well as analgesia, lead to longer ICU stay, more adverse events and eventually poorer clinical outcomes [4].

(corresponding author, phone: +1 650-862-1154; mwestover@mgh.harvard.edu).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Various clinical sedation assessment tools have been designed to monitor sedation levels in the ICU [5], including the Richmond Agitation-Sedation Scale (RASS) [6] and Sedation Agitation Scales (SAS) [7], among others. The inter-rater reliability and the correlation between different assessment tools are described in [5]. Despite having relatively high reliability and being easy to operate, there are several major disadvantages of these clinical assessment tools, including (1) the sedation level is only available at the point of assessment, thus having a low temporal resolution usually no better than once per hour in practice; (2) assessments are based on patient behaviors, which do not directly reflect the brain state; and (3) the patient needs to be periodically stimulated which can interfere with sleep and increase discomfort.

To overcome these disadvantages, several commercially available anesthesia depth and sedation monitors that track electroencephalogram (EEG) features has emerged as a physiologically-based, real-time alternative to clinical sedation assessments. These include the bispectral index (BIS) (Aspect Medical Systems, Norwood, MA, USA), Patient State Index (PSI) (Hospira, Lake Forest, IL, USA) and Nacrotrend (Monitor Technik, Bad Bramstedt, Germany), etc [8, 9]. All existing monitors extract various hand-crafted EEG features from both time and/or frequency domains, which vary from monitor to monitor [8]. Existing processed EEG monitors obtain a noisy sedation level for each EEG segment and then smooth them to get a stable index, where the smoothing time span is an extra parameter to tradeoff. Smoothing improves the stability of the index, but leads to long and in some cases (e.g. BIS) unpredictable time delays in response to changes in EEG; Short smoothing time span leads to faster response to EEG changes but is more noisy [10]. The time delay also creates difficulties for closed loop control and pharmacodynamics modeling [8, 11]. Finally, all prior monitors have been designed for non-ICU use, whereas the physiological variability due to the effects of severe illness and polypharmacy on brain activity makes brain monitoring more challenging than in the operating room environment.

Here we develop a sedation monitoring system based on Gated Recurrent Units (GRU, a type of recurrent neural network) [12]. GRU is a nonlinear autoregressive model, where the hidden state at t is updated based on the last hidden state at t-1 and the input at t, both of which are gated. The gates are learned through optimization. We train GRU end-to-end using the EEG spectrogram as the input without any other hand-crafted features. GRU can utilize both long and short term temporal contexts by remembering important patterns and forgetting irrelevant patterns in the past. The utilization of temporal context removes the need for ad hoc smoothing, so that it is expected to have both stable tracking of sedation and fast response to neurological changes in the brain. Unlike existing monitors which have been tested almost exclusively in the surgical setting, here we optimize for use in the ICU setting [13, 14]

The remainder of the paper is organized as follows. Section II provides details about patients, EEG processing and the GRU model. Section III presents a performance evaluation, and compares to a feed-forward model with smoothing. We also analyze the effect of sedative types on prediction error. Section IV describes the significance of this work as well as important limitations and future directions. Section V concludes the paper.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

This study was conducted under a protocol approved by the Institutional Review Board of Massachusetts General Hospital.

II. Method

A. RASS as Reference Sedation Level

The Richmond agitation-sedation scale (RASS) is one of the mostly used clinical sedation assessment tools in mechanically ventilated ICU patients. It has been validated to show robust inter-rater reliability [6]. It has 10 levels from −5 to +4. Values between −5 and −1 refer to decreasing sedation from unarousable (−5), deep sedation (−4), moderate sedation (−3), light sedation (−2) to drowsy (−1). Level 0 indicates an alert and calm state (0). Values from +1 to +4 represent increasing levels of agitation. RASS scores in the ICUs in our study were assessed about every 2 hours by nurses, and were additionally assessed 1–2 daily by research staffs.

