Heuristic Approximation of the MAP Estimator for Automatic Two-channel Sleep Staging

Shirin Riazy[1], Tilo Wendler[1], Jürgen Pilz[2], M. Glos[3] and T. Penzel[3]
[1]*Hochschule für Technik und Wirtschaft Berlin, Treskowallee 8, 10318 Berlin, Germany*
[2]*Alpen-Adria Universität Klagenfurt, Universitätsstraße 65-67, 9020 Klagenfurt am Wörthersee, Austria*
[3]*Charité - Universitätsmedizin Berlin, 10117 Berlin, Germany*

Keywords: Automatic Sleep Staging, Two-channel Measurement, Bayesian Statistics, Hidden Markov Model, MAP.

Abstract: In this paper, we shall introduce an algorithm that classifies EEG data into five sleep stages, relying only on two-channel sleep measurements. The sleep of a patient (divided into intervals of 30 seconds) is assumed to be a Markov chain on the five-element state space of sleep stages and our aim is to compute the most probable chain of this hidden Markov model by a maximum a posteriori (MAP) estimation in the Bayesian framework. Both the prior distribution of the chains and the likelihood model have to be trained on manual classifications made by professionals. For this purpose, the data is first preprocessed by a Fourier transform, a log transform and a principal component analysis for dimensionality reduction. Since the number of possible chains is immense (roughly 10^{335}), a heuristic approach for the computation of the MAP estimator is introduced, that systematically discards unlikely chains. The sleep stage classification is then compared to the classification of a professional, who scores according to the AASM and uses a full polysomnography. The overall structure of the hypnogram can adequately be reconstructed with error rates around 25%.

1 INTRODUCTION

In the context of a research project of the Hochschule für Technik und Wirtschaft Berlin, the Alpen-Adria Universität Klagenfurt and the Charité Universitätsmedizin Berlin, the authors have investigated on a mathematical model for the automatic classification of sleep stages using very little measurements. The measurement technique used for said classification is the Electroencephalography (EEG), a method for the measurement of differences of electrical potential on the surface of the head. This method has been widely used for the diagnosis of diseases of the central nervous system, as well as sleep processes. Using EEG data as one of several parameters in a full polysomnography, one can diagnose pathological states of dormancy, various forms of insomnia as well as dysfunctions of the circadian rhythm. To do so, sleep is classified into sleep stages in order to monitor the course of the sleep throughout the night.

For the classification of sleep stages, the EEG data has mostly been used for the visually distinctive signal structures, ranging in a certain frequency domain, which can be measured in certain areas of the head, usually in frontal, central and occipital positions. The classification procedure is done mostly by hand, with the help of a polysomnographic software. To simplify this process, we would like to propose an algorithm for the automatic sleep-staging using only two-channel measurements as input data.

In our case, the positions A_1 and A_2, typically measured at or right behind the ears at the so-called *mastoids* shall be the only input for the classification in contrast to a full polysomnography.

1.1 Automatic Two-channel Sleep Staging

The term sleep staging describes the classification of 30-second intervals of sleep (so-called *epochs*) into different sleep stages. The method for assigning a sleep stage to the measurement of an epoch was first systematized by Rechtschaffen and Kales (Kales et al., 1968) and has since evoked generations of sleep-staging manuals that contain detailed instructions as to how measurements should be adequately matched with a sleep stage. In these manuals, the percentage of certain frequencies, appearances of certain signal structures and the location of said signal structures play a vital role in the sleep-staging process.

Riazy S., Wendler T., Pilz J., Glos M. and Penzel T.
Heuristic Approximation of the MAP Estimator for Automatic Two-channel Sleep Staging.
DOI: 10.5220/0006242802360241
In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2017)*, pages 236-241
ISBN: 978-989-758-212-7

