Sleep-wake stages classification using heart rate signals from pulse oximetry

Ramiro Casal[a,b,d,*], Leandro E. Di Persia[b,c], Gastón Schlotthauer[a,b,d]

[a] *Lab. de Señales y Dinámicas no Lineales, Facultad de Ingeniería, Universidad Nacional de Entre Ríos (UNER), Argentina*
[b] *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina*
[c] *Instituto de Investigacion en Señales, Sistemas e Inteligencia Computacional - Universidad Nacional del Litoral - CONICET*
[d] *Instituto de Investigación y Desarrollo en Bioingeniería y Bioinformática - UNER - CONICET*

Abstract

The most important index of obstructive sleep apnea/hypopnea syndrome (OSAHS) is the apnea/hyponea index (AHI). The AHI is the number of apnea/hypopnea events per hour of sleep. Algorithms for the screening of OSAHS from pulse oximetry estimate an approximation to AHI counting the desaturation events without consider the sleep stage of the patient. This paper presents an automatic system to determine if a patient is awake or asleep using heart rate (HR) signals provided by pulse oximetry. In this study, 70 features are estimated using entropy and complexity measures, frequency domain and time-scale domain methods, and classical statistics. The dimension of feature space is reduced from 70 to 40 using three different schemes based on forward feature selection with support vector machine and feature importance with random forest. The algorithms were designed, trained and tested with 5000 patients from the Sleep Heart Health Study database. In the test stage, 10-fold cross validation method was applied obtaining performances up to 85.2% accuracy, 88.3% specificity, 79.0% sensitivity, 67.0% positive predictive value, and 91.3% negative predictive value. The results are encouraging, showing the possibility of using HR signals obtained from the same oximeter to determine the sleep stage of the patient, and thus potentially improving the estimation of AHI based on only pulse oximetry.

Keywords: Sleep apnea, Pulse oximetry, Heart rate, Automatic sleep staging

1. Introduction

Sleep plays a very important role in well-being and physiological recovery. The gold standard test for the study of sleep pathologies is polysomnography (PSG), which consists of the simultaneous recording of several physiological signals such as electroencephalography (EEG), electrocardiography (ECG), electromiography (EMG), respiratory effort, oronasal airflow, peripheral oxygen saturation (SpO_2), and electrooculography (EOG), among others. The PSG is supervised by a technician in a sleep medical center specially conditioned, and its analysis requires a tedious scoring, often by hand with the help of a software [1]. The scoring has a lot of variability among different professionals [2]. Due to these characteristics and the limited number of beds, the PSG is cost intensive and its availability is scarce, generating long waiting lists. Furthermore, many patients are reluctant to spend the night in the sleep laboratory or have difficulty to falling sleep [3].

Due to complexity, limited capacity, and high costs associated to the PSG, there is an increasing interest in reducing the need for complete PSG studies. Many approaches have been proposed to perform screening of sleep pathologies. Pulse oximeter is an ideal choice for the screening due to its low cost, accessibility and simplicity [3]. For this, it stands out among other techniques, such as cardiac and respiratory sounds [4], ECG [5], nasal airway pressure [6] and combinations of various signals [7].

Total sleep time is an important outcome of PSG for diagnosing several sleep disorders. One of these pathologies, in which the authors are currently interested, is the obstructive sleep apnea/hypopnea syndrome (OSAHS). OSAHS is one of the most prevalent sleep disorders[8] and it is characterized by repetitive interruptions of the respiratory flow, caused by pharyngeal collapses during sleep. These upper airway obstructions produce partial or total reduction in the airflow. This syndrome causes increased frequency of awakenings, reduced blood oxygen saturation, sleep fragmentation and, consequently, excessive daytime sleepiness [9]. Furthermore, it is associated with a high risk of acute pulmonary and systemic hypertension, nocturnal arrhythmias, ventricular failure and stroke, cognitive decline, and sudden death [10]. The potential social consequences of this disease, such as accidents, increased morbidity and unproductiveness, among others, make it one of the main public health problems in the world. The number of patients diagnosed and treated for OSAHS has increased drastically in the last few years [11, 12]. Sleep apnea can be easily treated applying a continuous positive airway pressure through the nose using a tight mask [9].

