Patient-Specific Classification of ICU Sedation Levels from Heart Rate Variability

Sunil B. Nagaraj, PhD[*,1], **Siddharth Biswal, MS**[1], **Emily J. Boyle, BS**[1], **David W. Zhou, MS**[1], **Lauren M. McClain, BA**[1], **Ednan K. Bajwa, MD, MPH**[2], **Sadeq A. Quraishi, MD, MHA, MMSc**[3], **Akeju Oluwaseun, MD, MMSc**[3], **Riccardo Barbieri**[3,4], **Patrick L. Purdon, PhD**[3], and **M. Brandon Westover, MD, PhD**[1]

[1]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, U.S.A

[2]Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, U.S.A

[3]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, U.S.A

[4]Department of Electronics, Informatics and Bioengineering, Politecnico di Milano, Italy

Abstract

Rationale—Intensive care unit (ICU) patients are at risk for both over- and under sedation, presently. An automated, physiologically-based method for monitoring sedation levels might help reduce this risk by allowing more precise titration of sedation in critically ill individuals.

Objective—To develop a peronsalizable algorithm to discriminate between sedation levels in ICU patients based on heart rate variability (HRV).

Methods—We gathered 21,912 hours of routine electrocardiogram (ECG) recordings from a heterogenous group of 70 adult ICU patients. All patients included in the study were mechanically ventilated and were receiving sedatives. As "ground truth" for developing our method we used Richmond agitation-sedation scale (RASS) scores grouped into four levels denoted "comatose" (−5), "deeply sedated" (−4 to −3), "lightly sedated" (−2 to 0), and "agitated" (+1 to +4). We trained a support vector machine learning algorithm to calculate the probability of each sedation level from HRV measures derived from the ECG. To estimate algorithm performance we calculated leave-one-subject-out cross-validated accuracy.

Measurements and Main Results—The patient-independent version of the proposed system discriminated between the 4 sedation levels with an overall accuracy of 59%. Upon personalizing the system supplementing the training data with patient-specific calibration data, consisting of an individual's labeled HRV epochs from the preceding 24 hours, accuracy improved to 67%. The personalized system discriminated between light- and deep-sedation states with an average accuracy of 75%.

Conclusions—Our results suggest that HRV information can be used to monitor sedation levels in medical and surgical ICU patients with varied pathophysiologies. Future work incorporating disease pathology, pharmacological information and other physiological variables may further improve the accuracy of this model.

INTRODUCTION

Managing sedation safely in critically ill mechanically ventilated patients requires accurate monitoring of the level of consciousness. ICU patients are often sedated to relieve stress, prevent injuries and to facilitate ventilation and analgesia (1). Over- and under sedation increase the risk of adverse patient outcomes, including prolonged mechanical ventilation, extended ICU stays, and delirium (2). Current practices rely on behavioral scoring systems based on the patient's response to verbal or physical stimulation. However, these systems rely on experience and clinical observation and are thus subjective, and are inaccurate at deep levels of sedation (3). Physiologically-based monitoring might be able to overcome these limitations, enabling more judicious titration of sedatives and reduced sedative-related adverse events.

Electroencephalogram (EEG)-based methods have been developed to assess the level of consciousness during general anesthesia (4–10). Applications of these EEG based methods have received little attention in the ICU, in part because EEG is not routinely available for general ICU patients. The electrocardiogram (ECG), however, is universally available in the ICU setting. ECG-based heart rate variability (HRV) has been shown to relate systematically to level of consciousness in some groups of patients (11). Whether sedation can be reliably inferred from HRV in individual ICU patients remains unclear.

In preliminary work we found that HRV measures show potential value as features in an automated monitoring system to predict the level of consciousness of ICU patients (12). Here we extend our previous work by developing a sedation level assessment system based on HRV measures that is *patient-specific*. Starting from 31 different time, frequency, nonlinear and complexity measures of HRV, we used an optimization procedure to determine an optimized set of features and used them to train a machine learning algorithm to predict patients' level of consciousness. We rigorously tested and validated the patient-specific system on a large set of ICU ECG recordings. The proposed system provides a foundation for developing HRV-based sedation assessment monitors.

