RAPIDEST: A Framework for Obstructive Sleep Apnea Detection

Xin-Xue Lin, Phone Lin, *Fellow, IEEE*, En-Hau Yeh, Gi-Ren Liu, Wan-Ching Lien, and Yuguang Fang, *Fellow, IEEE*

***Abstract*— The traditional polysomnography (PSG) examination for Obstructive Sleep Apnea (OSA) diagnosis needs to measure several signals, such as EEG, ECG, EMG, EOG and the oxygen level in blood, of a patient who may have to wear many sensors during sleep. After the PSG examination, the Apnea-Hypopnea Index (AHI) is calculated based on the measured data to evaluate the severity of apnea and hypopnea for the patient. This process is obviously complicated and inconvenient. In this paper, we propose an AI-based framework, called RAre Pattern Identification and DEtection for Sleep-stage Transitions (RAPIDEST), to detect OSA based on the sequence of sleep stages from which a novel rarity score is defined to capture the unusualness of the sequence of sleep stages. More importantly, under this framework, we only need EEG signals, thus significantly simplifying the signal collection process and reducing the complexity of the severity determination of apnea and hypopnea. We have conducted extensive experiments to verify the relationship between the rarity score and AHI and demonstrate the effectiveness of our proposed approach.**

***Index Terms*— Apnea-Hypopnea Index (AHI), EEG signal, machine learning, Obstructive Sleep Apnea (OSA).**

Manuscript received 11 January 2022; revised 2 June 2022 and 27 October 2022; accepted 18 November 2022. Date of publication 24 November 2022; date of current version 31 January 2023. The work of Phone Lin was supported in part by the Ministry of Science and Technology (MOST) of Taiwan under Grant MOST 111-2221-E-002-126-MY3, Grant MOST 110-2622-8-002-018, Grant MOST 110-2221-E-002-189, and Grant MOST 109-3111-8-002-002; and in part by the Ministry of Education Higher Education Sprout Project under Grant NTU-111L880603. *(Corresponding author: Phone Lin.)*

Xin-Xue Lin and En-Hau Yeh are with the Department of Computer Science and Information Engineering, National Taiwan University, Taipei 10617, Taiwan (e-mail: xinxue@pcs.csie.ntu.edu.tw; enhauyeh@pcs.csie.ntu.edu.tw).

Phone Lin is with the Internet of Things Research Center, the Department of Computer Science and Information Engineering, the Graduate Institute of Networking and Multimedia, and the Institute of Medical Device and Imaging, National Taiwan University, Taipei 10617, Taiwan (e-mail: plin@csie.ntu.edu.tw).

Gi-Ren Liu is with the Department of Mathematics, National Cheng Kung University, Tainan City 701, Taiwan (e-mail: girenliu@mail.ncku.edu.tw).

Wan-Ching Lien is with the Department of Emergency Medicine, National Taiwan University Hospital, Taipei 100229, Taiwan, and also with the Department of Emergency Medicine, College of Medicine, National Taiwan University, Taipei 10617, Taiwan (e-mail: wanchinglien@ntu.edu.tw).

Yuguang Fang is with the Department of Computer Science, City University of Hong Kong, Hong Kong, China (e-mail: my.fang@cityu.edu.hk).

Digital Object Identifier 10.1109/TNSRE.2022.3224474

I. Introduction

THE Obstructive Sleep Apnea (OSA) is a disorder, where a person experiences periodic disruptions in breathing during sleep. OSA may result in major health implications, such as excessive daytime sleepiness, nonrestorative sleep, memory impairment, depression, and serious cardiac arrhythmias [1]. For the diagnosis of sleep apnea and hypopnea, each patient (or the examinee) has to sleep in a laboratory setting and wears various devices to monitor the electroencephalography (EEG), the electrooculogram (EOG), the electromyogram (EMG), the electrocardiography (ECG), the oxygen level in blood, heart rate and breathing pattern, as well as eye and leg movements during sleep [2]. This process is called the overnight polysomnography (PSG) examination. Obviously, the signal process is very complicated and inconvenient to patients.

After the PSG examination, the technicians calculate the Apnea-Hypopnea Index (AHI) [3], which is used to evaluate the severity of apnea and hypopnea: (1) Normal: AHI $<$ 5; (2) Mild sleep apnea: 5 $\leq$ AHI $<$ 15; (3) Moderate sleep apnea: 15 $\leq$ AHI $<$ 30; and (4) Severe sleep apnea: AHI $\geq$ 30. AHI is obtained by dividing the number of apnea and hypopnea events by the number of hours of sleep, where an apnea event is defined as the absence of airflow for at least 10 seconds, and a hypopnea event is a 30% or greater decrease in airflow lasting at least 10 seconds with oxygen desaturation of at least 4%. The details of the AHI evaluation can be found in [3]. Based on the standard definition [4], when a patient has AHI $\geq$ 15, the patient is known as an OSA patient.

However, the long-term monitoring of upper airway airflows may make patients feel uncomfortable and inconvenient. The AHI evaluation heavily relies on visualized pattern recognition by experts or doctors, which is time consuming and requires effort [5]. Hence, how to extract features from more accessible physiological signals to infer the upper airway airflows has attracted intensive attention [6].

Previous works proposed algorithms to automatically detect OSA based on different signal measurements without involving experts or doctors. The signal measurements include airflow sensing signals [7], [8], rib cage movements together with abdomen movements [9], blood oxygen saturation (SaO2) [10], EMG together with EOG [11], ECG [12], and EEG [13]. In the following, we have a literature survey for the existing works that used these signals for OSA detection.

Obviously, the most intuitive way to detect OSA is to use the pattern of the respiratory signals and oxygen desaturation index. For example, based on the airflow signal measured overnight, Lakhan et al. [7] proposed a Deep Neural Network (DNN)-based method to classify the Sleep Apnea-Hypopnea Syndrome (SAHS) patients into 4 classes: no SAHS, mild SAHS, moderate SAHS, and severe SAHS. One of the typical sensors to measure the airflow of a patient is the airflow thermal thermistor. Behbehani et al. [8] designed an automatic positive airway pressure (APAP) device, which allows a patient to set an initial pressure while awake, and automatically elevates mask pressure during sleep when OSA is detected. However, a patient may not feel comfortable to wear these kinds of sensors while sleeping. Abdomen and rib cage movements can be acquired by Respiratory Inductive Plethysmography (RIP), which includes two belts to obtain both abdomen and rib cage signals. Staats et al. [9] used rib cage and abdomen motion obtained from PSG examination signals to classify the type of disorder breathing event including obstructive apnea, obstructive hypopnea, mixed apnea, and central apnea. The blood oxygen saturation (SaO2) can be measured on the finger using a pulse oximeter directly. Alvarez et al. [10] used SaO2 to classify the subjects into normal or OSA. The EOG signals can be measured by pairs of electrodes placed above and below the eye. The EMG signals can be obtained by attaching small disk electrodes to the skin surface over the muscle. Kalevo et al. [11] used EEG, EOG, and EMG signals to identify the four OSA severity levels (i.e., normal, mild OSA, moderate OSA, and severe OSA) in home sleep apnea testing. The ECG signals can be collected by ten electrodes placed on the patient's limbs and the surface of the chest. Hilmisson et al. [12] used ECG signals to classify the subjects into four OSA severity levels.

The measurement of EEG signals from a patient can be easily done through small electrodes attached to the scalp [14], and it is more comfortable for a patient to wear small electrodes during sleep. There also exist portable or wireless small electrodes [15], which can easily acquire EEG signals at home. However, most of the existing works (e.g., [16], [17], [18], [19], and [20]) focused on using the EEG signal to identify sleep apnea events during the overnight PSG examination, which cannot be applied to detect whether a patient has OSA or not according to the standard definition [4]. There are seldom works that used the EEG signals for OSA detection.

We discuss the following existing work that used EEG signal for OSA detection: In [13], based on the EEG signal and the corresponding sleep stages (manually labeled by the sleep medicine physicians), the authors extracted the Analysis of Brain Recurrence (ABR) features, and then applied Fisher's Linear Discriminant Analysis (LDA) to classify the patients into mild OSA (i.e., with $5 \leq \text{AHI} < 15$) or moderate OSA (i.e., $16 < \text{AHI} \leq 30$). However, in the approach in [13], it requires to determine the sleep stages by the sleep medicine physicians. The process is time-consuming and labor-intensive, which results in that the proposed approach in [13] cannot be applied for online OSA detection.

Most of the signals discussed above (e.g., EEG, ECG, SaO2, EMG, and EOG) can be obtained from the sensors attached on the patients (i.e., wearing-comfortability). However, among these different signals, only the EEG signal can be used not only for OSA detection but also for sleep stage classification (to be elaborated later). That is, the EEG signal can be used to evaluate sleep stages and sleep quality [16]. Other signals can be used only for OSA detection. For the above reasons, we are motivated to identify whether there exists any relationship between the sequence of sleep stages and OSA.

