SLEEPJ, 2021, 1–11

doi: 10.1093/sleep/zsaa167
Advance Access Publication Date: 29 August 2020
Original Article

ORIGINAL ARTICLE

Dexmedetomidine-induced deep sedation mimics non-rapid eye movement stage 3 sleep: large-scale validation using machine learning

Sowmya M. Ramaswamy[1,*], Maud A. S. Weerink[1], Michel M. R. F. Struys[1,2] and Sunil B. Nagaraj[3]

[1]University of Groningen, University Medical Center Groningen, Department of Anesthesiology, Groningen, The Netherlands, [2]Department of Basic and Applied Medical Sciences, Ghent University, Ghent, Belgium and [3]University of Groningen, University Medical Center Groningen, Department of Clinical Pharmacy & Pharmacology, Groningen, The Netherlands

*Corresponding author. Sowmya Muchukunte Ramaswamy, University of Groningen, University Medical Center Groningen, Department of Anesthesiology, Groningen, The Netherlands. Email: s.muchukunte.ramaswamy@umcg.nl.

Abstract

Study Objectives: Dexmedetomidine-induced electroencephalogram (EEG) patterns during deep sedation are comparable with natural sleep patterns. Using large-scale EEG recordings and machine learning techniques, we investigated whether dexmedetomidine-induced deep sedation indeed mimics natural sleep patterns.

Methods: We used EEG recordings from three sources in this study: 8,707 overnight sleep EEG and 30 dexmedetomidine clinical trial EEG. Dexmedetomidine-induced sedation levels were assessed using the Modified Observer's Assessment of Alertness/Sedation (MOAA/S) score. We extracted 22 spectral features from each EEG recording using a multitaper spectral estimation method. Elastic-net regularization method was used for feature selection. We compared the performance of several machine learning algorithms (logistic regression, support vector machine, and random forest), trained on individual sleep stages, to predict different levels of the MOAA/S sedation state.

Results: The random forest algorithm trained on non-rapid eye movement stage 3 (N3) predicted dexmedetomidine-induced deep sedation (MOAA/S = 0) with area under the receiver operator characteristics curve >0.8 outperforming other machine learning models. Power in the delta band (0–4 Hz) was selected as an important feature for prediction in addition to power in theta (4–8 Hz) and beta (16–30 Hz) bands.

Conclusions: Using a large-scale EEG data-driven approach and machine learning framework, we show that dexmedetomidine-induced deep sedation state mimics N3 sleep EEG patterns.

Clinical Trials: Name—Pharmacodynamic Interaction of REMI and DMED (PIRAD), URL—https://clinicaltrials.gov/ct2/show/NCT03143972, and registration—NCT03143972.

Statement of Significance

Dexmedetomidine is commonly used in intensive care units for sedating critically ill patients. Neurophysiologically, sedation using dexmedetomidine is known to promote sleep like electroencephalogram (EEG) patterns. Studies evaluating this relationship are performed through visual assessments of the spectrogram of the EEG from a limited set of healthy volunteers and does not capture heterogeneity. In this study, we use machine learning techniques and publicly available sleep EEG datasets to identify the sleep stage that is highly correlated with dexmedetomidine-induced sedation. We also show that the presence of spindles is not sufficient to distinguish between awake state and dexmedetomidine-induced sedation levels.

Key words: electroencephalogram; sedation monitoring; machine learning; dexmedetomidine; sleep

Submitted: 11 December, 2019; **Revised:** 17 August, 2020

Introduction

Dexmedetomidine is an intravenous α2a adrenergic receptor agonist that is widely used in intensive care units (ICUs) for sedation and in clinical anesthesia during surgical procedures [1]. Recent findings that dexmedetomidine neurophysiologically approximates non-rapid eye movement (NREM) sleep [2–5], and reduces the incidence of delirium [6–8], have led to a renewed interest in administering dexmedetomidine to critically ill patients in ICUs. Investigations looking at the association between dexmedetomidine-induced sedation state and sleep stages relied mainly on the univariate (or visual) analysis of the electroencephalogram (EEG) spectrogram on a limited sample size and lacked external validation [2–5]. In addition, these studies do not account for inter- (and intra-) participant variability commonly seen in sleep and sedation EEG patterns.

Dexmedetomidine sedation is characterized by slow oscillations (0–4 Hz) and spindle oscillations (12–16 Hz) similar to sleep spindles [2, 4]. At clinically recommended doses, dexmedetomidine induces a light state of sedation during which the patient is arousable and has hemodynamic and respiratory stability [9]. When combined with propofol, it provides cardiovascular stability, improves patient safety by decreasing the incidence of airway obstruction and hypoxia [10] and demonstrates propofol sparing effect [11]. These properties have led to growing interest in the use of dexmedetomidine in ICUs and during surgical procedures, and understanding the exact mechanism of dexmedetomidine sedation is important. Though studies suggest that dexmedetomidine-induced deep sedation mimics NREM stage 2^2 and stage 3^3 sleep, a large-scale validation at a population level is necessary to explore these findings.

In our previous study [12], we developed a novel data-repurposing framework to predict anesthetic drug-induced deep sedation state from sleep EEG using deep learning algorithms. This framework was designed to eliminate the necessity for new clinical trials for developing automatic sedation level monitors. We observed that the deep learning algorithm provided the highest accuracy to predict dexmedetomidine-induced deep sedated state (Modified Observer's Assessment of Alertness/Sedation [MOAA/S] score = 0) using NREM stage 3 sleep EEG. However, the exact reason and features learnt by neural networks in the deep learning model could not be revealed due to its "black-box" nature. To further explore this finding, in this study, we developed a framework using large-scale sleep EEG data and machine learning techniques to validate if dexmedetomidine-induced sedation indeed mimics NREM sleep stages. We trained machine learning algorithms on two large online sleep EEG datasets (N = 8,707) to predict dexmedetomidine-induced sedation levels on an external clinical trial dataset (N = 30). We investigated the robustness of several machine learning models, trained on individual sleep stages, to predict different levels of dexmedetomidine-induced sedation. In this way, we identified the sleep stage that is highly correlated with the dexmedetomidine-induced deep sedation.

Methods

Dataset

Polysomnography (PSG) dataset from three different sources was used in this study: the online Sleep Heart Health Study (SHHS) public dataset (N = 5,804) [13–16], the osteoporotic fractures in men study (MrOS; N = 2,097) [13, 14, 17–19], and the University Medical Center Groningen (UMCG) dexmedetomidine clinical trial dataset (N = 30) [20]. The SHHS study (mean age: 63.1 ± 11.2 years, M = 2,728, F = 2,993) had two rounds of PSG recordings and we only used the first round (SHHS-1) due to the availability of a large number of patients and consistent sampling frequency (125 Hz) in all recordings. Similarly, the MrOS study (mean age 73.5 ± 5.8 years, M = 2,907, F = 0) had two visits and we used EEG recordings from the first visit (visit-1) with successful overnight PSG recordings. Prior permission was obtained from the online portal (www.sleepdata.org) for the analysis.

We used EEG recordings from two bipolar central EEG channels: C4/A1 and C3/A2 from the SHHS and the MrOS datasets. In both datasets, each 30 s EEG segment was manually scored into one of the six sleep states using Rechtschaffen and Kales (R&K) guidelines by sleep technicians: Wake (W), NREM stage 1(N1), stage 2 (N2), stage 3 (N3), stage 4 (N4), and rapid eye movement (R). This was converted into five stages according to the current American Academy of Sleep Medicine (AASM) guidelines [21]: W, N1, N2, N3 (N3 + N4), and R. We did not perform sleep stage scoring in the UMCG data, since the goal of this study was not to develop a sleep staging prediction algorithm.

