Sleep Spindle Detection Using RUSBoost and Synchrosqueezed Wavelet Transform

Takafumi Kinoshita, Koichi Fujiwara, *Member, IEEE*, Manabu Kano, *Member, IEEE*, Keiko Ogawa, Yukiyoshi Sumi, Masahiro Matsuo, and Hiroshi Kadotani

***Abstract*— Sleep spindles are important electroencephalographic (EEG) waveforms in sleep medicine; however, it is burdensome even for experts to detect spindles, so automatic spindle detection methodologies have been investigated. Conventional methods utilize waveforms template matching or machine learning for detecting spindles. In the former approach, it is necessary to tune thresholds for individual adaptation, while the latter approach has the problem of imbalanced data because the amount of sleep spindles is small compared with the entire EEG data. The present work proposes a sleep spindle detection method that combines wavelet synchrosqueezed transform (SST) and random under-sampling boosting (RUSBoost). SST is a time-frequency analysis method suitable for extracting features of spindle waveforms. RUSBoost is a framework for coping with the imbalanced data problem. The proposed SST-RUS can deal with the imbalanced data in spindle detection and does not require threshold tuning because RUSBoost uses majority voting of weak classifiers for discrimination. The performance of SST-RUS was validated using an open-access database called the Montreal archives of sleep studies cohort 1 (MASS-C1), which showed an F-measure of 0.70 with a sensitivity of 76.9% and a positive predictive value of 61.2%. The proposed method can reduce the burden of PSG scoring.**

***Index Terms*— Sleep spindle detection, electroencephalography, polysomnography, synchrosqueezed wavelet transform, random under sampling boosting.**

I. Introduction

POLYSOMNOGRAPHY (PSG) is a gold standard test for sleep disorders, which records many channels simultaneously during sleep, e.g., electroencephalogram (EEG), electrocardiogram (ECG), electrooculography (EOG), electromyography (EMG), airflow, and oxygen saturation. PSG data analysis includes sleep stage scoring based on visual inspection of the EEG data. In particular, sleep spindle and k-complex detection are important for sleep stage scoring since they characterize non-rapid eye movement (NREM) sleep stage 2 although some sleep spindles and k-complexes also occur in NREM sleep stage 3.

Manuscript received August 11, 2019; revised December 8, 2019; accepted January 3, 2020. Date of publication January 7, 2020; date of current version February 12, 2020. This work was supported in part by JST PRESTO under Grant JPMJPR1859 and in part by JSPS KAHENHI under Grant 17H00872. *(Corresponding author: Koichi Fujiwara.)*

Takafumi Kinoshita and Manabu Kano are with the Department of Systems Science, Kyoto University, Kyoto 606-8501, Japan.

Koichi Fujiwara is with the Department of Material Process Engineering, Nagoya University, Nagoya 464-8601, Japan (e-mail: fujiwara.koichi@material.material-u.ac.jp).

Keiko Ogawa is with the Graduate School of Integrated Arts and Sciences, Hiroshima University, Higashihiroshima 739-8511, Japan.

Yukiyoshi Sumi and Masahiro Matsuo are with the Department of Psychiatry, Shiga University of Medical Science, Shiga 520-2192, Japan.

Hiroshi Kadotani is with the Department of Sleep and Behavioral Sciences, Shiga University of Medical Science, Shiga 520-2192, Japan.

Digital Object Identifier 10.1109/TNSRE.2020.2964597

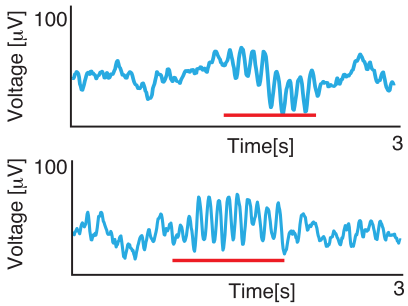

Fig. 1. Examples of sleep spindle waveforms: the red underlines denote spindle waveforms.

According to a sleep scoring manual published by the American Academy of Sleep Medicine (AASM), sleep spindles are defined as a train of distinct sinusoidal waves having a frequency of 11-16 Hz (most commonly 12-14 Hz) and a duration $\geq$ 0.5 s, usually maximal in amplitude in central derivations [1]. Figure 1 illustrates typical sleep spindle waveforms.

It is suggested that sleep spindles are linked to occurrence of neurologic diseases. Latreille *et al.* reported that spindle density, which is defined as the number of sleep spindles observed in unit time, significantly decreased in Parkinson's disease (PD) patients and was in association with cognitive impairment [2]. In addition, the number of sleep spindles was significantly lower in schizophrenia patients than that of non-schizophrenic patients [3].

Although accurate detection of sleep spindles is clinically essential, manual detection thereof is difficult and time-consuming even for experts. Wendt *et al.* reported that the average intra-expert and inter-expert F-measure agreements in sleep spindle detection were 72$\pm$7% and 61$\pm$6%, respectively [4]. Thus, automatic sleep spindle detection methods have been investigated in order to relieve the burden on experts as well as to improve detection accuracy.

Existing automatic sleep spindle detection methods are classified into two approaches: template matching approach and machine learning approach [5]. The former approach

divides EEG data into multiple segments and calculates the similarity between the divided EEG data and templates of the spindle waveforms. The EEG segments are detected as sleep spindles when the calculated similarities exceed a predefined threshold [6]–[8]. Since there is individuality in EEG data, the threshold needs to be tuned for individual adaptation. Some self-adjustment methods of the threshold have been proposed, which use prior knowledge of spindles included in the EEG data or information about previously detected spindles [9], [10].

The machine learning approach extracts multiple features from the EEG data and discriminates the EEG data between spindles and non-spindles from the extracted features by using a trained classifier [5]. Various machine learning methodologies have been adopted for sleep spindle detection, such as support vector machine (SVM), neural network (NN), hidden Markov model (HMM), and the Bayesian model [11]–[14].

The main weakness of the machine learning approach is the difficulty in constructing an appropriate dataset for classifier training. In order to construct reliable classifiers, sufficient volumes of spindle data and non-spindle data are needed [15]. Since it is much easier to collect non-spindle data than spindle data, the training dataset tends to be imbalanced – the non-spindle data constitute the majority while the spindle data constitute the minority. Coping with such imbalanced data is a challenging problem for standard machine learning algorithms since most of them are designed for balanced data [16], [17].