The EEG signals recorded during the overnight can be utilized to identify what sleep stage a patient experiences during sleep. The American Academy of Sleep Medicine (AASM) classifies the sleep stages into five categories [21]: Awake, Rapid-Eye-Movement (REM), N1, N2, and N3. The proportion of the sleep entering the N3 and REM stages has been proved to be associated with the physical and mental health [22], particularly in fatigue recovery [23], memory consolidation [24] and emotional distress [25]. Based on the changes in sleep stages, doctors can evaluate the effect of the medical treatment on the quality of sleep, which could serve as a reference for the subsequent treatment plan. The identification of the sleep stages using the signals obtained from the PSG examination has been discussed in many previous works [26], [27], [28]. AI technologies have also been widely adopted to classify the sleep stages using the EEG signals. It has been shown [29], [30], [31], [32] that AI technologies can significantly improve the accuracy of the sleep stage classification. For example, using the dataset provided by St. Vincent's University Hospital and University College Dublin (UCDDB), Zhang and Wu [33] showed that the overall accuracy of the sleep stage classification could reach 90.8% when using a convolutional neural network (CNN).

In this paper, based on EEG signals, we propose an AI-based framework, namely RAre Pattern Identification and DEtection for Sleep-stage Transitions (RAPIDEST), to analyze the sequence of sleep stages for OSA detection. We discover that there exists rare patterns indeed in the sequence of sleep stages. The rare pattern appears in the sequence of sleep stages of healthy patients with lower probability. On the other hand, the rare pattern appears in the sequence of sleep stages of OSA patients with higher probability. With this assumption, we define a scoring function to evaluate the unusualness of a sequence of sleep stages for a patient, based on which we obtain a novel rarity score for a patient. For OSA detection, we use the rarity score to detect OSA for a patient. In the proposed framework, we treat the OSA detection as a semi-supervised anomaly detection problem. That is, based on a training set of only normal data, we identify anomalies for the future. We only use data collected from healthy patients in the training phase, and then we classify the patients into healthy or OSA classes in the testing phase. To the best of our knowledge, no prior studies have examined semi-supervised anomaly detection for OSA classification. In this way, we can significantly reduce the complexity and overhead for the severity determination of apnea and hypopnea.

The remainder of this paper is organized as follows. In Section II, we introduce the proposed CNN-based method for sleep stage classification and the proposed framework RAPIDEST for rarity evaluation of sequence of sleep stages

and OSA detection. Section III conducts the performance of the proposed algorithm. We conclude this paper in Section IV.

II. RAPIDEST: RAre Pattern Identification and DEtection for Sleep-Stage Transitions

In this section, we propose our AI-based framework, RAre Pattern Identification and DEtection for Sleep-stage Transitions (RAPIDEST), to determine the rarity score for OSA detection. The execution of the proposed RAPIDEST consists two phases: the training phase and online detection phase.

The training phase exercises as follows: Given a training dataset $\mathbf{D}_T$ containing N_t patients, their corresponding EEG signals, and their AHI values. The EEG signals for each patient are already labeled with a sequence of sleep stages by experts. Suppose that among the N_t patients, there are $N_{t,h}$ healthy patients and $N_{t,o}$ OSA patients, i.e., $N_t = N_{t,h} + N_{t,o}$. The training phase consists of the following two steps, and the details can be found in Sections II-A and II-B, respectively.

Step T-1: Sleep Stage Classification. We randomly select $n_{t,h,1}$ healthy patients and $n_{t,o,1}$ OSA patients from the dataset $\mathbf{D}_T$ (where $n_{t,h,1} < N_{t,h}$ and $n_{t,o,1} < N_{t,o}$). The data of the $n_{t,h,1}$ healthy patients and the $n_{t,o,1}$ OSA patients forms a subset $\mathbf{d}_1 \subset \mathbf{D}_T$. In this step, we use a Convolutional Neural Networks (CNN) to build up a sleep stage classifier using the EEG signals of the subset $\mathbf{d}_1$.

Step T-2: OSA Detection Model. Let the subset $\mathbf{d}_2 = \mathbf{D}_T - \mathbf{d}_1$. For every patient in the subset $\mathbf{d}_2$, we transform the original EEG signals to a sequence of sleep stages using the sleep stage classifier built in Step T-1. Then based on the sequences of sleep stages of the patients in $\mathbf{d}_2$, we build up the OSA detection model (that consists of a sleep pattern table with two parameters, θ and η).

As shown in Fig. 1 (a), the online detection phase exercises as follows: Given a testing dataset $\mathbf{D}_D$ containing N_d patients, their corresponding EEG signals, and their AHI values, $\mathbf{D}_D \cap \mathbf{D}_T = \emptyset$. Suppose that among the N_d patients, there are $N_{d,h} > 0$ healthy patients and $N_{d,o} > 0$ OSA patients, i.e., $N_d = N_{d,h} + N_{d,o}$. In this phase, we execute the following two steps to determine whether a patient in $\mathbf{D}_D$ has OSA or not, based on his original EEG signals E.

Step D-1. We transform the EEG signals E to a sequence of sleep stages $\mathbf{X}_E$ using the sleep stage classifier built in Step T-1.

Step D-2. Based on the OSA detection model built in Step T-2, by referencing the sleep pattern table, we calculate a rarity score for the patient by scanning the sequence of the sleep stages $\mathbf{X}_E$. We determine the patient as a healthy patient if his rarity score is smaller than η. Otherwise (i.e., his rarity score is larger than or equal to η, the patient is determined as an OSA patient.

A. Step T-1: Sleep Stage Classification

In this section, we elaborate on the sleep stage classification in our RAPIDEST. In [29], [30], and [34], the authors applied neural networks to perform feature learning for the classification of the five sleep stages. The overall accuracy was demonstrated to be 74.8% using the Sleep-EDF SC* dataset. In [29], Supratak et al. proposed a deep learning model, DeepSleepNet, where Convolutional Neural Networks (CNN) is applied to extract features, and bidirectional Long Short-Term Memory (LSTM) is used to learn the stage transition rules from EEG epochs. The DeepSleepNet algorithm could achieve the state-of-the-art 82.0% of overall accuracy using the Sleep-EDF SC* dataset. In this paper, we also use a CNN model as the classification algorithm. Through the experiments, we justify that the CNN model does function well, and show that our CNN model has less computation complexity.

During sleep, a patient experiences at different sleep stages as time goes on. The transitions of sleep stages form a sleep stage transition sequence. A sleep stage can be determined by using a 30-second EEG signal (also known as "30-s epoch"). We set the length of a time slot as 30 seconds. Let $T^{(i)}$ be the total number of time slots during a sleep period for patient $\mathbf{u}_i$. To simplify our discussion, we denote the sleep stages Awake, REM, N1, N2, and N3, as **A**, **R**, **1**, **2**, and **3**, respectively. The set of sleep stages is denoted as $\mathcal{S} = \{\mathbf{A}, \mathbf{R}, \mathbf{1}, \mathbf{2}, \mathbf{3}\}$. We denote the sequence of sleep stages during sleep for patient $\mathbf{u}_i$ as $X^{(i)} = \langle x_1^{(i)}, x_2^{(i)}, \ldots, x_{T^{(i)}}^{(i)} \rangle = \langle x_t^{(i)} \rangle_{t=1}^{T^{(i)}}$, where $x_t^{(i)} \in \{\mathbf{A}, \mathbf{R}, \mathbf{1}, \mathbf{2}, \mathbf{3}\}$, and t is the index of the time slot from 1 to $T^{(i)}$.

Fig. 1 (b) shows the EEG signal measured from a patient during sleep for eight hours. The whole EEG signal are divided into multiple records of 30-second period, i.e., the time is divided into multiple 30-s epoches. In the dataset $\mathbf{d}_1$, the EEG signals of every 30-second record are labeled with a stages. The label is the ground truth for the sleep stage classification. The sleep stage classification processes the EEG signals of all patients $\mathbf{d}_1$ by executing the following three steps as shown in Fig. 2:

Step T-1-1: Feature Extraction (see ① in Fig. 2). We transform each 30-s epoch to two types of features including local features and global features. The details can be found in Section II-A.1.

Step T-1-2: Feature Normalization (see ② in Fig. 2). We normalize the extracted features to build up the CNN model, as shown in Section II-A.2.

Step T-1-3: Sleep Stage Classification (see ③ in Fig. 2). We use the CNN model to classify the sleep stages at each time slot. Then we can generate the sequences of sleep stages by the results of the classification, as shown in Section II-A.3.

The details for the three sub-steps are given in Sections II-A.1, II-A.2 and II-A.3, respectively.

