Sleep Stage Classification Using Wavelets, HOS and Cepstral Analysis

Christodoulos Michaelidis, Chrysoula Doulou, Dimitrios Orfanos and Stergios Grigoriou

***Abstract*—An approach, using multi-resolution analysis (MRA), higher order statistics (HOS) and cepstral analysis to extract features from polysomnographic sleep recordings (PSGs) suitable for automatic sleep scoring. An open-source dataset is utilized, consisting of 151 PSGs (4xEEG, chin EMG,2xEOG, ECG). Features extracted from EEG using MRA, HOS and cepstral analysis are mixed with sleep-related statistical features from EOG, EMG and ECG to form a rich feature-set. The analysis is combined with the implementation of different machine learning classification models. The best model is an SVM which scored 0.63 kappa in 5-fold validation. Findings show that although classification reaches substantial agreement with annotators, further study is needed either for enriching the feature-set with more complex EOG/EMG features or for tuning the classification scheme.**

***Index Terms*—Sleep Scoring, MRA, Bicoherence, Cepstrum, SVM**

I. Introduction

SLEEP is a big and important part of our lives. It is estimated that sleeping or attempting to sleep makes around a third of our lifetimes. Sleep is essential for our normal cognitive functioning and for our survival. Having a good night sleep makes a huge impact in someone's life both behaviourally and physiologically. This is why a lot of research has been focused on sleep over the years. [1] From a scientific standpoint, sleep is defined on the basis of both the behaviour of the person while asleep and related physiological changes that occur to the waking brain's electrical rhythms in sleep. [2] Although body movements are largely suppressed during sleep, resulting in reduced external behavior, the internal activity of the brain is rich and diverse, throughout a sleep session. [3] Thus it would be an inaccuracy to assume that sleep is a monotonous process. On contrary, sleep consists of distinct stages which alternate through a session. According to AASM v2.4 [4] sleep occurs in five stages: wake, N1, N2, N3, and REM. Stages N1 to N3 are considered non-rapid eye movement (NREM) sleep, with each stage a progressively deeper sleep. Each phase and stage of sleep includes variations in muscle tone, brain wave patterns, and eye movements. The body cycles through all of these stages approximately 4 to 6 times each night, averaging 90 minutes for each cycle. [5]

The amount of time in a session that each stage occupies together with sleep patern data is a very useful feature for diagnosis of a variety of sleep disorders. [6] Thus, sleep staging is one of the most important tasks during the clinical examination of polysomnographic sleep recordings (PSGs). A PSG records the relevant biomedical signals of a patient in the context of Sleep Medicine studies, representing the basic tool for the diagnosis of many sleep disorders. [7] In general that would be several channels of electroencephalography (EEG), electrooculography (EOG), electromyography (EMG) and electrocardiography (ECG). Each of these signals gives as an insight about the sleep stage because each sleep stage is classified by standard different physiological markers.

In particulaar, when awake people tend to move more, exhibit slow eye movement (SEM), have a risen heart-rate (HR), and usually develop high activity in the beta wave band of the EEG. Stage N1 normally lasts just between one to five minutes. During N1 sleep, the body and brain activities start to slow with periods of brief movements (twitches). Light changes in brain activity are observed during this stage, which are associated with falling asleep. It is easy to wake someone up during this sleep stage, but if unless a person is disturbed, they can move quickly into stage N2. As the sleep goes on, an uninterrupted sleeper may not spend much more time in stage N1 as they move throughout further sleep cycles.

During stage N2, the body enters a more subdued state which includes a drop in temperature, relaxed muscles, and slowed breathing and heart rate. At the same time, brain waves show a new pattern and eye movement stops. On the whole, brain activity slows, but there are short bursts of activity that actually help resist being woken up by external stimuli. Stage N2 sleep can last for 10-25 minutes during the first sleep cycle, and each N2 stage can become longer during the night. Collectively, a person typically spends about half their sleep time in N2 sleep.

Stage N3 sleep is also known as deep sleep, and it is harder to wake someone up if they are in this phase. Muscle tone, pulse, and breathing rate decrease in N3 sleep as the body relaxes even further. The brain activity during this period has an identifiable pattern of what are known as delta waves. For this reason, stage N3 may also be called delta sleep or slow-wave sleep (SWS). Even though brain activity is reduced, there is evidence that deep sleep contributes to insightful thinking, creativity, and memory. We spend the most time in deep sleep during the first half of the night. During the early sleep cycles, N3 stages commonly last for 20-40 minutes. As you continue sleeping, these stages get shorter, and more time gets spent in REM sleep instead.

During REM sleep, brain activity picks up, nearing levels seen when you're awake. At the same time, the body experiences atonia, which is a temporary paralysis of the muscles, with two exceptions: the eyes and the muscles controlling breathing. Even though the eyes are closed, they can be seen moving quickly, which is how this stage gets its name. REM sleep is known for the most vivid dreams, which is explained

by the significant uptick in brain activity. Dreams can occur in any sleep stage, but they are less common and intense in the NREM periods. Under normal circumstances, a REM sleep stage is not entered until being asleep for about 90 minutes. As the night goes on, REM stages get longer, especially in the second half of the night. While the first REM stage may last only a few minutes, later stages can last for around an hour. In total, REM stages make up around 25% of sleep in adults. [6]

The EEG signals have proven to be the most reliable type of signal to detect the various stages of sleep [8] and a lot of work on automatic sleep scoring has been focused on analysing and assessing the features extracted from a single EEG channel. This method has gained popularity over the years due to the limitations of polysomnography, but as concluded in, [9] PSG recordings for automatic sleep scoring algorithms are still in need when dealing with inconsistent sleep patterns.

