Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automatic Sleep Stage Classification Based on Two-channel EOG and One-channel EMG

Yanjun LI (yanjunli211@yahoo.cn)
China Astronaut Research and Training Center https://orcid.org/0000-0001-9732-5175
Xianglin Yang
China Astronaut Research and Training Center
Zhi Xu
China Astronaut Research and Training Center
Yu Zhang
China Astronaut Research and Training Center
Zhongping Cao
China Astronaut Research and Training Center

Research Article

Keywords: automatic sleep stage classification, sleep scoring, EEG-free sleep monitoring, EOG, EMG, sleep quality

Posted Date: June 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-491468/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

The sleep monitoring with PSG severely degrades the sleep quality. In order to simplify the hygienic processing and reduce the load of sleep monitoring, an approach to automatic sleep stage classification without electroencephalogram (EEG) was explored. Totally 108 features from two-channel electrooculogram (EOG) and 6 features from one-channel electromyogram (EMG) were extracted. After feature normalization, the random forest (RF) was used to classify five stages, including wakefulness, REM sleep, N1 sleep, N2 sleep and N3 sleep. Using 114 normalized features from the combination of EOG (108 features) and EMG (6 features), the Cohen's kappa coefficient was 0.749 and the accuracy was 80.8% by leave-one -out cross-validation (LOOCV) for 124 records from ISRUC-Sleep. As a reference for AASM standard, the Cohen's kappa coefficient was 0.801 and the accuracy was 84.7% for the same dataset based on 438 normalized features from the combination of EEG (324 features), EOG (108 features) and EMG (6 features). In conclusion, the approach by EOG+EMG with the normalization can reduce the load of sleep monitoring, and achieves comparable performances with the "gold standard" EEG+EOG+EMG on sleep classification.

Introduction

Sleep is the process of the body's self-repairing and self-recovery. A good quality sleep will eliminate fatigue, restore physical strength and energy. Generally the quality of sleep is assessed quantitatively from the macro-sleep structure, sleep breathing, and body movement during sleep. Polysomnography (PSG) is used as the basis in the clinic for the evaluation of macro-sleep structure. Following American Academy of Sleep Medicine (AASM) standard, sleep stages are divided into wakefulness, rapid eye movement (REM) sleep, and Non-REM (NREM) sleep. Furthermore, NREM sleep is divided into N1, N2 and N3 stage. Deep sleep (N3 stage) is mainly conducive to physical recovery. REM sleep is mainly conducive to the backtracking, encoding, consolidation, trimming and even strengthening of the knowledge and the memory. Breathing and blood oxygen are another dimension to sleep quality assessment. Sleep apnea includes central apnea, obstructive apnea and mixed apnea. Frequent apnea can cause reduction in Oxygenation and PH value in blood. Besides, frequent movement during sleep is a characteristic of some types of sleep disorders.

Sleep Stage Classification (SSC) from PSG provides sleep stage information for studying sleep patterns. Nowadays, there are two main research areas of Automatic Sleep Stage Classification (ASSC). Correctly identifying sleep stages are important in diagnosing and treating sleep disorders. Hence, researchers have tried to obtain higher accuracy with respect to manual SSC. As the most accurate way, PSG abides the decisive approach in many cases [1]. A joint classification-and-prediction framework based on convolutional neural networks (CNNs) yielded an accuracy of 82.3% on Sleep-EDF Expanded (Sleep-EDF), and accuracy of 83.6% on Montreal Archive of Sleep Studies (MASS) dataset [2]. Yan et al [3] developed a versatile deep-learning architecture to automatic sleep scoring when using raw PSG recordings, and got an accuracy of 86% and a kappa coefficient of 0.82 on ISRUC sleep. A Hybrid Stacked LSTM Neural Network got an accuracy of 83.1% and a kappa coefficient of 0.78 on 994 subjects when half subjects were randomly assigned to the training set, and the other half as the testing set [4].

Specifically, the inter-scorer agreement following the AASM standard [5] is only approximately 82.6% [6]. Too high accuracy from ASSC may be caused by over-fitting. It is reasonable to get a trade-off between boosting the classification performance by integrating as much features as possible, and not over-fitting sleep scoring model in certain sleep stage type. On the other hand, studies emphasis on reducing the burden from data recording during the whole night sleep, such as wearable, on-bed, and actigraphy devices [7]. Researchers have tried to use feasible devices for sleep monitoring in the community, such as ear-EEG [8], wireless sensor on the neck for tracheal breathing sounds [9], actigraphy-based devices [7] and pressure sensor mattress [1].

Although PSG is currently the "gold standard" for sleep monitoring, it requires attaching at least 10 electrodes to the head, the face and the body, which seriously interferes the subjects' natural sleep. The sleep monitoring with PSG severely degrades the sleep quality. As a result, PSG is mainly used in hospitals to monitor patients with severe sleep disorders because of its complex operation and high level of discomfort. After sleep experiments, it is not easy to clean up the residual conductive paste within the hair for EEG collecting.

Six-channel electroencephalogram (EEG), two-channel electrooculogram (EOG) and one-channel electromyogram (EMG) are recommended for sleep scoring according to AASM standard. If the EEG is not collected during sleep monitoring, the problem of injecting conductive paste into the electrode cap before EEG collection and washing hair after EEG collection can be avoided. As a result, the hygienic treatment is simplified. Besides, signal acquisition without EEG, also reduces the load of sleep monitoring on subjects.

Therefore, the study of EEG-free sleep monitoring method is of great significance to degrade the load of sleep monitoring. EOG and EMG signals can be acquired in a more comfortable way in comparison with EEG. Can comparable performances on sleep classification be achieved by EOG + EMG when in comparison with that of EEG + EOG + EMG? Focusing on this problem, this paper studies a method for low-load sleep monitoring based on EOG and EMG, and evaluates the role of EOG and EMG signals in sleep staging.

Method

1.1 Data acquisition

A public dataset called ISRUC-Sleep [10] with AASM standard was used, including the sleep disorder groups (two subsets, namely subgroup-I and subgroup-II) and the health group (subgroup-III). The database was provided by the Sleep Medicine Centre of Coimbra. It can be downloaded freely from

the web site “http://sleeptight.isr.uc.pt/ISRUC_Sleep/”. The data provide 126 PSG records and sleep stages labels from two experts. There are 19 channels of physiological data for most PSG records. However, record ‘8’ and record ‘40’were excluded from subgroup-I for the analysis. The former record does not provide EEG channels of F3 and F4, while the latter one suffers some electrode problems. As a result, totally 124 records were used in this research, including 98 records from subgroup-I, 16 records from subgroup-II and 10 records from subgroup-III.

Only six-channel electroencephalogram (EEG), two-channel electrooculogram (EOG) and one-channel electromyogram (EMG) were used in this paper. The sampling frequency is 200 Hz for each channel. All these channels in ISRUC-Sleep had been filtered to eliminate noise and undesired background by the dataset itself, aiming to enhance the PSG signal quality and increase the SNR. The filtering stage comprised: (1) a notch filter to eliminate the 50 Hz electrical noise; (2) a band-pass Butterworth filter with a lower cutoff frequency of 0.3 Hz and a higher cutoff frequency of 35 Hz for EEG and EOG channels, and a lower cutoff frequency of 10 Hz and higher cutoff frequency of 70 Hz for EMG channels.

According to ASSM rules, the sleep stages of each subject in the dataset were labeled by two experts individually. Therefore, small differences existed in annotations between two experts. If sleep scores from only one expert were used, a bias would produce from a rater’s style. As a result, only 30-s sequences with consist annotations from the two sleep diagrams were extracted for analysis in this paper.

