A Novel Multi-Class EEG-Based Sleep Stage Classification System

Pejman Memar and Farhad Faradji

***Abstract*— Sleep stage classification is one of the most critical steps in effective diagnosis and the treatment of sleep-related disorders. Visual inspection undertaken by sleep experts is a time-consuming and burdensome task. A computer-assisted sleep stage classification system is thus essential for both sleep-related disorders diagnosis and sleep monitoring. In this paper, we propose a system to classify the wake and sleep stages with high rates of sensitivity and specificity. The EEG signals of 25 subjects with suspected sleep-disordered breathing, and the EEG signals of 20 healthy subjects from three data sets are used. Every EEG epoch is decomposed into eight subband epochs each of which has a frequency band pertaining to one EEG rhythm (i.e., delta, theta, alpha, sigma, beta 1, beta 2, gamma 1, or gamma 2). Thirteen features are extracted from each subband epoch. Therefore, 104 features are totally obtained for every EEG epoch. The Kruskal–Wallis test is used to examine the significance of the features. Non-significant features are discarded. The minimal-redundancy-maximal-relevance feature selection algorithm is then used to eliminate redundant and irrelevant features. The features selected are classified by a random forest classifier. To set the system parameters and to evaluate the system performance, nested 5-fold cross-validation and subject cross-validation are performed. The performance of our proposed system is evaluated for different multi-class classification problems. The minimum overall accuracy rates obtained are 95.31% and 86.64% for nested 5-fold and subject cross-validation, respectively. The system performance is promising in terms of the accuracy, sensitivity, and specificity rates compared with the ones of the state-of-the-art systems. The proposed system can be used in health care applications with the aim of improving sleep stage classification.**

***Index Terms*— Sleep stages, classification, minimal-redundancy-maximal-relevance (mRMR), random forest (RF), electroencephalogram (EEG), gamma, subject cross-validation.**

I. Introduction

SLEEP has a significant role in healthiness and well-being throughout our life. Getting adequate sleep at nights can help protect our mental health, physical health, and quality of life. Sleep screening and analysis is a considerable tool in assessment of sleep-related disorders, such as sleep apnea syndrome, schizophrenia, depression, insomnia, narcolepsy, and other neural abnormalities.

Sleepers cyclically pass through five different stages of sleep, i.e., stage 1, stage 2, stage 3, stage 4, and stage 5 or rapid eye movement (REM) sleep. The typical length of a complete sleep cycle is about 90 to 110 minutes [1], [2]. Sleep in stage 1 is light. The eyes move slowly and muscle activity is slow. It is in stage 2 that eyes stop moving and the brain waves become slower. Deep sleep occurs in stages 3 and 4 when no eye movement and muscle activity exist [3]. In stage 5 or REM period, the rate of breathing increases and eyes move rapidly.

Sleep screening is normally performed using polysomnographic recordings which include the electrocardiogram (ECG), electroencephalogram (EEG), electromayogram (EMG), electrooculogram (EOG), and respiration signals [4]. Amongst the physiological signals, EEG signals are used frequently as they represent the brain activities.

The EEG rhythms or waves, which are delta, theta, alpha, sigma, beta, and gamma waves, demonstrate different characteristics of the sleep stages. Please refer to Table I [5]. Theta wave (4-8 Hz) and alpha wave (8-12 Hz) are present in stage 1 of sleep. During stage 2, the EEG signal amplitude increases and K-complexes appear. Theta waves are also more prominent in this stage. Theta wave and delta wave (0-4 Hz) are more noticeable in stage 3. The frequency of the EEG signal during stage 4 varies between 0.5 and 2 Hz. In REM period, sigma wave (12-15 Hz), beta wave (15-30 Hz), and gamma wave (>30 Hz) are more dominant; hence, the frequency content of the EEG signal is greater than 12 Hz. Beta waves are also more predominant during wakefulness [6].

Analyzing recorded EEG signals manually is quite exhaustive since physicians have to review several hours of EEG signals while the signals are displayed on a screen in 5- to 10-second frames. In order to assist physicians in

TABLE I
Frequency Bands of EEG Rhythms

Rhythm	Frequency Band (Hz)
Delta (δ)	0–4
Theta (θ)	4–8
Alpha (α)	8–12
Sigma (σ)	12–15
Beta 1 (β_1)	15–22
Beta 2 (β_2)	22–30
Gamma 1 (γ_1)	30–40
Gamma 2 (γ_2)	40–49.5

Manuscript received April 17, 2017; revised September 8, 2017; accepted November 17, 2017. Date of publication November 21, 2017; date of current version January 8, 2018. *(Corresponding author: Farhad Faradji.)*

The authors are with the Department of Electrical Engineering, K. N. Toosi University of Technology, Tehran 19697, Iran (e-mail: p.memar@email.kntu.ac.ir; faradji@kntu.ac.ir).

Digital Object Identifier 10.1109/TNSRE.2017.2776149

real-time diagnosis, it is essential to develop a sleep stage classification system.

II. Related Works

Many systems using single-channel EEG signals have been proposed in literature for sleep stage classification. Sharma *et al.* [7] propose a sleep stage classification system based on iterative filtering. Poincaré plot descriptors and statistical measures are the features extracted. Five different classifiers, i.e., naïve Bayes, k-nearest neighbor, multilayer perceptron, C4.5 decision tree, and random forest are used to perform classification. The accuracy rates obtained are between 90.02% and 98.02%. In a study by Tian *et al.* [1], the multi-scale entropy is used as the feature. Using a hierarchical classification method, extracted features are then classified. The accuracy rate reported is 91.4%. In [8], EEG signals are decomposed using ensemble empirical mode decomposition (EEMD). Several statistical moment-based features are extracted and classified using random under sampling boosting (RUSBoost). The accuracy rate obtained is 88.07%. Tunable-Q factor wavelet transform (TQWT) is applied to sleep EEG signal segments in [9]. The EEG segments are decomposed into TQWT sub-bands. Features are extracted using normal inverse Gaussian (NIG) distribution modeling. The features are then classified using an adaptive boosting (Adaboost) technique. The accuracy rate of the system is 90.01%.

In [10], the maximum-minimum distance and EnergySis features are used. Different classifiers are employed. The best accuracy rate obtained is 93.1% for 6-stage classification. In [11], the empirical mode decomposition (EMD) and statistical features are classified using different classifiers and the best accuracy rate for 6-stage classification is obtained 88.62%. In [12], a system based on complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating (Bagging) is proposed. After signal decomposition into intrinsic mode functions, higher order statistical moments are selected as features. Bagged decision trees are used for classification. The reported accuracy rate is 86.89%.

Lajnef *et al.* [13] extract various features such as the variance, skewness, kurtosis, linear prediction error energy, and permutation entropy from two EEG, two EOG, and one EMG signals. The features are classified using a multi-class support vector machine (SVM) with an accuracy rate of 92%. In [14], standard statistics and spectral features are classified using a Bagging classifier. The best accuracy rate obtained for 6-stage classification is 85.57%. The sleep stages are classified using graph domain features in [15]. An SVM is employed as the classifier. The accuracy rate reported for 6-state classification is 87.5%. In [16], various features based on the fractal dimension, detrended fluctuation analysis, shannon entropy, approximate entropy, sample entropy, multi-scale entropy, and principal component analysis are extracted. An accuracy rate of 80% is obtained for a neural network classifier.

The system proposed in [17] is based on the average spectral power, preferential frequency band, and cross frequency coupling. The reported accuracy rate for a linear discriminant analysis classifier is 75%. In [18], the time-frequency image (TFI) of EEG signals is obtained from the smoothed pseudo Wigner-Vill distribution based time-frequency representation of the EEG signals. The features are extracted from the histogram of the segmented TFI. The classifier is a multi-class least squares SVM with a radial basis function (RBF) kernel. The accuracy rate is 88.47%. Hsu *et al.* [19] use energy statistic features and Elman network classifier. An accuracy rate of 87.2% is obtained for 5-stage classification.

In [20], the proposed system is based on the power spectral density (PSD) using the short time Fourier transform and the sequential method. EEG signals from two forehead channels are used. The obtained accuracy rate is 77.1% when an SVM classifier is employed. In a similar study [21], the accuracy rate of the system based on the PSD, fuzzy C-means, and an SVM classifier is 70.92%. The system proposed in [22] is based on three different time-frequency analysis of the EEG signal, i.e., the Hilbert-Hough spectrum, the Wigner-Ville distribution, and the continuous wavelet transform. Features are extracted using the time-frequency entropy. Classification is performed by a feed-forward back-propagation neural network. The accuracy rate is obtained 84%.

In [23], a nonlinear technique, higher order spectra (HOS), is used to extract the information from the sleep EEG signal. For different sleep stages, unique bispectrum and bicoherence plots are first proposed. Several HOS-based features are then obtained from these plots. Classification is based on a Gaussian mixture model (GMM). The accuracy rate reported is 88.7%. The features employed in [24] are extracted using the EMD and Hilbert spectrum based on the Hilbert-Huang transform. Classification is based on the k-nearest neighbors algorithm. The obtained accuracy rate is 81.7%.

In [25], a sleep stage classification system is proposed based on EEG, EOG, and EMG signals. A sparse deep belief net is used for feature extraction and a combination of multiple classifiers is used for classification. The total accuracy rate is 91.31%. In [26], the proposed system is based on time-frequency analysis and stacked sparse autoencoders. The overall accuracy rate reported is 78%. In [27], Hjorth parameters, wavelet transformation, and symbolic representation are used for feature extraction. Classification is performed using SVM classifiers. Stacked sequential learning is then employed to improve the system performance. The accuracy rate obtained is 80.07%.

It is noteworthy that many research works providing very high accuracy rates do not present enough robustness in the problem setup. They select sometimes a subset of epochs, or a subset of subjects from a dataset, and they perform cross-validation, merging all epochs from all subjects. This is unrealistic from the point of view of performing sleep stage classification over unseen subject(s) signals. In [1], [13], [20], [21], and [25]–[27], a subject cross-validation process is used to evaluate the system performance. However, in [1], a subset of subjects and a subset of epochs for each subject is selected. In [20] and [25], a subset of subjects is selected and in [26], a subset of epochs, corresponding to the time in bed, is used.

The contribution of this paper is threefold. First, we propose an effective sleep staging system to classify between different sleep stages and wake using a single-channel EEG signal. Second, the proposed system is sufficiently robust since its

performance is evaluated using whole datasets via subject cross-validation. Third, some features extracted are from gamma band that is normally filtered.

III. Dataset

The first dataset used in this paper is the Sleep-EDF database (SEDFDB) which is available on PhysioNet [28], [29]. In this dataset, EEG signals have been recorded from Caucasian male and female subjects between the ages of 21 and 35. There exist 8 recordings in two subsets (marked as sc* and st*). The first four recordings (i.e., sc4002e0, sc4012e0, sc4102e0, and sc4112e0) are the signals of ambulatory healthy subjects during 24 hours in the normal daily life. These recordings correspond to the wake state (between 15.1 and 17.5 hours) and the sleep state. The remaining four recordings (i.e., st7022j0, st7052j0, st7121j0, and st7132j0) are the signals recorded from the subjects with mild difficulty in falling sleep during a night in the hospital. However, they were healthy otherwise. The dataset contains the EEG and horizontal EOG signals. The EEG signals were recorded from two bipolar channels, Pz-Oz and Fpz-Cz. The sampling rate was 100 Hz.

