OPEN

# Predicting Deep Hypnotic State From Sleep Brain Rhythms Using Deep Learning: A Data-Repurposing Approach

Sunil Belur Nagaraj, PhD,* Sowmya M. Ramaswamy, Msc,† Maud A. S. Weerink, MD,† and Michel M. R. F. Struys, MD, PhD, FRCA†‡

**BACKGROUND:** Brain monitors tracking quantitative brain activities from electroencephalogram (EEG) to predict hypnotic levels have been proposed as a labor-saving alternative to behavioral assessments. Expensive clinical trials are required to validate any newly developed processed EEG monitor for every drug and combinations of drugs due to drug-specific EEG patterns. There is a need for an alternative, efficient, and economical method.

**METHODS:** Using deep learning algorithms, we developed a novel data-repurposing framework to predict hypnotic levels from sleep brain rhythms. We used an online large sleep data set (5723 clinical EEGs) for training the deep learning algorithm and a clinical trial hypnotic data set (30 EEGs) for testing during dexmedetomidine infusion. Model performance was evaluated using accuracy and the area under the receiver operator characteristic curve (AUC).

**RESULTS:** The deep learning model (a combination of a convolutional neural network and long short-term memory units) trained on sleep EEG predicted deep hypnotic level with an accuracy (95% confidence interval [CI]) = 81 (79.2–88.3)%, AUC (95% CI) = 0.89 (0.82–0.94) using dexmedetomidine as a prototype drug. We also demonstrate that EEG patterns during dexmedetomidine-induced deep hypnotic level are homologous to nonrapid eye movement stage 3 EEG sleep.

**CONCLUSIONS:** We propose a novel method to develop hypnotic level monitors using large sleep EEG data, deep learning, and a data-repurposing approach, and for optimizing such a system for monitoring any given individual. We provide a novel data-repurposing framework to predict hypnosis levels using sleep EEG, eliminating the need for new clinical trials to develop hypnosis level monitors. (Anesth Analg 2020;130:1211–21)

## KEY POINTS

- **Question:** Because anesthetic drugs exhibit sleep-like patterns during deep hypnosis, can we predict hypnosis level from sleep brain rhythms?
- **Findings:** Deep learning algorithms when trained on nonrapid eye movement stage 3 sleep electroencephalogram can predict dexmedetomidine-induced deep hypnotic level.
- **Meaning:** Anesthetic-induced hypnosis levels can be predicted using sleep electroencephalogram and artificial intelligence techniques, eliminating the need for clinical trials to develop hypnotic level monitors.

## GLOSSARY

**1D** = 1-dimensional; **AASM** = American Academy of Sleep Medicine; **AUC** = area under the receiver operator characteristic curve; **CI** = confidence interval; **CNN** = convolutional neural networks; **DL** = deep learning; **EEG** = electroencephalogram; **ICU** = intensive care unit; **LSTM** = long short-term memory; **MOAA/S** = Modified Observer's Assessment of Alertness/Sedation Scale; **NREM** = nonrapid eye movement; **ReLU** = rectified linear unit; **REM** = rapid eye movement; **R&K** = Rechtschaffen and Kales; **SHHS** = Sleep Heart Health Study; **TW** = time-bandwidth product; **UMCG** = University Medical Center Groningen

Current practice for monitoring the hypnotic component of anesthesia relies mainly on intermittently obtained patient's response to a verbal and/or tactile stimulus.[1] Brain monitors that track quantitative electroencephalogram (EEG) signatures to monitor anesthesia have been proposed as an alternative to clinical hypnosis assessments.[2,3] Although they have widespread use in clinical practice, their performance is limited (lack of consistency and reliability) and is drug-specific.[2–4] One main reason for such limited performance is that these monitors are developed using a small data set from controlled clinical trials using specific drugs and do not capture large heterogeneity between patients. In addition, expensive clinical trials are required to develop and/or validate any newly developed processed EEG monitor for every drug and combinations of drugs due to drug-specific EEG patterns.

In recent years, large publicly available heterogeneous expert labeled data sets have provided several benefits for developing clinical decision tools using deep learning (DL) algorithms. One such application is EEG-based sleep scoring systems where the DL algorithm is trained to automatically score 5 sleep stages and have already reached expert-level performance.[5–7] Recent clinical studies suggest that anesthetic drugs also induce specific sleep-like EEG patterns at different levels of hypnosis.[8] For example, propofol induces slow waves in EEG at deep hypnotic levels resembling slow waves of nonrapid eye movement (NREM) sleep EEG[9]; dexmedetomidine approximates NREM sleep with slow waves and spindle-like patterns in the EEG during deep hypnotic state.[10–13]

Motivated by numerous studies demonstrating sleep-like inhibition of anesthetic drugs, and major breakthroughs in the application of DL algorithms for hypnosis monitoring[14–16] and sleep staging[5–7] using EEG, we propose a novel data-repurposing framework to predict Anesthesia-induced hypnotic levels from sleep EEG using DL in this study. DL algorithms learn patterns directly from the raw EEG data eliminating the necessity to extract hand-crafted engineering features from EEG for prediction. We demonstrate this framework by using dexmedetomidine as a prototype drug. We train the DL algorithm on a publicly available sleep EEG data set (5723 subjects) to predict different levels of hypnosis on the independent dexmedetomidine clinical trial EEG data set (30 subjects). We hypothesized that the DL model trained on sleep data set should be able to track dexmedetomidine-induced hypnosis. This will enable the development of a clinical EEG monitor with much broader application possibilities without the need for validating every drug-induced EEG change with a clinical trial.

## METHODS
### Data Set
EEG recordings used in this study were obtained from 2 different sources: A dexmedetomidine clinical study data set (N = 30, mean age: 40.7 ± 15.8 years, male = 15, female = 15) from The University Medical Center Groningen (UMCG) and the publicly available Sleep Heart Health Study (SHHS) data set (N = 5723, mean age: 63.1 ± 11.2 years, male = 2728, female = 2993).[17–20] The dexmedetomidine clinical trial was conducted in accordance with the Declaration of Helsinki and applicable good clinical practice and regulatory requirements. The study had ethical approval from the "The Independent Ethics Committee" (Medisch Ethische Toetsings Commissie) of the Foundation "Evaluation of Ethics in Biomedical Research" (Stichting BEBO), Assen, the Netherlands. The dexmedetomidine clinical trial study was registered before patient enrollment at Clinical Trials.gov (Identifier: NCT03143972, principal investigator: Michel M. R. F. Struys, date of registration: June 28, 2017). Informed written consent was obtained from all volunteers before EEG recordings. Permission to use the SHHS data set was obtained from the online portal: www.sleepdata.org. A detailed description of UMCG dexmedetomidine data set and experimental protocol can be found elsewhere.[21]

