Received March 26, 2021, accepted April 21, 2022, date of publication May 16, 2022

Comparison of Automatic Sleep Stage Classification Methods for Clinical Use

By Alexei Labrada, Elsa Santos Febles, José Manuel Antelo

Cuban Neuroscience Center (CNEURO), Cuba

ABSTRACT

Sleep stage scoring is necessary for diagnosing several sleep disorders. However, it is an intensive and repetitive task and a vital automation candidate. This work seeks to evaluate different kinds of Machine Learning based classification algorithms available in the scientific literature to determine which one fits better the clinical practice requirements. The comparison is made with a predefined experimental design, using electroencephalography, electrooculography, and electromyography signals from the polysomnographic records of the Sleep-EDFx dataset. The comparison considers the accuracy and speed of algorithms based on Linear Discriminate Analysis, Support Vector Machines, Random Forests, and Artificial Neural Networks. The latter group includes the Deep Neural Networks DeapFeatureNet, based on Convolutional Neural Networks, and DeepSleepNet, additionally based on Recurrent Neural Networks. It is determined that several of the tested algorithms boast high accuracy levels (85%). From them, DeepSleepNet is chosen as the fittest due to its considerable advantage in execution time. Nevertheless, the final result should always be reviewed by the experts.

Keywords – *polysomnography, sleep stage scoring, machine learning, deep learning, signal processing.*

INTRODUCTION

Sleep stage scoring is necessary for diagnosing several sleep disorders, including insomnia, sleep apnea, narcolepsy, and hypersomnia. According to the American Academy of Sleep Medicine (AASM), this operation entails the division of a polysomnographic record (PSG) in consecutive 30-second windows, called epochs. Each epoch has to be classified as wakefulness (W), REM sleep (R), or one of three non-REM sleep stages: N1, N2, or N3. * Additionally, AASM defines the rules that have to be followed to perform the scoring based on the visual examination of each epoch of the PSG record.

A PSG record shows the behavior throughout the time of various electrophysiological signals. The three most

important signals are (1) electrical activity in the cerebral cortex, measured using electroencephalography (EEG); (2) in the face muscles, using electromyography (EMG); and (3) the eye movements, using electrooculography (EOG). It may also include the cardiac activity or electrocardiogram (ECG), the respiratory activity, and the body movements. The scoring rules rely on identifying various patterns in the signals, including the Alpha, Beta, Theta, and Delta Activity, K complexes, Spindles, REM, and SEM.** Table 1 summarizes some of these patterns.

TABLE 1. Common Patterns in Polysomnographic Signals

Pattern	Stage	Signal	Frequency	Morphology
Alpha Activity	W, N1	EEG	8 - 13 Hz	
Beta Activity	W, N1, R	EEG	14 - 30 Hz	
Theta Activity	NREM, R	EEG	4 - 8 Hz	
Delta Activity	N3, R	EEG	0.5 - 4 Hz	
Spindle	N2, N3	EEG	12 - 14 Hz	
K Complex	N2, N3	EEG	0.5 - 1.5 Hz	Biphasic high amplitude peak
Slow waves	N3	EEG	0.5 - 2 Hz	High amplitude waves

EEG = electroencephalography.

The PSG records may last for 8 hours, so the number of epochs is close to a thousand. Therefore, the scoring process is intensive, repetitive, and prone to errors. The scientific literature describes many algorithms that allow the automation of the process by using various Machine Learning techniques. However, the low inter-scorer agreement level,[1,2] among other limitations, has limited the accuracy of the algorithms and, hence, the reach of the automation process.

For instance, Fraiwan et al. [3] use the Continuous Wavelet Transform of the EEG signals as features and a Linear Discriminant Analysis (LDA) based classifier. As a result, they reach an 84% accuracy level with the MIT-BIH[4,5] dataset records. Susmakova & Krakovska[6] also use an LDA-based classifier, but their algorithm extracts a wider variety of features from different signals. Furthermore, they prove the importance of the information contained within the EOG and EMG signals to discriminate some of the stages.