A detailed description of UMCG dexmedetomidine dataset (mean age 40.7± 15.8 years, M = 15, F = 15) and experimental protocol can be found elsewhere [20]. In short, the EEG signals contain 17 channels and were recorded at a sampling frequency of 5 kHz from 30 healthy volunteers. Each volunteer received dexmedetomidine administered through target-controlled infusion with increasing targeted effect site concentrations: 1 ng/mL (0–40 minutes), 2 ng/mL (40–90 minutes), 3 ng/mL (90–130 minutes), 5 ng/mL (130–170 minutes), and 8 ng/mL (170–220 minutes), and was ceased afterwards. During this session, the six-level MOAA/S score [22] was used to estimate the sedation level ranging from 5 (awake) to 0 (nonresponsive). An example from the SHHS and the UMCG dataset is shown in Figure 1.

EEG preprocessing

EEG signals from two bipolar channels: C4/A1 and C3/A2 were used in this study as these specific channels were present in all three datasets. First, we used a bandpass filter to filter EEG between 0.1 and 30 Hz and then down sampled to 125 Hz (to match the SHHS dataset sampling frequency). We restricted the upper frequency limit to 30 Hz to minimize the influence of muscular and movement artifacts. We then segmented the EEG into 30 s nonoverlapping epochs corresponding to the duration of the sleep scoring for further analysis. Epochs satisfying any one of the following conditions were considered as artifacts and were not analyzed: (1) amplifier saturation/movement artifacts – abnormally high signal amplitude (>500 μV); (2) loose electrode artifacts – mean amplitude of the first channel = 2 × mean amplitude of the sum of signals. After artifact rejection, in total 5,767,772 and 4,101,684 epochs were obtained in the SHHS and MroS datasets, respectively. Similarly, for the UMCG data, we obtained 10,528 epochs corresponding to different MOAA/S scores. The distribution of the 30 s epochs in the different groups is shown in Figure 2.

Spectral feature extraction

Each 30 s EEG epochs was further divided into 4 s small subepochs (with 0.1 s shift) and the spectrogram was estimated

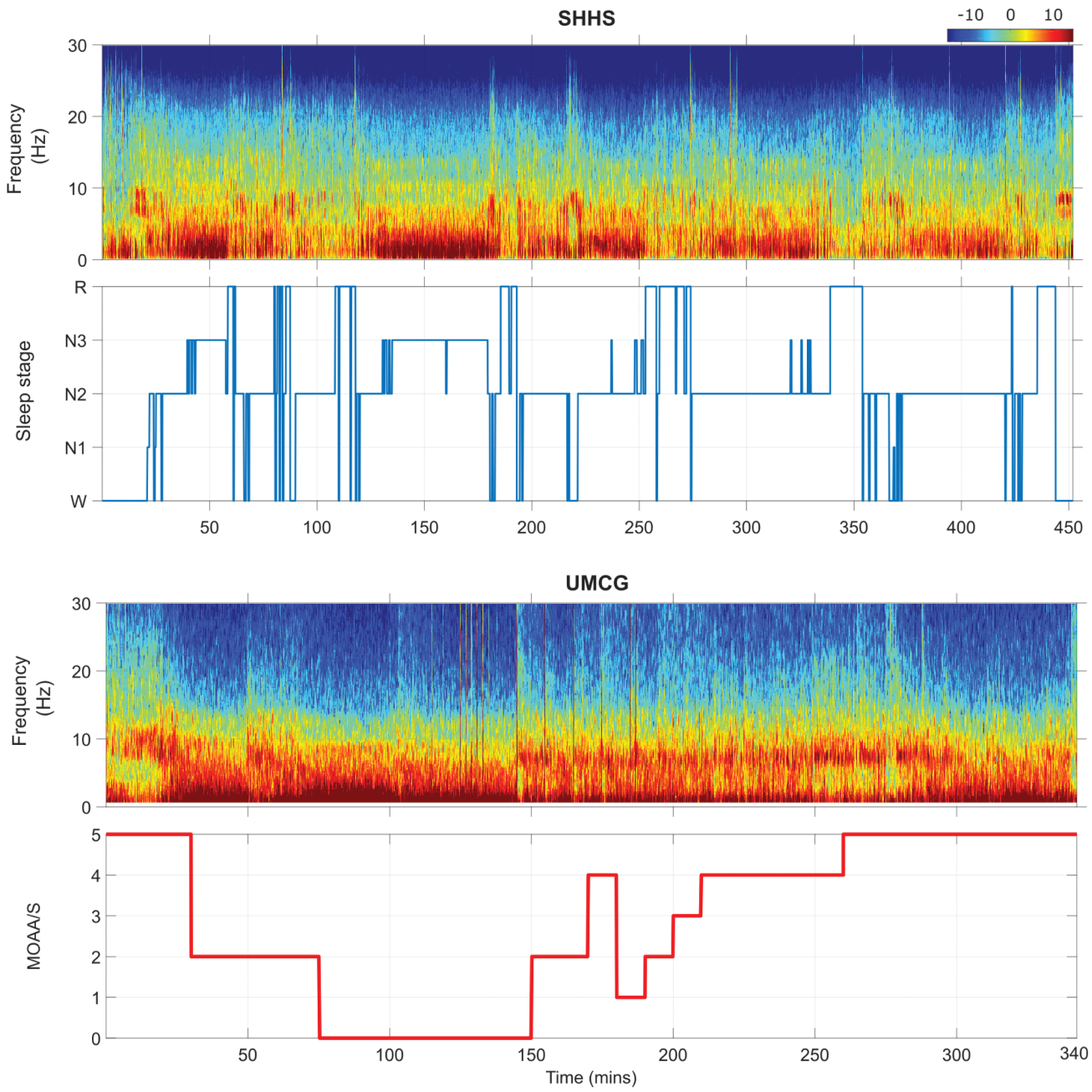

Figure 1. Example comparison of the multitaper EEG power spectrogram and their corresponding annotations from the SHHS and UMCG datasets. The spectrogram was obtained using multitaper spectral estimation via the chronux toolbox with the following parameters: time–bandwidth product TW = 3, window length T = 4 s (with 3.9 s overlap), number of tapers K = 5, and spectral resolution 2W of 1.5 Hz.

from each subepoch using Thompson's multitaper spectral estimation method [23] via Chronux toolbox [24] from each channel using the following parameters: length of the window T = 4 s with 3.9 s overlap, time–bandwidth product TW = 3, number of tapers K = 5, and spectral resolution 2W of 1.5 Hz. We estimated 22 spectral features from the spectrogram of each epoch: power in different subbands: delta (0.1–4 Hz; p_δ), theta (4–8 Hz; p_θ), alpha (8–12 Hz; p_α), spindle (12–16 Hz; p_σ), lower beta (16–20 Hz; $p_{\beta L}$), and upper beta (20–30 Hz; $p_{\beta U}$) bands, total power (0.1–30 Hz; p_T), normalized by total power— $\frac{p_\delta}{p_T}, \frac{p_\theta}{p_T}, \frac{p_\alpha}{p_T}, \frac{p_\sigma}{p_T}, \frac{p_{\beta L}}{p_T}, \frac{p_{\beta U}}{p_T}$, normalized by delta power— $\frac{p_\theta}{p_\delta}, \frac{p_\alpha}{p_\delta}, \frac{p_\sigma}{p_\delta}, \frac{p_{\beta L}}{p_\delta}, \frac{p_{\beta U}}{p_\delta}$, and normalized by theta power— $\frac{p_\alpha}{p_\theta}, \frac{p_\sigma}{p_\theta}, \frac{p_{\beta L}}{p_\theta}, \frac{p_{\beta U}}{p_\theta}$. We obtained median across channels and then averaged features from 4 s subepochs. Each 30 s EEG epoch (30 × 125 = 3,750 samples), was, therefore, represented by 22 spectral features. We decided to use only spectral features for a straightforward interpretation of the results since spectral analysis is commonly used in the clinical environment for EEG interpretation.