The levels of hypnosis in the UMCG data set were scored by 3 expert anesthesiologists using the Modified Observer's Assessment of Alertness/Sedation (MOAA/S) score.[22] MOAA/S scores denote 6 levels of hypnosis ranging from 5 (responding readily to name spoken in normal tone) to 0 (not responding to a painful trapezius squeeze/deep hypnotic state). The initial sleep scores of SHHS data set (using the Rechtschaffen and Kales [R&K] guidelines[23]) were converted to the American Academy of Sleep Medicine (AASM) guidelines[24] by combining NREM stages 3 and 4 as single NREM stage 3: wake (W), NREM sleep: stages 1 (N1), 2 (N2), and 3 (N3), and rapid eye movement (REM)—R. EEG recordings with <5 sleep stages in SHHS data set were excluded from the analysis, resulting in a total of 5723 EEG recordings (initial SHHS data set consisted of 5804 EEG recordings). We excluded EEG recordings with <5 sleep stages to remove patients with severe sleep disorders. Sleep stage scoring was not performed in the UMCG data set because the goal of this study was not to develop another automatic sleep scoring system but to develop a framework predicting dexmedetomidine-induced hypnosis levels using sleep EEG. A priori power analysis was not performed to guide sample size in data collection.

### EEG Recordings
The UMCG data set consisted of 17 channel scalp EEG (Fp1, Fp2, F3, Fz, F4, T7, C3, Cz, C4, T8, P3, Pz, P4, O1, O2, A1, A2) and SHHS data set had 2 central

EEG channels: primary (C4/A1) and secondary (C3/A2). EEG recordings from subjects in UMCG data set were collected using BrainAmp DC32 amplifier with a BrainVision recorder at a sampling frequency of 5 kHz. For the entire study duration, subjects were instructed to close their eyes. Subjects with neurological/cardiovascular/pulmonary/gastric/endocrinological disorders, history of psychoactive medications usage, >20 g/d alcohol consumption, or pregnancy were not included in the study.

Dexmedetomidine was administered in a step-up dosing regimen by using the effect-site target-controlled infusion using the Hannivoort-Colin model.[25,26] First, 5-minute baseline data were obtained in which subjects were asked to relax and close their eyes. Later, dexmedetomidine was administered using the following effect-site target concentration: 1, 2, 3, 5, 8 ng/mL for 40 (0–40), 50 (40–90), 40 (90–130), 40 (130– 70), and 50 (170–220) minutes, respectively. This dosage regimen allowed all effect sites to reach a steady state. Dexmedetomidine infusion was ceased after the 220th minute. The MOAA/S assessment was performed at baseline, at each infusion step, and during the recovery phase after cessation. In addition, before each increase in infusion step, laryngoscopy was performed if the MOAA/S score was <2. Except for the MOAA/S assessments, volunteers were not stimulated and ambient noise was kept low throughout the study session. Figure 1 shows the behavioral response (MOAA/S scores) and the corresponding EEG spectrogram of a subject from the UMCG data set. More details about the dexmedetomidine data set can be found in Weerink et al.[21]

## EEG Preprocessing and Epoch Extraction

For the present study, we used 2 EEG channels common to UMCG and SHHS data set: C4/A1 and C3/A2. We first bandpass filtered the EEG signals between 0.5 and 30 Hz and then downsampled to 125 Hz (to match SHHS data set sampling frequency). To reduce the impact of differences in the amplifiers during EEG acquisition (which may have significantly affected the amplitude of the EEG signals), we standardized the EEG to have zero median and unit interquartile range for the entire recording in both data sets. We restricted the upper-frequency range to 30 Hz to eliminate the majority of muscle artifacts during the awake state. The EEG data were divided into nonoverlapping 30-second segments resulting in a total of 5,767,772 and 10,528 segments in SHHS and UMCG data set, respectively. Supplemental Digital Content, Figure 1, http://links.lww.com/AA/D7, shows the distribution of 30-second segments in different classes for both data sets.

## DL Architecture

We used LSTM-CNN architecture: a combination of a convolutional neural network (CNN) and long short-term memory units (LSTM), which was recently used for EEG-based expert-level sleep stage scoring[5] as shown in Figure 2 to predict levels of hypnosis. The CNN module extracts discriminative features from the raw EEG, and the LSTM module captures temporal dynamics in the EEG. To obtain the probability score, we used a final dense layer with sigmoid activation. We used glorot uniform initializer to initialize the weights of the neural network and trained the LSTM-CNN from scratch. To avoid overfitting, we used $L_2$ weight regularization and the model was
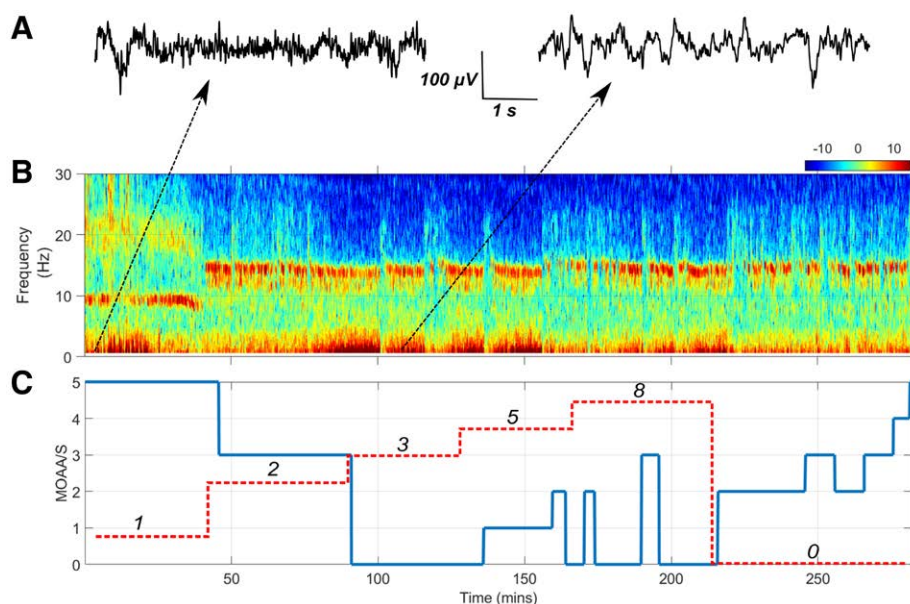