Table 1
Records used in this paper from ISRUC-Sleep

group	Type of participants	Number of records	Number of subjects (gender)	age
subgroup-I	participants with sleep disorder	98*	98 subjects (54 male, 44 female)	20–85, Avg. = 50.7, std. = 15.9 years
subgroup-II	participants with sleep disorder	16	8 subjects (6 male, 2 female)	26–79, Avg. = 46.9, std. = 18.7 years
subgroup-III	Healthy participants	10	10 subjects (9 male,1 female)	30–58, Avg. = 39.6, std. = 10.1 years
*note: record ‘8’ and record ‘40’ were excluded in the analysis. The former record does not provide EEG channels of F3 and F4, while the latter one has some electrode problems.				

1.2 Feature extraction

1.2.1 Features from single-channel EOG

Two EOG channels are unipolar, namely'LOC-A2' and 'ROC-A1'. FFT is applied to each EOG channel to get the power spectral density (PSD). The sum of energy in sub-bands delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz) and beta (13–30 Hz) in each 30-s period is defined as E_{delta}, E_{theta}, E_{alpha} and E_{beta}, while the sum of E_{delta}, E_{theta}, E_{alpha} and E_{beta} is defined as E_{sum}. The entropy derived from E_{delta}, E_{theta}, E_{alpha} and E_{beta} is defined as $E_{Entropy}$. Similarly, the sum of the absolute value in these sub-bands in each 30-s period is defined as S_{delta}, S_{theta}, S_{alpha} and S_{beta}, while the sum of them is defined as S_{sum}. The entropy derived from S_{delta}, S_{theta}, S_{alpha} and S_{beta} is defined as $S_{Entropy}$. Feature vector within four sub-bands is defined as

$$EogBand4\_Ft16= [E_{Entropy}\ E_{beta}/E_{delta}\ E_{delta}/E_{sum}\ E_{theta}/E_{sum}\ E_{alpha}/E_{sum}\ E_{beta}/E_{sum}\ S_{Entropy}\ S_{beta}/S_{delta}\ S_{delta}/S_{sum}\ S_{theta}/S_{sum}\ S_{alpha}/S_{sum}\ S_{beta}/S_{sum}\ S_{delta}\ S_{theta}\ S_{alpha}\ S_{beta}] \quad (1)$$

In the same way, for eleven sub-bands (0.4-4) Hz, (4–8) Hz, (8–10) Hz, (10–13) Hz, (13–18) Hz, (18–25) Hz, (25–30) Hz, (30–36) Hz, (36–41) Hz, (41–46) Hz and (46–50) Hz [11], there are 11 ratios of 2-Norm within each band to the sum of them, 11 ratios of 1-Norm within each band to the sum of them, and the energy themselves. Consequently, the feature vector with 33 features within eleven sub-bands is defined as EogBand11_Ft33. The number of features for single-channel EOG is 49. The feature vector is as follows,

$$OneLeadEog=[\ EogBand4\_Ft16\ EogBand11\_Ft33] \quad (2)$$

while OneLeadEog represents LOC_LeadEog for lead 'LOC-A2' and ROC_LeadEog for lead 'ROC-A1'.

1.2.2 Correlation features between two-channel EOG

Temporal signals within sub-bands delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz) and beta (13–30 Hz) are derived from individual FIR band-pass filter from original EEG within the frequency band 1–4 Hz, 4–8 Hz, 8–13 Hz and 13–30 Hz, respectively. The correlation coefficients [12] between two-channel EOG in four sub-bands during each 30-s period are defined as r_{delta}, r_{theta}, r_{alpha} and r_{beta}, respectively. The correlation coefficient between two-

channel EOG with the original waveform during each 30-s period is defined as r_{org}. In the same way, phase-locking value (PLV) is obtained, including PLV_{beta}, PLV_{alpha}, PLV_{theta}, PLV_{delt} and PLV_{org}.

The number of features between two-channel EOG is 10, and the feature vector is as follows,

$$EogBtwn= [r_{beta}\ r_{alpha}\ r_{theta}\ r_{delt}\ r_{org}\ PLV_{beta}\ PLV_{alpha}\ PLV_{theta}\ PLV_{delt}\ PLV_{org}\ ] \quad (3)$$

The number of features for one-channel EOG is 49, and the number of features between two-channel EOG is 10. Therefore, the total number of features for two-channel EOG 'LOC-A2'and'ROC-A1' is 49×2 + 10 = 108, and the whole vector of EOG is defined as,

EogFeat=[LOC_LeadEog ROC_LeadEog EogBtwn] (4)

1.2.3 Features from single-channel EMG

The fractal dimension of EMG is defined as EmgFD, and the root mean square is defined as EmgStd in every 30 seconds.

The EMG signals of every 30-s period are transformed by Hilbert to obtain the enveloping signal. After that, the enveloping mean is defined as EnvlpMean, the enveloping maximum is defined as EnvlpMax, the enveloping root mean square is defined as EnvlpStd, and the ratio of EnvlpMax to EnvlpMean is defined as RtMaxdMean. The total number of features for single-channel EMG is 6, and the whole vector of EMG is defined as,

EmgFeat=[EmgFD EmgStd RtMaxdMean EnvlpMean EnvlpMax EnvlpStd] (5)

1.2.4 Features from six-channel EEG

This method is compared with AASM standard in this paper. Therefore, EEG features are calculated. Six-channel EEG can be divided into three groups, including {F3, F4}, {C3, C4} and {O1, O2}. For each group, a total number of 108 features can be obtained in the same way as formula (4), which is defined as F34Feat, C34Feat and O12Feat, respectively. Consequently, there are 108×3 = 324 features for all six-channel EEG.

1.2.5 Whole feature vector

For classification from EOG + EMG, the whole feature vector with 114 features from two-channel EOG (108 features) and one single-channel EMG (6 features) is defined as follows,

Feat1=[EogFeat EmgFeat] (6)

For comparison, the whole feature vector with 438 features from two-channel EOG (108 features), one single-channel EMG (6 features) and six-channel EEG (324 features) is defined as follows,

Feat2=[EogFeat EmgFeat F34Feat C34Feat O12Feat] (7)

1.3 Characteristic normalization

Physiological signals often have significant individual characteristics. For example, although the lowest EMG amplitudes in most subjects occurred during deep or REM sleep, a few subjects tended to be different, and they have the highest EMG amplitudes during wakefulness.

One normal sleep in adults may last 8 hours. During such a long period, the recording conditions variation such as skin humidity, body temperature, body movements or even worse as electrode contact loss. Besides, the discriminant information for the considered sleep stage classification lies in relative amplitudes rather than the absolute amplitudes.

If the maximum and the minimum values in the feature sequence are taken as the reference for feature normalization, it may cause an error; because both the maximum and the minimum values may be noise points. For example, most values in a feature sequence are near 1, but one noise point is 100 and the other noise point is -10. If the normalized scale is according to the maximum and the minimum values, i.e., 100 + 10 = 110, then most of the values in the normalized feature series are clustered around 0.01. Only the former noise point is 1 and the latter noise point is 0, which is obviously not the expected result of normalization.

A new ‘quasi-normalization’ method is designed in this paper. First, the original feature sequences {a(n)} are arranged in order from small to large, which is defined as {f(n)}. Set the series number of {f(n)} at the position of 10% length from the beginning as n1, the series number of {f(n)} at the position of 50% length from the beginning as n2, and the series number of {f(n)} at the position of 90% length from the beginning as n3.

