An Automatic Sleep Scoring Toolbox: Multi-modality of Polysomnography Signals' Processing

Rui Yan[1, 2], Fan Li[3], Xiaoyu Wang[2], Tapani Ristaniemi[1] and Fengyu Cong[1, 2]
[1]*Faculty of Information Technology, University of Jyväskylä, 40014, Jyväskylä, Finland*
[2]*School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, 116024, Dalian, China*
[3]*School of Information Science and Engineering, Dalian Polytechnic University, 116034, Dalian, China*

Keywords: Polysomnography, Multi-modality Analysis, MATLAB Toolbox, Automatic Sleep Scoring.

Abstract: Sleep scoring is a fundamental but time-consuming process in any sleep laboratory. To speed up the process of sleep scoring without compromising accuracy, this paper develops an automatic sleep scoring toolbox with the capability of multi-signal processing. It allows the user to choose signal types and the number of target classes. Then, an automatic process containing signal pre-processing, feature extraction, classifier training (or prediction) and result correction will be performed. Finally, the application interface displays predicted sleep structure, related sleep parameters and the sleep quality index for reference. To improve the identification accuracy of minority stages, a layer-wise classification strategy is proposed according to the signal characteristics of sleep stages. The context of the current stage is taken into consideration in the correction phase by employing a Hidden Markov Model to study the transition rules of sleep stages in the training dataset. These transition rules will be used for logic classification results. The performance of proposed toolbox has been tested on 100 subjects with an average accuracy of 85.76%. The proposed automatic scoring toolbox would alleviate the burden of the physicians, speed up sleep scoring, and expedite sleep research.

1 INTRODUCTION

Sleep covers almost one-third of human lifespan. Adequate and high-quality sleep is vital to our physical and mental well-being (Pagel and Pandi-Perumal, 2014). However, and likely because of our ephemeral lifestyle in modern society, sleep disorder complaints increase dramatically among people. Assessing sleep behaviour and analysing the sleep structure, therefore, become more and more crucial. Up to now, the conventional visual scoring method is still the main method in most clinical and sleep research labs worldwide.

Visual scoring, mainly based on the rules of Rechtschaffen & Kales (R&K) (Rechtschaffen and Kales 1968) and the recently updated American Academy of Sleep Medicine rules (AASM) (Berry et al., 2012), requires at least one registered sleep technologist (RST) who has sufficient expertise and experience in sleep scoring. Generally, the annotation of 8-h recording requires approximately 2-4 hours (Hassan and Bhuiyan, 2016a), which is rather time-consuming. Besides, visual scoring in some degree is subjective, as the inter-scorer reliability among trained technologists is less than 90% (Danker-Hopfe et al., 2009). In contrast, automatic sleep scoring has demonstrated advantages of cost-effective and preferable scoring performance.

Electroencephalogram (EEG) signals are mainly used in automatic sleep scoring since they contain valuable and interpretable information resembling brain activities (Boostani et al., 2017). According to the morphological characteristics of EEG signals, sleep EEG waves are mainly composed by α wave, β wave, θ wave and δ wave, K complex, sleep spindles and saw-tooth (Niedermeyer and da Silva, 2005). These rhythm waves form the foundation of sleep scoring. Some studies (Hassan et al., 2015; Hassan and Bhuiyan, 2016b) tried to extract statistical and spectral features from these rhythm waves to perform an automatic sleep scoring. Cross frequency coupling estimated between rhythm waves also showed high classification accuracy (Dimitriadis et al., 2018). Instead of traditional linear features, multiscale entropy and autoregressive models for single-channel EEG were employed in Liang et.al' s study, obtaining a good scoring performance (Liang et al., 2012).

Sleep is a complex process involving multiple

Yan, R., Li, F., Wang, X., Ristaniemi, T. and Cong, F.
An Automatic Sleep Scoring Toolbox: Multi-modality of Polysomnography Signals' Processing.
DOI: 10.5220/0007925503010309
In *Proceedings of the 16th International Joint Conference on e-Business and Telecommunications (ICETE 2019)*, pages 301-309
ISBN: 978-989-758-378-0

