Automated identification of sleep stages from EEG signals using machine learning

Smayan Das , Jayant Choudhary and Abdulhamit Subasi

Abstract— Sleep stage classification plays an important role in the diagnosis of sleep-related diseases. Manual sleep stage scoring is a time-consuming task for sleep experts and is limited by inter-rater reliability. Therefore, we need an automated way for sleep stage scoring. Automated identification of sleep stages is a key step in the study of sleep disorders and sleep research. In this paper, we propose an automatic sleep stage annotation method using a single-channel EEG signal. MNE (a python library) was used for this purpose. Various Machine Learning models were applied on the data .The sleep stage classification methods will help in improving the study of sleep disorders and help the sleep specialists in accurate diagnosis.

I. Introduction

Sleep plays an important role in physical health and quality of life.Sleep diseases , such as Insomnia and Obstructive sleep apnea, may cause daytime sleepiness, depression , or even death [1]. Therefore , there is an urgent demand for an effective way to diagnose and cure sleep-related diseases. Sleep related diseases research, defined as sleep medicine, is already an important branch of medicine and has been involved in several clinical problems.

Sleep stage classification is the first step in the diagnosis of sleep-related diseases [2] [3]. The electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG) signals are widely used to diagnose the sleep disorders (e.g., sleep apnea, parasomnias, and hypersomnia). These signals are typically recorded by placing sensors on different parts of the patient's body. In an overnight polysomnography (PSG) (also called as sleep study), the EEG signal is usually the main collected signal being used to monitor the brain activities to diagnose different sleep disorders [4]and other common disorders such as epilepsy [5].

The most general categorization of sleep includes the stages of wakefulness, REM sleep and non-REM sleep. Apart from that, non-REM sleep is furtherly divided into 4 sub-stages. Stage 1 represents light sleep and moving towards stage 4 the sleep gets deeper. The scoring of 5 sleep stages is done by the proposal of a committee led by A. Rechtschaffen and A.Kales aiming at standardizing the process of recording and scoring sleep stages. Rechtschaffen and Kales scoring system suggests scoring sleep stages by epochs. The duration of the epoch is set at 20 or 30 seconds (nowadays 30 second epochs are used in most recordings). The duration of the epoch does not change during the whole recording. When two stages occur during one epoch, the one that takes up the largest portion of the epoch should be scored as the stage of the epoch.

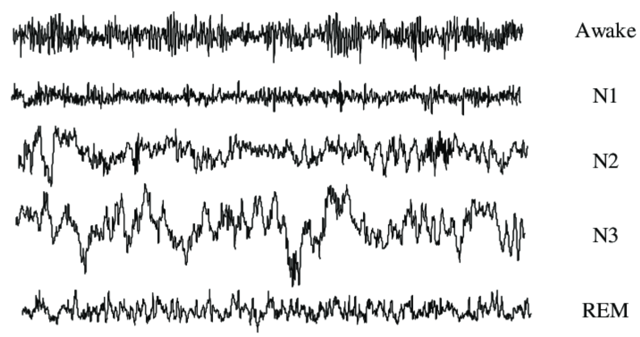

Fig. 1. EEG pattern of various sleep stages

II. Subjects and Data Collection

Data used in this paper to evaluate the performances of proposed system was obtained from the sleep European data format (EDF) [Expanded] database [6] which is publicly which is publicly available online on physionet website . The sleep-edf database contains 197 whole-night PolySomnoGraphic sleep recordings, containing EEG, EOG, chin EMG, and event markers. Some records also contain respiration and body temperature. Corresponding hypnograms (sleep patterns) were manually scored by well-trained technicians according to the Rechtschaffen and Kales manual. The ***PSG.edf** files are whole-night polysmnographic sleep recordings containing EEG (from Fpz-Cz and Pz-Oz electrode locations), EOG (horizontal), submental chin EMG, and an event marker. The **SC*PSG.edf** files often also contain oro-nasal respiration and rectal body temperature. The ***Hypnogram.edf** files contain annotations of the sleep patterns that correspond to the PSGs. These patterns (hypnograms) consist of sleep stages W, R, 1, 2, 3, 4, M (Movement time) and ? (not scored). All hypnograms were manually scored by well-trained technicians (identified by the eighth letter of the hypnogram filename) according to the 1968 Rechtschaffen and Kales manual , but based on Fpz-Cz/Pz-Oz EEGs instead of C4-A1/C3-A2 EEGs. All EDF header fields also comply with the EDF+ specs, and unrecorded signals were removed from the ST*PSG.edf files.