Training and testing

Since there are four sleep stages (N1, N2, N3, and R) and a wake stage (W), we trained four separate models for binary classification: WN1 = trained on W and N1, WN2 = trained on W and N2, WN3 = trained on W and N3, and WR = trained on W and

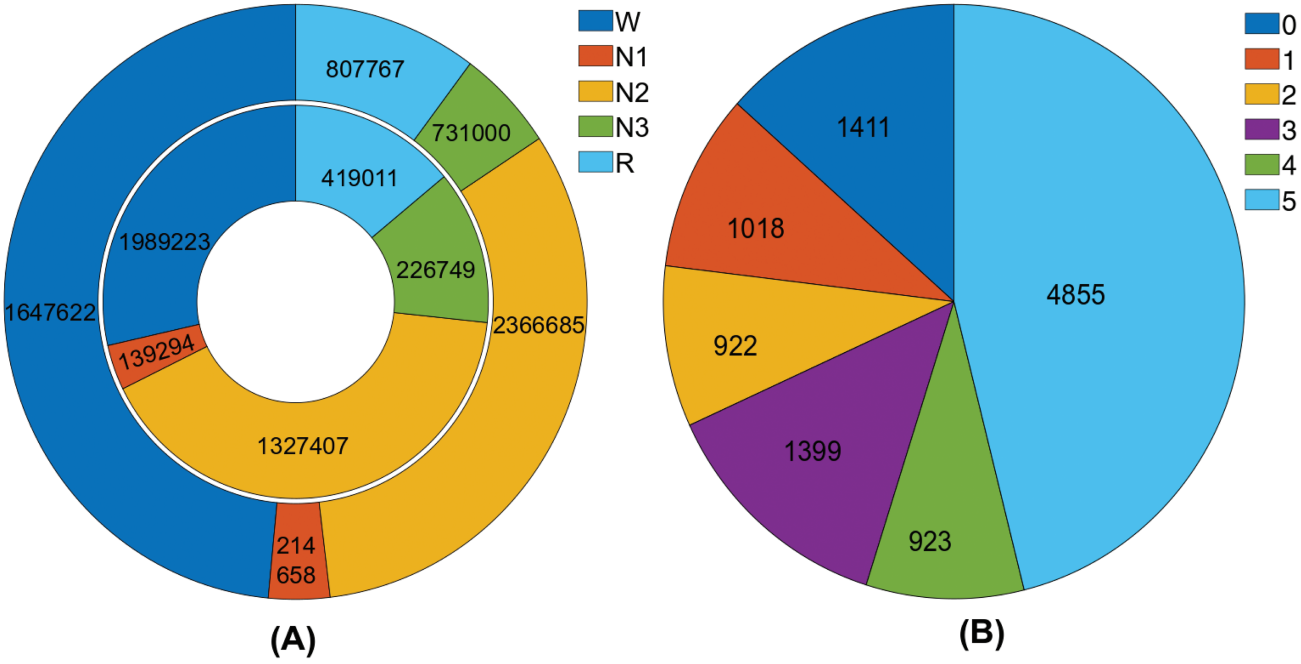

Figure 2. Distribution of (A) 30 s epochs in different sleep stages in SHHS (outer ring), MrOS (inner ring), (B) 30 s epochs in different levels of sedation (MOAA/S scores) in the UMCG datasets used in this study.

R. Each model was then used to differentiate between awake (MOAA/S = 5) and individual dexmedetomidine-induced hypnotic levels. For example, WN1 was used to differentiate between MOAA/S = 5 and MOAA/S = 4 (AM4), MOAA/S = 5 and MOAA/S = 3 (AM3), and so on for all training and testing combinations. Wake (W) EEG segments from SHHS dataset (and MOAA/S =5 from UMCG dataset) were labeled as 0, EEG segments corresponding to sleep stages (N1, N2, N3, and R) and hypnotic states (MOAA/S = 4, 3, 2, 1, and 0) were labeled as 1. Since the training data were highly imbalanced with unequal distribution of EEG segments in two classes, it can severely bias the performance of machine learning models [25]. To address the issue of class imbalance, we randomly selected an equal number of EEG segments corresponding to the length of the smallest class during training. For example, if there are K_1 segments from class 0, K_2 segments from class 1, and $K_1 > K_2$, then we randomly selected K_2 class 1 segments. In this way, we (1) identified which sleep stage is homologous with the dexmedetomidine-induced sedation, and (2) performed external validation by using completely independent datasets for the training and testing. This process is illustrated in Figure 3.

All features were z-score standardized before training and the testing set features were standardized with respect to the training set before using them for classification.

Internal cross validation

To evaluate the robustness of the prediction performances of machine learning models, we also evaluated the performance of the models using a 10-fold internal cross-validation technique within each dataset, i.e. we divided the data into 10 approximately equal folds. In each fold, we used 90% of the data for training the random forest classifier and the remaining unseen 10% data for testing. Since the training dataset consisted unequal number of epochs from two classes, we created a balanced training set as described in the previous section. In each fold, features in the training set were normalized (by subtracting the mean and dividing by the standard deviation) and the testing set features were normalized with respect to the mean and standard deviation of the training set before using them for classification. This process was repeated until each 10% unseen data was used for testing.

Algorithm selection

Since many features were highly correlated, we used elastic-net (EN) regularization technique [26] for feature selection before using them for classification with the *glmnet* package [27]. EN technique involves the selection of optimal hyperparameters: ridge coefficient (α) and coefficient shrinkage (λ). Setting $\alpha = 0$ is equivalent to ridge regression and can result in nonzero weights assigned to nondiscriminatory/irrelevant features due to L2 loss function. Setting $\alpha = 1$ is equivalent to LASSO regression which used L1 loss function and is not an ideal choice in dataset with highly correlated features. We set $\alpha = 0.5$ for equal contribution from L1/L2 loss function (EN regression), and selected λ that provided the lowest mean squared error in prediction during training (using 10-fold cross validation).

We used three traditional machine learning algorithms: logistic regression, support vector machine with the Gaussian kernel ($c = 1$, $\sigma = 1$) and random forest (500 trees) for classification. We did not perform any model hyperparameter selection since the goal of this study was not to develop an accurate prediction system but to validate a clinical hypothesis. The area under the receiver operator characteristics curve (AUC) was used to evaluate the performance of machine learning models. AUC is commonly used in binary classification problems which provides the performance ability of models to distinguish between two classes. All simulations were performed using MATLAB 2018a scripting language