**Figure 1.** Sample dexmedetomidine data. Illustration of (A) 15 s sample EEG at minute 5 and minute 108, (B) C4/A1 channel EEG spectrogram, and (C) MOAA/S score of a subject from UMCG data set and red-dotted line shows target-controlled infusion of dexmedetomidine in nanogram per milliliter. We can see the presence of spindle waves with an increase in the level of hypnosis. The following values were set to perform spectral estimation using multitaper spectral estimation via the chronux toolbox: length of the window $T = 4$ s with 0.1 s shift, time-bandwidth product TW = 3, number of tapers $K = 5$, and spectral resolution 2 W of 1.5 Hz. EEG indicates electroencephalogram; MOAA/S, Modified Observer's Assessment of Alertness/Sedation Scale; TW, time-bandwidth product.
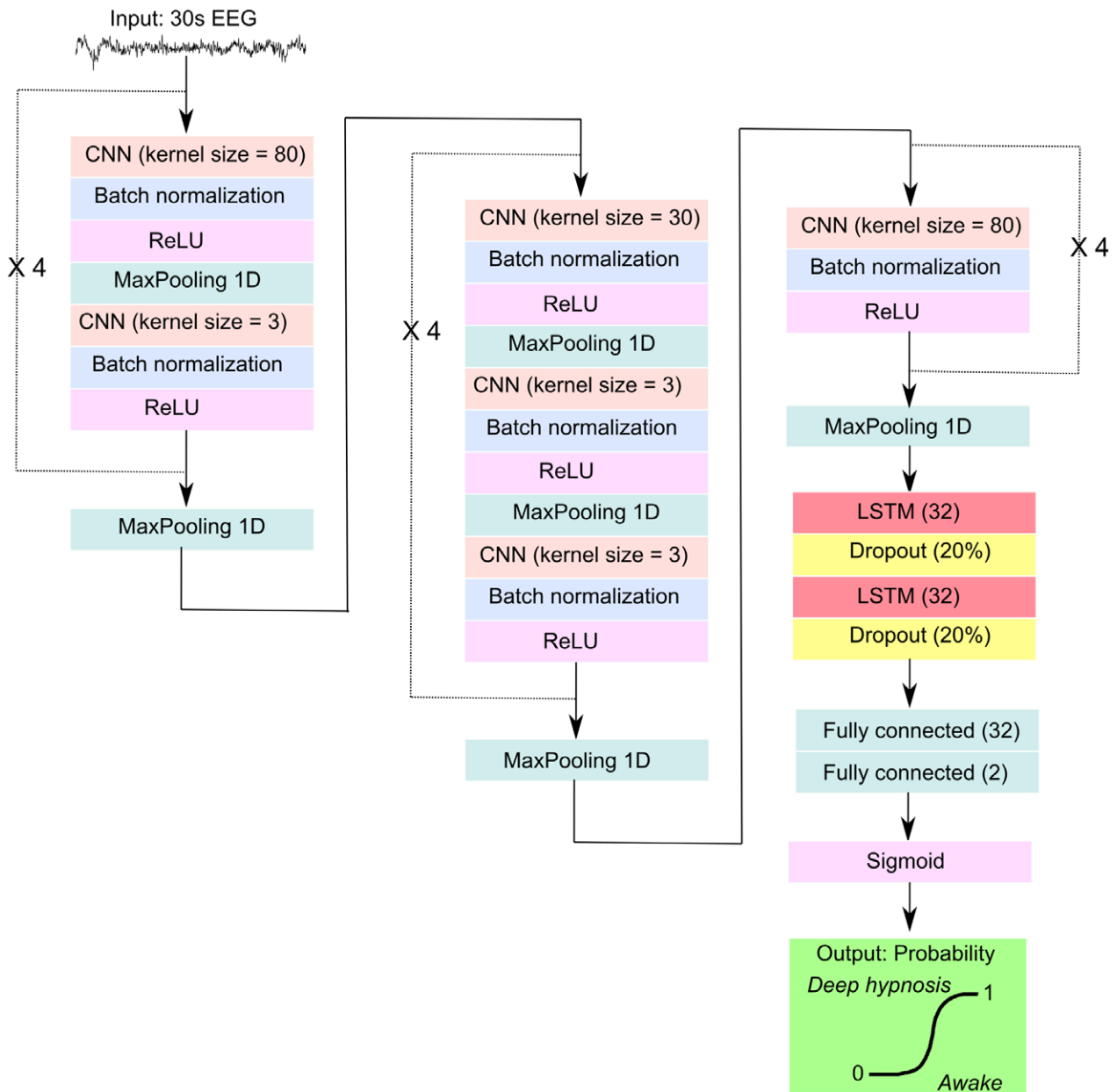
**Figure 2.** The architecture of the LSTM-CNN model[5] used in this study. The length of the input 1D EEG segment is 125 (samples) × 30 (seconds). The output of the model provides a probability score of a given EEG segment belonging to deep hypnotic state. "x4" refers to number of layers of residual network. In this architecture there are 4 + 4 + 4 = 12 layers of residual network. 1D indicates 1-dimensional; CNN, convolutional neural networks; EEG, electroencephalogram; LSTM, long short-term memory; ReLU, rectified linear unit.

trained using the stochastic gradient descent algorithm (learning rate = 0.01, momentum = 0.9, weight decay = 0.0001), and binary cross-entropy as a loss function. This architecture demonstrated to perform expert-level sleep scoring using large-scale sleep EEG data with rigorous hyper-parameter tuning similar to the architecture used in Biswal et al.[5] All experiments were performed on a local computer with Intel Xeon 4116, 32GB RAM, NVidia 1080Ti GPU, and CUDA 9.0. LSTM-CNN models were implemented using Keras wrapper with Tensorflow 2.0 backend in Python scripting language.

**Training and Testing**

To identify which sleep stage predicts different levels of hypnosis induced by the dexmedetomidine infusion, we performed the following binary classifications:

1. Label awake stage as 0 and sleep stage as 1, that is, W = 0, N1 = 1 in SHHS data set (denoted as WN1). Similarly, label awake state as 0 and hypnotic state as 1, that is, MOAA/S score 5 = 0, MOAA/S score 4 = 1 in the UMCG data set (denoted as M54).

2. Balance the data using undersampling group equalization strategy (select random epochs from both groups corresponding to the length of minority group) to set random chance level prediction accuracy to 50% in both data sets.
3. Train the LSTM-CNN model on WN1.
4. Predict the probability of hypnotic level in M54 using the trained model.
5. Repeat steps 1–3 until all MOAA/S states are used for prediction in step 4 (M53, M52, M51, M50).
6. Repeat steps 1–5 until all sleep stages are used for training (WN2, WN3, WR).