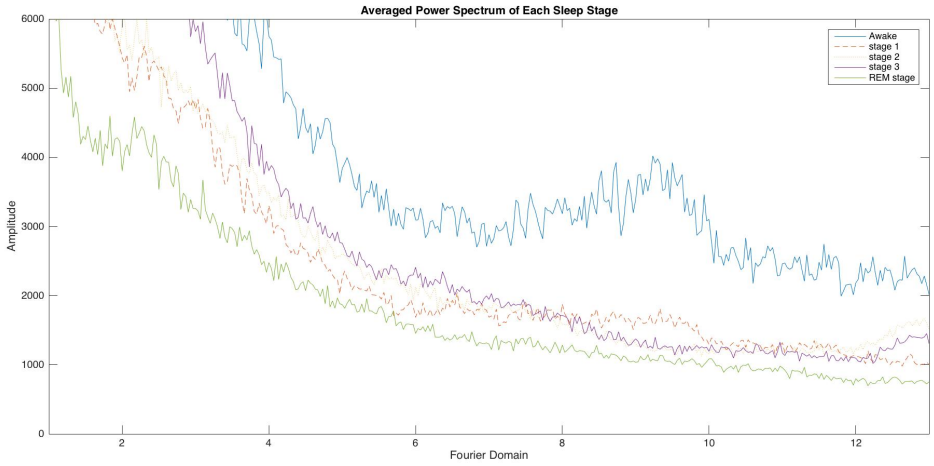

Figure 1: Averaged power spectrum of each sleep stage with the frequencies given in Hz.

The American Academy of Sleep Medicine's manual (Berry et al., 2012) differentiates between five stages of sleep: Awake, Stage 1, Stage 2, Stage 3, REM sleep, where the stages of light sleep (Stage 1 and Stage 2) are sometimes called NREM 1 and NREM 2 (for Non-REM sleep 1 and 2) and the deep sleep stage, Stage 3, is sometimes called NREM 3 (for Non-REM sleep 3).

A measurement of a full night of sleep results in approximately 960 epochs (8 hours) per patient. Further, as recommended by AASM (Berry et al., 2012), professional scorers rely on a full polysomnography, which consists of several EEG nodes, EOG derivations, a Chin EMG and several other parameters for an optimal rating of the sleep of the patient. The performance of these measurements is not only painstaking and expensive, but can also have a strong influence on the sleep of the patient, leading to distorted data.

As the classification of sleep stages via professionals tends to be tedious, there have been various approaches to automatize the procedure (Tagliazucchi et al., 2012; Redmond and Heneghan, 2003; Liang et al., 2012; Pan et al., 2012; Wang et al., 2015; Anderer et al., 2010; Gudmundsson et al., 2005), using methods such as SVMs, neural networks or Hidden Markov Models. Mostly, full polysomnographic data was used and reliabilities ranging between 70% and 90% were achieved (Penzel and Conradt, 2000); however, there has also been automatic single-channel or two-channel sleep staging (Zibrandtsen et al., 2016; Wang et al., 2015; Koley and Dey, 2012). In (Koley and Dey, 2012), the measurements of the electrodes at the mastoid (often referred to as A_1 and A_2 or M_1 and M_2) were used for the classification into light sleep, deep sleep and REM sleep, thus offering proof that measurements at the mastoid yield enough reliable information for the classification into three sleep stages. In this paper, we would like to take it one step further by classifying into five different sleep stages, as is usual in sleep medicine. This task proves to be highly difficult, since the transitional stage 1 is often uncertainly classified and can easily be mistaken for the sleep stages 2 and Awake.

1.2 Overview and Notation

Our data set consists of A_1- and A_2- measurements over 8 hours of sleep of several persons with a sampling rate of 256Hz. This data is divided into blocks of 30 seconds (so-called epochs), that contain $256 \cdot 30 = 7680$ measurements each for A_1 and A_2:

$$\mathcal{E}^k = \left\{E_1^{A_k}, \ldots, E_T^{A_k}\right\} \subseteq \mathbb{R}^{7680}, \quad T \approx 960,\ k = 1,2.$$

For each patient, we divide the set of epochs into a training set and a test set

$$\begin{aligned}
\mathcal{E} &= \mathcal{E}_{\text{train}} \cup \mathcal{E}_{\text{test}}, \\
\mathcal{E}_{\text{train}} &= \bigcup_{k=1,2} (E_1^{A_k}, \ldots, E_{T_{\text{train}}}^{A_k}), \\
\mathcal{E}_{\text{test}} &= \bigcup_{k=1,2} (E_{T_{\text{train}}+1}^{A_k}, \ldots, E_T^{A_k}),
\end{aligned}$$

In our case, we divided the 8-hour-measurement into two halves and used the first half as the training set (which equates to $T_{\text{test}} \approx 480$) and the second half as the test set. This choice shall be improved in the next stages of this project, as it is known that REM sleep typically occurs more often in the second half of the night. The sequence of sleep stages corresponding to these epochs will be denoted by random variables

$$C = (\underbrace{C_1, \ldots, C_{T_{\text{train}}}}_{=: C_{\text{train}}}, \underbrace{C_{T_{\text{train}}+1}, \ldots, C_T}_{=: C_{\text{test}}}) \in \mathcal{S}^T,$$