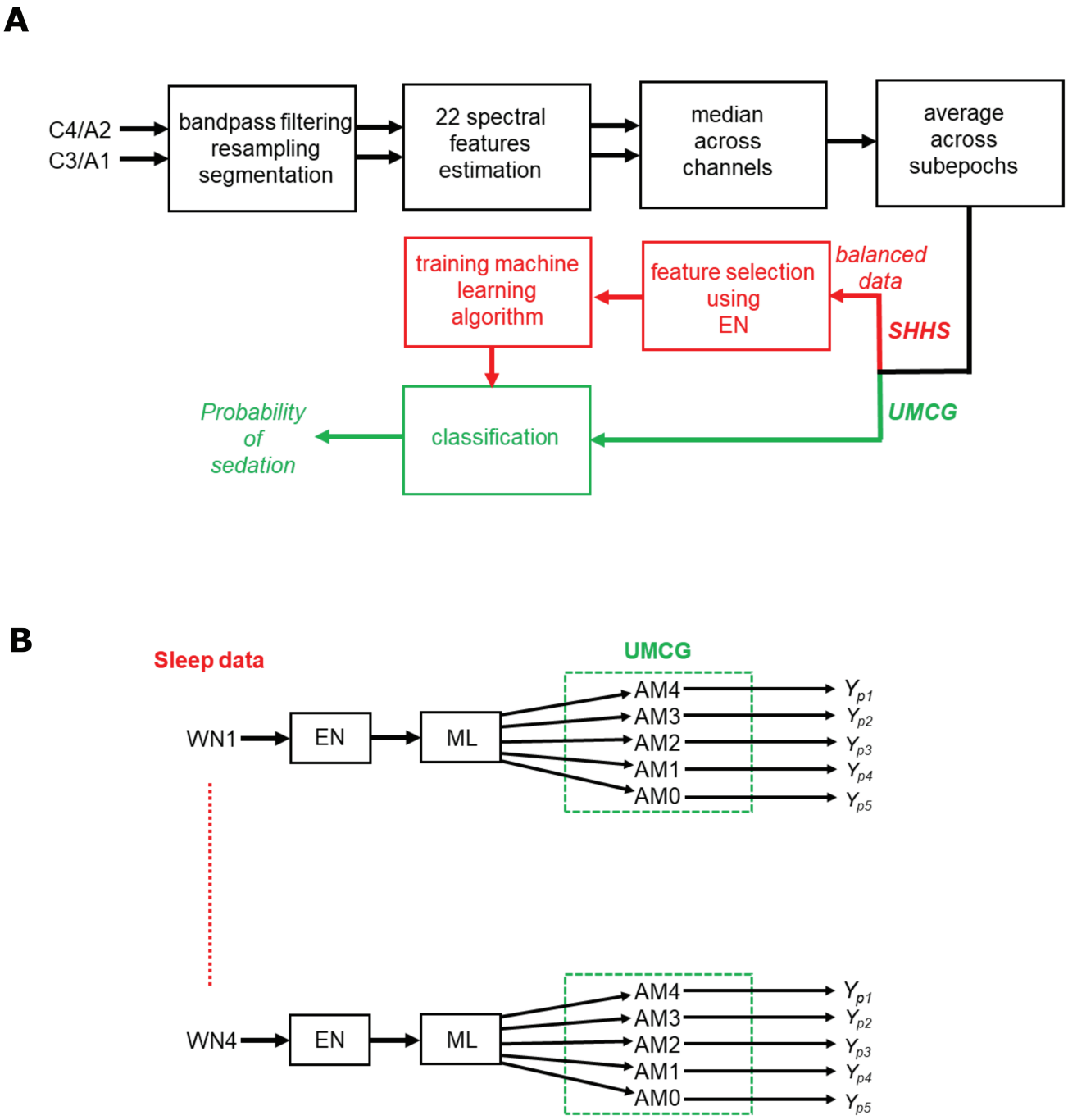

Figure 3. Illustration of the (A) proposed validation model used in this study and (B) different training and testing combinations performed in this study. Here, the machine learning model is trained in individual sleep states to predict different levels of sedation induced during dexmedetomidine infusion.

(Natick, MA). All results are reported as mean (95% confidence interval) unless explicitly stated otherwise. To estimate the 95% confidence interval, we used bootstrapping with 1,000 samplings. To test the statistical significance, we performed Wilcoxon rank sum test on the probability output of the machine learning models and p-values <0.05 were considered significant.

Results

Performance of machine learning models

The prediction performance of different models using SHHS and MrOS is summarized in Tables 1 and 2, respectively. The models trained on the WN3 sleep stage provided the best performance to predict dexmedetomidine-induced deep sedation (AM0). The highest prediction was obtained using random forest with an overall AUC = 0.83 (0.78–0.88) and 0.86 (0.85–0.87) using SHHS and MrOS datasets for training, respectively, outperforming other machine learning models.

Discriminatory features

Figure 4 is the heatmap of a list of features selected by the EN technique for all training testing combinations. Features that are more discriminative/informative selected by the algorithm

Table 1. Mean AUC of the machine learning models to predict different levels of sedation using different sleep stages in SHHS data

		Testing				
Model	Training	AM4	AM3	AM2	AM1	AM0
LR	WN1	0.57 (0.55–0.60)	0.53 (0.51–0.55)	0.58 (0.56–0.61)	0.55 (0.53–0.57)	0.54 (0.52–0.56)
	WN2	0.51 (0.50–0.54)	0.56 (0.54–0.58)	0.54 (0.51–0.56)	0.54 (0.52–0.57)	0.55 (0.53–0.57)
	WN3	0.56 (0.53–0.59)	0.58 (0.56–0.60)	0.59 (0.54–0.63)	0.62 (0.57–0.66)	0.76 (0.71–0.78)
	WR	0.51 (0.50–0.55)	0.55 (0.53–0.57)	0.53 (0.50–0.55)	0.52 (0.50–0.55)	0.51 (0.50–0.54)
SVM	WN1	0.56 (0.54–0.58)	0.54 (0.52–0.56)	0.52 (0.50–0.54)	0.54 (0.52–0.56)	0.57 (0.55–0.59)
	WN2	0.53 (0.51–0.54)	0.52 (0.50–0.55)	0.53 (0.51–0.54)	0.59 (0.58–0.61)	0.57 (0.56–0.59)
	WN3	0.55 (0.54–0.56)	0.55 (0.54–0.57)	0.54 (0.52–0.55)	0.60 (0.59–0.62)	0.80 (0.75–0.84)
	WR	0.51 (0.50–0.53)	0.58 (0.57–0.60)	0.51 (0.50–0.53)	0.55 (0.53–0.57)	0.58 (0.56–0.60)
RF	WN1	0.58 (0.55–0.60)	0.53 (0.51–0.56)	0.59 (0.57–0.61)	0.51 (0.50–0.54)	0.53 (0.51–0.55)
	WN2	0.51 (0.50–0.54)	0.59 (0.57–0.61)	0.59 (0.56–0.61)	0.53 (0.51–0.55)	0.58 (0.56–0.60)
	WN3	0.59 (0.56–0.61)	0.56 (0.54–0.59)	0.59 (0.56–0.62)	0.66 (0.64–0.68)	0.83 (0.78–0.88)
	WR	0.54 (0.52–0.57)	0.57 (0.55–0.59)	0.58 (0.55–0.60)	0.52 (0.50–0.54)	0.61 (0.59–0.63)

In all three algorithms, the model trained on WN3 had the highest accuracy to predict deep sedation (MOAA/S = 0). To estimate the 95% confidence interval, we used bootstrapping with 1,000 samplings. LR = logistic regression; SVM = support vector machine; RF = Random forest; WN1 = model trained on wake (W) and N1 sleep state; WN2 = model trained on W and N2 sleep state; WN3 = model trained on W and N3 sleep state; WR = model trained on W and rapid eye movement sleep state; AM4 = model tested to predict MOAA/S score 4; AM3 = model tested to predict MOAA/S score 3; AM2 = model tested to predict MOAA/S score 2; AM1 = model tested to predict MOAA/S score 1; AM0 = model tested to predict MOAA/S score 0.

Table 2. Mean AUC of the machine learning models to predict different levels of sedation using different sleep stages from MrOS data

		Testing				
Model	Training	AM4	AM3	AM2	AM1	AM0
LR	WN1	0.52 (0.51–0.54)	0.52 (0.50–0.53)	0.52 (0.51–0.54)	0.60 (0.58–0.61)	0.54 (0.52–0.55)
	WN2	0.54 (0.53–0.55)	0.55 (0.54–0.56)	0.52 (0.51–0.54)	0.55 (0.53–0.56)	0.60 (0.58–0.61)
	WN3	0.55 (0.54–0.56)	0.54 (0.53–0.55)	0.54 (0.52–0.55)	0.61 (0.59–0.62)	0.71 (0.69–0.72)
	WR	0.53 (0.52–0.54)	0.52 (0.51–0.53)	0.53 (0.52–0.54)	0.51 (0.50–0.52)	0.55 (0.54–0.57)
SVM	WN1	0.54 (0.51–0.57)	0.55 (0.53–0.57)	0.52 (0.50–0.54)	0.58 (0.55–0.60)	0.60 (0.58–0.63)
	WN2	0.54 (0.51–0.56)	0.51 (0.50–0.54)	0.51 (0.50–0.53)	0.52 (0.50–0.55)	0.58 (0.56–0.60)
	WN3	0.58 (0.55–0.61)	0.56 (0.54–0.58)	0.60 (0.58–0.63)	0.67 (0.65–0.70)	0.79 (0.77–0.80)
	WR	0.55 (0.53–0.58)	0.51 (0.50–0.53)	0.57 (0.54–0.60)	0.54 (0.52–0.57)	0.61 (0.59–0.63)
RF	WN1	0.53 (0.50–0.55)	0.59 (0.57–0.61)	0.59 (0.57–0.62)	0.60 (0.57–0.62)	0.54 (0.52–0.56)
	WN2	0.56 (0.53–0.59)	0.51 (0.50–0.53)	0.51 (0.450–0.53)	0.60 (0.57–0.62)	0.74 (0.72–0.76)
	WN3	0.55 (0.53–0.58)	0.59 (0.57–0.61)	0.67 (0.65–0.69)	0.74 (0.72–0.76)	0.86 (0.83–0.87)
	WR	0.53 (0.50–0.55)	0.57 (0.55–0.59)	0.54 (0.52–0.57)	0.57 (0.55–0.59)	0.57 (0.55–0.59)