1) Step T-1-1: Feature Extraction: The feature extraction in our proposed framework is a general procedure widely used in CNN-based models in the previous works (e.g., [35]). It includes the local feature extraction and the global feature extraction. In general, the local feature is the input of each kernel function in the convolutional layer, and the global feature is the input of the fully-connected layer. Thus, we extract the local features for each shorter time window. The global

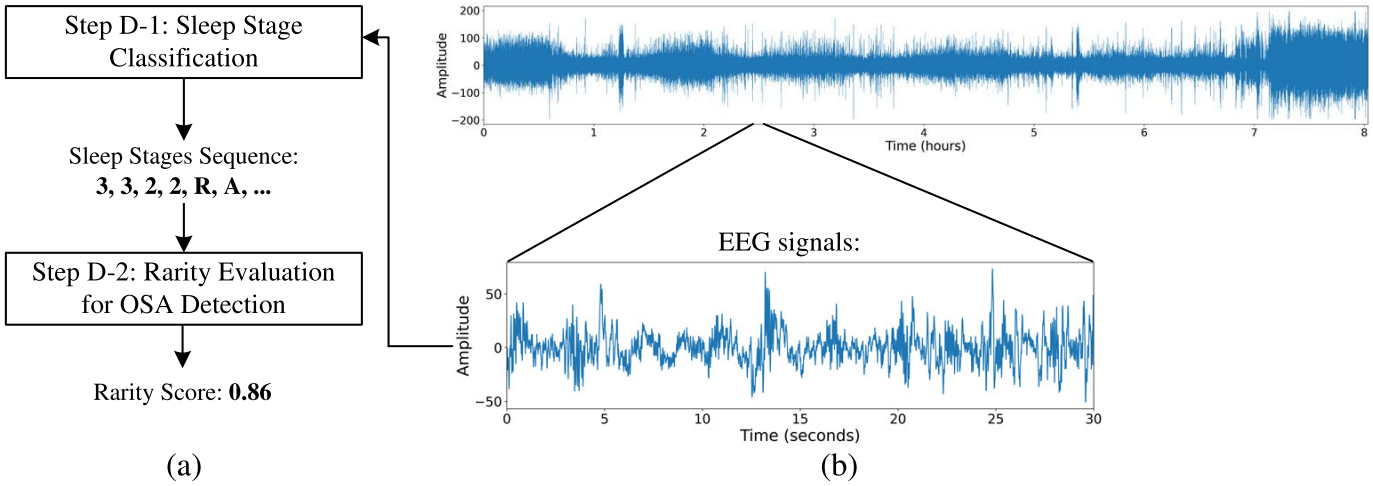

Fig. 1. (a) The online detection flow chart of our proposed RAPIDEST; (b) The EEG signal for a patient recorded for one night.

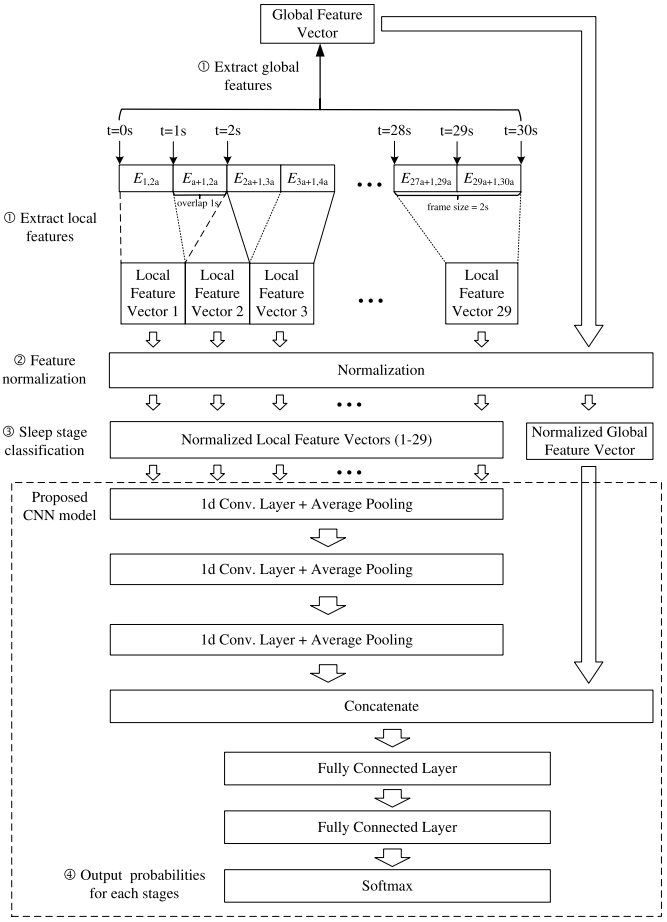

Fig. 2. Feature extraction and CNN architecture.

feature is the statistical results of the whole 30s EEG signals, which does not contain the time information of the original signals. Let $\mathbb{E}$ be the set of sampling points in an 30-s epoch EEG signal, and let the sampling rate for the EEG signals be a Hz. There are total $30 \times a$ sample points. We denote $\mathbf{e}_i$ as the i-th sampling point in $\mathbb{E}$. Let $\mathbf{E}_{i,j}$ (where $i \leq j$) be the subset of $\mathbb{E}$, i.e., $\mathbf{E}_{i,j}$ contains $\mathbf{e}_i$, $\mathbf{e}_{i+1}$, ..., $\mathbf{e}_j$.

In the local feature extraction, we apply the sliding window to scan the EEG signal in a 30-s epoch as shown in ① in Fig. 2. A 30-s epoch EEG signal is further divided into 30 1-second frames. We scan the 30-s epoch EEG signal frame by frame, and set a local window with window size ω_l equal to 2 frames. We can obtain 29 local windows, i.e., $\mathbf{E}_{1,2a}$, $\mathbf{E}_{a+1,3a}$, $\mathbf{E}_{2a+1,4a}$, ..., $\mathbf{E}_{28a+1,30a}$.

For the EEG signals in each window $\mathbf{E}_{x,y}$ (where $x = ia + 1$, $y = (i+2)a$, and $i = 0, 1, 2, 3, \ldots, 28$), we use the Hamming window function [31] (denoted as $\mathcal{H}(\mathbf{X})|_{\mathbf{X}=\mathbf{E}_{x,y}}$) to reduce spectral leakage, and then apply the Fast Fourier Transform (denoted as $\mathcal{F}(\mathbf{Y})|_{\mathbf{Y}=\mathcal{H}(\mathbf{X})}$) to each window. Then for $\mathbf{X} = \mathbf{E}_{x,y}$, we obtain a vector of $\lfloor \frac{y-x+1}{2} \rfloor$ dimensions, e.g., $V_{x,y} = \mathcal{F}(\mathcal{H}(\mathbf{X}))|_{\mathbf{X}=\mathbf{E}_{x,y}}$ for each window, and $V_{x,y} = (v_1, v_2, v_3, \ldots, v_{\lfloor \frac{y-x+1}{2} \rfloor})$. Then we take the square of the magnitude to estimate the corresponding power spectral densities

$$(\|v_1\|^2, \|v_2\|^2, \ldots, \|v_{\lfloor \frac{y-x+1}{2} \rfloor}\|^2) \tag{1}$$

for each window, which is used as the local feature for each window. For a 30-s epoch EEG signal, there are total 29 corresponding local feature vectors.

Following [36] and [37], from a 30-s epoch EEG signals, we extract the features as shown in ② in Fig. 2, including power spectral density [31], coefficients of AutoRegressive (AR) model [37] and statistical features [36].

We obtain the power spectral density by using the same procedure as that in local feature extraction. We extend the widow size to 30 frames, i.e., a window contains $\mathbf{E}_{1,30a}$, and obtain the vector $V_{1,30a} = \mathcal{F}(\mathcal{H}(X))|_{X=\mathbf{E}_{1,30a}}$, and the corresponding power spectral densities $(\|v_1\|^2, \|v_2\|^2, \ldots, \|v_{15a}\|^2)$.

The coefficients of the AR model can be obtained as follows. Let p be the order of the AR model, $\hat{\mathbf{e}}_t$ be the prediction value of $\mathbf{e}_t$ (where $p+1 \leq t \leq 30a$) based on $\mathbf{e}_{t-i}$ (where $i = 1, 2, \ldots, p$). Then we define the AR model with coefficients α_i and β as $\hat{\mathbf{e}}_t = \sum_{i=1}^{p} \alpha_i \mathbf{e}_{t-i} + \beta$ where α_i and β are the coefficients of the AR model. We attempt to find the coefficients α_i and β such that $(\hat{\mathbf{e}}_t - \mathbf{e}_t)^2$ is minimized. Based on the Bayesian Information Criterion [38], we set the

order of the AR model p to 28. There are total 29 coefficients of AR used as global features.

The four standard statistical features, including mean, standard deviation, minimum and maximum, are also used as the global features, which are obtained by $\mathsf{mean}(\mathbb{E}) = \frac{1}{30a}\sum_{i=1}^{30a}\mathbf{e}_i$; $\mathsf{std}(\mathbb{E}) = \sqrt{\frac{1}{30a-1}\sum_{i=1}^{30a}(\mathbf{e}_i - \mathsf{mean}(\mathbb{E}))^2}$; $\min(\mathbb{E}) = \min_{1\le i\le 30a}(\mathbf{e}_i)$; $\max(\mathbb{E}) = \max_{1\le i\le 30a}(\mathbf{e}_i)$.

To summarize, for a 30-s EEG signal, we extract the global features including $15a$ power spectral densities, 29 coefficients of AR, and 4 statistic features. Then we combine those features as one global feature vector:

$$(\|v_1\|^2, \|v_2\|^2, \ldots, \|v_{15a}\|^2, \alpha_1, \ldots, , \alpha_{28}, \beta, \mathsf{mean}(\mathbb{E}), \mathsf{std}(\mathbb{E}), \min(\mathbb{E}), \max(\mathbb{E})) \quad (2)$$

2) Step T-1-2: Feature Normalization: We perform the feature standardization [39] on each feature vector, including 29 local feature vectors (see Eq. (1)) and a global feature vector (see Eq. (2)).

