HHS Public Access

Author manuscript

IEEE Trans Neural Syst Rehabil Eng. Author manuscript; available in PMC 2019 May 31.

Published in final edited form as:

IEEE Trans Neural Syst Rehabil Eng. 2019 February ; 27(2): 118–128. doi:10.1109/TNSRE.2018.2890272.

Automatic Sleep Stage Classification Based on Subthalamic Local Field Potentials

Yue Chen,
National Engineering Laboratory for Neuromodulation, School of Aerospace Engineering, Tsinghua University, Beijing 100084, China.

Chen Gong,
National Engineering Laboratory for Neuromodulation, School of Aerospace Engineering, Tsinghua University, Beijing 100084, China.

Hongwei Hao,
National Engineering Laboratory for Neuromodulation, School of Aerospace Engineering, Tsinghua University, Beijing 100084, China.

Yi Guo,
Department of Neurosurgery, Peking Union Medical College Hospital, Beijing 100032, China.

Shujun Xu,
Department of Neurosurgery, Qilu Hospital of Shandong University, Shandong 250012, China.

Yuhuan Zhang,
Department of Otolaryngology, Head and Neck Surgery, Beijing Tsinghua Changgung Hospital, Beijing 102218, China.

Guoping Yin,
Department of Otolaryngology, Head and Neck Surgery, Beijing Tsinghua Changgung Hospital, Beijing 102218, China.

Xin Cao,
Department of Otolaryngology, Head and Neck Surgery, Beijing Tsinghua Changgung Hospital, Beijing 102218, China.

Anchao Yang,
Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China.

Fangang Meng,
Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China.

Jingying Ye,
Department of Otolaryngology, Head and Neck Surgery, Beijing Tsinghua Changgung Hospital, Beijing 102218, China.

Corresponding authors: Yanan Sui ysui@cs.stanford.edu; Luming Li lilm@mail.tsinghua.edu.cn.

Hesheng Liu,
Athinoula A. Martinos Center for Biomedical Imaging, Harvard Medical School, MA 02129, USA.

Jianguo Zhang,
Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China.

Yanan Sui, and
Computer Science Department, Stanford University, Stanford, CA 94305 USA.

Luming Li
National Engineering Laboratory for Neuromodulation, School of Aerospace Engineering, Tsinghua University, Beijing 100084, China, also with the School of Aerospace Engineering, Man-Machine- Environment Engineering Institute, Tsinghua University, Beijing 100084, China, also with the Precision Medicine & Healthcare Research Center, Tsinghua-Berkeley Shenzhen Institute, Shenzhen 518071, China, and also with the Center of Epilepsy, Beijing Institute for Brain Disorders, Beijing 100069, China.

Abstract

Deep brain stimulation (DBS) is an established treatment for patients with Parkinson's disease (PD). Sleep disorders are common complications of PD and affected by subthalamic DBS treatment. To achieve more precise neuromodulation, chronic sleep monitoring and closed-loop DBS toward sleep-wake cycles could potentially be utilized. Local field potential (LFP) signals that are sensed by the DBS electrode could be processed as primary feedback signals. This is the first study to systematically investigate the sleep-stage classification based on LFPs in subthalamic nucleus (STN). With our newly developed recording and transmission system, STN-LFPs were collected from 12 PD patients during wakefulness and nocturnal polysomnography sleep monitoring at one month after DBS implantation. Automatic sleep-stage classification models were built with robust and interpretable machine learning methods (support vector machine and decision tree). The accuracy, sensitivity, selectivity, and specificity of the classification reached high values (above 90% at most measures) at group and individual levels. Features extracted in alpha (8–13 Hz), beta (13–35 Hz), and gamma (35–50 Hz) bands were found to contribute the most to the classification. These results will directly guide the engineering development of implantable sleep monitoring and closed-loop DBS and pave the way for a better understanding of the STN-LFP sleep patterns.

Keywords

Deep brain stimulation (DBS); local field potential (LFP); sleep; classification; Parkinson's disease

I. Introduction

DEEP brain stimulation (DBS) has become a well-established surgical treatment for Parkinson's disease (PD) [1]. However, the current widely used continuous high frequency stimulation (HF-DBS) is not sufficient to meet growing clinical demands of more precise

neuromodulation to improve clinical outcomes. Sleep disorders are commonly observed in PD patients [2], [3]. Although the general quality of sleep is reported to be improved, the specific role of subthalamic nucleus (STN) HF-DBS in sleep is being questioned [4], [5]. Rapid eye movement (REM) sleep behavior disorder, restless legs syndrome, and periodic limbs movements are often found unimproved or even deteriorated following STN-DBS [4], [6], [7]. Recent studies have shown that normal movement control was restored to some extent in PD patients during REM sleep without stimulation, suggesting that HF-DBS may not be necessary during the whole night for efficient sleep [8], [9]. Given that sleep disorders are common among PD patients, chronic sleep monitoring would be useful and closed-loop DBS strategy is a potential solution for precisely controlling exacerbated sleep disorders, without sacrificing treatment efficacy for motor symptoms.

Local field potential (LFP) is the summed synchronous activities in a population of neurons, which can be detected by the contacts from the DBS leads. In this work, STN-LFP was studied for automatic sleep stage classification in PD patients with DBS. LFP recordings were taken instead of traditional electroencephalography (EEG) in sleep considering the following aspects:

- Motivation 1: LFPs-based methods have a better integrated capacity for a fully implanted closed-loop DBS system, as shown in Fig. 1. In a previous study, we developed an implantable deep brain stimulator with the ability to record LFPs using the DBS electrode [17]. Compared with the traditional EEG-based methods, LFP-based methods are more practical in reducing the engineering price of an additional sensing device, data transmission, and data synchronization.
- Motivation 2: The LFP-based methods have better practicality for the chronic daily sleep-monitoring for DBS patients. The current advanced DBS system provides a unique opportunity to investigate LFP sleep patterns without influence of surgical interferences including edema, infection, and micro-lesion in the acute surgery period [18], [19]. Validating LFP as a solid biomarker of sleep stages is of great potential for expanding the healthcare for DBS patients. Compared with non-invasive EEG, LFP-based methods will not need any additional sensors to achieve sleep monitoring.
- Motivation 3: LFP recorded in the brain target might be a better decoder for the clinical state during sleep. In previous studies, LFP has been reported as an indicator for the motor symptoms of PD patients and DBS has been reported as a modulator of LFP activities in the target [11], [12]. It has been recognized as the first choice for feedback signals in the current investigation of closed-loop DBS [10]. It is foreseeable that in future closed-loop DBS studies, LFPs would be of great use to decode the real-time clinical and stimulation effects during sleep.

Finding solid algorithms for sleep stage classification is the next key. Whereas most studies have focused on the LFP and motor function in the daytime, some have reported LFP patterns during sleep [13], [14]. However, the LFP patterns were only observed during the acute surgical periods with externalized extension leads and amplifiers, which need to be verified with long-term recording. More importantly, whether it could be a solid biomarker

for the classification of sleep stage are still largely unknown. Although it has been suggested there is an opportunity for taking LFP spectrums as an indicator for sleep sleep-stage transitions [14], little is known about the design of features, methods and the robustness of the result. Creating an effective feedback for implantable closed-loop DBS remains a big challenge from both clinical and engineering perspectives [15], [16].

As we know, LFPs in deep brain are coherent with the neural activities in cortical regions [20]. Therefore, we have learned from the studies which have worked on the automatic classification of sleep stages from EEG, electrocorticography (ECoG), and intracranial EEG (iEEG) signals [21]–[25]. Classification studies involving the support vector machines (SVM), decision trees (DT) and other machine learning methods pave the way for building a new solid and interpretable classifier based on LFPs from STN.

In this study, subthalamic LFPs were recorded by implanted neurostimulators at 1 month after DBS implantation, which is the start of long-term DBS. We built machine learning classifiers which are potentially promising for long-term usage. Our machine learning methods are robust, interpretable, and achieved high performance at both individual and group levels with cross-validation. Model interpretability was in turn used to improve the understanding of STN-LFP sleep patterns. To our knowledge, this is the first systematic study to realize classification of awake and sleep states based on STN-LFPs. The results will directly guide the development of the long-term closed-loop DBS application in the future.