Similar to SHHS, the model trained on WN3 had the highest accuracy to predict deep sedation (MOAA/S = 0). To estimate the 95% confidence interval, we used bootstrapping with 1,000 samplings. For abbreviations please refer to Table 1.

are represented by the shaded region. Different features were selected by the EN technique in different sleep stages. The model trained on W and N3 mainly used powers in the delta, theta, and beta bands to predict the dexmedetomidine deep sedation state (AM0). It should be noted that power in the spindle band in both SHHS and MrOS was not discriminatory.

Internal cross-validation performance

Next, we performed 10-fold internal cross validation to differentiate between Wake and N3 sleep stage in the SHHS and MrOS datasets; awake (MOAA/S = 5) and sedated (MOAA/S = 0) state in UMCG data using random forest algorithm. The algorithm classified two groups with an AUC of 0.94 (0.85–0.97), 0.96 (0.84–0.98), and 0.84 (0.79–0.90) in the SHHS, MrOS, and UMCG datasets, respectively.

Performance of individual features

To evaluate the performance of individual spectral EEG features, we performed classification within each dataset to discriminate between two groups: SHHS, MrOS (W vs N3), and UMCG (MOAA/S = 5 vs MOAA/S = 0). Figure 5 summarizes the prediction performance. The following top five features provided the highest discriminatory performance (mean prediction AUC): $\frac{p_{\beta L}}{p_T}$ (0.88), $\frac{p_{\beta U}}{p_T}$ (0.86), $\frac{p_{\beta U}}{p_\theta}$ (0.83), $\frac{p_\delta}{p_T}$ (0.80), and $\frac{p_\sigma}{p_T}$ (0.80) in SHHS; $\frac{p_{\beta U}}{p_T}$ (0.77), $\frac{p_{\beta L}}{p_T}$ (0.75), $\frac{p_{\beta U}}{p_\theta}$ (0.75), $\frac{p_\delta}{p_T}$ (0.74), and $\frac{p_\alpha}{p_\theta}$ (0.73) in MrOS; $\frac{p_{\beta U}}{p_\theta}$ (0.75), $\frac{p_{\beta L}}{p_T}$ (0.75), $\frac{p_\delta}{p_T}$ (0.74), and $\frac{p_\alpha}{p_\theta}$ (0.74) in UMCG datasets.

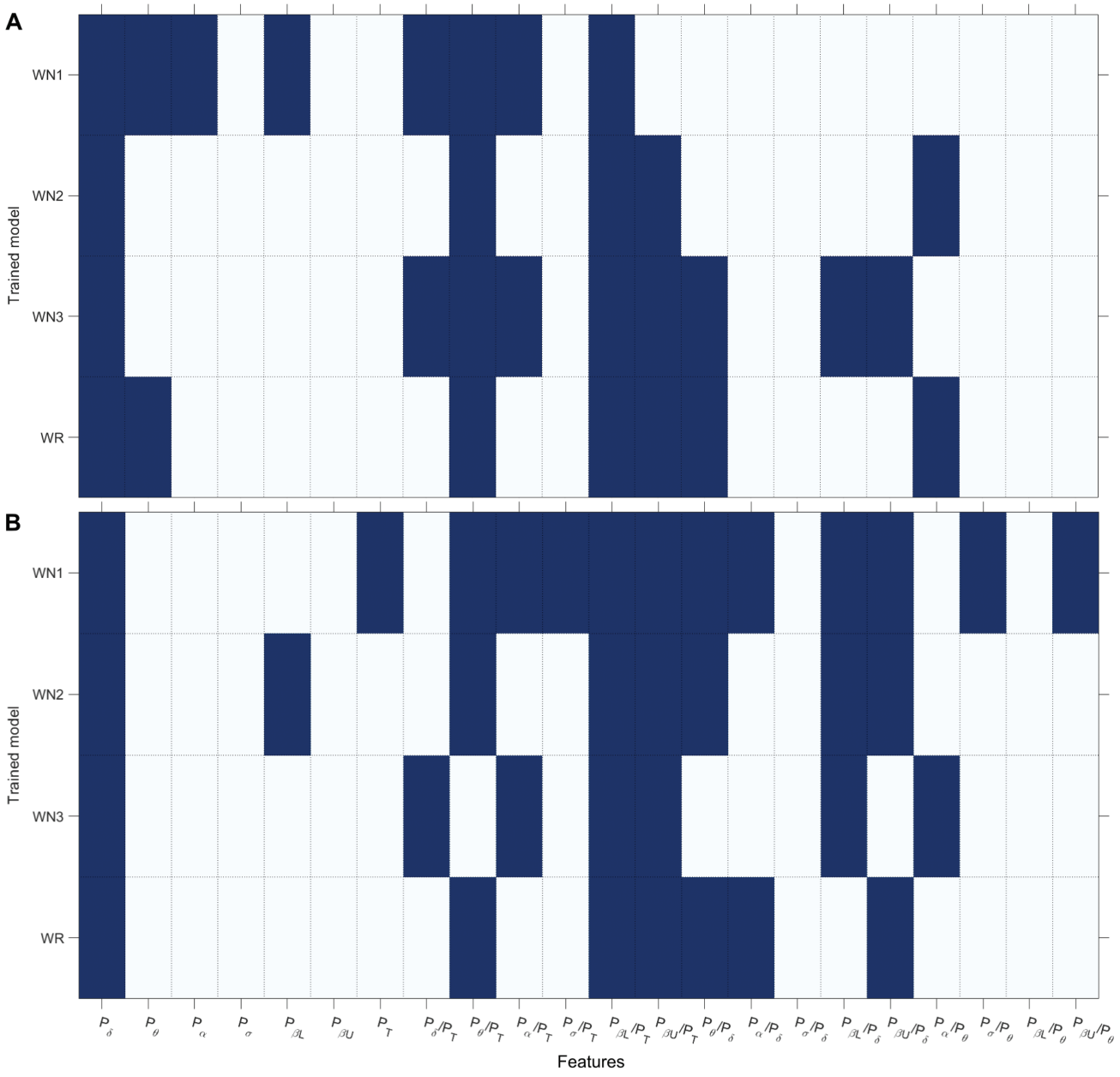

Figure 4. Heatmap of features selected by the EN regularization algorithm during the different training process. Here, the features selected by the EN algorithm are represented by the dark blue color. Different features were selected in (A) SHHS and (B) MrOS datasets. The power in the spindle band (normalized and un-normalized) was not selected as an important feature in both datasets.