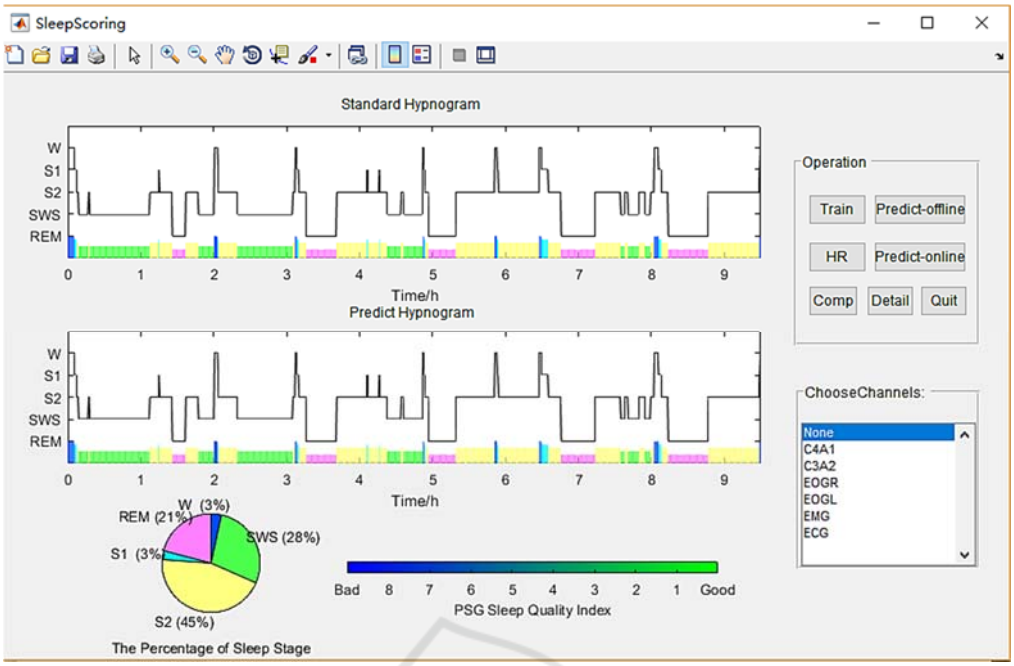


Figure 1: The interface of sleep scoring toolbox.

organs. Signals recorded from different physical areas change with the sleep cycle. The multi-modality signals' contribution to sleep scoring has been explored in several studies (Gharbali et al., 2018; Yan et al., 2019; Šušmáková and Krakovská, 2008). Özşen concluded that as the sleep deepen, the frequency of EEG signals attenuated gradually, along with rare eye movements, low electromyography (EMG) activity and slow heart rate (Özşen, 2013). Ebrahimi and his colleagues found that under the control of parasympathetic nervous system and sympathetic nervous system, cardiovascular and respiratory behaviours fluctuated with the alternation of sleep stage (Ebrahimi et al., 2015). It has demonstrated that features from multi-modality signals were beneficial to the improvement of scoring accuracy (Boostani et al., 2017).

Although there are many studies on automatic sleep scoring, the available software and toolbox is limited. Given that, this study aims to develop an automatic sleep scoring toolbox with the capability of multi-signal processing, see Figure 1. The main contributions of this work are presented as following:

a) An automatic sleep scoring toolbox is proposed which supports multiple sleep signals and two data formats.
b) An interactive interface is provided which allows the user to select the number of target classes, change signal types and visualize various analysis results.
c) A layer-wise classification strategy is proposed which can significantly improve the classification accuracy of minority stages without compromising the accuracy of other classes.
d) A correction procedure is proposed to make classification results logical.

The article is organized as follows: Section 2 explains the details of experimental data and methodology of this study. Section 3 demonstrates the performance of proposed toolbox. Section 4 provides discussions of results and limitations of this study. Finally, section 5 gives conclusions of this paper.

2 MATERIALS AND METHODS

2.1 System Overview

The proposed toolbox consists of a training module, an offline prediction module, an online prediction module and several parameter panels, as shown in Figure 1. Their functions are briefly described in the following lines. The specific model structure will be introduced in detail in section 2.5.

Training Module: The objective of the training module is to train a classifier based on the user's selection. The user can choose signal types and the number of target stages as required. The software automatically performs signal pre-processing, feature

extraction and classifier training. The output of this module is a trained model which can be used to predict sleep structures.

Prediction Module: The aim of this module is to predict sleep structure based on the predefined model or user-specified model. The module automatically checks if the user has trained a model, and allows the user to determine if the predefined model is needed. Once the model selected, the module automatically processes the test data based on model parameters. Finally, the application interface displays the predicted sleep structure, related sleep parameters and a sleep quality index as a reference to sleep quality. If a hypnogram (e.g., labels scored by RST) is available for the test data, the interface would display both the hypnogram and predicted labels together, and highlight the disagreement by pressing the button named "Comp".

Online Prediction Module: The module is similar to the offline prediction process except for the real-time updating results. The module can be connected to a sleep monitoring device in order to realize the real-time analysis of sleep signals and to visualize sleep structures. The updated sleep signal will be saved as a TXT file in storage.