3) Step T-1-3: Sleep Stage Classification: In this section, we propose our Convolutional Neural Networks (CNN) architecture (Fig. 2) for the sleep stage classification, which consists of three convolutional layers and three fully-connected (FC) layers. The input to the proposed CNN as shown in ③ in Fig. 2 includes 29 normalized local feature vectors (i.e., Eq. (1)) and one normalized global feature vector (i.e., Eq. (2)). The numbers of filters of the first three convolutional layers are 64, 128, and 256, respectively, and the filter sizes are 8, 4, and 2, respectively, with the same stride size equal to 1. Two max-pooling layer are followed by the first two convolutional layer with pooling size of 2. An average pooling layer is followed by the last convolutional layer. The output vector of the average pooling layer and the normalized global feature vector are concatenated to a new vector, and the three FC layers are followed by the concatenated vector. The numbers of neurons of the three FC layers are 512, 256, and 5, respectively. Finally, the softmax activation function is applied in the last FC layer to output the probabilities of the five sleep stages. Following [40], the scaled exponential linear units (SeLU) is applied to the output of every convolutional layer and FC layer. The alpha dropout [40] is applied to the output of every pooling layer and first two FC layers with dropout rate 0.25, 0.25, 0.25, 0.25, and 0.5, respectively. The weights of the convolutional layer and FC layer use the same random initialization described in [40]. For the hyperparameter selection, we follow the setups in [29] and [41] and try many different setups (including the number of layers, the number of filters, the filter size, the learning rate, and the activation function). We randomly select a patient from the training dataset $\mathbf{d}_1$ for validation, and we fine-tune the hyperparameters based on the loss of the validation. Due to the page limitation, we omit the details of the setups, and only present the results of the best performance.

B. Step T-2: OSA Detection Model

Given a training dataset $\mathbf{d}_2$, there are $n_{t,h,2} = N_{t,h} - n_{t,h,1}$ healthy patients (who form the subset $\mathbf{d}_{2,h}$) and $n_{t,o,2} = N_{t,o} - n_{t,o,1}$ OSA patients (who form the subset $\mathbf{d}_{2,o}$).

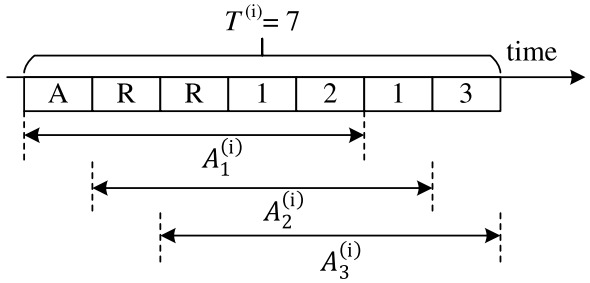

Fig. 3. An example of the sleep patterns.

To simplify our description, we suppose that $\mathbf{d}_{2,h}$ consists of N patients (i.e., $N = n_{t,h,2}$). We perform the sleep stage classification using the sleep stage classifier built in Step T-1, to obtain the sequence of the sleep stages for each patient in $\mathbf{d}_{2,h}$. Then we have the data set $H_X = \{X^{(1)}, X^{(2)}, \ldots, X^{(N)}\}$, where $X^{(i)}$ is the sequence of the sleep stages for patient $\mathbf{u}_i$. Define a sleep pattern as a subsequence of $X^{(i)}$ of length L, denoted as $A_k^{(i)} = \langle x_t^{(i)}\rangle_{t=k}^{k+L-1}$, where $t = k$ is the start index of the subsequence, and $t = k + L - 1$ is the end index of the subsequence. There are a total of $T^{(i)} - L + 1$ sleep patterns in $X^{(i)}$. Denote the set $\mathcal{A}_L(X^{(i)})$ of all sleep patterns of length L in the sequence $X^{(i)}$ as $\mathcal{A}_L(X^{(i)}) = \{A_1^{(i)}, A_2^{(i)}, \ldots, A_{T^{(i)}-L+1}^{(i)}\}$. Then we have the set $H_A = \bigcup_{h\in H_X}\mathcal{A}_L(h)$, which contains all sleep patterns of length L in H_X. Fig. 3 shows an example for the sleep patterns. In this figure, for the sequence of sleep stages $X^{(i)} = \langle \mathbf{A}, \mathbf{R}, \mathbf{R}, \mathbf{1}, \mathbf{2}, \mathbf{1}, \mathbf{3}\rangle$ with $T^{(i)} = 7$, there are 3 sleep stage patterns of length $L = 5$, namely, $A_1^{(i)} = \langle \mathbf{A}, \mathbf{R}, \mathbf{R}, \mathbf{1}, \mathbf{2}\rangle$, $A_2^{(i)} = \langle \mathbf{R}, \mathbf{R}, \mathbf{1}, \mathbf{2}, \mathbf{1}\rangle$, and $A_3^{(i)} = \langle \mathbf{R}, \mathbf{1}, \mathbf{2}, \mathbf{1}, \mathbf{3}\rangle$.

We apply the lookahead pair method [42] to define the rare sleep patterns. A "lookahead" pair $\langle x, y\rangle_k$ exists if there is a sleep pattern $\boldsymbol{a} = \langle a_t\rangle_{t=1}^{L} \in H_A$ such that the two sleep stages $x, y \in S$ appears in a, and y occurs at the k-th location after x. In other words, given a sleep pattern $\boldsymbol{a} = \langle a_t\rangle_{t=1}^{L} \in H_A$, $\langle x, y\rangle_k$ exists if the following rule holds:

$$\exists m, n, k \in \mathbb{N}, 1 \le m, n, k \le L, \ni x = a_m, y = a_n, k = n - m. \quad (3)$$

Given a sleep pattern $\boldsymbol{a}$, let $\mathcal{B}_{lo}(\boldsymbol{a})$ be the set containing all lookahead pairs by checking Eq. (3), i.e.,

$$\mathcal{B}_{lo}(\boldsymbol{a}) = \{\langle a_m, a_n\rangle_k : \exists m, n, k \in \mathbb{N} \text{ such that } 1 \le m, n, k \le L \text{ and } k = n - m\} \quad (4)$$

Fig. 4 shows an example, where all lookahead pairs in a sleep pattern $\langle \mathbf{W}, \mathbf{R}, \mathbf{W}, \mathbf{R}\rangle$ are $\langle \mathbf{W}, \mathbf{R}\rangle_1$, $\langle \mathbf{R}, \mathbf{W}\rangle_1$, $\langle \mathbf{W}, \mathbf{W}\rangle_2$, $\langle \mathbf{R}, \mathbf{R}\rangle_2$, and $\langle \mathbf{W}, \mathbf{R}\rangle_3$. Then we have

$$\mathcal{B}_{lo}(\langle \mathbf{W}, \mathbf{R}, \mathbf{W}, \mathbf{R}\rangle) = \{\langle \mathbf{W}, \mathbf{R}\rangle_1, \langle \mathbf{R}, \mathbf{W}\rangle_1, \langle \mathbf{W}, \mathbf{W}\rangle_2, \langle \mathbf{R}, \mathbf{R}\rangle_2, \langle \mathbf{W}, \mathbf{R}\rangle_3\}$$

We search the sleep stage sequences in the H_A to find all lookahead pairs in H_A. Then we count the number (denoted by $C(\langle x, y\rangle_k, H_A)$) of occurrences of sleep patterns containing lookahead pair $\langle x, y\rangle_k$ in H_A by:

$$C(\langle x, y\rangle_k, H_A) = |\{\boldsymbol{a} : \boldsymbol{a} \in H_A \text{ and } \langle x, y\rangle_k \in \mathcal{B}_{lo}(\boldsymbol{a})\}| \quad (5)$$

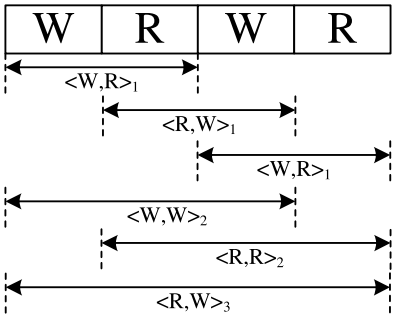

Fig. 4. An example of all lookahead pairs in the sleep pattern $\langle \mathbf{W}, \mathbf{R}, \mathbf{W}, \mathbf{R} \rangle$.

For example, $C(\langle \mathbf{W}, \mathbf{R} \rangle_1, \{\langle \mathbf{W}, \mathbf{R}, \mathbf{W}, \mathbf{R} \rangle, \langle \mathbf{W}, \mathbf{W}, \mathbf{W}, \mathbf{W} \rangle\}) = |\{\langle \mathbf{W}, \mathbf{R}, \mathbf{W}, \mathbf{R} \rangle\}| = 1$. We maintain a lookahead pair table that stores the numbers of occurrences for all lookahead pairs $\langle x, y \rangle_k$ in $\bigcup_{\boldsymbol{a} \in H_A} \mathcal{B}_{lo}(\boldsymbol{a})$.