The present work proposes a new spindle detection algorithm that overcomes the drawbacks of the existing approaches – the drawbacks being the necessity for threshold tuning for individual adaptation in the template matching approach and classifier training from the imbalanced data in the machine learning approach. The proposed method utilizes random under-sampling boosting (RUSBoost) as a classifier. Random under-sampling (RUS) is a framework for coping with the imbalanced data problem by discarding randomly selected majority samples so that the numbers of majority and minority samples become balanced when a classifier is built. Boosting is an ensemble method that constructs multiple weak classifiers and combines the outputs of the weak classifiers into the final output by means of majority voting [18]. RUSBoost is a combination of RUS and boosting, which is able to cope with the imbalanced data problem [19], because ensemble learning algorithms tend to outperform other machine learning algorithms when dealing with the imbalanced data problem [20]. Moreover, RUSBoost does not require threshold tuning for individual adaptation because it uses the majority voting of weak classifiers.

In order to construct a good classifier, appropriate input features have to be selected. In this research, synchrosqueezed wavelet transform (SST) is used for extracting features from the EEG data. SST is a time-frequency analysis method suitable for analyzing multicomponent signals with oscillating modes. Since SST decomposes a signal into multiple frequency components with high-resolution, it can be expected to extract specific features of spindle waveforms appropriately.

In the proposed method, referred to as SST-RUS, multiple features are extracted from the EEG data by SST, and sleep spindles are detected from the extracted features by RUSBoost. Spindle detection performance is improved by combining SST and RUSBoost.

The performance of the proposed SST-RUS is demonstrated through its application to an open-access database called Montreal archives of sleep studies cohort 1 (MASS-C1) [21].

II. Spindle Detection

The present work proposes a new method for detecting sleep spindles that can cope with the imbalanced data problem in the machine learning approach and the threshold tuning problem in the template matching approach. The proposed SST-RUS uses SST for feature extraction and RUSBoost for classifier construction. First, an explanation of SST is provided.

A. Synchrosqueezing Wavelet Transform (SST)

EEG signals are assumed to consist of several components – background activities, temporary waves, and sleep spindles – where sleep spindles are temporary waves whose frequency range is 11-16 Hz. In order to separate spindles from background activities, accurate frequency estimation of the EEG signals is crucial. Previous studies have adopted various time-frequency approaches for spindle detection, such as continuous wavelet transform (CWT) [22] and empirical mode decomposition (EMD) [8].

SST gives plausible frequency estimation of a signal when it consists of several components with different frequency ranges.

CWT is defined as inner products between a signal $s(t)$ and a mother wavelet $\psi(\cdot)$ [23]:

$$W_s(b,a) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{|a|}} s(t) \psi^* \left(\frac{t-b}{a} \right) dt \tag{1}$$

where a and b denote a scaling parameter and a translating parameter, respectively. $\psi^*(\cdot)$ is the complex conjugate of $\psi(\cdot)$. $W_s(b,a)$ is interpreted as the power of $s(t)$ at time b and scale a.

SST is an extended version of CWT for analyzing multicomponent signals with high-resolution [24], [25]. In SST, the phase transform $f_s(b,a)$ is obtained as

$$f_s(b,a) = \frac{1}{2\pi j W_s(b,a)} \frac{\partial}{\partial b} W_s(b,a) \tag{2}$$

where j denotes the imaginary unit. Equation (2) represents mapping from time-scale representation to time-frequency representation. SST with accuracy δ and threshold ϵ is defined as

$$S(b,f) = \int_{A_{s,\epsilon}(b)} W_s(b,a) \frac{1}{\delta} h \left(\frac{f - f_s(b,a)}{\delta} \right) a^{-\frac{3}{2}} da \tag{3}$$

where f is a frequency and $A_{s,\epsilon}(b) = \{a \in \mathbb{R}_+; |W_s(b,a)| > \epsilon\}$. $h(t)$ is a function with $\int h(t)dt = 1$. $S(b,f)$ is referred to as the SST coefficient.

This transformation gives the squeezed CWT because the instantaneous frequency band is reassigned at the centroid of the CWT time-frequency region. Thus, this frequency

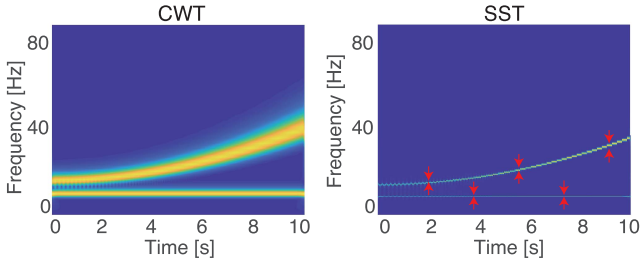

Fig. 2. Scalograms of CWT (left) and SST (right) of a chirp signal. These examples show that SST gives a sharper time-frequency representation than CWT.

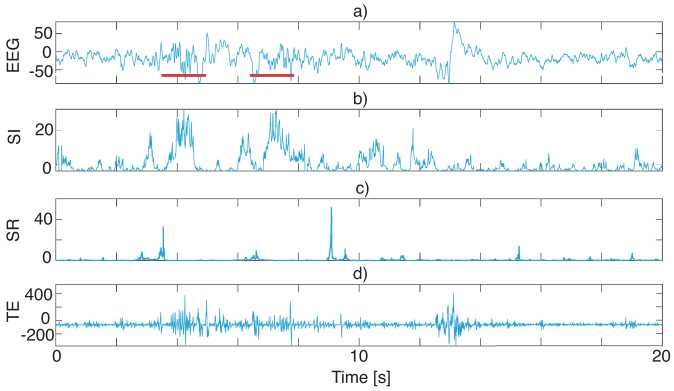

Fig. 3. Extracted features: a) original EEG, b) *SI*, c) *SR*, and d) *TE*. *SI* and *TE* significantly fluctuated when the spindle waveforms generated.

reassignment results in a sharpened output in comparison with the CWT.

A comparison example of CWT and SST is shown in Fig. 2. The original signal consists of two components: a chirp signal, whose frequency increases from 16 Hz to 36 Hz, and a sinusoidal signal having a frequency of 8Hz. SST gives a sharper time-frequency representation than CWT. This characteristic of SST may be beneficial for spindle detection.

Kabir *et al.* proposed an SST-based spindle detection method that derives an original index and detects spindles when the derived index exceeds a threshold [26]. According to the method, the threshold has to be tuned appropriately in advance.