METHODS AND MATERIALS

Dataset

For this study we used 21,912 hours of ECG data from 70 patients (43 males; 27 females), recorded using GE bedside patient monitors and archived using BedMaster software (Excel Medical Electronics, Jupiter FL, USA) at a sampling frequency of 240 Hz. Recordings took place in several ICUs at Massachusetts General Hospital (MGH), Boston, USA, under an IRB-approved protocol. Table 1 provides patient demographic characteristics.

Sedation assessment

To categorize sedation levels, we used the Richmond agitation-sedation scale (RASS), which has been shown to exhibit good inter-rater reliability in medical and surgical ICU patients (13). RASS scores range from −5 (deeply sedated/unresponsive) to +4 (agitated/violent) with 0 representing a calm awake state. ICU nurses performed RASS assessments as part of routine care at approximately 2-hour intervals. A member of the study staff later collected RASS scores from the medical record. Trained clinical research staff also performed RASS assessments once daily as part of the research protocol. In total our data included 4,258 RASS assessments from 70 patients. In addition, we recorded time and dosages for sedative and analgesic medications (both intravenous infusion and bolus).

We grouped RASS scores into four categories (see Supplemental Material - Table 1). For convenience, from here on we refer to RASS groupings as A = Agitated = RASS >0; B = "Lightly sedated", RASS 0, −1, −2; C = "Deeply sedated", RASS −3, −4; and D = "Comatose", RASS −5. We chose to group RASS scores to balance two considerations. First, grouping increases the number of samples in each category for classifier training. Second, assignment of RASS scores is imperfect, giving rise to some degree of "label noise". That is, scores assigned to patients in clinically similar states may vary slightly: A patient who is lightly sedated with minimal stimulation may be assigned to either RASS −1 or −2, but will not be scored as RASS −5. We designed our RASS grouping such that categories were distinct enough to enable reliable classifier performance while still defining distinct clinical states.

Overview of the automatic sedation level classification system

Figure 1a provides an overview of the processing steps for the proposed automatic HRV-based sedation-level inference system (AHSISt). In the following sections we provide details for each step.

Artifact reduction and feature extraction

Obtaining informative HRV measures depends critically on first pre-processing the RRI signal to reduce artifacts. To obtain candidate R-wave peaks, we applied the Pan-Tompkins algorithm to the ECG signal, yielding an initial series of RRI (14). As an initial step we remove abnormal ectopic beats using a threshold-based method (15). However, numerous spurious RR peaks remain, thus we further clean the RRI signal using a threshold rejection method based on the interquartile range, adapted from Kauffman et al. (16). In this method, after obtaining differences between successive RRI ($RRI_{diff} = RRI_{n+1} − RRI_n$ where n is the

sample number), we calculate the absolute value of RRI_{diff}, and discard the largest 2% of the values. Gaps created by the 2% discarded data are filled in using linear interpolation. After pre-processing, we regard the cleaned and interpolated RRI as the cleaned heart rate time series. All subsequent HRV features are based on the cleaned RRI signal. An example illustrating the process of artifact reduction is shown in Supplemental digital content - Figure 1. Examples of the cleaned heart rate time series and accompanying RASS scores from four patients are shown in Supplemental digital content - Figure 2.

Due to limited prior knowledge regarding optimal features for predicting sedation levels, we extracted a large set of features from the cleaned RRI signal that have been used in previous studies in adults (17–20). Table 2 describes the 31 features that we extracted. We normalized all features using the box-cox transformation (21) to have uniform mean and standard deviation before feeding them to the SVM training and classification algorithms.

Nested cross validation for model training and testing

We used a nested cross validation (CV) strategy, consisting of distinct inner and outer CV loops to train the model (inner CV loop), and to obtain approximately unbiased estimates of model prediction performance (outer CV loop). Technical details regarding model training and evaluation are discussed briefly here; for further details see the Supplemental Material.