Koley & Dey[7] evaluate the performance of a Support Vector Machine (SVM) based classifier with different combinations of features. Their algorithm has an 89% accuracy on their own dataset, close to the inter-scorer agreement level. Aboalayon et al.[8] also use an SVM classifier, reaching a 92.5% accuracy on records from the Sleep-EDF[5,9] dataset.

Set et al. [10] compare the performance of different classifiers, including Decision Trees (DT), Random Forests (RF), SVM, and Artificial Neural Networks (ANN). Moreover, they employ various feature extraction techniques, counting the Discrete Wavelet Transform (DWT). As a result, they determine that the RF obtains the best results, reaching a 97% accuracy with their own records. Finally, Aboalayon et al.[11] compare the DT, SVM, ANN, K-Nearest Neighbors, Naive Bayes (NB), and LDA classifiers. In their work, the DT classifier obtained the best results with a 93% accuracy on records from the Sleep-EDF dataset.

Finally, the Deep Learning techniques also have gained a foothold in sleep stage scoring. For example, Zhang et al.[12] propose using a Recurrent Neural Network (RNN) as a classifier but using conventional feature extraction methods. Their algorithm reaches 80.25% accuracy on the SHHS5 dataset records. Alternatively, Yildirim et al.[13] present a Convolution Neural Network-based algorithm that uses convolutional layers for feature extraction, with a 91% accuracy on Sleep-EDF records. Additionally, Supratak et al.[14] use a Convolutional Neural Networks (CNN) combined with an RNN, reaching an 82% accuracy on the same records.

The goal of this work is to select a sleep stage scoring algorithm to facilitate the work of the experts. Furthermore, the algorithm should be included in a software system

for the clinical analysis of polysomnographic records. Therefore, the selection should be based on the accuracy of the predictions and consider execution time and the general availability of the input data. With that in mind, the performance of several algorithms from the scientific literature will be compared using the same records and in similar conditions.

MATERIALS

The work uses PSG records from the Sleep Cassette dataset belonging to Sleep-EDFx.[5,9] The dataset has 153 subjects between 25 and 101 years old and was scored by several experts following the Rechtschaffen and Kales (R & K)[15] rules. The records include two EEG and one EOG signal, samples at 100 Hz, and one EMG signal at 1 Hz. Both EOG and EMG signals are considered in this work, but only the Fpz-Cz channel is used from the EEG signals. That way, all the implemented algorithms depend only on the minimum parameters of any PSG record.[1]

The dataset is split into two parts of approximately the same number of records. The first half contains the subjects with identifications 00 through 38 and is reserved for training the scoring algorithms. The second one, with subjects 40 through 82, is used to evaluate and compare the performance of said algorithms.

METHODS

The analyzed algorithms' execution time can be split into three main phases: Data preprocessing, feature extraction, and classification. The preprocessing and feature extraction phases are implemented in the Python and C# programming languages. For the classification, the work additionally employs the Weka software system[16,17] from the University of Waikato, New Zealand.

Preprocessing

The goal of the preprocessing phase is to prepare the data for the feature extraction phase. To achieve it, all signals are uniformly sampled at 100 Hz, and no digital filtering is applied beyond what is already included in the dataset: 0.5 to 100 Hz range for EEG and EOG and 0.7 to 16 Hz, for EMG. The records are segmented in 30-second windows that match the epochs that will be classified later. Also, the third and fourth non-REM sleep stages from R & K are combined into one Slow Wave Sleep or N3 stage[1,7] to fit better the AASM stages. Additionally, the unknown or invalid sleep stages are excluded from consideration. The wake stages before the first and posterior to the last sleep stages are also excluded from the training dataset records. The latter operation reduces the disparities in the amounts of epochs classified with each sleep stage. Besides, more importantly, for the RNN classifiers, it does not affect the continuity of a record's epochs.

Feature Extraction

The feature extraction phase obtains descriptive values that reflect the information inside the relevant signals for the classification process. The values or features used in this work are obtained by analyzing the signals in each epoch in the time domain, frequency domain, time-frequency domain, and other nonlinear means.