2.2 Description of Experiment Data

The sleep data for this investigation was provided by the Sleep Heart Health Study (SHHS) database. We used only the first round (SHHS-1) due to its wide age range. The recordings employed in this study were selected by considering a Respiratory Disturbance Index 3 Percent (RDI3P) < 5 to have near-normal characteristics. Moreover, subjects did not use beta-blockers, alpha-blockers, inhibitors, and did not suffer documented hypertension, heart disease, or history of stroke. Given that, a total number of 100 subjects were selected with the total duration of 816 hours and 43 minutes. The age of subjects ranged from 40 to 54 years, with a mean value of 47 years and a standard deviation of 4.3 years. Each record was scored by the experienced research assistant or sleep technologist according to the R&K rules. The sleep recordings were segmented into 30-second per epoch and labelled as wakefulness (W), non-rapid eye movement stage (NREM, containing S1, S2, S3 and S4) and rapid eye movement stage (REM). The deepest NREM stage, namely S3 and S4, were collectively referred to as "slow wave sleep" (SWS), based on a prevalence of low-frequency oscillations (Berry et al., 2012). A detailed description of SHHS was given in the study (Quan et al., 1997) .

2.3 Pre-processing

For the predefined model and the following experiments, four modalities of polysomnography (PSG) signals were considered: EEG channels (C4-A1 and C3-A2, following the 10-20 international electrode placement system), two electrooculography (EOG) channels (named: ROC, LOC), one submental electromyography (EMG) channel and one electrocardiography (ECG) channel. All the aforementioned signals were fully included within the evaluation process without discarding any recorded segments, thereby to have a near-clinical situation.

In order to remove noise and artefacts, a notch filter, a high-pass filter with a cut-off frequency of 0.3Hz and a low-pass filter with a cut-off frequency of 30Hz were applied to the signals of EEG, EOG and ECG. In terms of EMG, a notch filter, a high-pass filter with a cut-off frequency of 10Hz and a low-pass filter with a cut-off frequency of 75Hz were performed. The whole night recordings were smoothed by its mean value ±5×standard deviation to remove the outliers. In order to eliminate individual differences, the sleep signals were normalized to [-100, 100]. Afterwards, all the signals were divided into 30-second epochs, each epoch corresponding to a single sleep stage.

2.4 Feature Extraction

The features, employed in this study, involves a variety of traditional and modern characteristics serving as distinctive markers for various psycho-physiological states. They are summarized in Table 1. Some of the parameters are introduced in the following, and the others can be found in Yan et al.'s research (Yan et al. 2019).

2.4.1 Time Domain Parameters

Some statistical parameters, such as minimum value, maximum value, standard deviation, arithmetic mean, variance, skewness, kurtosis and median are derived from signal segments. These statistical parameters are good indicators of the amplitude and distribution of time series (Şen et al. 2014). Percentile analysis is known as the most effective time domain measures for EEG signals (Boostani et al. 2017). Hjorth parameters (i.e., activity, mobility and complexity) represent the signal power, the mean frequency and frequency changes (Vidaurre et al. 2009).

Table 1: Parameter list.

Type	Feature Name
Statistical measures	Minimum Value (MinV), Maximum Value (MaxV), Arithmetic Mean(AM), Median(M), Standard Deviation (SD), Variance(V), Skewness(S), Kurtosis(K), The 5th Percentile (Pre5), The 25th Percentile (Pre25), The 75th Percentile (Pre75), The 95th Percentile (Pre95), Hjorth Parameters（HA, HM, HC）, Zero-Crossing(ZC)
Spectral measures	Power Spectral Density(PSD), Mean Value of PSD (mPSD), Median Value of PSD (mdPSD), Power Ratio(PR), Absolute and Relative Spectral Power (APSD, RPSD), Brain Rate (BR), Spectral Centroid (Sc), Spectral Width (Sw), Spectral Asymmetry (Sa), Spectral Flatness (Sk), Spectrum Flatness (Sf), Spectral Slope (Ss), Spectral Decrease (Sd), Edge_D, Spectral Edge Frequency at 90% and 50%
Nonlinear measures	Mean teager energy (MTE), Mean Energy (E), Mean curve length (CL), SecD, The 4th Power,
Fractal measures	Petrosian fractal dimension (PFD)
Entropy measures	Spectral Entropy(SpE)
Mutual measures	Coherence

2.4.2 Spectral Features

The calculation of spectral measures is based on Fourier transform using hamming window in the time domain. The following spectral measures are considered.

Power spectral density is calculated based on the following formula. Meanwhile, its mean value and median value are also considered.

$$PSD = F(\omega) \times F^*(\omega)/N \quad (1)$$

where ω is the frequency, * representing the complex conjugate, and N is the length of time series.

Spectral edge is defined as the frequencies corresponding to 90% and 50% of the total spectral power (Imtiaz and Rodriguez-Villegas 2014). The difference between the two frequencies (edge_D) is also considered.

$$\sum_{f=f_{min}}^{edge} P(f) = p \sum_{f=f_{min}}^{30Hz} P(f) \quad (2)$$

where p is equal to 0.9 or 0.5, f_{min} is 0.3Hz in terms of EEG, EOG and ECG, and 10Hz in EMG.