In the outer CV loop data is separated into separate training and testing sets. Model training and optimization is performed entirely on the training data while leaving the testing set untouched. The resulting model is then evaluated on the left-out testing data. For the outer CV loop we use a leave-one-subject-out CV strategy (LOSOCV). That is, we perform 70 rounds of CV, where in each round data from 69 of the 70 subjects is used for training, and the remaining data from one subject is used for testing. In a second round of experiments, to personalize the algorithm for an individual patient, we allowed a portion of the left-out patient's *past* data to be included in the training set (see below).

The inner CV loop is used for model parameter optimization. This is performed on the training data from the 69 patients included in each outer CV loop. In each inner CV loop we determine an optimal window duration (dur_) for extracting HRV features from the ECG, an optimal set of features to include in the final model ("feature selection"), and optimal values of parameters (σ, C) required by the support vector machine (SVM) algorithm.

Training of the patient-independent AHSISt system—Initially, each training set consisted of all data except that from a single test subject. This resulted in a split of 69:1 between training and testing sets. For the patient-*independent* version of AHSISt, we strictly excluded all data from the testing set from the procedures used for classifier training and parameter tuning.

Training of the patient-specific AHSISt system—For creating a patient-*specific* version of AHSISt, we added the first 24 hours of data (day *1*) from the testing set (i.e. 12 epochs, one RASS assessment every 2 hours) to the training data set; we used the remaining data for testing (days 2 to day *M*, where *M* is the total number of days in the ICU). In this way we allowed past observations regarding the relationship between an individual's HRV

features and their RASS level to inform future HRV-based predictions of RASS scores for the same individual. This technique of using prior data from an ECG recording to train a patient-specific prediction algorithm is commonly used in arrhythmia classification systems (22, 23). This procedure allows the classifier to become calibrated, or individualized, to provide patient-specific prediction of ICU sedation levels (see figure 1b).

Handling of class imbalance in the training data—The training dataset had an unequal number of epochs from different RASS groups (see Supplemental digital content-Figure 3). This is potentially problematic, as SVM classifier training is sensitive to class imbalance (24). To overcome this challenge, we used a sample-size equalization strategy in which we selected random epochs from each group in the training set corresponding to the size of the class with the smallest number of available samples, creating a series of balanced training sets. For the patient-specific AHSISt we ensured that all data from the patient's first recording day was included in the balanced training set.

RESULTS

An example of the proposed system output shown as continuous probabilistic values for a single patient using the patient-independent and patient-specific versions of AHSISt is shown in Supplemental digital content - Figure 4. We see that by calibrating AHSISt with the initial 24hrs of data, the ability to correctly classify sedation levels (Supplemental digital content - Figure 4C) is significantly improved compared to the patient-independent system (Supplemental digital content - Figure 4A).

The training procedure used in 10-fold cross validation explored 70 different values of window size dur_{-}. A value of 15 minutes provided the maximum cross-validated accuracy during the training process. The proposed system in this work is patient specific, in that it creates a new classifier for each individual patient. SFS identified different optimal subsets of features for each patient in the LOSOCV procedure. However, mean heart rate (MHR), root mean square of SDNN (RMSDD), high frequency spectral power (PHF), poincaré plot measures (SD1, SD2) and kolmogorov complexity features were consistently selected by the automated feature selection process across all patients.

The overall estimated cross validated accuracy of the patient-independent and patient-specific AHSISt were 59% and 67%, respectively, substantially better than chance level accuracy (38%). Figure 3 shows the distribution of classification decisions made by AHSISt as a function of the true underlying RASS group. AHSISt identifies RASS group C ("deep sedation") most accurately (74%). Accuracy for RASS group D ("comatose") was lowest. The majority of misclassifications that occur are due to assignments to neighboring sedation levels.

Figure 3b shows the accuracy obtained using patient-specific AHSISt for individual patients. Except for 2 patients, the accuracy was better than chance for all other patients. For the two outliers, most misclassifications occurred between neighboring sedation levels e.g. between group C (deeply sedated) and D (comatose); or between groups A (awake and agitated) and B (lightly sedated).