A. Sleep Cassette Study and Data

The 153 **SC*** files (SC = Sleep Cassette) were obtained in a 1987-1991 study of age effects on sleep in healthy Caucasians aged 25-101,

.
[1] Kang et al., "A State Space and Density Estimation Framework for Sleep Staging in Obstructive Sleep Apnea."
[2] Mendonça et al., "A Portable Wireless Device Based on Oximetry for Sleep Apnea Detection."
[3] Cheung et al., "Screening for Obstructive Sleep Apnea in the Assessment of Coronary Risk."
[4] Biswal et al., "SLEEPNET."
[5] "(PDF) Deep Convolutional Neural Network for the Automated Detection and Diagnosis of Seizure Using EEG Signals."
[6] Kemp et al., "The Sleep-EDF Database [Expanded]."

without any sleep-related medication . Two PSGs of about 20 hours each were recorded during two subsequent day-night periods at the subjects homes. The EOG and EEG signals were each sampled at 100 Hz. The submental-EMG signal was electronically highpass filtered, rectified and low-pass filtered after which the resulting EMG envelope expressed in uV rms (root-mean-square) was sampled at 1Hz. Oro-nasal airflow, rectal body temperature and the event marker were also sampled at 1Hz.

B. *Sleep Telemetry Study and Data*

The 44 **ST*** files (ST = Sleep Telemetry) were obtained in a 1994 study of temazepam effects on sleep in 22 Caucasian males and females without other medication. Subjects had mild difficulty falling asleep but were otherwise healthy. The PSGs of about 9 hours were recorded in the hospital during two nights, one of which was after temazepam intake, and the other of which was after placebo intake. EOG, EMG and EEG signals were sampled at 100 Hz, and the event marker at 1 Hz.

III. Methodology

For doing this project we used a brilliant library called **MNE-python** [7]which is a Open-source Python software for exploring, visualizing, and analyzing human neurophysiological data: MEG, EEG and more.

A. *Loading the data*

In the first step the PSG and EDF data are read to create a raw object. Combining the PSG and EDF data in a object then we extract events based on the descriptions of the annotations to obtain the epochs.

B. *Feature extraction*

Observing the power spectral density (PSD) plot of the epochs grouped by sleeping stage we can see that different sleep stages have different signatures. EEG features are created based on relative power in specific frequency bands to capture this difference between the sleep stages in our data. Then we create a function to extract EEG features based on relative power in specific frequency bands to be able to predict sleep stages from EEG signals.

[7] "Installation — Contents — MNE 0.21.Dev0 Documentation."

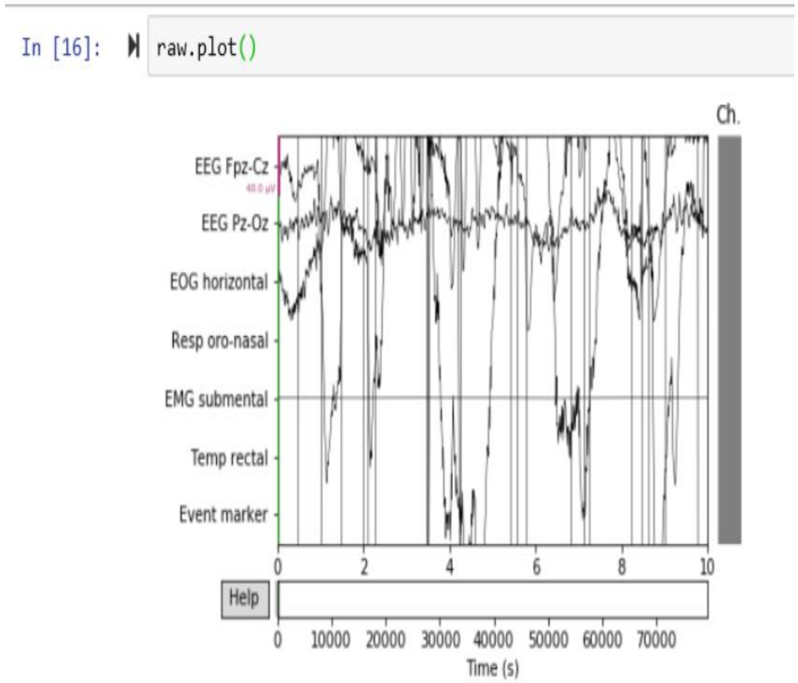

Fig.2. Different channels of EEG signal

IV. Experimental results

Claasification performance was compared using various Machine Learning models. The dataset used in the models is same to get a better idea that which model is most accurate. The obtained accuracies are as shown in the figures below –

A. *XGBoost*

Training accuracy score obtained for this model is 1 and test accuracy was 0.8960. Training accuracy obtained from this model is pretty good but the test accuracy is not that good.