Absolute and relative spectral power are obtained from seven frequency bands of EEG, namely, 0.3-4Hz (delta), 2-3.9Hz (K complex), 2-6Hz (saw-tooth), 4-8Hz (theta), 8-12Hz (alpha), 14-30Hz (beta), and 12-16Hz (spindle). Absolute spectral power is spectral power within the specific frequency bands. The relative value is defined as the ratio of the absolute value to the total spectral power. The total spectral powers of EEG, EOG and ECG signals are computed within the range of 0.3-30Hz, and 10-30Hz for EMG signals.

Power ratios are computed based on absolute spectral powers in aforementioned frequency bands. The following power ratios are computed: delta/theta, delta/alpha, delta/beta, theta/alpha, theta/beta, alpha/beta, alpha/(theta + delta), delta/(theta + alpha) and theta/(beta + delta).

Brain rate estimates the EEG mean frequency weighted over the brain spectrum distribution (Pop-Jordanova and Pop-Jordanov 2005).

$$BR = \sum_i^M f_i \times P_i \,/\, \sum_i^M P_i \quad (3)$$

where M is the number of frequency bins, i the sub-band, P_i the power of the spectral distribution corresponding to frequency band i, and f_i is the frequency at bin i.

Spectral centroid is defined as the frequency-weighted sum of the magnitude spectrum of the signal normalized by its unweighted sum, indicating the

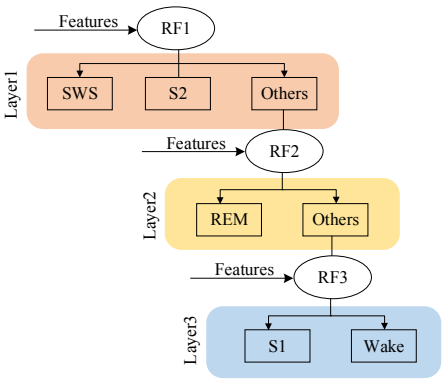


Figure 2: Layer-wise classifier.

location of the spectrum centre (Hassan et al. 2015). Spectral width is the wavelength interval over which the magnitude of all spectral components is equal to or greater than a specified fraction of the magnitude of the component having the maximum value. Spectral asymmetry represents the asymmetry in the distribution of the spectrum of eigenvalues of an operator. Spectral flatness, measured in decibels, provides a way to quantify how noise-like a sound is (Dubnov 2004). Spectrum flatness defines the planeness properties from an audio signal's spectrum, which shows how the power spectrum of a signal deviates from a frequency of a flat shape (Lazaro et al. 2017). Spectral slope is a measure of the slope of the spectral shape (Hassan et al. 2015). The steepness of the decrease of the spectral envelope of the signal with respect to its frequency is defined as spectral decrease (Hassan et al. 2015). The detailed definition of these parameters can be found in Chen et al.'s study (Chen et al. 2018).

2.5 Classification

It is well-known that the distribution of epochs among sleep stages are highly imbalanced. Unfortunately, the traditional classifier is kind of sensitive to the distribution of data sets. When instances of one class in the training set vastly outnumber the instances of other classes, the classifier inclines to classify instances as belonging to the majority class and ends up creating suboptimal classification models in the process (Hassan and Bhuiyan 2016c). After studying the characteristics of sleep stages, we find that the REM, S1 and wakefulness present a certain similarity leading to misclassification. For example, the level of brain activity and eye movements increase in REM stage which is similar to the waking period. In addition, S1 is a transition phase of wakefulness and sleep, along with the ambiguous neuronal oscillation, that makes the detection of S1 is the most problematic of the sleep stages. For S2 and SWS, with the deepening of sleep, the activity levels of various organs decrease to some extent.