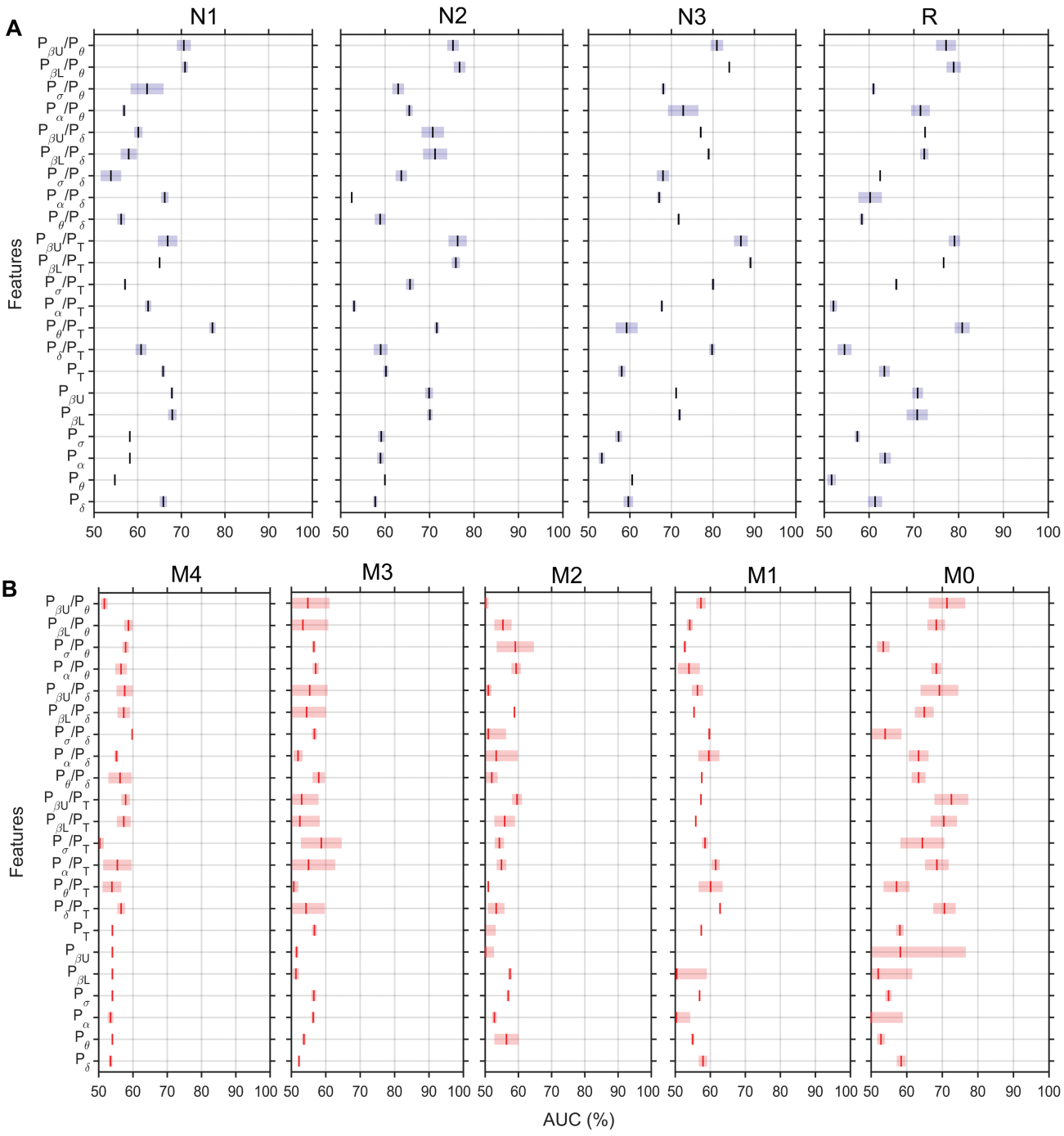

Figure 5. Prediction performance of individual features to differentiate between wake (W) and individual sleep stages in SHHS (red) and MrOS (blue) datasets; between awake (MOAA/S = 5) and individual sedated states in UMCG (blue) datasets. The prediction performance of power in the spindle band (normalized and un-normalized) to differentiate between awake and different sedation levels was poor (AUCs ≤0.6) in dexmedetomidine when compared with other spectral features. N1 = features tested to differentiate between wake and NREM stage 1 sleep state; N2 = features tested to differentiate between wake and NREM stage 2 sleep state; N3 = features tested to differentiate between wake and NREM stage 3 sleep state; R = features tested to differentiate between wake and rapid eye movement sleep state; M4 = features tested to differentiate between awake (MOAA/S score 5) and MOAA/S score 4; M3 = features tested to differentiate between awake (MOAA/S score 5) and MOAA/S score 3; M2 = features tested to differentiate between awake (MOAA/S score 5) and MOAA/S score 2; M1 = features tested to differentiate between awake (MOAA/S score 5) and MOAA/S score 1; M0 = features tested to differentiate between awake (MOAA/S score 5) and MOAA/S score 0.

Effect of age

To evaluate the effect of age on the prediction, we performed case controlled analysis matched by age. First, we grouped the EEG recordings into three groups in the UMCG data as group 1 = 18–35 years; group 2 = 35–50 years; and group 3 = 50–70 years. We then performed a binary classification between AM0 and WN3 using the random forest model for each group. The results are summarized in Table 3. In all three groups, the prediction performance was similar suggesting that the findings in this study can be generalized to all age groups (except below 18 years due to unavailability of data).

Effect of sex

To evaluate the effect of sex, we used the random forest model trained on the WN3 sleep stage to predict deep sedation separately for males and females. We only used SHHS data for training since MrOS study was conducted only in males. The classification results are summarized in Table 4. While the performance was better when trained and tested within the same sex compared with cross-training and testing, it was not significant ($p = 0.76$).

Discussion

In this study, we developed a framework using machine learning and large-scale EEG datasets to validate a clinical hypothesis: are EEG patterns during dexmedetomidine-induced deep sedation similar to NREM sleep EEG patterns? Results obtained in this study using both linear and nonlinear machine learning algorithms demonstrate that dexmedetomidine-induced deep sedation mimics the N3 sleep stage. Particularly, our best performing model using random forest algorithm predicted deep sedation with AUCs = 0.83 (0.78–0.88) and 0.86 (0.85–0.87) using features estimated from N3 sleep stage EEG in the SHHS and MrOS datasets, respectively. Two strengths of this study are (1) the use of a large number of heterogenous EEG recordings from two sleep studies, and (2) validation on an external dataset using machine learning algorithms. This type of validation is necessary to capture the inter- (and intra-) participant variability seen in the EEG recordings as shown in Figure 6. We believe that the framework and findings of this study using large heterogenous datasets can be generalized to external populations across different age groups and sex.

Table 3. Prediction performance (AUCs) obtained using the random forest model

	Training	
Testing	SHHS	MrOS
Group 1	0.81 (0.75–0.89)	0.84 (0.74–0.90)
Group 2	0.84 (0.76–0.91)	0.87 (0.72–0.90)
Group 3	0.82 (0.76–0.87)	0.86 (0.71–0.91)

The binary classification was performed between WN3 and AM0 for individual groups (in UMCG data). Here, group 1 = 18–35 years; group 2 = 35–50 years; and group 3 = 50–70 years. There was no significant difference in the prediction performance between individual groups ($p = 0.81$). To estimate the 95% confidence interval, we used bootstrapping with 1,000 samplings. To test the statistical significance, we performed Wilcoxon rank sum test on the probability output of the random forest model and p-values <0.05 were considered significant. For abbreviations please refer to Table 1.

Table 4. Prediction performance (AUCs) obtained using the random forest model across different groups. Though the performance reduced during cross training and testing, it was not significant.

	Training	
Testing	Male	Female
Male	0.85 (0.74–0.91)	0.81 (0.70–0.89)
Female	0.80 (0.71–0.90)	0.84 (0.72–0.90)

The binary classification was performed between WN3 and AM0 for males and females separately. The model was trained using only SHHS data. Though the performance dropped approximately by 4% during cross-training and testing, it was not significant ($p = 0.76$). To estimate the 95% confidence interval, we used bootstrapping with 1,000 samplings. To test the statistical significance, we performed Wilcoxon rank sum test on the probability output of the random forest model and p-values <0.05 were considered significant. For abbreviations please refer to Table 1.