Descriptive Statistics

These features are obtained by computing descriptive statistics from the signal's samples. The Mean, Variance, Kurtosis, Skewness, and 75th Percentile have been employed in this work.

Entropy

Entropy is a measure of the irregularity of a signal in the time domain.[18] Equation 1 shows the formula proposed by Shannon for this measure:

where p(x_i) is the probability of a signal sample having the value x_i.

$$ShEn = -\sum_{i=1}^{N} p(x_i)\log p(x_i) \quad (1)$$

Other estimation methods, including the Approximate Entropy, are displayed in equation 2.

$$ApEn(r,m) = \phi_r^m - \phi_r^{m+1} \quad (2)$$

The values of ϕ can be obtained using an algorithm that represents the signal in the phase domain $X_i=\{x_i, x_{(i+1)}, ..., x_{(i+(m-1))}\}$ and calculates the distance between those patterns using the L1 norm. Then,

$$\phi_r^m = \frac{1}{M}\sum_{i=1}^{M} \log \frac{N_r^m(i)}{M} \quad M = N - m + 1 \qquad (3)$$

where N_r^m is the amount of X_j patterns that satisfy $\|X_i - X_j\|_1 \leq r$.

In this work, the pattern length (m) is 2 and r is the standard deviation of the signal in the epoch, multiplied by 0.1, as estimated in.[18]

Largest Lyapunov Exponent

The Largest Lyapunov Exponent (LLE) indicates how unpredictable a signal is. It has been demonstrated that it can help discriminate the N1 and N2 stages.[7] The algorithm proposed by[19] allows estimating LLE by calculating the distances between the most similar trajectories, which are also distant in the time domain. Equation 4 describes this distance,

$$d_j(0) = \min_k \| X_j - X_k \|, \quad |i - j| > \tau \qquad (4)$$

where τ is the threshold in time domain and $X_i = \{x_i, x_{(i+J)}, \ldots, x_{(i+(m-1)J)}\}$ is a trajectory in phase domain. Once the distances have been calculated, the LLE can be obtained using linear regression with equation 5.

$$y(i) = \sum_{j=1}^{M} \frac{\log d_j(i)}{T_s M} \quad M = N - (m-1)J \qquad (5)$$

In our work we use the values 10 and 7 for m and J, respectively, while τ is the mean period of the signal (MNF^{-1}).

Fractal Dimension

The fractal dimension estimates the fractional dimensions of the geometric shape of a signal in the time domain.[18] This measure is especially useful for recognizing the N3 stage.[7]

The Higuchi algorithm calculates the fractal dimension as the slope of the mean squares fit of the values of log(L(k)) against log(1/k) for k between 1 and k_{max}. The values of L(k) are calculated using the equation 6:

$$L(k) = \sum_{m=1}^{k} L_m(k) \qquad (6)$$

where $L_m(k)$ is the mean length of the sequence

$$x_m^k = (x_m, x_{m+k}, x_{m+2k}, \ldots, x_{m+N_m^k k}), \; N_m^k = \lfloor (N-m)/k \rfloor k$$

calculated with equation 7:

$$L_m(k) = \frac{(N-1)\sum_{i=1}^{N_m^k} x_{m+i\,k} - x_{m+(i-1)\,k}}{N_m^k\, k} \qquad (7)$$

In this work we use the value 40 for k_{max}, that was estimated in.[18]

Discrete Fourier Transform

The Fast Fourier Transform (FFT) algorithm efficiently estimates the frequency spectrum. The spectrum can be used to obtain the mean frequency of the signal, the spectral entropy, and the relative spectral density of the relevant frequency bands (Table 1).