Based on the lookahead table with a predefined threshold θ ($0 \leq \theta < 1$), we identify whether a sleep pattern $\boldsymbol{x}$ is a rare sleep pattern or not, and the result is stored in the flag $f(\boldsymbol{x})$. A rare sleep pattern is one that occurs in the sequences of sleep stages with low probability. We design and implement the following function $f_r(\cdot)$ for the rare sleep pattern identification. Given a sleep pattern $\boldsymbol{x}$ and the threshold θ,

$$f_r(\boldsymbol{x}) = \begin{cases} 1, & \text{if } |\{l : l \in \mathcal{B}_{lo}(\boldsymbol{x}) \text{ and } \frac{C(l, H_A)}{|H_A|} < \theta\}| > 0; \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

$f_r(\boldsymbol{x}) = 1$ implies that $\boldsymbol{x}$ is a rare pattern, and $f_r(\boldsymbol{x}) = 0$ implies that $\boldsymbol{x}$ is a normal pattern. Note that the calculation of $f_r(\boldsymbol{x})$ is based on the data of all patients in $\mathbf{d}_{2,h}$ and the threshold θ. We generate a sleep pattern table that stores the flag $f_r(\boldsymbol{x})$ for all sleep pattern $\boldsymbol{x}$ of the patients in $\mathbf{d}_{2,h}$, where we can set a small value for θ (e.g., $\theta = 0.00001$).

Based on $f_r(\cdot)$, given a sequence of sleep stages $\boldsymbol{x}$ for a patient, we design and implement a scoring function $V(\boldsymbol{x}, f_r, L)$, called *rarity score*, to evaluate the rarity of the sequence of sleep stages of a patient, where $0 \leq V(\boldsymbol{x}, f_r, L) \leq 1$. The rarity score $V(\boldsymbol{x}, f_r, L)$ is defined as:

$$V(\boldsymbol{x}, f_r, L) = \frac{1}{|\boldsymbol{x}| - L + 1} \sum_{a \in \mathcal{A}_L(\boldsymbol{x})} f_r(a) \quad (7)$$

We calculate the rarity score for every patient in $\mathbf{d}_2$ using the sequence of sleep stages of the patient. The dataset $\mathbf{d}_2$ contains both healthy patients (i.e., the subdataset $\mathbf{d}_{2,h}$) and OSA patients (i.e., the subdataset $\mathbf{d}_{2,o}$). Thus we compare the rarity scores of every patients in $\mathbf{d}_2$, then set a threshold η. If a patient with the rarity score $> \eta$, the patient is predicted as an OSA patient. Otherwise (i.e., the patient with the rarity score $\leq \eta$), he is predicted as a healthy patient.

We execute the following steps to fine-tune the parameters θ and η. Let Θ be the set of all candidate parameters θ, and H be the set of all candidate parameters η. Let the set of all (θ, η) pairs be $\Theta \times H = \{(\theta, \eta) : \theta \in \Theta \text{ and } \eta \in H\}$. For each pair (θ', η') in $\Theta \times H$, we use the θ' to obtain f_r and generate the corresponding sleep pattern table. Then we calculate the rarity scores for all patients in $\mathbf{d}_2$. We calculate the number e_h of the patients in $\mathbf{d}_{2,h}$ with rarity score $> \eta'$ and the number e_o of the patients in $\mathbf{d}_{2,o}$ with rarity score $\leq \eta'$. Let $e(\theta', \eta') = e_h + e_o$. We select the pair $(\theta', \eta') \in \Theta \times H$ with lowest (θ', η') as the parameters θ and η of the OSA detection model.

III. Performance Study

In this section, we study the performance of our RAPIDEST, in terms of the accuracy for sleep stage classification, and the accuracy of OSA detection. We also study the relationship between the rarity score and AHI.

A. Datasets

In this paper, to evaluate the performance of our RAPIDEST, we use the following three open datasets: the Sleep-EDF dataset [34], the University College Dublin Sleep Apnea Database (UCDDB) dataset [34], and Wisconsin Sleep Cohort (WSC) dataset [43], which are available from the public repository PhysioNet [44], [45] and National Sleep Research Resource [46], [47].

The Sleep-EDF dataset contains two subsets denoted as SC* and ST*, respectively. The subset SC* was obtained from the healthy patients without any sleep-related medication, which contains the Fpz-Cz/Pz-Oz EEG signals recorded from 10 males and 10 females without any sleep-related medication. The age range of the patients is 25-34 year-old. There are two approximately 20-hour recordings per patient except one patients who has only a single recording. EEG signals were recorded during two consecutive nights at the patients' home. The sampling rate is 100 Hz.

The subset ST* was obtained in a study in 1994, which studied the temazepam effects on the sleep of patients with mild difficulty in falling asleep. It contains the Fpz-Cz/Pz-Oz EEG signals recorded from 7 males and 15 females, who had mild difficulty in falling asleep. The EEG signals were recorded in the hospital for two nights. The sampling rate is 100 Hz.

In the SC* and ST* subsets, each 30-s epoch of the EEG signals has been annotated into the classes Awake, REM, N1, N2, N3, and N4, respectively. The epochs corresponding to movement and unknown stages were excluded and the epochs labeled by N4 are merged to N3 according to the AASM standard.

The UCDDB dataset contains the overnight PSG recordings of 25 patients with sleep-disordered breathing [34], including of 21 males and 4 females with an average age of 50. Each recording contains two EEG channels and an annotation file with detailed onset time and duration of every hypopnea event. The sampling rate of the EEG signals is 128 Hz.

The WSC dataset consists of the overnight PSG recordings of 1123 patients, including 608 males and 515 females with an average age of 56. The patients in the WSC dataset are random samples of the general population [43]. Each recording contains the two EEG channels, namely, C3-M2 and O1-M2, and the sleep stages for each 30-s epoch. The sampling rate of the EEG signals is 100 Hz. The dataset contains AHI value for each recording. We randomly select 269 patients (including 77 patients with AHI ≥ 15 and 192 patients with AHI < 15) from the WSC dataset.

B. Performance Study for Sleep Stage Classification

We evaluate the sleep stage classification performance for RAPIDEST by adopting the Leave-One-Out Cross-Validation (LOOCV) scheme [48]. The LOOCV scheme exercises as follows: In each run of the experiment, we use the EEG signals of one patient as the testing dataset. The EEG signals of other patients are used to train the classification model. The experiment terminates until all patients are tested. Then, we calculate the performance metrics for the sleep stage classification. For example, if there are 20 patients in the data set, the experiment is run for 20 runs.

There are significant differences between EEG signals of different patients. In this study, we use the EEG signals from different patients for testing (i.e., to classify the sleep stages for a new arrival patient from a given dataset that the sleep stages are already annotated) by which the performance can approximate that in the real-world scenario as the ground truth.

We measure the performance in terms of the confusion matrix $M \in \mathbb{R}^{5\times 5}$ [49]. To simplify our discussion, we denote Sleep Stages **A**, **R**, **1**, **2**, **3** as the 1st class, the 2nd class, the 3rd class, the 4th class, and the 5th class, respectively. Let $m_{i,j}$ be the element at the ith row and the jth column of the confusion matrix M, where $1 \le i, j \le 5$. $m_{i,j}$ is obtained from the classification results for the EEG signals of the testing patient. $m_{i,j}$ is the number of 30-s EEG signals that are labeled as the ith class in the original data set, but are classified as the jth class by the classification model.

There are many performance metrics in assessing a classification model [50]. In this paper, we will adopt the following few metrics suitable for our problem. Following [50], we start with precision ($n_p(i)$), recall ($n_r(i)$), and then we have F1-score ($n_{F1}(i)$) for the ith class:

$$n_p(i) = \frac{m_{i,i}}{\sum_{j=1}^{5} m_{j,i}}; n_r(i) = \frac{m_{i,i}}{\sum_{j=1}^{5} m_{i,j}};$$
$$n_{F1}(i) = \frac{2n_p(i)n_r(i)}{n_p(i)+n_r(i)}.$$

The precision $n_p(i)$ measures correct classification among all samples that are classified into the ith-class. The recall $n_r(i)$ measures the ratio of the number of the samples in the ith class that is correctly classified into the ith class to the number of the samples in the ith-class to be classified. The F1-score $n_{F1}(i)$ is the harmonic mean of $n_p(i)$ and $n_r(i)$ for the ith class. Furthermore, we obtain the overall accuracy ACC, the macro F1 score Macro-F1 and the Cohen's kappa coefficient κ as follows [50], [51]:

$$\text{ACC} = \frac{\sum_{i=1}^{5} m_{i,i}}{\sum_{i,j=1}^{5} m_{i,j}}; \quad \text{Macro-F1} = \frac{1}{5}\sum_{i=1}^{5} n_{F1}(i). \tag{8}$$

Let EA be the expected accuracy, which is given by [51]

$$\text{EA} = \frac{\sum_{i=1}^{5}\left(\sum_{j=1}^{5} m_{i,j}\right) \times \left(\sum_{j=1}^{5} m_{j,i}\right)}{\left(\sum_{i,j=1}^{5} m_{i,j}\right)^2}.$$

Then we obtain the Cohen's kappa (κ) coefficient by

$$\kappa = \frac{\text{ACC} - \text{EA}}{1 - \text{EA}}. \tag{9}$$

TABLE I
SLEEP STAGE CLASSIFICATION PERFORMANCES RESULT FOR THE DATASETS SLEEP-EDF SC*, SLEEP-EDF ST*, UCDDB, AND WSC

Dataset	Stages	Per-class Metrics			Overall Metrics		
		n_p	n_r	n_{F1}	ACC	Macro-F1	κ
Sleep-EDF SC*	Awake	87	91	89	82.90%	74.69%	76.4%
	REM	77	78	78			
	N1	37	27	31			
	N2	88	87	88			
	N3	86	90	88			
Sleep-EDF ST*	Awake	78	88	83	76.53%	72.24%	67.3%
	REM	83	79	81			
	N1	44	44	44			
	N2	82	81	82			
	N3	71	72	71			
UCDDB	Awake	75	85	79	70.29%	66.53%	61.2%
	REM	61	77	68			
	N1	42	20	27			
	N2	74	77	75			
	N3	82	83	82			
WSC	Awake	83	90	86	74.46%	65.61%	63.42%
	REM	71	84	77			
	N1	36	49	41			
	N2	96	70	81			
	N3	28	94	43			

The Macro-F1 is the average F1-score $n_{F1}(i)$ of all classes. The Cohen's kappa coefficient κ measures the consistency between the classification results and the labels. A higher κ implies that the classification result and the label are more consistent. The Cohen's kappa coefficient is more useful than overall accuracy for imbalanced data.

