

Automated sleep staging based on multi-module neural network using simpler signal: Respiratory signal

Yinqing Que^a, Pengyi Jiang^b, Tianyi Zhang^a, Yunzhang Cheng^{*,a}

^aSchool of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, P.R. China; ^bSchool of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, P.R. China

ABSTRACT

Sleeping is a vital biological state which help maintaining the homeostasis of organisms of all biological lives. A full sleep can be divided into different repeating stages, rapid eye movement sleep (REM) stage, non-rapid eye movement sleep (NREM) stage 1-4. An effective sleep staging system can help patients improving their sleep quality. In the past, patients are required to wear polysomnogram (PSG) for the whole night to collect signals like electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG) for diagnosis. And the traditional sleep staging system use one or more signals above to predict a sleep stage. In this paper, we introduce a new sleep staging algorithm based on machine learning. Our model has two main inputs: patients' respiratory signal and their physical data, like age, gender, and weight. The strategy is to use two CNNs to extract features from raw respiratory signal in time domain and frequency domain, several Word2vec layers are built to extract features from patients' meta data and a transformer encoder to collect all the features. Using the MIT-BIH Polysomnographic database, our model achieves a result of 81.96% accuracy. This shows that it is completely feasible to classify patients' sleep stage with their respiratory signal and meta information.

Keywords: Sleep staging, respiratory, machine learning

1. INTRODUCTION

Sleep is an important biological activity to most creatures. With a sufficient sleep, living organisms have a chance to recover from the day's activities and maintain a normal physiological balance. While sleeping, several physiological changes occur. For example, an organism's sensory awareness and metabolic rate will decrease, muscle activities will be inhibited. As a result, the organism will produce less metabolic waste than they can be eliminated¹. Thus, sleeping is an important way for organisms to restore the normal function to various systems, which allows them for better performance during wakeful activities.

An effective and healthy sleep is closely related to the healthy life of human beings. Some people may suffer from poor sleep quality, and accurate classification of sleep status and a detailed analysis of sleep quality have become vital prerequisite for doctors to assist in improving sleep. According to scientists' research on the human sleep process, human sleep activity is an irregular biological state². Sleep can be divided into a series of repeated stages, the cycle varies from person to person, between 60 and 90 minutes. The series of stages can be divided into wakefulness, NREM sleep stage and REM sleep stage. NREM sleep can be further subdivided into four stages: I, II, III, and IV. This is the R&K standard which is widely used to classify sleep stages around the globe³.

*Yunzhang Cheng: cyz2008@usst.edu.cn; phone 021-55571660

One of the authoritative references on sleep scoring and staging is the AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. According to this manual, patients are required to wear PSG with the help of an experienced doctor when assessing sleep stage. Typically, there are three to six electrodes for collecting EEG signals, three electrodes for EMG signals and another two electrodes for EOG signals in addition to pulse oximeters, breathing detectors⁴. After the data are collected, an experienced doctor will complete the final sleep stage classification based on the EEG curve characteristics, combined with other supporting features, such as patients' physical performance, moving events, respiratory behaviors. Although PSG is currently the key standard for sleep staging, the PSG device is large and complicated to use. Patients are equipped with dozens of signal sensors on the body when they are sleeping. Their sleep quality will also be affected by the equipment, which introduces additional variables beyond themselves. Raw data must be subjected to long-term manual monitoring and analysis by doctors, resulting in potential errors and deviations from actual values during staging analysis.

The surge in the occurrence of sleep problems has spurred extensive research in sleep science. Sleep staging is a widely researched and multidisciplinary application that encompasses various fields, including medicine, sensor technology, and computer science. The traditional method of sleep staging involves manual marking by experienced doctors, which is a time-consuming and expensive process. Consequently, several research studies and efforts have been conducted to solve this problem. There are a variety of automatic sleep staging systems been developed, including those based on EEG⁵⁻⁷, those based on all PSG signals⁸⁻¹⁰, and those based on pre-feature extraction based on EEG and EOG signals¹¹

In this paper, we build an automatic sleep staging model using deep learning. We only take patients' respiratory signal collected via PSG and basic physical data as inputs. Firstly, a low-pass filter is implemented to reduce high-frequency noise from the original signal. Next, we extract features from patients' respiratory signal in time domain and frequency domain simultaneously using short-time Fourier transform, wavelet packet processing and convolutional neural network. Finally, we use a transformer encoder is used to combine the features extracted above with additional metadata from patients, including their gender, age, and weight. This model is a comprehensive application of convolutional neural network, transformer encoder, Word2vec to unveil hidden features within the signal data.