In this paper, the EEG signal recorded from the Pz-Oz channel is used, since it has been shown in previous studies [15], [30], [31] that the accuracy rate obtained is higher if the signal of this bipolar channel is used instead of that of the Fpz-Cz channel. Each 30-s epoch of the EEG signal is scored according to the Rechtschaffen and Kales (R&K) guidelines by an expert as stage 1 (S1), stage 2 (S2), stage 3 (S3), stage 4 (S4), REM, wake (W), movement time, or unscored.

The second dataset used in this paper has been provided by St. Vincent's University Hospital and University College Dublin (UCDDB). It is also available on PhysioNet [32]. The polysomnography (PSG) signals of 25 adult individuals with suspected sleep-disordered breathing have been recorded overnight. The participants (i.e., 21 males and 4 females) are between ages of 28 and 68. The dataset contains two EEG signals (i.e., C3-A2 and C4-A1), two EOG signals (i.e., left EOG and right EOG), the submental EMG signal, 3-channel ECG signals, the oronasal airflow, ribcage movements, abdomen movements, the oxygen saturation, snoring and the body position. The sampling rate was 128 Hz and the average recording time was 6.9 hours. The sleep stages have been scored by an experienced sleep technologist according to the standard R&K rules. Each 30-s epoch is scored as wake (W), stage 1 (S1), stage 2 (S2), stage 3 (S3), stage 4 (S4), REM, artifact, or indeterminate. In this paper, the signal of the C3-A2 channel is used.

The third dataset used in this paper is the Expanded Sleep-EDF database (XSEDFDB) which is available on PhysioNet [28], [33]. In this dataset, which is an expanded version of the SEDFDB, PSG signals have been recorded from 10 male and 10 female subjects between the ages of 25 and 34. The subjects are healthy Caucasian, without any sleep-related medication. The dataset contains the PSG signals of about 20 hours recorded during two subsequent day-night periods at the subjects' homes. The dataset is similar to the SEDFDB otherwise. In this paper, the EEG signal recorded from the Pz-Oz channel is used.

TABLE II
The Number of EEG Epochs Available for the Wake and Sleep Stages in the Three Datasets

	Wake	Stage 1	Stage 2	Stage 3	Stage 4	REM	Total
SEDFDB	8053	599	3616	663	645	1599	15175
UCDDB	4467	3403	7005	683	2002	3115	20675
XSEDFDB	72523	2813	17833	3385	2357	7630	106541

The six sleep states considered here are W, S1, S2, S3, S4, and REM. The number of epochs available for the wake and sleep stages in the three datasets are presented in Table II.

IV. Feature Extraction

A brief explanation of the features used in this paper is given in this section. Please refer to Table III for the list of the features.

A. Standard Deviation

The standard deviation (SD) can be used as a time-domain feature that quantifies the amount of variation or dispersion of the data. It is usually estimated by the following expression for signal $x(n)$:

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x(n)-\bar{x})^2} \tag{1}$$

where:

$$\bar{x} = \frac{1}{N}\sum_{n=1}^{N} x(n) \tag{2}$$

B. Hjorth Parameters

Hjorth parameters are used to determine the activity, mobility, and complexity of the EEG signals [34]. They are calculated using the signal, say $x(n)$, and its first and second derivatives (i.e., $x'(n)$ and $x''(n)$, respectively) as follows:

$$Hjorth\ Activity = \sigma_x^2 \tag{3}$$

$$Hjorth\ Mobility\ (HM) = \sigma_{x'}/\sigma_x \tag{4}$$

$$Hjorth\ Complexity\ (HC) = (\sigma_{x''}/\sigma_{x'})/(\sigma_{x'}/\sigma_x) \tag{5}$$

where σ_x, $\sigma_{x'}$, and $\sigma_{x''}$ are the standard deviations of $x(n)$, $x'(n)$, and $x''(n)$, respectively.

C. Maximum-Minimum Distance

The maximum-minimum distance (MMD) is a feature which is based on a distance formula derived from the Pythagorean Theorem [10]. The distance between the maximum and minimum points in the k^{th} sliding window is calculated using the Pythagorean distance formula as:

$$D_k = \sqrt{\Delta x_k^2 + \Delta y_k^2} \tag{6}$$

TABLE III
FEATURES EXTRACTED FROM THE EEG SIGNAL

Feature set	No.	δ rhythm	No.	θ rhythm	No.	α rhythm	No.	σ rhythm	No.	β_1 rhythm	No.	β_2 rhythm	No.	γ_1 rhythm	No.	γ_2 rhythm
Standard Deviation (SD)	1	SD_δ	14	SD_θ	27	SD_α	40	SD_σ	53	SD_{β_1}	66	SD_{β_2}	79	SD_{γ_1}	92	SD_{γ_2}
Hjorth Mobility (HM)	2	HM_δ	15	HM_θ	28	HM_α	41	HM_σ	54	HM_{β_1}	67	HM_{β_2}	80	HM_{γ_1}	93	HM_{γ_2}
Hjorth Complexity (HC)	3	HC_δ	16	HC_θ	29	HC_α	42	HC_σ	55	HC_{β_1}	68	HC_{β_2}	81	HC_{γ_1}	94	HC_{γ_2}
Maximum-Minimum Distance (MMD)	4	MMD_δ	17	MMD_θ	30	MMD_α	43	MMD_σ	56	MMD_{β_1}	69	MMD_{β_2}	82	MMD_{γ_1}	95	MMD_{γ_2}
Petrosian Fractal Dimension (PFD)	5	PFD_δ	18	PFD_θ	31	PFD_α	44	PFD_σ	57	PFD_{β_1}	70	PFD_{β_2}	83	PFD_{γ_1}	96	PFD_{γ_2}
Normalized Line Length (NLL)	6	NLL_δ	19	NLL_θ	32	NLL_α	45	NLL_σ	58	NLL_{β_1}	71	NLL_{β_2}	84	NLL_{γ_1}	97	NLL_{γ_2}
Generalized Hurst Exponent (GHE)	7	GHE_δ	20	GHE_θ	33	GHE_α	46	GHE_σ	59	GHE_{β_1}	72	GHE_{β_2}	85	GHE_{γ_1}	98	GHE_{γ_2}
Log Root Sum of Sequential Variations (LRSSV)	8	$LRSSV_\delta$	21	$LRSSV_\theta$	34	$LRSSV_\alpha$	47	$LRSSV_\sigma$	60	$LRSSV_{\beta_1}$	73	$LRSSV_{\beta_2}$	86	$LRSSV_{\gamma_1}$	99	$LRSSV_{\gamma_2}$
Normalized Spectral Entropy (NSE)	9	NSE_δ	22	NSE_θ	35	NSE_α	48	NSE_σ	61	NSE_{β_1}	74	NSE_{β_2}	87	NSE_{γ_1}	100	NSE_{γ_2}
Rényi Entropy (RE)	10	RE_δ	23	RE_θ	36	RE_α	49	RE_σ	62	RE_{β_1}	75	RE_{β_2}	88	RE_{γ_1}	101	RE_{γ_2}
Kraskov Entropy (KE)	11	KE_δ	24	KE_θ	37	KE_α	50	KE_σ	63	KE_{β_1}	76	KE_{β_2}	89	KE_{γ_1}	102	KE_{γ_2}
Phase Mean (PM)	12	PM_δ	25	PM_θ	38	PM_α	51	PM_σ	64	PM_{β_1}	77	PM_{β_2}	90	PM_{γ_1}	103	PM_{γ_2}
Phase Standard Deviation (PSD)	13	PSD_δ	26	PSD_θ	39	PSD_α	52	PSD_σ	65	PSD_{β_1}	78	PSD_{β_2}	91	PSD_{γ_1}	104	PSD_{γ_2}

where Δx_k and Δy_k refer to the x-axis difference and the y-axis difference, respectively, of the maximum and minimum points in the k^{th} window. The MMD is obtained by summing the distances of all sliding windows as follows:

$$MMD = \sum_{k=1}^{N} D_k \tag{7}$$

where N refers to the total number of the sliding windows in an epoch of a signal.

D. Petrosian Fractal Dimension

The complexity of a signal can be measured by the fractal dimension. Transforming the signal into binary sequences, the Petrosian fractal dimension or PFD simplifies calculation of this chaotic method [35]. The PFD can be estimated as follows:

$$PFD = \frac{log_{10}N}{log_{10}N + log_{10}(N/(N+0.4M))} \tag{8}$$

where N refers to the number of the signal samples and M is the number of the sign changes in the signal derivative.

E. Normalized Line Length

The normalized line length (NLL) is used to estimate the Katz fractal dimension [36]. The NLL has been proposed by Esteller *et al.* [37] as:

$$NLL(n) = \frac{1}{M} \sum_{m=n-N+1}^{n} |x(m) - x(m-1)| \tag{9}$$

where $x(n)$ is an EEG signal epoch, N is the length of the sliding window, and M, which serves as a normalizing factor, is the number of the window shifts that fit into the sliding window. In this paper, $M = 1$.

F. Generalized Hurst Exponent

Scaling properties of a signal can be studied using the generalized Hurst exponent (GHE) [38]. If $x(n)$ is a time series, the GHE of $x(n)$ is defined as follows:

$$\frac{\langle |x(n+d) - x(n)|^m \rangle}{\langle |x(n)|^m \rangle} \propto d^{mH(m)} \tag{10}$$

where d is a time lag, $\langle \cdot \rangle$ is the sample average operator, and $H(m)$ is the GHE of $x(n)$.

In this paper, $5 \leq d \leq 19$, $m = 1$ [39], and $x(n)$ is proposed to be the cumulative sum of the EEG epoch. For more information on GHE, please refer to [38]–[42].

G. Log Root Sum of Sequential Variations

The log root sum of sequential variations (LRSSV) is proposed in this paper to measure the sequential variations between the samples of the signal. The LRSSV is calculated using the following expression:

$$LRSSV = log_{10} \sqrt{\sum_{n=1}^{N-1} (x(n) - x(n-1))^2} \tag{11}$$

where N is the length of the signal $x(n)$.

H. Normalized Spectral Entropy

The normalized spectral entropy (NSE) can be estimated from the normalized power spectrum of the signal [43], [44]. If the power spectrum is normalized, then $\sum S(f) = 1$. The NSE shows the irregularity of the spectrum [43], [44]. The NSE is calculated as follows [44]:

$$NSE = \frac{1}{log_2 N_f} \sum_{f=f_1}^{f_2} S(f) log_2 \frac{1}{S(f)} \tag{12}$$

where f_1 and f_2 are the lower frequency and upper frequency of the power spectrum, respectively, S is the power density, and N_f is the number of frequencies within the band $[f_1, f_2]$. In this paper, f_1 and f_2 are set to 0 Hz and 50 Hz, respectively.

I. Rényi Entropy

The Rényi entropy (RE) can be considered as a measure of the entropy of the distribution $P = (p_1, p_2, \ldots, p_n)$ [45]. The RE is defined by:

$$RE^{(m)} = \frac{1}{1-m} log_2 (\sum_{i=1}^{n} p_i^m) \tag{13}$$

where m is called the order of the RE, $m > 0$, and $m \neq 1$. In this paper, $m = 2$.