The mean frequency can be calculated using equation 8:

$$MNF = \sum_{i=1}^{M} f_i \; P_i \qquad (8)$$

where M in the amount of frequency bins, f_i are the frequency values and P is the normalized spectral frequency ($\sum P_i = 1$).[20] Similarly, the spectral entropy of a frequency band can be obtained from equation 9:

$$SpEn = -\sum_{i=f_l}^{f_h} \frac{P_i \log P_i}{\log N_f} \qquad (9)$$

where f_l and f_h are the minimum and maximum frequencies, respectively and N_f is the amount of frequency bins in the range $[f_l, f_h]$.[18]

High Order Spectra

The High Order Spectra analysis can extract features related to third-order statistics of a signal.[21] Before calculating the features, the Bispectrum has to be estimated using equation 10,

$$B(f_1, f_2) = \sum_{i=1}^{W} \frac{X_i(f_1)X_i(f_2)X_i(f_1 + f_2)}{W} \quad (10)$$

where X_i is the Short-Time Fourier Transform (STFT) of the signal on the i-th window and W is the number of windows. The STFT in a vicinity of x_i is the FFT of the product of the signal and a window function centered on x_i. [22] In our work, we use 2 seconds long Haan windows, with 1 second (50%) of overlap between consecutive windows. The Bispectrum is symmetric in both axes, so its domain of interest is defined in the expression 11.

$$\Omega = \{(f_1, f_2) |\ f_1 \geq 0\ , f_1 \geq f_2\ , f_1 + f_2 \leq 0.5\} \quad (11)$$

Once the Bispectrum is calculated, it is possible to calculate its mean amplitude, the Normalized Bispectral Entropy (equation 12), its logarithmic sum (equation 13) and its mean frequency (equation 14):

$$BiEn = -\sum_{n=1}^{N} p_n \log p_n$$
$$p_n = \frac{|B(f_1, f_2)|}{\sum_{\Omega} |B(f_1, f_2)|} \quad (12)$$

$$H_1 = \sum_{\Omega} \log |B(f_1, f_2)| \quad (13)$$

$$WCOB_1 = \frac{\sum_{\Omega} f_1\ B(f_1, f_2)}{\sum_{\Omega} B(f_1, f_2)} \quad (14)$$

Wavelet Transform

The Wavelet Transforms translate a signal into the time-frequency domain. The transformation approximates the signal inside a time window by a Wavelet base (ψ) using different time scales.[22] The scale factors are inversely proportional to the frequency of the Wavelet base, as stated in equation 15,

$$\omega = \frac{f_\psi}{a\, T_s} \quad (15)$$

where T_s is the sampling period and f_ψ is the mean frequency of the Wavelet base (3).

The DWT decomposes the signal in two coefficient vectors with N/2 values, satisfying

$$a_1 = H_\psi\, x \qquad d_1 = G_\psi\, x \quad (16)$$

where H_ψ and G_ψ are dual filters with sub-sampling, related to the Wavelet base.[22] The a_1 vector contains an approximation of the original signal in the frequency range [0,1/4 f_s], while d_1 is a detail vector in the frequency range [1/4 f_s,1/2 f_s], where f_s is the sampling frequency.[10] The DWT can be computed again from vector a_1, in order to obtain the vectors a_2 and d_2 with frequency ranges [0,1/8 f_s] and [1/8 f_s,1/4 f_s], respectively. Thus, successively, the signal can be decomposed in L levels, after which the vectors d_1,d_2,...,d_L,a_L belong to different frequency bands.

The entropy of each relevant frequency band (Table 1) along the epoch in question can be calculated from the transform. We use the Daubechies function (*db1*) as the Wavelet base for the EOG signals and the reverse biorthogonal function (*rbio3.3*) for the EEG signals. Given the 100 Hz sampling frequency of the signals, once they are decomposed into 5 levels, the frequencies of the coefficient vectors approximately match the frequency bands in Table 1.

Classification

The classification phase is responsible for assigning a sleep stage to each epoch contingent on the features extracted from it. In our work, we use classifiers based on Linear Discriminate Analysis,[3] SVMs,[23] RF,[23,24] ANN, and NB.[23]

Several kinds of Neural Networks have been analyzed, including Multilayer Perceptrons (MLP),[10,25] CNN, and RNN. Specifically, we have tested the networks *DeepFeatureNet* (DFN) and *DeepSleepNet* (DSN),[14] implemented on Python using Tensorflow. The former is a CNN, while the latter is a hybrid network combining a CNN and an RNN. Both algorithms use CNN for feature extraction, so they do not require the methods described in section *Feature Extraction*.