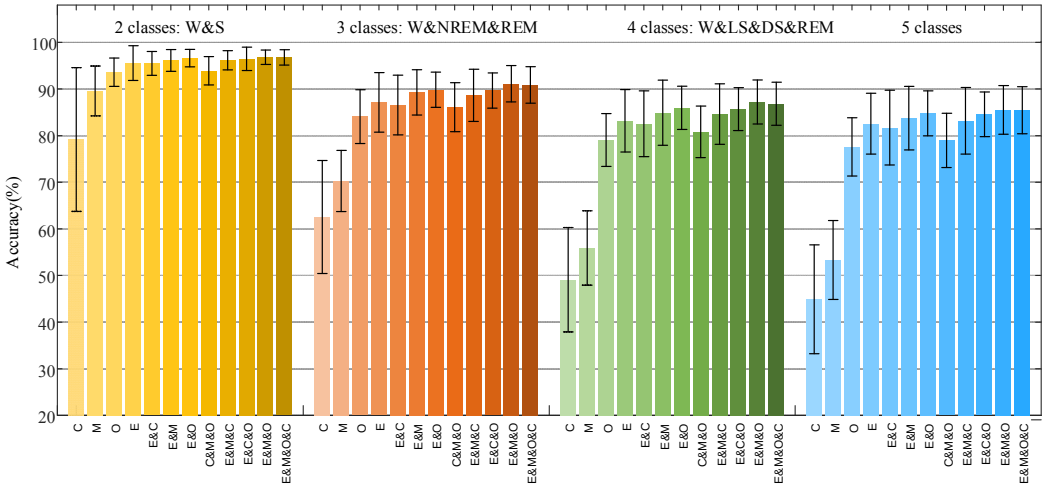
Based on these characteristics of the sleep stages, we develop a layer-wise classification strategy (See Figure 2) which is used in this toolbox to train and predict sleep structures. The strategy uses three random forest classifiers. The first layer is a multi-class classifier dividing the sleep sequences into SWS, S2, and others. The second layer is a two-class classifier, which aims to distinguish the REM stage according to its lowest EMG activity and obvious eye movements. The third layer discriminates the characteristics of S1 and awake stage. Experiments have confirmed that the structure can significantly improve the recognition accuracy of the minority sleep stages, such as S1, without significantly reducing the classification accuracy of other classes.

2.6 Result Correction

Studies have found that sleep transition is not a random process. However, the traditional classifier can only give its decision according to the information of the current stage, but can't remember the context. Therefore, a correction process is applied to classification results. Firstly, the Hidden Markov Model is used to learn the transition rules among sleep stages in the training data. Then, the correction rule can be derived according to these transition rules and some natural characters of sleep. These rules refer to the epochs prior to and posterior to the current epoch. The development of correction rules is inspired by the studies of Liang et al. (Liang et al. 2012) and Li et al. (Li et al. 2018). More specifically, the stage sequences, like $[S_{i-1}, S_i, S_{i+1}]$, are smoothed by the rules proposed in the study (Liang et al. 2012) to correct some sudden changes in predicted results. For the stage sequences that do not meet the aforementioned smooth rules, the transition rules derived from Hidden Markov Model will be used to analyse the rationality of the stage transitions.

Table 2: Sleep parameters and its definition.

Sleep Parameters	Definition
Time in bed	From light off to getting up
Sleep period time	From sleep onset to sleep end, in minutes
Sleep efficiency	Total sleep time / Bed time
Sleep onset latency	From recording start to sleep onset, in minutes
REM latency	From sleep onset to the occurrence of the first REM period, in minutes
Stage shifts/h	Number of sleep stage shifts after sleep onset per hour
Waking times	Number of awakenings after sleep onset per hour
Waking time	Wakefulness after sleep onset, percentage of sleep period time
Number of REM	Number of REM periods
Stage time	Specific stage time after sleep onset, in minutes
Stage percentage	Specific stage time in percentage of sleep period time



*C: single-modality ECG; M: single-modality EMG; O: single-modality EOG; E: single-modality EEG; &: combination signals; 2 classes: wakefulness and sleep (W&S); 3 classes: wakefulness, non-rapid eye movement sleep and rapid eye movement sleep (W&NREM&REM); 4 classes: wakefulness, light sleep (containing S1 and S2), deep sleep (SWS) and rapid eye movement sleep (W&LS&DS &REM); 5 classes: W, S1, S2, SWS and REM

Figure 3: The classification accuracy for different signal fusions and target class.

Table 3: Selected features for distinguishing specific pair of sleep stages.