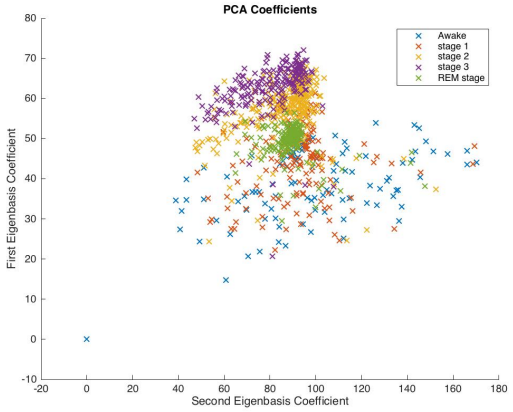

Figure 2: PCA Coefficients of the first two dominant eigenvectors of measurements of different sleep stages.

where the state space $\mathcal{S}$ consists of the five sleep stages:

$$\mathcal{S} = \{\text{'Awake', 'Stage 1', 'Stage 2', 'Stage 3', 'REM'}\}.$$

We will assume its classification by a professional scorer (who had access to the full polysomnography),

$$(C_{\text{train}}, C_{\text{test}}) \stackrel{!}{=} (c^{\text{prof}}_{\text{train}}, c^{\text{prof}}_{\text{test}})$$

to be correct. Our aim is to implement an algorithm $\mathcal{A}$, that classifies the test set $\mathcal{E}_{\text{test}}$ using only the professional classification $c^{\text{prof}}_{\text{train}}$ of the training set $\mathcal{E}_{\text{train}}$,

$$\mathcal{A} : (\mathcal{E}_{\text{train}}, \mathcal{E}_{\text{test}}, c^{\text{prof}}_{\text{train}}) \mapsto c^{\text{algo}}_{\text{test}} \in \mathcal{S}^{T-T_{\text{train}}},$$

such that $c^{\text{algo}}_{\text{test}} = c^{\text{prof}}_{\text{test}}$ in many components with high probability.

2 THE ALGORITHM

2.1 Preprocessing

The frequencies in a measurement play a central role in the classification of sleep stages (Berry et al., 2012; Penzel and Conradt, 2000). Signals of certain frequencies indicate conscious behaviour (such as the highly-frequent beta waves) or unconsciousness (such as delta waves). Therefore, it is meaningful to utilize the power spectrum of a measurement. As it can be seen in Figure 1, the average power spectra of different sleep stages behave visibly different, except for sleep stages one and two, which are not easily distinguishable. Furthermore, the power spectra of measurements $E^{A_k}_i \in \mathbb{R}^{7680}$ of epochs might be easier to compare, since they are invariant under shifts in time.

Additionally, it is known that certain confounding measurements may hinder the classification of sleep stages. These signals, which are called *artifacts*, can be caused for example by sweat on the skin of the proband or movements of body parts, eyes etc.. However, it is also known that these artifacts typically evoke signals that have different frequencies than the signals, which are used for the classification of sleep stages. Through a low pass filter, we ignore all the frequencies above a fixed threshold (approximately 66 Hertz). We do so in the following manner:

$$E^{A_k}_t \in \mathbb{R}^{7680} \mapsto F^{A_k}_t = |\text{FFT}(E^{A_k}_t)| \in \mathbb{R}^{7680}.$$

and

$$F^{A_k}_t \mapsto \tilde{F}^{A_k}_t = \text{proj}(F^{A_k}_t) \in \mathbb{R}^{2000},$$

where $\text{proj}(\cdot)$ is the projection of a vector onto its first 2000 coordinates. We assume this to be adequate, as the signals used for professional scoring range in the same frequency interval and it is known that artifacts created by bodily movement typically involve high-frequent signals.

We would like to remark that we have purposely not filtered out all of the frequencies that are not used in traditional sleep-staging in an attempt to retain as much information as possible while still removing artifacts.

To further reduce the dimension of the data and to make the data usable for classification, we used a principal component analysis (PCA), where we projected the data onto the space spanned by the 15 dominant eigenvectors of the covariance matrix. As it can be seen in Figure 2, the procedure retains a considerable amount of information from the measurements and the differences of the sleep stages can be seen even when using only the first two dominant eigenvectors. A scree plot was used to determine the optimal amount of eigenspaces, which was 15 in our case.