The most important index of OSAHS severity is the apnea/hypopnea index (AHI), which represents the number of apnea/hypopnea events per hour of sleep. The OSAHS is classified as normal, mild, moderate or severe if it belong to the intervals [0,5),[5,15),[15,30) or greater than 30 apnea or hypop-

*Corresponding author. Tel: +54-(0)343-4975100 (122)
Email address: rcasal@conicet.gov.ar (Ramiro Casal)

nea events per hour of sleep, respectively [13]. This implies the need to know if the patient was sleeping (in any stage of sleep) or awake when an respiratory event was detected.

The upper airway obstructions associated with apnea/hypopnea events results in a drop of oxygen saturation levels [14]. Several works has been carried out with the aim of detecting this events using pulse oximeter signals [15, 16]. In these studies, the oxygen desaturation index (ODI) is estimated as an approximation to the AHI. However, it is important to point out that these works do not take into account whether the patient is or not asleep. In some of these works, the ODI was reported as a relation between the number of detected desaturations and the total sleep time (TST) estimated using the EEG, which was previously assumed as not accessible. In some other publications, the total time of the study (TT) was used, which introduces a significant bias (the value of AHI will be underestimated by this approach) [17]. TST estimation from the same signals used to estimate the number of apneas/hypopnea events could improve the reported AHI without increasing the complexity of the study. Being able to estimate the total sleep time from signals obtained with a pulse oximeter will be a great complement to improve these screening devices.

Although we are focused on the diagnosis of apnea, the results of this work may be useful for many other applications. For example, drowsy drivers is an important factor in most traffic accidents. Automatic systems with the goal of detect and prevent sleep are an active research field. Most of them use cameras to assess the level of sleepiness by detection of physiological events related to fatigue and drowsiness [18]. Due to its characteristics, our algorithm can be part of one of these systems and provide complementary information. Daily life applications related with sleep measures from personal health monitoring devices are currently under spotlight [19]. In summary, any critical work in which the sleepiness can cause accidents and material or human losses can benefit from applications such as the one developed in this paper.

In the literature, there are many researchers addressing the automatic sleep staging problem. The state of the art results are obtained using EEG signals, sometimes extracting information from other complementary signals such EOG, EMG, an others. In pursuit of obtain home-based diagnosis devices, there are many studies on automatic sleep staging with signals whose recording and processing is simpler than EEG.

Many authors have studied the dynamic of HR variability, obtaining by processing the ECG, during sleep [20, 21]. These works have made posible that Adnane et al. [22], Xiao et al. [23] and Yücelbaş [24] used ECG to classify the sleep stage. Adnane only considered two classes, awake and asleep, while Yücelbaş and Xiao considered three, awake, REM and non-REM stages. When their results are analyzed, it can be seen that the works that considered longer periods obtained better results than those that used 30-seconds segments, based on the rules published by the AASM [1]. However, considering longer periods is an unrealistic situation since as the length increases. There is a greater probability that the segments contain a mixture of awake and asleep stages. This problem will be addressed in more detail in the discussion section.

There are other works that try to exploit the relationship between HR and sleep stages in the same way as those that use ECG, but using photoplethysmography (PPG). Beattie et al. [25] used PPG signals and accelerometer, considering 5 classes. The database was composed by self-reported normal sleepers. The Uçar et al. research [26] used PPG and heart rate variability (HRV) from PPG. They classified in awake and asleep. An important limitation of all these works is that the size of the databases used is small, so it is difficult to clearly determine their generalization capability.

Motivated by the drawbacks of screening devices for sleep disorders, as well as by the current challenges to estimate sleep measures through mobile and wearable devices, and being inspired by these previous researchers, the aim of this work is to classify the sleep stage in awake (W) or asleep (S), regardless of the corresponding sleep stage. We only have an estimate of the heart rate (HR) from PPG, instead of ECG, which is affected by the lower temporal and frequency resolutions. Further, we use a large database with the intention that the results have the minimum risk of overfitting. The classification will be done applying machine learning techniques. In the feature extraction stage, we use information theory tools, such as dispersion [27], approximate [28], sample [29], fuzzy [30] and Renyi [31] entropies, and methods for frequency and time-frequency analysis [32]. Further, classical statistics were calculated. We suppose these features can be able to discriminate the different dynamics presented in HR series corresponding to awake and sleep stages [20, 21]. Then, we applied a feature selection scheme together with the classification. Finally, the selected system is tested with patients data never used in training.