B. Patient Selection

The dataset consists of 195 mechanically ventilated ICU patients without any pre-existing neurological deficits, enrolled between 2014 and 2016. In the present work we consider only assessments with RASS −5/−4 (deeply sedated) vs. −1/0 (not sedated), which reduces the number of eligible patients to 154 who have at least one of the two levels. There are 1510 RASS assessments in total, where 623 assessments have RASS −5/−4 and 887 have −1/0. Demographic characteristics of the patients are shown in TABLE I.

C. EEG and Preprocessing

EEG signals were recorded using Sedline brain function monitors (Masimo Corporation, Irvine, CA, USA) at a sampling rate of 250Hz. The signals consist of 4 frontal channels at Fp1, Fp2, F7 and F8, and a reference channel at Fpz (using 10–20 system naming conventions). We re-reference the EEG to bipolar montage: Fp1-F7, Fp2-F8, Fp1-Fp2 and F7-F8 and band-pass the signal between 0.5Hz and 16Hz using zero-phase FIR filtering. The range of pass band is relatively restrictive compared to other studies [8] to reduce the influence of multiple noise sources from various machines in the ICU setting. We only evaluate the model using the 10min before each RASS assessment. We estimate EEG spectrograms using the multitaper method with the following parameters: window length T = 4s with 2s overlap, number of tapers K = 7 and spectral resolution of 2 Hz.

For artifact removal, we remove segments with: amplitude larger than 500uV; standard deviation smaller than 0.2uV for more than 2s; or spectrum is spuriously staircase-like, defined by the maximum value of the convolution with a predefined stair-like shape is larger than a predefined threshold, which indicates nonphysiologic artifacts from ICU machines (e.g. pumps or cooling blankets). About 10% of the data is removed due to artifacts.

D. Classifier Training and Testing

In our approach there are no hand-crafted features extracted from the spectrogram. Instead, the raw spectrogram is fed directly to a 2-layer GRU with 32 hidden nodes at each layer and dropout rate 0.5. Each time step in each layer consists of four components

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$h_t = z_t h_{t-1} + (1 - z_t) n_t; \quad (1)$$

$$z_t = \text{sigmoid}(W_{iz} x_t + b_{iz} + W_{hz} h_{t-1} + b_{hz}); \quad (2)$$

$$n_t = \tanh(W_{in} x_t + b_{in} + r_t (W_{hn} h_{t-1} + b_{hn})); \quad (3)$$

$$r_t = \text{sigmoid}(W_{ir} x_t + b_{ir} + W_{hr} h_{t-1} + b_{hr}); \quad (4)$$

where x_t, h_t, r_t, n_t, z_t are the input, hidden state, reset gate, new gate and input gate at time step t respectively; $W_{i\cdot}$ and $b_{i\cdot}$ are the input-to-hidden weight and bias for different gates; $W_{h\cdot}$ and $b_{h\cdot}$ are the hidden-to-hidden weight and bias for different gates. The hidden state of GRU is fed to a logistic regression layer for binary classification (RASS −5/−4 vs. −1/0).

Data from different patients is randomly split into 10 folds with approximately equal size. Data from the same patient appears in either the training or testing set, never both at the same time. Classification performance is measured by area the under the receiver operator curve (AUC). Testing performance is obtained from each fold while trained on the other 9 folds. The reported result is the average AUC on the testing set from the 10 folds. Strict separation of training patients from testing patients is maintained throughout all experiments.

E. Correcting Label Noise and Burst Suppression

There are a few corrections to the model prediction in ICU setting. First, RASS assessments are sometimes recorded after a delay or in anticipation of a change of sedation level following changes in sedative infusion rate. We correct the RASS scores with inconsistent EEG spectrograms near a change in sedative infusion rate, which make up 3% of assessments. Second, a separate burst suppression detection [15] is carried out. In our case, about 3% of the RASS assessments are detected during EEG burst suppression. In this dataset we observe several patients having burst suppression-like EEG patterns while is still at RASS −1/0. Due to the noisy electrical environment in ICU, it is not clear if this is due to patient heterogeneity or bad electrode connection, thus they are excluded from the analysis.