This process is illustrated in Figure 3. We performed a binary classification instead of multiclass prediction for 2 reasons. First, the primary goal of this study was to identify which individual sleep stage corresponded to a different level of MOAA/S score and not to predict 6 levels of MOAA/S scores from 5 sleep stages. Training the model to track different levels of hypnosis of sedation based on varying stages of sleep is not ideal because the annotation systems are different in the 2 data sets. Second, a multiclass prediction model will again result in discrete hypnotic level scores. Because hypnotic level is continuous, it is desirable to obtain a continuous score and we achieved this by means of probabilistic estimation using a sigmoid layer.

We fixed the batch size to 500 and numbers of epochs to 100 for model training, which means that the training data were provided 100 times to the network in chunks of 500 segments. We used 90% of the SHHS data (5150 patients) for training LSTM-CNN model and 10% (573 patients) for validation, and the UMCG data were held out as a completely independent test set. Since multiple EEG segments from the same patient were included in the analysis,

we ensured that the EEG segments in both sets were independent, that is, no overlap of patients in training and validation sets. Model training was terminated if (1) the validation accuracy reached 100%, or (2) finished 100 epochs, or (3) no change in the loss function of the validation set. After training, from each trained model, we predicted the hypnosis level on the UMCG data set. The accuracy and the $Y_p$ of the awake or hypnotic state of the EEG segment were estimated. Here, $Y_p = 1$ and $Y_p = 0$ corresponds to deep hypnotic and awake states, respectively. The classification was performed separately for 2 channels.

### Internal Cross-Validation
To evaluate how well LSTM-CNN model performs when trained and tested on the same data, we also performed internal 5-fold cross-validation, that is, trained and tested the model within same data (train and test on SHHS data; train and test on UMCG data) when compared to trained on one data (SHHS) and tested on other (UMCG).

### Continuous Hypnotic Level Assessment
Because hypnosis level is continuous, it is important to obtain a continuous probabilistic estimation of level of hypnosis. The proposed framework in this study raises an important question: given the output of the sleep stage prediction model, which MOAA/S score does the model predict for a new EEG segment? To obtain a continuous level of hypnosis, we performed the following: for each subject, we predicted all levels of MOAA/S scores using the best performing sleep model to assign probability score to each 30-second EEG epoch. By this way, we map discrete levels MOAA/S scores to continuous probability scores as shown in Figure 4A. As the probability score → 1, the subject enters into deep hypnotic state. We then
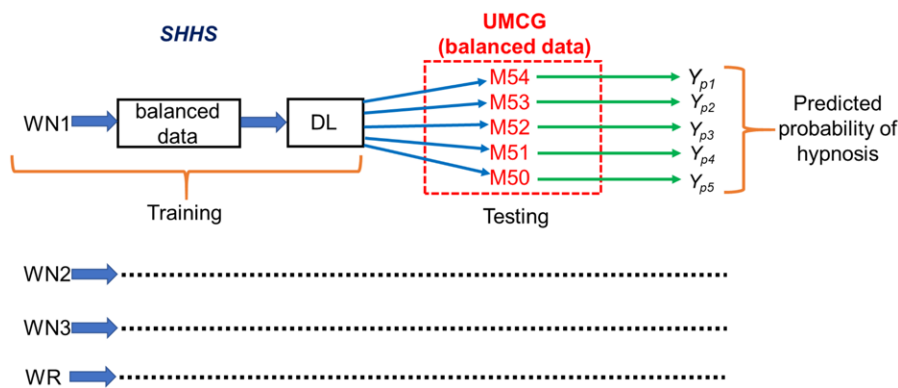


**Figure 3.** Illustration of the training testing experiment performed in this study. Because there are 4 sleep stages (N1, N2, N3, R) and a wake stage (W), we trained 4 separate DL models for binary classification: WN1, trained on W and N1; WN2, trained on W and N2; WN3, trained on W and N3; and WR, trained on W and 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 4 (M54), MOAA/S = 5 and 3 (M53), and so on until MOAA/S = 5 and 0 (M50) to estimate the probability of hypnosis $Y_p$. This process was repeated until all sleep stage DL models were used for predicting hypnosis levels. DL indicates deep learning; MOAA/S, Modified Observer's Assessment of Alertness/Sedation Scale; UMCG, University Medical Center Groningen.
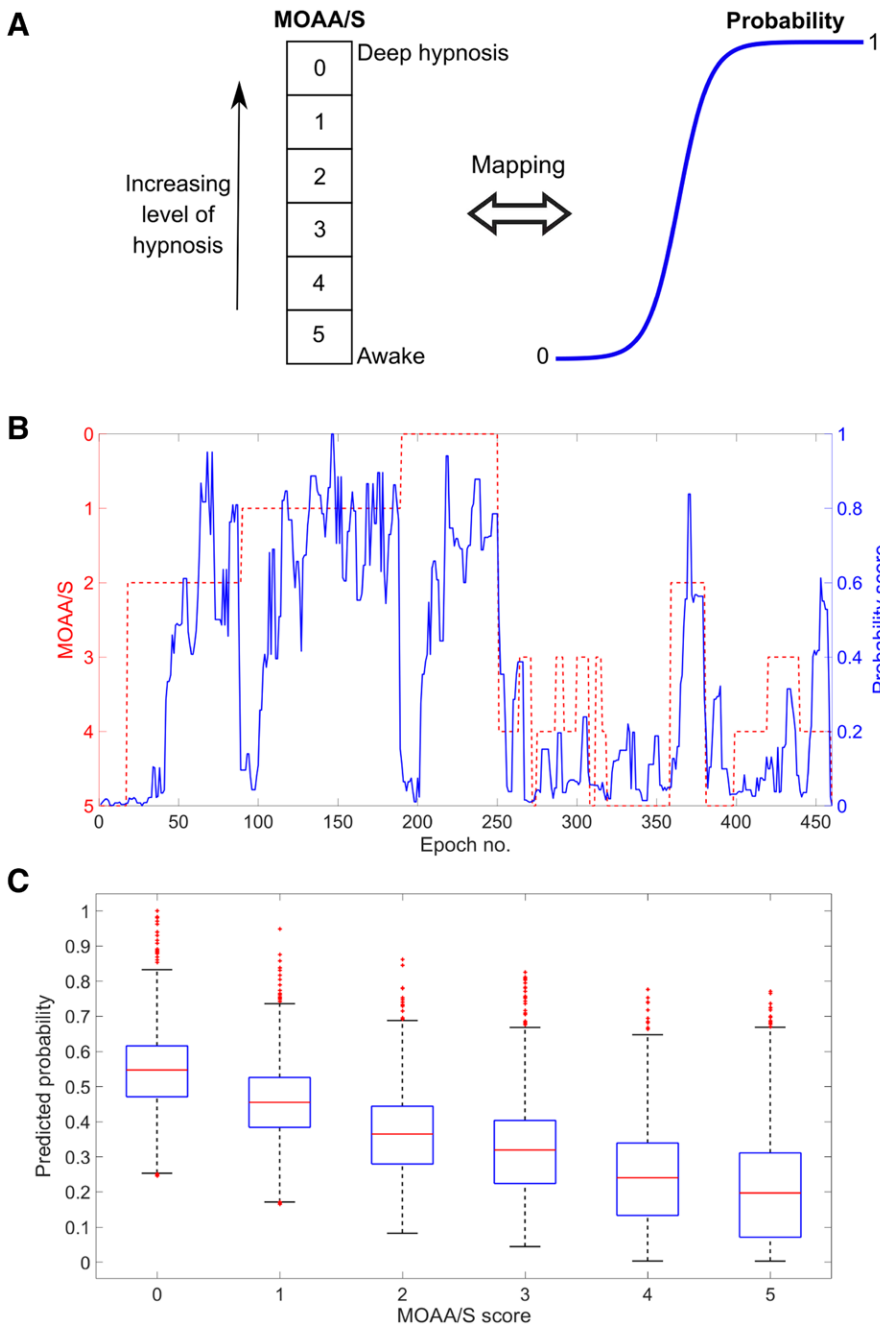
**Figure 4.** Hypnosis level prediction output. A, Illustration of mapping discrete MOAA/S score onto a continuous probability score via sigmoid transformation. Here probability score = 0 and 1 correspond to awake and deep hypnotic state, respectively. B, Illustration of correlation ($\rho$ = 0.53 in this example) between the probability score predicted by the DL model (blue) and MOAA/S scores (red), and (C) box plot comparing the distribution of predicted probability scores across all MOAA/S scores. Here the probability score is obtained by the WN3 LSTM-CNN model tested on all MOAA/S scores. The predicted probability score tends toward zero with increase in level of consciousness. Here the DL model is trained on wake and NREM stage 3 EEG segments and is used to predict all levels of MOAA/S scores (MOAA/S 0, 1, 2, 3, 4, 5) to obtain continuous levels of hypnosis. CNN indicates convolutional neural networks; DL, deep learning; EEG, electroencephalogram; LSTM, long short-term memory; MOAA/S, Modified Observer's Assessment of Alertness/Sedation Scale; NREM, nonrapid eye movement.