Sleep stages		SWS-S2	SWS-S1	SWS-R	SWS-W	S2-S1	S2-R	S2-W	S1-R	S1-W	R-W
Top 15 features	Top1	C.Per25	C.Per25	C.Per25	M.ZC	C.PFD	O.ZC	C.ZC	O.ZC	E.SPE	O.ZC
	Top2	M.ZC	C.Per75	C.Per75	C.ZC	E.PSD	O.Sw	M.ZC	O.Per75	C.PSD	O.Per75
	Top3	C.Per75	M.ZC	C.Per95	M.PFD	C. mPSD	E.PR	E.SPE	E.K	O.Ss	O.Per25
	Top4	C.ZC	C.Per95	M.Per25	C.Per25	E.PR	E.PFD	C.PFD	O.Per25	O.Sf	O.PFD
	Top5	C.Per5	M.Sf	M.ZC	C.PFD	C.HM	O.edge90	O.Sf	O.PFD	E.MTE	M.ZC
	Top6	C.Per95	E.PSD	C.PSD	E.PSD	M.PFD	C. mPSD	M.PFD	O.HC	O.MTE	C.ZC
	Top7	C.K	C.Per5	E.Per95	C.Per75	O.Sd	C.K	O.Ss	O.HM	C.ZC	E.Per5
	Top8	O.K	M.Per25	C.Per5	M.Sf	M.Sf	O.Sd	E.PFD	O.SPE	C.Ss	O.Ss
	Top9	E.Per5	O.PFD	O.Power4	O.edge.D	E.PFD	O.PFD	C.HM	O.K	O.MaxV	O.RPSD
	Top10	M.Per25	E.RPSD	M.Per75	O.Sk	O.BR	E.S	C.PSD	O.K	E.D	O.CL
	Top11	M.K	C.ZC	E.PR	E.PR	O.PFD	O.BR	E.edge90	O.CL	C.Sf	C.PSD
	Top12	O.Per75	C.HM	C.K	E. mdPSD	C.ZC	C. mdPSD	E.PR	O.RPSD	O.RPSD	E.CL
	Top13	E.Per25	M.HM	C.HA	M.HM	O.edge90	O.HC	C.R R	O.SecD	E.MaxV	O.SecD
	Top14	M.HM	E.RPSD	O.Sf	E.PR	C.mPSD	O.S	C.mPSD	O.MaxV	M.ZC	O.Sf
	Top15	E.K	O.edge.D	C.PFD	E.PR	O.Sw	E.PR	M.HM	O.Sc	C.PFD	E.Sc

EEG features (colour: yellow); EOG features (colour: green); EMG features (colour: red); ECG features (colour: blue).

2.7 PSG Sleep Quality Index

Usually, sleep quality is evaluated by the standardized questionnaire, such as the Pittsburgh Sleep Quality Index (PSQI), the Berlin Questionnaire, and so on. These self-report questionnaires are subjective, and can be easily exaggerated or minimized by the person completing them. Furthermore, some items of questionnaires are challenging to self-evaluation. For example, the PSQI needs to evaluate the time it takes to fall asleep and the actual sleep time per night. Some papers claimed that the correspondence between the objective measurement and a person's subjective assessment of the sleep quality is surprisingly small, if existent (Sohn et al., 2012). In order to overcome the uncertainty of subjective assessment, the toolbox proposed a sleep quality index. The algorithm will calculate various sleep parameters (summarized in Table 2) according to the predicted sleep stages. Based on these sleep parameters, PSG sleep quality index is statistically calculated and displayed in a bar in the lower-right corner of the interface. Detailed sleep parameters can be obtained by pressing the button "Detail".

3 PERFORMANCE ASSESSMENT

3.1 Influence of Signal Types

In order to explore the relationship between signal types and classification accuracy, we performed a greedy search for several signal fusions referring to different target classes. The result was shown in Figure 3 where the column denoted the mean accuracy of 10-fold cross-validation and the bars represented the standard deviation. Four categories were considered, highlighted in different colours in Figure 3. For each category, twelve signal fusions were listed along X-axis where signals' names were abbreviated to its middle letter.

Figure 3 depicted the uncertainty or variation of classification accuracy under each condition. Collectively, with the enrichment of signal types, the mean value of accuracy increased, and the uncertainty decreased to some extent. More specifically, Figure 3 indicated that the required signal types varied with the number of target classes. If sleep recordings were classified into two classes, namely wakefulness (W) and sleep (S), all considered signal fusions gave satisfactory results. With the increasing number of target classes, the number of required signals increased accordingly.

From the perspective of signal types, the signal fusions containing EEG signals showed better identification accuracy, indicating a crucial role of EEG signals in sleep scoring. Furthermore, the discriminative information provided by ECG and EMG channels was inferior to that from EEG and EOG signals.

3.2 Feature Evaluation

To further elucidate the contributions of features and signals, the important features, measured by their contribution to distinguishing each pair of sleep stages, were derived from random forest classifier. The top 15 features were shown in Table 3, where features sorted in descending order of discriminative capability. As can be seen from Table 3, the features from EEG contributed to the recognition of most stages. Meanwhile, ECG features demonstrated its contribution to the discrimination of SWS from the others. For EOG signal, its features were good at distinguishing REM stage and wakefulness. In terms of feature types, the top 15 features indicated that the optimal feature subset was a fusion of statistical measures (e.g. Percentiles, Hjorth parameters, Zero-Crossing), spectral measures (e.g. spectral edge, power spectral density), entropy measures (e.g. spectral entropy), fractal measures (e.g. Petrosian fractal dimension) and nonlinear measures (e.g. mean curve length, the 4th Power).