Several previous studies have already looked at the association between natural sleep stages and dexmedetomidine-induced sedation states using EEG. In Huupponen et al. [4], EEG from 11 healthy participants undergoing dexmedetomidine sedation, and during physiological sleep in 10 healthy control participants were used to visually compare the sleep spindles between datasets. In this study, authors observe that dexmedetomidine produces a state closely resembling physiological N2 sleep state. However, since self-reported sedation assessment was not performed in these volunteers it is not clear from which sedation state the EEG spindles were used for comparison (these EEG segments were selected at times with high spindle density). In Purdon et al. [28], dexmedetomidine was administered as a low-dose infusion inducing a level of sedation in which the patient responds to minimal auditory or tactile stimulation (which translates to MOAA/S score = 4). It was shown that the dexmedetomidine spindles are consistent with a light state of sedation (MOAA/S score = 4 according to their definition of patient response). In Akeju et al. [5], authors show that the EEG dynamics induced by dexmedetomidine more closely approximates N2 sleep stage but in another study by the same authors [3], it is shown that dexmedetomidine promotes biomimetic N3 sleep stage in dose dependent manner. All these studies were performed through visual analysis (univariate analysis) of EEG spectrogram from limited set of healthy volunteers in a controlled environment which do not capture heterogeneity commonly seen in sleep physiology. Due to this, there are discrepancies between different studies and a large-scale validation of these findings was necessary. We took a data-driven approach to validate these findings and explored how dexmedetomidine-induced sedation states correspond with different sleep stages. Our results using two large external sleep EEG datasets suggest that at population level, dexmedetomidine-induced deep sedation (MOAA/S = 0) is synonymous to N3 sleep stage. Findings of our study are in line with several previous studies that have already demonstrated the relationship between dexmedetomidine-induced and natural sleep EEG patterns. This is due to the close proximity between the mechanisms of NREM sleep induction and how dexmedetomidine acts on brain circuits to induce sedation [29–32]. Using large sleep EEG datasets and machine learning algorithms, we further strengthened this relationship by capturing large heterogeneity seen in sleep EEG patterns.

Different anesthetics induce distinct anesthetic states by targeting related neural circuitries [31, 32]. The anesthetic states can be identified and monitored by drug-specific signatures in unprocessed EEG as well as spectrogram [28]. Although the spindles are brief and episodic, dexmedetomidine sedation is characterized by slow oscillations and spindle-like activity with frequency range and spatial distribution similar to propofol-induced frontal alpha oscillations [4, 33, 34]. The EN feature selection algorithm selected

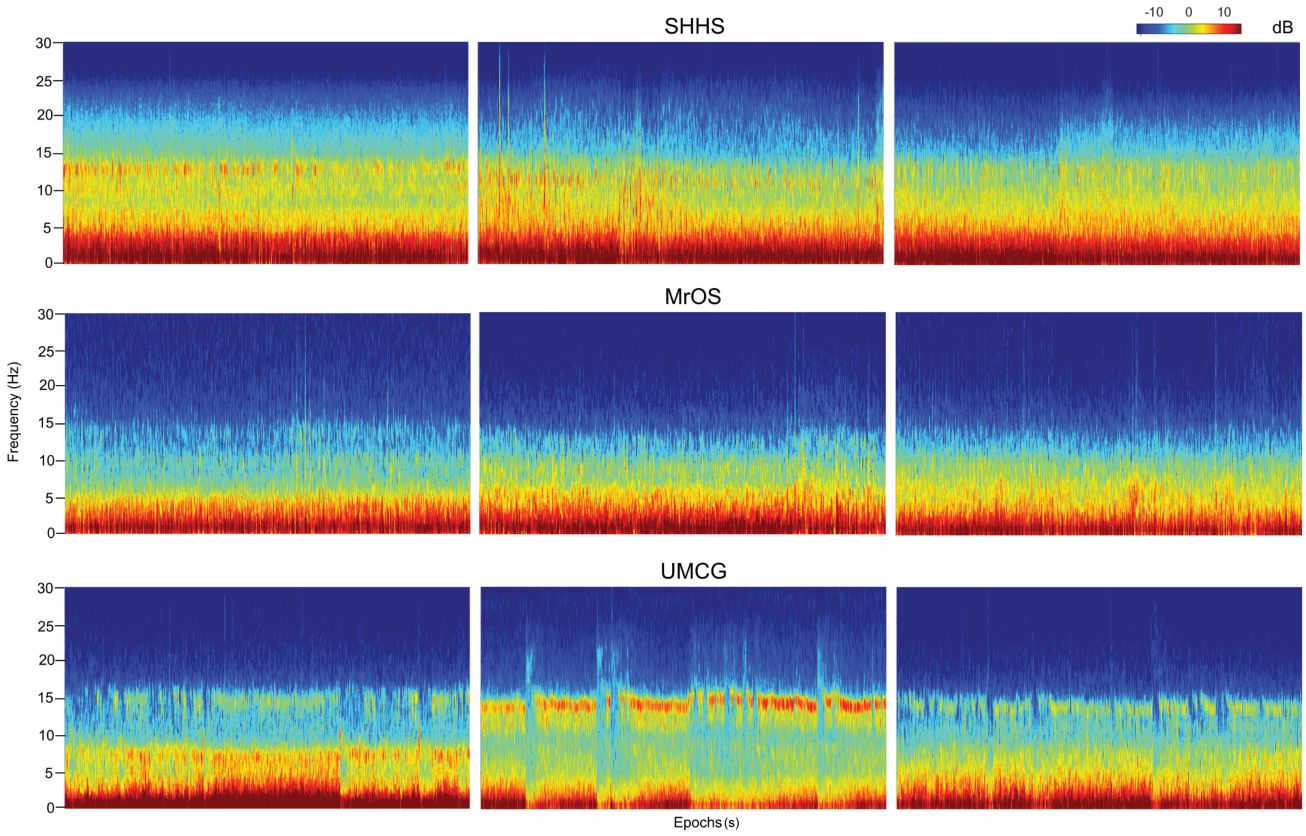

Figure 6. Comparison of EEG power spectrogram from three participants during N3 sleep state in SHHS, MrOS and deep sedated state (MOAA/S = 0) in UMCG datasets. Here, the multitaper spectrogram was obtained using all 30 s EEG epochs from N3 and deep sedated state. Large variability can be seen in delta (0–4 Hz) and spindle (12–16 Hz) spectral regions in all three datasets. This suggests that the morphology of spindles seen during the biological NREM stage 3 sleep is different from spindles seen during dexmedetomidine-induced deep sedation. N3 = NREM stage 3 sleep state.

different features across different sleep stages to predict deep sedation. Power in the delta, theta, alpha, and beta bands from the N3 sleep stage were selected by the EN algorithm to predict deep sedation state, which supports these findings (see Figure 4). It should be noted that power in the spindle band was not selected by the EN algorithm for prediction. When compared with other spectral features, power in the spindle band (normalized and un-normalized) was also not a good predictor to differentiate between awake and sedated states during dexmedetomidine infusion as a standalone feature as shown in Figure 5. For clinicians, this finding suggests that dexmedetomidine drug titration should not be performed only by visualizing the presence of spindles in the spectrogram. For manufacturers, it is important that sedation level monitors should not be developed solely based on the spectrogram/spectral features for tracking dexmedetomidine sedation levels.