To avoid an a priori bias towards the lower frequencies, the Fourier transformed data is logarithmized before applying the PCA. This makes sense as higher frequencies tend to have smaller amplitudes. Let

$$M^k_t = \log \tilde{F}^{A_k}_t$$

be the logarithmized projection of the power spectrum for a $t \in \{1, \dots, T\}$. We used the data matrix (M^1_t, M^2_t) for the PCA

$$(\tilde{F}^{A_k}_t, \tilde{F}^{A_k}_t) \stackrel{\log}{\longmapsto} (M^1_t, M^2_t) \stackrel{\text{PCA}}{\longmapsto} m_t \in \mathbb{R}^{15},$$

where m_t is the projected data resulting from the PCA. This leads to the following overall preprocessing steps

$$\begin{aligned} &(\mathbb{R}^{7680})^2 \longrightarrow (\mathbb{R}^{2000})^2 \longrightarrow (\mathbb{R}^{2000})^2 \longrightarrow \mathbb{R}^{15} \\ &(E^{A_k}_t)_k \underset{\text{proj}}{\overset{\text{FFT}}{\longmapsto}} (\tilde{F}^{A_k}_t)_k \stackrel{\log}{\longmapsto} (M^k_t)_k \stackrel{\text{PCA}}{\longmapsto} m_t. \end{aligned}$$

We then define

$$X_{\text{train}} := \{m_1, \ldots, m_{T_{\text{train}}}\}$$
$$X_{\text{test}} := \{m_{T_{\text{train}}+1}, \ldots, m_T\}$$

2.2 Hidden Markov Model and the MAP Estimator

We assume the sequence of sleep stages $C = (C_1, ..., C_T) \in \mathcal{S}^T$ introduced in Section 1.2 to be a Markov chain with transition matrix $\mathcal{P} \in \mathbb{R}^{5\times 5}$, which will be approximated by the empirical probabilities of all possible transitions

$$\mathcal{P}_{s,s'} = \frac{\#\{x_{t+1} = s,\ x_t = s' \mid t = 1, \ldots, T_{\text{train}} - 1\}}{\#\{x_t = s' \mid t = 1, \ldots, T_{\text{train}} - 1\}}$$

for $s, s' \in \mathcal{S}$ (it is thereby not considered part of the inference process). Accordingly, the invariant measure π of the Markov chain will be approximated by the empirical probabilities

$$\pi(s) = \frac{\#\{x_\tau = s \mid \tau = 1, \ldots, T_{\text{train}}\}}{T_{\text{train}}}, \quad s \in \mathcal{S}.$$

For the inference process, our prior konwledge $\mathbb{P}(C_{\text{test}})$ about C_{test} is therefore modeled by Markov chain of length $T_{\text{test}} = T - T_{\text{train}}$ with $\mathbb{P}(C_t) = \pi$ for each $t = T_{\text{train}} + 1, \ldots, T$ and transition matrix $\mathcal{P}$.

Bayes' rule is the proper tool to update our knowledge about C_{test} given the measurements X_{test}:

$$\mathbb{P}(C_{\text{test}} \mid X_{\text{test}}) = \frac{\mathbb{P}(X_{\text{test}} \mid C_{\text{test}})\,\mathbb{P}(C_{\text{test}})}{\mathbb{P}(X_{\text{test}})}.$$

We will mainly focus on estimating the maximum a posteriori (MAP) estimator of C_{test},

$$\begin{aligned} c_{\text{test}}^{\text{MAP}} &= \arg\max_c \mathbb{P}(C_{\text{test}} = c \mid X_{\text{test}}) \\ &= \arg\max_c \mathbb{P}(X_{\text{test}} \mid C_{\text{test}} = c)\,\mathbb{P}(C_{\text{test}} = c). \end{aligned}$$

Since each measurement X_t is assumed to depend only on C_t, we assume that all measurements

$$X_t \mid C_{\text{test}} = X_t \mid C_t, \quad t = T_{\text{train}} + 1, \ldots, T,$$

are independent. Therefore, the likelihood model $\mathbb{P}(X_{\text{test}} \mid C_{\text{test}})$ is be given by the product of the individual likelihoods $\mathbb{P}(X_t \mid C_t)$,

$$\mathbb{P}(X_{\text{test}} \mid C_{\text{test}}) = \prod_{t=T_{\text{train}}+1}^{T} \mathbb{P}(X_t \mid C_t)$$

the latter being trained by a pre-installed MATLAB classifier. Theoretically, we now have all the necessary ingredients to compute the MAP estimator.