III. Results

A. Performance on the Holdout Testing Patients

As shown in Figure 1A, the average AUC across 10 folds in the testing set is 0.73 (SD 0.05). After correcting label noise and burst suppression, the testing AUC is 0.80 (SD 0.06). The

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

different responses to sedatives (i.e. different EEG patterns for the same behvaioral state / level of sedation) is a primary source of the imperfect AUC value.

In Figure 2 we show examples of EEG spectrogram and the predicted sedation level in both RASS −5/−4 and RASS −1/0 cases. The initial part, about 200s (note the time unit is ×2s), is a mixing period where the GRU model flushes the initial zero hidden state. This should not be an issue in real-time monitoring, since the mixing period only occurs at the beginning of each monitoring session. The trace of predicted sedation level is robust to noisy perturbations in the spectrogram, as exemplified in Figure 2B. The trace of sedation level is smooth due to the intrinsic recurrent structure, without ad hoc smoothing.

B. Comparison with Feed-Forward Neural Network with Smoothing

To further illustrate the benefit of our recurrent model for sedation monitoring, we treat the spectra in the spectrogram as independent samples and train a feed-forward neural network with the same number of layers and the same number of hidden nodes with the recurrent counterpart, GRU. For a feed-forward network, due to its lack of temporal context, sedation level estimates are noisy. We can smooth these estimates by averaging the trace using the most recent T seconds, where T is the length of the smoothing window. In Figure 3, we compare the performance of our RNN against the feed-forward network at different smoothing window lengths. The performance of the RNN is not affected by the length of smoothing window, while the performance of feed-forward network is improved by longer smoothing windows. However, the longer smoothing window introduces a longer time delay to sedation monitoring in terms of neurologically meaningful changes, which is not desirable for ICU setting.

C. The Effect of Sedative Drugs

Many ICU patients receive continuous infusions or bolus doses of sedative drugs. A major reason for misclassifications in the current model appears to be patient-specific differences in EEG signatures of sedatives. To show this, in Figure 4, we compare the prediction error when there is no drug given, and when there is a certain drug or combination of a sedative and an opioid given to the patient. Here we only show combinations with more than 50 RASS assessments.

Assessments where no drug is being administered show a relatively low prediction error. Dexmedetomidine, as well as its combination with hydromorphine and fentanyl, gives the lowest prediction error, suggesting relatively consistent brain responses across patients. A combination of propofol and hydromorphine gives larger prediction errors compared to several other cases, suggesting large patient-specific variations for this combination. Propofol and its combination with other opioids have relatively large prediction errors. Ketamine is not included in Figure 3 due to an insufficient number of examples to allow reliable evaluation.

IV. Discussion

We have developed a sedation monitoring system based on a recurrent neural network, trained from the EEG spectrograms of 154 ICU patients. This model obtains an average

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

testing AUC at 0.8 without any additional feature extraction or smoothing. A feed-forward model followed by smoothing obtains lower AUC (Figure 3), where smoothing leads to stable estimates of sedation levels while creating a response delay. In contrast, the RNN model achieves both low-variance estimates of sedation level and short delay time.

The ability of the GRU RNN model to learn and appropriately forget temporal context leads to less strict requirements on artifact removal. For minor perturbations, such as the one exemplified in Figure 2B, the model learns to ignore these fluctuations. Therefore, the current model should be more robust compared to existing sedation monitors in the ICU setting. However, a standardized validation protocol, such as [16], will be required to comprehensively validate the model.

There are several limitations and future refinements needed for the current model. First, burst suppression is treated separately. Second, propofol and its combination with other opioids in many cases lead to relatively large prediction errors as shown in Figure 4. To alleviate these limitations, a patient-specific calibration should be used where one can determine a personalized drug tolerance for achieving certain RASS scores. Another important limitation is that we treat the problem as binary classification, i.e. RASS −5/−4 vs. −1/0. Since RASS score is ordinal, ordinal regression should be used to make use of all RASS scores in real scenario applications such as closed-loop sedation level control.