The implementation proposed for a single signal has been expanded to process the EOG, EMG, and EEG signals.[14] This was achieved by taking advantage of the

capacity of CNN layers to process several input channels and by increasing the size of the filters proportionally to the number of channels. The DFN network has been trained with 75 epochs, while DSN has required 25 more in fine-tuning. The source code is available at https://github.com/ALabrada/deepsleepnet.

For the remaining classifiers, it has been used the implementations available in Weka, using their respective default parameters.

Evaluation

The performance of each algorithm has been analyzed, considering the accuracy (Acc) and Cohen's *kappa* coefficient. Additionally, the classification performance of the individual stages is considered using the Precision (PR) and Recall (RE) metrics.

RESULTS

The classification algorithms have been trained with the first half of the PSG records of the Sleep Cassette dataset. The set has 76 records that belong to 39 different subjects with identifiers 00 through 38. Table 2 shows the distribution of the stages assigned by the experts to the 74354 epochs that have been used from those records.

The 10-fold cross-validation technique has been used to estimate the hyper-parameters of the models and the validation error. Table 3 shows the estimated errors.

The trained classifiers have been tested using the second half of the *Sleep Cassette* dataset, and the results have been compared. The set has 77 records that belong to 39 subjects with identifiers 40 through 82. A total of 68.8% of the 208349 epochs belong to the wake stage.

TABLE 2. Sleep Stage Distribution of the Analyzed Epochs

Stage	Training		Testing (partial)		Testing (full)	
	Count	Percent	Count	Percent	Count	Percent
W	14884	20.0	33410	33.9	143265	68.8
N1	7536	10.1	14013	14.2	14013	6.7
N2	30143	40.5	33906	34.4	33906	16.3
N3	7954	10.7	5104	5.2	5104	2.4
R	13837	18.6	12062	12.2	12062	5.8
Total	74354	100.0	98495	100.0	208349	100.0

TABLE 3. Validation Error using the Training Records

Type	Acc	Kappa	PR					RE				
			W	N1	N2	N3	R	W	N1	N2	N3	R
LDA	77.29	0.6882	0.902	0.438	0.775	0.817	0.785	0.804	0.398	0.869	0.843	0.695
NB	64.79	0.5324	0.748	0.320	0.738	0.507	0.659	0.689	0.249	0.668	0.925	0.619
RF	83.09	0.7674	0.868	0.623	0.830	0.896	0.822	0.904	0.365	0.906	0.863	0.825
SVM	79.49	0.7155	0.858	0.501	0.788	0.872	0.787	0.867	0.286	0.894	0.848	0.745
MLP	80.60	0.7334	0.883	0.515	0.808	0.867	0.790	0.874	0.342	0.885	0.838	0.794
DFN	74.27	0.6630	0.969	0.287	0.883	0.652	0.822	0.789	0.692	0.721	0.912	0.658
DSN	78.10	0.7055	0.906	0.326	0.854	0.756	0.911	0.901	0.427	0.812	0.726	0.817
AVG	76.80	0.6865	0.876	0.430	0.811	0.767	0.797	0.833	0.394	0.822	0.851	0.736

Following the procedure that has been described in section *Preprocessing*, the disparity between stages can be decreased by reducing this quantity to the 33.9%. Table 4 shows a performance comparison between the algorithms using only the selected epochs, while Table 5 shows the same comparison, but with all the epochs.

Finally, Table 6 compares the execution time of the algorithms while classifying the whole test dataset. The execution time of the algorithms that use classifiers implemented in Weka is further split into the feature extraction and classification phases. The data has been collected in a personal computer with an Intel Core i5-4570 processor (CPU), 16 GB of DDR3-1600 memory (RAM), and executed in Microsoft .NET Framework.

DISCUSSION

The results show that the test error is less than the validation error when using the full records, but it is greater when using the selected subset of the epochs. This apparent discrepancy can be explained due to the previously mentioned high proportion of epochs classified with wake stages. Every one of the analyzed algorithms obtains relatively high precision and recall results classifying this stage.