A typical PSG examination comprises 8 up to 24 hours of continuous signal recording, and its analysis is usually carried out manually by an expert clinician. The scoring process is consequently expensive and highly demanding, due to the involved clinician's time, and the complexity of the analysis itself. Moreover, the demand for PSG investigations is growing in relation with the general public awareness, motivated by clinical findings over the last years uncovering the negative impact that sleep disorders exert over health. This represents a challenge for the already congested sleep centers, with steadily increasing waiting lists.

Automatic analysis of the sleep macrostructure is thus of interest, given the potential great savings in terms of time and human resources. An additional advantage is the possibility of providing deterministic (repeatable) diagnostic outcomes, hence contributing to the standardization and quality improvement in the diagnosis. [7]

In this study an ensemble of signal processing methods is used to extract sleep scoring related features. These methods include multiresolution analysis (MRA), use of higher order statistics (HOS) and cepstral analysis. Combined with machine learning (binary trees ensemble and SVM), it is an attempt to create a robust automatic sleep scoring model based on PSG data from an open-source dataset.

II. Mathematical Background

A. Multi-resolution Analysis (MRA)

Wavelet transform is a really useful tool for time-frequency analysis of a signal, because of its ability to successfully address the possible non-stationarity of it. Thus it is particularly useful on the non-stationary signals that are recorded during a polysomnography, such as electroencephalographic (EEG), electrooculographic (EOG), electromyographic (EMG), and electrocardiographic (ECG) activities. [10] The multi-resolution analysis (MRA) is a method developed by Mallat [11], [12] that provides a good time resolution and low frequency resolution at high frequencies and low time resolution and high frequency resolution at low frequencies. Thus it is uniquely suited for the analysis of signals with most of their transient phenomena at low frequencies as the signals at hand.

MRA is based on the dual-scale equation representation [13], [14] of the discrete wavelet transform (DWT):

$$\phi(t) = \sum_n h(n)\phi_{-1,n}(t) = \sqrt{2}\sum_n h(n)\phi(2t-n) \quad (1)$$

$$\psi(t) = \sum_n g(n)\phi_{-1,n}(t) = \sqrt{2}\sum_n g(n)\phi(2t-n) \quad (2)$$

where ϕ is the scaling function ψ is the wavelet function and the coefficients

$$h(n) = \langle \phi, \phi_{-1,n} \rangle$$
$$g(n) = \langle \psi, \phi_{-1,n} \rangle$$

are a pair of low-pass and high-pass wavelet filters. Essentially, through MRA we decompose a signal $x(t)$ to its approximate coefficient $a_{j,k}$ and its detailed coefficient $d_{j,k}$.

$$a_{j,k} = \langle x(t), \phi_{j,k} \rangle = \sum_m h(m-2k)a_{j-1,m} \quad (3)$$

$$d_{j,k} = \langle x(t), \psi_{j,k} \rangle = \sum_m g(m-2k)a_{j-1,m} \quad (4)$$

So the approximate coefficient represents the low-frequency component of the signal, while the detailed the high-frequency component and the coefficients of each scale are obtained by convolving the previous-scale approximate coefficients with the low-pass and the high-pass filter respectively. Inferentially, each level (scale) of the MRA splits the approximate coefficients into to sub-bands $[0, \frac{F}{2}]$ and $[\frac{F}{2}, F]$, where F is the Nyquist frequency of $a_{j,k}$.

B. Bispectrum and Bicoherence

The power spectrum does not contain any information about the non-linear properties of a signal. On the other hand, Higher-Order-Spectra have been successfully used to study such non-linear phenomena. [15] Out of all the Higher-Order-Spectra, the bispectrum is by far the most widely used, since it provides an easy way to detect quadratic-phase-coupling (QPC) between two frequency components of a signal. [16], [17], [18] QPC is the sum of phases at two frequency variables given by f1 + f2. Bispectrum can be estimated through two approaches: direct and indirect methods. The direct method follows 5 steps [19]. Let $x(k)$ be a discrete random process:

1) Divide the process into K segments of M samples.
2) Subtract the mean value from each segment.
3) Compute the DFT of each segment

$$F_x^i(\lambda) = \sum_{k=0}^{M-1} x^i(k)e^{-j\frac{2\pi}{M}k\lambda} \quad (5)$$

where $i = 1, 2, \ldots, K$, $\lambda = 0, 1, \ldots, M-1$

4) Then an estimation of bispectrum in each segment i is given by:

$$\hat{b}_3^{x^i}(\lambda_1, \lambda_2) = \frac{1}{\Delta_0^2} F_x^i(\lambda_1)F_x^i(\lambda_2)F_x^{i*}(\lambda_1+\lambda_2) \quad (6)$$