J. Kraskov Entropy

The Kraskov entropy (KE) is an estimate for Shannon entropy using N samples of an m-dimensional random vector $\mathbf{x}$. The KE is defined as:

$$KE = -\psi(k) + \psi(N) + \log(v_m) + \frac{m}{N} \sum_{i=1}^{N} \log(2r_i) \tag{14}$$

where ψ is the digamma function (i.e., the logarithmic derivative of the gamma function), k is the number of the nearest neighbors, N is the number of samples, m is the dimension of $\mathbf{x}$, v_m is the volume of the m-dimensional unit sphere, and r_i is the distance from $\mathbf{x}_i$ to its k-th nearest neighbor [46], [47].

In this paper, k is set to be the square root of the EEG epoch length.

K. Phase Mean and Phase Standard Deviation

The phase mean (PM) and the phase standard deviation (PSD) are the last features used in this paper. The phase of the signal is obtained from the analytic signal associated with the (real) signal. The analytic signal is a complex signal whose imaginary part is the Hilbert transform [48] of its real part [49]. The analytic signal, $z(t)$, is hence obtained from the real EEG signal, $x(t)$ as:

$$z(t) = x(t) + jy(t) \tag{15}$$

where $y(t)$ is the Hilbert transform of $x(t)$.

If we write $z(t)$ in the polar form, we have:

$$z(t) = A(t)e^{j\phi(t)} \tag{16}$$

where $A(t)$ and $\phi(t)$ are the magnitude and phase of $z(t)$, respectively.

$A(t)$ and $\phi(t)$ also refer to the envelope and phase of the real signal $x(t)$, respectively [49]. In this paper, the mean and standard deviation of $\phi(t)$ are used as features.

V. Feature Selection

The statistical significance of the features discussed in the previous section is examined using the Kruskal-Wallis test [50], which is a method for nonparametric distribution-free one-way ANOVA. The statistically non-significant features resulting $p > 0.01$ are then discarded. The minimal-redundancy-maximal-relevance (mRMR) algorithm is employed to select the best features remained after the statistical test. The mRMR feature selection is a method which selects the features with the highest relevance to the target classes while it minimizes the redundancy among the selected features [51], [52]. For more information on the mRMR algorithm, please refer to [51] and [52]. The code provided by [53] is used in this paper.

VI. Classification

Classification process in this paper is performed using a random forest (RF). An RF is a combination of tree-structured classifiers where the random vectors generated for trees are independent identically distributed and each tree has one vote for each input [54]. An RF works as follows:

1. Sampling (or resampling) with replacement is performed on the original dataset. This type of (re)sampling is called bootstrapping. The new sub-datasets are hence formed.
2. Different trees are built (grown) on each of the new sub-datasets without pruning.
3. Once the trees are grown, majority voting is applied to the outputs (i.e., classes predicted) by the trees [54], [55].

TABLE IV
Different n-Class Classification Problems Considered in This Paper

Classification	Wake and Sleep Stages
6-class problem	W, S1, S2, S3, S4, and REM
5-class problem	W, S1, S2, S3+S4, and REM
4-class problem	W, S1+S2, S3+S4, and REM
3-class problem	W, S1+S2+S3+S4 (NREM), and REM
2-class problem	W and S1+S2+S3+S4+REM (Sleep)

At each node of a tree, some features are randomly selected from the feature set. The split of the node is then performed based on the selected features. In this paper, the number of the features selected randomly for each node is set to be the square root of the total number of the features.

VII. System Performance Evaluation

Four criteria (measures) are used to evaluate the performance of the proposed sleep stage classification system. These criteria are the classification accuracy, sensitivity, specificity, and confusion matrix.

The classification accuracy, sensitivity, and specificity are defined by:

$$Accuracy\ (Ac) = \frac{TP + TN}{TP + FN + TN + FP}\ (\%) \tag{17}$$

$$Sensitivity\ (Sn) = \frac{TP}{TP + FN}\ (\%) \tag{18}$$

$$Specificity\ (Sp) = \frac{TN}{TN + FP}\ (\%) \tag{19}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

In order to set the system parameters and to evaluate the system performance, we use a nested k-fold cross-validation process. To do this, the dataset is divided randomly into k equal-size subsets. At each fold, the $(k-1)$ subsets are used as the training and validating data and 1 subset is used for testing. This process is repeated k times (the number of folds), with each subset used exactly once for testing. This makes the outer folds. For each outer fold, k inner folds are made such that the training and validating data is further divided into k equal-size subsets. At each inner fold, the $(k-1)$ subsets are used as the training data and 1 subset is used for validating. This process is repeated k times, with each subset used exactly once for validating. The results of the $k \times k$ inner folds are used for setting the system parameters. The results of the k outer folds are averaged and reported as the system performance.

We also use a k-fold subject cross-validation process, where k is the number of subjects in the dataset. At each fold, the data of one subject is used for testing, while the data of other subjects is used as the training and validating data. This process is repeated k times. The results of the k folds are averaged and reported as the system performance.

VIII. Experimental Results and Discussion

Sleep is divided into REM and non-REM (NREM) sleep. Stages S1, S2, S3, and S4 are pertained to non-REM (NREM)

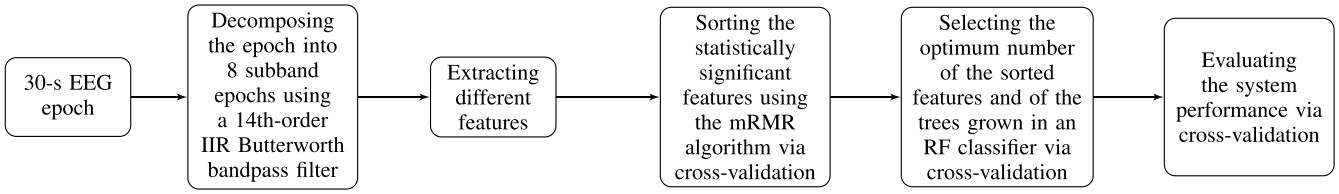

Fig. 1. Block diagram of the proposed sleep stage classification system.

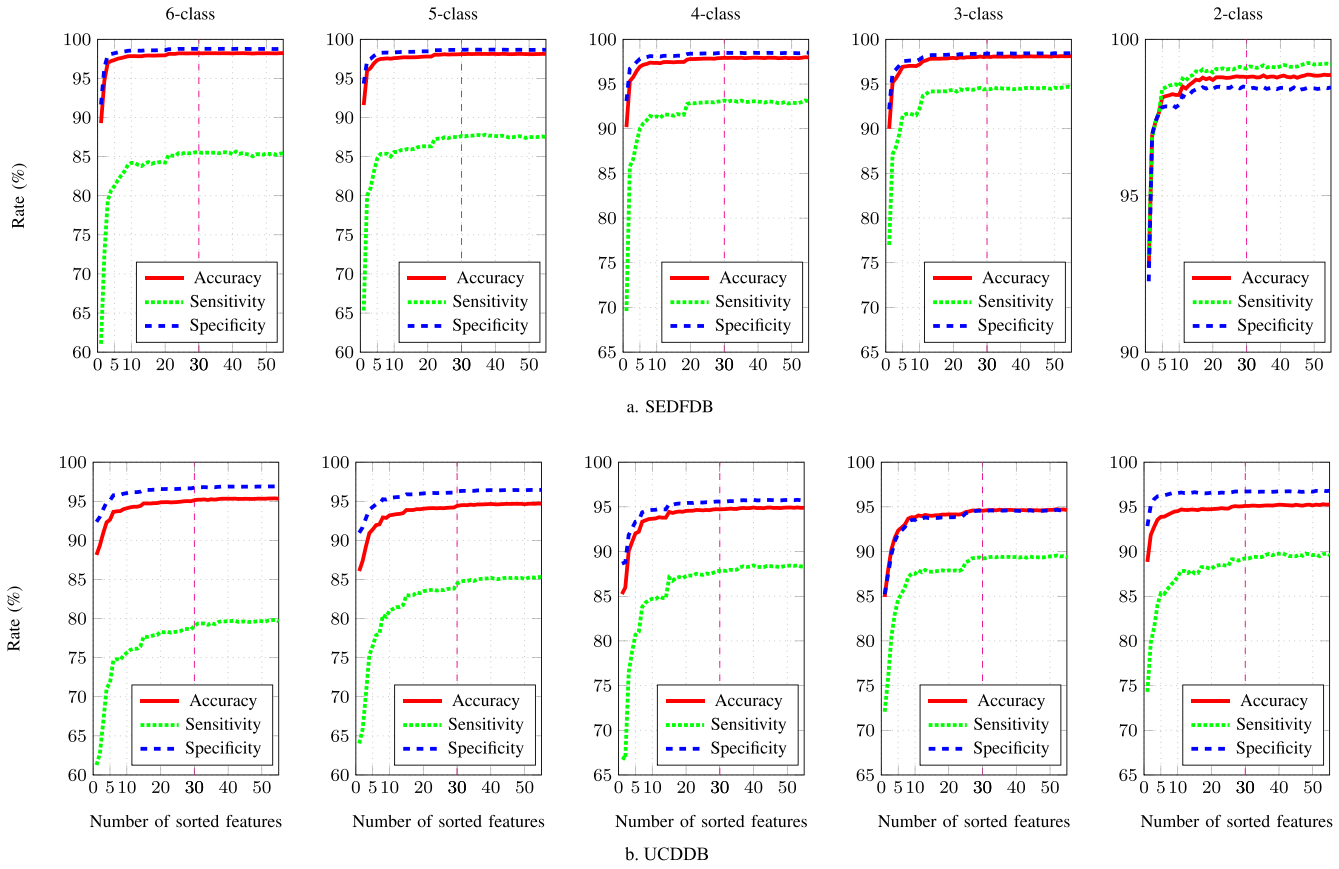

Fig. 2. The average of accuracy, sensitivity, and specificity rates versus the number of sorted features in the *n*-class problems. (a) SEDFDB. (b) UCDDB.

sleep. NREM sleep is further divided into shallow sleep and deep sleep [56]. Shallow sleep is during stages S1 and S2 while deep sleep occurs during S3 and S4.

In order to diagnose sleep-related disorders, different sleep stages must be detected and classified. For example, REM sleep detection is a necessary task to diagnose REM sleep behavior disorder (RBD) [57], or classification between wake and sleep must be done for sleep monitoring. In this paper, classification between different sleep stages and wake is done to address these needs. The multi-class classification problems considered here are given in Table IV.

The block diagram of the proposed multi-class sleep stage classification system is illustrated in Figure 1. The three datasets used in this paper contain EEG signals during 6 different states of wake (W), S1, S2, S3, S4, and REM.

A. Nested 5-Fold Cross-Validation Results

The EEG signal is first divided into non-overlapping 30-s epochs. The EEG epoch is decomposed into 8 subband epochs using a 14th-order IIR Butterworth bandpass filter such that each subband epoch has a frequency band of one of the eight EEG rhythms (i.e., δ, θ, α, σ, β_1, β_2, γ_1, or γ_2). This is done since during different stages of sleep and wake, different EEG rhythms become more dominant. Please refer to Table I. The 14th-order IIR Butterworth bandpass filter is used because of its simplicity, however, a more complex method can be used instead.

Thirteen features are extracted for each subband epoch. Hence, there exist a total of 104 features. Please refer to Table III. Using the Kruskal-Wallis statistical test, the non-significant features, which result in $p > 0.01$, are discarded. Using the mRMR algorithm [53] via nested 5-fold cross-validation, the remaining features are sorted for the different multi-class classification problems.