The standard deviation of the sequences {f(n1:n3)} is defined as Sd.

$$Sd = std(f(n1:n3)) \quad (8)$$

$$Ku = f(n3) - f(n2) \quad (9)$$

Kd = f(n2)-f(n1) (10)

s = 2*min([Sd Ku Kd]) (11)

b(n)=(a(n)- f(n2))/s (12)

Then using formula (12) for 'quasi-normalization', most elements in {b(n)} are transformed into the interval [-1, 1], but a few elements are out of that range. In order to make all elements locate into the interval [-2, 2], the following transform is applied,

$$c(n) = \begin{cases} 2, when\ \ b(n) > 2 \\ b(n), when -1 \le b(n) \le 1 \\ -2, when\ \ b(n) < -2 \end{cases} \quad (13)$$

Finally, the feature sequences {c(n)} are used for classification.

Figure 1 is an example of the quasi-normalization for EmgStd of EMG. For original index EmgStd as Fig. 1b, most elements are lower than 2, but none is lower than 0. After using formula (12) for 'quasi-normalization', most elements in {b(n)} are transformed into the interval [-1, 1], as Fig. 1c, but some elements are still higher than 2. After using formula (13) for data truncation, elements that higher than 2 are reset as 2.

Figure 2 quasi-normalization for PLV_{org} of EOG (a) Manual scoring from the first expert as blue line, and manual scoring from the second expert as red line; (b) original index PLV_{org} of EOG, different stages with different colors; (c) sequences {b(n)}; (d) sequences {c(n)}

1.4 Classification model selection

Random Forest (RF) [13] has some wonderful advantages, including strong generalization ability, strong anti-over-fitting ability, rapid model training, simple structure and easy constructing, which is suitable for processing high-dimensional data sets without feature selection.

1.5 Comparison of classification results

Leave-one-record-out (LOOCV) strategy was applied to the mixed group (10 healthy recordings and 114 sleep disorder recordings). The training dataset contained 123 records while the rest one record was used as the validation set. This step repeated 124 times until each record had been tested. The whole 124 times' testing formed the final results.

Furthermore, the results were compared that derived from each signal type among EEG, EOG and EMG. Evaluation indices are employed, including accuracy, the multi-class weighted F1 score [14] and Cohen's kappa coefficient.

Results

2.1 Classification results for individual record

Sleep stages classification results on subgroup-III from the combination of EOG (108 features) and EMG (6 features) are shown in Table 2. According to Cohen's kappa coefficients, results from 8 out of 10 records are in substantial agreement, with kappa coefficients in the range [0.6, 0.8]. Record No.1 is almost perfect agreement (0.801), but record No.10 is only in moderate agreement (0.542).

Classification results of No.6 from subgroup-III are shown in Fig. 3. Classification results are similar with manual scoring, as N3 F-score (0.930), N2 F-score (0.862), Awake F-score and REM F-score (0.772). Sleep quality (percentage of each stage) is also similar with manual scoring, as shown in the row of No.6 in Table 3.

Classification results of No.3 from subgroup-III are shown in Fig. 4. Classification results are to some extent similar with manual scoring, as REM F-score (0.822) and N3 F-score (0.778). Percentage of REM stage is similar with manual scoring, as shown in the row of No.3 in Table 3. However, percentage of N3 stage (22.0%) derived from the classification is significantly lower than the manual scoring (38.5% and 35.0%). Visual evidence is very obvious in Fig. 4, as the second N3 stage in the second expert's manual scoring is mistaken for N2 stage, and almost half quantity of the last two N3 stage in the second expert's manual scoring is also mistaken for N2 stage. Besides, the percentage of wakefulness stage (17.7%) is significantly higher than the manual scoring (10.8% and 10.0%).

Worst-case result from Table 4 is No.10, which is shown in Fig. 5. Classification results are to some extent similar with manual scoring, as N3 F-score (0.867) and wakefulness F-score (0.808). Percentage of N3 stage (13.9%) is as similar as the manual scoring (14.1%), as shown in the row of No.10 in Table 3. However, percentage of N1 stage (6.9%) derived from the classification is significantly lower than the manual scoring (27.8%). Besides, the percentage of N2 stage (42.7%) is significantly higher than the manual scoring (22.7%).

Table 2
Classification results from EMG + EOG on subgroup-III

Number of Record	Acc (%)	Kappa	Balanced F-score	Awake F-score	REM F-score	N1 F-score	N2 F-score	N3 F-score
1	85.2	0.801	0.826	0.875	0.901	0.625	0.871	0.858
2	80.8	0.745	0.748	0.771	0.884	0.404	0.822	0.861
3	70.9	0.613	0.686	0.695	0.822	0.482	0.654	0.778
4	78.8	0.724	0.734	0.898	0.743	0.368	0.767	0.895
5	81.2	0.752	0.751	0.797	0.792	0.327	0.901	0.940
6	83.1	0.769	0.769	0.811	0.772	0.469	0.862	0.930
7	78.1	0.708	0.664	0.946	0.810	0.109	0.649	0.806
8	78.4	0.715	0.729	0.912	0.896	0.358	0.617	0.862
9	77.4	0.697	0.735	0.824	0.682	0.484	0.813	0.871
10	63.4	0.542	0.643	0.808	0.627	0.339	0.575	0.867

Table 3
Percentage of each stage from EMG + EOG on subgroup-III

	Percentage of each stage based on visual scoring from Expert-1					Percentage of each stage based on visual scoring from Expert-2					Percentage of each stage based on the proposed method				
Subject	W	N1	N2	N3	REM	W	N1	N2	N3	REM	W	N1	N2	N3	REM
1	17.3	12.5	39.1	18.7	12.5	15.4	11.5	41.1	19.9	12.1	19.1	8.5	40.0	18.7	13.7
2	12.2	15.3	34.3	20.9	17.2	9.4	11.3	34.2	25.8	19.3	17.7	6.3	42.8	16.7	16.5
3	10.8	7.9	30.3	38.5	12.5	10.0	8.9	31.9	35.0	14.3	17.7	11.4	35.9	22.0	13.0
4	21.9	17.3	29.5	20.0	11.3	19.8	16.9	27.5	19.9	16.0	21.4	6.4	41.3	21.4	9.4
5	32.0	7.6	31.1	20.7	8.6	30.1	11.3	32.9	15.4	10.3	22.1	14.7	34.3	21.1	7.7
6	9.1	15.9	34.7	29.0	11.3	8.4	8.1	39.7	27.7	16.1	11.4	13.6	35.5	29.7	9.8
7	27.3	8.4	19.8	32.2	12.4	24.6	2.1	31.7	29.0	12.7	28.6	11.7	30.8	19.7	9.2
8	37.6	11.9	20.1	14.3	16.1	37.6	11.9	20.1	14.3	16.1	33.5	2.8	30.1	19.6	14.0
9	15.4	17.2	37.7	23.2	6.5	14.9	17.3	37.9	23.4	6.5	18.2	8.5	39.0	23.3	11.0
10	18.5	27.8	22.7	14.1	17.0	18.5	27.8	22.7	14.1	17.0	25.8	6.9	42.7	13.9	10.7

2.2 Classification results for all records

Sleep stages classification results from the combination of EOG (108 features) and EMG (6 features) are shown in Table 4, and results from the combination of EEG (324 features), EOG (108 features) and EMG (6 features) are shown in Table 5. More detail is shown in Table 6. Using 114 normalized features from the combination of EOG and EMG, the Cohen's kappa coefficient was 0.749 and the accuracy was 80.8% by LOOCV for 124 records from ISRUC-Sleep. As a reference for AASM standard, the Cohen's kappa coefficient was 0.801 and the accuracy was 84.7% for the same dataset based on 438 normalized features from the combination of EEG, EOG and EMG.