In contrast, all algorithms attain poor precision and recall results that classify the N1 stage in absolute and relative terms. This behavior is consistent with other studies from the scientific literature,[24] especially those

TABLE 4. Performance Comparison of the Classifiers using the Partial Test Dataset

Type	Acc	Kappa	PR					RE				
			W	N1	N2	N3	R	W	N1	N2	N3	R
LDA	69.43	0.5776	0.911	0.385	0.664	0.465	0.723	0.759	0.279	0.840	0.752	0.563
NB	55.09	0.4109	0.841	0.329	0.594	0.241	0.555	0.604	0.231	0.582	0.954	0.515
RF	73.98	0.6335	0.858	0.504	0.692	0.637	0.737	0.853	0.183	0.887	0.756	0.652
SVM	72.93	0.6213	0.866	0.433	0.697	0.591	0.718	0.842	0.215	0.863	0.774	0.619
MLP	71.22	0.6009	0.817	0.399	0.724	0.558	0.682	0.856	0.244	0.784	0.770	0.634
DFN	67.78	0.5670	0.968	0.303	0.743	0.517	0.898	0.697	0.640	0.734	0.776	0.454
DSN	73.88	0.6308	0.864	0.347	0.772	0.812	0.959	0.894	0.419	0.744	0.553	0.675
AVG	69.19	0.5774	0.875	0.386	0.698	0.546	0.753	0.786	0.316	0.776	0.762	0.587

TABLE 5. Performance Comparison of the Classifiers using the Full Test Dataset

Type	Acc	Kappa	PR					RE				
			W	N1	N2	N3	R	W	N1	N2	N3	R
LDA	83.45	0.6804	0.981	0.340	0.644	0.429	0.655	0.913	0.279	0.840	0.752	0.563
NB	69.02	0.4682	0.966	0.222	0.504	0.197	0.380	0.766	0.231	0.582	0.954	0.515
RF	86.43	0.7263	0.966	0.466	0.666	0.622	0.711	0.947	0.183	0.887	0.756	0.652
SVM	85.73	0.7147	0.969	0.382	0.679	0.563	0.675	0.942	0.215	0.863	0.774	0.619
MLP	85.10	0.699	0.955	0.347	0.709	0.532	0.666	0.947	0.244	0.784	0.770	0.634
DFN	80.14	0.6366	0.991	0.270	0.687	0.415	0.879	0.862	0.579	0.774	0.772	0.432
DSN	85.30	0.6973	0.953	0.318	0.773	0.812	0.967	0.970	0.537	0.688	0.442	0.474
AVG	82.17	0.6604	0.969	0.335	0.666	0.510	0.705	0.908	0.324	0.774	0.746	0.556

using the Sleep-EDFx dataset.[13,14,26–28] The DFN and DSN algorithms reach around 20% higher recall measures for this stage, but its influence is mitigated by lower values in other stages. The low classification accuracy of the N1 stage can affect the result of the sleep quality analysis,[29] which makes the algorithms unsuitable for standalone usage and, thus, require the intervention of the experts.

From the first five algorithms, the ones using more conventional strategies, the RF-based classifier obtains the best results. This confirms the conclusions that were reached by previous studies.[10,30] Furthermore, SVM, MLP, and LDA also obtain satisfactory results according to both performance metrics.

From the two last algorithms based on Deep Learning, DSN reaches superior results in all metrics other than DFN. However, during validation, our implementation of DSN is 4% lower in accuracy and 6% lower in Kappa score than the one reported by Supratak et al.[14] with the same dataset, but using different hyper-parameters and half of the PSG records. Regarding the traditional algorithms, the accuracy of DSN classifying the test dataset is equivalent to the accuracy of RF within 1%.

Considering that several of the algorithms reach similar accuracy levels, their execution times are used as tie-breakers. The results in Table 6 prove that, from the analyzed algorithms, the ones based on Deep Learning require a significantly lower amount of time to identify the sleep stages of a PSG record.