B. SST-Based Feature Extraction

The present work introduces three SST-based features for spindle detection: sigma index SI, sigma ratio SR, and Teager energy operator TE. SST is applied to each EEG segment to obtain the SST coefficient $S(b, f)$, and features are calculated by using $S(b, f)$. The Morlet wavelet, which has a similar form to sleep spindles, is adopted as the mother wavelet in this study. Details of the three features, SI, SR, and TE, are as follows.

1) Sigma Index: The sigma index SI represents the ratio of the peak power of a spindle (11-16 Hz) to the peak powers of the frequency range around the spindle range. The original sigma index proposed by Huupponen *et al.* is as follows [7]:

$$SI(t) = \frac{2 \cdot \max(|P_{\mathrm{sp}}(t)|)}{\mathrm{mean}(|P_{\mathrm{low}}(t)|) + \mathrm{mean}(|P_{\mathrm{high}}(t)|)} \tag{4}$$

where $P_{\mathrm{low}}(t)$, $P_{\mathrm{high}}(t)$, and $P_{\mathrm{sp}}(t)$ are the power spectra of frequency ranges $4-10$ Hz, $20-40$ Hz, and $11-16$ Hz, respectively. Each $P(t)$ is obtained using Fourier transform with a 1 s window centered at t. Max and mean denote the maximum and the mean values in the specific frequency range.

In this research, the sigma index is modified by replacing $P(t)$ with $S(t, f)$, which makes it possible to distinguish the sleep spindles from the alpha activity. The modified $SI(t)$ becomes a large value when sleep spindles occur, as shown in Fig. 3 b). The red horizontal lines in Fig. 3 a) denote spindle waveforms labeled by an expert.

2) Sigma Ratio: The original sigma ratio SR was proposed by Patti *et al.* [27]:

$$SR(t) = \frac{R(t)}{R(t-1) + R(t+1}) \tag{5}$$

where $R(t)$ is the ratio of the power of the spindle frequency range (11-16 Hz) to the whole signal power at time t. When a waveform with a frequency range of $11-16$ Hz appears, $SR(t)$ becomes large.

This research uses a modified version of the sigma ratio $SR(t)$, by replacing $R(t)$ with the modified sigma index $SI(t)$, because it is thought to be more sensitive to spindle occurrence than $R(t)$. In Fig. 3 c), $SI(t)$ became large exactly when the first spindle appearance started.

3) Teager Energy Operator: The Teager energy operator TE is defined as follows [27]:

$$TE(t) = s(t)^2 - s(t-1)s(t+1) \tag{6}$$

where $s(t)$ is the analyzed signal. $TE(t)$ represents instantaneous changes in the signal amplitude. That is, the signal amplitude becomes high when temporary waves such as sleep spindles appear. A large fluctuation of $TE(t)$ is observed when a sleep spindle occurs, as shown in Fig. 3 d).

C. Random Under-Sampling Boosting (RUSBoost)

RUSBoost is a boosting algorithm that can deal with the imbalanced data problem [19]. Since the number of spindle samples is smaller than that of non-spindle samples, the adaptation of RUSBoost as a classifier for spindle detection is reasonable.

Algorithm 1 describes the procedure of RUSBoost. A dataset consists of samples $\boldsymbol{x}_i$ and their labels y_i as $\mathcal{D} = \{\boldsymbol{x}_i, y_i\}$ $(i = 1, \cdots, N)$ and $\mathcal{Y} = \{-1, +1\}$. N denotes the total number of samples, and the ratio of positive and negative samples is unbalanced. Before training, it is necessary to determine the number of iterations T and the ratio of the number of the majority samples to that of samples in an under-sampled dataset $\mathcal{D}'$, $\tilde{c}$.

In step 1, the weight of each sample is initialized equally to $1/N$. Randomly-selected majority samples are discarded so that the ratio of the number of the majority samples to that of samples in $\mathcal{D}'$ becomes $\tilde{c}$ in step 3. The tth weak classifier h_t is trained from $\mathcal{D}'_t$ and the hypothesis $h_t(\boldsymbol{x}_i)$ $(i = 1, \ldots, N)$ is obtained in steps 4 and 5. In steps 6 and 7, the error rate ε_t of the tth weak classifier h_t is calculated and the weight update parameter β_t is obtained. The boosting weights w_t are updated

Algorithm 1 RUSBoost

Input: $\mathcal{D}$, T, $\tilde{c}$

1: Initialize boosting weights $w_1(\boldsymbol{x}_i) = 1/N$ for all i.
2: **for** $t = 1, 2, \ldots, T$ **do**
3: Perform random under-sampling to get $\mathcal{D}'_t$.
4: Train weak classifier h_t from $\mathcal{D}'_t$.
5: Get hypothesis $h_t(\boldsymbol{x}_i)$ for all i in $\mathcal{D}$.
6: Calculate the error of $h_t : \varepsilon_t = \sum_{i:h_t(\boldsymbol{x}_i)\neq y_i} w_t(\boldsymbol{x}_i)$.
7: Set the weight update parameter $\beta_t = \varepsilon_t/(1-\varepsilon_t)$.
8: Update the boosting weight $w_{t+1}(i)$:

$$w_{t+1}(i) = w_t(i) \times \begin{cases} \beta_t & h_t(\boldsymbol{x}_i) = y_i \\ 1 & \text{otherwise} \end{cases}.$$

9: Normalize w_{t+1}:

$$w_{t+1}(i) = w_{t+1}/Z_t \text{ where } Z_t = \sum_{i=1}^{N} w_{t+1}(i).$$

10: **end for**

Output: The final classifier:

$$h_{\text{fin}}(\boldsymbol{x}) = \arg\max_{y\in\mathcal{Y}} \sum_{t:h_t(\boldsymbol{x})=y} \log(1/\beta_t).$$

in steps 8 and 9. After T iterations, the final classifier $h_{\text{fin}}(\boldsymbol{x})$ is returned as a weighted vote of the T weak hypotheses. A classification and regression tree (CART) is adopted as a weak classifier in this work [28].

Because the final classifier is the weighted sum of multiple CARTs, the variable importance of the pth variable, VI_p, is defined as the weighted sum of the decreases by the pth variable splitting:

$$VI_p = \frac{1}{Z_{\text{VI}}} \sum_t \log(1/\beta_t) \Delta I_G^t(p) \tag{7}$$

where $\Delta I_G^t(p)$ $(t = 1, \cdots, T)$ is the Gini coefficient decrease by the pth variable splitting in the tth CART, and Z_{VI} is a normalization constant.