2. METHODOLOGY

Data

We use the respiratory signal collected via PSG from the public MIT-BIH Polysomnographic Database¹² for training and evaluation. The signal is sampled at 250Hz. After slicing the raw record into 30 seconds pieces, we achieve 15,575 slices, 3,893,750 data points in total. Before we start training, due to the imbalance of the original data distribute, we perform some data expansion and enhancement. After preprocessing, 85% of the dataset is selected as the training set, and the rest is used as the verification and test set.

Preprocessing

Because of some reasons when collecting the raw data, the signal is full of high-frequency noise. Firstly, we designed a Butterworth low-pass filter to remove the noise. The comparison is shown in figure 1.

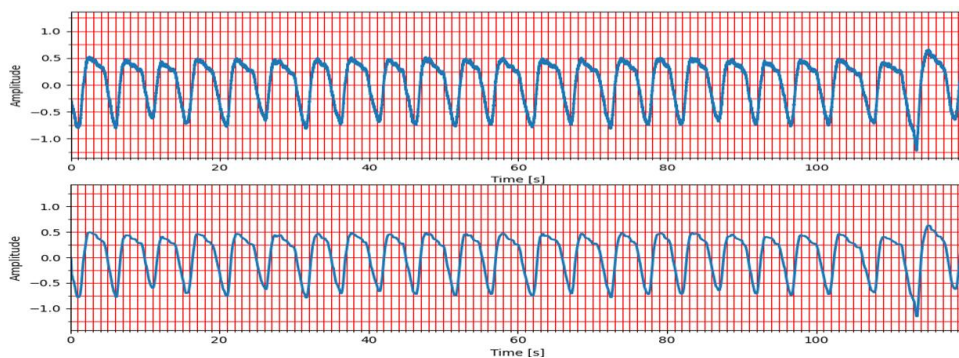


Figure 1. Signal before (up) filtering and after (down) filtering

After the signal is filtered, we divide the signal into 30 seconds part and save it with the label annotated by experienced doctor provided by the database.

Feature extraction

The short-time Fourier transform (STFT) is a method of extracting partial frequency features in a signal. Firstly, it uses a fixed window function to divide a long continuous time-domain signal into several shorter segments of the same length. Then, for each segment, a Fourier transform is performed separately. Finally, overlap these segments and generate a spectrogram of the entire signal.

Given a time-domain signal $x[n]$, a window $w[n]$, and a hop size H . The calculation process is as follows.

$$x_t[n] = x[n]w[n - tH] \quad (1)$$

$$X_t[n] = \frac{\sum_t x_t[n]w[n-tH]}{\sum_t w^2[n-tH]} \quad (2)$$

Continuous wavelet transform (CWT) is also a method of extracting and compressing information from the original time-domain signal. It uses several wavelet functions to convolve the original signal, decomposes the complex information carried by the original signal into several basic waves, and achieves the original signal. To achieve the purpose of analyze the signal in time-domain and frequency-domain simultaneously.

Given a mother wavelet which is a continuous function in both the time-domain and the frequency-domain $\psi(t)$ to generate child wavelet in the further calculation.

$$\psi(t) = \frac{2}{\sqrt{3\sigma\pi^4}} \left(1 - \left(\frac{t}{\sigma}\right)^2\right) e^{-\frac{t^2}{2\sigma^2}} \quad (3)$$

By shifting and translating the mother wavelet, and convolving with the original signal. We can get:

$$X_w(a, b) = \frac{1}{|a|^2} \int_{-\infty}^{\infty} x(t) \bar{\psi}\left(\frac{t-b}{a}\right) dt \quad (4)$$

We use STFT and CWT to extract features in frequency-domain from the original signal, and save them in a 224x224 color map as shown in figure 2.