II. Methods and Materials

A. LFP Acquisition

1) Subjects—This study was approved and registered at ClinicalTrials.gov (Identifier NCT02937727). A total of 13 PD patients were involved in this study. One patient was excluded due to the abnormal impedance of the contacts (> 100KΩ), suspected to be caused from a break in the extension. The inclusion criteria were determined based on the main indication for STN-DBS and the ability of coordination during preoperative tests. Written informed consents were provided by all subjects. Experimental procedures were approved by the ethical committees of the surgical hospitals (Beijing Tiantan Hospital of Capital Medical University, Peking Union Medical College Hospital, and Qilu Hospital of Shandong University). All of the subjects were implanted with bilateral STN-DBS (G106R, PINS, China). The electrode consisted of four platinum-iridium cylindrical contacts (Model L301C, PINS, China, 1.3 mm diameter and 1.5 mm length spaced 0.5 mm apart). Locations of the electrodes were confirmed by the intraoperative frame-based stereotaxy and micro electrode recording, and postoperative DBS improvement. Demographic information of the 12 subjects is listed in Table I.

2) Experimental Protocol—Experiments were conducted at night before the initial activation of DBS, 1 month after surgery. The acute effects of surgery were examined by impedance measures and MRI scan results. The primary objective of this study was to evaluate whether intrinsic STN-LFP sleep patterns could be used to systematically classify awake and sleep stages. Therefore, to minimize medication effects, subjects were withdrawn from long-acting dopaminergic medication for at least 24 hours and were withdrawn from

short-acting dopaminergic medication for at least 8 hours. STN-LFPs in the wakeful rest state were recorded for 6 minutes before sleep. STN-LFPs in the sleep state were recorded approximately 4–6 hours at night, depending on sleep periods and battery life of the stimulator.

Polysomnography (PSG) monitoring were conducted during the whole night. The sleep stages were justified every 30s according to the manual guideline from the American Academy of Sleep Medicine (AASM) [26], including non-REM stage 1 (N1), non-REM stage 2 (N2), and non-REM stage 3 (N3), and REM. The montages of EEG (F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, O2-M1), EOG (E1-M2, E2-M2), and EMG placed on the chin were mainly used for sleep staging. Final sleep stages were scored by one expert and validated by another. The experiments were conducted in the Sleep Medicine Center of Beijing Tsinghua Changgung Hospital.

3) LFP Recording Instrumentation—The LFP recording setup is illustrated in Fig.2. The advanced deep brain stimulator with LFP recording function was used [17]. Eight channels (24-bit resolution) differentiated between pairs of contacts in the electrodes were wirelessly and synchronically transmitted to the recording platform outside the body through the radio frequency (RF) modules. The transmission rate is 250 kbps and the delay is less than 10ms. Wireless communication distance from the stimulator to the RF modules is about 2m. The stimulator could continuously sample and transmit LFPs and the chargable battery life is 10 years. Before sampling, the analog LFPs were preprocessed by a built-in 0.3 Hz highpass filter and a 250 Hz low-pass filter to cut down the baseline drift and the aliasing harmonics in the high frequency range. The sampling rate was 500 Hz.

As shown in Fig.2, wireless data transmission for all body positions during sleep were guaranteed by four RF modules (403 MHz medical radio frequency carrier). The recording platform PC was connected to the LFP monitoring PC in the central control room. A PSG system (GRAEL™, Australia or SOMNOscreen™, Germany) was equipped to record all the necessary signals for sleep staging. To avoid interrupting the sleep of the subjects, all devices in the PSG monitoring room were operated remotely.

B. Data Preprocessing

Data from the 12 PD subjects were all included for the analysis. LFP recordings in the right STN of Case 1 was interrupted by system artifacts. Therefore, data from 23 STNs were included in the study.

LFP recordings from the pair of contacts adjacent to the therapeutic cathode were chosen for analysis. The time of PSG monitoring system and LFP storage on the computer was manually synchronized with an error of less than 1 s. Aligned to the PSG recording, LFPs were marked by the edge of each sleep stage. The data were then processed by a 0.5–50 Hz band-pass filter and segmented as 30s epochs. Mean time sequence value of each 30s epoch was subtracted. To exclude the synchronization error and sleep state transition periods, for the sleep stage N1, the data selected for analysis were longer than 60s and the first 30s epoch was excluded. For the sleep stage N2, N3, and REM, data selected for analysis were longer than 150s and the two 30s epochs at both ends were excluded. Any epochs that contained

movement or system artifacts were manually excluded. The number of the prepared epochs are listed in Table I. Following this, the data was segmented to 1s lengths.

MATLAB (version 2016a) were used for calculations. Spectrogram was calculated using a Short-Time Fourier Transform (STFT) with a Hamming window of 1s and an overlap of 50%. Power spectrum density (PSD) of each 1s epoch in the data set were generated by the Welch's method with a Hamming window of 0.2 s and an overlap of 75%. The range of the frequency analysis was unified to 0–50 Hz (Filter stop band below 0.5 Hz was ignored).

C. Feature Selection

We extracted different combinations of features from time domain, frequency domain, and entropies.

In time domain, three types of features were examined. The first feature was the mean absolute error (MAE). MAE was calculated as the averaged distance between every two adjacent points in a time sequence. The value of MAE reflects the variance within each time sequence. The second feature was the root mean square, which reflects the averaged power of each time sequence. The third was a group of coefficients from the autoregressive model (AR model) fitted to the data. A fourth order AR model was calculated and the coefficients represented the autocorrelation of the points in a time sequence. Sample entropy was also considered as a possible feature.

In the frequency domain, PSD was extracted as a group of features. Power in the five typical bands (delta band, 1–4 Hz; theta band, 4–8 Hz; alpha band, 8–13 Hz; beta band, 13–35 Hz; part of gamma band, 35–50 Hz) was tested. To increase the effective amount of information for classifiers, refined densities of the frequency power calculation were tested. The width of the bands was scaled from 1 Hz, 2 Hz, 4 Hz to inhomogeneous width.

For each group of the features, the discriminant analysis procedure [32] was repeated 50 times to evaluate the capacity of the individual group to differentiate among various sleep stages. The best single group is features in the frequency domain. In the next step, a forward selection procedure (FSP) was conducted [32]. The group of features in the time domain and the group of entropy were added respectively and the multi-class classification errors were evaluated. Features in frequency domain was subdivided with a step of 2 Hz. The procedure was repeated 50 times. The total error rate (percentage of incorrect epochs) and error rate in each sleep stage were calculated, showing in Fig.3. The results suggest that as the features in frequency domain increased, error rates significantly decreased. No significant difference was found from adding features in the groups of time domain and entropy. Finally, a total of 25 features in the range of 0–50 Hz were selected.

D. Classification and Prediction

1) Overall Process of Classification and Prediction—The overall process is illustrated in Fig.4. Features were extracted from the 1s data segments from each class. The 2 Hz bandwidth power of PSD and sleep stage labels constituted the samples. 90% in each class were randomly selected for training and the remaining 10% in each class were for testing. A running average rule was applied to generate the final decisions.

To evaluate the method in a simulated online scene, we separately analyzed the performance in a test set which is completely isolated with the training set in time. The so-called "prediction" in the results means the data segments in the training set and test set are selected in time order. For prediction, the first 90% strictly in time domain of the raw data in each class were chosen for training. Thus, in each class, the remaining 10% were for testing.

N3 and N2 sleep recordings were merged in this procedure based on three considerations: (1) Based on our primary design of the closed-loop DBS system, both N2 and N3 stages represent a continuously stable non-REM sleep period during which the DBS strategy might be the same; (2) The LFP characteristics in N3 stage in frequency domain are similar with those in N2 stage, as shown in Fig.7(a); (3) The numbers of N3 LFP epochs are significantly unbalanced across all the subjects. To represent complete results, the subset from the three subjects with N3 stage recorded were additionally constructed and the same classification procedure was repeated.

We performed the overall process of classification and prediction on the binary combinations, including wakefulness against N1, wakefulness against N2, wakefulness against REM, N1 against N2, N1 against REM, N2 against REM, and wakefulness against sleep (N1, N2, and REM) and the multi-class together.

In each class, the training and testing sets were partially separated with a proportion of 9:1. Before both training and testing, samples from each class were randomly selected to build balanced training and testing sets. To verify the robustness, the whole procedure that started with feature extraction was repeated 50 times. Cross-validation method was used for the model optimization and training to unbiasedly estimate performance. Details of the process are described in the following sub sections.