An RF classifier with 55 trees is trained via inner folds of a nested 5-fold cross-validation process using m features of the (first 55) sorted features ($m = 1, 2, \ldots, 55$). The average of the accuracy, sensitivity, and specificity rates of the wake

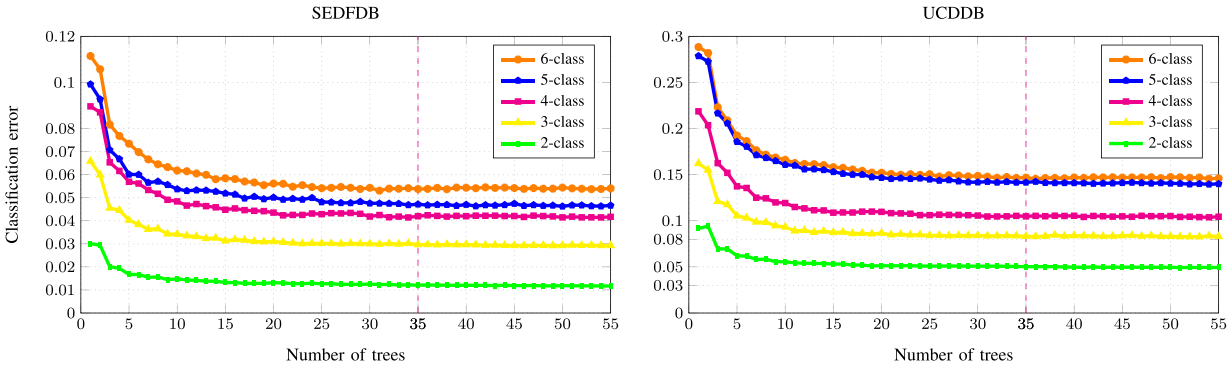

Fig. 3. Classification error versus the number of trees grown in the RF.

TABLE V
THE PERFORMANCE OF THE PROPOSED SYSTEM FOR THE 6-CLASS PROBLEM

		SEDFDB							UCDDB							
		Proposed system score								Proposed system score						
		W	**S1**	**S2**	**S3**	**S4**	**REM**	**Overall**		**W**	**S1**	**S2**	**S3**	**S4**	**REM**	**Overall**
Expert's score	**W**	8027	6	5	0	0	15	8053	**W**	4159	200	84	0	1	23	4467
	S1	83	408	41	0	1	66	599	**S1**	402	2496	376	0	0	129	3403
	S2	21	15	3470	57	4	49	3616	**S2**	141	170	6524	8	52	110	7005
	S3	5	3	85	520	50	0	663	**S3**	8	5	206	395	69	0	683
	S4	3	0	7	43	592	0	645	**S4**	10	6	172	13	1800	1	2002
	REM	23	26	76	0	0	1474	1599	**REM**	59	99	223	1	7	2726	3115
Sn (%)		99.68	68.11	95.96	78.43	91.78	92.18	87.69		93.10	73.35	93.13	57.83	89.91	87.51	82.47
Sp (%)		98.10	99.66	98.15	99.31	99.62	99.04	98.98		96.17	97.22	92.24	99.89	99.31	98.50	97.22
Ac (%)		98.94	98.41	97.63	98.40	99.29	98.32	98.50		95.51	93.29	92.54	98.50	98.40	96.85	95.85

TABLE VI
THE PERFORMANCE OF THE PROPOSED SYSTEM FOR THE 5-CLASS, 4-CLASS, 3-CLASS, AND 2-CLASS PROBLEMS

		5-class						**4-class**					**3-class**				**2-class**
		W	**S1**	**S2**	**S3+S4**	**REM**	**Overall**	**W**	**S1+S2**	**S3+S4**	**REM**	**Overall**	**W**	**NREM**	**REM**	**Overall**	**W/Sleep**
SEDFDB	Sn (%)	99.61	67.88	95.84	93.13	92.19	89.73	99.48	94.16	93.73	89.75	94.28	99.45	97.09	89.56	95.37	99.23
	Sp (%)	98.10	99.68	98.16	99.47	99.10	98.90	98.35	97.83	99.46	99.23	98.72	98.39	98.24	99.35	98.66	98.69
	Ac (%)	98.91	98.41	97.61	98.92	98.37	98.44	98.95	96.81	98.97	98.23	98.24	98.96	97.82	98.32	98.37	98.98
UCDDB	Sn (%)	92.97	74.06	92.62	87.94	87.60	87.04	91.31	93.40	87.90	85.34	89.49	91.42	95.55	84.90	90.62	89.87
	Sp (%)	96.35	97.13	92.90	99.32	98.61	96.86	96.93	89.58	99.39	98.97	96.22	97.14	89.61	98.95	95.23	97.38
	Ac (%)	95.61	93.33	92.80	97.85	96.95	95.31	95.72	91.50	97.91	96.92	95.51	95.91	93.37	96.84	95.37	95.75

and sleep stages are calculated for $m = 1, 2, \ldots, 55$. This procedure is performed for all n-class problems (i.e., 6-class, 5-class, 4-class, 3-class, and 2-class problems) using both the SEDFDB and the UCDDB. Please refer to Figure 2 for the plots. Based on the plots, the optimum number of features is selected to be 30.

After setting the number of features, the classification error of the RF classifier for different number of trees $(1, 2, \ldots, 55)$ in 6-class, 5-class, 4-class, 3-class, and 2-class problems is measured and plotted for both the SEDFDB and the UCDDB. Please refer to Figure 3. Based on the classification error curves, the optimum number of trees in the RF is set to be 35.

The system performance for the 6-class problem is given in Table V. The table shows the confusion matrix and sensitivity, specificity, and accuracy rates for the two datasets. The system performance for the 5-class, 4-class, 3-class, and 2-class problems is shown in Table VI in terms of the sensitivity, specificity, and accuracy rates.

For the 6-class problem, using the SEDFDB, the sensitivity rates obtained for W, S1, S2, S3, S4, and REM states are 99.68%, 68.11%, 95.96%, 78.43%, 91.78%, and 92.18%, respectively, with an overall value of 87.69%. The specificity rate is higher than 98.10% for each state.

For the 6-class problem, when the UCDDB is used, the sensitivity rates obtained for states W, S1, S2, S3, S4, and REM are 93.10%, 73.35%, 93.13%, 57.83%, 89.91%, and 87.51%, respectively, with an overall value of 82.47%. The specificity rate is higher than 92.24% for each state.

As it can be seen, the performance of the system is very good in terms of the sensitivity and specificity rates. The system also performs very well for the other n-class problems.

TABLE VII
THE PERFORMANCE OF SLEEP STAGE CLASSIFICATION SYSTEMS USING THE SEDFDB

System		6-class W	S1	S2	S3	S4	REM	5-class W	S1	S2	S3+S4	REM	4-class W	S1+S2	S3+S4	REM	3-class W	NREM	REM	2-class W/Sleep
[58]	Sn (%)	41.07	33.70	92.99	76.36	**96.36**	79.69	—	—	—	—	—	—	—	—	—	—	—	—	—
	Sp (%)	**99.77**	95.02	84.68	97.56	98.68	**99.20**	—	—	—	—	—	—	—	—	—	—	—	—	—
[59]	Sn (%)	—	—	—	—	—	—	88.43	35.12	87.01	90.80	90.51	—	—	—	—	—	—	—	—
	Sp (%)	—	—	—	—	—	—	97.06	97.38	92.47	98.78	95.55	—	—	—	—	—	—	—	—
[30]	Sn (%)	83.71	35.83	89.03	28.66	75.97	80.38	—	—	—	—	—	82.58	84.72	75.04	78.06	80.32	93.36	77.19	96.90
	Sp (%)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
[19]	Sn (%)	—	—	—	—	—	—	70.83	36.36	**97.16**	89.66	89.50	—	—	—	—	—	—	—	—
	Sp (%)	—	—	—	—	—	—	**99.32**	98.75	87.57	98.12	98.13	—	—	—	—	—	—	—	—
[15]	Sn (%)	—	—	—	—	—	—	98.77	15.77	90.14	77.30	76.21	—	—	—	—	—	—	—	—
	Sp (%)	—	—	—	—	—	—	96.83	99.57	93.26	98.89	96.58	—	—	—	—	—	—	—	—
[14]	Sn (%)	97.93	39.07	90.45	42.56	82.17	78.51	98.21	39.74	89.56	79.85	78.14	—	—	—	—	—	—	—	—
	Sp (%)	94.59	99.30	94.15	99.20	99.13	97.65	95.26	99.41	94.49	99.05	97.73	—	—	—	—	—	—	—	—
[12]	Sn (%)	95.06	47.35	89.01	78.27	38.85	78.39	95.28	47.02	92.38	90	80.87	—	—	—	—	—	—	—	—
	Sp (%)	96.05	98.56	94.35	98.35	99.42	96.16	96.47	99.10	95.06	99.02	97.63	—	—	—	—	—	—	—	—
[8]	Sn (%)	95.16	42.05	79.51	**86.61**	48.09	80.50	—	—	—	—	—	—	—	—	—	—	—	—	—
	Sp (%)	96.52	97.96	96.20	98.28	98.41	94.52	—	—	—	—	—	—	—	—	—	—	—	—	—
[9]	Sn (%)	98.81	40.73	90.78	48.21	81.85	83.35	98.88	39.74	90.23	82.31	82.98	—	—	—	—	—	—	—	—
	Sp (%)	96.41	99.14	95.75	98.57	99.20	97.64	96.44	99.26	95.90	98.95	97.57	—	—	—	—	—	—	—	—
[1]	Sn (%)	—	—	—	—	—	—	91.69	**84.23**	93.87	88.12	91.36	—	—	—	—	—	—	—	—
	Sp (%)	—	—	—	—	—	—	98.73	98.75	94.98	97.55	98.02	—	—	—	—	—	—	—	—
[7]	Sn (%)	99.2	18.9	92.5	52.54	83.3	83.70	99.30	18.7	92.10	83.80	83.2	—	—	—	—	—	—	—	—
	Sp (%)	95.9	**99.80**	95.20	98.70	99.10	97.80	95.90	**99.80**	95.30	98.9	97.80	—	—	—	—	—	—	—	—
Proposed system	Sn (%)	**99.68**	**68.11**	**95.96**	78.43	91.78	**92.18**	**99.61**	67.88	95.84	**93.13**	**92.19**	**99.48**	**94.16**	**93.73**	**89.75**	**99.45**	**97.09**	**89.56**	**99.23**
	Sp (%)	98.10	99.66	**98.15**	**99.31**	**99.62**	99.04	98.10	99.68	**98.16**	**99.47**	**99.10**	**98.35**	**97.83**	**99.46**	**99.23**	**98.39**	**98.24**	**99.35**	**98.69**