V. Conclusion

Reliable sedation monitoring in ICU patients can be achieved using a recurrent neural network trained end-to-end from EEG spectrograms. The sedation level predictions are stable (low variance) without ad hoc smoothing. Patient-specific calibration for specific responses to sedatives, such as propofol, is expected to further improve the performance of sedation monitoring.

References

[1]. Reade MC and Finfer S, “Sedation and delirium in the intensive care unit,” New England Journal of Medicine, vol. 370, no. 5, pp. 444–454, 2014. [PubMed: 24476433]

[2]. Hariharan U and Garg R, “Sedation and Analgesia in Critical Care,” J Anesth Crit Care Open Access, vol. 7, no. 3, p. 00262, 2017.

[3]. Jackson DL, Proudfoot CW, Cann KF, and Walsh TS, “The incidence of sub-optimal sedation in the ICU: a systematic review,” Critical Care, vol. 13, no. 6, p. R204, 2009. [PubMed: 20015357]

[4]. Hughes CG, McGrane S, and Pandharipande PP, “Sedation in the intensive care setting,” Clinical pharmacology: advances and applications, vol. 4, p. 53, 2012. [PubMed: 23204873]

[5]. Sessler CN, Grap MJ, and Ramsay MA, “Evaluating and monitoring analgesia and sedation in the intensive care unit,” Critical Care, vol. 12, no. 3, p. S2, 2008.

[6]. Sessler CN et al., “The Richmond Agitation–Sedation Scale: validity and reliability in adult intensive care unit patients,” American journal of respiratory and critical care medicine, vol. 166, no. 10, pp. 1338–1344, 2002. [PubMed: 12421743]

[7]. Vivien B, Di Maria S, Ouattara A, Langeron O, Coriat P, and Riou B, “Overestimation of Bispectral Index in sedated intensive care unit patients revealed by administration of muscle relaxant,” Anesthesiology: The Journal of the American Society of Anesthesiologists, vol. 99, no. 1, pp. 9–17, 2003.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[8]. Musizza B and Ribaric S, “Monitoring the depth of anaesthesia,” Sensors, vol. 10, no. 12, pp. 10896–10935, 2010. [PubMed: 22163504]

[9]. Sheahan C and Mathews D, “Monitoring and delivery of sedation,” British journal of anaesthesia, vol. 113, pp. ii37–ii47, 2014. [PubMed: 25498581]

[10]. Zanner R, Pilge S, Kochs E, Kreuzer M, and Schneider G, “Time delay of electroencephalogram index calculation: analysis of cerebral state, bispectral, and Narcotrend indices using perioperatively recorded electroencephalographic signals,” British journal of anaesthesia, vol. 103, no. 3, pp. 394–399, 2009. [PubMed: 19648154]

[11]. Abdulla SA and Wen P, “The effects of time-delay on feedback control of depth of anesthesia,” in Biomedical and Health Informatics (BHI), 2012 IEEE-EMBS International Conference on, 2012, pp. 956–959: IEEE.

[12]. Chung J, Gulcehre C, Cho K, and Bengio Y, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” arXiv preprint arXiv:1412.3555, 2014.