2) Machine Learning Methods—We developed our models with two learning methods in this study, the support vector machine (SVM) and decision tree (DT). Both of these are widely used in the classification for a variety of physiological signals. A SVM model is a space construction for mapping the examples, aiming at separating the classes by a gap that is as wide as possible [27]. By properly choosing the kernel function, SVM is efficient for achieving both linear and non-linear classification. In this study, a kernelized SVM classification model with penalty factor C was employed. The optimization goal of our SVM model is:

$$\text{maximize } f\left(\alpha_1,\ldots,\alpha_n\right)=\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \times k\left(\boldsymbol{x_i},\boldsymbol{x_j}\right)$$

$$\textit{subject to } 0\le\alpha_i\le \mathrm{C} \;\; \textit{and} \;\; \sum_{i=1}^{n}\alpha_i y_i=0 \quad (1)$$

where α_i's are the Lagrange multipliers, m is the number of samples, $\boldsymbol{x}_i$'s and $\boldsymbol{x}_j$'s are the samples, y_i's and y_j's are the class labels of the samples, k represents the kernel function, C represents the penalty factor. The kernel function $k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ in our model is radial basis function (RBF):

$$k\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \mathrm{e}^{-\gamma \left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2} \quad (2)$$

where γ is the hyper-parameters of the RBF kernel. Output function of the SVM model is:

$$y = \sum_{h=1}^{m} \alpha_h y_h k\left(\boldsymbol{x}_h, \boldsymbol{x}\right) + b \quad (3)$$

where $\boldsymbol{x}_h$'s represent the vectors from the training set, y_h's represent the class labels of the vectors, and α_h's represent the Lagrange multipliers, $\boldsymbol{x}$ represents sample of a test point. For the support vectors, α_h is non-zero and for the rest vectors in training set, α_h is zero. The sign of y decides the classification result. The LibSVM toolbox [28] for MATLAB was used for calculation.

A decision tree is a group of conditional statements for improving the purity of each leaf. A DT model could provide a clear path for a classification [29]. We used a standard classification and regression tree (CART) model [31] to generate a binary tree for classification. All the predictors were continuous. Gini's diversity index was used to calculate the impurity of nodes in the decision tree. The Gini's index of each feature is:

$$G\left(k_i\right) = 1 - \sum_{j=1}^{2} p_j^2 \quad (4)$$

where $G(k_i)$ represents the Gini index of each node i split by feature k's value, p_j's represent the observed fraction of class j in the split nodes. The split information of feature k is calculated as Gini gain:

$$Gain\ (k) = \frac{N_1}{N} G\left(k_1\right) + \frac{N_2}{N} G\left(k_2\right) \quad (5)$$

where N_1 and N_2 represent the sizes of the split nodes, and N represents the size the node before split.

3) Optimization and Training—The hyper-parameters were optimized via grid search method. For the SVM model, cost C represents the weight of how many samples inside the margin contribute to the overall error. C was optimized in the range of 2^2–^{10}. For the RBF kernel function of SVM model, γ was optimized in the range of and 2^{-15}–2. For the DT

model, the minimum leaf number was optimized in the range of 1–100. Five-fold for SVM and ten-fold for DT cross-validations were performed. Models were trained with the hyper-parameters that achieved the highest classification accuracy in the training set.

4) Running Average—For the final step, a running average method was used to decrease the influence caused by the fluctuation of the signals in each second. For a voting length of N s, (N-1) samples which were time-adjacent with the sample at the test point were reorganized together as a voting group. A majority vote strategy was performed and the final decision generated from the voting group was recognized as the classifiers' output.

With a larger voting length, the composition of the general rhythm during this period will be highlighted and the independent influence of nonstationary randomness will be attenuated. In this work, the voting length N was scaled from 1 s (no voting), 3 s, 5 s, 7 s to 9 s to evaluate the performance and determine the trade-off between the classification accuracy and the time delay of decision making.

5) Performance Measure—For binary classification and prediction, the performance was measured by the overall accuracy, sensitivity, selectivity, and specificity. The overall accuracy was calculated by equation (6):

$$Accuracy = \frac{TN_1 + TN_2}{N_1 + N_2} \times 100\% \quad (6)$$

where N_1 and N_2 represent the numbers of samples in class 1 and class 2 from the test set. TN_1 and TN_2 represent the numbers of correct predictions of class 1 and class 2. The sensitivity was calculated by equation (7):

$$Sensitivity = \frac{TN_1}{N_1} \times 100\% \quad (7)$$

The specificity was calculated by equation (8):

$$Specificity = \frac{TN_2}{N_2} \times 100\% \quad (8)$$

The sensitivity and specificity, respectively, stand for the accuracy of class 1 and class 2.

The selectivity was calculated by equation (9):

$$Selectivity = \frac{TN_1}{N_1'} \times 100\% \quad (9)$$

where N_1' represents the number of class 1 in the predicted results. The selectivity stands for the ratio of actual class 1 instances in the predicted class 1 instances.

For multi-class classification and prediction, the performance was measured by confusion matrixes [32].

III. Results

In this study, STN-LFPs from 12 PD patients were post-operatively recorded and analyzed during sleep state. Firstly, we investigated the signal features in the frequency domain. Based on the features, SVM and DT models were built for classification of sleep stages and wakefulness. Finally, the model structure was analyzed.

A. STN-LFP Sleep Frequency Patterns

A piece of typical LFP data (Case7, left STN) for different sleep-wake stages is shown in Fig.5(a). In the time-frequency spectrogram, beta oscillation decreased in sleep stage N1 and almost disappeared in sleep stage N2. Once the REM sleep began, beta oscillation recovered immediately to a level similar to wakefulness. On the contrary, activities in the low frequency band (below 15 Hz) increased in N1 sleep and became stronger in N2 sleep, whereas attenuated in REM and wakefulness.

Fig.5(b) shows the averaged PSDs across subjects. The differences among sleep stages were located in both low frequency band (delta, theta, and alpha band) and high frequency band (beta and gamma band). Fig.5(c) further shows the power in each band. Paired Wilcoxon signed rank tests were conducted and the results were FDR corrected [30]. Delta, theta, and alpha bands significantly increased from wakefulness to N2 ($N=12$, $p<0.01$), and decreased in REM sleep ($p<0.05$). On the contrary, beta and gamma bands decreased from wakefulness to N2 ($p<0.01$) and recovered in the stage of REM ($p<0.05$).

The general rhythms of STN-LFP are clarified in Fig.5. However, from the shadow in Fig. 5(b) and error bars in Fig.5(c), the standard deviations of PSDs and band power during wakefulness or each sleep stage were very large. From Fig.5(a), it was found that the LFP rhythms were fluctuating and not always consistent in every second. Besides, from individual PSDs, there is a big difference in the spectrum power across subjects. Therefore, the PSDs and band power in different sleep stages were overlapped. This indicates that a simple threshold classification or an amplitude-responsive algorithm would not reach a satisfactory performance. Machine learning models are necessary to build a more general profile of the STN-LFP sleep patterns for the classification and prediction. A running average method are worthy to confront the fluctuation of the signal without sacrificing the responding speed.

B. Classification of Wakefulness and Sleep Stages

Personalized and unified (across all the subjects) classification models were developed. The voting length for the running average method was fixed to 5 s in this section. Numbers of epochs included for analysis were shown in Table II. Table III shows the averaged accuracies of both personalized and group level SVM and DT classification. Generally speaking, the

personalized models reached the performance above 90% and the unified models reached the performance above 80% except for N1 against N2 and REM. Best performance was achieved at wakefulness and N2 detection.

Based on the data from single subject, SVM and DT models were developed for personalized classification. Performance was high for the classification of wakefulness against N1, N2, REM. Accuracy of the classification of wakefulness against sleep (including N1, N2, and REM) was 97.4% (std=2.6%) for the SVM models and 94.7% (std=3.6%) for the DT models. The results in SVM models and DT models are consistent. Taking SVM models as an example, the sensitivity, selectivity, and specificity were in the range of 89.3% to 100%, except for N1 against REM (67.4%, 78.3%, and 81.4%).

Unified models were developed across subjects. The accuracy could reach above 94% for wakefulness against N2. Accuracies of the classifications of wakefulness against sleep were high (SVM, accuracy=90.2%, std=1.8%; DT, accuracy=88.5%, std=1.7%). The results in SVM models and DT models are consistent. Taking SVM models as an example, the sensitivity, selectivity, and specificity were in the range of 81.7% to 96.5%, except for sensitivity for N1 against N2 (70.0%), and all the measures for N1 against REM (76.6%, 68.9%, and 65.0%).

Fig.6 shows the multi-class classification results. In general, the accuracy, sensitivity, and selectivity were all above chance level ($p < 0.001$). Performance of the wakefulness and N2 were high in both unified and personalized models. N1 and REM were most difficult to be classified, which was consistent with the binary classification results. The general accuracies of unified classification are 65.6%(std=1.8%, SVM) and 61.4%(std=1.6%, DT). The general accuracies of personalized classification are 80.8%(std=3.6%, SVM) and 75.6%(std=4.4%, DT). The overall performance was improved by considering personalized models.