D. SST-RUS

The present work proposes a new automatic spindle detection method that combines SST and RUSBoost, which is referred to as SST-RUS.

Since the length of sleep spindles ranges from 0.5 s to 1.5 s, the extracted features are segmented by using a 0.5 s moving window with a 0.25 s overlap. The maximum, median, and mean values of each of the three features in each segment are used as input variables of a classifier constructed by RUSBoost. That is, the total number of input variables is nine.

The procedure for SST-RUS training is described in Algorithm 2. K is the number of subjects adopted in the training dataset. The input features are extracted from the PSG recording of the kth subject, that is, the EEG data are segmented by a moving window, and the SST-based features are extracted in each EEG segment in steps 2 and 3. In step 4, the SST-based features are arranged as the kth feature matrix $\boldsymbol{X}^{(k)} \in \mathbb{R}^{N_k \times p}$ $(k = 1, \cdots, K)$, where p and N_k are the numbers of input variables and feature samples collected from the kth PSG recording, an $p = 9$ in this research. In step 5, a label vector $\boldsymbol{y}^{(k)} \in \mathcal{Y}^{N_k}$ $(\mathcal{Y} = \{-1, +1\})$ is prepared, of which the sample is annotated as a sleep spindle when more than 75% of one segment is occupied by a sleep spindle according to expert labeling. In this algorithm, the label "+1" denotes a segment with a spindle.

Algorithm 2 Classifier Training

Input: The PSG recordings collected from K subjects, T, $\tilde{c}$.

1: **for** $k = 1, 2, \ldots, K$ **do**
2: Segment the EEG data by the moving window.
3: Apply the SST to the segmented EEG recording.
4: Extract a feature matrix $\boldsymbol{X}^{(k)}$.
5: Construct a label vector $\boldsymbol{y}^{(k)}$.
6: **end for**
7: Merge the matrices $\boldsymbol{X}^{(k)}$ into one matrix $\boldsymbol{X}$.
8: Merge the label vectors $\boldsymbol{y}^{(k)}$ into one matrix $\boldsymbol{y}$.
9: Construct dataset $\mathcal{D} = \{\boldsymbol{X}, \boldsymbol{y}\}$.
10: Train a classifier by RUSBoost with $\mathcal{D}$.

Output: The final classifier h.

In steps 7 - 9, the feature matrices $\boldsymbol{X}^{(k)}$ and the label vectors $\boldsymbol{y}^{(k)}$ $(k = 1, \cdots, K)$ are merged into one feature matrix $\boldsymbol{X} \in \mathbb{R}^{N\times p}$ and one label vector $\boldsymbol{y} \in \mathcal{Y}^N$ respectively, where $N = \sum_{k=1}^{K} N_k$. In addition, the training dataset $\mathcal{D}$ is constructed as $\mathcal{D} = \{\boldsymbol{X}, \boldsymbol{y}\}$. Finally, a classifier is trained with the training dataset $\mathcal{D}$ in step 10.

In order to construct a highly-accurate classifier by RUSBoost, appropriate hyperparameters – the number of iterations T and the ratio of the majority samples in the data $\tilde{c}$ – need to be tuned. This study uses 70% of samples in the whole $\mathcal{D}$ for classifier training, and the other 30% of samples are for searching optimal hyperparameters that minimize misclassification errors.

Since the proposed SST-RUS uses a 0.5 s moving window with a 0.25 s overlap, spindle detection is performed every 0.25 s, and each 0.25 s segment spans two moving windows. When and only when two successive segments are classified as positives, their overlapping part is regarded as a sleep spindle candidate. Finally, successive sleep spindle candidates, whose durations are higher than 0.5 s, are accepted as spindle waveforms because the minimum duration of a sleep spindle is 0.5 s [1].

Algorithm 3 shows a procedure for sleep spindle detection. In steps 1-3, feature extraction is conducted in the same manner as Algorithm 2. Spindle candidates are marked in steps 4 and 5, and, in step 6, it is judged whether or not their durations exceed 0.5 s to output the final result.

III. Case Study

The usefulness of the proposed SST-RUS was validated using an open-access database, Montreal archives of sleep studies cohort 1 (MASS-C1) [21], which consists of five subsets of PSG recordings. This study used only subset 2 (SS2) because it has sleep spindle annotation in the EEG data.

A. Data Description

The SS2 dataset in the MASS-C1 data consists of 19 PSG recordings (eight males and eleven females; 18-33 y.o.), and each recording includes EEG, EOG, ECG, EMG, and

Algorithm 3 Spindle Detection

Input: The PSG recording and classifier h.

1: Segment the EEG data by the moving window.
2: Apply the SST to the EEG recording.
3: Extract a feature matrix $\boldsymbol{X}$.
4: Predict labels $\hat{\boldsymbol{y}}$ using h and $\boldsymbol{X}$.
5: Mark the overlapping parts where both successive segments are classified as positives as spindle candidates.
6: Search the spindle candidates whose successive duration > 0.5 s.

Output: Detected spindles.

TABLE I
THE NUMBER OF SLEEP SPINDLES INCLUDED IN EACH SUBJECTS

Subject	E1	E2	E1 ∩ E2
Sub 1	901	1979	980
Sub 2	1095	1797	1084
Sub 3	130	246	128
Sub 5	314	644	313
Sub 6	133	258	132
Sub 7	840	1291	794
Sub 9	765	1298	756
Sub 10	748	1406	736
Sub 11	563	1020	557
Sub 12	655	923	620
Sub 13	641	996	619
Sub 14	666	1133	645
Sub 17	428	700	422
Sub 18	1075	1390	984
Sub 19	294	475	294

respiratory data. The EEG data were recorded according to the international 10-20 system. An expert identified sleep stages of successive 20 s durations following the Rechtschaffen and Kales (R&K) rules [29], and we analyzed the NREM sleep stage 2 EEG data that contain spindle waveforms.

Two different experts performed manual sleep spindle detection independently using the C3 channel. While expert 1 annotated sleep spindles using the R&K rule, expert 2 adopted a wider frequency band in sleep spindle definition than the R&K rule. Since the second expert scored only 15 out of 19 recordings, subjects 4, 8, 15, and 16 were excluded from this analysis. This work adopts EEG segments agreed upon by both experts to be sleep spindles. The number of the detected sleep spindles in each subject is summarized in Table I, in which $E1$ and $E2$ denote the numbers of sleep spindles detected by experts 1 and 2, and $E1 \cap E2$ is the number of the agreed-upon sleep spindles. The total numbers of $E1 \cap E2$ and non-spindle samples were 9,066 and more than 150,000, respectively, which indicates that spindle detection is a typical imbalanced data problem. Finally, we constructed the dataset $\mathcal{D}_{E1 \cap E2}$ for analysis.