[13]. Nagaraj SB et al., “Automatic Classification of Sedation Levels in ICU Patients Using Heart Rate Variability,” Critical care medicine, vol. 44, no. 9, pp. e782–9, 2016. [PubMed: 27035240]

[14]. Nagaraj SB et al., “Patient-Specific Classification of ICU Sedation Levels From Heart Rate Variability,” Critical care medicine, vol. 45, no. 7, pp. e683–e690, 2017. [PubMed: 28441231]

[15]. Chemali J, Ching S, Purdon PL, Solt K, and Brown EN, “Burst suppression probability algorithms: state-space methods for tracking EEG burst suppression,” Journal of neural engineering, vol. 10, no. 5, p. 056017, 2013. [PubMed: 24018288]

[16]. Heyse B, Van Ooteghem B, Wyler B, Struys M, Herregods L, and Vereecke H, “Comparison of contemporary EEG derived depth of anesthesia monitors with a 5 step validation process,” Acta Anaesthesiologica Belgica, vol. 60, no. 1, p. 19, 2009. [PubMed: 19459551]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

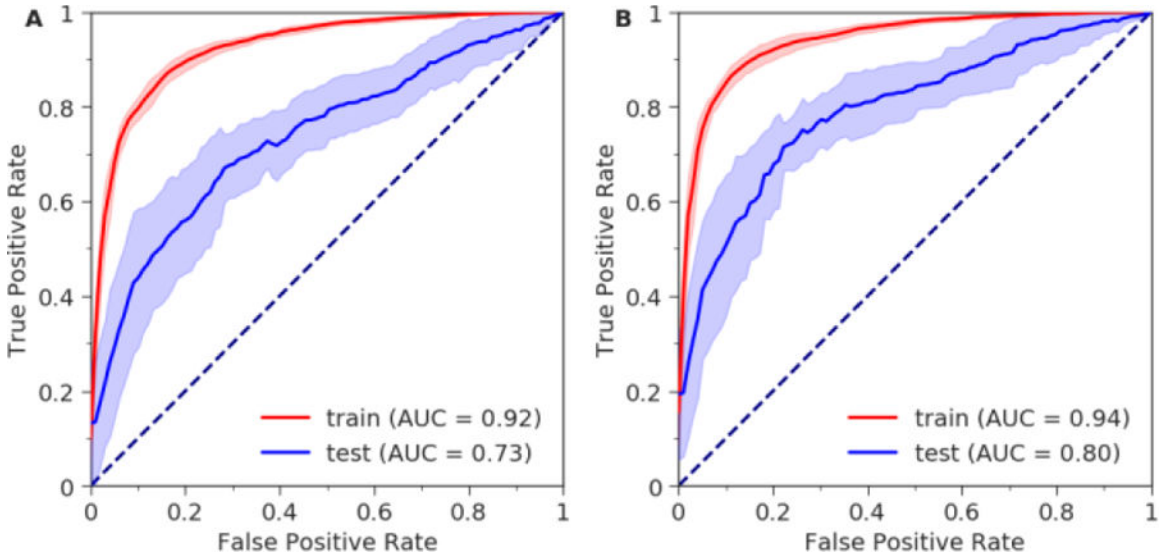

Figure 1.

ROC curves for the training and testing patients averaged from 10 folds. (A) Without any correction. (B) After correcting label noise and burst suppression.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