estimated a Spearman rank correlation ($\rho$) between different level of MOAA/S scores and WN3 model probability output.

### Spectrogram Analysis

To compare the performance of LSTM-CNN model with traditional spectrogram analysis, we estimated 5 spectral features from each 30-second EEG segment in the UMCG data set: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), spindle (12–16 Hz), and beta (16–30 Hz) power in decibel scale. Spectral estimation was performed using the Thompson multitaper spectral estimation method via chronux toolbox[27] with the

following parameters: length of the window $T$ = 4 seconds with 0.1-second shift (3.9 seconds overlap), time-bandwidth product time-bandwidth product (TW) = 3, number of tapers $K$ = 5, and spectral resolution 2 W of 1.5 Hz.

### Evaluation Metrics

We used the overall classification accuracy to evaluate the performance of the LSTM-CNN algorithm. We also report the area under the receiver operator characteristic curve (AUC). All results are reported as mean (95% confidence interval [CI]) unless otherwise stated. The

**Table. Performance (Accuracy [95% CI]) of the LSTM-CNN Model Trained on Individual Sleep Stages to Predict Different Levels of MOAA/S Scores**

| Testing Training | M54 | M53 | M52 | M51 | M50 |
|---|---|---|---|---|---|
| WN1 | 51.2 (45.3–54.4) | 46.5 (41.2–51.3) | 46.3 (41.5–52.4) | 40.4 (35.4–48.5) | 47.1 (41.2–53.3) |
| WN2 | 52.2 (46.1–57.2) | 53.7 (46.3–58.8) | 49.6 (41.5–55.4) | 57.6 (51.2– 61.5) | 57.1 (51.3–60.4) |
| WN3 | 56.4 (47.5–59.3) | 56.4 (48.6–60.5) | 59.7 (52.2–61.4) | 66.1 (59.7– 71.4) | 80.8 (79.2–88.3)[a] |
| WR | 53.4 (45.8–58.2) | 50.6 (42.4–56.3) | 50.3 (41.8–56.1) | 49.4 (42.5–55.7) | 64.8 (59.8–69.5) |

The WN3 model had the highest accuracy in predicting deep hypnotic state (MOAA/S = 0).

Abbreviations: CI, confidence interval; CNN, convolutional neural networks; MOAA/S, Modified Observer's Assessment of Alertness/Sedation Scale; LSTM, long short-term memory; WN1, model trained on wake (W) and N1 sleep state; WN2, trained on W and N2 sleep state; WN3, trained on W and N3 sleep state; WR, trained on W and rapid eye movement (R) sleep state; M54, model tested to discriminate between MOAA/S = 5 and 4; M53, model tested to discriminate between MOAA/S = 5 and 3; M52, model tested to discriminate between MOAA/S = 5 and 2; M51, model tested to discriminate between MOAA/S = 5 and 1; M50, model tested to discriminate between MOAA/S = 5 and 0.

[a]Highest performance obtained by the model.

95% CI was estimated using bootstrapping with 1000 samplings (BCa method) on the test data set.

## RESULTS

### Cross–Data Set Experiment

The Table summarizes the prediction performance of the LSTM-CNN model when trained on individual sleep states to predict different levels of MOAA/S scores. The LSTM-CNN WN3 model (trained on W and N3 stage) achieved an accuracy = 81 (79.2–88.3)%, AUC = 0.89 (0.82–0.94) in predicting dexmedetomidine deep hypnotic state (in channel C4/A1) much better than the random chance level accuracy of 50%. The LSTM-CNN model discriminated W and N3 during training with an accuracy of 98% and 95% in the training and validation set, respectively. The performance was poor for other models suggesting that dexmedetomidine deep hypnotic state is analogous to N3 sleep patterns. Similar performance was obtained in the secondary C3/A2 channel (accuracy = 81 [78.2–87.6]%, AUC = 0.88 [0.80–0.93]). Examples of EEG epochs and their corresponding predicted probabilities are shown in Supplemental Digital Content, Figure 2, http://links.lww.com/AA/D7. Supplemental Digital Content, Figure 3, http://links.lww.com/AA/D7, shows the confusion matrices for predicting M50 using individual sleep stages. Prediction performance for individual subjects is given in Supplemental Digital Content, Table 1, http://links.lww.com/AA/D7, and the performance on the raw data (without balancing the testing test) is summarized in Supplemental Digital Content, Table 2, http://links.lww.com/AA/D7.