B. Evaluation Criteria

This study evaluated the spindle detection performance using by-event-performance measure [30]. Let E and A denote exact sleep spindle segments defined in Sec. III-A and spindle waveforms detected by the algorithm, respectively, and $l(A \cap E)$ and $l(A \cup E)$ are the lengths of the intersection and the union between E and A, as shown in Fig. 4. Note that multiple overlaps between E and A are not allowed here, that is, one A must be matched to one E.

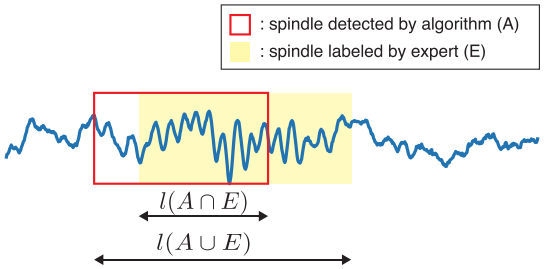

Fig. 4. The intersection and the union between *E* and *A*.

		Classification	
		spindle	non-spindle
Reference	spindle	2576 (True Positive)	817 (False Negative)
	non-spindle	1448 (False Positive)	- (True Negative)

Fig. 5. Confusion matrix of spindle detection by the proposed SST-RUS. True negative was not described here because the number of non-spindle samples was much larger than that of spindle samples.

The degree of overlap between E and A is defined as follows:

$$O_{\mathrm{AE}} = \frac{l(A \cap E)}{l(A \cup E)}. \tag{8}$$

When O_{AE} exceeds a predefined overlap threshold $\bar{O}$, we discriminate A as true positive (TP). Otherwise, E and A are regarded as false negative (FN) and false positive (FP), respectively. In this study, the threshold $\bar{O}$ was set to 0.20 following previous research [27], [30].

The present work adopted three types of performance metrics: sensitivity (SEN), positive predictive value (PPV), and F-measure.

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{9}$$

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{10}$$

$$\mathrm{F} = \frac{2\mathrm{TP}}{\mathrm{FP} + \mathrm{FN} + 2\mathrm{TP}} \tag{11}$$

Specificity was not used in this study because true negatives cannot be defined in the by-event-performance measure.

C. Spindle Detection

In this section, the application results of the proposed SST-RUS to spindle detection in the MASS-C1 data are reported. Eight randomly selected PSG recordings out of fifteen in the SS2 dataset were used for training, and the rest were used for testing. 30% of the training data were split into a hyperparameter tuning dataset in SST-RUS. Thus, we used about 4,000 spindle samples for training and about 1,000 spindle samples for tuning. All of the calculations were repeated ten times for precise evaluation.

The overall performance of the proposed SST-RUS resulted in a sensitivity of 76.9±6.0%, PPV of 61.2±6.4%, and an F-measure of 0.70±0.015 on average. A confusion matrix of this result is shown in Fig. 5, in which true negative was not calculated because the number of non-spindle samples was much larger than that of spindle samples.

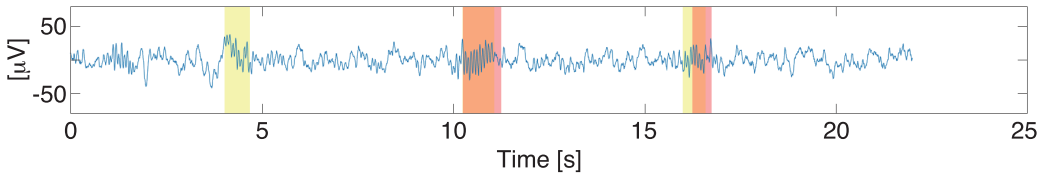

Fig. 6. Spindle detection result. The red-colored and yellow-colored bands denote spindles detected by SST-RUS, and the true sleep spindles, respectively. The spindle waveforms occurred around 11 and 17 s were successfully detected by SST-RUS, and the spindle waveform around 4 s was not detected.

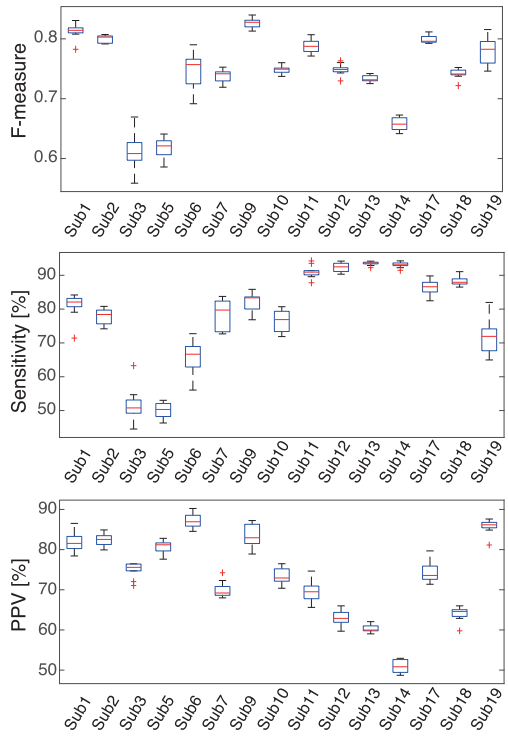

Fig. 7. Spindle detection performances for each subject. SST-RUS achieved good detection performance in almost all subjects except subjects 3, 5, and 14.

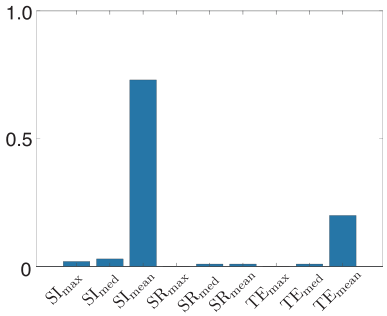

Fig. 8. Variable importance calculated by Eq. (7), showing that $\mathrm{SI}_{\mathrm{mean}}$ was the most important feature for spindle detection.