2. Materials

2.1. Oximetry signals

PPG is an optic measurement technique widely used in both clinic and research. It detects changes in blood volume through a device consisting of a light source and a photodetector. The PPG signal results from the light interaction with biological tissues, namely, the balance between scattering, absorption, reflection, transmission and fluorescence of the signal. Several physiological variables can be estimated directly and indirectly from the PPG signal [33].

The arterial oxygen saturation (SaO_2) is the fraction of saturated hemoglobin relative to total hemoglobin in blood. The pulse oximeters, devices based on PPG, allow a noninvasive estimation of SaO_2, commonly referred as peripheral oxygen saturation (SpO_2), using two light sources (red and infrared) that presents absorption differences due to the hemoglobin presence [34]. The SpO_2 is very useful for the screening of OSAHS, since the number of apnea/hypopnea events can be approximate counting the number of desaturations.

In addition, pulse oximeters provide a HR estimation from the pulsatile component of PPG. The most common algorithms consist of digital filters and zeros crossing detector [33], although there are many research works with the objective of

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

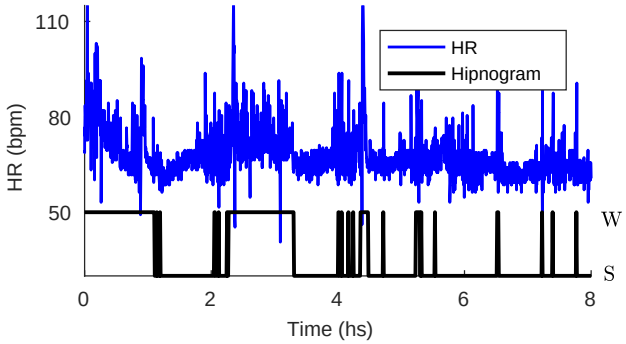

Figure 1: Hypnogram and HR. The states of awake (W) and asleep (S) are shown in the hypnogram (black). The dynamical changes between this states can be noticed in the HR signal (blue).

developing algorithms to reduce the movement artifacts that greatly affect the signal [35].

2.2. HR and sleep stages

The sleep has an orderly internal structure in which different stages are determined. The sleep stages are classified in *wakefulness*, two stages of *light sleep*, two of *deep sleep* and *rapid eye movement sleep* (REM), which are differentiated in the basis of typical patterns and waveforms in signals of EEG, EOC and EMG. These sleep stages are labeled in consecutive 30 seconds long segments. This results in a sleep profile or hypnogram.

The regulation of the autonomic nervous system changes with the sleep stages. In this way, HR, blood pressure, and respiratory rate decrease to adapt a reduced metabolism during sleep. The average HR falls steadily from the waking states to deep sleep. During REM, HR increases lightly and presents greater variability than during wakefulness [20]. The relationship between sleep stages and HR is shown in the Fig. 1. This work is based on these physiological phenomena in order to discriminate the states of awake and asleep.

2.3. Database

The set of biomedical signals used in this article was obtained from the *Sleep Heart Health Study* dataset. This dataset was designed to investigate the relationship between sleep-disordered breathing and cardiovascular consequences. SHHS database is divided into two subsets of PSG records, the SHHS Visit 1 and SHHS Visit 2, obtained several years later with the aim of studying the evolution of patients. The PSG records were acquired automatically at home of patients with supervision of specialized technicians [36]. Full details can be found in [37].

The SHHS contains several signals corresponding to a PSG study collected on twelve channels: SpO_2, HR, chest wall and abdomen movement, nasal/oral airflow, body position, EEG (two central, one for redundancy in case of failure/loss), bilateral EOG, chin EMG and ECG. The oximeter provides two signals, HR and SpO_2, and it also gives a quality status signal that provides information about the sensor connection status. In this work only the HR and the quality status signal are used.