TABLE VIII
THE PERFORMANCE OF SLEEP STAGE CLASSIFICATION SYSTEMS USING THE UCDDB

System		6-class W	S1	S2	S3	S4	REM	Overall	5-class W	S1	S2	S3+S4	REM	Overall	4-class W	S1+S2	S3+S4	REM	Overall	3-class W	NREM	REM	Overall	2-class W/Sleep
[60]	Sn (%)	**100**	**97.55**	**94.23**	**96.20**	**97.12**	**96.60**	**96.95**	**100**	**98.22**	**95.85**	**97.94**	**97.84**	**97.97**	—	—	—	—	—	—	—	—	—	—
	Sp (%)	**100**	**99.25**	**98.92**	99.02	**99.55**	**99.67**	**99.40**	**100**	**99.34**	**99.30**	99.19	**99.63**	**99.49**	—	—	—	—	—	—	—	—	—	—
	Ac (%)	**100**	**98.95**	**98.02**	98.43	**99.19**	**99.48**	**99.01**	**100**	**99.15**	**98.64**	**98.74**	**99.52**	**99.21**	—	—	—	—	—	—	—	—	—	—
[8]	Sn (%)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	Sp (%)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	Ac (%)	—	—	—	—	—	—	82.75	—	—	—	—	—	84.18	—	—	—	—	92.83	—	—	—	90.90	**99.02**
[9]	Sn (%)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	Sp (%)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	Ac (%)	—	—	—	—	—	—	80.08	—	—	—	—	—	81.24	—	—	—	—	92.65	—	—	—	**97.76**	98.56
Proposed system	Sn (%)	93.10	73.35	93.13	57.83	89.91	87.51	82.47	92.97	74.06	92.62	87.94	87.60	87.04	**91.31**	**93.40**	**87.90**	**85.34**	**89.49**	**91.42**	**95.55**	**84.90**	**90.62**	**89.87**
	Sp (%)	96.17	97.22	92.24	**99.89**	99.31	98.50	97.22	96.35	97.13	92.90	**99.32**	98.61	96.86	**96.93**	**89.58**	**99.39**	**98.97**	**96.22**	**97.14**	**89.61**	**98.95**	**95.23**	**97.38**
	Ac (%)	95.51	93.29	92.54	**98.50**	98.40	96.85	95.85	95.61	93.33	92.80	97.85	96.95	95.31	**95.72**	**91.50**	**97.91**	**96.92**	**95.51**	**95.91**	**93.37**	**96.84**	95.37	95.75

The overall sensitivity rate for the 5-class, 4-class, 3-class, and 2-class problems using the SEDFDB are 89.73%, 94.28%, 95.37%, and 99.23%, respectively. The corresponding overall specificity rates are, respectively, 98.90%, 98.72%, 98.66%, and 98.69%.

The pair of (overall sensitivity rate, overall specificity rate) for the 5-class, 4-class, 3-class, and 2-class problems using the UCDDB are (87.04%, 96.86%), (89.49%, 96.22%), (90.62%, 95.23%), and (89.87%, 97.38%), respectively. For more details, please refer to Table VI.

As shown in the tables, the performance of the system in REM detection (in order to diagnose RBD) and in wake/sleep classification (to be used in sleep monitoring) is very good.

The subjects of the SEDFDB are healthy, while the subjects of the UCDDB have suspected sleep-disordered breathing. The performance of the proposed system on the two datasets implies that the proposed system can be used for sleep monitoring of both healthy subjects and subjects with suspected sleep-disordered breathing.

The performance of some of the existing sleep stage classification systems are given and compared with the performance of the proposed system in Table VII and Table VIII. All of the systems reported in Table VII have used the SEDFDB

TABLE IX

p-VALUES OBTAINED FOR FEATURES USING THE KRUSKAL-WALLIS TEST IN THE 4-CLASS PROBLEM (SUBJECT CROSS-VALIDATION RESULTS)

δ rhythm	*p*-value	θ rhythm	*p*-value	α rhythm	*p*-value	σ rhythm	*p*-value	β_1 rhythm	*p*-value	β_2 rhythm	*p*-value	γ_1 rhythm	*p*-value	γ_2 rhythm	*p*-value
1. SD_δ	0 (0)	14. SD_θ	0 (0)	27. SD_α	0 (0)	40. SD_σ	0 (0)	53. SD_{β_1}	0 (0)	66. SD_{β_2}	0 (0)	79. SD_{γ_1}	0 (0)	92. SD_{γ_2}	0 (0)
2. HM_δ	0 (0)	15. HM_θ	0 (0)	28. HM_α	0 (0)	41. HM_σ	0 (0)	54. HM_{β_1}	0 (0)	67. HM_{β_2}	0 (0)	80. HM_{γ_1}	0 (0)	93. HM_{γ_2}	0 (0)
3. HC_δ	0 (0)	16. HC_θ	0 (0)	29. HC_α	0 (0)	42. HC_σ	0 (0)	55. HC_{β_1}	0 (0)	68. HC_{β_2}	0 (0)	81. HC_{γ_1}	0 (0)	94. HC_{γ_2}	0 (0)
4. MMD_δ	0 (0)	17. MMD_θ	0 (0)	30. MMD_α	0 (0)	43. MMD_σ	0 (0)	56. MMD_{β_1}	0 (0)	69. MMD_{β_2}	0 (0)	82. MMD_{γ_1}	0 (0)	95. MMD_{γ_2}	0 (0)
5. PFD_δ	0 (0)	18. PFD_θ	0 (0)	31. PFD_α	0 (0)	44. PFD_σ	0 (0)	57. PFD_{β_1}	0 (0)	70. PFD_{β_2}	0 (0)	83. PFD_{γ_1}	0 (0)	96. PFD_{γ_2}	0 (0)
6. NLL_δ	0 (0)	19. NLL_θ	0 (0)	32. NLL_α	0 (0)	45. NLL_σ	0 (0)	58. NLL_{β_1}	0 (0)	71. NLL_{β_2}	0 (0)	84. NLL_{γ_1}	0 (0)	97. NLL_{γ_2}	0 (0)
7. GHE_δ	0 (0)	20. GHE_θ	0 (0)	33. GHE_α	0 (0)	46. GHE_σ	0 (0)	59. GHE_{β_1}	0 (0)	72. GHE_{β_2}	0 (0)	85. GHE_{γ_1}	0 (0)	98. GHE_{γ_2}	0.001 **(0.023)**
8. $LRSSV_\delta$	0 (0)	21. $LRSSV_\theta$	0 (0)	34. $LRSSV_\alpha$	0 (0)	47. $LRSSV_\sigma$	0 (0)	60. $LRSSV_{\beta_1}$	0 (0)	73. $LRSSV_{\beta_2}$	0 (0)	86. $LRSSV_{\gamma_1}$	0 (0)	99. $LRSSV_{\gamma_2}$	0 (0)
9. NSE_δ	0 (0)	22. NSE_θ	0 (0)	35. NSE_α	0 (0)	48. NSE_σ	0 (0)	61. NSE_{β_1}	0 (0)	74. NSE_{β_2}	0 (0)	87. NSE_{γ_1}	0 (0)	100. NSE_{γ_2}	0 (0)
10. RE_δ	0 (0)	23. RE_θ	0 (0)	36. RE_α	0 (0)	49. RE_σ	0 (0)	62. RE_{β_1}	0 (0)	75. RE_{β_2}	0 (0)	88. RE_{γ_1}	0 (0)	101. RE_{γ_2}	0 (0)
11. KE_δ	0 (0)	24. KE_θ	0 (0)	37. KE_α	0 (0)	50. KE_σ	0 (0)	63. KE_{β_1}	0 (0)	76. KE_{β_2}	0 (0)	89. KE_{γ_1}	0 (0)	102. KE_{γ_2}	0 (0)
12. PM_δ	0 (0)	25. PM_θ	**0.029 (0.281)**	38. PM_α	**0.259 (0.970)**	51. PM_σ	**0.231 (0.891)**	64. PM_{β_1}	**0.145 (0.872)**	77. PM_{β_2}	0 (0)	90. PM_{γ_1}	0 (0)	103. PM_{γ_2}	0.006 **(0.091)**
13. PSD_δ	0 (0)	26. PSD_θ	0 (0)	39. PSD_α	**0.136 (0.681)**	52. PSD_σ	0 (0)	65. PSD_{β_1}	0 **(0.011)**	78. PSD_{β_2}	0 (0)	91. PSD_{γ_1}	0 (0)	104. PSD_{γ_2}	0 (0)

* All *p*-values calculated for 20 subjects of the XSEDFDB are expressed in the form of "Mean (Maximum)".
* *p*-values less than 0.0001 are rounded to zero.

TABLE X

SYSTEM PERFORMANCE WITH DIFFERENT FEATURES FOR THE 4-CLASS PROBLEM (SUBJECT CROSS-VALIDATION RESULTS)