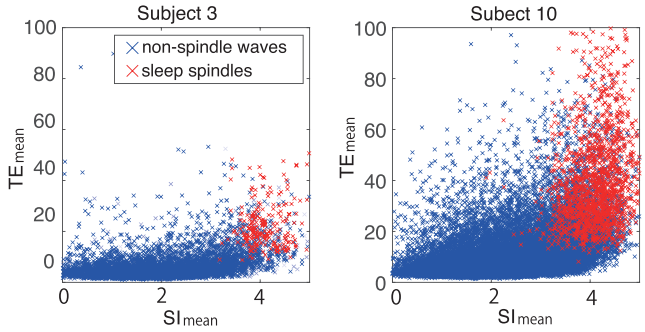

Fig. 9. Scatter plots of SI_{mean} vs. E_{mean} in subjects 3 and 10. The red dots and the blue dots denotes features extracted from sleep spindle and non-spindle waves. The distribution of TE_{mean} varied greatly between two subjects.

An example of a sleep spindle result is shown in Fig. 6. The red-colored bands represent spindles detected by the proposed SST-RUS, and the yellow-colored areas are true sleep spindles. That is, the figure contains one FN and two TP.

Figure 7 illustrates the spindle detection performance of SST-RUS for each subject, showing that the proposed SST-RUS achieved good detection performance in almost all subjects except subjects 3, 5, and 14. This individuality among subjects is discussed in Sec. IV.

IV. Discussion

Overall, the proposed SST-RUS detected about 80% of true sleep spindles when the spindle dataset $\mathcal{D}_{E1\cap E2}$ was adopted; however, spindles in subjects 3, 5, and 14 were not appropriately detected according to Fig. 7. In order to investigate its cause, the variable importance defined by Eq. (7) was calculated, which is shown in Fig. 8.

The variable importance of the mean of the sigma index SI_{mean} was the highest of all variables, while the mean of the Teager energy operator TE_{mean} was the second highest. This result is reasonable since $SI(t)$ was the most sensitive to spindle occurrence, according to Fig. 3. Figure 9 shows the scatter plots of SI_{mean} vs. TE_{mean} of subjects 10 and 3, in whom the proposed method respectively did and did not accurately detect spindles. The red dots and the blue dots represent features extracted from sleep spindle and non-spindle waves. Although the distribution of SI_{mean} did not differ much, the distribution of TE_{mean} varied greatly between two subjects. Although detailed data of subjects 5 and 14 are not described here, the same tendency was observed. Thus, the low value of TE_{mean} might lead to deteriorating the spindle detection performance. The Teager energy operator $TE(t)$ represents an instantaneous change in the EEG amplitude, which becomes high when sleep spindles appear with a large amplitude. This indicates that the amplitude of spindles in subjects 3, 5, and 14 might be small in comparison with the background EEG activities.

TABLE II
SUMMARY OF PERFORMANCES IN THE MASS-C1 DATABASE

Authors	Sensitivity	PPV	F-measure
Kabir *et al.* [26]	74%	58%	0.65
Tsanas *et al.* [22]	83%	16%	
Patti *et al.* [27]	74%		0.69
SST-RUS	77%	61.2%	0.7

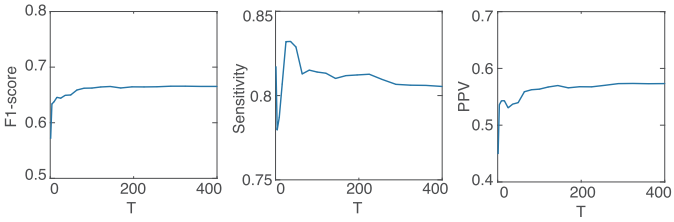

Fig. 10. Effect of the number of CARTs on the detection performance. Although the F-measure and PPV were stabilized from around $T = 200$, the sensitivity slightly deteriorated in the area over $T = 200$.

The spindle detection performance of SST-RUS was compared with various conventional methods using the MASS-C1 data. The Kabir algorithm, which uses SST for feature extraction [26], achieved an F-measure, sensitivity, and PPV of 0.65, 74%, and 58% on average when the detection threshold was optimized so that the F-measure in the training set was maximized. This comparison result clearly illustrates that the proposed SST-RUS achieved better performance than the Kabir algorithm in all performance metrics. Tsanas *et al.* proposed a CWT-based sleep spindle detection algorithm, whose sensitivity and PPV were 83% and 16%, respectively, when the MASS-C1 data was used [22]. This result indicates that the number of FP by the proposed SST-RUS was smaller than that of the Tsanas method. Patti *et al.* combined the sigma index and the sigma ratio with the Gaussian mixture model (GMM) for sleep spindle detection and validated it with the MASS-C1 data. Its detection performance was an F-measure of 0.69 and sensitivity of 74.1% [27]. Thus, the proposed SST-RUS achieved a spindle detection performance superior to the Patti method. These comparison results show the usefulness of the proposed SST-RUS in automatic spindle detection. Table II summarizes the spindle detection performances of various conventional methods when the MASS-C1 data was used.

We investigated the effect of the changes in the number of CARTs T on the spindle detection performance in the proposed SST-RUS. Figure 10 shows the detection performance change of the proposed SST-RUS when T increased from 1 to 400. Although the F-measure and PPV were stabilized from around $T = 200$, the sensitivity slightly deteriorated in the area over $T = 200$. Thus, the number of CARTs T should be tuned carefully for good spindle detection.

The detection performance of RUSBoost was compared with other ensemble machine learning methods – random forest (RF) [31] and AdaBoost [32] – which do not use under-sampling for data balancing. The input features and the parameter tuning manner were the same as the proposed SST-RUS. Figure 11 illustrates the F-measure, the sensitivity, and the PPV of each method. RUSBoost performed the best

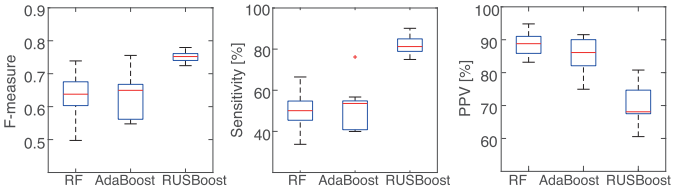

Fig. 11. Performance comparison of RF, AdaBoost, and RUSBoost. RUSBoost achieved the highest F-measure and sensitivity in three methods.