Based on the subset from the three subjects with independent N2 and N3 recordings, the same classification procedure was conducted. The accuracies of awake against N3, N1 against N3, and REM against N3 is above 85% for classification. The general SVM multi-class classification accuracy is 58.3% (std=3.1%). The results are shown in Fig.7(b) and are all above chance level($p < 0.001$). Note that N3 and N2 stages were mostly confused with each other. From Fig. 7(a), the N3 stage frequency features are very similar with N2 stage, therefore it is difficult for the classifiers to distinguish between the two stages.

C. Prediction of Wakefulness and Sleep Stages

The prediction means the training and testing sets are strictly time isolated. Personalized and unified (across all the subjects) classification models were respectively developed. The voting length was fixed to 5 s in this section. Numbers of epochs were equal with Table II. Table IV shows the averaged unified and personalized prediction accuracies. Generally speaking, the prediction for wakefulness against sleep stages reached performance above 84% at group level and 87% at personalized level. Best prediction performance was achieved at wakefulness and N2 detection.

For the unified models, the prediction performance of wakefulness against N1, N2, REM, and sleep was higher than the other groups. Wakefulness against N2 was best predicted and the accuracy was above 93%. The results across all prediction groups in SVM models and DT models are consistent. Taking SVM models as an example, the sensitivity was above 80%, except for wakefulness against REM (78.0%) and N1 against N2 (68.1%). The selectivity was above 91%, except for N1 against N2 (79.7%), N1 against REM (58.6%), and N2 against REM (66.6%). The specificity was above 92%, except for N1 against N2 (82.5%), N1 against REM (44.0%), and N2 against REM (56.7%).

For the personalized models, the accuracies of prediction were generally improved. Prediction of wakefulness against N2 was much higher for both SVM models (accuracy= 100.0%, std=0) and DT (accuracy=99.9%, std=0.5%). For wakefulness against sleep, performance reached an accuracy of 95.4% (std=2.4%) for SVM models and 93.1% (std=3.8%) for DT models. The results in SVM models and DT models are consistent. Taking SVM models as an example, the sensitivity, selectivity, and specificity were in the range of 85.6% to 100%, except for N1 against REM (52.3%, 76.5%, and 83.7%, respectively).

Fig.8 shows the confusion matrix of multi-class prediction results. In general, the accuracy, sensitivity, and selectivity were all above chance level ($p<0.001$). Wakefulness and N2 were best predicted. For unified models, both of these two stages were mostly confused with N1. The general accuracies of unified prediction are 58.8%(std=1.1%, SVM) and 57.1% (std=1.0%, DT). The general accuracies of personalized prediction are 76.3%(std=2.2%, SVM) and 73.7%(std=3.3%, DT). The personalized model could improve the overall performance of multi-class prediction.

D. Zooming of Running Average

Voting length of the running average method controls the size of the vote group. The cost of large voting length was that the classifier would need more raw-segments and the output decision would be delayed for longer. To investigate the influence of the signal fluctuation and find an optimal compromise, the voting length was zoomed from 1 s to 9 s. T-tests were conducted. Fig.9 shows that as the voting length increased, the overall performance of the unified model for wakefulness against sleep classification (Fig.9(a)) and prediction (Fig. 9(b)) was largely improved. The best accuracy of classification of the models was 91.0% (std = 2.1%) with the voting length 9 s. We consider the voting length of 5 s as the optimal zoom of running average.

E. Model Structure Analysis

The unified classification training of wakefulness against sleep stages was repeated 5120 times. The structure of SVM and DT models were summarized for feature importance analysis.

Fig.10(a) shows the distribution of the features in the top four levels (root node, Level 1, Level 2, and Level 3) of the decision trees generated from the repeated training models. The results show that beta band occupied 99.3% root nodes in the 5120 trees and gamma band occupied the rest 0.7%. Band power in 25–27 Hz occupied 96.3% of the root nodes among

all of the trees. Delta, theta, and alpha band power never appeared as the root nodes. Alpha band power was the most frequently used feature in Level 1 to Level 3. It was found that 64.8% of the nodes in Level 1, 31.0% of the nodes in Level 2, and 19.4% of the nodes in Level 3 were the feature of band power in 7–9 Hz. Gamma band power occupied 31.0% of the nodes in Level 1. These results indicate that the power change in beta band (especially in the sub-band of 25–27 Hz) was the most important feature in the decision paths. Alpha band (especially in the sub-band of 7–9 Hz) and gamma band were found to be second and third, respectively, in levels of importance.

Fig.10(b) shows the average weights of the support vectors across the 5120 SVM models. The general roles of beta, alpha and gamma band were consistent with those in DT models. The increase of the weights in the band around 13 Hz shows that large weights were needed for the features which are relatively close in PSDs, to maximize the gap to the hyperplane.

In general, both DT and SVM models show that the most important features for the classification of wakefulness against sleep are located in the beta, alpha, and gamma bands. Among them, the beta band consistently plays the most important role.

IV. Discussion

This is the first systematic study to put forward a solution for automatic sleep stage classification based on STN-LFP postoperatively recorded in PD patients. The robust and interpretable SVM and DT models achieved high performance with an accuracy above 90%. Performance at the group level highlighted the reliable and ubiquitous sleep patterns. Benefiting from model interpretability, we analytically clarified that beta, alpha, and gamma rhythms in STN-LFP are critical for the future development of closed-loop DBS.

A. Taking STN-LFP as an Indicator for Sleep Stage

High frequency stimulation in the STN generally improves sleep quality of PD patients by decreasing wakefulness [33], whereas other reports suggest some sleep disorders related to movement are not improved or even deteriorated [4], [6], [7]. Therefore, to find a balance between the improvement of motor function and sleep disorders, closed-loop STN-DBS towards sleep cycles is a promising solution.

The use of EEG or ECoG to classify sleep stages has been widely discussed with a large spatial resolution in neural activity. However, for implantable closed-loop DBS, both EEG and ECoG recording are not yet viable for integration in daily use. LFP, which could be directly recorded from the DBS electrode, is an important feedback signal for the primary investigation of closed-loop DBS [16]. Although the specific role is not clear, the projections from the basal ganglia to pedunculopontine nucleus may involve the STN in the ascending activation network in sleep modulation [34]. Based on the results from LFPs and microelectrode recordings in the nucleus, STN neural activities are synchronously changed with sleep [13], [14], [35], [36]. Given the correlation with sleep modulation, STN-LFP which could be directly recorded from the DBS electrode is irreplaceable in the current field.

Recently, advancements in technology for recording LFPs with the implanted stimulator have opened a window for long-term observations and paved the way for the development of closed-loop DBS [37]. Taking STN-LFP as an indicator for sleep stage classification represents a new area but indeed a critical step for closed-loop DBS.

It could be intrinsically more challenging to use STN-LFPs for sleep classification. The STN-LFP signals are generated by fewer neurons in a more localized space than EEG or ECoG, which incorporate more randomness and variability [38], [39]. The modulation rules of sleep states in single channel STN-LFP may be more complicated and more individually diverse.

In this study, the performance of the SVM and DT models has reached the same level with previous EEG, ECoG, and iEEG studies [22], [40], [41], indicating the flexibility of taking STN-LFP for sleep stage classification. A previous study reported a primary trial for SVM classification with 30s STN-LFP epochs [14]. However, 30s are too long as the responding time of a closed-loop system and the complete method for building the models is not clarified. In this study, classification models based on 1s epochs were systematically built. A running average method was applied for confronting the data fluctuation in each second. With a voting length of 5 s, the models' performance is largely improved without sacrificing the responding speed. The classification in a simulated online scene was separately discussed at group and individual levels, which will directly guide the development of closed-loop DBS system. The same classification process was applicable to the EEG signals.

B. STN-LFP Activity Patterns During Sleep State

This study also reveals STN-LFP patterns during sleep state with robust and interpretable machine learning models. It was first reported by Urrestarazu et al. [13] (2009) that beta oscillations in the STN of PD patients decreased in non-REM sleep and recovered in REM sleep. The modulation of beta oscillation during movements in REM sleep was shown to be different with that during wakefulness [35], [42]. It was also suggested by Thompson et al. [14] (2017) that all frequency bands were significantly different between sleep stages. In our study, group analysis in the full frequency band showed that three low frequency bands (delta, theta, and alpha band) have a consistent trend of power increase in non-REM sleep and power decrease to a similar degree with wakefulness in REM sleep. The remaining high frequency band (beta and gamma band) has an opposite trend of power decrease in non-REM sleep and increase in REM sleep. These results are generally consistent with previous studies conducted during DBS surgery, but elucidate sleep patterns with full-band analysis at the group level and without acute surgical interferences [18], [19]. Furthermore, from the feature importance analysis of the classifiers in Fig. 9, we found that the largest modulation effects of wakefulness and sleep lie in the beta, alpha, and gamma bands, which are critical for the development of closed-loop DBS.