### Internal Cross-Validation Within Each Data Set

To further evaluate the prediction performance of the LSTM-CNN model within each data set, we performed 5-fold cross-validation to discriminate between (1) W and N3 stage in the SHHS data set, and (2) awake (MOAA/S = 5) and deep hypnotic state (MOAA/S = 0) in the UMCG data set. The following performances were obtained: accuracy = 95.5 (91.2–99.4)%, AUC = 0.98 (0.91–0.99), and accuracy =

85.4 (79.3–89.6)%, AUC = 0.93 (0.89–0.96) for SHHS and UMCG data set (in channel C4/A1), respectively. Similarly, in the secondary channel (C3/A2), accuracy of 94.2 (90.8–99.1)% and AUC = 0.97 (0.90–0.99) in SHHS data set, and accuracy of 85.1 (78.5–88.7)% and AUC = 0.92 (0.87–0.95) in UMCG data sets were obtained. Because the training and testing were performed within UMCG data set during cross-validation, there was a 4% increase in the prediction accuracy in UMCG data set when compared to the cross–data set prediction accuracy (85% vs 81%). However, this increase in accuracy was not significant (P = .764).

### Continuous Hypnosis Level Estimation

Next, using WN3 model that was trained only using wake and N3 sleep stages, we predicted all MOAA/S scores for each subject. This resulted in a mean $\rho = 0.40$ (0.34–0.78), suggesting that the proposed method can be useful in developing continuous hypnotic level prediction system. Intermediate probability scores will provide an estimate of the deep hypnotic level of a subject. An example illustrating this is shown in Figure 4B. Here, $Y_p = 0$ indicates awake state (MOAA/S = 5) and $Y_p = 1$ indicates deep hypnotic state. $Y_p = 0.6$ indicates that the probability of patient being in deep hypnotic state is 0.6 and the drug infusion should be increased to increase the level of hypnosis (or reach MOSS/S score 0). The distribution of all predicted probability scores versus MOAA/S scores is shown in Figure 4C. With a decrease in the level of hypnosis (or increasing MOAA/S scores), the predicted probability score tends toward zero. Though promising, the proposed mapping method needs to be further validated/tested in another external data set.

### Comparison With Spectral Analysis

To evaluate the performance of individual spectral features, we performed a binary classification between 2 extreme levels of hypnosis: MOAA/S = 5 and MOAA/S = 0. The following prediction accuracies were obtained using individual spectral features: delta power = 54.4 (51.3–58.4)%, theta power = 50.6 (45.2–55.3)%, alpha power = 51.3 (44.2–57.5)%, spindle

power = 50 (47.2–54.1)%, and beta power = 52.7 (41.5–61.4)%. When all spectral features were used together in the traditional linear discriminant analysis, support vector machine (linear kernel, box constraint = 1) and random forest (100 trees) models to predict deep hypnosis, the system achieved an overall accuracy of 61.2 (55.3–63.2), 70.5 (65.8–74.4), and 72.8 (67.2–78.3)%, respectively. This suggests that the traditional spectral analysis alone is not suitable to predict deep hypnosis during dexmedetomidine infusion.

## DISCUSSION

Our study provides a novel data-repurposing framework using DL and large-scale EEG data to track hypnotic levels from sleep brain rhythms. The LSTM-CNN model predicted a deep hypnotic state with accuracy >80% when trained on the publicly available SHHS sleep data set and tested on the independent UMCG dexmedetomidine clinical trial data set. We also demonstrate using the DL algorithm that EEG patterns in dexmedetomidine-induced deep hypnotic state mimic NREM sleep stage 3 EEG patterns. To the best of our knowledge, this is the first study to explore the potential of DL algorithms to predict hypnotic levels using sleep brain rhythms.

The classical approach to developing EEG-based hypnosis level tracking systems is to extract information from frontal EEG channels mounted on the forehead to capture dynamic changes in the EEG oscillations at different level of hypnosis. This requires expensive clinical trials to record and analyze EEG data, develop techniques to monitor hypnotic levels for each drug class. Another major limitation with such techniques is that they are dependent on feature engineering and several potential discriminative features may not be included in the analysis. DL algorithms do not require any prior hand-crafted features and can learn potential discriminative features directly from the raw data. Our results suggest that DL algorithms, when trained on a sleep data set, can predict the hypnotic level and obtain nearly similar performance when trained on a dexmedetomidine data set (81% vs 85%, $P = .74$), eliminating the need for clinical trials to develop hypnotic level monitors.

Several previous studies using traditional spectrogram analysis have shown that dexmedetomidine hypnotic EEG patterns are characterized by slow oscillations in the slow-delta band (0–4 Hz) and spindle-like activities in spindle band (12–16 Hz), similar to NREM sleep EEG patterns. Though it is evident that dexmedetomidine hypnotic EEG mimics NREM sleep EEG, it was unclear which NREM sleep stage (N2 or N3) is homologous with a deep hypnotic state. Oto et al[28] demonstrated that nighttime infusion of dexmedetomidine-induced hypnosis is synonymous with N2 sleep stage in 10 mechanically ventilated intensive

care unit (ICU) patients. A study by Alexopoulou et al[29] also demonstrated that dexmedetomidine infusion increases N2 sleep stage in 13 ICU patients. In both these studies, continuous infusion of dexmedetomidine was given targeting a light hypnosis level (Richmond Agitation-Sedation Scale between −1 and −4). A recent study by Akeju et al[11] demonstrated that dexmedetomidine infusion significantly increased N3 sleep stage in a dose-dependent manner when compared to natural sleep in 10 healthy volunteers. Though intrasubject variability in these EEG patterns is minimal, there is considerable intersubject variability (for both sleep and dexmedetomidine) due to factors such as sex,[30] age,[31,32] or genetic factors[33,34]; an example is shown in Figure 5. Using large-scale EEG data and DL, we demonstrate that dexmedetomidine-induced deep hypnotic level is synonymous to N3 sleep stage. This
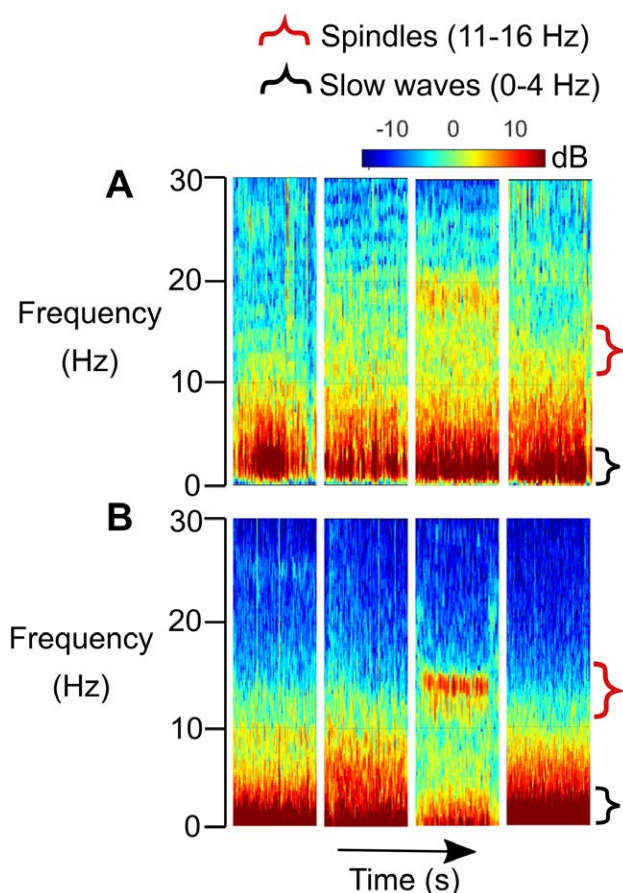