5) The bispectrum is estimated by averaging across the K segments, i.e.

$$\hat{C}_3^x(\omega_1, \omega_2) = \frac{1}{K}\sum_{i=1}^{K} \hat{b}_3^{x^i}(\omega_1, \omega_2) \quad (7)$$

where $\omega_1 = (\frac{2\pi f_s}{N_0})\lambda_1, \omega_2 = (\frac{2\pi f_s}{N_0})\lambda_2$

Since the Fourier transform of a signal is generally speaking a complex number, so is its bispectrum. Assuming that a frequency pair (f_1, f_2) exhibits perfect quadratic-phase-coupling, then the bispectrum of said pair should in theory be infinite in magnitude and its phase/angle should be zero. In practice, it is more convenient to use a normalized version of the bispectrum to detect quadratic phase coupling, called the bicoherency index. Many methods have been proposed to normalize the bispectrum. The one used here [20] is obtained as follows:

$$b^2(f_1, f_2) = \frac{E[|F_x(f_1)F_x(f_2)F_x^*(f_1 + f_2)|]^2}{E[|F_x(f_1)F_x(f_2)|^2] \cdot E[|F_x^*(f_1 + f_2)|^2]} \quad (8)$$

By its definition, the bicoherence index is a real number bounded between 0 and 1, which can be interpreted as the amount of energy in f1+f2 contributed by quadratic-phase-coupling between f1 and f2. Therefore, values near 1 indicate that the non-linear interaction between f1 and f2 is strong, whereas values near 0 indicate that the interaction is non-existent.

C. Cepstrum

In order to enrich the feature representation making use of EEG periodic patterns, this paper applies a preliminary research based on cepstrum and period detection that can be further expanded using cepstral coefficients. Cepstrum Analysis is a tool for the detection of periodicity in a frequency spectrum, and so far it is mainly used in speech analysis for voice pitch determination and related questions. [21] The cepstrum is defined as the power spectrum of the logarithmic power spectrum (i.e. in dB amplitude form) [22], and is thus related to the auto-correlation function, which can be obtained by inverse Fourier transformation of the power spectrum with linear ordinates. The real cepstrum of signal x(t) is calculated as follows:

$$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} log|X(\omega)|e^{i\omega}\, d\omega \quad (9)$$

III. Materials and Methods

A. Dataset

We used the Haaglanden Medisch Centrum sleep staging database , [7] which is an open database in physionet [23] that consists of 151 (whole-night polysomnographic (PSG)) sleep recordings (85 Male, 66 Female, mean Age of 53.9 ± 15.4) collected during 2018 at the Haaglanden Medisch Centrum (HMC, The Netherlands) sleep center. The PSG recordings contain four EEG channels (F4/M1, C4/M1, O2/M1, and C3/M2), two EOG channels (E1/M2 and E2/M2), one bipolar chin EMG and one ECG (single modified lead II). Along with the recordings, the dataset includes also the corresponding hypnogram annotations. Every 30s of sleep are accompanied by an annotation according to the 2.4 version of the AASM guidelines. All signals were sampled at 256 Hz. Signals were recorded using SOMNOscreen PSG, PSG+, and EEG 10-20 recorders (SOMNOmedics, Germany) using AgAgCl electrodes. Recordings were randomly selected from a heterogeneous population which was referred for PSG examination on the context of different sleep disorders. Though, annotations of these disorders have not been made public.

B. Preprocessing

A band-pass elliptic filter is used, in order to face some issues inherent to our dataset; mainly lack of homogeneity on prefilter parameters and high interference caused by the power grid. The response of the filter is band-pass between 0.4 and 33.5 Hz, with higher attenuation around the power grid frequency (in our dataset 50Hz [7]). Forward-backward filtering is used in order to ensure zero-phase distortion.

C. EEG

We performed a variety of analysis methods on all EEG channels. Though the most useful features came out of the central derivations as expected.

Firstly 5 level MRA was performed using the symlet 5 wavelet. The selection of this particular wavelet was based on empirical data and past researches [10] in this area. This process decomposed our signal in 6 sub-bands similar to the EEG bands tabulated in Table I. The resulted scales are shown in Figure 1.

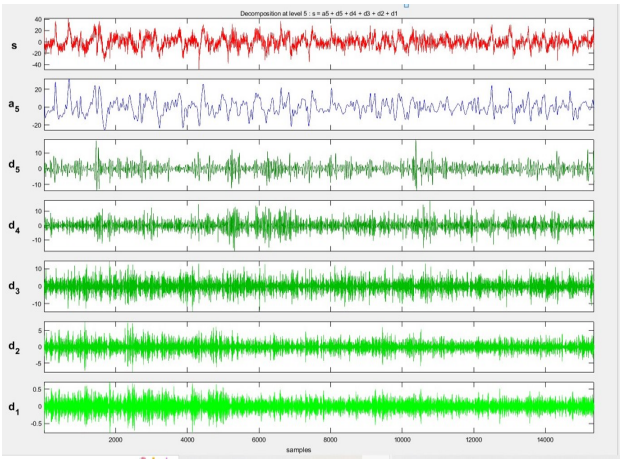
TABLE I: The EEG decomposition

Decomposition Level	Frequency Band (Hz)	Coefficients	EEG Band
1	64–128	D1	Gamma and noise
2	32–64	D2	Lower gamma (γ)
3	16–32	D3	Beta (β)
4	8–16	D4	Alpha (α)
5	4–8	D5	Theta (θ)
5	0–4	A5	Delta (δ)