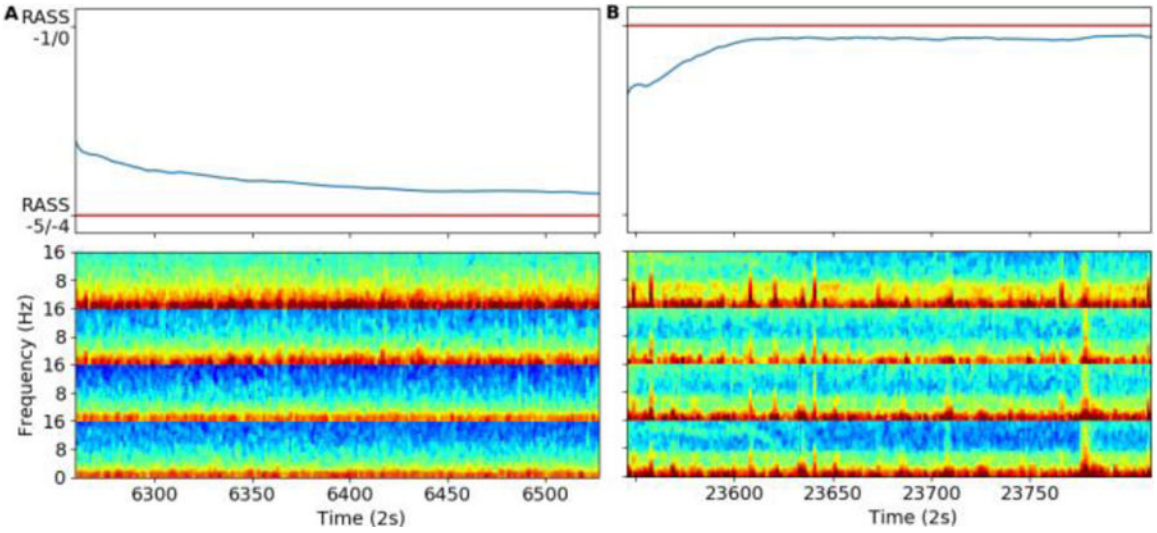

Figure 2.

(A) Top: The red line indicates the sedation level given by RASS assessment at −5/−4; The blue line indicates the predicted sedation level, expressed as the probability of being RASS −1/0. Bottom: The EEG spectrogram from four bipolar frontal channels. The unit of x axis is ×2s. (B) Similar to (A), but showing an example with RASS at −1/0.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

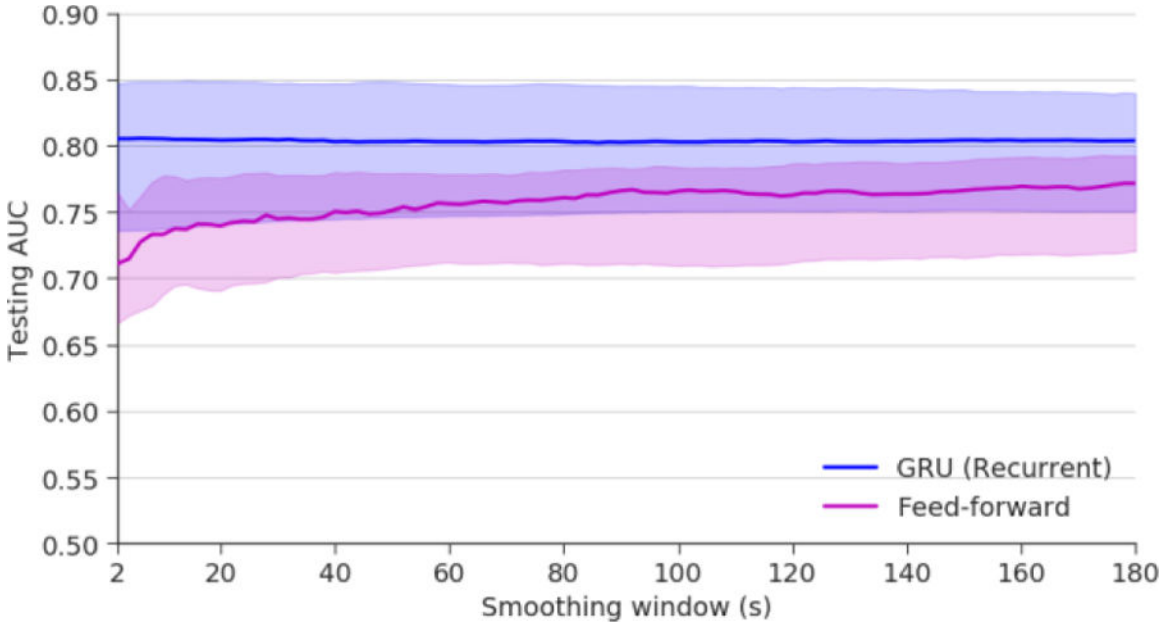

Figure 3.