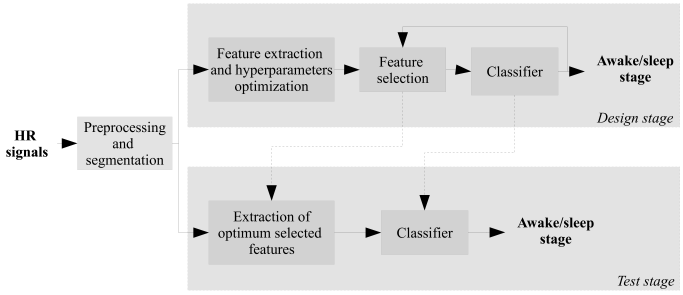

Figure 2: Scheme of the algorithm. In the design stage, the feature extraction and selection is performed. In the test stage, the design system is tested with new data..

In future works we will incorporate the SpO_2 signal to detect apnea/hypopnea events.

In SHHS database, SpO_2 signals have a sampling rate of 1 Hz, resolution of 1% and accuracy of ±2% in the range of 70% to 100%. Their performance significantly decreases for values below this range. The HR signal based on pulse oximeter has a sampling rate of 1 Hz and a precision of 3 beats per minute.

In this work, the SHHS visit 1 was used. Acording to the SHHS 1 Protocol, all the records were processed with a software system to provide preliminary estimates of the AHI. Then, the recordings were manually scored on screen, with annotations of sleep stages, arousals, oxygen desaturation, and respiratory events. Table 1 shows a summary of the characteristics of the database used. 5000 patients were randomly selected to be used in the experiments detailed below. For detail on the sleep stage annotation protocol, refer to [37, 36].

	SHHS 1 (min, max)
n	5804
Age (years)	63.1 ± 11.2 (39.0, 90.0)
Female (%)	52.3%
Epworth sleepiness scale	7.8 ± 4.4 (0.0, 24.0)
Arousal index (/hr)	19.2 ± 10.7 (0.0, 110.4)
AHI (/hr)	9.6 ± 12.7 (0.0, 115.8)
TST (min)	587.7 ± 107.6 (35.0, 858.0)
BMI (kg/m^2)	28.2 ± 5.1 (18.0, 50.0)
TST/TT (%)	74.2%

Table 1: Characteristics of the study population in SHHS Visit 1.

3. Methods

3.1. System overview

The scheme of the algorithm is shown in 2. First, the HR signals from pulse oximeter were preprocessed and segmented into windows of length L for the 5000 patients, as indicated below. SHHS dataset was splitted into two subsets: 500 subjects were randomly selected in order to optimize the design and to select the features (top of figure 2), and the remainig 4500 subjects were used to train and test the classifier (botton of figure 2).

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

In the design stage, a set of features was extracted from each window, optimizing its hyperparameters in order to maximize the area under ROC curve (AUC) [38]. Then, the features were standardized to have zero mean and unit variance. The feature selection was performed along with the classification, using two different schemes: forward feature selection (FFS) and support vector machine (SVM), and variable selection based on random forest (RF). The *k-fold* cross validation technique was used to validate the classifier performance.

Finally, the best system obtained in the design stage was trained and validated with the remaining 4500 subjects using a *k-fold* approach. There is not a formal rule to choose k as long as the size of the database allows to obtain k partitions with a sufficient number of observations to calculate the reliable statistics. We set $k = 10$ for this problem based on recommendations from [39, 40].

3.2. Preprocessing

The pulse oximeter signals available in SHHS dataset provide a complementary signal with information about the state of the oximeter. The status signal was used to mask the HR signal, removing the invalid data. Then, we linearly interpolate between the previous and posterior valid data.

Then, the HR records were standardized in order to reduce inter-subject variability. Then, the signals were segmented into non-overlap windows of length L, considering only the segments corresponding completely to a single state: awake or asleep. The values of L were varied from $L = 30$ to $L = 300$ in steps of 30. No other processing for artifact or noise reduction was used, as our objective is to keep the method as simple as possible, to operate in the raw signal, aiming at low power wearable devices.