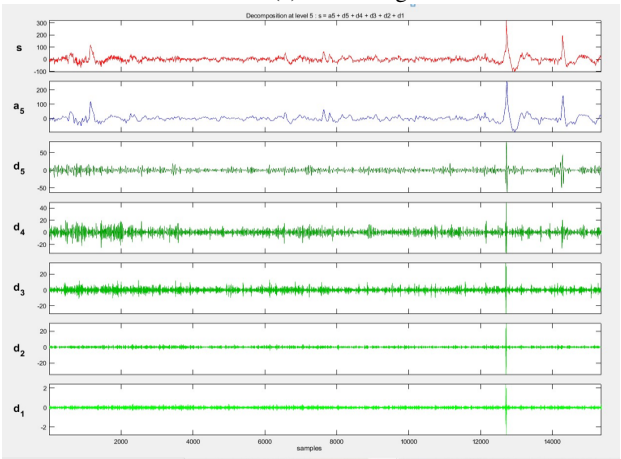
Some of the pre-filters on EEG recordings were set at 35 Hz. So for reasons of homogeneity only the lower 4 bands were kept, which (roughly) correspond to delta, theta , alpha and beta waves respectively. Then statistical features were computed on each band.

Through performing wavelet decomposition on the EEG channels we obtain very conveniently the frequency bands that prevail in each stage of sleep. During Wakefulness alpha (8-12) and beta waves (12-32) appear in the majority of time, in N1 we observe theta waves (4-8). Delta waves are present mostly in the deepest Non-REM stage (0-4) and in the REM stage beta waves are back again. The dynamics of each frequency band in every stage of sleep can be described via various statistical measures. In this study variance, skewness and kurtosis of the wavelet coefficients are calculated, giving us a better understanding of the complexity of the EEG signals.

When first implementing cepstral analysis of the EEGs our primary goal was to model the brain transitions from one sleep stage to another. This proved fruitless, though in the process we extracted some additional useful features. For the cepstral analysis of the EEGs we use only the alpha and beta rhythm bands (D4 and D3 respectively). In each thirty-second epoch we use a hamming window of the same length on each scale (D3, D4). Then we calculate the real cepstrum using (9) on each scale. The first 2 seconds of the quefrency are assumed to represent the impulse response of the system (the brain) at

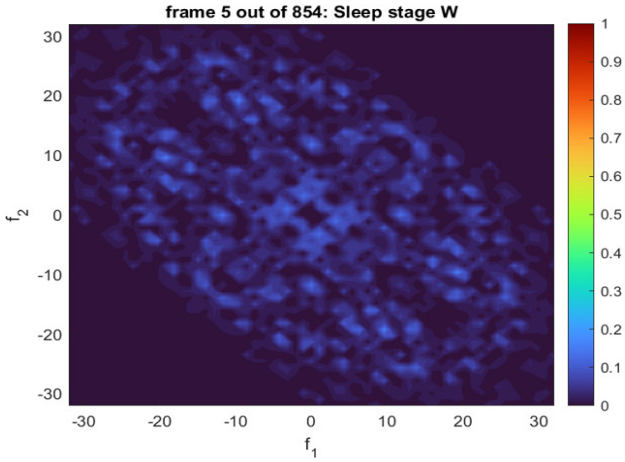


(a) Wake stage

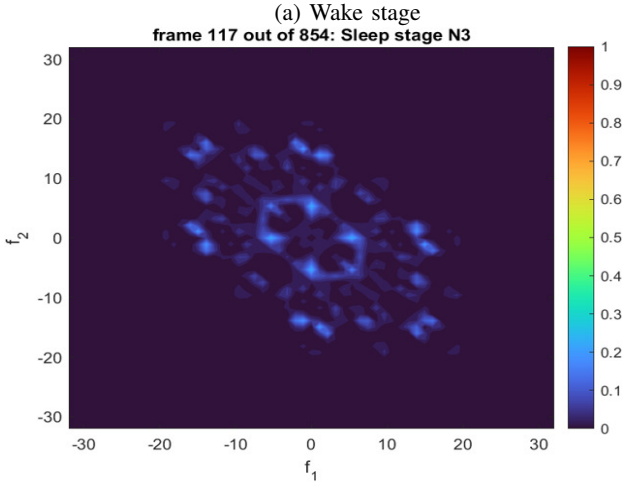


(b) N3 stage

Fig. 1: Shows two 60s-epoch of EEG (F4/M1) decomposed using the symlet 5 wavelet.



(a) Wake stage



(b) N3 stage

Fig. 2: Shows two 30s-epoch plots of bicoherence on EEG-C3/M2.

the given sleep stage and the remaining 13 are assumed to be the input. For each scale we calculate (both for the impulse response and for the input) the first four statistical moments (mean, variance, skewness, kurtosis) and the zero-crossing rate of the cepstral coefficients.