There are several limitations to this study. First, we did not tune the model hyperparameters since the goal of this study was to validate a clinical hypothesis using linear and nonlinear machine learning algorithms instead of developing an accurate prediction model. Though different models performed differently in terms of prediction, all models provided highest performance to predict deep sedation state using N3 sleep stage. Second, we only used spectral features in this study for straightforward analysis and interpretation of results. The inclusion of additional features from time and entropy domain can improve the overall performance of the system for dexmedetomidine-induced sedation level prediction. Third, the results are yet to be validated in pediatric patients/group (<18 years). We could not perform this validation due to the unavailability of the data. Fourth, the SHHS and MroS datasets consisted of some patients with heterogenous pathophysiology and we did not exclude them from the analysis. However, we believe this is not an issue since the sleep stage scoring is performed using standard guidelines without any information about underlying pathophysiology.

To summarize, we demonstrate using a data-driven approach that dexmedetomidine-induced EEG patterns mimics NREM stage 3 EEG sleep patterns using linear and nonlinear machine learning algorithms. Our approach also provides a data-driven framework to validate clinical hypotheses at population level using big data and machine learning algorithms.

Acknowledgments

The authors wish to acknowledge the assistance of R. Spanjersberg, S. D. Atmosoerodjo, P. J. Colin, and A. R. Absalom (Department of Anaesthesiology, University Medical Center Groningen, The Netherlands).

Funding

This study was partially funded by the department of Anesthesiology, University of Groningen, University Medical Center Groningen, The Netherlands.

Conflict of interest statement. M.M.R.F.S.: His research group/department received (*over the last 3 years*) research grants and

consultancy fees from The Medicines Company (Parsippany, NJ), Masimo (Irvine, CA), Fresenius (Bad Homburg, Germany), Dräger (Lübeck, Germany), Paion (Aachen, Germany), and Medtronic (Dublin, Ireland). He receives royalties on intellectual property from Demed Medical (Temse, Belgium) and the Ghent University (Gent, Belgium). Other authors have no conflicts of interest to declare. Nonfinancial disclosure: All authors have no nonfinancial conflicts of interest to declare.

References

1. Mantz J, *et al*. Dexmedetomidine: new insights. *Eur J Anaesthesiol*. 2011;**28**(1):3–6.
2. Akeju O, *et al*. A comparison of propofol- and dexmedetomidine-induced electroencephalogram dynamics using spectral and coherence analysis. *Anesthesiology*. 2014;**121**(5):978–989.
3. Akeju O, *et al*. Dexmedetomidine promotes biomimetic non-rapid eye movement stage 3 sleep in humans: a pilot study. *Clin Neurophysiol*. 2018;**129**(1):69–78.
4. Huupponen E, *et al*. Electroencephalogram spindle activity during dexmedetomidine sedation and physiological sleep. *Acta Anaesthesiol Scand*. 2008;**52**(2):289–294.
5. Akeju O, *et al*. Spatiotemporal dynamics of dexmedetomidine-induced electroencephalogram oscillations. *PLoS One*. 2016;**11**(10):e0163431.
6. Pasin L, *et al*. Dexmedetomidine reduces the risk of delirium, agitation and confusion in critically ill patients: a meta-analysis of randomized controlled trials. *J Cardiothorac Vasc Anesth*. 2014;**28**(6):1459–1466.
7. McLaughlin M, *et al*. Dexmedetomidine and delirium in the ICU. *Ann Transl Med*. 2016;**4**(11):224.
8. Reade MC, *et al*.; DahLIA Investigators; Australian and New Zealand Intensive Care Society Clinical Trials Group. Effect of dexmedetomidine added to standard care on ventilator-free time in patients with agitated delirium: a randomized clinical trial. *JAMA*. 2016;**315**(14):1460–1468.
9. Shapiro FE. *Manual of Office-Based Anesthesia Procedures*. Lippincott Williams & Wilkins; 2007.
10. Kim KN, *et al*. Combined use of dexmedetomidine and propofol in monitored anesthesia care: a randomized controlled study. *BMC Anesthesiol*. 2017;**17**(1):34.
11. Kundra TS, *et al*. To evaluate dexmedetomidine as an additive to propofol for sedation for elective cardioversion in a cardiac intensive care unit: a double-blind randomized controlled trial. *Ann Card Anaesth*. 2017;**20**(3):337–340.
12. Belur Nagaraj S, *et al*. Predicting deep hypnotic state from sleep brain rhythms using deep learning: a data-repurposing approach. *Anesth Analg*. 2020;**130**(5):1211–1221.
13. Dean DA II, *et al*. Scaling up scientific discovery in sleep medicine: the National Sleep Research Resource. *Sleep*. 2016;**39**(5):1151–1164.
14. Zhang GQ, *et al*. The National Sleep Research Resource: towards a sleep data commons. *J Am Med Inform Assoc*. 2018;**25**(10):1351–1358.
15. Quan SF, *et al*. The Sleep Heart Health Study: design, rationale, and methods. *Sleep*. 1997;**20**(12):1077–1085.
16. Redline S, *et al*. Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. Sleep Heart Health Research Group. *Sleep*. 1998;**21**(7):759–767.
17. Blackwell T, *et al*.; Osteoporotic Fractures in Men Study Group. Associations between sleep architecture and sleep-disordered breathing and cognition in older community-dwelling men: the Osteoporotic Fractures in Men Sleep Study. *J Am Geriatr Soc*. 2011;**59**(12):2217–2225.
18. Blank JB, *et al*. Overview of recruitment for the osteoporotic fractures in men study (MrOS). *Contemp Clin Trials*. 2005;**26**(5):557–568.
19. Orwoll E, *et al*. Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study—a large observational study of the determinants of fracture in older men. *Contemp Clin Trials*. 2005;**26**(5):569–585.
20. Weerink MAS, *et al*. Pharmacodynamic interaction of remifentanil and dexmedetomidine on depth of sedation and tolerance of laryngoscopy. *Anesthesiology*. 2019;**131**(5):1004–1017.
21. Berry RB, *et al*. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules Terminol Tech Specif*. Darien, IL: Am Acad Sleep Med; Published online 2012.
22. Chernik DA, *et al*. Validity and reliability of the Observer's Assessment of Alertness/Sedation Scale: study with intravenous midazolam. *J Clin Psychopharmacol*. 1990;**10**(4):244–251.
23. Thomson DJ. Spectrum estimation and harmonic analysis. *Proc IEEE*. 1982;**70**(9):1055–1096.
24. Percival DB, *et al*. *Spectral Analysis for Physical Applications*. Cambridge University Press; 1993.
25. Chawla NV. Data mining for imbalanced datasets: an overview. In: *Data Mining and Knowledge Discovery Handbook*. Springer; 2009: 875–886.
26. Zou H, *and* Trevor H. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol*. 2005;**67**(2):301–320.
27. Qian J, *et al*. *Glmnet for Matlab*. 2013. Published online 2013. http://www.stanford.edu/~hastie/glmnet_matlab/
28. Purdon PL, *et al*. Clinical electroencephalography for anesthesiologists: Part I: Background and basic signatures. *J Am Soc Anesthesiol*. 2015;**123**(4):937–960.
29. Guldenmund P, *et al*. Brain functional connectivity differentiates dexmedetomidine from propofol and natural sleep. *Br J Anaesth*. 2017;**119**(4):674–684.
30. Sanders RD, *et al*. Contribution of sedative-hypnotic agents to delirium via modulation of the sleep pathway. *Can J Anaesth*. 2011;**58**(2):149–156.
31. Brown EN, *et al*. General anesthesia, sleep, and coma. *N Engl J Med*. 2010;**363**(27):2638–2650.
32. Brown EN, *et al*. General anesthesia and altered states of arousal: a systems neuroscience analysis. *Annu Rev Neurosci*. 2011;**34**:601–628.
33. Aksu R, *et al*. The comparison of the effects of dexmedetomidine and midazolam sedation on electroencephalography in pediatric patients with febrile convulsion. *Paediatr Anaesth*. 2011;**21**(4):373–378.
34. Mason KP, *et al*. Effects of dexmedetomidine sedation on the EEG in children. *Paediatr Anaesth*. 2009;**19**(12):1175–1183.