			XSEDFDB										UCDDB									
			With Gamma					Without Gamma					With Gamma					Without Gamma				
Block	**Feature**		**W**	**S1+S2**	**S3+S4**	**REM**	**Overall**	**W**	**S1+S2**	**S3+S4**	**REM**	**Overall**	**W**	**S1+S2**	**S3+S4**	**REM**	**Overall**	**W**	**S1+S2**	**S3+S4**	**REM**	**Overall**
		Sn (%)	97.77	76.78	55.39	37.65	66.90	97.70	55.46	49.61	16.27	54.76	75.28	74.83	29.72	41.22	55.26	77.59	75.68	21.87	1.92	44.27
1	**SD**	Sp (%)	89.56	91.75	99.27	97.71	94.57	63.75	93.02	99.06	99.42	88.81	88.48	60.13	98.45	94.47	85.38	78.81	52.12	99.10	99.50	82.38
		Ac (%)	94.97	88.64	97.12	93.33	93.51	86.53	85.40	96.66	93.40	90.50	86.32	68.30	89.78	87.69	83.02	79.28	64.26	88.95	85.41	79.47
		Sn (%)	93.98	71.35	36.52	45.04	61.72	81.93	73.17	30.60	47.40	58.28	68.16	71.85	28.20	49.74	54.49	56.09	57.98	30.48	61.35	51.48
2	**HM**	Sp (%)	88.07	89.96	98.49	96.14	93.17	92.61	82.28	98.55	92.64	91.52	89.66	59.45	96.44	91.85	84.35	91.65	65.53	95.13	79.03	82.84
		Ac (%)	92.23	86.48	95.20	92.40	91.58	85.40	80.53	94.89	89.33	87.54	84.94	65.31	87.37	86.44	81.02	83.63	61.31	86.66	77.02	77.16
		Sn (%)	93.50	72.29	55.40	46.87	67.01	84.17	74.94	52.80	44.62	64.13	58.58	75.75	46.29	38.70	54.83	41.98	66.51	47.66	49.20	51.34
3	**HC**	Sp (%)	88.57	90.92	98.49	96.06	93.51	93.16	84.46	98.50	93.51	92.41	91.34	55.58	96.31	93.44	84.17	94.14	56.78	94.93	83.57	82.35
		Ac (%)	92.10	87.35	96.10	92.55	92.02	87.11	82.62	95.96	89.95	88.91	83.61	65.50	90.17	85.95	81.31	81.71	61.19	89.19	79.15	77.81
		Sn (%)	89.59	71.24	60.66	5.21	56.68	61.36	74.12	63.45	8.39	51.83	24.04	86.43	6.71	0.99	29.54	24.43	82.05	6.84	2.38	28.92
4	**MMD**	Sp (%)	81.97	84.29	98.09	98.74	90.77	89.00	63.21	96.76	96.30	86.32	88.69	19.81	98.66	99.09	76.56	86.25	23.44	97.72	98.06	76.37
		Ac (%)	87.02	81.66	96.26	92.00	89.24	70.19	65.40	95.18	89.94	80.18	74.13	52.90	86.43	84.81	74.57	72.46	52.49	85.70	84.15	73.70
		Sn (%)	92.86	71.62	39.88	39.18	60.89	79.26	71.50	38.67	44.73	58.54	67.76	72.26	32.49	39.67	53.05	61.16	53.55	34.95	50.32	50.00
5	**PFD**	Sp (%)	87.91	89.35	98.21	95.96	92.86	91.81	81.74	97.92	91.93	90.85	90.55	58.11	96.07	91.15	83.97	92.16	66.82	93.13	77.01	82.28
		Ac (%)	91.41	85.89	95.10	91.91	91.08	83.24	79.75	94.80	88.46	86.56	86.07	64.81	87.57	84.27	80.68	85.76	59.87	85.41	74.05	76.27
		Sn (%)	86.62	78.85	67.89	34.58	66.98	50.41	71.74	64.93	29.24	54.08	72.00	75.05	51.88	50.75	62.42	50.99	78.11	48.55	37.86	53.88
6	**NLL**	Sp (%)	95.31	81.95	98.96	96.98	93.30	91.57	54.20	98.50	94.41	84.67	90.41	64.85	97.33	95.00	86.90	91.24	58.19	97.72	91.78	84.73
		Ac (%)	89.41	81.33	97.25	92.44	90.11	63.59	57.71	96.77	89.59	76.91	87.17	70.54	91.39	89.17	84.57	83.35	68.27	91.24	84.85	81.93
		Sn (%)	92.17	56.70	31.74	10.33	47.74	89.04	54.03	26.09	12.18	45.34	54.38	71.01	31.41	17.82	43.65	52.42	57.42	42.86	20.41	43.28
7	**GHE**	Sp (%)	64.47	89.93	97.56	97.74	87.43	62.01	89.06	97.52	96.31	86.22	87.49	49.83	92.86	94.81	81.25	85.68	57.59	86.15	92.43	80.46
		Ac (%)	83.52	83.61	94.04	91.43	88.15	80.55	82.28	93.65	90.26	86.69	80.62	60.51	84.59	84.00	77.43	78.78	57.91	80.36	82.44	74.87
		Sn (%)	89.65	78.99	62.82	35.53	66.75	49.74	69.31	65.38	30.87	53.82	70.50	75.63	42.95	46.06	58.78	52.56	77.69	41.04	37.57	52.21
8	**LRSSV**	Sp (%)	94.49	84.43	98.98	97.29	93.80	91.37	53.33	97.90	95.06	84.41	89.82	61.16	97.37	95.37	85.93	90.75	56.48	97.53	92.06	84.21
		Ac (%)	91.19	83.26	97.11	92.82	91.09	63.08	56.63	96.25	90.29	76.56	86.22	68.72	90.08	88.95	83.49	83.13	67.03	89.94	85.15	81.31
		Sn (%)	91.74	53.82	14.76	19.04	44.84	37.03	55.75	17.10	24.51	33.60	68.95	72.39	40.59	38.75	55.17	66.22	59.92	39.50	45.85	52.87
9	**NSE**	Sp (%)	59.05	90.17	98.73	97.15	86.27	70.63	48.19	98.42	90.88	77.03	88.70	59.92	97.24	92.90	84.69	84.75	65.13	97.05	86.52	83.36
		Ac (%)	81.46	83.40	94.23	91.52	87.65	47.89	49.99	94.04	86.08	69.50	84.69	66.38	90.00	85.48	81.64	80.93	62.59	89.75	80.80	78.52
		Sn (%)	97.90	70.69	29.23	25.11	55.73	97.20	22.79	11.56	18.97	37.63	69.82	78.90	19.58	15.82	46.03	78.33	41.52	21.18	4.09	36.28
10	**RE**	Sp (%)	81.03	90.62	99.46	97.70	92.20	34.98	94.38	99.67	98.12	81.79	86.83	46.63	98.14	98.28	82.47	53.35	65.48	94.79	99.08	78.18
		Ac (%)	92.24	86.41	95.76	92.38	91.70	77.09	80.37	94.95	92.36	86.19	84.23	63.12	87.69	86.73	80.44	60.12	53.74	84.85	85.46	71.04
		Sn (%)	98.06	77.20	52.61	35.17	65.76	97.54	24.86	40.15	14.54	44.27	77.07	73.05	37.78	43.95	57.96	85.83	64.12	12.20	0.66	40.70
11	**KE**	Sp (%)	88.46	91.79	99.43	97.87	94.39	39.17	94.38	99.25	99.48	83.07	87.95	63.30	98.54	93.97	85.94	66.41	59.08	99.23	99.92	81.16
		Ac (%)	94.84	88.82	97.17	93.21	93.51	78.52	80.61	96.44	93.29	87.22	86.61	68.96	90.92	87.39	83.47	71.75	62.78	87.66	85.50	76.92
		Sn (%)	97.54	3.65	0.05	0	25.31	95.82	4.34	0.24	0.20	25.15	15.85	88.65	5.59	1.48	27.89	17.41	85.20	6.52	2.89	28.00
12	**PM**	Sp (%)	3.67	97.40	99.99	99.98	75.26	4.55	96.06	99.90	99.82	75.08	92.03	13.20	98.95	98.59	75.69	90.29	17.29	98.06	97.58	75.81
		Ac (%)	67.62	79.27	94.66	92.77	83.58	66.72	78.30	94.58	92.64	83.06	74.68	51.02	86.56	84.44	74.18	73.67	51.28	85.93	83.76	73.66
		Sn (%)	98.42	4.02	0.00	0.06	25.63	98.80	1.21	0.03	0.05	25.02	17.51	83.96	15.40	3.21	30.02	19.51	81.17	19.27	4.46	31.10
13	**PSD**	Sp (%)	4.07	98.05	100.00	99.99	75.53	1.24	98.89	99.97	99.95	75.01	91.73	21.39	95.25	97.65	76.50	90.54	25.48	94.36	96.77	76.79
		Ac (%)	68.32	79.85	94.66	92.79	83.90	67.71	80.02	94.64	92.74	83.78	74.93	52.72	84.56	83.92	74.03	74.45	53.34	84.14	83.33	73.82
		Sn (%)	97.99	79.62	75.39	59.62	78.16	97.46	78.18	74.89	57.99	77.13	79.48	79.66	54.89	50.46	66.12	78.16	78.95	59.48	38.42	63.75
14	**All**	Sp (%)	93.69	94.39	99.11	97.63	96.21	90.64	94.86	99.16	97.51	95.54	90.80	69.12	97.99	96.67	88.65	90.44	68.62	98.24	94.75	88.01
		Ac (%)	96.55	91.47	97.85	94.85	95.18	95.12	91.56	97.92	94.57	94.79	88.90	74.47	92.56	90.64	86.64	88.14	73.79	93.39	87.50	85.71

while the systems reported in Table VIII have employed the UCDDB.

As shown in Table VII, the proposed system outperforms the other sleep stage classification systems which have used the SEDFDB, in terms of the sensitivity and specificity rates. In papers [1], [8], [14], [15], [19], [58], the confusion matrices and the sensitivity rates have been reported. In papers [9], [12], the specificity rates have been reported as well. However,

these rates are not consistent with the corresponding confusion matrices. Therefore, the corrected rates based on the confusion matrices are reported here. In paper [30], the sensitivity rates have only been reported. Paper [7], [59] has reported confusion matrices, sensitivity rates, and specificity rates correctly.

To the best of our knowledge, the proposed system in this paper also outperforms the other sleep stage classification systems which have used the whole UCDDB, in terms of the sensitivity, specificity, and accuracy rates (Table VIII). It seems that the system proposed in [60] has better performance than our system, however, it is not a fair comparison as a subset of the UCDDB, and not the whole dataset, has been used in [60] to evaluate the system performance. Therefore, the results provided in [60] are not as robust and realistic as our results in this paper. Furthermore, in [60], the sensitivity rates have been wrongly reported as the accuracy rates. The sensitivity, specificity, and accuracy rates are hence recalculated based on the confusion matrices provided in [60] and shown in Table VIII. In papers [8], [9], overall accuracy rates have been only reported, and confusion matrices, sensitivity rates, and specificity rates have not been given.

B. Subject Cross-Validation Results

The system performance is also evaluated via subject cross-validation for the 4-class problem. The procedure is similar to that followed in the previous subsection (Subsection VIII-A) except that the optimum number of features and trees in the RF is selected to be 40 and 100, respectively.

Table IX shows the results of the Kruskal-Wallis statistical test on the features using the XSEDFDB. The p-values are obtained at each fold. The mean and maximum of p-values are calculated from 20 folds and given in the table. Using a statistical significance level of 0.01, features PM_θ, PM_α, PM_σ, PM_{β_1}, and PSD_α are non-significant based on their mean values. Features PM_{γ_2}, PSD_{β_1}, and GHE_{γ_2} are non-significant based on their maximum values.

The system performance is reported in Table X (Block 14). The table gives the sensitivity, specificity, and accuracy rates which are averaged over 20 subjects of the XSEDFDB and averaged over 25 subjects of the UCDDB. The sensitivity rate for classes W, S1 + S2, S3 + S4, and REM using the XSEDFDB are 97.99%, 79.62%, 75.39%, and 59.62%, respectively. The corresponding specificity rates are, respectively, 93.69%, 94.39%, 99.11%, and 97.63%. The pair of (sensitivity rate, specificity rate) for classes W, S1+S2, S3+S4, and REM using the UCDDB are (79.48%, 90.80%), (79.66%, 69.12%), (54.89%, 97.99%), and (50.46%, 96.67%), respectively.

Table X also shows the performance of each feature set (from the 13 sets given in Table III) separately (Block 1–13). Based on Table X, feature sets SD, LRSSV, HC, NLL, and KE are among the best. HM, PFD, and RE are neither very good nor very bad. PM, PSD, and GHE are among the worst feature sets. It should be noted that each feature set contains 8 features belonging to the 8 subband epochs. Each feature set is hence an 8-dimensional vector.

EEG signals are traditionally low-pass filtered at the upper bound of beta band, i.e. 30 Hz. This is because there is not much evidence for gamma band (>30 Hz) activities in brain signals. However, we have found some studies that show the existence of gamma band activities during waking and sleep [61]–[64]. Therefore, in this paper, some features are extracted from gamma band. The performance of the system with and without features from gamma band is reported in Table X, Block 14. The effect of gamma band on the performance of each of the 13 feature sets is also investigated. The results are shown in Table X, Block 1–13.

The difference in performance between using and not using gamma band is examined by means of the Kruskal-Wallis statistical test. The resultant p-values are given in Table XI. Using a 5% significance level, a p-value less than 0.05 shows a significant difference in performance. These cases are shown in bold in the table. For most feature sets, it is evident that not using gamma band results in significant degradation in performance. However, there is no significant difference when all feature sets are used. This might be because features from other bands compensate for not using gamma band.