To gain insight into the intrinsic limits associated with inferring sedation levels from HRV in individual patients, we trained a binary classifier to discriminate between two RASS groups that are well separated clinically: light-sedation [RASS 0, −1] versus deep-sedation [RASS −4, −5]. The overall accuracy obtained was 75% and the area under the receiver-operator characteristic (ROC) curve (AUC) was 0.73. At the operating point on the ROC curve that maximizes accuracy, the proposed system correctly classified 66% of deep-sedation (sensitivity = 66%) and 86% of light-sedation epochs (specificity = 86%). This analysis provides a measure of the distinguishability of "deep-sedation" and "light-sedation" states from the standpoint of HRV.

From the predicted SVM probability output, we observe that the binary classifier is able to provide, as byproduct, a continuous score, that tends to be higher for deeper sedation levels and lower for lighter levels. As a preliminary test of whether this score might serve as a proxy for sedation level, we used the binary classifier, which is trained only on epochs of light and deep sedation, to assign a score to all epochs corresponding to 4-class RASS scores. We then calculated the Spearman's rank-correlation (ρ) between our binary classifier score and the RASS scores and obtained an overall ρ=0.5. These results are promising and suggest that with further improvement measures of HRV may ultimately be able to provide a useful continuous measure of sedation depth.

DISCUSSION

In a previous investigation we found that HRV features showed promise for distinguishing between different levels of sedation in mechanically ventilated ICU patients (12). Our present results extend these findings by demonstrating that HRV-based classification of sedation levels can be accomplished with an accuracy substantially above chance at the level of individual patients. To the best of our knowledge, this is the first attempt to automatically classify sedation levels using HRV features in a large heterogeneous set of mechanically ventilated ICU patients.

The proposed system using HRV has advantages over conventional purely behaviorally assessment-based methods for assessing depth of sedation. First, HRV-based sedation monitoring is objective, free from inter-observer variability. Second, it is based on ECG measurements that are readily available and a standard of care within the ICU. Third, it can be performed continuously. Finally, interpreting the output of the proposed system in a probabilistic way makes it flexible and suitable for following HRV trends.

However, the current system is not yet ready for clinical deployment; several limitations need to be addressed in future work. First, we did not take into account infusion rates and types of sedative drugs administered, which can influence HRV (25, 26) independent of changes in sedative state. We were unable to incorporate pharmakokinetic and pharmacodynamics information (PKPD) in the present study because our data is observational, based on patients managed according to usual care, with all of the variation in care that this implies; and our data did not provide sufficient power to estimate PKPD models for each drug or for drug-drug interactions and drug-disease interactions and their relationship to HRV and RASS levels. Estimating prediction models that incorporate such

information would best be performed under controlled conditions to limit patient heterogeneity (e.g. using a distinct sedation strategy in a homogenous patient population), or would need to include a far larger number of subjects. Overall, the heterogeneity of our patient cohort is a strength and a weakness: A strength because it suggests that our results apply broadly and robustly across a wide range of patients; a weakness due to the inability to further account for PKPD and disease effects that translates into imperfect prediction performance. Future refinements of our approach, enabled by larger datasets and/or controlled trials, will allow including PKPD information and personalized information about the patient's illness and are expected to further improve performance.

Second, we did not consider associations between deeper levels of sedation and increasing levels of organ failure. Several studies have shown the influence of multiple organ dysfunction on the autonomic nervous system (27–32). Therefore, additional modeling to account for the severity of organ dysfunction or medical illness might improve classifier performance. In addition, other physiological signals such as blood pressure, respiration rate, temperature, and SpO_2 may carry complementary information, and inclusion of features derived from these signals could improve classifier performance (33).

Third, we grouped the RASS scores, which can take on 10 different values, into 4 groups. A preferable future approach might be to treat level of sedation as a continuous rather than categorical variable.

Fourth, because RASS scores rely on clinical behavioral observations, they are subject to some degree of human error. We tried to address this by grouping scores into four clinically distinct levels, but some "annotation noise" probably remained and may have limited performance. In addition, our data does not allow us to assess the accuracy of a continuous measure of sedation, because RASS scores are discrete-valued, and because the number of examples of each RASS score is imbalanced (necessitating our choice to group scores into 4 levels). In other words, rigorously validating a continuous measure of sedation will require obtaining continuous measures of the response variables, e.g. measures of threshold stimulus intensities required to elicit a response from the patient.