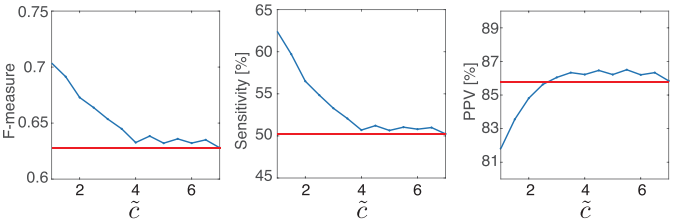

Fig. 12. Spindle detection performance by changing $\tilde{c}$. These results indicate that the F-measure and the sensitivity were the highest when $\tilde{c} = 1$.

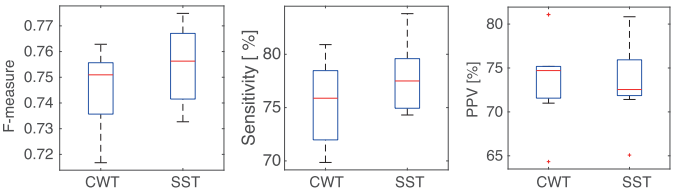

Fig. 13. Spindle detection performances of SST-based and CWT-based features. SST was more appropriate than CWT for spindle detection.

of the three methods from the viewpoint of the F-measure and sensitivity. Because the input features were the same, it is suggested that RUSBoost used in the proposed method is useful as a classifier for detecting sleep spindles.

To confirm the effectiveness of random under-sampling, the spindle detection performance was compared when the ratio of the number of the majority class samples to that of samples in the whole dataset, $\tilde{c}$, was changed. Figure 12 shows changes in spindle detection performances of the proposed method with $\tilde{c} = 1$ - 7. The red line in each plot denotes the AdaBoost performance as a benchmark. The F-measure and the sensitivity were the highest when $\tilde{c} = 1$, which means that the number of the minority class samples used for training equaled to that of the under-sampled majority class samples. Although the PPV was the lowest when $\tilde{c} = 1$, it was not much inferior to AdaBoost, and the performance of the proposed method approached that of AdaBoost as $\tilde{c}$ increased. These results indicate that the numbers of spindle and non-spindle samples should be balanced for spindle detection algorithm construction.

RUSBoost with SST-based features and RUSBoost with standard CWT-based features were compared. The SST-based features performed better than the CWT-based features from the viewpoint of the F-measure and sensitivity, as shown in Fig. 13. Since the sigma index expresses the intensity of the peak power in the spindle frequency range, a sharp time-frequency ridge obtained from SST emphasized the

sigma index when sleep spindles occurred. Thus, SST is more effective than CWT for spindle feature extraction.

This study adopted spindles that two experts agreed to be "true" spindles, constructed the spindle dataset $\mathcal{D}_{E1\cap E2}$, and trained the classifier SST-RUS$_{E1\cap E2}$; however, we constructed another dataset based on spindles labeled by at least one of the experts, $\mathcal{D}_{E1\cup E2}$, and trained SST-RUS using $\mathcal{D}_{E1\cup E2}$, which is referred to as SST-RUS$_{E1\cup E2}$. In $\mathcal{D}_{E1\cup E2}$, the number of spindle samples was 15,817. The performance of SST-RUS$_{E1\cup E2}$ resulted in a sensitivity of 73.5±8.1%, PPV of 71.6±7.8%, and an F-measure of 0.72±0.013. Since the number of spindle samples used for training in $\mathcal{D}_{E1\cup E2}$ was 1.75 times larger than $\mathcal{D}_{E1\cap E2}$, the sensitivity and the F-measure of $\mathcal{D}_{E1\cup E2}$ were improved in comparison with SST-RUS$_{E1\cap E2}$. On the other hand, the PPV slightly deteriorated, which means that the number of false positives increased in SST-RUS$_{E1\cup E2}$. This suggests that many false spindles might be contained in $\mathcal{D}_{E1\cup E2}$. A high-quality spindle dataset is required for accurate spindle detection.

In addition, we applied the proposed SST-RUS to another an open-access database, the DREAMS Spindles Database [33], Although the DREAMS database consists of eight PSG recordings and usually two different experts performed manual sleep spindle detection independently, only one expert performed spindle detection in subjects 7 and 8. Thus, we excluded subjects 7 and 8 from this analysis. This work adopts EEG segments agreed upon by both experts to be sleep spindles, and the total number of spindle waveforms was 180. Four randomly selected PSG recordings out of six in the dataset were used for training, and the rest were used for testing. 30% of the training data were split into a hyperparameter tuning dataset in SST-RUS. All of the calculations were repeated ten times for precise evaluation.

The application result showed a sensitivity of 72.3±8.8%, PPV of 55.6±14.2%, and an F-measure of 0.61±0.074 on average. Although the total performance in the DREAMS dataset was slightly worse than the MASS-C1 dataset and their standard deviations were larger than the MASS-C1 dataset due to the small number of data size, the detection performance of the proposed SST-RUS will be improved through further parameter tuning. Therefore, the result of applying SST-RUS to the DREAMS database showed that the applicability of the proposed method in spindle detection.

The limitations of this study include the data used for analysis, such as the limited number of subjects included in the MASS-C1 data. The proposed SST-RUS should be applied to different PSG data for further validation.

It is concluded that the combination of SST and RUSBoost is effective for spindle detection. The proposed SST-RUS is more promising than conventional spindle detection methods with respect to detection performance. Therefore, the proposed SST-RUS can contribute to mitigating the PSG scoring burden on experts.

V. Conclusion

The present work proposed a new sleep spindle detection method that combines synchrosqueezed wavelet transform (SST) and random under-sampling boosting (RUSBoost). In the proposed method, referred to as SST-RUS, SST extracts spindle features from the segmented EEG data, and RUSBoost discriminates spindles from the extracted features. Since SST decomposes a signal into multiple frequency components with high-resolution, it extracts the features of spindle waveforms, specifically. Also, RUSBoost can detect spindle waveforms adequately because it constructs balanced datasets for binary classification by discarding majority samples appropriately. The proposed SST-RUS does not require any threshold tuning for individual adaptation. The present work evaluated the performance of the proposed SST-RUS through its application to the MASS-C1 data. The mean spindle detection performance of SST-RUS resulted in a sensitivity of 76.9%, PPV of 61.2%, and an F-measure of 0.70, which are higher than the conventional methods. In addition, the proposed SST-RUS was applied to the DREAMS database, showing that the sensitivity, PPV, and the F-measure were 72.3%, 55.6%, and 0.61, respectively. Theses results showed the applicability of the proposed method in spindle detection.