1) Performance Study for Sleep Stage Classification of RAPIDEST: We run experiments based on the datasets, Sleep-EDF SC*, Sleep-EDF ST*, UCDDB, and WSC, and we have 20, 22, 25, and 269 LOOCV runs for the four datasets, respectively. Table I shows the per-class performance metrics, $n_p(i)$, $n_r(i)$ and $n_{F1}(i)$, and the corresponding ACC, Macro-F1, and κ, for the datasets Sleep-EDF SC*, Sleep-EDF ST*, UCDDB, and WSC. As shown in Table I, our RAPIDEST has the highest ACC (82.%), Macro-F1 (74.79%), and κ (76.4%) for the Sleep-EDF SC* dataset. The patients in the Sleep-EDF SC* dataset do not have any sleep-related medication, some of the patients in the Sleep-EDF ST* dataset have mild difficulty in falling asleep, and all patients in the UCDDB dataset have sleep apnea. The WSC dataset contains random samples from the general population, and the patients may be healthy or have sleep apena. We conclude that the RAPIDEST has the best ACC performance on the sleep stage classification for the dataset Sleep-EDF SC* that contains patients who do not have any sleep-related medication. For the other datasets, Sleep-EDF ST*, UCDDB, WSC, our RAPIDEST has ACC higher than 70%.

2) Performance Comparison With Existing Works for Sleep Stage Classification: We also compare the sleep stage classification performance of our RAPIDEST with two existing sleep stage classification algorithms, including the diffusion map [52] and the DeepSleepNet [29]. Due to page limitation, we do not offer the details of the two existing algorithms. For more details, readers may refer to [29] and [52].

TABLE II
PERFORMANCE COMPARISON BETWEEN RAPIDEST, DIFFUSION MAP, AND DEEPSLEEPNET FOR THE SLEEP-EDF SC* DATASET

algorithm	ACC	Macro-F1	κ	Execution Time
Our RAPIDEST	82.9%	74.69%	76.4%	0.81 seconds
diffusion map	83.68%	77.11%	78.14%	27.88 seconds
DeepSleepNet	82.01%	76.86%	75.7%	1.97 seconds

The diffusion map and DeepSleepNet have shown their corresponding ACC, Macro-F1, and Cohen's kappa performances in [29] and [52] for the Sleep-EDF SC* dataset. Based on the performance results in [29] and [52], Table II compares our proposed RAPIDEST, diffusion map, and DeepSleeNet in terms of ACC, Macro-F1, and κ. We observe that the ACC, Macro-F1, and κ performances of the three algorithms fall into the ranges 82.01%-83.68%, 74.69%-77.11%, and 75.7%-78.14%, respectively. To conclude, the three algorithms have similar performance for the sleep stage classification.

Based on the Sleep-EDF SC* dataset, we have also studied the average execution time performance for the sleep stage classification for our RAPIDEST, diffusion map, and DeepSleepNet. We have acquired the code directly from the authors for the diffusion map [52] implemented using MATLAB. We obtain the source codes of the DeepSleepNet implemented using python from the open link [53] as shown in [29]. We implement our RAPIDEST using python. We set up a PC as the computation environment with OS: Ubuntu 22.04; CPU: Intel(R) Core(TM) i7-6700 CPU@3.40GHz (4 cores); and Memory: 32GB.

As shown in Table II, the average execution times taken to classify the sleep stages for a patient using RAPIDEST, diffusion map, and DeepSleepNet are 0.81s, 27.88s and 1.97s, respectively. Obviously, among the three algorithms, our RAPIDEST has the best performance in terms of the execution time. To summarize, in terms of the stage classification, our RAPIDEST, diffusion map, and DeepSleepNet have similar performance, but our RAPIDEST has the best performance in complexity.

3) Effects of OSA and Insomnia on Sleep Stage Classification in RAPIDEST: We also study how the OSA and the insomnia affect the sleep stage classification performance of our RAPIDEST. We set up the following experiment to study the effects. Based on the WSC dataset, we obtain the following sub-datasets:

- The $\mathbf{S}_N$ dataset: We include the 104 patients (who have neither OSA nor insomnia) in the WSC dataset into the $\mathbf{S}_N$ dataset.
- The $\mathbf{S}_{N1}$ and $\mathbf{S}_{N2}$ datasets: We form the $\mathbf{S}_{N1}$ dataset by randomly selecting 70 subjects from $\mathbf{S}_N$. The remaining 34 patients in the $\mathbf{S}_N$ dataset form the $\mathbf{S}_{N2}$ dataset.
- The $\mathbf{S}_O$ dataset: There are a total of 77 patients (who have OSA but do not have insomnia) in the WSC dataset. We accommodate the 77 patients into the $\mathbf{S}_O$ dataset.
- The $\mathbf{S}_I$ dataset: There are a total of 129 patients (who have insomnia but do not have OSA) in the WSC dataset. We include the 129 patients into the $\mathbf{S}_I$ dataset.
- The $\mathbf{S}_{O,I}$ dataset: There are a total of 41 subjects (who have both OSA and insomnia) in the WSC dataset. We put the 41 subjects into the dataset $\mathbf{S}_{O,I}$ dataset.

By applying our RAPIDEST, we build up a sleep stage classification model (i.e., Step T-1) using the $\mathbf{S}_{N1}$ dataset. We run the sleep stage classification model to test the four datasets, $\mathbf{S}_{N2}$, $\mathbf{S}_O$, $\mathbf{S}_I$, and $\mathbf{S}_{O,I}$. We investigate the overall accuracy ACC performance (i.e., Eq. (8)) for the four datasets. The ACCs for $\mathbf{S}_{N2}$, $\mathbf{S}_O$, $\mathbf{S}_I$, and $\mathbf{S}_{O,I}$ are 75.5%, 76.2%, 75.6%, and 74.6%, respectively, which shows that the effects of OSA and insomnia on the sleep stage classification in our RAPIDEST are minor.

For the effects of other kinds of sleep disorder (e.g., the narcolepsy, etc.) on the sleep stage classification performance of our RAPIDEST, because currently we cannot find the datasets that contain labels for the other kinds of sleep disorder, we do not come out the study of this part, and we will treat it as a future work.

In the next section, we will study the performance for the OSA detection in the proposed RAPIDEST, and show that OSA detection in our RAPIDEST works well when there exists errors in the sleep stage classification.

C. Performance Study for OSA Detection

In this section, we first investigate the performance of our RAPIDEST for OSA detection. Then we investigate the relationship between the rarity score and AHI. Finally, we study the effects of the sleep stage classification on the OSA detection.

1) OSA Detection: Based on the datasets, UCDDB and WSC, we set up experiments to study the performance of OSA detection. We label the patients (with AHI $\geq$ 15) in UCDDB and WSC as the OSA class. For those with AHI $<$ 15, they are labeled as the healthy class. We evaluate the performance of our RAPIDEST for OSA detection in terms of accuracy, recall, and precision. We treat OSA class as positive class and healthy class as negative class. We define True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) as:

- TP: the number of the patients of the OSA class that are correctly classified as the OSA class;
- FP: the number of the patients of the healthy class that are incorrectly classified as the OSA class;
- TN: the number of the patients of the healthy class that are correctly classified as the healthy class;
- FN: the number of the patients of the OSA class that are incorrectly classified as the healthy class.

The accuracy, recall, and precision are obtained by $\frac{TP+TN}{TP+FP+TN+FN}$, $\frac{TP}{TP+FN}$, and $\frac{TP}{TP+FP}$, respectively.

We set up the following two experiments using UCDDB and WSC, respectively.

Experiment I (UCDDB): There are 11 healthy patients and 14 OSA patients in UCDDB. We randomly select 5 healthy patients and 5 OSA patients to form the testing dataset, $\mathbf{D}_D$, i.e., $N_{d,h} = 5$ and $N_{d,o} = 5$. The remaining 6 healthy patients and 9 OSA patients form the training dataset, $\mathbf{D}_T$, i.e., $N_{t,h} = 6$ and

$N_{t,o} = 9$. We randomly select $n_{t,h,1} = 2$ healthy patients from the $N_{t,h} = 9$ healthy patients and randomly select $n_{t,o,1} = 5$ OSA patients from the $N_{t,o} = 6$ patients to form dataset $\mathbf{d}_1$, and the dataset $\mathbf{d}_2 = \mathbf{D}_T - \mathbf{d}_1$.