Fifth, our method of evaluating classifier performance may be overly harsh: in our method, we penalize equally classifying RASS 0 as RASS −1 or as RASS −5. Future work should explore a "soft" classification strategy, assigning probabilities to possible sedation levels rather than making hard classification decisions. This would allow performance scoring using a metric that penalizes misclassifications to nearby states less than far away states. In addition, RASS scoring is intrinsically limited, and this may set a ceiling on the performance we are able to obtain with our methodology. We are exploring two approaches to overcome this limitation in future work: (i) obtaining higher quality assessments using a small, specially trained set of raters with extensive measures to ensure high inter-rater agreement and measurement validity; (ii) using our method in a prospective fashion to predict behavioral responses to a battery of stimuli (e.g. increasing grades of verbal, tactile, and noxious stimuli), in effect bypassing the RASS scoring system.

Author Manuscript

Finally, a sixth limitation is that AHSISt uses observations of RASS and HRV from the first 24 hours of monitoring as input to the training procedure. In this sense AHSISt might be considered "semi-automatic", because it requires some human input for optimal performance. In practice, on the first day AHSISt could be started in the patient-independent mode (which had an accuracy of 57%). For subsequent days, it would be possible to further improve performance by repeatedly retraining with new incoming RASS+HRV data as the patient continues to be monitored. Since the nurse records this information every two hours in current practice, this would not pose an additional burden, and would provide the added value of continuous monitoring between clinical assessments.

As described above, there may be room for making the proposed system more or even fully automatic by developing a multimodal AHSISt algorithm which includes additional information from other sources, including EEG.

CONCLUSION

This work presents a novel framework to infer levels of sedation in critically-ill mechanically ventilated ICU patients using HRV features as input to a machine learning algorithm. Using a set of signal descriptors extracted from the ECG, the proposed system obtains promising prediction results across a heterogeneous cohort of 70 ICU patients. The proposed sedation level classification system represents a positive step towards developing a practical physiologically-based sedation monitoring system for ICU patients.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Pomeranz B, Macaulay RJ, Caudill MA, et al. Assessment of autonomic function in humans by heart rate spectral analysis. Am J Physiol. 1985; 248:H151–153. [PubMed: 3970172]

2. Gehlbach BK, Kress JP. Sedation in the intensive care unit. Curr Opin Crit Care. 2002; 8:290–298. [PubMed: 12386488]

3. De Jonghe B, Cook D, Appere-De-Vecchi C, et al. Using and understanding sedation scoring systems: a systematic review. Intensive Care Med. 2000; 26:275–285. [PubMed: 10823383]

4. Jospin M, Caminal P, Jensen EW, et al. Detrended fluctuation analysis of EEG as a measure of depth of anesthesia. Biomed Eng IEEE Trans On. 2007; 54:840–846.

5. Li X, Li D, Liang Z, et al. Analysis of depth of anesthesia with Hilbert–Huang spectral entropy. Clin Neurophysiol. 2008; 119:2465–2475. [PubMed: 18812265]

6. Bouillon TW, Bruhn J, Radulescu L, et al. Pharmacodynamic interaction between propofol and remifentanil regarding hypnosis, tolerance of laryngoscopy, bispectral index, and electroencephalographic approximate entropy. Anesthesiology. 2004; 100:1353–1372. [PubMed: 15166553]

7. Kreuer S, Biedler A, Larsen R, et al. The Narcotrend™–a new EEG monitor designed to measure the depth of anaesthesia A comparison with bispectral index monitoring during propofol-remifentanil-anaesthesia. Anaesthesist. 2001; 50:921–925. [PubMed: 11824075]

8. Viertiö-Oja H, Maja V, Särkelä M, et al. Description of the Entropy algorithm as applied in the Datex-Ohmeda S/5 Entropy Module. Acta Anaesthesiol Scand. 2004; 48:154–161. [PubMed: 14995936]

9. Sigl JC, Chamoun NG. An introduction to bispectral analysis for the electroencephalogram. J Clin Monit. 1994; 10:392–404. [PubMed: 7836975]

10. Bruhn J, Myles PS, Sneyd R, et al. Depth of anaesthesia monitoring: what's available, what's validated and what's next? Br J Anaesth. 2006; 97:85–94. [PubMed: 16751211]

11. Janz BA, Clifford GD, Mietus JE, et al. Multivariable analysis of sedation, activity, and agitation in critically ill patients using the Riker scale ECG, blood pressure, and respiratory rate. Computers in Cardiology. 2005; 2005:735–738.