C. Error Sources and Limitations

The models proposed in this study exhibited good performance in the classification of wakefulness and sleep stages, whereas the performance of the classification of N1 and other sleep stage (especially REM) is moderate. Similar problems exist in sleep stage

classification by EEG [43], [44]. The reason for this may be the inherent similarities of N1 and REM signals in the STN which make it difficult to classify by the current features. The classification results for the independent N2 and N3 stage are also moderate. The performance is limited by the sample size of the N3 stage. Besides, the various locations of the DBS electrodes across subjects would directly influence the recorded LFP signals [45]. Therefore, the extracted features of the signal rhythms may be influenced. Other errors may come from the LFP sampling process which is highly integrated in the stimulator. The signal quality is restricted by the implantable device, including the space and power consumption [46].

Several LFPs were recorded during the symptoms of REM sleep behavior disorder and periodic limb movements of two subjects. However, limited by the sample size and contaminated signals, we did not find any significant conclusion. Another limitation is that the LFPs in the current work were all recorded during DBS off state. For clinical application, LFPs recorded during DBS on state should be considered. Nevertheless, one of the potential solutions provided from this work is to record the signal during a few seconds after DBS is switched off.

V. Conclusion and Future Work

For the first time, this study provides a systematic method for the classification of sleep stages based on STN-LFP in PD patients. The classification models are able to achieve high cross-validated performance at both group level and individual level, without losing interpretability and generalizability. The interpretable model structure highlights that beta, alpha, and gamma bands contribute the most to classification, taking a step towards uncovering the neuromodulation mechanisms of DBS and sleep disorders.

Future work will be focused on the development of the implantable closed-loop DBS system with integration of the methods from this study. From the perspective of engineering, embeddable algorithms with less computing cost and memory requirements should be considered. From the perspective of improving the classification performance, more precise correlation between the STN-LFP and awareness state needs to be clarified, especially for the sleep state transition periods. Other methods such as deep neural networks, wavelet transform, random forest, and multiple ensemble methods may increase the performance. Prediction of the awareness state in the next few seconds with current signals should be explored. Furthermore, for clinical application, the question to whether STN-LFP could be used as a robust indicator during medication on and/or DBS on state should be considered. More samples need to be recorded to study the influence of sleep disorder symptoms to LFP signals and the corresponding improvement for achieving a sufficient performance of classification. For the implantable closed-loop DBS system, recordings over a longer time scale, such as 6 months to several years need to be explored and the long-term stability of the LFP sleep patterns needs to be verified.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 81527901 and Grant 51777115, in part by the National Key Research and Development Program of China under Grant

2016YFC0105901 and Grant 2016YFC0105904, and in part by the Major Achievements Transformation Project of Beijing's College.

References

[1]. Perlmutter JS and Mink JW, "Deep brain stimulation," Annu. Rev. Neurosci, vol. 29, pp. 229–257, Jul. 2006. [PubMed: 16776585]

[2]. Comella CL, "Sleep disorders in Parkinson's disease: An overview," Movement Disorders, vol. 22, no. S17, pp. S367–S373, Dec. 2007. [PubMed: 18175398]

[3]. Tekriwal A et al., "REM sleep behaviour disorder: Prodromal and mechanistic insights for Parkinson's disease," J. Neurol. Neurosurg. Psychiatry, vol. 88, no. 5, pp. 445–51, Dec. 2016. [PubMed: 27965397]

[4]. Baumann-Vogel H et al., "The impact of subthalamic deep brain stimulation on sleep-wake behavior: A prospective electrophysiological study in 50 Parkinson patients," Sleep, vol. 40, no. 5, pp. 1–9, Mar. 2017.

[5]. Chahine L, Ahmed A, and Sun Z, "Effects of STN DBS for Parkinson's disease on restless legs syndrome and other sleep-related measures," Parkinsonism Rel. Disorders, vol. 17, no. 3, pp. 208–211, Mar. 2011.

[6]. Umemura A, Oka Y, Yamamoto K, Okita K, Matsukawa N, and Yamada K, "Complications of subthalamic nucleus stimulation in Parkinson's disease," Neurol. Medico-Chirurgica, vol. 51, no. 11, pp. 749–755, Nov-Jul 2011.

[7]. Kim YE et al., "Rapid eye movement sleep behavior disorder after bilateral subthalamic stimulation in Parkinson's disease," J. Clin. Neurosci, vol. 22, no. 2, pp. 315–319, Feb. 2014. [PubMed: 25439757]

[8]. De Cock VC et al., "Restoration of normal motor control in Parkinson's disease during REM sleep," Brain, vol. 130, no. 2, pp. 450–456, Feb. 2007. [PubMed: 17235126]

[9]. Amara AW et al., "Effects of subthalamic nucleus deep brain stimulation on objective sleep outcomes in Parkinson's disease," Movement Disorders Clin. Pract, vol. 4, no. 2, pp. 183–190, Mar-Apr 2016.

[10]. Little S and Brown P, "What brain signals are suitable for feedback control of deep brain stimulation in Parkinson's disease?" Ann. York Acad. Sci, vol. 1265, no. 1, pp. 9–24, Jul. 2012.

[11]. Kühn AA et al., "High-frequency stimulation of the subthalamic nucleus suppresses oscillatory ß activity in patients with Parkinson's disease in parallel with improvement in motor performance," J. Neurosci, vol. 28, no. 24, pp. 6165–6173, Jun. 2008. [PubMed: 18550758]

[12]. Brittain J-S and Brown P, "Oscillations and the basal ganglia: Motor control and beyond," Neuroimage, vol. 85, pp. 637–647, Jan. 2014. [PubMed: 23711535]

[13]. Urrestarazu E et al., "Beta activity in the subthalamic nucleus during sleep in patients with Parkinson's disease," Movement Disorders, vol. 24, no. 2, pp. 254–260, Oct. 2008.

[14]. Thompson JA et al., "Sleep patterns in Parkinson's disease: Direct recordings from the subthalamic nucleus," J. Neurol. Neurosurg. Psychiatry, vol. 89, no. 1, pp. 95–104, Jan. 2017. [PubMed: 28866626]

[15]. Beudel M and Brown P, "Adaptive deep brain stimulation in Parkinson's disease," Parkinsonism Rel. Disorders, vol. 22, pp. S123–S126, Jan. 2015.

[16]. Meidahl AC, Tinkhauser G, Herz DM, Cagnan H, Debarros J, and Brown P, "Adaptive deep brain stimulation for movement disorders: The long road to clinical therapy," Movement Disorders, vol. 32, no. 6, pp. 810–819, Jun. 2017. [PubMed: 28597557]

[17]. Qian X, Chen Y, Ma B, Hao H, and Li L, "Chronically monitoring the deep brain rhythms: From stimulation to recording," Sci. Bull, vol. 61, no. 19, pp. 1522–1524, Oct. 2016.

[18]. Mann JM et al., "Brain penetration effects of microelectrodes and DBS leads in STN or GPi," J. Neurol., Neurosurgery Psychiatry, vol. 80, no. 7, pp. 794–798, Feb. 2009.

[19]. Deogaonkar M, Nazzaro JM, Machado A, and Rezai A, "Transient, symptomatic, post-operative, non-infectious hypodensity around the deep brain stimulation (DBS) electrode," J. Clin. Neurosci, vol. 18, no. 7, pp. 910–915, Jul. 2011. [PubMed: 21571534]

[20]. Marsden JF, Limousin-Dowsey P, Ashby P, Pollak P, and Brown P, "Subthalamic nucleus, sensorimotor cortex and muscle interrelationships in Parkinson's disease," Brain, vol. 124, pp. 378–388, Feb. 2001. [PubMed: 11157565]

[21]. Kremen V et al., "Behavioral state classification in epileptic brain using intracranial electrophysiology," J. Neural Eng, vol. 14, no. 2, p. 026001, Jan. 2017. [PubMed: 28050973]

[22]. Zempel JM et al., "Characterization of scale-free properties of human electrocorticography in awake and slow wave sleep states," Frontiers Neurol, vol. 3, p. 76, Jun. 2012.

23. [] Reed CM et al., "Automatic detection of periods of slow wave sleep based on intracranial depth electrode recordings," J. Neurosci. Methods, vol. 282, pp. 1–8, Apr. 2017. [PubMed: 28238858]

24. [] Boostani R, Karimzadeh F, and Nami M, "A comparative review on sleep stage classification methods in patients and healthy individuals," Comput. Methods Programs Biomed, vol. 140, pp. 77–91, Mar. 2017. [PubMed: 28254093]

[25]. Motamedi-Fakhr S, Moshrefi-Torbati M, Hill M, Hill CM, and White PR, "Signal processing techniques applied to human sleep EEG signals—A review," Biomed. Signal Process. Control, vol. 10, pp. 21–33, Mar. 2014.