Then we proceed on reconstructing the signal using the 4 lower scales (A5,D5,D4,D3), which appeared to be the most useful ones. On the reconstructed signal we calculated the bicoherence using (8) to search for non-linear phenomena. EEG signals exhibit such non-linear phenomena. However, the intensity of these phenomena seems to reduce as a person moves to deeper sleep stages, which means that the bicoherence could be used to quantify the depth of sleep. Figure 2 shows an estimate of the EEG-bicoherence during wakefulness and deep sleep (N3) respectively.

From the bicoherence (only primary I area) we extract four features for every 30s epoch:

- Bicoherence entropy calculated as:

$$H_b = -E_{|b|}[log_2(\sqrt{b^2})]$$

- Squared bicoherence entropy calculated as:

$$H_{b^2} = -E_{b^2}[log_2(b^2)]$$

- The mean of bicoherence distribution $E_b[\sqrt{b^2}]$.
- The 80th-percentile frequency of squared bicoherence b^2. As a measure of how spread out is the bicoherence in each epoch.

Then we perform a hard thresholding on the bicoherence estimates in order to detect QPC. We keep only the points that exceed a threshold of 10% of total bicoherence (or have a value larger than 0.01 in normalized b^2 in our case) to locate the local maxima in each epoch. These local maxima are indicators of QPC. This process is mathematically expressed as:

$$b^2(f_1, f_2) = \begin{cases} b^2(f_1, f_2), & b^2(f_1, f_2) \geq 0.01 \\ 0, & b^2(f_1, f_2) < 0.01 \end{cases}$$

Figure 3 presents the results of this procedure on a N1 annotated epoch. Lastly, from this thresholded squared bicoherence (only primary I area), we extracted an additional five features, which are used to quantify and qualify the QPC in each epoch *(Let d_i be the euclidean distance of each peak from the origin of the frequency axes)*:

- Number of detected peaks (N).

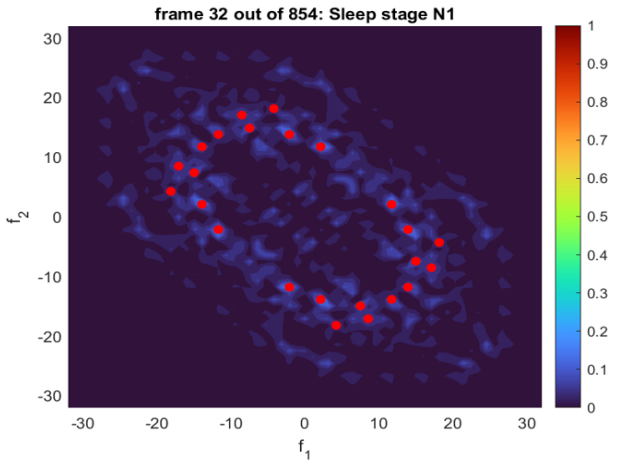


Fig. 3: QPC detection on a N1 30s-epoch plot. *With red colour are marked the kept peaks, which indicate QPC.*

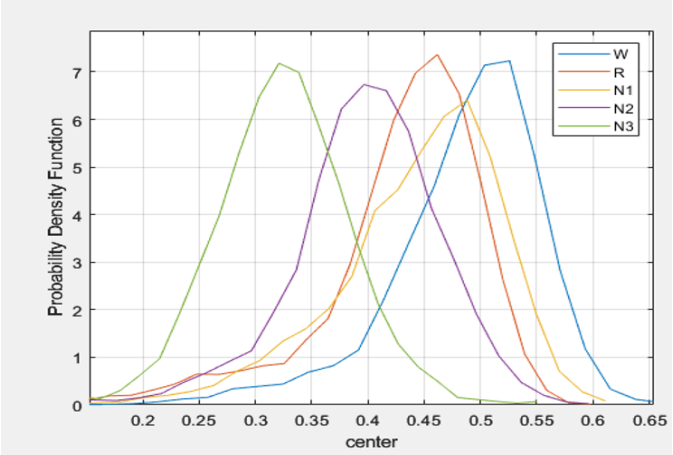


Fig. 4: The center of mass distributions of detected QPC for the five sleep stages. *(Blue = Wake, Yellow = N1, Purple = N2, Green = N3, Red = R)*

- The weighted mean of the root values of detected peaks (prioritizing the furthest from the origin peaks).

$$A = \frac{\sum_i^N (d_i \sqrt{b_i})}{N \cdot \sum_i^N d_i}$$

- The weighted mean of the euclidean distance of each peak from the origin of the frequency axes

$$\bar{d} = \frac{1}{N} \sum_i^N (d_i \sqrt{b_i})$$

.

- The weighted maximum of the euclidean distances the peaks from the origin of the frequency axes.

$$d_{max} = \max_i (d_i \sqrt{b_i})$$

- The center of mass of the detected peaks calculated as:

$$C = \sqrt[r]{\frac{(\sum_i^N (b_i f_1))^r + (\sum_i^N (b_i f_2))^r}{(\sum_i^N b_i)^r}}$$

The results for r = 1.2 can be seen in Figure 4.