Table 4
Classification results from EMG + EOG (Number of 30-s segments)

		Classification results					
		awake	REM	N1	N2	N3	all
Reference	awake	20376	464	671	902	338	22751
	REM	705	11917	418	538	164	13742
	N1	1668	593	3191	1933	141	7526
	N2	1778	997	1397	22463	2611	29246
	N3	293	163	7	1787	15755	18005
	all	24820	14134	5684	27623	19009	

Table 5
Classification results from EEG + EMG + EOG (Number of 30-s segments)

		Classification results					
		awake	REM	N1	N2	N3	all
Reference	awake	20991	468	614	624	54	22751
	REM	625	12292	362	422	41	13742
	N1	1403	645	3520	1882	76	7526
	N2	1029	652	1300	24157	2108	29246
	N3	170	0	2	1447	16386	18005
	all	24218	14057	5798	28532	18665	/

When classification results were derived from only one kind of signal from EEG, EOG and EMG, Cohen's kappa coefficients from high to low were in the order that C-EEG > F-EEG > O-EEG > EOG > EMG, as shown in Table 6. Rahman et al [16] analyzed single channel EOG in Discrete Wavelet Transform (DWT) domain employing various statistical features, and got an accuracy of 86.0% with RF for ISRUC-Sleep data. The main shortcoming of their study is that only 10 records were used, as 5 records for training and the other 5 records for testing.

In multi-class sleep staging, the best discrimination was achieved by the combination of EEG + EMG + EOG, in which the highest F1-score was 0.894 for both awake and N3 stages in Table 6, followed by REM (0.884) and N2 (0.836) stages. However, the lowest F1-score resided in the detection of stage N1 (0.528). The order in F1-score for detection of single stage was consist with the order in the accuracy of the research of Khalighi et al [14], with awake (88.59%) > N3 (87.13%) > REM (86.99%) > N2 (79.06%) > N1 (66.91%).

Most studies in Table 6 did not use the whole dataset of ISRUC-Sleep for validation. It is not a fair comparison when one study uses more than 100 records for validation and another one only use 5 records for validation. Using 99 records for validation from 6EEG + 2EOG + 1EMG + 1ECG, deep learning [3] got an accuracy of 86% and Cohen's kappa coefficient of 0.82. In comparison with that, our proposed method used 124 records for validation, and got an accuracy of 84.7% and Cohen's kappa coefficient of 0.807 from 6EEG + 2EOG + 1EMG. The main difference is that the F-score is 0.528 from our method with 6EEG + 2EOG + 1EMG, but the deep learning [3] got an F-score of 0.67 for N1 stage.

Furthermore, for training and testing set as Table 7, our method obtains comparable performance with the method of Md Mosheyur Rahman [16].

Table 6
Comparison of the performance by LOOCV on the dataset ISRUC-Sleep

method	Number of record	Signal types	model	Number of features	Acc (%)	Kappa	Balanced F-score	Awake F-score	REM F-score	N1 F-score	N2 F-score	N3 F-score
proposed method	124	6EEG + 2EOG + 1EMG	RF	438	84.7	0.801	0.807	0.894	0.884	0.528	0.836	0.894
		6EEG	RF	324	83.5	0.784	0.790	0.886	0.848	0.496	0.824	0.894
		C-EEG	RF	108	82.1	0.766	0.774	0.871	0.837	0.469	0.808	0.884
		F-EEG	RF	108	82.0	0.764	0.771	0.867	0.843	0.458	0.807	0.881
		O-EEG	RF	108	81.2	0.755	0.764	0.879	0.804	0.457	0.797	0.882
		2EOG + EMG	RF	114	80.8	0.749	0.767	0.857	0.855	0.483	0.790	0.851
		2EOG	RF	108	80.3	0.742	0.759	0.854	0.835	0.465	0.791	0.848
		EMG	RF	6	50.1	0.359	0.469	0.622	0.627	0.272	0.364	0.462
[15]	10	C3–A2 EEG	state space model	NP	81.7	0.763	NP	0.903	0.833	0.577	0.811	0.875
[14]	40	6EEG + 2EOG + 1EMG	support vector machine (SVM)	326	84.7	NP	0.747	NP	NP	NP	NP	NP
[17]	10	C3-A2 EEG	combining locality energy (LE) and dual state space models (DSSMs)	NP	81.7	0.763	NP	NP	NP	NP	NP	NP
[14]	40	6EEG	SVM	200	80.2	NP	0.671	NP	NP	NP	NP	NP
[11]	10	C3-A2 EEG	Stockwell transform, SVM	44	82.3	0.771	NP	NP	NP	NP	NP	NP

Note: leave-one-out cross-validation (LOOCV)

Table 7
Comparison of the performance with the training and testing set on the dataset ISRUC-Sleep

method	Number of record	Validation way	Signal types	model	Number of features	Acc (%)	Kappa	Balanced F-score	Awake F-score	REM F-score	N1 F-score	N2 F-score	N3 F-score
[3]	99	5-fold cross-validation	6EEG + 2EOG + 1EMG + 1ECG	Deep learning	/	86	0.82	NP	0.94	0.84	0.67	0.86	0.89
[16]	10	5 records for training, 5 records for testing	left EOG channel	RUSBoost	30	84.7	NP	NP	NP	NP	NP	NP	NP
[16]	10	5 records for training, 5 records for testing	left EOG channel	RF	30	86.0	NP	NP	NP	NP	NP	NP	NP
[16]	10	5 records for training, 5 records for testing	left EOG channel	SVM	30	85.4	NP	NP	NP	NP	NP	NP	NP
proposed method	10	5 records for training, 5 records for testing	2EOG + EMG	RF	114	83.5	0.779	0.805	0.861	0.892	0.573	0.856	0.844
[18]	126	106 records for training, 20 records for testing	6EEG + 2EOG + 1EMG	Long Short-Term Memory (LSTM)	NP	81	NP	0.80	0.82	0.89	0.78	0.70	0.82
[18]	126	106 records for training, 20 records for testing	6EEG + 2EOG + 1EMG	LSTM with Fuzzy entropy	NP	86	NP	0.84	0.88	0.86	0.90	0.70	0.84

Discussion

3.1 The necessity of feature normalization

As shown in Table 8, our proposed way for feature normalization improves the performance when in comparison with original features. For traditional feature normalization, i.e., y(n)=(x(n)-mean(X))/(max(X)-min(X)), the performance is even worse than the original features. The reason is that physiological signals often have significant individual characteristics. For example, although the lowest EMG amplitudes in most subjects occurred during deep or REM sleep, a few subjects tended to be different, so did the highest EMG amplitudes during wakefulness. This is the reason why traditional feature normalization should be reconsidered for noised physiological signals.

If the features are normalized by traditional way, features have the risk of being greatly influenced by noise; that is, the test accuracy will be greatly reduced. Therefore, the proposed feature normalization is superior to the traditional way in terms of generalization ability.

The influence of positions of index *n1*and *n3* in formula (8 ~ 10) is also tested. As shown in Table 8, the changes of their positions induce fluctuation in accuracy less than 1%. Obviously, the proposed classification method benefited from feature normalization.