The average testing AUC from 10 folds for GRU (blue) and Feed-forward neural network (magenta) when using smoothing windows with different lengths. The shaded area indicates interquantile range.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

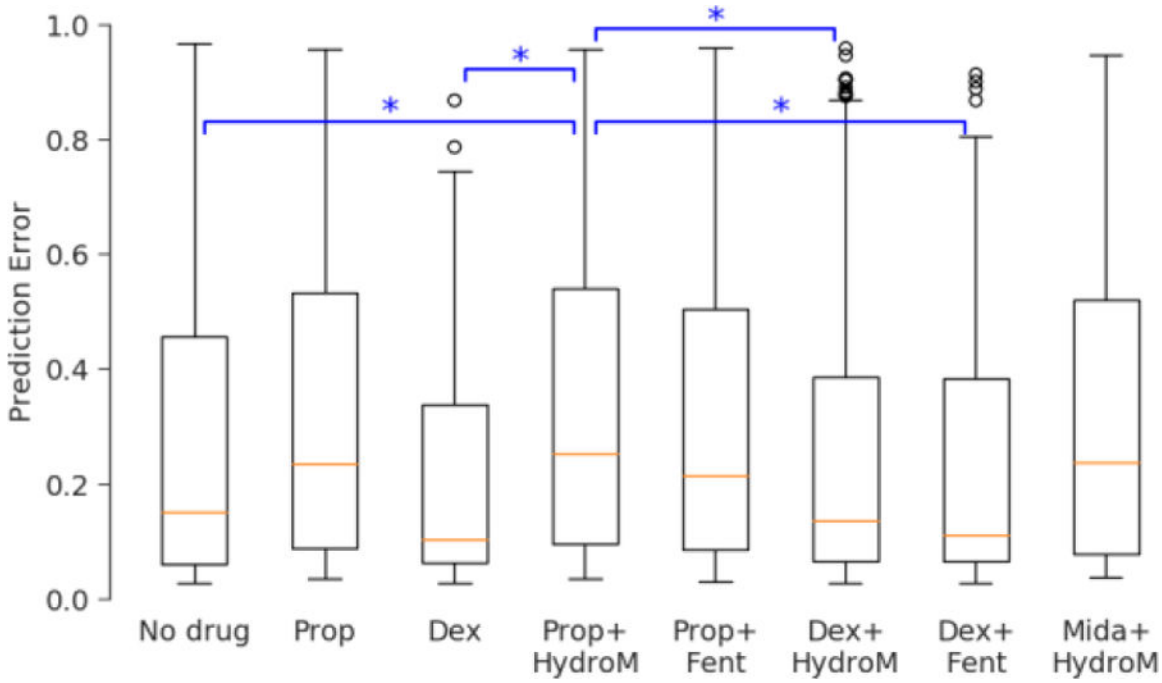

Figure 4.

The prediction error (absolute difference between target and prediction probability) with drug and without drug in testing patients. Prop: propofol. Dex: dexmedetomidine. HydroM: hydromorphine. Fent: fentanyl. Mida: midazolam. Stars indicate significant difference ($p < 0.05$) using Kruskal-Wallis test followed by Dunn's post-hoc test.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE I.

PATIENT CHARACTERISTICS

Characteristic	Median (IQR[a]) or Number (Percentage)
Age (year)	60 (51 – 75)
Sex	Female 49 (32%); Male 105 (68%)
Race	White 135 (88%); Black 9 (6%); Asian 2 (1%); Unknown 8 (5%)
BMI[b] (kg/m^2)	29 (23 – 35)
Days in ICU	12 (7 – 19)
APACHE II[c] at admission	22 (15 – 28)
Charlson comorbidity index	3 (2 – 4)
Diagnosis at ICU admission	
Sepsis	22 (14%)
Acute respiratory failure	92 (60%)
Surgery	31 (20%)
Cardiac shock, myocardial ischemia, or arrhythmia	10 (6%)
Liver or renal failure	35 (23%)

[a] interquartile range

[b] body mass index

[c] Acute Physiology And Chronic Health Evaluation II