[26]. Iber C et al., The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology Technical Specifications, 1st ed. Darien, IL, USA: Amer. Acad. Sleep Med., 2007.

[27]. Suykens JAK and Vandewalle J, "Least squares support vector machine classifiers," Neural Process. Lett, vol. 9, no. 3, pp. 293–300, Jun. 1999.

[28]. Chang C-C and Lin C-J, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[29]. Landgrebe D, "A survey of decision tree classifier methodology," IEEE Trans. Syst., Man Cybern, vol. 21, no. 3, pp. 660–674, 5 1991.

[30]. Benjamini Y and Hochberg Y, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," J. Roy. Stat. Soc. B, Methodol), vol. 57, no. 1, pp. 289–300, 1995.

[31]. Loh W-Y, "Classification and regression trees," Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery, vol. 1, no. 1, pp. 14–23, 12011.

[32]. Imtiaz SA and Rodriguez-Villegas E, "Recommendations for performance assessment of automatic sleep staging algorithms," in Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc, Aug. 2014, pp. 5044–5047.

[33]. Arnulf I et al., "Improvement of sleep architecture in PD with subthalamic nucleus stimulation," Neurology, vol. 55, no. 11, pp. 1732–1735, Dec. 2000. [PubMed: 11113233]

[34]. Takakusaki K, Saitoh K, Harada H, and Kashiwayanagi M, "Role of basal ganglia-brainstem pathways in the control of motor behaviors," Neurosci. Res, vol. 50, no. 2, pp. 137–151, Oct. 2004. [PubMed: 15380321]

[35]. Fernandez-Mendoza J et al., "Evidence of subthalamic PGO-like waves during REM sleep in humans: A deep brain polysomnographic study," Sleep, vol. 32, no. 9, pp. 1117–1126, Sep. 2009. [PubMed: 19750916]

[36]. Stefani A et al., "Spontaneous sleep modulates the firing pattern of Parkinsonian subthalamic nucleus," Exp. Brain Res, vol. 168, nos. 1–2, pp. 277–280, Jan. 2006. [PubMed: 16328297]

[37]. Shen H, "Neuroscience: Tuning the brain," Nature, vol. 507, pp. 290–292, Mar. 2014. [PubMed: 24646978]

[38]. Kajikawa Y and Schroeder CE, "How local is the local field potential?" Neuron, vol. 72, no. 5, pp. 847–858, Dec. 2011. [PubMed: 22153379]

[39]. Buzsaki G, Anastassiou CA, and Koch C, "The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes," Nature Rev. Neurosci, vol. 13, no. 6, pp. 407–420, 5 2012. [PubMed: 22595786]

[40]. Chambon S, Galtier MN, Arnal PJ, Wainrib G, and Gramfort A, "A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series," IEEE Trans. Neural Syst. Rehabil. Eng, vol. 26, no. 4, pp. 758–769, Apr. 2018. [PubMed: 29641380]

[41]. Supratak A, Dong H, Wu C, and Guo Y, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," IEEE Trans. Neural Syst. Rehabil. Eng, vol. 25, no. 11, pp. 1998–2008, Nov. 2017. [PubMed: 28678710]

[42]. Hackius M, Werth E, Sürücü O, Baumann CR, and Imbach LL, "Electrophysiological evidence for alternative motor networks in REM sleep behavior disorder," J. Neurosci, vol. 36, no. 46, pp. 11795–11800, Nov. 2016. [PubMed: 27852786]

[43]. Liang S-F, Kuo Y-H, Hu Y-H, Pan Y-H, and Wang Y-H, "Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models," IEEE Trans. Instrum. Meas, vol. 61, no. 6, pp. 1649–1657, Jun. 2012.

[44]. Himanen S-L and Hasan J, "Limitations of rechtschaffen and kales," Sleep Med. Rev, vol. 4, no. 2, pp. 149–167, Apr. 2000. [PubMed: 12531164]

[45]. Horn A et al., "Toward an electrophysiological 'sweet spot' for deep brain stimulation in the subthalamic nucleus," Hum. Brain Mapping, vol. 38, no. 7, pp. 3377–3390, Apr. 2017.

[46]. Qian X, Chen Y, Feng Y, Ma B, Hao H, and Li L, "A method for removal of deep brain stimulation artifact from local field potentials," IEEE Trans. Neural Syst. Rehabil. Eng, vol. 25, no. 12, pp. 2217–2226, Dec. 2016. [PubMed: 28113981]

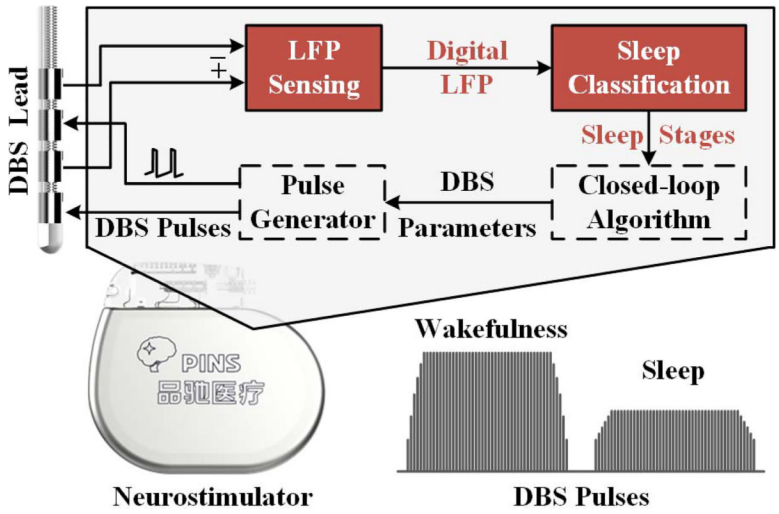


Fig. 1.

Illustration of the use-case of conceptual implantable sleep monitoring and closed-loop DBS system. LFPs are detected by the DBS lead. With integrated classifiers, sleep stages are predicted and relative monitoring and closed-loop algorithms can adjust the DBS pulses, such as decrease the amplitude during sleep stages. The solid blocks with deep colors represent the work in this study.

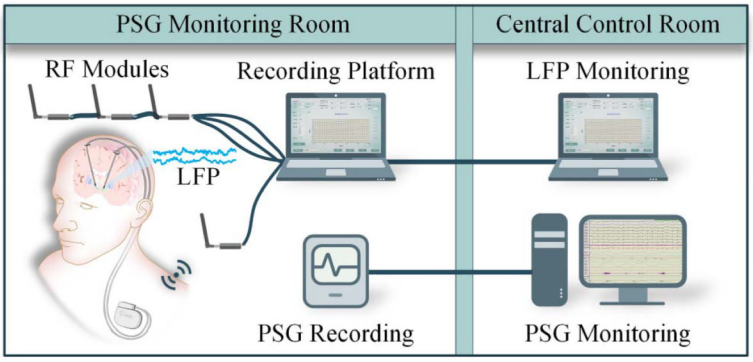


Fig. 2.

Illustration of the STN-LFP recording and sleep monitoring. LFPs were recorded by the implanted deep brain stimulator (G106R, PINS, China) and wirelessly transmitted to the LFP recording PC through one of the RF modules. Data were monitored in the central control room during the experiments. PSG-polysomnography; STN- subthalamic neuclus; LFP-local field potential; RF-radio frequency.

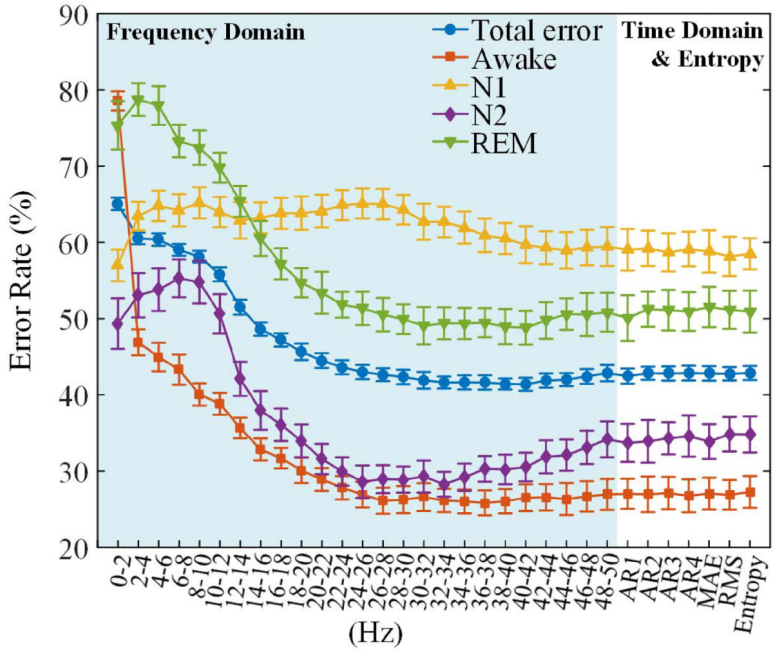


Fig. 3.