4 DISCUSSION

PSG, the golden standard for measuring sleep qualitatively, is a traditional technology which is time-consuming and has barely changed over the years. Burgeoning public interest in sleep quality improves a strong impetus for a robust, easily implemented and rapid sleep scoring system. Limited toolbox or software is available for automatic sleep scoring, although there are many theoretical researches in this field. In previous studies, some portable devices were developed based on ECG and respiration. For example, Hermawan et al. (Hermawan et al., 2012) developed a real-time sleep stage classification device which classified sleep recordings into 2 stages (wakefulness and sleep) with an average precision of 0.941. Recently, some deep learning-based scoring tools sprouted out, such as SLEEPNET (Biswal et al., 2017) and SeqSleepNet (Phan et al., 2019). The classification accuracy of these deep learning-based tools was about 0.85 with the support of tremendous training data and highly configured computer (like GPU or server).

Compared with previous studies, the proposed toolbox provides comparable precision and greater freedom. The toolbox, based on MATLAB, allows users to select the available signal types and the number of target classes according to their condition and need. Meanwhile, it supports two popular data formats (MAT file and EDF file) that make data transfer easy. This offline prediction module is helpful for researchers, especially the newcomers in this field, to accelerate their understanding of sleep structures. It can also be used in clinic to speed up the annotation of PSG records, thus alleviating the burden of the physicians. The online prediction module provides the potential to control sleep tasks automatically by combining the toolbox with sleep experiments.

Even though our results are encouraging, our model still has several limitations. One of them is that the performance of proposed toolbox is affected by the data property. As our model learns from training data, it might not perform well when the trained model is applied to the data with different properties. For example, a scoring model trained by healthy subjects may not perform well for the analysis of patients' sleep structure. To achieve better results in

that condition, the model might have to be re-trained or fine-tuned.

5 CONCLUSIONS

This paper proposed an automatic sleep scoring toolbox that supported four types of sleep signals and two data formats. The toolbox provided an interface for user-friendly operation. Sleep recordings could be automatically analysed to reveal multiple sleep parameters and sleep quality index. A layer-wise classification strategy was proposed to improve the classification accuracy of minority stages. In addition, a Hidden Markov Model was used to make classification results logic. Compared with manual scoring, the proposed automatic scoring toolbox is cost-effective, which would alleviate the burden of the physicians, speed up sleep scoring and expedite sleep research.

ACKNOWLEDGEMENTS

The authors would like to thank the SHHS for providing the polysomnographic data. This work was supported by the scholarships from China Scholarship Council (Nos. 201606060227).

REFERENCES

Berry, R.B., Budhiraja, R., Gottlieb, D.J., Gozal, D., Iber, C., Kapur, V.K., Marcus, C.L., Mehra, R., Parthasarathy, S., Quan, S.F. and Redline, S., 2012. Rules for scoring respiratory events in sleep: update of the 2007 AASM manual for the scoring of sleep and associated events. *Journal of clinical sleep medicine*, *8*(05), pp.597-619.

Biswal, S., Kulas, J., Sun, H., Goparaju, B., Westover, M.B., Bianchi, M.T. and Sun, J., 2017. SLEEPNET: automated sleep staging system via deep learning. *arXiv preprint arXiv:1707.08262*.

Boostani, R., Karimzadeh, F. and Nami, M., 2017. A comparative review on sleep stage classification methods in patients and healthy individuals. *Computer methods and programs in biomedicine*, *140*, pp.77-91.

Chen, T., Huang, H., Pan, J. and Li, Y., 2018, May. An EEG-based brain-computer interface for automatic sleep stage classification. In *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)* (pp. 1988-1991). IEEE.

Danker-hopfe, H., Anderer, P., Zeitlhofer, J., Boeck, M., Dorn, H., Gruber, G., Heller, E., Loretz, E., Moser, D., Parapatics, S. and Saletu, B., 2009. Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard. *Journal of sleep research*, *18*(1), pp.74-84.

Dimitriadis, S.I., Salis, C. and Linden, D., 2018. A novel, fast and efficient single-sensor automatic sleep-stage classification based on complementary cross-frequency coupling estimates. *Clinical Neurophysiology*, *129*(4), pp.815-828.

Dubnov, S., 2004. Generalization of spectral flatness measure for non-gaussian linear processes. *IEEE Signal Processing Letters*, *11*(8), pp.698-701.

Ebrahimi, F., Setarehdan, S.K. and Nazeran, H., 2015. Automatic sleep staging by simultaneous analysis of ECG and respiratory signals in long epochs. *Biomedical Signal Processing and Control*, *18*, pp.69-79.

Gharbali, A.A., Najdi, S. and Fonseca, J.M., 2018. Investigating the contribution of distance-based features to automatic sleep stage classification. *Computers in biology and medicine*, *96*, pp.8-23.

Hassan, A.R., Bashar, S.K. and Bhuiyan, M.I.H., 2015, August. On the classification of sleep states by means of statistical and spectral features from single channel electroencephalogram. In *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 2238-2243). IEEE.

Hassan, A.R. and Bhuiyan, M.I.H., 2016. A decision support system for automatic sleep staging from EEG signals using tunable Q-factor wavelet transform and spectral features. *Journal of neuroscience methods*, *271*, pp.107-118.