Table 8
Comparison of the performance with the normalized or original features by LOOCV on subgroup-III of the dataset ISRUC-Sleep using 2EOG + EMG

feature	n1 (%)	n3 (%)	Acc (%)	Kappa	Balanced F-score	Awake F-score	REM F-score	N1 F-score	N2 F-score	N3 F-score
normalized by proposed way	10	90	77.7	0.709	0.739	0.851	0.809	0.412	0.762	0.862
	5	95	77.5	0.706	0.734	0.845	0.805	0.413	0.761	0.863
	15	85	77.3	0.703	0.733	0.850	0.810	0.392	0.758	0.855
	25	75	77.1	0.701	0.732	0.849	0.806	0.392	0.757	0.855
	20	80	77.1	0.700	0.734	0.848	0.799	0.413	0.754	0.857
original feature	/	/	77.0	0.702	0.741	0.834	0.820	0.462	0.754	0.836
normalized by traditional way	/	/	74.5	0.665	0.693	0.822	0.782	0.281	0.735	0.843

3.2 Selection of signal type for sleep monitoring

Currently, the gold standard for sleep monitoring is PSG that recorded in the hospital or in a sleep laboratory. PSG remains a complex, high demanding and obtrusive procedure [1]. Shortcomings limits the utility of PSG, such as discomfort sleep assessment, high cost and being labour-intensive. On the contrary, unattended and portable non-medical devices deliver highly unobtrusive measurements at the expense of accuracy and reliability, also referred as electronic gadgets [1]. The focus is that “how to assess sleep stages and sleep quality less intrusively bur more reliably” [1]. It is contradictory to improve the classification accuracy and reduce the intrusion degree during sleep.

When signals for sleep recording are selected from EEG, EOG and EMG signals, the more the signal types, the higher the classification accuracy. However, for reducing the burden on subjects and reducing sleep disturbance, the fewer the signal types, the better. According to the AASM, six- channel unipolar EEG acquisition needs 8 electrodes, two- channel EOG acquisition needs 3 electrodes, one- channel EMG acquisition needs 2 electrodes. These electrodes and their cables will greatly influence the sleep quality.

The brain provides the most useful information about sleep regulation. EEG is the most important signal that directly reflects the state of the brain in sleep [19]. However, due to the structure of the scalp and the effects from the hair, novel materials or electrodes are still hard to obtain that can be used for EEG acquisition with comfort and high signal-to-noise ratio [19].

Using only one type of physiological signal can further reduce the number of electrodes attached to the body and the physiological load. When only one physiological signal is selected from EOG and EMG, as shown in Table 6, the precision for sleep staging derived from EOG is much higher than that from EMG, but lower than EEG alone. EOG-based sleep staging is relatively poorer than the method of using EEG, because the criterion of scoring the sleep stage mainly depends on the characteristics of EEG signals.

Unlike EEG, EOG electrodes can be placed below the hairline with self-adhesive electrode without the assistance of experts [16]. EOG is highly useful to identify the wakefulness and REM sleep, since there are major eye movements during these stages. Eye movements tend to slow down with the depth of sleep. EOG signals record the movement of the eyes is a fundamental indicator to distinguish between REM and NREM stages [19]. As shown in Fig. 2b, PLV_{org} of EOG is usually bigger than 0 during N2 and N3 sleep, and it is usually smaller than 0 during wakefulness and REM sleep. Visual examples are shown in Fig. 6 and Fig. 7. During REM sleep in Fig. 6, the polar between left EOG and right EOG is opposite, i.e., when a peak shows in left EOG, a valley will show in right EOG, such as the position of 12-s.

During N3 sleep in Fig. 7, the polar between left EOG and right EOG is consist, i.e., when a peak shows in left EOG, a similar peak will also show in right EOG, such as the position of 5-s. Furthermore, waveforms in EOG are to some extent similar to that of EEG, which is obvious in Fig. 7.

When compared with the microvolts of EEG’s small amplitude variations, EOG and EMG show in millivolts and requires less stable contact with the body, which make them less sensitive to noise and more suitable for unobtrusive measurement apparatus. Hence, the combination of EOG and EMG is a good choice for sleep monitoring.

3.3 sleep monitoring by EOG and EMG

The EOG makes the main contribution for sleep scoring derived from the combination of EOG and EMG. The placements of EOG electrodes are close to the placements of Fp1 and Fp2 EEG electrodes [19]. Hence, the EOG recordings are highly influenced by a portion of EEG activities [19]. During NREM sleep, EOG has a similar waveform with the EEG signals recorded at the frontal poles [19].

Using 114 normalized features from a combination of EOG (108 features) and EMG (6 features), Cohen's kappa coefficient by RF from LOOCV (N = 124) was 0.749 and the accuracy was 80.8%. The F1-scores were 0.857, 0.855, 0.483, 0.790 and 0.851 for wakefulness, REM sleep, N1 sleep, N2 sleep and N3 sleep, respectively.

In addition, using 438 normalized features from the combination of EEG (324 features), EOG (108 features) and EMG (6 features), Cohen's kappa coefficient by RF from LOOCV (N = 124) was 0.801 and the accuracy was 84.7%. The F1-scores were 0.894, 0.884, 0.528, 0.836 and 0.894 for wakefulness, REM sleep, N1 sleep, N2 sleep and N3 sleep, respectively. Consequently, the performances on sleep classification that achieved by EOG + EMG are comparable with that of EEG + EOG + EMG.

Conclusion

On a public data set called ISRUC-Sleep, comparative analysis suggests that the performances on sleep classification that achieved by EOG+EMG are comparable with that of EEG+EOG+EMG. The EOG makes the main contribution for sleep scoring derived from the combination of EOG and EMG. The proposed method from the combination of EOG and EMG can achieve comparable performance as using EEG signals for sleep staging.

Declarations

Acknowledgements

This study was funded by State Key Laboratory of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center (SMFA15B06, SMFA15A01). We acknowledge the support from the public dataset ISRUC-Sleep, provided by the Sleep Medicine Centre of Coimbra.

Compliance with ethical standards

Conflict of interest

The authors have no financial interest in any related entities.

Ethical approval

Ethical approval was not required because all data used in this paper were from the public dataset ISRUC-Sleep web site "http://sleeptight.isr.uc.pt/ISRUC_Sleep/".

References

1. GEORGES MATAR, JEAN-MARC LINA ,JULIE CARRIER, AND GEORGES KADDOUM.Unobtrusive Sleep Monitoring Using Cardiac,Breathing and Movements Activities: An Exhaustive Review [J].Digital Object Identifier 10.1109/ACCESS.2018.2865487
2. Huy Phan, Fernando Andreotti, Navin Cooray, Joint Classification and Prediction CNN Framework for Automatic Sleep Stage Classification [J].IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 66, NO. 5, MAY 2019 1285-1296.
3. Rui Yan,Fan Li,Dong Dong Zhou,Tapani Ristaniemi,Fengyu Cong.Automatic sleep scoring: A deep learning architecture for multi-modality time series[J]Journal of Neuroscience Methods,2021,348:108971
4. CHIH-EN KUO, GUAN-TING CHEN.Automatic Sleep Staging Based on a Hybrid Stacked LSTM Neural Network: Verification Using Large-Scale Dataset [J].Digital Object Identifier 10.1109/ACCESS.2020.3002548
5. Iber C, Ancoli-Israel S, Chesson A et al (2007) The AASM manual for the scoring of sleep and associated events: rules,terminology and technical specifications [M]. Westbrook Corporate Center AASM, USA, Westchester, pp 19–37
6. R. S. Rosenberg and S. Van Hout, "The american academy of sleep medicine inter-scorer reliability program: Sleep stage scoring," J. Clin. Sleep Med., vol. 9, no. 1, pp. 81–87, Jan. 2013.
7. Fabio Mendona, Sheikh Shanawaz Mostafa, Fernando Morgado-Dias. A Review of Approaches for Sleep Quality Analysis[J]Digital Object Identifier 10.1109/ACCESS.2019.2900345
8. Takashi Nakamura, Valentin Goverdovsky, Mary J. Morrell, Danilo P. Mandic.Automatic Sleep Monitoring Using Ear-EEG [J]. Digital Object Identifier 10.1109/JTEHM.2017.2702558
9. Marcel Myczak, Tulio A. Valdez , Wojciech Kukwa. Joint Apnea and Body Position Analysis for Home Sleep Studies Using a Wireless Audio and Motion Sensor [J].Digital Object Identifier 10.1109/ACCESS.2020.3024122
10. Khalighi S, Sousa T, Santos J M, et al. ISRUC-Sleep: a comprehensive public dataset for sleep researchers [J]. Computer Methods and Programs in Biomedicine, 2016,124:180 -192.
11. Peyman Ghasemzadeh; Hashem Kalbkhani; Mahrokh G. Shayesteh. Sleep stages classification from EEG signal based on Stockwell transform [J]. IET Signal Processing, 2019, 13(2): 242-252
12. Li Yanjun, Tang Xiaoying, Xu Zhi, Liu Weifeng, Li Jing. Temporal correlation between two channels EEG of bipolar lead in the head midline is associated with sleep-wake stages [J]. Australasian Physical & Engineering Sciences in Medicine, 2016, 39(1):147-155.