TABLE XI
p-VALUES OBTAINED TESTING THE PERFORMANCE DIFFERENCE BETWEEN USING AND NOT USING GAMMA BAND

		XSEDFDB					UCDDB				
Feature		W	S1+S2	S3+S4	REM	Overall	W	S1+S2	S3+S4	REM	Overall
SD	Sn	0.52	**0.00**	0.52	**0.00**	**0.00**	0.47	1.00	0.31	**0.00**	**0.00**
	Sp	**0.00**	0.14	0.57	**0.00**	**0.00**	**0.00**	0.05	0.18	**0.00**	**0.00**
	Ac	**0.00**	**0.01**	0.45	0.80	**0.00**	**0.00**	0.09	0.88	0.20	**0.01**
HM	Sn	**0.00**	0.53	0.32	0.59	0.17	**0.01**	**0.01**	0.68	0.07	0.23
	Sp	**0.02**	**0.00**	0.63	**0.02**	0.05	0.27	0.14	0.26	**0.00**	**0.05**
	Ac	**0.01**	**0.00**	0.55	**0.02**	**0.00**	0.48	**0.03**	0.88	**0.00**	**0.00**
HC	Sn	**0.00**	0.33	0.83	0.50	0.21	**0.00**	**0.01**	0.90	0.07	0.29
	Sp	**0.01**	**0.00**	0.97	0.05	0.26	0.10	0.71	0.13	**0.00**	0.07
	Ac	0.06	**0.01**	0.99	**0.01**	**0.01**	0.38	**0.04**	0.55	**0.00**	**0.00**
MMD	Sn	**0.00**	0.19	0.50	**0.00**	**0.01**	0.79	**0.05**	0.53	**0.00**	0.57
	Sp	**0.00**	**0.00**	**0.01**	**0.00**	**0.00**	0.26	0.07	**0.00**	**0.00**	0.58
	Ac	**0.00**	**0.00**	0.05	**0.01**	**0.00**	0.48	0.93	0.51	0.66	0.43
PFD	Sn	**0.00**	0.94	0.80	0.26	0.42	**0.05**	**0.00**	0.59	0.08	0.11
	Sp	**0.04**	**0.01**	0.70	**0.00**	0.05	0.38	**0.01**	**0.03**	**0.00**	**0.01**
	Ac	**0.00**	**0.00**	0.79	**0.00**	**0.00**	0.95	**0.02**	0.14	**0.00**	**0.00**
NLL	Sn	**0.00**	**0.03**	0.58	0.42	**0.00**	**0.00**	0.36	0.63	0.10	**0.00**
	Sp	0.16	**0.00**	0.27	**0.00**	**0.00**	0.66	0.07	0.91	0.41	**0.00**
	Ac	**0.00**	**0.00**	0.30	**0.01**	**0.00**	0.06	0.24	0.82	**0.04**	**0.03**
GHE	Sn	**0.00**	0.43	0.29	0.26	0.21	0.62	**0.00**	0.10	0.52	0.85
	Sp	0.34	0.19	0.63	**0.01**	0.11	0.32	0.09	**0.01**	**0.00**	0.17
	Ac	**0.04**	0.12	0.55	0.09	**0.02**	0.33	0.31	0.07	0.32	**0.03**
LRSSV	Sn	**0.00**	**0.02**	0.74	0.55	**0.00**	**0.00**	0.42	0.68	0.14	0.05
	Sp	0.25	**0.00**	**0.04**	**0.02**	**0.00**	0.51	0.17	0.96	0.25	**0.03**
	Ac	**0.00**	**0.00**	0.10	**0.01**	**0.00**	0.13	0.45	0.88	0.08	0.05
NSE	Sn	**0.00**	0.70	0.74	0.14	**0.00**	0.60	**0.00**	0.88	0.22	0.33
	Sp	**0.00**	**0.00**	0.10	**0.00**	**0.00**	**0.05**	0.03	0.49	**0.00**	**0.04**
	Ac	**0.00**	**0.00**	0.68	**0.00**	**0.00**	0.08	**0.02**	0.92	**0.01**	**0.00**
RE	Sn	0.36	**0.00**	**0.00**	0.14	**0.00**	**0.02**	**0.00**	0.69	**0.00**	**0.00**
	Sp	**0.00**	**0.00**	**0.05**	0.27	**0.00**	**0.00**	**0.00**	**0.00**	0.07	**0.00**
	Ac	**0.00**	**0.00**	0.24	0.94	**0.00**	**0.00**	**0.00**	0.11	0.52	**0.00**
KE	Sn	0.34	**0.00**	0.11	**0.00**	**0.00**	**0.01**	0.14	**0.00**	**0.00**	**0.00**
	Sp	**0.00**	**0.00**	0.42	**0.00**	**0.00**	**0.00**	0.36	0.15	**0.00**	**0.00**
	Ac	**0.00**	**0.00**	0.15	0.94	**0.00**	**0.00**	0.07	**0.04**	0.36	**0.00**
PM	Sn	**0.00**	0.09	**0.00**	**0.00**	**0.04**	0.31	**0.00**	**0.03**	**0.02**	0.93
	Sp	**0.01**	**0.00**	**0.00**	**0.00**	**0.01**	**0.01**	**0.00**	**0.00**	**0.00**	0.35
	Ac	0.42	0.33	0.81	0.61	0.36	0.61	0.87	0.61	0.71	0.59
PSD	Sn	**0.01**	**0.00**	0.15	0.67	**0.00**	0.31	**0.01**	0.05	0.11	0.11
	Sp	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.11	**0.01**	0.32	**0.04**	0.23
	Ac	0.52	0.83	0.85	0.82	0.75	0.73	0.78	0.76	0.76	0.80
All	Sn	0.96	0.29	0.98	0.85	0.65	0.75	0.92	0.64	0.16	0.43
	Sp	0.22	0.45	0.96	0.89	0.26	0.73	0.87	0.94	0.54	0.33
	Ac	0.11	0.54	0.98	0.85	0.70	0.48	0.65	0.37	0.08	0.31

* p-values less than 0.0001 are rounded to zero.

TABLE XII
THE ELAPSED TIME AT DIFFERENT SYSTEM STAGES

Stage	Dimension	Elapsed time (s)
Feature extraction for a 30-s epoch	3000×1	0.801
mRMR algorithm	101162×104	26.547
Training 19 subjects of the XSEDFDB	101162×104	191.250
Testing a 30-s epoch	1×104	0.422

* System Info: Manjaro Linux, Intel i5-3230M (4) @ 3.200 GHz, 6 GB RAM.
* All the experiments in this paper were performed using Matlab R2015b.

For each feature set, the few number of features are not able to compensate for not using gamma band. Hence, gamma band features have a significant effect on performance.

The following information can be extracted from the feature selection process. Feature sets PM and PSD are discarded by the Kruskal-Wallis test. Feature sets KE, RE, NSE, HM, and HC are selected by the mRMR algorithm for both the XSEDFDB and the UCDDB. LRSSV is also selected for the UCDDB and SD and PFD are selected for the XSEDFDB as well. Eight and nine features selected are pertaining to gamma band using the XSEDFDB and the UCDDB, respectively.

The proposed system is implemented using Matlab R2015b running on a laptop with an Intel i5-3230M (4) 3.200 GHz CPU, 6 GB RAM, Manjaro Linux operating system. The elapsed time at the different stages of the system is given in Table XII.

Finally, it is worth mentioning that there is a lack of agreement between sleep experts manually scoring PSG signals. This agreement is sometimes less than 90%. As it is reported in [65], 96% and 75% of PSG epochs scored have agreement of at least 3 and 4 (out of 5) experts, respectively. Therefore, this level of interobserver agreement indicates approximately an upper limit in the accuracy rate that an automatic sleep stage classification system might be able to reach. In other words, an accuracy rate much higher than the level of interobserver agreement is most likely due to overfitting.

IX. CONCLUSION

In this paper, we proposed a multi-class sleep stage classification system using the EEG signals. Each EEG epoch was first decomposed into 8 subband epochs with different frequency contents. Thirteen features from different domains were then extracted from each subband epoch. The mRMR algorithm were used to sort the statistically significant features based on minimal redundancy and maximal relevance. Classification was based on RF. The system proposed outperforms the existing systems considering the important factors such as the datasets used, the amount of data selected from the datasets (i.e., whether or not the whole datasets are used), the number and the type of the input signals, and the type of cross-validation (i.e., whether or not subject cross-validation is performed).

The proposed system has some advantages as follows. First, a single EEG channel is only needed. Second, the high performance of the proposed system implies that, it can be used to assist the experts with the aim of improving the sleep stage classification process. Third, it is not computationally demanding. It can be hence used in portable devices to perform online and real-time classification. Fourth, the system performance does not degrade much using the signals recorded from the subjects with suspected sleep-disordered breathing.

An important limitation of the proposed system is the sensitivity rate obtained in subject cross-validation for REM which is lower than the ones for other states. Improving the sensitivity rate for REM is left for future work. Due to rapid eye movements during REM, one can investigate the effect of movement-related potential features in this regard.

Although we have shown that there is useful information in gamma band for sleep stage classification, we could not prove its significance in the overall system performance. Finding other features extracted from gamma band to show its significance is also left for future work.

REFERENCES

[1] P. Tian *et al.*, "A hierarchical classification method for automatic sleep scoring using multiscale entropy features and proportion information of sleep architecture," *Biocybern. Biomed. Eng.*, vol. 37, no. 2, pp. 263–271, 2017.
[2] M. A. Carskadon and W. C. Dement, "Normal human sleep: An overview," in *Principles and Practice of Sleep Medicine*, M. Kryger, T. Roth, and W. C. Dement, Eds., 6th ed. Amsterdam, The Netherlands: Elsevier, 2017, pp. 15–24. [Online]. Avaialble: https://doi.org/10.1016/B978-0-323-24288-2.00002-7
[3] R. A. U., O. Faust, N. Kannathal, T. Chua, and S. Laxminarayan, "Non-linear analysis of EEG signals at various sleep stages," *Comput. Methods Programs Biomed.*, vol. 80, no. 1, pp. 37–45, 2005.
[4] J. V. Holland, W. C. Dement, and D. M. Raynal, "Polysomnography: A response to a need for improved communication," presented at the 14th Annu. Meeting Assoc. Psychophysiol. Study Sleep, 1974. [Online]. Available: https://www.lib.uchicago.edu/e/scrc/findingaids/view.php?eadid=ICU.SPCL.SLEEP#idp90478928
[5] K. A. I. Aboalayon and M. Faezipour, "Multi-class SVM based on sleep stage identification using EEG signal," in *Proc. IEEE Healthcare Innov. Conf. (HIC)*, Oct. 2014, pp. 181–184.
[6] U. R. Acharya *et al.*, "Nonlinear dynamics measures for automated EEG-based sleep stage detection," *Eur. Neurol.*, vol. 74, nos. 5–6, pp. 268–287, 2015.
[7] R. Sharma, R. B. Pachori, and A. Upadhyay, "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals," *Neural Comput. Appl.*, vol. 28, no. 10, pp. 2959–2978, Mar. 2017.
[8] A. R. Hassan and M. I. H. Bhuiyan, "Automated identification of sleep states from EEG signals by means of ensemble empirical mode decomposition and random under sampling boosting," *Comput. Methods Programs Biomed.*, vol. 140, pp. 201–210, Mar. 2017.
[9] A. R. Hassan and M. I. H. Bhuiyan, "An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting," *Neurocomputing*, vol. 219, pp. 76–87, Jan. 2017.
[10] K. A. I. Aboalayon, M. Faezipour, W. S. Almuhammadi, and S. Moslehpour, "Sleep stage classification using EEG signal analysis: A comprehensive survey and new investigation," *Entropy*, vol. 18, no. 9, p. 272, 2016.
[11] A. R. Hassan and M. I. H. Bhuiyan, "Automatic sleep scoring using statistical features in the EMD domain and ensemble methods," *Biocybern. Biomed. Eng.*, vol. 36, no. 1, pp. 248–255, 2016.
[12] A. R. Hassan and M. I. H. Bhuiyan, "Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating," *Biomed. Signal Process. Control*, vol. 24, pp. 1–10, Feb. 2016.
[13] T. Lajnef *et al.*, "Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines," *J. Neurosci. Methods*, vol. 250, pp. 94–105, Jul. 2015.
[14] A. R. Hassan, S. K. Bashar, and M. I. H. Bhuiyan, "On the classification of sleep states by means of statistical and spectral features from single channel electroencephalogram," in *Proc. IEEE Int. Conf. Adv. Comput., Commun. Inform. (ICACCI)*, Aug. 2015, pp. 2238–2243.
[15] G. Zhu, Y. Li, and P. P. Wen, "Analysis and classification of sleep stages based on difference visibility graphs from a single-channel EEG signal," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 6, pp. 1813–1821, Nov. 2014.