12. Nagaraj SB, McClain LM, Zhou DW, et al. Automatic Classification of Sedation Levels in ICU Patients Using Heart Rate Variability. Crit Care Med. 2016

13. Sessler CN, Gosnell MS, Grap MJ, et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. Am J Respir Crit Care Med. 2002; 166:1338–1344. [PubMed: 12421743]

14. Pan J, Tompkins WJ. A Real-Time QRS Detection Algorithm. IEEE Trans Biomed Eng. 1985; BME-32:230–236.

15. Clifford, GD., McSharry, PE., Tarassenko, L. Computers in Cardiology. Vol. 2002. IEEE; 2002. Characterizing artefact in the normal human 24-hour RR time series to aid identification and artificial replication of circadian variations in human beat to beat heart rate using a simple threshold [Internet]; p. 129-132.

16. Kaufmann T, Sütterlin S, Schulz SM, et al. ARTiiFACT: a tool for heart rate artifact processing and heart rate variability analysis. Behav Res Methods. 2011; 43:1161–1170. [PubMed: 21573720]

17. Moser M, Lehofer M, Sedminek A, et al. Heart rate variability as a prognostic tool in cardiology. A contribution to the problem from a theoretical point of view. Circulation. 1994; 90:1078–1082. [PubMed: 8044922]

18. Stein PK, Kleiger RE, Domitrovich PP, et al. Clinical and demographic determinants of heart rate variability in patients post myocardial infarction: insights from the cardiac arrhythmia suppression trial (CAST). Clin Cardiol. 2000; 23:187–194. [PubMed: 10761807]

19. Toichi M, Sugiura T, Murai T, et al. A new method of assessing cardiac autonomic function and its comparison with spectral analysis and coefficient of variation of R–R interval. J Auton Nerv Syst. 1997; 62:79–84. [PubMed: 9021653]

20. de Chazal P, Heneghan C, Sheridan E, et al. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. IEEE Trans Biomed Eng. 2003; 50:686–696. [PubMed: 12814235]

21. Box GE, Cox DR. An analysis of transformations. J R Stat Soc Ser B Methodol. 1964:211–252.

22. Recommended Practice for Testing and Reporting Performance Results of Ventricular Arrhythmia Detection Algorithms (proposed). Association for the Advancement of Medical Instrumentation; 1986.

23. Ince* T, Kiranyaz S, Gabbouj M. A Generic and Robust System for Automated Patient-Specific Classification of ECG Signals. IEEE Trans Biomed Eng. 2009; 56:1415–1426. [PubMed: 19203885]

24. Akbani, R., Kwek, S., Japkowicz, N. Machine Learning: ECML 2004. Springer; 2004. Applying support vector machines to imbalanced datasets [Internet]; p. 39-50.[cited 2015 Dec 9] Available from: http://link.springer.com/chapter/10.1007/978-3-540-30115-8_7

25. Nakatsuka I, Ochiai R, Takeda J. Changes in heart rate variability in sevoflurane and nitrous oxide anesthesia: effects of respiration and depth of anesthesia. J Clin Anesth. 2002; 14:196–200. [PubMed: 12031752]

26. Kanaya N, Hirata N, Kurosawa S, et al. Differential effects of propofol and sevoflurane on heart rate variability. Anesthesiology. 2003; 98:34–40. [PubMed: 12502976]

27. Schmidt HB, Werdan K, Müller-Werdan U. Autonomic dysfunction in the ICU patient. Curr Opin Crit Care. 2001; 7:314–322. [PubMed: 11805528]

28. Schmidt H, Müller-Werdan U, Hoffmann T, et al. Autonomic dysfunction predicts mortality in patients with multiple organ dysfunction syndrome of different age groups*. Crit Care Med. 2005; 33:1994–2002. [PubMed: 16148471]