Error rates of multi-class classification as a function of increasing feature dimensions. Bars represent one standard deviation.

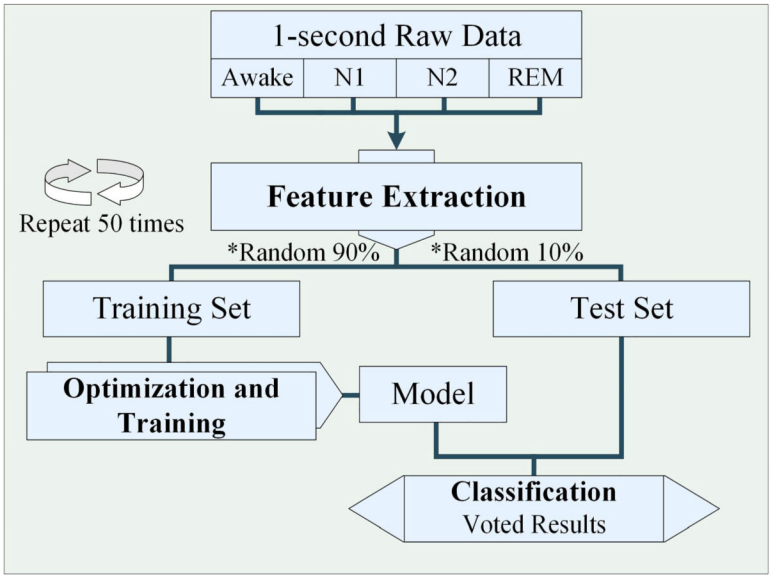


Fig. 4.

Flow diagrams of classification and prediction. The difference of classification and prediction is marked by the asterisk. For classification, 90% of the raw data set was randomly selected as the training data set and the remaining 10% was the test data set. For prediction, the first 90% in time order was selected as the training data set and the remaining 10% was the test data set.

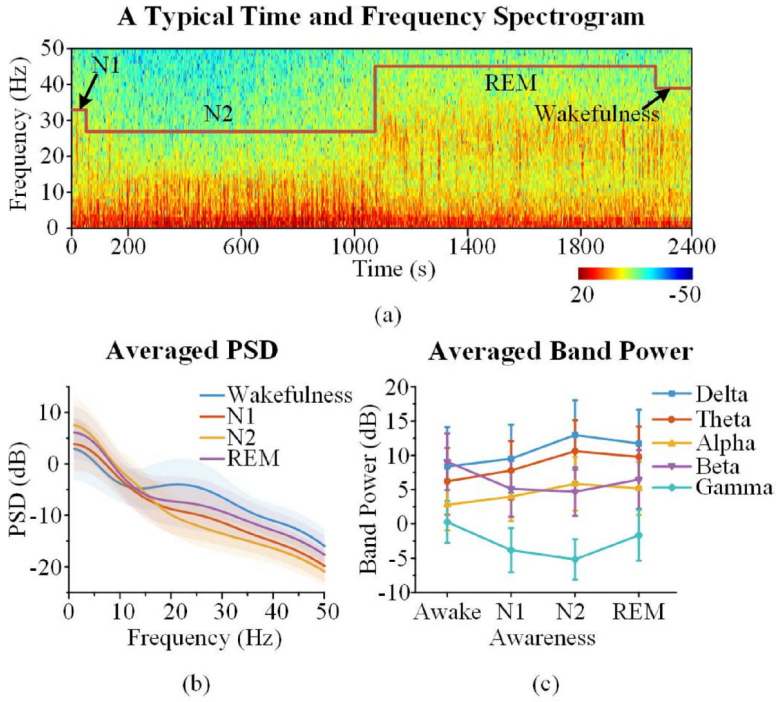


Fig. 5.

STN-LFP sleep patterns in frequency domain. (a) Time-frequency spectrogram of a typical piece of LFP data recorded during wakefulness and different sleep stages. The red line labels the sleep stages from N1, N2, REM to wakefulness. (b) Averaged PSDs of LFPs across subjects. The shadows represent one standard deviations regions. (c) Averaged power in different the stages. The bars represent one standard deviation values. PSD-power spectrum density.

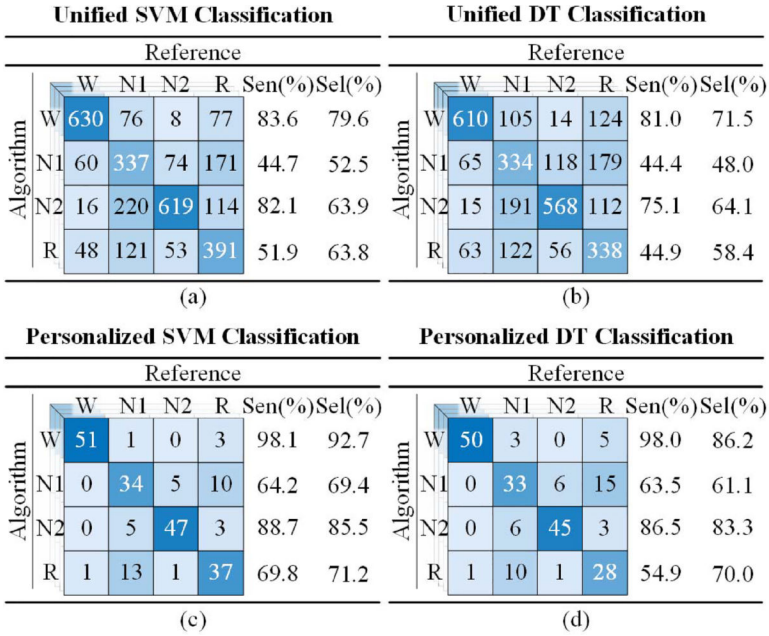
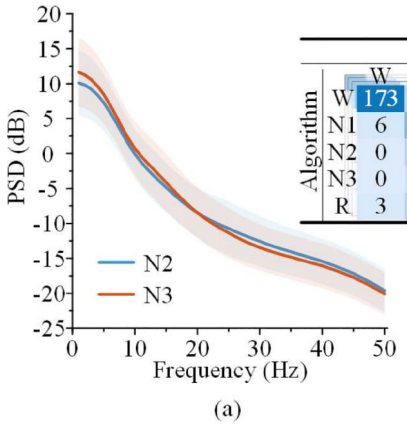


Fig. 6.

Averaged confusion matrix of multi-class classification for the 50 repeats. Voting length was fixed to 5s.



(a)

SVM Classification

Algorithm \ Reference	W	N1	N2	N3	R	Sen (%)	Sel (%)
W	173	26	1	0	26	95.1	76.5
N1	6	76	27	8	40	41.8	48.4
N2	0	34	65	37	22	35.5	41.1
N3	0	12	78	132	8	72.9	57.4
R	3	34	12	4	85	47.0	61.6

(b)

Fig. 7.

Results of classification with independent N2 and N3 recordings. (a) Averaged PSD of N2 and N3 stages. Shadows represent for one stand deviation. (b) Averaged SVM multi-class classification confusion matrix.

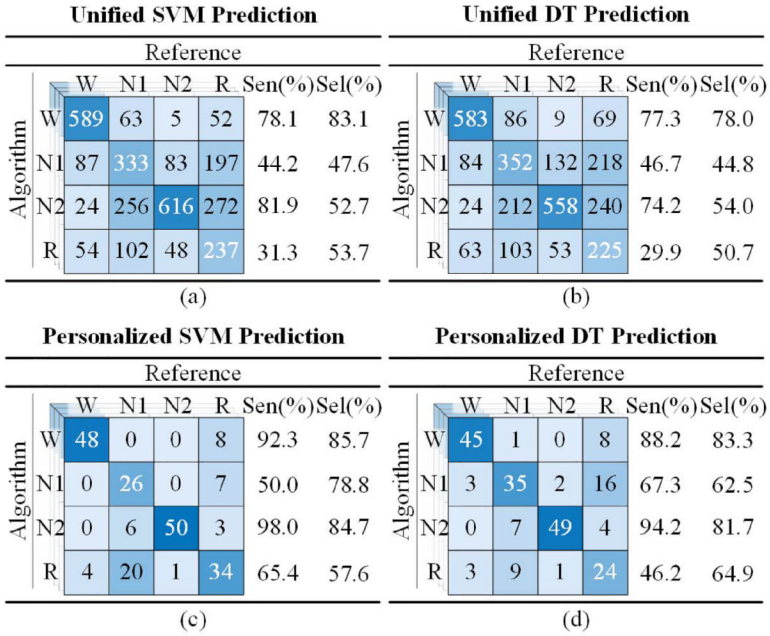


Fig. 8.