In future works, the proposed SST-RUS will be applied to spindle density visualization, which may contribute to neurological disorder diagnosis.

References

[1] *The AASM Manual for the Scoring of Sleep and Associated Events*, 2nd ed., AASM, Darien, IL, USA, 2018.

[2] V. Latreille *et al.*, "Sleep spindles in Parkinson's disease may predict the development of dementia," *Neurobiol. Aging*, vol. 36, no. 2, pp. 1083–1090, Feb. 2015.

[3] F. Ferrarelli *et al.*, "Thalamic dysfunction in schizophrenia suggested by whole-night deficits in slow and fast spindles," *Amer. J. Psychiatry*, vol. 167, no. 11, pp. 1339–1348, Nov. 2010.

[4] S. L. Wendt *et al.*, "Inter-expert and intra-expert reliability in sleep spindle scoring," *Clin. Neurophysiol.*, vol. 126, no. 8, pp. 1548–1556, Aug. 2015.

[5] D. Coppieters't Wallant, P. Maquet, and C. Phillips, "Sleep spindles as an electrographic element: Description and automatic detection methods," *Neural Plast.*, vol. 2016, Apr. 2016, Art. no. 6783812.

[6] E. Olbrich and P. Achermann, "Analysis of oscillatory patterns in the human sleep EEG using a novel detection algorithm," *J. Sleep Res.*, vol. 14, no. 4, pp. 337–346, Dec. 2005.

[7] E. Huupponen, G. Gómez-Herrero, A. Saastamoinen, A. Värri, J. Hasan, and S.-L. Himanen, "Development and comparison of four sleep spindle detection methods," *Artif. Intell. Med.*, vol. 40, no. 3, pp. 157–170, Jul. 2007.

[8] L. Causa *et al.*, "Automated sleep-spindle detection in healthy children polysomnograms," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 9, pp. 2135–2146, Sep. 2010.

[9] A. Nonclercq, C. Urbain, D. Verheulpen, C. Decaestecker, P. Van Bogaert, and P. Peigneux, "Sleep spindle detection through amplitude–frequency normal modelling," *J. Neurosci. Methods*, vol. 214, no. 2, pp. 192–203, Apr. 2013.

[10] L. B. Ray *et al.*, "Expert and crowd-sourced validation of an individualized sleep spindle detection method employing complex demodulation and individualized normalization," *Frontiers Hum. Neurosci.*, vol. 9, p. 507, Sep. 2015.

[11] D. Gorur, U. Halici, H. Aydin, G. Ongun, F. Ozgen, and K. Leblebicioglu, "Sleep spindles detection using short time Fourier transform and neural networks," in *Proc. IEEE IJCNN*, vol. 2, May 2002, pp. 1631–1636.

[12] N. Acır and C. S. Güzeliş, "Automatic recognition of sleep spindles in EEG by using artificial neural networks," *Expert Syst. Appl.*, vol. 27, no. 3, pp. 451–458, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417404000557

[13] E. M. Ventouras *et al.*, "Sleep spindle detection using artificial neural networks trained with filtered time-domain EEG: A feasibility study," *Comput. Methods Programs Biomed.*, vol. 78, no. 3, pp. 191–207, Jun. 2005.

[14] B. Babadi, S. M. Mckinney, V. Tarokh, and J. M. Ellenbogen, "DiBa: A data-driven Bayesian algorithm for sleep spindle detection," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 2, pp. 483–493, Feb. 2012.

[15] P. Y. Ktonas and E. C. Ventouras, "Automated detection of sleep spindles in the scalp EEG and estimation of their intracranial current sources: Comments on techniques and on related experimental and clinical studies," *Frontiers Hum. Neurosci.*, vol. 8, p. 998, 2014.

[16] Y. Sun, A. K. Wong, and M. S. Kamel, "Classification of imbalanced data: A review," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 4, pp. 687–719, 2009.

[17] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, "Classification with class imbalance problem: A review," *Int. J. Adv. Soft Comput. Appl.*, vol. 7, no. 3, pp. 176–204, 2015.

[18] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.

[19] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: A hybrid approach to alleviating class imbalance," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 1, pp. 185–197, Jan. 2010.

[20] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Jun. 2009.

[21] C. O'Reilly, N. Gosselin, J. Carrier, and T. Nielsen, "Montreal archive of sleep studies: An open-access resource for instrument benchmarking and exploratory research," *J. Sleep Res.*, vol. 23, no. 6, pp. 628–635, Dec. 2014.

[22] A. Tsanas and G. D. Clifford, "Stage-independent, single lead EEG sleep spindle detection using the continuous wavelet transform and local weighted smoothing," *Frontiers Hum. Neurosci.*, vol. 9, p. 181, 2015.

[23] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. New York, NY, USA: Academic, 2008.

[24] I. Daubechies, J. Lu, and H.-T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 243–261, Mar. 2011.

[25] G. Thakur, E. Brevdo, N. S. Fučkar, and H.-T. Wu, "The synchrosqueezing algorithm for time-varying spectral analysis," *Signal Process.*, vol. 93, no. 5, pp. 1079–1094, 2013.

[26] M. M. Kabir, R. Tafreshi, D. B. Boivin, and N. Haddad, "Enhanced automated sleep spindle detection algorithm based on synchrosqueezing," *Med. Biol. Eng. Comput.*, vol. 53, no. 7, pp. 635–644, Jul. 2015.

[27] C. R. Patti, T. Penzel, and D. Cvetkovic, "Sleep spindle detection using multivariate Gaussian mixture models," *J. Sleep Res.*, vol. 27, no. 4, Aug. 2018, Art. no. e12614.

[28] W.-Y. Loh, "Classification and regression trees," *WIREs Data Mining Knowl. Discovery*, vol. 1, no. 1, pp. 14–23, Jan. 2011.

[29] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Arch. Gen. Psychiatry*, vol. 20, no. 2, pp. 246–247, 1968.

[30] S. C. Warby *et al.*, "Sleep-spindle detection: Crowdsourcing and evaluating performance of experts, non-experts and automated methods," *Nature Methods*, vol. 11, no. 4, pp. 385–392, Apr. 2014.

[31] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[32] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. 13th Int. Conf. Int. Conf. Mach. Learn. (ICML)*, San Francisco, CA, USA, 1996, pp. 148–156.

[33] S. Devuyst. *The DREAMS Dataset*. [Online]. Available: http://www.tcts.fpms.ac.be/~devuyst/Databases/DatabaseSpindles