[16] J. L. Rodríguez-Sotelo, A. Osorio-Forero, A. Jiménez-Rodríguez, D. Cuesta-Frau, E. Cirugeda-Roldán, and D. Peluffo, "Automatic sleep stages classification using EEG entropy features and unsupervised pattern analysis techniques," *Entropy*, vol. 16, no. 12, pp. 6573–6589, 2014.

[17] T. H. Sanders, M. McCurry, and M. A. Clements, "Sleep stage classification with cross frequency coupling," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. (EMBC)*, Aug. 2014, pp. 4579–4582.

[18] V. Bajaj and R. B. Pachori, "Automatic classification of sleep stages based on the time-frequency image of EEG signals," *Comput. Methods Programs Biomed.*, vol. 112, no. 3, pp. 320–328, 2013.

[19] Y.-L. Hsu, Y.-T. Yang, J.-S. Wang, and C.-Y. Hsu, "Automatic sleep stage recurrent neural classifier using energy features of EEG signals," *Neurocomputing*, vol. 104, pp. 105–114, Mar. 2013.

[20] C.-S. Huang, C.-L. Lin, L.-W. Ko, S.-Y. Liu, T.-P. Sua, and C.-T. Lin, "A hierarchical classification system for sleep stage scoring via forehead EEG signals," in *Proc. IEEE Symp. Comput. Intell., Cognit. Algorithms, Mind, Brain (CCMB)*, Apr. 2013, pp. 1–5.

[21] C.-S. Huang, C.-L. Lin, W.-Y. Yang, L.-W. Ko, S.-Y. Liu, and C.-T. Lin, "Applying the fuzzy c-means based dimension reduction to improve the sleep classification system," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ)*, Jul. 2013, pp. 1–5.

[22] L. Fraiwan, K. Lweesy, N. Khasawneh, M. Fraiwan, H. Wenz, and H. Dickhaus, "Time frequency analysis for automated sleep stage identification in fullterm and preterm neonates," *J. Med. Syst.*, vol. 35, no. 4, pp. 693–702, 2011.

[23] U. R. Acharya, E. C.-P. Chua, K. C. Chua, L. C. Min, and T. Tamura, "Analysis and automatic identification of sleep stages using higher order spectra," *Int. J. Neural Syst.*, vol. 20, no. 6, pp. 509–521, 2010.

[24] Y. Li, F. Yingle, L. Gu, and T. Qinye, "Sleep stage classification based on EEG Hilbert–Huang transform," in *Proc. 4th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, May 2009, pp. 3676–3681.

[25] J. Zhang, Y. Wu, J. Bai, and F. Chen, "Automatic sleep stage classification based on sparse deep belief net and combination of multiple classifiers," *Trans. Inst. Meas. Control*, vol. 38, no. 4, pp. 435–451, 2016.

[26] O. Tsinalis, P. M. Matthews, and Y. Guo, "Automatic sleep stage scoring using time-frequency analysis and stacked sparse autoencoders," *Ann. Biomed. Eng.*, vol. 44, no. 5, pp. 1587–1597, May 2016.

[27] L. J. Herrera *et al.*, "Combination of heterogeneous EEG feature extraction methods and stacked sequential learning for sleep stage classification," *Int. J. Neural Syst.*, vol. 23, no. 3, p. 1350012, 2013.

[28] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[29] *The Sleep-EDF Database*. Accessed: Aug. 10, 2016. [Online]. Available: https://physionet.org/physiobank/database/sleep-edf/

[30] M. Ronzhina, O. Janoušek, J. Kolářová, M. Nováková, P. Honzík, and I. Provazník, "Sleep scoring using artificial neural networks," *Sleep Med. Rev.*, vol. 16, no. 3, pp. 251–263, 2012.

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

[32] *St. Vincent's University Hospital/University College Dublin Sleep Apnea Database*. Accessed: Aug. 10, 2016. [Online]. Available: https://physionet.org/pn3/ucddb/

[33] *The Sleep-EDF Database [Expanded]*. Accessed: Aug. 8, 2017. [Online]. Available: https://physionet.org/pn4/sleep-edfx/

[34] B. Hjorth, "Time domain descriptors and their relation to a particular model for generation of EEG activity," in *CEAN: Computerized EEG Analysis*, G. Dolce and H. Künkel, Eds. Stuttgart, Germany: Gustav Fischer-Verlag, 1975, pp. 3–8.

[35] A. Petrosian, "Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns," in *Proc. 8th IEEE Symp. Comput.-Based Med. Syst.*, Jun. 1995, pp. 212–217.

[36] M. J. Katz, "Fractals and the analysis of waveforms," *Comput. Biol. Med.*, vol. 18, no. 3, pp. 145–156, 1988.

[37] R. Esteller, J. Echauz, T. Tcheng, B. Litt, and B. Pless, "Line length: An efficient feature for seizure onset detection," in *Proc. 23rd Annu. Int. Conf. IEEE Eng. Med. Biol. (EMBC)*, vol. 2. Oct. 2001, pp. 1707–1710.

[38] R. Morales, T. Di Matteo, R. Gramatica, and T. Aste, "Dynamical generalized hurst exponent as a tool to monitor unstable periods in financial time series," *Phys. A, Statist. Mech. Appl.*, vol. 391, no. 11, pp. 3180–3189, 2012.

[39] T. Di Matteo, T. Aste, and M. M. Dacorogna, "Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development," *J. Banking Finance*, vol. 29, no. 4, pp. 827–851, 2005.

[40] L. Cabrera-Brito, G. Rodriguez, L. García-Weil, M. Pacheco, E. Perez, and J. J. Waniek, "Fractal analysis of deep ocean current speed time series," *J. Atmos. Ocean. Technol.*, vol. 34, no. 4, pp. 817–827, 2017.

[41] R. Hasan and M. Mohammed Salim, "Power law cross-correlations between price change and volume change of Indian stocks," *Phys. A, Statist. Mech. Appl.*, vol. 473, pp. 620–631, May 2017.

[42] T. Di Matteo, "Multi-scaling in finance," *Quantitative Finance*, vol. 7, no. 1, pp. 21–36, 2007.

[43] T. Inouye *et al.*, "Quantification of EEG irregularity by use of the entropy of the power spectrum," *Electroencephalogr. Clin. Neurophysiol.*, vol. 79, no. 3, pp. 204–210, 1991.

[44] M. Sabeti, S. Katebi, and R. Boostani, "Entropy and complexity measures for EEG signal classification of schizophrenic and control participants," *Artif. Intell. Med.*, vol. 47, no. 3, pp. 263–274, 2009.

[45] A. Renyi, "On measures of information and entropy," in *Proc. 4th Berkeley Symp. Math. Statist. Probab.*, 1960, pp. 547–561.

[46] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, p. 066138, Jun. 2004.

[47] S. Patidar and T. Panigrahi, "Detection of epileptic seizure using Kraskov entropy applied on tunable-Q wavelet transform of EEG signals," *Biomed. Signal Process. Control*, vol. 34, pp. 74–80, Apr. 2017.

[48] A. D. Poularikas, *The Handbook of Formulas and Tables for Signal Processing*, vol. 73. Boca Raton, FL, USA: CRC Press, 1998, p. 79.

[49] E. Bedrosian, "The analytic signal representation of modulated waveforms," *Proc. IRE*, vol. 50, no. 10, pp. 2071–2076, Oct. 1962.

[50] J. D. Spurrier, "On the null distribution of the Kruskal–Wallis statistic," *J. Nonparametric Statist.*, vol. 15, no. 6, pp. 685–691, 2003.

[51] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J. Bioinf. Comput. Biol.*, vol. 3, no. 2, pp. 185–205, 2005.

[52] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[53] G. Roffo, S. Melzi, and M. Cristani, "Infinite feature selection," in *Proc. IEEE Int. Conf. Comptut. Vis. (ICCV)*, Dec. 2015, pp. 4202–4210.

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

[55] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, 2006.

[56] H. T. Wu, R. Talmon, and Y. L. Lo, "Assess sleep stage by modern signal processing techniques," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1159–1168, Apr. 2015.

[57] A. Iranzo *et al.*, "Rapid-eye-movement sleep behaviour disorder as an early marker for a neurodegenerative disorder: A descriptive study," *Lancet Neurol.*, vol. 5, no. 7, pp. 572–577, 2006.

[58] C. Vural and M. Yildiz, "Determination of sleep stage separation ability of features extracted from EEG signals using principle component analysis," *J. Med. Syst.*, vol. 34, no. 1, pp. 83–89, 2010.

[59] S.-F. Liang, C.-E. Kuo, Y.-H. Hu, and Y.-S. Cheng, "A rule-based automatic sleep staging method," *J. Neurosci. Methods*, vol. 205, no. 1, pp. 169–176, 2012.

[60] B. Şen, M. Peker, A. Çavuşoğlu, and F. V. Çelebi, "A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms," *J. Med. Syst.*, vol. 38, no. 3, p. 18, 2014.

[61] J. L. Cantero, M. Atienza, J. R. Madsen, and R. Stickgold, "Gamma EEG dynamics in neocortex and hippocampus during human wakefulness and sleep," *NeuroImage*, vol. 22, no. 3, pp. 1271–1280, 2004.

[62] S. M. Montgomery, A. Sirota, and G. Buzsáki, "Theta and gamma coordination of hippocampal networks during waking and rapid eye movement sleep," *J. Neurosci.*, vol. 28, no. 26, pp. 6731–6741, Jun. 2008.

[63] U. Voss, R. Holzmann, I. Tuin, and A. J. Hobson, "Lucid dreaming: A state of consciousness with features of both waking and non-lucid dreaming," *Sleep*, vol. 32, no. 9, pp. 1191–1200, Sep. 2009.

[64] U. Voss *et al.*, "Induction of self awareness in dreams through frontal low current stimulation of gamma activity," *Nature Neurosci.*, vol. 17, no. 6, pp. 810–812, Jun. 2014.

[65] R. G. Norman, I. Pal, C. Stewart, J. A. Walsleben, and D. M. Rapoport, "Interobserver agreement among sleep scorers from different centers in a large dataset," *Sleep*, vol. 23, no. 7, pp. 901–908, Nov. 2000.