D. ECG

Electrocardiogram data would be really useful in classifying sleep apnea/hypopnea trhough cepstral analysis, [24] but relative annotations are not available in the chosen dataset. Thus some more general attributes, which slightly improve classification of sleep stages are estimated. These are, heart rate (HR), short-term (5 min) heart rate variability (ST-HRV) and ultra-short-term (each 30s epoch) heart rate variability (UST-HRV) [25].

At first, an additional preprocessing is required in order to revert some of the ECGs that had different polarity and to discard 2 of them as corrupted (SN036.edf and SN098.edf). Then, MRA is performed and we keep D5,D4 and D3 scales (4-32 Hz) using symlet 4 wavelet. [26] We continue with implementation of the Pan-Tompkins algorithm [27] as described here [28], in order to retrieve the time positions of the QRS complexes. At last, we estimate HR, ST-HRV and UST-HRV using these moments.

E. EOG and EMG

In the goal towards automatic sleep scoring, particularly challenging was found to be the segregation of Wakefulness, REM stage and first stage of sleep (N1). That happens because when the brain enters the REM stage of sleep it displays similar alertness to when awake, alongside with N1 being a passage from Wakefulness to sleep. [29] EOG signals have been used vastly in identifying the stages of sleep. As highlighted in [8] they are the second more effective type of signal after the EEG signals. They provide useful information for distinguishing between REM and Non-REM sleep. EMG signals give useful information towards that decision. The lowest muscle tone is observed during REM stage, and during stage N1 there is less movement than what is observed during wakefulness. Their efficiency combined with the easy placement that can be done outside a sleep monitoring facility and without professional assistance have resulted in the development of automatic sleep scoring algorithms using features from single EOG channel [30] to the exploitation of the EOG features along with EEG and EMG signals, achieving a maximum accuracy of 91.8%.

Five level MRA with Daubechies' 5 wavelet is performed on both EOG channels and only the approximation coefficients (0-4 Hz) are used in the reconstruction of the signal. Then two features are extracted from the reconstructed signals in each 30s-epoch. These are:

1) Total energy of the EOG channels inside the 0.10Hz-1.00Hz frequency band (ECB).
2) Cross-correlation coefficient of the two EOG channels. A useful feature for identifying mirrored EOGs. A phenomenon that occurs almost specifically in REM sleep.

On the EMG, aside from the universal preprocessing filter, no other frequency analysis is performed, since the useful data lies within its amplitude. Thus we calculated, for every epoch, the standard deviation of the signal and its peak to peak voltage every 30 seconds.

F. Classification Procedure

Using the results of the previous analysis techniques, it is possible to train simple classifiers for automatic sleep stage scoring. The classifiers are trained using a 5-fold cross-validation partinioning scheme. Namely, 80% of the available PSG recordings are used for training and 20% for cross-validation. The performance of the classifiers is evaluated by using the average classification accuracy. The EEG features which have been proposed during the bispectral, cepstral and multi-resolution analysis as well as the features from the EOG, ECG and EMG recordings are estimated in 30 second time windows for every available recording and then fed into the classifiers.

In total, three classifier systems were trained for the purpose of automatic sleep stage scoring. Our first approach utilises a simple decision tree which achieves an average cross-validation accuracy of 59,23%.

Plain decision trees are rarely used in practice. Over the last years, several boosting techniques have been proposed to train ensembles of multiple weak learners. Our second classifier uses Adaboost, a fairly well-known boosting algorithm, [31] to train an ensemble of 20 decision trees. This classifier correctly recognises 65,71% of the available cross-validation recordings.

Finally, a group of 10 SVMs with gaussian kernels was used to reduce this multi-class classification problem into a series of simpler binary-classification problems. This prediction model achieved an accuracy of 72,4%.

IV. Results and Discussion

To begin with, the chosen dataset -as literature would suggest- is not at all balanced. To be specific we have 23519 W, 15415 N1, 49256 N2, 26407 N3 and 20963 REM 30s epochs out of a total of 135560 (not including subjects 36 and 98 due to corrupted ECG). Therefore, it would be an inaccuracy to use simple accuracy as our sole criterion. A more suitable statistic, widely used in automatic sleep stage scoring evaluation [7], [32], [33], [34], is Cohen's kappa [35]. Cohen's kappa measures inter-rater agreement rather than simple accuracy and excludes the possibility of agreeing by chance. Thus the results of the three different classifier architectures proposed are presented in table II.

TABLE II: Classifier Performance

Classifier Architecture	Classification Accuracy	kappa
Plain Decision Tree	59,23%	45.02%
Ensemble of 20 Decision Trees	65.71%	51.09%
One-vs-One SVM	72.18%	62.8%

The SVM model seems superior in terms of performance, and achieves substantial agreement with the annotators of the dataset. Though it is far from being perfect. We can infer that our proposed set of features does indeed provide some useful information for recognising different sleep stages. However, the rather mediocre performance of the best model leaves a lot of room for further improvements.

In figure 5 the confusion matrix of the model is presented, from which we can deduce where the model is weak. We see that there is weak to none (we could say random) classification of the stage N1. There is also some substantial mix of REM with N2 and N3 with N2.