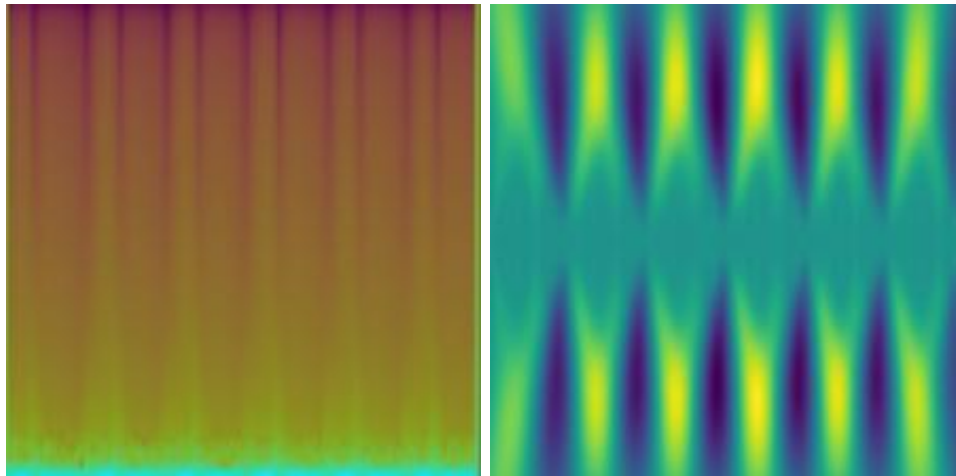


Figure 2. Signal features (left: STFT, right: CWT)

Model

In general, a person's respiratory signal is related to his physical status and basic information such as age, weight, and gender. Some traditional deep learning model may only take one feature into account, for example, time-domain signal series, while ignoring some other details. Inspired by the model introduced by Minh¹³. Our model can accept the time-domain signal, frequency-domain features, and other meta data simultaneously. There will be auto-encoder to mine

potential relationship between the features extracted separately by different model. The model design is shown in figure 3:

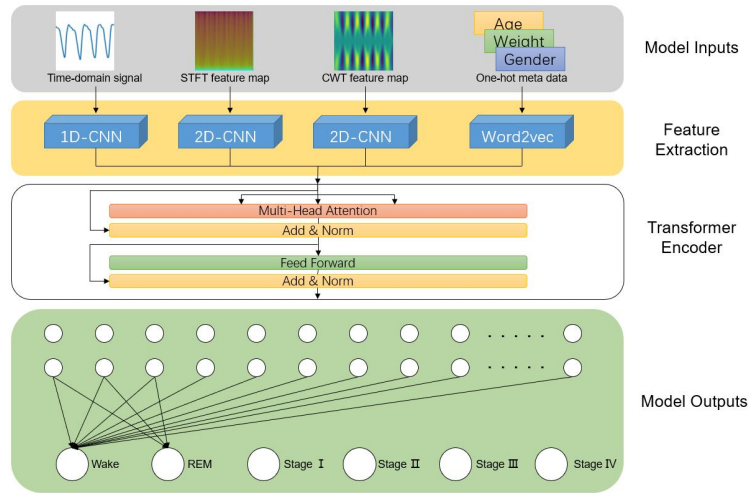


Figure 3. Model structure

Feature extraction module This module aims to extract feature vector from the inputs separately. We use 1D Resnet-34 to extract features from the time-domain signal and two independent 2D Resnet-34 to process the STFT feature map and the CWT feature map. For the meta data part, an auto encoder Word2vec is used to extract features. These four feature extraction components will generate four 64-dimensional features.

Transformer encoder module This module aims to generate encodings contained in the inputs and between the inputs. For each part of the input, the multi-head attention module will weigh the relevance from the four feature vectors and extract information from them. The feed-forward module will perform additional processing of the outputs from the multi-head attention module, such as normalization and residual connects.

Model output module This module aims to integrate the feature vector produced from the transformer encoder module and give a prediction. In this module, we use several full-connect layers and dropout layers, ReLU and SoftMax as the activation functions, to generate the result.

3. RESULTS

Datasets The MIT-BIH Polysomnographic dataset consists of 130 hours long respiratory signal records of 18 patients. The recordings are digitized at a sampling frequency of 250Hz. There are a total of 16 different labeled data in the database tag. In our experiments we choose 6 common labels related to sleep staging. The labels and their meaning are shown in the following table 1.

Table 1. Labels and their meanings

Label	Meaning
W	Wake
R	REM sleep
1	NREM sleep stage 1
2	NREM sleep stage 2
3	NREM sleep stage 3
4	NREM sleep stage 4

Before training, we found that there is a data imbalance in the original database. We use the following strategies to enhance the database data to get a better performance:

1: For two adjacent segments with the same label. Combining the two segments into a 1-minute-long segment and take the center 30-second as a new segment. The label remains unchanged.