Sleep stages	Precision	Recall	F1-score
Sleep stage W	0.98	0.98	0.98
Sleep stage 1	0.15	0.31	0.20
Sleep stage 2	0.76	0.91	0.83
Sleep stage 3/4	0.96	0.73	0.83
Sleep stage R	0.69	0.5	0.58
Macro Average	0.71	0.69	0.69
Weighted Average	0.91	0.90	0.90

Fig.3. Results obtained from XGBoost

B. SVM

Training accuracy score obtained for this model is 0.930 and test accuracy was 0.904. In this case both training and testing accuracy obtained are good and hence we can say that this model works good and can classify sleep stages much accurately.

Sleep Stages	Precision	Recall	F1-score
Sleep stage W	0.97	0.98	0.98
Sleep stage 1	0.00	0.00	0.00
Sleep stage 2	0.73	0.89	0.80
Sleep stage 3/4	0.95	0.76	0.84
Sleep stage R	0.64	0.72	0.68
Macro Average	0.66	0.67	0.66
Weighted Average	0.89	0.90	0.90

Fig.4. Results obtained from SVM

C. Random Forest

Training accuracy score obtained for this model is 1 and test accuracy was 0.905. The training accuracy obtained is very high as compared to the test accuracy.

Sleep Stages	Precision	Recall	F1-score
Sleep stage W	0.98	0.98	0.98
Sleep stage 1	0.20	0.32	0.25
Sleep stage 2	0.80	0.89	0.85
Sleep stage 3/4	0.95	0.76	0.84
Sleep stage R	0.68	0.62	0.65
Macro Average	0.72	0.71	0.71
Weighted Average	0.92	0.91	0.91

Fig.5. Results obtained from Random Forest

D. ANN

Training accuracy score obtained for this model is 0.901 and test accuracy was 0.864 . In this model both training and test accuracy obtained are low . Hence this is not the most accurate model.

Sleep Stages	Precision	Recall	F1-score
Sleep stage W	0.92	0.97	0.94
Sleep stage 1	0.00	0.00	0.00
Sleep stage 2	0.71	0.79	0.75
Sleep stage 3/4	0.91	0.79	0.85
Sleep stage R	0.52	0.42	0.47
Macro Average	0.61	0.59	0.60
Weighted Average	0.84	0.86	0.85

Fig.6. Results obtained from ANN

E. KNN

Training accuracy score obtained for this model is 0.947 and test accuracy was 0.8568 .

Sleep Stages	Precision	Recall	F1-score
Sleep stage W	0.9	1.00	0.95
Sleep stage 1	0.12	0.03	0.04
Sleep stage 2	0.82	0.73	0.77
Sleep stage 3/4	0.57	0.90	0.70
Sleep stage R	0.76	0.26	0.39
Macro Average	0.63	0.58	0.57
Weighted Average	0.83	0.86	0.83

Fig.7. Results obtained from KNN

F. Bagging

Training accuracy score obtained for this model is 0.9985 and test accuracy was 0.8204 .

Sleep Stages	Precision	Recall	F1-score
Sleep stage W	0.87	1.00	0.93

Sleep stage 1	0.05	0.08	0.06
Sleep stage 2	0.91	0.54	0.68
Sleep stage 3/4	0.76	0.94	0.84
Sleep stage R	0.70	0.21	0.32
Macro Average	0.66	0.55	0.57
Weighted Average	0.83	0.82	0.81

Fig.8. Results obtained from Bagging

G. Adaboost

Training accuracy score obtained for this model is 0.913 and test accuracy was 0.842 .

Sleep Stages	Precision	Recall	F1-score
Sleep stage W	0.98	0.98	0.98
Sleep stage 1	0.20	0.32	0.25
Sleep stage 2	0.80	0.89	0.85
Sleep stage 3/4	0.95	0.76	0.84
Sleep stage R	0.68	0.62	0.65
Macro Average	0.72	0.71	0.71
Weighted Average	0.92	0.91	0.91

Fig.9. Results obtained from Adaboost

H. Discussion

This paper proposes a method for automatic classification of five different sleep stages, namely, wakefulness, NREM1, NREM2, SWS, and REM, using single-channel classification. For that reason, proposed system can be considered as promising tool for sleep monitoring.

V. Conclusion

In this paper, automated system for sleep stage classification system using only one channel EEG is proposed. MNE python was used for this purpose. The Kohen Kappa score obtained in various methods – 0.7 for Adaboost , 0.617 for Bagging , 0.698 for KNN , 0.727 for ANN , 0.82 for Random Forest , 0.814 for SVM, 0.801 for XGBoost. Since manual classification of sleep stages is very time consuming so the method proposed in this paper could be very useful. The proposed method can be very useful as hospitals or sleep clinics equipped with automatic sleep scorers could faster and more easily diagnose sleep related issues to patients, since a human scorer (doctor) would no longer be necessary.