TABLE 6. Comparison of the Execution Time of the Algorithms

Type	Time			
	Extraction	Classification	Total	Average per record
LDA	54758.69 s	6.08 s	54764.77 s	711.23 s
NB	54758.69 s	5.07 s	54763.76 s	711.22 s
RF	54758.69 s	10.97 s	54769.66 s	711.29 s
SVM	54758.69 s	1.41 s	54760.10 s	711.17 s
MLP	54758.69 s	2.64 s	54761.33 s	711.19 s
DFN	-	-	766.55 s	9.66 s
DSN	-	-	1695.68 s	22.02 s

CONCLUSIONS

As part of our work, we have compared the performance of a wide range of sleep stage scoring algorithms available in the scientific literature to find the one that better matches clinical use requirements. With that in mind, accuracy and speed are used as the selection criteria for the comparison. The results prove that the RF, SVM, MLP, and DSN algorithms reach the greater accuracy levels while classifying, exceeding 85% in this metric and 0.69 in Cohen's *kappa*. Moreover, from them, DSN is significantly faster, requiring less than 30 seconds to score a record's epochs on average. The combination of both criteria determines that DSN is the most appropriate sleep stage scoring algorithms for the context of the clinical practice, from the set of candidates taken into consideration. Nevertheless, the algorithms are much less accurate in classifying the N1 stage, so the experts should review the sleep stage scoring performed by DSN.

REFERENCES

1. Malhotra RK, Avidan AY. Atlas of sleep medicine. In: 2nd ed. Elsevier; 2014. pp. 77–99.
2. Daker-Hopfe H, Anderer P, Zeitlhofer J, Boeck M, Dorn H, Gruber G, et al. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. J Sleep Res. 2009;78–84. https://doi.org/10.1111/j.1365-2869.2008.00700.x
3. Fraiwan L, Lweesy K, Khasawneh N, Fraiwan M, Wenz H, Dickhaus H. Classification of sleep stages using multi-wavelet time frequency entropy and LDA. Methods of information in medicine. 2010 Jan;49:230–7. https://doi.org/10.3414/ME09-01-0054
4. Ichimaru Y, Moody GB. Development of the polysomnographic database on CD-ROM. Psychiatry and Clinical Neurosciences. 1999;53:175–7. https://doi.org/10.1046/j.1440-1819.1999.00527.x
5. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet : Components of a new research resource for complex physiologic signals. Circulation. 2000 Jun 13;101(23). https://doi.org/10.1161/01.CIR.101.23.e215