Averaged confusion matrix of multi-class prediction for the 50 repeats. Voting length was fixed to 5s.

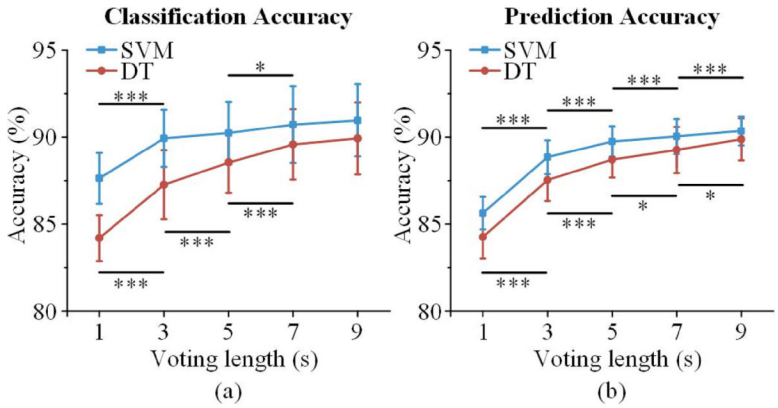


Fig. 9.

Increase of the voting length could improve the performance of classification and prediction for wakefulness against sleep (N1, N2, and REM). Cubes and dots represent the average accuracy for the repeated trained models. Bars represent standard deviations. *, $0.01<p<0.05$, **, $0.001<p<0.01$, ***, $p<0.001$, N=50.

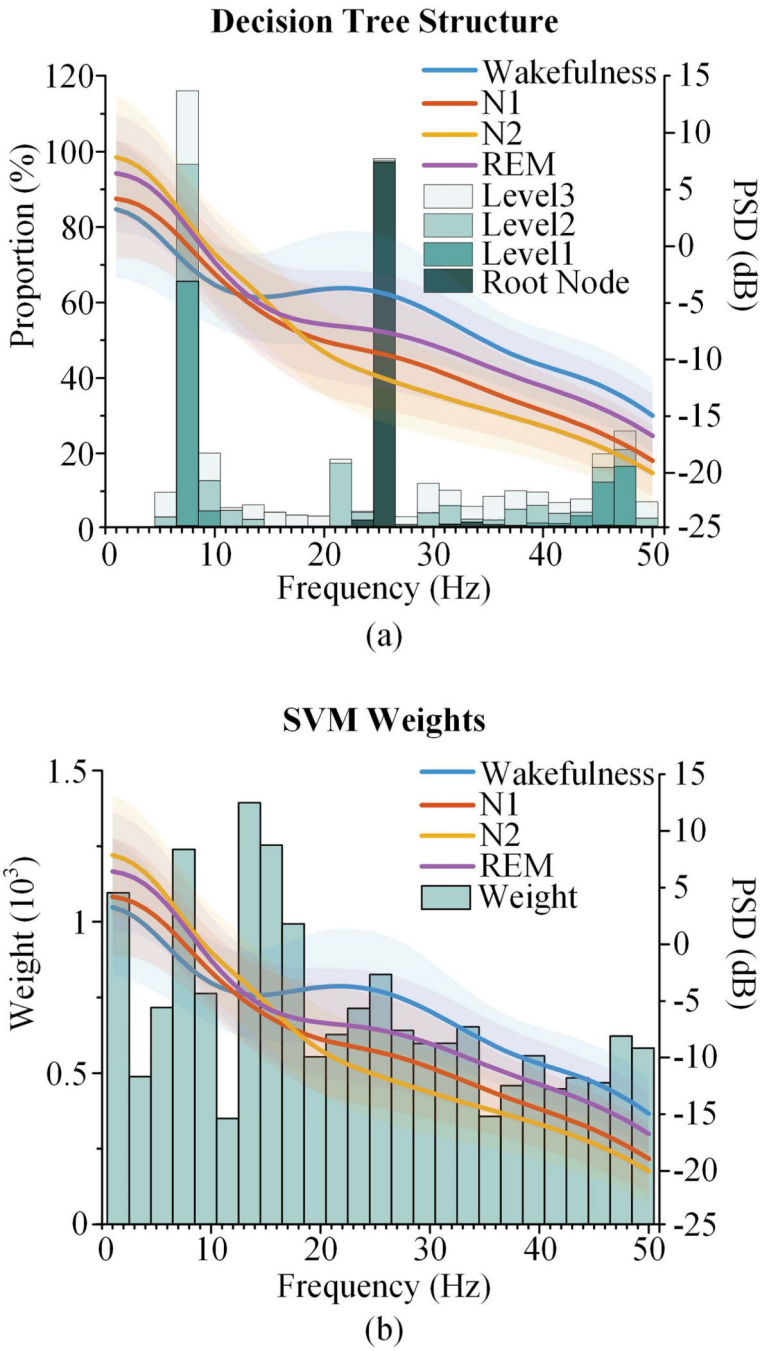


Fig. 10.

Model structure analysis of DT and SVM models. (a) Distribution of the features as nodes in the classification trees for wakefulness against sleep. Bins represent the percentages of each feature (band power of every 2 Hz of PSD) locating at root node, Level 1, Level 2, and Level 3. (b) Averaged weights of features in support vectors of SVM models.

TABLE I

Demographic Information of Subjects Involved in the Study and the Amount of the Data Segments for Analysis

No.	Age	Gender	PD History (years)	H-Y [a]	DBS Improvement [b] (%)	Recording Impedance [c] (Ω)	Number of 1s epochs for analysis				
							Awake×30	N1×30	N2×30	N3×30	REM×30
1	40	M	8	3	42	L: 2528, R: 3293	6	63	51	0	11
2	53	M	13	3	72	L: 1636, R: 1787	34	24	38	0	6
3	67	M	8	3	88	L: 2528, R: 3086	16	24	168	6	126
4	51	M	20	4	23	L: 3086, R: 3086	12	10	508	0	58
5	65	M	8	3	45	L: 2528, R: 1723	18	82	352	0	250
6	60	F	7	3	68	L: 1693, R: 2991	22	106	302	0	40
7	46	M	7	4	62	L: 2740, R: 2464	14	20	428	44	238
8	61	F	8	4	48	L: 2050, R: 2595	22	24	414	116	156
9	47	F	12	5	53	L: 2464, R: 2528	12	18	424	0	100
10	56	F	15	4	64	L: 2049, R: 1723	22	2	208	0	6
11	61	F	8	4	41	L: 2595, R: 2404	24	62	324	0	320
12	51	M	8	4	77	L: 2008, R: 3186	20	48	292	0	190

[a] H-Y, Hoehn and Yahr Scale

[b] The DBS improvement was calculated as the percentage of the difference between preoperative and postoperative DBS on UPDRS III scores.

[c] The impedance was measured between the pair of contacts for recording. The therapeutic impedance measured between the case and stimulation contact was around 1000 Ω.

TABLE II

Numbers of 1 s Epochs Included in Analysis (Voting Length is 5 s)

		Awake	N1	N2	REM
Unified	Training	5790	128701404	99030	40440
	Testing	754	1404	9724	3978
Personalized	Training	540	1290	7860	5130
	Testing	52	130	780	494

TABLE III

Averaged Classification Accuracy for 50 Repeats

Accuracy (%)	Unified		Personalized	
	SVM	DT	SVM	DT
Awake-N1	87.1	84.4	97.8	96.1
Awake-N2	95.5	94.7	99.8	98.4
Awake-REM	86.8	83.5	96.6	94.8
N1-N2	77.7	75.8	91.4	90.7
N1-REM	70.8	69.1	74.4	73.3
N2-REM	83.9	81.9	93.5	93.1
Awake-Sleep [a]	90.2	88.5	97.4	94.7

[a] Sleep including stages of N1, N2, and REM.

TABLE IV

Averaged Prediction Accuracy for 50 Repeats

Accuracy (%)	Unified		Personalized	
	SVM	DT	SVM	DT
Awake-N1	86.4	84.4	99.2	94.9
Awake-N2	93.0	93.2	100.0	99.9
Awake-REM	86.2	85.2	88.8	87.0
N1-N2	75.3	73.3	92.4	91.2
N1-REM	62.2	62.3	68.0	69.6
N2-REM	71.6	70.8	95.9	95.3
Awake-Sleep[a]	89.8	88.7	95.4	93.1

[a]Sleep including stages of N1, N2, and REM.