**Figure 5.** Spectrogram comparison of deep hypnosis and N3 sleep stage. Comparison of 5-min EEG power spectrogram from 4 subjects during (A) N3 sleep state in SHHS and (B) dexmedetomidine deep hypnotic state in UMCG. We can clearly see large variability in the slow-wave delta band (0–4 Hz) and spindle band (11–16 Hz) across subjects in both SHHS and UMCG data set. The following values were set to perform spectral estimation using multitaper spectral estimation via the chronux toolbox: length of the window $T = 4$ s with 0.1 s shift, time-bandwidth product TW = 3, number of tapers $K = 5$, and spectral resolution 2 W of 1.5 Hz. EEG indicates electroencephalogram; SHHS, Sleep Heart Health Study; TW, time-bandwidth product; UMCG, University Medical Center Groningen.

kind of external validation, as proposed in this study, is important to capture heterogeneity commonly seen in EEG recordings.

It should be noted that though the DL model was trained on SHHS data set and later used to predict hypnosis level on the UMCG data set, the proposed data-repurposing framework should not be confused with the typical transfer learning problem. In transfer learning, the pretrained model from data set A is used as a starting model, retrained on data set B to perform a prediction task within data set B. However, in the proposed data-repurposing approach, we used existing data set (SHHS) that is used to answer clinical questions in 1 domain (in this case sleep staging) to answer clinical questions in another domain (hypnosis level prediction) on a different data set (UMCG). The DL algorithm was trained from scratch using the SHHS data set and is completely different from transfer learning. However, any model developed for 1-dimensional (1D) physiological signal classification can be used for this application. Because different platforms are used to develop DL models (keras, python versions, architecture selection), it is difficult and requires substantial time and effort to implement. Because this was out of the scope of the current study, we did not perform transfer learning.

An automated approach to monitoring dexmedetomidine as proposed in this study is presumably well suited for patients in ICUs. These patients have comorbid conditions that, in principle, will significantly affect their sleep cycles which influence their EEG dynamics as a function of time. By training the DL model on large heterogeneous sleep EEG data capturing dynamic variations in the time-frequency properties of the EEG signal, it is possible to monitor deep hypnotic levels in the ICU. To implement the proposed framework in clinical settings as a patient independent system, we first train the DL model on W and N3 EEG segments from all available sleep data. The raw EEG signal from a new patient will be used as an input to this trained model which will provide a continuous probability of being either conscious or deeply hypnotized once every 30 seconds. This framework can also be used as a patient-specific (or personalized) hypnosis level monitoring system where the model is retrained repeatedly with new incoming 30-second EEG segments for initial few hours and then calibrate it for the underlying patient using reinforcement learning. By this way, the EEG of hypnosis monitoring will be based on the dynamic changes in the EEG that adaptively update the DL model specific to the underlying patient.

Imbalanced data can severely bias the model prediction results during both training and testing.[35,36] In our study, we balanced both training and testing data for 2 reasons: (1) straightforward interpretation of the model performance when compared with a random chance level accuracy (50%) and (2) consistent metric during both training and testing. Since we used all epochs corresponding to hypnosis (MOSS/S scores 4, 3, 2, 1) and random epochs corresponding to awake state (MOAA/S score 0), the model takes into account both inter- and intrasubject variability of EEG patterns.

Though results obtained in this study are promising, several limitations need to be addressed in the future study. First, we only used 2 EEG channels (C4/A1 and C3/A2) since the SHHS data set only included these 2 channels. Investigating hypnotic effects on other regions of the brain can reveal new insights about the anesthetic hypnosis mechanism. Second, we used dexmedetomidine data set from healthy volunteers and the results obtained should be validated in EEG recordings from patients in the ICU or undergoing surgery. Third, we only performed a hypnotic level prediction using dexmedetomidine as a prototype drug. Further validation is required to test this hypothesis and, as a future study, we will assess the performance of the system in other hypnotic drugs. Fourth, several epochs were misclassified (Supplemental Digital Content, Figure 3, http://links.lww.com/AA/D7) and we could not achieve a perfect prediction (100%). Because this is a proof-of-concept study, we did not perform rigorous model selection for best prediction performance and the current model is not yet ready for clinical deployment to predict individual patient's sedation level. An ideal system should accurately predict awake and hypnotic state and we believe that with more data and complex DL models, it is possible to develop such system.

To summarize, we provide a novel data-repurposing framework to predict anesthetic drug-induced hypnotic levels using sleep EEG data, which can be useful in developing hypnosis level monitoring systems. We also show using a data-driven approach that dexmedetomidine-induced deep hypnotic state mimics NREM sleep stage 3 and demonstrates the feasibility of DL algorithms to validate and verify the robustness of clinical hypothesis using large-scale EEG data instead of visual assessments using traditional EEG spectrogram. We also demonstrate that the DL model developed from archived cases ("training data") generally allows reliable monitoring of hypnosis levels in new patients whose data were not included during the training process, thus the system can be used "out of the box." ■■