3.3. Features

The information contained in the HR signal was summarized into a set of features based on information theory, frequency and time-frequency domain, and classic statistics. In total, 70 features were extracted. A brief description of the most relevant ones is given in this section.

3.3.1. Approximate Entropy

Approximate entropy (ApEn), introduced by Pincus [28], is a measure of data regularity. A greater irregularity in a signal produces a higher ApEn value, and vice versa. For an N-dimensional time series, ApEn depends on three parameters: the embedding dimension m, the embedding delay τ and the threshold r. ApEn has been widely used as a non-linear feature to classify different dynamics.

Let $x[n]$ a time series of length N. Then, $M = N - (m - 1)\tau$ state vectors can be reconstructed doing $\mathbf{x}_i^m = [x[i], x[i + \tau], \ldots, x[i + (m - 1)\tau]]$, where $i = 1, 2, \ldots, M$ [41]. Then, the ApEn is defined as [28]:

$$\mathrm{ApEn}(m, \tau, r, N) = \phi^m(r) - \phi^{m-1}(r), \tag{1}$$

with

$$\phi^m(r) = \frac{1}{M}\sum_{i=1}^{M} \ln \frac{1}{M}\sum_{j=1}^{M} \theta(d(\mathbf{x}_i^m, \mathbf{x}_j^m), r), \tag{2}$$

where $d(\cdot)$ is a distance measure between state vectors and $\theta(\cdot)$ is a kernel function. Usually the distance measure is the Euclidean norm or the Maximum norm. The most used kernel functions are the Heaviside step function [28] and the Gaussian kernel [42].

In this work, the Euclidean norm and Gaussian kernel were used. Three features related to ApEn were extracted. The first one is the value of ApEn estimated with the parameters r, τ and m maximizing the AUC. The other two features are the maximum of ApEn and the value of r where this maximum is located [43]. In these cases, both m and τ were selected in order to maximize the AUC, as before.

3.3.2. Sample Entropy

ApEn is a highly biased estimator due to the inclusion of self-matches, and this bias is more noticeable in short data lengths. To overcome this limitation, Richman and Moorman proposed the Sample Entropy (SampEn). SampEn is largely independent of record length and shows a higher consistency than ApEn [29]. Following a notation similar to that used in ApEn, the equation for determining SampEn is given as

$$\mathrm{SampEn}(m, \tau, r, N) = -\ln \frac{B^{m+1}(r)}{B^m(r)}, \tag{3}$$

with

$$B^m(r) = \frac{1}{M}\sum_{i=1}^{M} \frac{1}{M-1}\sum_{j=1, i\neq j}^{M} \theta(d(\mathbf{x}_i^m, \mathbf{x}_j^m), r), \tag{4}$$

where $d(\cdot)$ is a distance measure between state vectors and $\theta(\cdot)$ is a kernel function. In this work, we chose an Euclidean norm and a Heaviside step function.

For high r thresholds, SampEn $\rightarrow$ ApEn, and for small values of r, SampEn has a high variance. As the authors propose, we considered only the first $N - m$ vectors of length m when computing $B^m(r)$, ensuring that, for $1 \leq i \leq N - m$, $\mathbf{x}_i^m$ and $\mathbf{x}_i^{m+1}$ were defined.

3.3.3. Fuzzy Entropy

Fuzzy entropy (FuzEn) is analogous to SampEn, but the similarity degree (kernel function $\theta(\cdot)$) is calculated through a fuzzy function defined by $e^{-d(\mathbf{x}_i^m, \mathbf{x}_j^m)^q / r}$. According with [30], FuzEn is more consistent and less dependent on the data length than SampEn.

3.3.4. Dispersion entropy

SampEn is a powerful tool to assess the dynamical characteristics of time series, but it is computationally expensive. On the other hand, permutation entropy (PermEn) quantify the irregularity of the time series based on analysis of permutation patterns, which depends on comparisons of neighboring values [44]. PermEn does not consider differences between amplitudes.