Based on the training dataset $\mathbf{D}_T$, we build up the OSA detection model with $\theta = 0.0354$ and $\eta = 0.0337$.

Experiment II (WSC): There are 192 healthy patients and 77 OSA patients in WSC. We randomly select 47 healthy patients and 47 OSA patients to form the testing dataset, $\mathbf{D}_D$, i.e., $N_{d,h} = 47$ and $N_{d,o} = 47$. The remaining 145 healthy patients and 30 OSA patients form the training dataset, $\mathbf{D}_T$, i.e., $N_{t,h} = 145$ and $N_{t,o} = 30$. We randomly select $n_{t,h,1} = 30$ healthy patients from the $N_{t,h} = 145$ healthy patients to form dataset $\mathbf{d}_1$, i.e., $n_{t,o,1} = 0$

Based on the training dataset $\mathbf{D}_T$, we build up the OSA detection model with $\theta = 0.032$ and $\eta = 0.0308$.

Table III shows the performance result for OSA detection for the datasets, UCDDB and WSC. We observe that the recall, precision, and accuracy for UCDDB are 100%, 71.43%, and 80%, respectively. Obviously, our RAPIDEST has good OSA detection performance for the UCDDB dataset. However, for the WSC dataset, the recall, precision, and accuracy are 80.85%, 65.52%, and 69.15%, respectively. The performance is not as good as the UCDDB dataset. As mentioned in Section III-A, the patients in the WSC dataset are random samples of the general population [43]. On the other hand, the patients in the UCDDB are with sleep disorder [34], [54]. Our RAPIDEST has better OSA detection performance for the patients who have sleep disorder.

2) Relationship Between Rarity Score and AHI: We have also studied the relationship between Rarity Score and AHI. The points (x_i, y_i) represents the rarity score x_i and the AHI value y_i of every patient. We use the Pearson's correlation coefficient to measure the strength of the relationship between the rarity score and AHI. The Pearson's correlation coefficient ρ [55] for the n pairs of samples $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ is calculated by

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Figs. 5 and 6 show the scatter plots for 10 patients in $\mathbf{D}_D$ of the UCDDB dataset (Experiment I) and 94 patients in $\mathbf{D}_D$ of the WSC dataset (Experiment II), respectively. As shown in [55], a $\rho > 0.7$ implies a high positive correlation between the rarity score x_i and the AHI value y_i, i.e., the higher rarity score implies the higher AHI. In this study, we observe that the Pearson's correlation coefficient is 0.87 for UCDDB, which implies a positive correlation between the rarity score and AHI. For the WSC dataset, the Pearson's correlation coefficient is 0.576 for the 94 patients, which implies a moderate positive correlation between the rarity score and AHI. The result for UCDDB may not be convinced because of fewer number of patients in UCDDB. This phenomenon shows that our proposed PAPIDEST can offer a rarity score that reflects the severity of sleep disorder (i.e., AHI) of a patient especially when the patient has sleep disorder.

TABLE III
PERFORMANCE OF OSA DETECTION FOR THE DATASETS, UCDDB AND WSC

Dataset	Recall	Precision	Accuracy
UCDDB	100.00%	71.43%	80.00%
WSC	80.85%	65.52%	69.15%

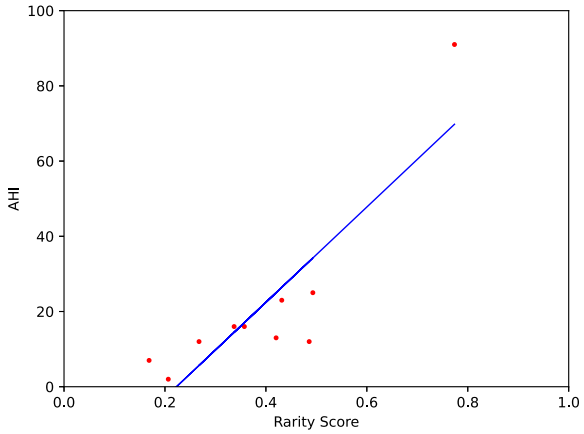

Fig. 5. Scatter plot of rarity score and AHI of the 10 patients for the UCDDB dataset.

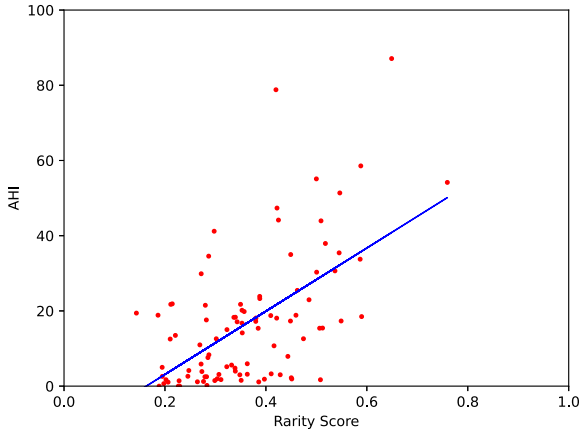

Fig. 6. Scatter plot of rarity score and AHI of the 94 patients for the WSC dataset.

3) Effects of Sleep Stage Classification on OSA Detection: In addition, we run experiments to study how the sleep stage classification performance affects OSA detection performance. Based on Experiment I (UCDDB dataset), we use the ground truth of the sequence of sleep stages (that has been labeled) in the original dataset $\mathbf{d}_2$ to construct the OSA detection model. Then using the OSA detection model and the ground truth of the sequence of sleep stages (that has been labeled) in $\mathbf{D}_D$, we have OSA detection on the patients in the dataset $\mathbf{D}_D$. We obtain that the recall, precision, and accuracy are 100%, 83.33%, and 90%, respectively. Comparing the results with that shown in Table III (UCDDB), we observe that using the

ground truth for sleep stage classification improves both the precision and the accuracy by about 10%.

IV. Conclusion

In this paper, we have designed and implemented RAPIDEST: a rare pattern identification and detection for sleep stage transitions that only uses the EEG signals for OSA detection, which can significantly reduce the complexity of signal collection and the overhead for OSA diagnosis. In our RAPIDEST, we define a rarity score to evaluate the unusualness of sequences of sleep stages for a patient, and identify the relationship between the rarity score of sequence of sleep stages and AHI. We have run experiments to thoroughly evaluate the performance of our RAPIDEST in terms of the sleep stage classification and the OSA detection. Our study shows that our RAPIDEST (1) has good performance for the sleep stage classification with lower computation time complexity, (2) offer a rarity score that has a positive correlation with the severity of sleep disorder (i.e., AHI) of the patient, and (3) has good performance for OSA detection for the patients who has sleep disorder. Since only EEG signals are needed in the OSA detection and the resulting metric (the rarity score) has close correlation to the well-known AHI, our rarity score can serve as a potential effective metric for OSA detection with much more convenient data collection process, potentially revolutionizing the OSA detection technologies.

Unfortunately, our proposed framework RAPIDEST is only to identify rare patterns in the sequence of sleep stages. When it comes to different types of sleep disorders, our proposed RAPIDEST cannot determine what type the disorder is. If we could acquire sufficient data for sleep disorders labeled with types, we could extend our framework to provide such a function, which forms a promising research direction. It is also hard to find the rarity score of individual segment because we compute the rarity score over the whole-night sleep stages. Since a healthy patient may also contain some rare patterns in the sequence of sleep stages over a longer period, rarity score obtained from the whole-night EEG signals may cause significant false alarm. To overcome this problem, we may need the dataset labeled with anomaly event to determine the rarity score for EEG signals for each 30-s epoch, which demands further research.

Acknowledgment

The authors submit their applications to obtain the Wisconsin Sleep Cohort (WSC) dataset provided by the National Sleep Research Resource. The authors would like to thank the National Sleep Research Resource. The Wisconsin Sleep Cohort (WSC) dataset used in this work is provided by the National Sleep Research Resource. The National Sleep Research Resource was supported by the U.S. National Institutes of Health, National Heart Lung and Blood Institute (R24 HL114473, 75N92019R002). The WSC study was supported by the U.S. National Institutes of Health, National Heart, Lung, and Blood Institute (R01HL62252), National Institute on Aging (R01AG036838, R01AG058680), and the National Center for Research Resources (1UL1RR025011).

References

[1] L. R. Kline. *Clinical Presentation and Diagnosis of Obstructive Sleep Apnea in Adults*. Accessed: Jan. 2022. [Online]. Available: https://www.uptodate.com/contents/clinical-presentation-and-diagnosis-of-obstructive-sleep-apnea-in-adults

[2] A. L. Chesson, "Practice parameters for the indications for polysomnography and related procedures," *Sleep*, vol. 20, no. 6, pp. 406–422, 1997.

[3] W. R. Ruehland, P. D. Rochford, F. J. O'Donoghue, R. J. Pierce, P. Singh, and A. T. Thornton, "The new AASM criteria for scoring hypopneas: Impact on the apnea hypopnea index," *Sleep*, vol. 32, no. 2, pp. 150–157, Feb. 2009.

[4] *The International Classification of Sleep Disorders: Diagnostic and Coding Manual*, 2nd ed., American Academy of Sleep Medicine, Westchester, IL, USA, 2005.

[5] E. A. Wolpert, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Archives General Psychiatry*, vol. 20, no. 2, pp. 246–247, 1969.