2.3 Heuristic Approximation of the MAP Estimator

Though we are now able to compute the probability of each realization $C_{\text{test}} = c \in \mathcal{S}^{T_{\text{test}}}$ given the measurements X_{test}, the straightforward approach of computing all these probabilities is impractical due to the huge number of possible chains,

$$\left|\mathcal{S}^{T_{\text{test}}}\right| = 5^{T_{\text{test}}} \approx 3 \cdot 10^{335}.$$

Therefore, a heuristic approach is necessary in order to compute these probabilities $\mathbb{P}(C_{\text{test}} = c \mid X_{\text{test}})$ only for "realistic" or "probable" chains. Since we are only interested in the MAP estimator $c_{\text{test}}^{\text{MAP}}$ and not in the whole probability distribution of $C_{\text{test}}|X_{\text{test}}$, this restriction is meaningful, however, it will only yield an approximation of $c_{\text{test}}^{\text{MAP}}$.

For the heuristic approach let us first introduce the notation

$$\begin{aligned} C^t &:= (C_{t_{\text{Train}}+1}, \ldots, C_{t_{\text{Train}}+t}), \\ X^t &:= (X_{t_{\text{Train}}+1}, \ldots, X_{t_{\text{Train}}+t}), \\ \pi_t(c^t) &:= \mathbb{P}(X^t \mid C^t = c^t)\,\mathbb{P}(C^t = c^t), \end{aligned}$$

the latter being the value we want to maximize for $t = T_{\text{test}}$.

We will proceed as follows:

1. Compute all possible ($5^4 = 625$) values of π_4.
2. For $t = 5, \ldots, T_{\text{test}}$, iterate
   (a) Restrict the set of considered subchains $(c_1, \ldots, c_{t-1})$ to 125 subchains with the highest values of π_{t-1}.
   (b) Compute π_t for all possible subchains $(c_1, \ldots, c_t)$ that contain the above subchains ($5 \cdot 125 = 625$ values).

This way, chains with unlikely subchains are discarded successively and we end up with 625 very probable (but possibly not most probable) chains. Step (b) in the above iteration can be performed very efficiently via the formula

$$\mathbb{P}(C^t \mid X^t) \ \propto\ \mathbb{P}(C^{t-1} \mid X^{t-1})\,\mathbb{P}(C_t \mid C_{t-1})\,\mathbb{P}(X^t \mid C^t).$$

Through this heuristic, the number of evaluations of this formula is reduced to 125 in each timestep.

3 RESULTS

The method was tested on five healthy subjects at the sleep laboratory of the Advanced Sleep Research GmbH. Each patient's sleep was measured with a full polysomnography for a duration of approximately 8

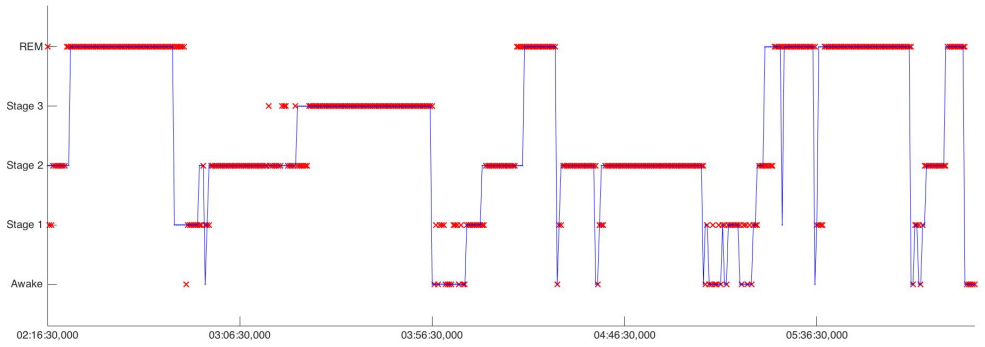

Figure 3: Sleep stage classification of the second half of the measurement. Comparison of the MAP estimator method (blue) and the classification of a professional scorer (red).

Table 1: Confusion matrix of patient 3, where the classifications of a professional scorer are compared to the output of the algorithm.