Hassan, A.R. and Bhuiyan, M.I.H., 2016. Automatic sleep scoring using statistical features in the EMD domain and ensemble methods. *Biocybernetics and Biomedical Engineering*, *36*(1), pp.248-255.

Hassan, A.R. and Bhuiyan, M.I.H., 2016. Computer-aided sleep staging using complete ensemble empirical mode decomposition with adaptive noise and bootstrap aggregating. *Biomedical Signal Processing and Control*, *24*, pp.1-10.

Hermawan, I., Alvissalim, M.S., Tawakal, M.I. and Jatmiko, W., 2012, December. An integrated sleep stage classification device based on electrocardiograph signal. In *2012 International Conference on Advanced Computer Science and Information Systems (ICACSIS)* (pp. 37-41). IEEE.

Imtiaz, S.A. and Rodriguez-Villegas, E., 2014. A low computational cost algorithm for REM sleep detection using single channel EEG. *Annals of biomedical engineering*, *42*(11), pp.2344-2359.

Lazaro, A., Sarno, R., Andre, R.J. and Mahardika, M.N., 2017, October. Music tempo classification using audio spectrum centroid, audio spectrum flatness, and audio spectrum spread based on MPEG-7 audio features. In *2017 3rd International Conference on Science in Information Technology (ICSITech)*(pp. 41-46). IEEE.

Li, X., Cui, L., Tao, S., Chen, J., Zhang, X. and Zhang, G.Q., 2017. Hyclasss: A hybrid classifier for automatic

sleep stage scoring. *IEEE journal of biomedical and health informatics*, *22*(2), pp.375-385.

Liang, S.F., Kuo, C.E., Hu, Y.H. and Cheng, Y.S., 2012. A rule-based automatic sleep staging method. *Journal of neuroscience methods*, *205*(1), pp.169-176.

Liang, S.F., Kuo, C.E., Hu, Y.H., Pan, Y.H. and Wang, Y.H., 2012. Automatic stage scoring of single-channel sleep EEG by using multiscale entropy and autoregressive models. *IEEE Transactions on Instrumentation and Measurement*, *61*(6), pp.1649-1657.

Niedermeyer, E. and da Silva, F.L. eds., 2005. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

Özşen, S., 2013. Classification of sleep stages using class-dependent sequential feature selection and artificial neural network. *Neural Computing and Applications*, *23*(5), pp.1239-1250.

Pagel, J.F. and Pandi-Perumal, S.R. eds., 2014. *Primary Care Sleep Medicine: A Practical Guide*. Springer.

Phan, H., Andreotti, F., Cooray, N., Chén, O.Y. and De Vos, M., 2019. SeqSleepNet: End-to-End Hierarchical Recurrent Neural Network for Sequence-to-Sequence Automatic Sleep Staging. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

Pop-Jordanova, N. and Pop-Jordanov, J., 2005. Spectrum-weighted EEG frequency ("brain-rate") as a quantitative indicator of mental arousal. *Prilozi*, *26*(2), pp.35-42.

Quan, S.F., Howard, B.V., Iber, C., Kiley, J.P., Nieto, F.J., O'Connor, G.T., Rapoport, D.M., Redline, S., Robbins, J., Samet, J.M. and Wahl, P.W., 1997. The sleep heart health study: design, rationale, and methods. *Sleep*, *20*(12), pp.1077-1085.

Rechtschaffen, A. & Kales, A., 1968. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. *Washington DC: US National Institute of Health Publication*.

Şen, B., Peker, M., Çavuşoğlu, A. and Çelebi, F.V., 2014. A comparative study on classification of sleep stage based on EEG signals using feature selection and classification algorithms. *Journal of medical systems*, *38*(3), p.18.

Sohn, S.I., Kim, D.H., Lee, M.Y. and Cho, Y.W., 2012. The reliability and validity of the Korean version of the Pittsburgh Sleep Quality Index. *Sleep and Breathing*, *16*(3), pp.803-812.

Šušmáková, K. and Krakovská, A., 2008. Discrimination ability of individual measures used in sleep stages classification. *Artificial intelligence in medicine*, *44*(3), pp.261-277.

Vidaurre, C., Krämer, N., Blankertz, B. and Schlögl, A., 2009. Time domain parameters as a feature for EEG-based brain–computer interfaces. *Neural Networks*, *22*(9), pp.1313-1319.

Yan, R., Zhang, C., Spruyt, K., Wei, L., Wang, Z., Tian, L., Li, X., Ristaniemi, T., Zhang, J. and Cong, F., 2019. Multi-modality of polysomnography signals' fusion for automatic sleep scoring. *Biomedical Signal Processing and Control*, *49*, pp.14-23.