29. Baguley IJ, Heriseanu RE, Felmingham KL, et al. Dysautonomia and heart rate variability following severe traumatic brain injury. Brain Inj. 2006; 20:437–444. [PubMed: 16716989]

30. Korach M, Sharshar T, Jarrin I, et al. Cardiac variability in critically ill adults: influence of sepsis. Crit Care Med. 2001; 29:1380–1385. [PubMed: 11445691]

31. Buchman TG, Stein PK, Goldstein B. Heart rate variability in critical illness and critical care. Curr Opin Crit Care. 2002; 8:311–315. [PubMed: 12386491]

32. Bradley BD, Green G, Ramsay T, et al. Impact of sedation and organ failure on continuous heart and respiratory rate variability monitoring in critically ill patients: A pilot study*. Crit Care Med. 2013; 41:433–444. [PubMed: 23263615]

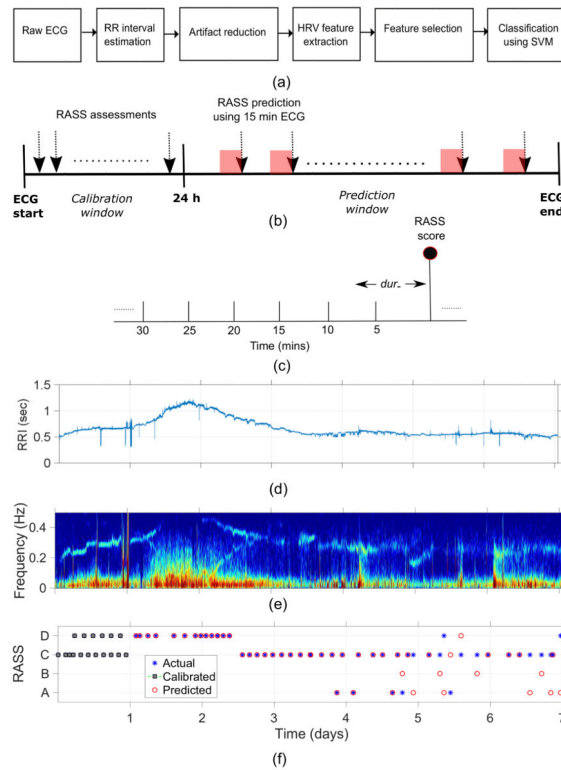33. Guignard B. Monitoring analgesia. Best Pract Res Clin Anaesthesiol. 2006; 20:161–180. [PubMed: 16634423]