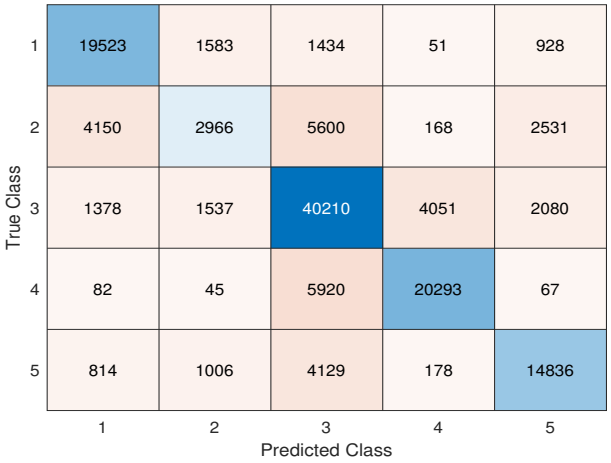


Fig. 5: Confusion Matrix of the SVM model. *(1 = W, 2 = N1, 3 = N2, 4 = N3, 5 = REM)*

These inaccuracies originate both from the classification optimization scheme and from weak EOG and EMG features. None of the proposed classifiers use any kind of temporal information. This is probably the main drawback of our prediction models, since it is a well known fact that people go through many cycles of non-REM and REM stages, while they sleep. We also observed that our classifiers were not particularly good at distinguishing between REM sleep and wakefulness. This is due to the fact that those stages exhibit relatively similar EEG activity, which means that the only reliable way of distinguishing between those two classes is by relying on the EOG patterns. Unfortunately, our proposed feature-set does a poor job at capturing useful EOG information. Last but not least, the performance of the classifiers is also hindered by the fact that our dataset does not provide any information about the pathologies of its participants.

A proposed action would be to implement further EOG analysis and make use of some novel features that exist in literature such as [36], which help classify eye movements as slow or fast. Further EMG analysis is also suggested. At first better filtering of low frequencies could appear useful. Even some traditional EEG features which were, in this study, excluded could be in fact useful such as sub-band relative energies.

Another, correctional direction is tuning the classification process. Although some dimension-reduction techniques (i.e PCA) were used without positive results, a more thorough feature selection would be proven beneficial. Furthermore, we could use some features like opposite EOG channels to immediately classify some epochs (in this example it would be REM) and train a model on the rest of the dataset.

V. Conclusion

To conclude this paper proposes an automatic sleep stage classification based on an ensemble of signal processing methods and machine learning. In particular, by combining MRA with HOS and with cepstral analysis on PSG recording signals,

a rich feature-set was formed. The classification performance using SVM reaches substantial agreement but is considered mediocre (k = 0.63), however useful data about some of the features is obtained. Apparently, further study is needed either for enriching the feature-set with useful EOG/EMG features or for tuning the classification scheme.

VI. Implementation

The code for the given analysis can be found at this public repository: https://github.com/cmichailidis/ASPT-Project

References

[1] M. Aminoff, F. Boller, and D. Swaab, "We spend about one-third of our life either sleeping or attempting to do so," *Handbook of clinical neurology / edited by P.J. Vinken and G.W. Bruyn*, vol. 98, p. vii, 01 2011.

[2] S. Chokroverty *et al.*, "Overview of sleep & sleep disorders," *Indian J Med Res*, vol. 131, no. 2, pp. 126–140, 2010.

[3] T. J. Sejnowski and A. Destexhe, "Why do we sleep?" *Brain research*, vol. 886, no. 1-2, pp. 208–223, 2000.

[4] R. B. Berry, R. Brooks, C. Gamaldo, S. M. Harding, R. M. Lloyd, S. F. Quan, M. T. Troester, and B. V. Vaughn, "Aasm scoring manual updates for 2017 (version 2.4)," pp. 665–666, 2017.

[5] A. K. Patel, V. Reddy, and J. F. Araujo, "Physiology, sleep stages," in *StatPearls [Internet]*. StatPearls Publishing, 2021.

[6] M. A. Carskadon, W. C. Dement *et al.*, "Normal human sleep: an overview," *Principles and practice of sleep medicine*, vol. 4, no. 1, pp. 13–23, 2005.

[7] D. Alvarez-Estevez and R. M. Rijsman, "Inter-database validation of a deep learning approach for automatic sleep scoring," *PloS one*, vol. 16, no. 8, p. e0256111, 2021.

[8] M. Sharma, A. Yadav, J. Tiwari, M. Karabatak, O. Yildirim, and U. R. Acharya, "An automated wavelet-based sleep scoring model using eeg, emg, and eog signals with more than 8000 subjects," *International Journal of Environmental Research and Public Health*, vol. 19, no. 12, p. 7176, 2022.

[9] B. P. Lucey, J. S. Mcleland, C. D. Toedebusch, J. Boyd, J. C. Morris, E. C. Landsness, K. Yamada, and D. M. Holtzman, "Comparison of a single-channel eeg sleep study to polysomnography," *Journal of sleep research*, vol. 25, no. 6, pp. 625–635, 2016.

[10] N. K. Al-Qazzaz, S. Hamid Bin Mohd Ali, S. A. Ahmad, M. S. Islam, and J. Escudero, "Selection of mother wavelet functions for multi-channel eeg signal analysis during a working memory task," *Sensors*, vol. 15, no. 11, pp. 29 015–29 035, 2015.