		Classification				
		Awake	Stage 1	Stage 2	Stage 3	REM
Stages	Awake	27	2	0	0	0
	Stage 1	17	31	11	0	3
	Stage 2	2	1	145	6	6
	Stage 3	1	0	5	64	0
	REM	3	7	7	0	145

hours. As stated before, the first half of the night was used as training data, whereas the second half of the night was used to test the algorithm. The hypnogram that was used as the "ground truth" to train and verify the algorithm was scored by a professional somnologist, who used a full polysomnography.

The error rates for the five subjects were 14%, 24%, 34%, 22% and 31%, respectively. Thus, they are ranging between 14%-34% with an average of 25%.

As one can see in Figure 3, the overall structure of the classification of the algorithm is very close to the scoring of a professional, who had the advantange of using a full polysomnography.

The evaluation of the confusion matrices showed that the classification of the Awake stage, as well as stages 2 and 3 are fairly adequate. Stage 1 is often mistaken for the Awake stage or stage 2 and REM stage is sometimes mistaken for one of the light sleep stages, as it can be seen in Table 1.

4 DISCUSSION

Summing things up, we can conclude that the measurements at A_1 and A_2 contain relevant information for automatic sleep-scoring. This can be viewed as a proof of concept, as these positions are usually only used as reference electrodes. Beyond the recognition of wakefulness and sleep, the MAP estimator offers differentiated classifications.

It appears that, in contrast to earlier methods used for the classification of sleep stages, the method presented in this paper seems to prefer continuous structures. Especially in phases of transition between stages, where professional scorers seem to be uncertain, the MAP estimation method prefers a single transition in contrast to the erratic behaviour of the professional scorer.

It could be argued that such an undecisive phase between the stages should be flattened with respect to the preceding and the following sleep stage. In this case, it is unclear, whether classification of the automatic method or the scorer is preferable.

Usually, patients are requested to spend two nights at a sleep laboratory, in order to minimize confounding effects. A minimized measurement system with the patient-specific algorithm at hand could replace the second measurement and thus halve the costs for sleep-related diagnosis at worst. At best, the algorithm could learn to transfer the classification from one person to another.

ACKNOWLEDGEMENTS

We would like to thank the Charité Universitätsmedizin Berlin as well as the Advanced Sleep Research GmbH for providing the data for testing the method.

REFERENCES

Anderer, P., Moreau, A., Woertz, M., Ross, M., Gruber, G., Parapatics, S., Loretz, E., Heller, E., Schmidt, A., Boeck, M., et al. (2010). Computer-assisted sleep classification according to the standard of the american academy of sleep medicine: validation study of the aasm version of the somnolyzer 24× 7. *Neuropsychobiology*, 62(4):250–264.

Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Marcus, C., and Vaughn, B. (2012). The aasm manual for the scoring of sleep and associated events. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine*.

Gudmundsson, S., Runarsson, T. P., and Sigurdsson, S. (2005). Automatic sleep staging using support vector machines with posterior probability estimates. In *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, volume 2, pages 366–372. IEEE.

Kales, A., Rechtschaffen, A., University of California, L. A., and (U.S.), N. N. I. N. (1968). *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Allan Rechtschaffen and Anthony Kales, editors*. U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network Bethesda, Md.

Koley, B. and Dey, D. (2012). An ensemble system for automatic sleep stage classification using single channel eeg signal. *Computers in biology and medicine*, 42(12):1186–1195.

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):169–176.

Pan, S.-T., Kuo, C.-E., Zeng, J.-H., and Liang, S.-F. (2012). A transition-constrained discrete hidden markov model for automatic sleep staging. *Biomedical engineering online*, 11(1):1.

Penzel, T. and Conradt, R. (2000). Computer based sleep recording and analysis. *Sleep medicine reviews*, 4(2):131–148.

Redmond, S. and Heneghan, C. (2003). Electrocardiogram-based automatic sleep staging in sleep disordered breathing. In *Computers in Cardiology, 2003*, pages 609–612. IEEE.

Tagliazucchi, E., von Wegner, F., Morzelewski, A., Borisov, S., Jahnke, K., and Laufs, H. (2012). Automatic sleep staging using fmri functional connectivity data. *Neuroimage*, 63(1):63–72.

Wang, Y., Loparo, K. A., Kelly, M. R., and Kaplan, R. F. (2015). Evaluation of an automated single-channel sleep staging algorithm. *Nature and science of sleep*, 7:101.

Zibrandtsen, I., Kidmose, P., Otto, M., Ibsen, J., and Kjaer, T. (2016). Case comparison of sleep features from ear-eeg and scalp-eeg. *Sleep Science*.