Figure 1.
Illustration of (a) the proposed automatic sedation classification system, (b) the calibration window (initial 24 hours of ECG) and prediction window using 15 minutes of ECG (15 minutes was found as optimal window length in the classification) for patient-specific system, (c) identifying optimal RRI epoch length for sedation level assessment (d) a sample RRI, (e) its corresponding spectrogram, and (f) sample probability output from the patient specific classification where day *1* recording is used for calibrating the system and the predicted on the remaining days.
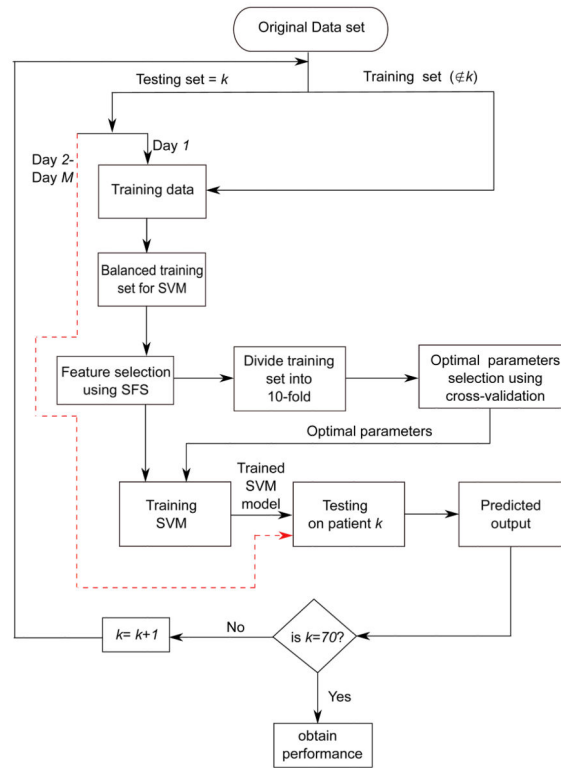
Figure 2.
Illustration of the proposed patient-specific AHSISt with LOSO performance assessment and 10-fold optimal parameter selection routines. Note that day *1* recordings are used during training process in case of patient-specific system. Data from the test patient was not used in case of patient-independent system.
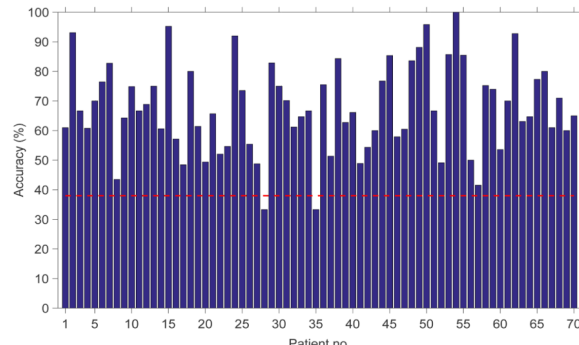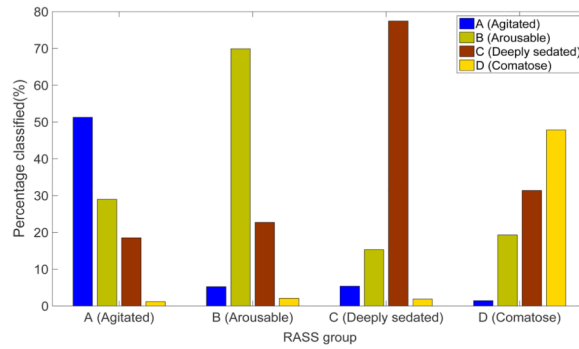
Figure 3.
The distribution of the epochs classified by the proposed automatic sedation system for different sedation levels, and (b) accuracy of the proposed patient-specific AHSISt for individual patients.

Table 1

Demographic and clinical characteristics of the patients used in this study.

Characteristic	Value
Age	
Median	58
Interquantile range	51–68
Weight (kg)	
Median	83
Interquantile range	71–102
Male sex – no. (%)	43 (61)
Days in ICU	
Median	13
Interquantile range	8–22
Charlson comorbidity index	
Median	3
Interquantile range	2–5
APACHE II score at enrollment	
Median	25
Interquantile range	18–31
Delirium	
No. of patients – (%)	56 (80)
No. of positive CAMICU assessments [*]	
Median	3
Interquantile range	1–8
Use of sedative or analgesic agent in ICU - (%) [+]	
Propofol	98
Benzodiazepines	84
Dexmedetomidine	34
Opiods	97

[*] number of positive CAMICU assessments (CAMICU=1) during the ICU stay.

[+] percentage of patients in this study receiving the drug.

Table 2

List of HRV features used in this work for the classification of sedation levels.

Domain	Features
Time	• Maximum, minimum, mean, median of NN.
	• Standard deviation of NN (SDNN), percentage of consecutive NNs that vary by 5, 10, 15, 20, 25, 30 ms (pNNx).
	• Root mean square of SDNN (RMSDD), mean heart rate (MHR), standard deviation of heart rate (SDHR), HRV triangular index, coefficient of variation.
	• Max change in NN (MCNN), Mean of the absolute value of first derivative of NN (MAFDNN), Nonlinear energy, Line length.
Frequency	• PVLF-Power in very low frequency spectrum (0.003–0.04 Hz).
	• PLF-Power in low frequency spectrum (0.04–0.15Hz).
	• PHF-Power in high frequency spectrum (0.15–0.4 Hz).
	• PLF/PHF.
	• PLF/PTOT $\times$ 100, PTOT is the total power spectrum.
	• PHF/PTOT $\times$ 100.
Nonlinear	• Poincaré plot measures (SD1, SD2).
Complexity	• Shannon entropy.
	• Spectral entropy.
	• Kolmogorov complexity.