6. Susmakova K, Krakovska A. Discrimination ability of individual measures used in sleep stages classification. Artificial Intelligence in Medicine. 2008 Nov;44(3):261–77. https://doi.org/10.1016/j.artmed.2008.07.005
7. Koley B, Dey D. An ensemble system for automatic sleep stage classification using single channel EEG signal. Computers in Biology and Medicine. 2012;42(12):1186–95. https://doi.org/10.1016/j.compbiomed.2012.09.012
8. Aboalayon K, Ocbagabir H, Faezipour M. Efficient sleep stage classification based on EEG signals. In: Applications and technology conference. 2014. https://doi.org/10.1109/LISAT.2014.6845193
9. Kemp B, Zwinderman AH, Tuk B, Kamphuisen HAC, Oberye JJL. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. IEEE Transactions on Biomedical Engineering. 2000;47(9):1185–94. https://doi.org/10.1109/10.867928
10. Sen B, Peker M, Cavusoglu A, Celebi FV. A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. J Med Syst. 2014; https://doi.org/10.1007/s10916-014-0018-0
11. Aboalayon K, Faezipour M, Almuhammadi W, Moslehpour S. Sleep stage classification using EEG signal analysis: A comprehensive survey and new investigation. Entropy. 2016 Aug;18. https://doi.org/10.3390/e18090272
12. Zhang Y, Yang Z, Lan K, Liu X, Zhang Z, Li P, et al. Sleep stage classification using bidirectional LSTM in wearable multi-sensor systems [Internet]. 2019. Available from: https://arxiv.org/abs/1909.11141
13. Yildirim O, Baloglu UB, Acharya UR. A deep learning model for automated sleep stages classification using PSG signals. Int J Environ Res Public Helth. 2019; https://doi.org/10.3390/ijerph16040599
14. Supratak A, Dong H, Wu C, Guo Y. DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2017; https://doi.org/10.1109/TNSRE.2017.2721116
15. Rechstaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. In: A manual of standardized terminology. Los Angeles, California: University of California; 1986.
16. Frank E, Hall MA, Holmes G, Kirkby R, Pfahringer B, Witten IH. Weka: A machine learning workbench for data mining. In: Data mining and knowledge discovery handbook: A complete guide for practitioners and researchers [Internet]. Berlin: Springer; 2005. pp. 1305–14. Available from: http://researchcommons.waikato.ac.nz/handle/10289/1497
17. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: An update. SIGKDD Explorations. 2009;11(1):10–8. https://doi.org/10.1145/1656274.1656278
18. Sabeti M, Katebi S, Boostani R. Entropy and complexity measures for EEG signal classification of schizophrenic and control participants. Artificial Intelligence in Medicine. 2009 Nov;47(3):263–74. https://doi.org/10.1016/j.artmed.2009.03.003
19. Rosenstein MT, Collins JJ, Luca CJD. A practical method for calculating largest Lyapunov exponents from small data sets. Physica D: Nonlinear Phenomena. 1993 Jun;65(1-2):117–34. https://doi.org/10.1016/0167-2789(93)90009-P
20. Phinyomark A, Thongpanja S, Hu H, Phukpattaranont P, Limsakul C. The usefullness of mean and median frequencies in electromiography analysis, computational intelligence in electromiography analysis - a perspective on current applications and future challenges. In IntechOpen; 2012. https://doi.org/10.5772/50639
21. Acharya R, Chua EC-P, Chua KC, Min LC, Tamura T. Analysis and automatic identification of sleep stages using higher order spectra. International Journal of Neural Systems. 2010 Nov;20(6):509–21. https://doi.org/10.1142/S0129065710002589
22. Stark H-G. Wavelets and signal processing: An application-based introduction. Springer; 2005. https://doi.org/10.1007/3-540-27481-2
23. Tzimourta KD, Tsilimbaris A, Tzioukalia K, Tzallas AT, Tsipouras MG, Astrakas LG, et al. EEG-based automatic sleep stage classification. Biomedical Journal of Scientific & Technical Research. 2018;7(4):6032–7. https://doi.org/10.26717/BJSTR.2018.07.001535

24. Fraiwan L, Lweesy K, Khasawneh N, Wenz H, Dickhaus H. Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier. Computer Methods and Programs in Biomedicine. 2012;108:10–9. https://doi.org/10.1016/j.cmpb.2011.11.005

25. Ronzhina M, Janousek O, Kolarova J, Novakova M, Honzik P, Provaznik I. Sleep scoring using artificial neural networks. Sleep Medicine Reviews. 2012;16(3):251–63. https://doi.org/10.1016/j.smrv.2011.06.003

26. Dong H, Supratak A, Pan W, Wu C, Matthews PM, Guo Y. Mixed neural network approach for temporal sleep stage classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2018; https://doi.org/10.1109/TNSRE.2017.2733220

27. Koushik A, Amores J, Maes P. Real-time sleep staging using deep learning on a smartphone for a wearable EEG. CoRR [Internet]. 2018;abs/1811.10111. Available from: http://arxiv.org/abs/1811.10111

28. Mousavi S, Afghah F, Acharya UR. SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach. Pławiak P, editor. PLOS ONE [Internet]. 2019 May;14(5). Available from: http://dx.doi.org/10.1371/journal.pone.0216456

29. Mendonça F, Mostafa SS, Morgado-Dias F, Ravelo-García AG, Penzel T. A review of approaches for sleep quality analysis. IEEE Access. 2019;7:24527–46. https://doi.org/10.1109/ACCESS.2019.2900345

30. Boostani R, Karimzadeh F, Torabi-Nami M. A comparative review on sleep stage classification methods in patients and healthy individuals. HAL. 2017 Mar;140:77–91. https://doi.org/10.1016/j.cmpb.2016.12.004