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

Dispersion Entropy (DE) was proposed to overcome these limitations of PermEn and SampEn[27]. For a given time series $x[n]$ of length N, the DE algorithm includes the following steps. First, $x[n]$ is assigned to c classes. For this, $x[n]$ is mapped to $y[n]$ from a normal cumulative distribution function. Then, each $y[n]$ is assigned to an integer from 1 to c as $z[n] = \text{round}(cy[n] + 0.5)$ with $n = 1, 2, \ldots, N$. Finally, given an embedding dimension m and delay τ, the state vectors are reconstructed by $\mathbf{z}_i^{m,c} = [z[i], z[i+\tau], \ldots, z[i+(m-1)\tau]]$.

Each time series $\mathbf{z}_i^{m,c}$ is mapped to a dispersion pattern $\pi_{v_0,v_1,\ldots,v_{m-1}}$, where $z^m[i] = v_0$, $z^m[i+\tau] = v_1$, $\ldots$, $z^m[i+(m-1)\tau] = v_{m-1}$. Then, for each of c^m potencial dispersion pattern, relative frequency is obtained by $p(\pi_{v_0,v_1,\ldots,v_{m-1}})$. Finally, based on the definition of the Shannon entropy, DE is calculated by

$$\text{DE}(x, m, c, \tau) = -\sum_{\pi=1}^{c^m} p(\pi_{v_0,v_1,\ldots,v_{m-1}}) \ln p(\pi_{v_0,v_1,\ldots,v_{m-1}}). \quad (5)$$

3.3.5. Extension of entropy measures to the joint time-frecuency (t, f) or time-scale (t, s) domains

A measure of the uniformity of signal energy distribution in the frequency domain can be defined by interpreting a power spectral density (PSD) as a quasi-probability distribution function [45], and using entropy concepts. A larger f-domain entropy value implies more uniformity and vice versa. These concepts can be extended to (t, f) or (t, s) domains in order to discriminate signals with similar bandwidth, but with different variations over time [46]. Let $x[n]$ be a real time series of length N and $z[k]$ its Fourier Transform (FT) of length M. The spectral entropy ($\text{SE}_{(f)}$) is defined as

$$\text{SE}_{(f)} = -\sum_{k=1}^{M} \mathcal{Z}[k] \ln \mathcal{Z}[k], \quad (6)$$

where $\mathcal{Z}[k] = \frac{|z[k]|^2}{\sum_{k=1}^{M} |z[k]|^2}$.

The (t, f) or (t, s) Shannon entropy is an extension of the ($\text{SE}_{(f)}$). It is obtained by replacing the FT with a time-frequency or time-scale distribution $\rho[n, k]$. The (t, s) Shannon entropy is

$$\text{SE}_{(t,s)} = -\sum_{n=1}^{N}\sum_{k=1}^{M} \rho_N[n, k] \ln\left(\rho_N[n, k]\right), \quad (7)$$

where $\rho_N[n, k] = \frac{\rho[n,k]}{\sum_n \sum_k \rho[n,k]}$.

These ideas can also be used to extend the Renyi Entropy as

$$\text{RE}_{(t,s)} = \frac{1}{1-q} \ln \sum_{n=1}^{N}\sum_{k=1}^{M} \rho_N[n, k]^q. \quad (8)$$

In this work, we set $q = 3$ [46].

Inspired by these concepts and multiresolution entropy [47], we propose a measure of entropy through the scales for this application. For each scale k of the (t, s) representation $\rho[n, k]$, we calculate an entropy value by

$$\text{RE}_{(t,s)}(k) = \frac{1}{1-q} \ln \sum_{n=1}^{N} \left(\frac{\rho[n, k]}{\sum_n \rho[n, k]}\right)^q. \quad (9)$$

This entropy measure allows to estimate the uniformity of signal energy distribution for a simple scale. This is maximum when the energy is constant over the time.

We normalize $\rho[n, k]$ in order to fulfill the probability density function properties. Thus, the entropy measure does not vary with the magnitude of energy, but is only based on its distribution over time. As in the previous case, we set $q = 3$. From now on, these features will be called *time-scale multiresolution Renyi entropy* (TSMRE).

The time-scale distribution was obtained using continuous wavelet transform (CWT) using 32 scales and Haar's wavelet. Although this wavelet is very simple, and it has the disadvantage of being discontinuous and therefore not derivable, for this work it was the wavelet that showed the best performance. This is due to the fact that the HR signal provided by the pulse oximeter has a large quantization step and thus can be well approximated by piecewise constant functions.