2: For some types with fewer samples. Shift the window forward and backward for 5 seconds to resample new records. The label is consistent with the source sample.

After removing partially overlapping signal records, the data is divided into training set and test set at a ratio of 85:15.

Precision and recall matrixes represent how well the model performs.

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

F1-score represents the harmonic mean of the precision score and recall score.

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{7}$$

The experiment result is shown in figure 4.

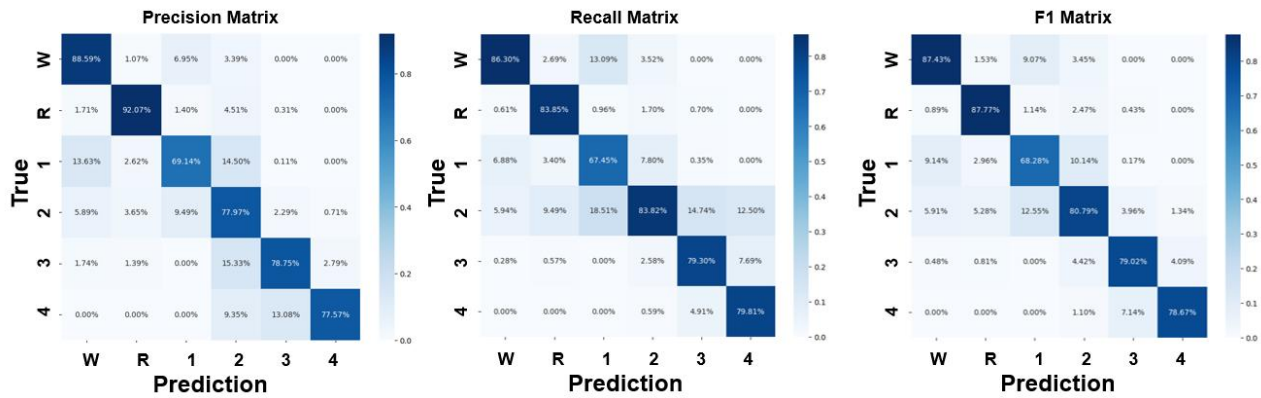


Figure 4. Model performance (6-categories)

To compare with other sleep staging models in table 2, we combine label 3 and label 4 as label 3. And the prediction result is shown below in figure 5.

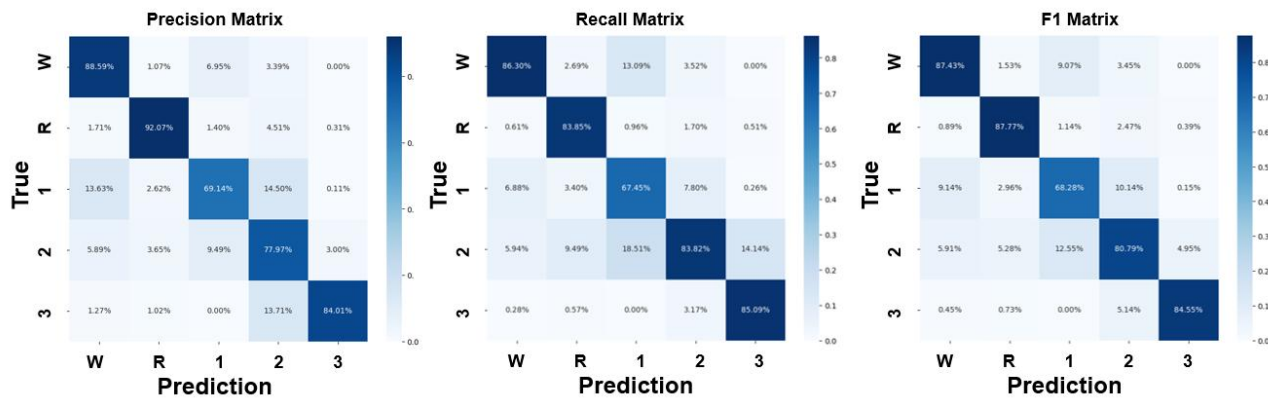


Figure 5. Model performance (5-categories)

Table 2. Performance comparison between our model with some other sleep staging model