[11] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[12] ——, "Multiresolution approximations and wavelet orthonormal bases of v^2," *Transactions of the American mathematical society*, vol. 315, no. 1, pp. 69–87, 1989.

[13] I. Daubechies, *Ten lectures on wavelets*. SIAM, 1992.

[14] D. L. Donoho, S. G. Mallat, R. von Sachs *et al.*, "Estimating covariances of locally stationary processes: rates of convergence of best basis methods," Departement of Statistics, Stanford University, Tech. Rep., 1998.

[15] D. P. Subha, P. K. Joseph, R. Acharya U, C. M. Lim *et al.*, "Eeg signal analysis: a survey," *Journal of medical systems*, vol. 34, no. 2, pp. 195–212, 2010.

[16] N. Arthur and J. Penman, "Induction machine condition monitoring with higher order spectra," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 1031–1041, 2000.

[17] C. Courtney, S. Neild, P. Wilcox, and B. Drinkwater, "Application of the bispectrum for detection of small nonlinearities excited sinusoidally," *Journal of Sound and Vibration*, vol. 329, no. 20, pp. 4279–4293, 2010.

[18] A. J. Hillis, S. A. Neild, B. W. Drinkwater, and P. D. Wilcox, "Global crack detection using bispectral analysis," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 462, no. 2069, pp. 1515–1530, 2006.

[19] W. Collis, P. White, and J. Hammond, "Higher-order spectra: the bispectrum and trispectrum," *Mechanical systems and signal processing*, vol. 12, no. 3, pp. 375–394, 1998.

[20] Y. Nagashima, K. Itoh, S. Itoh, K. Hoshino, A. Fujisawa, A. Ejiri, Y. Takase, M. Yagi, K. Shinohara, K. Uehara *et al.*, "Observation of coherent bicoherence and biphase in potential fluctuations around geodesic acoustic mode frequency on jft-2m," *Plasma physics and controlled fusion*, vol. 48, no. 5A, p. A377, 2006.

[21] R. Makhijani, U. Shrawankar, and V. M. Thakare, "Speech enhancement using pitch detection approach for noisy environment," *arXiv preprint arXiv:1305.2352*, 2013.

[22] A. V. Oppenheim and R. W. Schafer, "From frequency to quefrency: A history of the cepstrum," *IEEE signal processing Magazine*, vol. 21, no. 5, pp. 95–106, 2004.

[23] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000 (June 13), circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215.full PMID:1085218; doi: 10.1161/01.CIR.101.23.e215.

[24] S. Martín-González, J. L. Navarro-Mesa, G. Juliá-Serdá, J. F. Kraemer, N. Wessel, and A. G. Ravelo-García, "Heart rate variability feature selection in the presence of sleep apnea: An expert system for the characterization and detection of the disorder," *Computers in biology and medicine*, vol. 91, pp. 47–58, 2017.

[25] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, p. 258, 2017.

[26] R. Singh, R. Mehta, and N. Rajpal, "Efficient wavelet families for ecg classification using neural classifiers," *Procedia computer science*, vol. 132, pp. 11–21, 2018.

[27] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.

[28] H. Sedghamiz, "Matlab implementation of pan tompkins ecg qrs detector," *Code Available at the File Exchange Site of MathWorks*, 2014.

[29] R. Yan, F. Li, X. Wang, T. Ristaniemi, and F. Cong, "An automatic sleep scoring toolbox: multi-modality of polysomnography signals' processing," in *International Conference on Signal Processing and Multimedia Applications*. SCITEPRESS Science And Technology Publications, 2019.

[30] C.-E. Kuo, S.-F. Liang, Y.-C. Lee, F.-Y. Cherng, W.-C. Lin, P.-Y. Chen, Y.-C. Liu, and F.-Z. Shaw, "An eog-based automatic sleep scoring system and its related application in sleep environmental control," in *International Conference on Physiological Computing Systems*. Springer, 2014, pp. 71–88.

[31] Y. Freung and R. Shapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J Comput Syst Sci*, vol. 55, pp. 119–139, 1997.

[32] H. Danker-hopfe, P. Anderer, J. Zeitlhofer, M. Boeck, H. Dorn, G. Gruber, E. Heller, E. Loretz, D. Moser, S. Parapatics *et al.*, "Interrater reliability for sleep scoring according to the rechtschaffen & kales and the new aasm standard," *Journal of sleep research*, vol. 18, no. 1, pp. 74–84, 2009.

[33] U. J. Magalang, N.-H. Chen, P. A. Cistulli, A. C. Fedson, T. Gíslason, D. Hillman, T. Penzel, R. Tamisier, S. Tufik, G. Phillips *et al.*, "Agreement in the scoring of respiratory events and sleep among international sleep centers," *Sleep*, vol. 36, no. 4, pp. 591–596, 2013.

[34] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[35] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and psychological measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[36] D. Alvarez-Estevez, I. van Velzen, T. Ottolini-Capellen, and B. Kemp, "Derivation and modeling of two new features for the characterization of rapid and slow eye movements in electrooculographic sleep recordings," *Biomedical Signal Processing and Control*, vol. 35, pp. 87–99, 2017.