13. Pan H, Xu Z, Yan H, et al. Lying position classification based on ECG waveform and random forest during sleep in healthy people[J]. Biomedical Engineering Online, 2018, 17(1):116.
14. Sirvan Khalighi, Teresa Sousa, Gabriel Pires, Urbano Nunes. Automatic sleep staging: A computer assisted approach for optimal combination of features and polysomnographic channels [J]. Expert Systems with Applications 40 (2013) 7046–7059
15. Huaming Shen, Feng Ran, Meihua Xu, Allon Guez , Ang Li and Aiying Guo. An Automatic Sleep Stage Classification Algorithm Using Improved Model Based Essence Features [J]. Sensors 2020, 20, 4677; doi:10.3390/s20174677
16. Md Mosheyur Rahman, Mohammed Imamul Hassan Bhuiyan, Ahnaf Rashik Hassan. Sleep stage classification using single-channel EOG[J]. Computers in Biology and Medicine, 2018,102:211–220.
17. Huaming Shen,Feng Ran,Meihua Xu,Allon Guez,Ang Li,Aiying Guo.An Automatic Sleep Stage Classification Algorithm Using Improved Model Based Essence Features[J].Sensors, 2020, 20, 4677; doi:10.3390/s20174677
18. Peiying Shi , Xiangwei Zheng , Ping Du, and Feng Yuan. Automatic Sleep Stage Classification Based on LSTM [J]. Chinese CSCW 2018, CCIS 917, pp. 478-486, 2019.
19. Chenglu Sun, Chen Chen, Jiahao Fan, Wei Li, Yuanting Zhang, Wei Chen.A hierarchical sequential neural network with feature fusion for sleep staging based on EOG and RR signals [J].Journal of Neural Engineering,16 066020

Figures

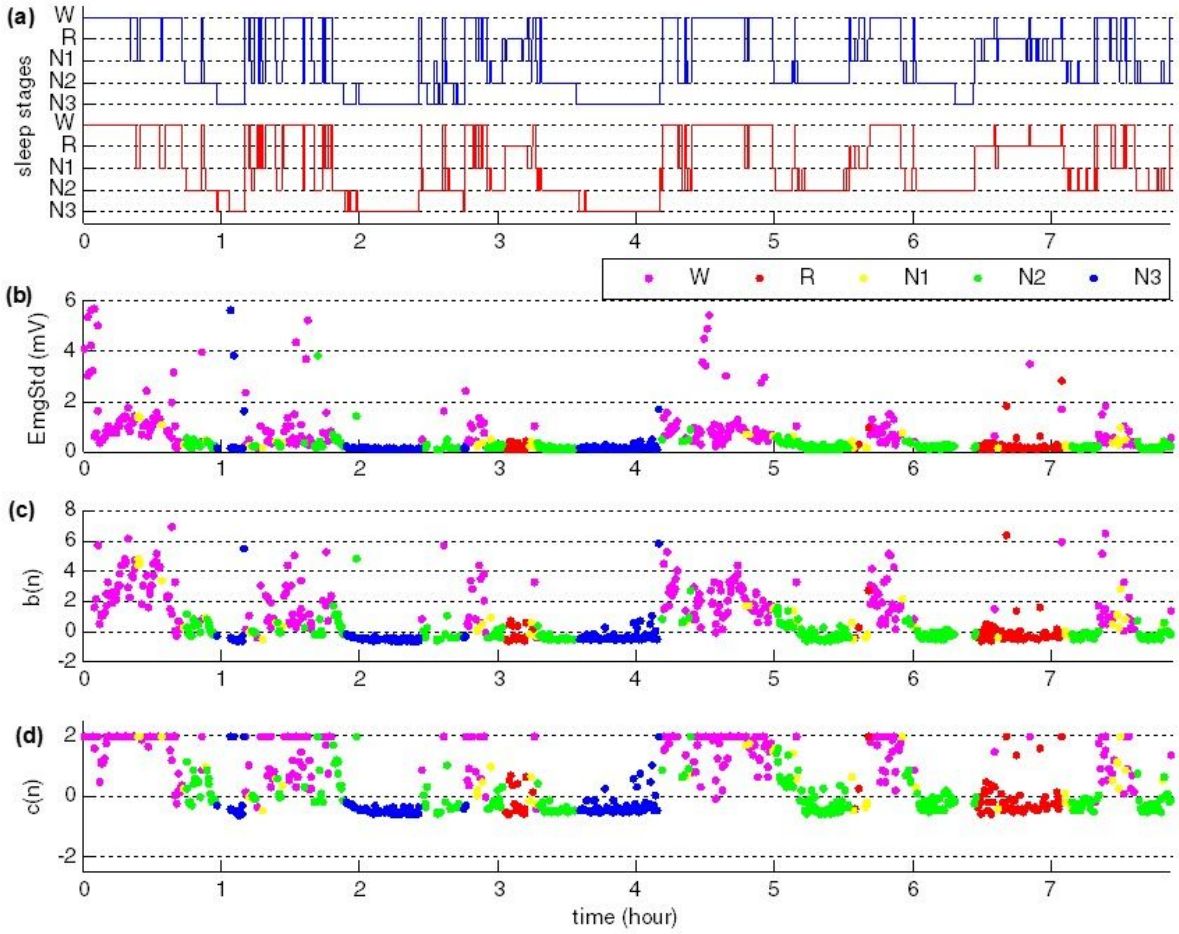

Figure 1

quasi-normalization for EmgStd of EMG (a) Manual scoring from the first expert as blue line, and manual scoring from the second expert as red line; (b) original index EmgStd, different stages with different colors; (c) sequences {b(n)}; (d) sequences {c(n)}