3.3.6. Lempel-Ziv Complexity

Lempel-Ziv complexity (LZ) [48] is a metric that has been widely used in biological signals for recognition of structural regularities. The LZ is nonparametric and simple to compute. First, the discrete-time signal $x[n]$ is converted into a symbol sequence $P = s_1 s_2 \ldots s_n$ by comparison with thresholds. Then, complexity measure $c[n]$ can be calculated as referenced in [49]. In the context of biomedical signal analysis, typically the signal is converted into a binary sequence using the median. In this work, we obtain better performance setting two thresholds using 0.33 and 0.66 quantiles. That is, we use a sequence of three different symbols.

The features described previously allow to quantify the regularity of the data. As we mentioned in the subsection 2.2, the HR shows greater variability and mean value during wakefulness. From this, these measures were proposed in order to exploit this difference in the regularity or complexity of the data. These measures have proven useful in several works on biomedical signals [50]. Further, our previous work [51] showed the potential of these features for classification.

3.3.7. Frequency domain based features

Spectral analysis of the heart rate variability signals allows to quantify the influence of the autonomic nervous system [52]. These characteristics are selected here on the assumption that they will provide information about the sleep stage, as discussed in [53]. Very low frequency (VLF) $(0.003, 0.04]$ Hz, low frequency (LF) $(0.04, 0.15]$ Hz and high frequency (HF) $(0.15, 0.4]$ Hz components were obtained. Further, the ratio LF/HF, normalized HF and VF and total power (TP) are used [52]. Spectral estimation is performed using periodogram.

3.3.8. (t, f) or (t, s) signal-based features

We extracted the 15 features discussed in [54] that allow characterize the non-stationary nature of the HR. This capability is potentially useful for discriminating sleep stages. The features are briefly explained below.

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

Several features are based on singular value decomposition (SVD) of the (t, f) or (t, s) representation $\rho[n, k]$. The SVD divides the $N \times M$ matrix ρ into two subspaces, signal subspace and an orthogonal alternate subspace of the form $\rho = \mathbf{USV}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are $N \times N$ and $M \times M$ unitary matrices, respectively. $\mathbf{S}$ is an $N \times M$ diagonal matrix with non-negative real numbers. The diagonal entries of $\mathbf{S}$ are known as the singular values of ρ.

The first and second features are the maximum ($\max \mathbf{S}$) and variance ($\mathrm{var}\,\mathbf{S}$) of the singular values of ρ. The third feature is a complexity measure given by

$$\mathrm{E_{SVD}} = -\sum_{i=1}^{N} \bar{S}_i \log \bar{S}_i, \tag{10}$$

where $\bar{S}_i = \frac{S_i}{\sum_{j=1}^{N} S_j}$.

The fourth feature is the energy concentration measure, and it is defined as

$$\mathrm{ECM} = \left(\sum_{n=1}^{N} \sum_{k=1}^{M} |\rho[n, k]|^{\frac{1}{2}} \right)^2. \tag{11}$$

Then, 8 features related to the sub-band energy are obtained integrating over time. That is

$$\mathrm{SBE}_\delta = \sum_{n=1}^{N} \sum_{k=\delta L+1}^{(\delta+1)L} \rho[n, k] \tag{12}$$

where $L = \frac{M}{8}$ and $\delta = 0, 1, \ldots, 7$.

Finally, in addition to these features, we calculated (t, s)-domain mean and standard deviation by extension from classical statistics and (t, s) Renyi entropy by equation 8.

3.3.9. Autocorrelation-based features

We extracted some features related to the autocorrelation serie. In general, the autocorrelation of awake segments has periodicities and is smoother than asleep segments. In order to differentiate them, we compute the first minimum and the first zero-crossing of the serie. Then, we determine the coefficients of an autorregresive (AR) model of order 4 that fits the signal. These coefficients were used as features. Finally, we calculate the LZ complexity of the autocorrelation serie in order to measure the regularity differences.

3.3.10. Statistical features

Further to the above-mentioned features, some classical statistics such as mean and standard deviation of the temporal signal were calculated. These features are useful for discriminate sleep stages (especially the mean value), since they vary markedly between sleep and wakefulness.