	Inputs	Classes	Acc.
Ref. ⁸	6-channel PSG	5	87.50%
Ref. ⁵	F4-EOG(Left)	5	86.20%
Ref. ⁹	All PSG channels	5	85.20%
Ref. ⁶	FP2-EOG(Left)	5	83.35%
Ref. ⁵	FPz-Cz	5	82.00%
Ours	Respiratory Signal	5	81.96%
Ref. ⁸	2-channel PSG	5	81.90%
Ours	Respiratory Signal	6	81.56%
Ref. ⁵	Pz-Oz	5	79.80%
Ref. ¹⁰	EEG & EOG	5	79.00%
Ref. ¹⁴	FPz_Cz	5	74.80%

4. DISCUSSION AND CONCLUSION

In this paper, we assume that a person’s sleep depth can be reflected from respiratory signal. To prove this concept, we build a multi-module neural network and use the patient's respiratory signal data collected from PSG during sleep from the MIT-BIH Polysomnographic dataset. And our model achieved an accurate score of 81.96%. By deploying our automated sleep staging system, the characteristics of respiratory signal can be fully mined, and get a similar effect of the model using EEG, EOG, ECG and many other signals collected by PSG.

Our experiments demonstrate the feasibility of performing an automatic sleep staging using only a patient's respiratory signal. This model implies that in some non-emergency medical scenarios, we can complete sleep staging without PSG. This will eliminate the influence carried by the complicated monitoring equipment. Our automated sleep staging system shows great potential for use in future scenarios, such as home-based medical treatment. The model can be deployed in some IoT devices, and enables people to monitor their sleep at home without the help of the doctors. Moving forward, we believe that our automated sleep staging system will provide people more accessible, cost-effective, and personalized sleep analytics.

REFERENCES

- [1] Xie L., Kang H., Xu Q., Chen M. J., Liao Y., Thiyagarajan M., et al., Sleep drives metabolite clearance from the adult brain, *Science*, 342(6165), 373-377 (2013).
- [2] Agarwal, R., Gotman J., Computer-assisted sleep staging, *IEEE Transactions on Biomedical Engineering*, 48(12), 1412-1423 (2001).
- [3] Rechtschaffen A., A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects, *Archives of General Psychiatry*, 1-55 (1968).
- [4] Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Marcus, C., Vaughn, B. V., AASM manual for the scoring of sleep and associated events, Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine, 176, 2012 (2012).
- [5] 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*, 25(11), 1998-2008 (2017).
- [6] Dong, H., Supratak, A., Pan, W., Wu, C., Matthews, P. M., Guo, Y., Mixed neural network approach for temporal sleep stage classification, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(2), 324-333 (2017).

- [7] Mousavi, S., Afghah, F., Acharya, U. R., SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach, *PloS one*, 14(5), e0216456 (2019).
- [8] Biswal, S., Sun, H., Goparaju, B., Westover, M. B., Sun, J., Bianchi, M. T., Expert-level sleep scoring with deep neural networks, *Journal of the American Medical Informatics Association*, 25(15), 1643-1650 (2018).
- [9] Guillot, A., Sauvet, F., During, E. H., Thorey, V., Dreem open datasets: Multi-scored sleep datasets to compare human and automated sleep staging, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28(9), 1955-1965 (2020).
- [10] Perslev, M., Darkner, S., Kempfner, L., Nikolic, M., Jennum, P. J., Igel, C., U-Sleep: resilient high-frequency sleep staging, *NPJ digital medicine*, 4(1), 72 (2021).
- [11] Kuo, C. E., Chen, G. T., Automatic sleep staging based on a hybrid stacked LSTM neural network: verification using large-scale dataset, *IEEE access* 8, 111837-111849 (2020).
- [12] Ichimaru, Y., Moody, G., Development of the polysomnographic database on CD - ROM, *Psychiatry and clinical neurosciences*, 53(2), 175-177 (1999).
- [13] Le, M. D., Rathour, V. S., Truong, Q. S., Mai, Q., Brijesh P., Le N., Multi-module recurrent convolutional neural network with transformer encoder for ECG arrhythmia classification, 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), IEEE, pp.1-5 (2021).
- [14] Tsinalis, O., Matthews, P. M., Guo, Y., Zafeiriou, S., Automatic sleep stage scoring with single-channel EEG using convolutional neural networks, *arXiv preprint arXiv:161001683* (2016).