## DISCLOSURES

## REFERENCES

1. Sheahan CG, Mathews DM. Monitoring and delivery of sedation. *Br J Anaesth*. 2014;113(suppl 2):ii37–ii47.
2. Bibian S, Dumont GA, Zikov T. Dynamic behavior of BIS, M-entropy and neuroSENSE brain function monitors. *J Clin Monit Comput*. 2011;25:81–87.
3. Li TN, Li Y. Depth of anaesthesia monitors and the latest algorithms. *Asian Pac J Trop Med*. 2014;7:429–437.
4. Bresson J, Gayat E, Agrawal G, et al. A randomized controlled trial comparison of NeuroSENSE and bispectral brain monitors during propofol-based versus sevoflurane-based general anesthesia. *Anesth Analg*. 2015;121:1194–1201.
5. Biswal S, Sun H, Goparaju B, Westover MB, Sun J, Bianchi MT. Expert-level sleep scoring with deep neural networks. *J Am Med Inform Assoc*. 2018;25:1643–1650.
6. Biswal S, Kulas J, Sun H, et al. SLEEPNET: automated sleep staging system via deep learning. ArXiv Prepr ArXiv170708262 2017.
7. Supratak A, Dong H, Wu C, Guo Y. DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Trans Neural Syst Rehabil Eng*. 2017;25:1998–2008.
8. Brown EN, Lydic R, Schiff ND. General anesthesia, sleep, and coma. *N Engl J Med*. 2010;363:2638–2650.
9. Murphy M, Bruno MA, Riedner BA, et al. Propofol anesthesia and sleep: a high-density EEG study. *Sleep*. 2011;34:283–91A.
10. Akeju O, Pavone KJ, Westover MB, et al. A comparison of propofol- and dexmedetomidine-induced electroencephalogram dynamics using spectral and coherence analysis. *Anesthesiology*. 2014;121:978–989.
11. Akeju O, Hobbs LE, Gao L, et al. Dexmedetomidine promotes biomimetic non-rapid eye movement stage 3 sleep in humans: a pilot study. *Clin Neurophysiol*. 2018;129:69–78.
12. Huupponen E, Maksimow A, Lapinlampi P, et al. Electroencephalogram spindle activity during dexmedetomidine sedation and physiological sleep. *Acta Anaesthesiol Scand*. 2008;52:289–294.
13. Akeju O, Kim SE, Vazquez R, et al. Spatiotemporal dynamics of dexmedetomidine-induced electroencephalogram oscillations. *PLoS One*. 2016;11:e0163431.
14. Lee HC, Ryu HG, Chung EJ, Jung CW. Prediction of bispectral index during target-controlled infusion of propofol and remifentanil: a deep learning approach. *Anesthesiol J Am Soc Anesthesiol*. 2018;128:492–501.
15. Sun H, Nagaraj SB, Akeju O, Purdon PL, Westover BM. Brain Monitoring of sedation in the intensive care unit using a recurrent neural network. *Conf Proc IEEE Eng Med Biol Soc*. 2018;2018:1–4.
16. Sun H, Nagaraj SB, Westover MB. Predicting Ordinal Level of Sedation from the Spectrogram of Electroencephalography. In: 2018 International Conference on Cyberworlds (CW). IEEE, 2018:292–295.
17. Dean DA II, Goldberger AL, Mueller R, et al. Scaling up scientific discovery in sleep medicine: the national sleep research resource. *Sleep*. 2016;39:1151–1164.
18. Zhang GQ, Cui L, Mueller R, et al. The national sleep research resource: towards a sleep data commons. *J Am Med Inform Assoc*. 2018;25:1351–1358.
19. Quan SF, Howard BV, Iber C, et al. The Sleep Heart Health Study: design, rationale, and methods. *Sleep*. 1997;20:1077–1085.
20. Redline S, Sanders MH, Lind BK, et al. Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. Sleep Heart Health Research Group. *Sleep*. 1998;21:759–767.
21. Weerink MAS, Barends CRM, Muskiet ERR, et al. Pharmacodynamic interaction of remifentanil and dexmedetomidine on depth of sedation and tolerance of laryngoscopy. *Anesthesiology*. 2019;131:1004–1017.
22. Chernik DA, Gillings D, Laine H, et al. Validity and reliability of the Observer's Assessment of Alertness/Sedation Scale: study with intravenous midazolam. *J Clin Psychopharmacol*. 1990;10:244–251.
23. Hori T, Sugita Y, Koga E, et al; Sleep Computing Committee of the Japanese Society of Sleep Research Society. Proposed supplements and amendments to 'A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects', the Rechtschaffen & Kales (1968) standard. *Psychiatry Clin Neurosci*. 2001;55:305–310.
24. Berry RB, Brooks R, Gamaldo CE, Harding SM, Marcus CL, Vaughn BV. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, version 2.0. Darien, IL: American Academy of Sleep Medicine; 2012.
25. Colin PJ, Hannivoort LN, Eleveld DJ, et al. Dexmedetomidine pharmacodynamics in healthy volunteers: 2. Haemodynamic profile. *Br J Anaesth*. 2017;119:211–220.
26. Weerink MAS, Struys MMRF, Hannivoort LN, Barends CRM, Absalom AR, Colin P. Clinical Pharmacokinetics and pharmacodynamics of dexmedetomidine. *Clin Pharmacokinet*. 2017;56:893–913.
27. Bokil H, Andrews P, Kulkarni JE, Mehta S, Mitra PP. Chronux: a platform for analyzing neural signals. *J Neurosci Methods*. 2010;192:146–151.
28. Oto J, Yamamoto K, Koike S, Onodera M, Imanaka H, Nishimura M. Sleep quality of mechanically ventilated patients sedated with dexmedetomidine. *Intensive Care Med*. 2012;38:1982–1989.
29. Alexopoulou C, Kondili E, Diamantaki E, et al. Effects of dexmedetomidine on sleep quality in critically ill patients: a pilot study. *Anesthesiology*. 2014;121:801–807.
30. Genzel L, Kiefer T, Renner, L et al. Sex and modulatory menstrual cycle effects on sleep related memory consolidation. *Psychoneuroendocrinology*. 2012;37:987–998.

31. Campbell IG, Feinberg I. Maturational patterns of sigma frequency power across childhood and adolescence: a Longitudinal Study. *Sleep*. 2016;39:193–201.

32. Sprecher KE, Riedner BA, Smith RF, Tononi G, Davidson RJ, Benca RM. High resolution topography of age-related changes in non-rapid eye movement sleep electroencephalography. *PLoS One*. 2016;11:e0149770.

33. De Gennaro L, Marzano C, Fratello, F et al. The electroencephalographic fingerprint of sleep is genetically determined: a twin study. *Ann Neurol*. 2008;64:455–460.

34. Adamczyk M, Genzel L, Dresler M, Steiger A, Friess E. Automatic sleep spindle detection and genetic influence estimation using continuous wavelet transform. *Front Hum Neurosci*. 2015;9:624.

35. Chawla NV. Data mining for imbalanced datasets: An overview. In: *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer; 2009:875–886.

36. Wei Q, Dunbrack RL Jr. The role of balanced training and testing data sets for binary classifiers in bioinformatics. *PLoS One*. 2013;8:e67863.