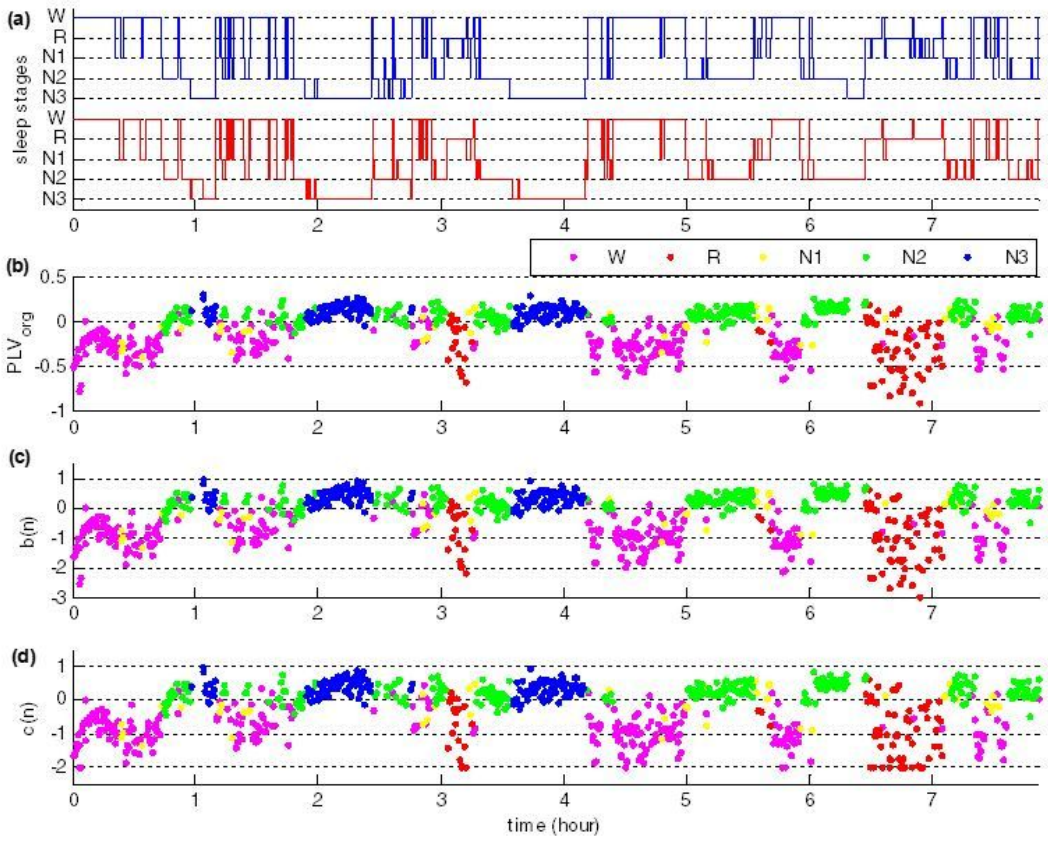

Figure 2

quasi-normalization for PLVorg of EOG (a) Manual scoring from the first expert as blue line, and manual scoring from the second expert as red line; (b) original index PLVorg of EOG, different stages with different colors; (c) sequences {b(n)}; (d) sequences {c(n)}