[6] G. Ayoub, T. H. Dang, T. I. Oh, S.-W. Kim, and E. J. Woo, "Feature extraction of upper airway dynamics during sleep apnea using electrical impedance tomography," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, Dec. 2020.

[7] P. Lakhan, A. Ditthapron, N. Banluesombatkul, and T. Wilaiprasitporn, "Deep neural networks with weighted averaged overnight airflow features for sleep apnea-hypopnea severity classification," in *Proc. TENCON IEEE Region Conf.*, Oct. 2018, pp. 441–445.

[8] K. Behbehani, F.-C. Yen, J. R. Burk, E. A. Lucas, and J. R. Axe, "Automatic control of airway pressure for treatment of obstructive sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 10, pp. 1007–1016, Oct. 1995.

[9] B. A. Staats, H. W. Bonekat, C. D. Harris, and K. P. Offord, "Chest wall motion in sleep apnea," *Amer. Rev. Respiratory Disease*, vol. 130, no. 1, pp. 59–63, 1984.

[10] D. Álvarez, R. Hornero, J. V. Marcos, and F. del Campo, "Multivariate analysis of blood oxygen saturation recordings in obstructive sleep apnea diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2816–2824, Dec. 2010.

[11] L. Kalevo et al., "Self-applied electrode set provides a clinically feasible solution enabling EEG recording in home sleep apnea testing," *IEEE Access*, vol. 10, pp. 60633–60642, 2022.

[12] H. Hilmisson, N. Lange, and S. P. Duntley, "Sleep apnea detection: Accuracy of using automated ECG analysis compared to manually scored polysomnography (apnea hypopnea index)," *Sleep Breath*, vol. 23, no. 1, pp. 125–133, Mar. 2019.

[13] P. Y. Kim, D. E. McCarty, L. Wang, C. Frilot, A. L. Chesson, and A. A. Marino, "Two-group classification of patients with obstructive sleep apnea based on analysis of brain recurrence," *Clin. Neurophysiol.*, vol. 125, no. 6, pp. 1174–1181, Jun. 2014.

[14] I. G. Campbell, "EEG recording and analysis for sleep research," *Current Protocols Neurosci.*, vol. 49, no. 1, pp. 2–10, Oct. 2009.

[15] S. Shustak et al., "Home monitoring of sleep with a temporary-tattoo EEG, EOG and EMG electrode array: A feasibility study," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026024.

[16] X. Zhao et al., "Classification of sleep apnea based on EEG sub-band signal characteristics," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, Dec. 2021.

[17] D. Liu, Z. Pang, and S. R. Lloyd, "A neural network method for detection of obstructive sleep apnea and narcolepsy based on pupil size and EEG," *IEEE Trans. Neural Netw.*, vol. 19, no. 2, pp. 308–318, Feb. 2008.

[18] Y. Wang, S. Ji, T. Yang, X. Wang, H. Wang, and X. Zhao, "An efficient method to detect sleep Hypopnea–apnea events based on EEG signals," *IEEE Access*, vol. 9, pp. 641–650, 2021.

[19] T. Mahmud, I. A. Khan, T. I. Mahmud, S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "Sleep apnea detection from variational mode decomposed EEG signal using a hybrid CNN-BiLSTM," *IEEE Access*, vol. 9, pp. 102355–102367, 2021.

[20] K. Feng, H. Qin, S. Wu, W. Pan, and G. Liu, "A sleep apnea detection method based on unsupervised feature learning and single-lead electrocardiogram," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[21] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, and B. V. Vaughn, *The AASM Manual for the Scoring of Sleep and Associated Events*. Darien, IL, USA: American Academy of Sleep Medicine, 2012.

[22] Z. M. Gotts, V. Deary, J. Newton, D. Van der Dussen, P. D. Roy, and J. G. Ellis, "Are there sleep-specific phenotypes in patients with chronic fatigue syndrome? A cross-sectional polysomnography analysis," *BMJ Open*, vol. 3, no. 6, 2013, Art. no. e002999.

[23] C. M. Shapiro, R. D. Griesel, P. R. Bartel, and P. L. Jooste, "Sleep patterns afted graded exercise," *J. Appl. Physiol.*, vol. 39, no. 2, pp. 187–190, Aug. 1975.
[24] R. Stickgold, "Sleep-dependent memory consolidation," *Nature*, vol. 437, pp. 1272–1278, Oct. 2005.
[25] R. Wassing et al., "Slow dissolving of emotional distress contributes to hyperarousal," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 9, pp. 2538–2543, Mar. 2016.
[26] J. Somanna, D. Joshi, H. Gundu, and G. Srinivasa, "Automated classification of sleep apnea and hypopnea on polysomnography data," in *Proc. 12th Biomed. Eng. Int. Conf. (BMEiCON)*, Nov. 2019, pp. 1–5.
[27] G. C. Gutiérrez-Tobal, M. L. Alonso-Álvarez, D. Álvarez, F. del Campo, J. Terán-Santos, and R. Hornero, "Diagnosis of pediatric obstructive sleep apnea: Preliminary findings using automatic analysis of airflow and oximetry recordings obtained at patients' home," *Biomed. Signal Process. Control*, vol. 18, pp. 401–407, Apr. 2015.
[28] E. Estrada, H. Nazeran, J. Barragan, J. R. Burk, E. A. Lucas, and K. Behbehani, "EOG and EMG: Two important switches in automatic sleep stage classification," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2006, pp. 2458–2461.
[29] A. Supratak, H. Dong, C. Wu, and Y. Guo, "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.
[30] O. Tsinalis, P. M. Matthews, Y. Guo, and S. Zafeiriou, "Automatic sleep stage scoring with single-channel EEG using convolutional neural networks," 2016, *arXiv:1610.01683*.
[31] H. Dong, A. Supratak, W. Pan, C. Wu, P. M. Matthews, and Y. Guo, "Mixed neural network approach for temporal sleep stage classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 324–333, Feb. 2018.
[32] N. Banluesombatkul et al., "MetaSleepLearner: A pilot study on fast adaptation of bio-signals-based sleep stage classifier to new individual subject using meta-learning," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 1949–1963, Jun. 2021.
[33] J. Zhang and Y. Wu, "Automatic sleep stage classification of single-channel EEG by using complex-valued convolutional neural network," *Biomed. Eng./Biomedizinische Technik*, vol. 63, no. 2, pp. 177–190, Mar. 2018.
[34] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. 215–220, Jan. 2000.
[35] C. Huang, Y. Xiao, and G. Xu, "Predicting human intention-behavior through EEG signal analysis using multi-scale CNN," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 5, pp. 1722–1729, Sep. 2021.
[36] P. Memar and F. Faradji, "A novel multi-class EEG-based sleep stage classification system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 1, pp. 84–95, Jan. 2018.
[37] A. C. Tsoi, D. S. C. So, and A. Sergejew, "Classification of electroencephalogram using artificial neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 6, 1994, pp. 1151–1158.
[38] A. Atyabi, F. Shic, and A. Naples, "Mixture of autoregressive modeling orders and its implication on single trial EEG classification," *Expert Syst. Appl.*, vol. 65, pp. 164–180, Dec. 2016.
[39] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 4, no. 4. Cham, Switzerland: Springer, 2006.
[40] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 971–980.
[41] K.-O. Cho and H.-J. Jang, "Comparison of different input modalities and network structures for deep learning-based seizure detection," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, Dec. 2020.
[42] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection for discrete sequences: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 5, pp. 823–839, May 2012.
[43] T. Young, M. Palta, J. Dempsey, P. E. Peppard, F. J. Nieto, and K. M. Hla, "Burden of sleep apnea: Rationale, design, and major findings of the Wisconsin Sleep Cohort study," *WMJ, Off. Publication State Med. Soc. Wisconsin*, vol. 108, no. 5, p. 246, 2009.
[44] *Sleep-EDF Database Expanded*. Accessed: Jan. 2021. [Online]. Available: http://www.physionet.org/physiobank/database/sleep-edfx/
[45] *St. Vincent's University Hospital/University College Dublin Sleep Apnea Database*. Accessed: Jan. 2021. [Online]. Available: https://physionet.org/content/ucddb/
[46] G.-Q. Zhang et al., "The national sleep research resource: Towards a sleep data commons," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 10, pp. 1351–1358, 2018.
[47] *National Sleep Research Resource*. Accessed: May 2022. [Online]. Available: https://sleepdata.org/
[48] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Cham, Switzerland: Springer, 2009, pp. 241–242.
[49] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012, pp. 180–181.
[50] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, 2009.
[51] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.
[52] G.-R. Liu, Y.-L. Lo, J. Malik, Y.-C. Sheu, and H.-T. Wu, "Diffuse to fuse EEG spectra—Intrinsic geometry of sleep dynamics for classification," *Biomed. Signal Process. Control*, vol. 55, Jan. 2020, Art. no. 101576.
[53] *DeepSleepNet Code*. Accessed: May 2022. [Online]. Available: https://github.com/akaraspt/deepsleepnet
[54] J. P. Bakker et al., "Gastric banding surgery versus continuous positive airway pressure for obstructive sleep apnea: A randomized controlled trial," *Amer. J. Respiratory Crit. Care Med.*, vol. 197, no. 8, pp. 1080–1083, Apr. 2018.
[55] M. M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Med. J.*, vol. 24, no. 3, pp. 69–71, 2012.