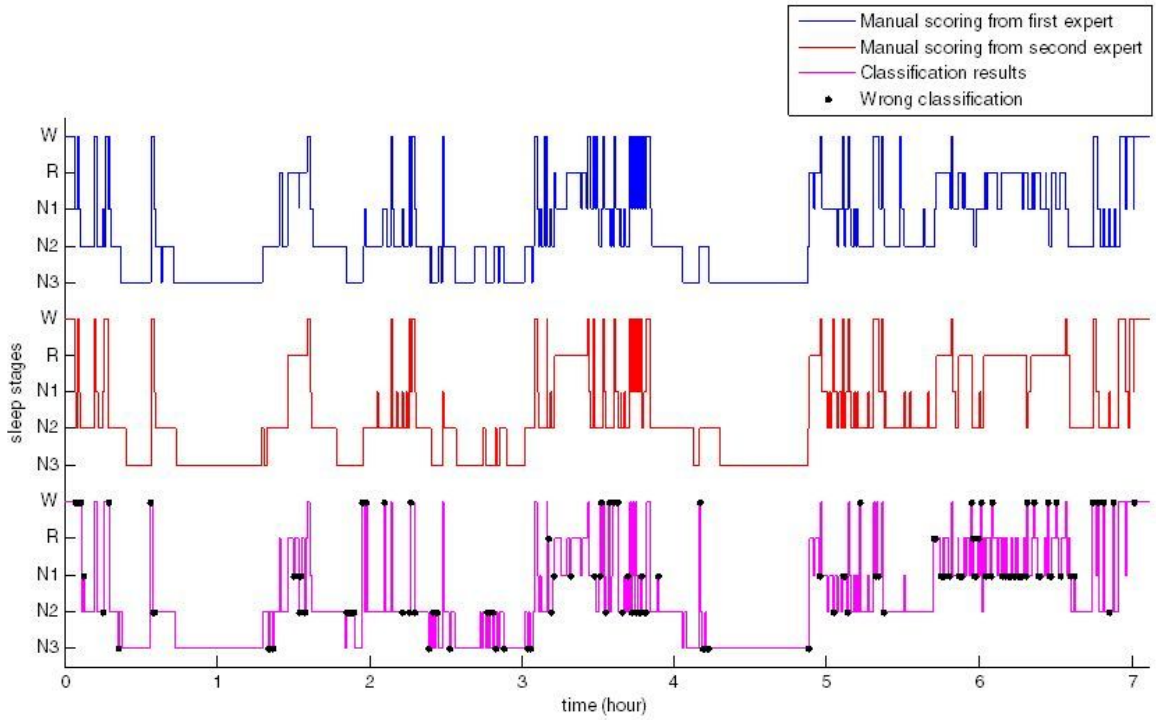

Figure 3

Classification results of No.6 from subgroup-III

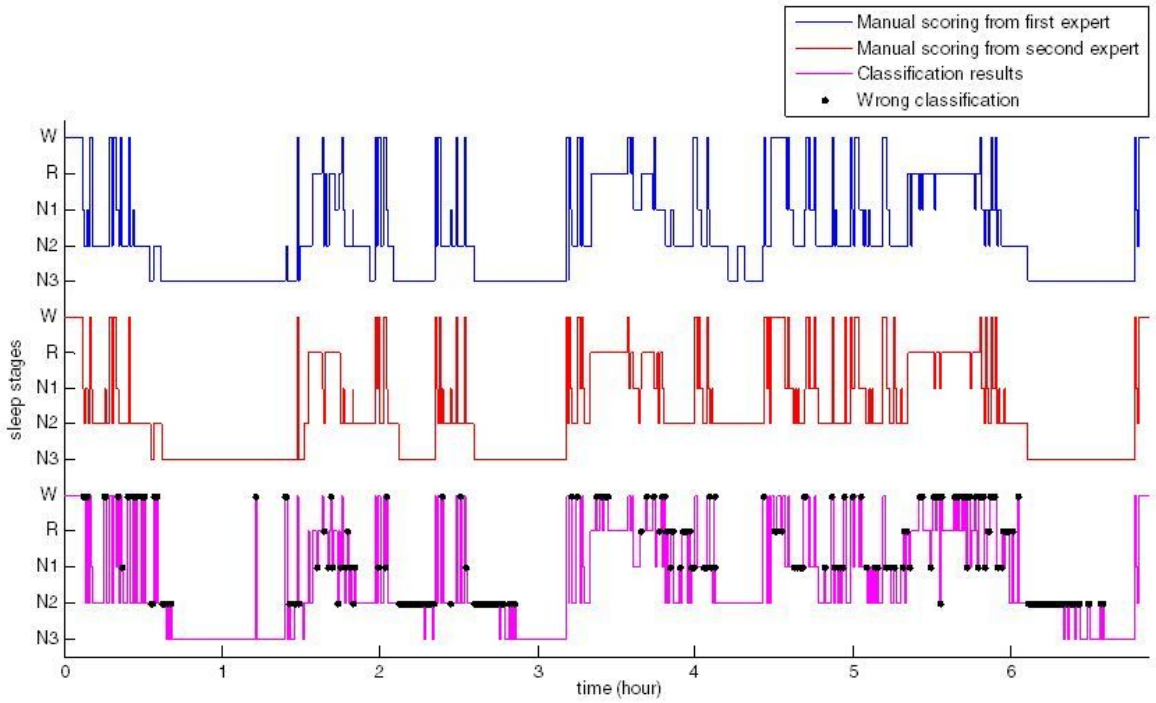

Figure 4

Classification results of No.3 from subgroup-III

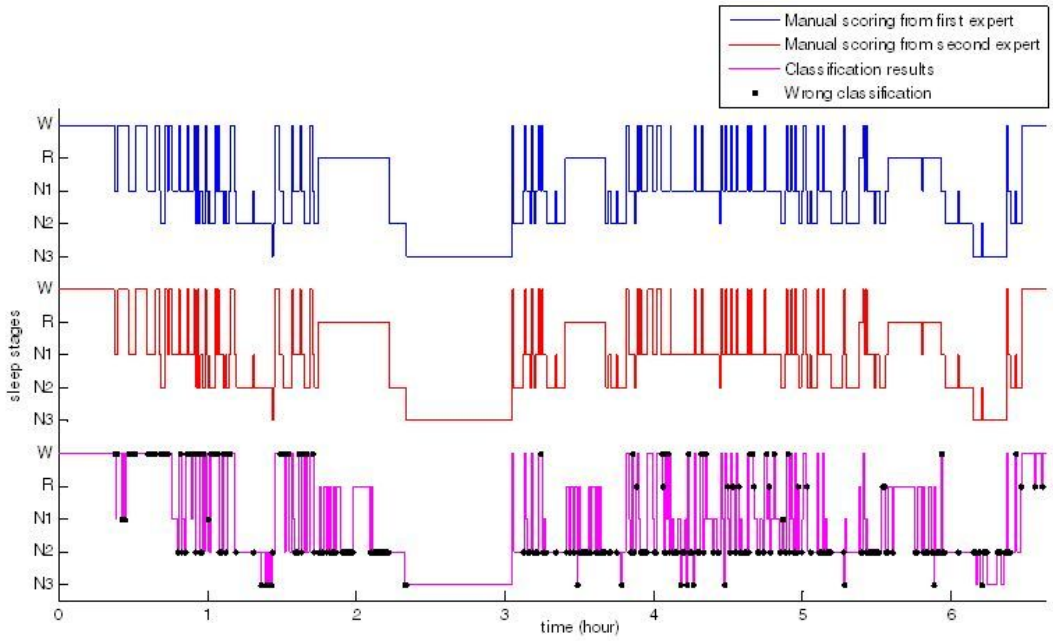

Figure 5

Classification results of No.10 from subgroup-III

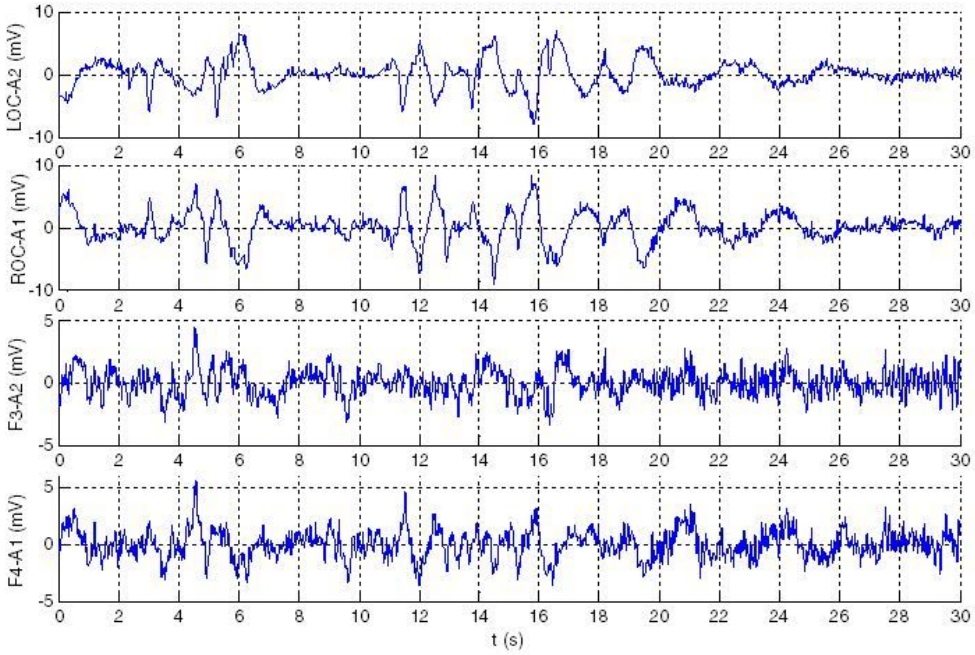

Figure 6

EOG and EEG during REM sleep

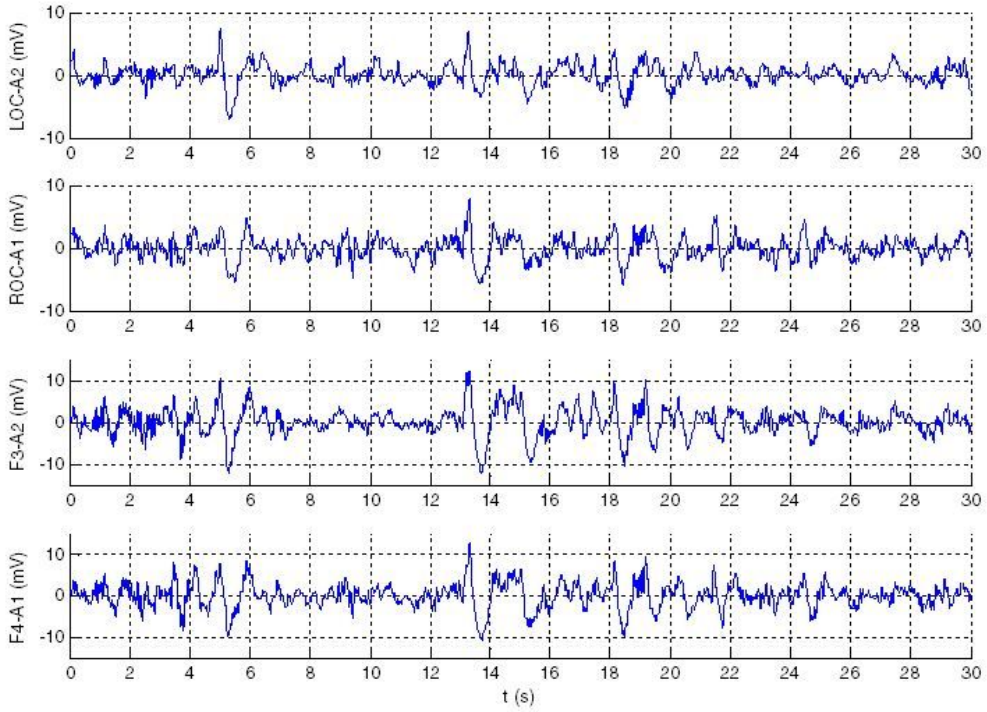

Figure 7

EOG and EEG during N3 sleep