3.3.11. Summary of features

The obtained features can be summarized in:

- 7 variants of entropy and complexity measures: ApEn, $\mathrm{ApEn}_{\max}$ and $r_{\max}$, SampEn, FuzEn, DispEn, and LZ.
- 32 entropy measures in the (t, s) domain: TSMRE.
- 7 frequency domain features: VLF, LF, HF, LF/HF, normalize HF and VF, and TP.
- 15 features in the (t, s) domain: $\max \mathbf{S}$, $\mathrm{var}\,\mathbf{S}$, $\mathrm{E_{SVD}}$, ECM, 8 SBE_δ, $\mathrm{mean}_{(t,s)}$, $\mathrm{std}_{(t,s)}$, and $\mathrm{RE}_{(t,s)}$.
- 7 autocorrelation-based features: first min, first ZC, 4 AR coefficients, and LZ of autocorrelation.
- 2 statistical features: mean and standard deviation.

3.4. Feature selection and classification

As mentioned above, a total of 70 features were obtained. However, system performance may vary with different combinations of features. In addition to potentially worsening performance, the presence of redundant or non-informative features increases the computational cost and makes the classifier harder to train by the "curse of dimensionality". A feature selection routine is a popular way to resolve this problems. In this work, we propose two different schemes for the feature selection. The first scheme is a *wrapper method* [55] of feature selection and SVM and the second scheme is an *embedding method* that uses RF as classifier.

3.4.1. Forward feature selection and SVM

We use a measure of classifier performance to select the optimal subset of features. According to Kohavi and John [56], it is necessary to define how to search the space of all possible variable subsets; what performance measure to use to guide the search; and which classifier to use. An exhaustive search find the global optimum, but the problem is NP-hard. To avoid this problem, we implemented a greedy solution called *forward feature selection*.

Let $\mathcal{S}_f^n$ and $\mathcal{R}_f^n$ be the optimal feature set selected and the remaining features, respectively, in the n-th iteration. Let N_f the number of total features and ε an error measure. Let $\mathcal{S}_f^1 = \emptyset$ and $\mathcal{R}_f^1$ the set of all features. The procedure can be summarized as follows:

1: **for** $n = 1$ **to** N_f **do**
2: **for** $R_f \in \mathcal{R}_f^n$ **do**
3: **Provisional set of features:** $\mathcal{P}_{R_f} = \mathcal{S}_f^n \cup R_f$
4: **Cross validation:** Randomly split data set into K parts.
5: **for** $k = 1$ **to** K **do**
6: Select k-th subset for **testing** and the rest for **training**.
7: **Train** classifier with train set.
8: **Test** classifier with test set.
9: Compute the **errors** $\varepsilon(k)$ for all k.
10: **end for**
11: Compute the mean error $\bar{\varepsilon}(R_f)$ for all $\mathcal{P}_{R_f}$.
12: **end for**
13: Find a minimizer $R_f^* = \arg\min_{R_f} \bar{\varepsilon}(R_f)$.
14: Update **best set of selected features** $\mathcal{S}_f^{n+1} = \mathcal{S}_f^n \cup R_f^*$.
15: Update **set of remaining features** $\mathcal{R}_f^{n+1} = \mathcal{R}_f^n - \{R_f^*\}$.
16: Save the best result $\varepsilon_{\min}(n) = \bar{\varepsilon}(R_f^*)$ for all n.

17: **end for**

In this algorithm, features are progressively incorporated into larger subsets. This method yields nested subsets of features. Finally, an error measurement is obtained for each subset and from this, the "optimum" set can be selected. While the computational load is less than in an exhaustive search, reaching the solution by this method may be slow.

In this scheme we use SVM as classifier [57]. SVM involves the optimization of a convex objective function with constraints and it is unaffected by local minima. SVM produces an optimum separation hyperplane through mapping the features in a hyperdimensional space. In this work, we use a Gaussian kernel given by $K(x_i, x_j) = \exp\{-\gamma\|x_i - x_j\|^2\}$. Detailed explanation of SVM can be found in [57, 58].

The measure error ε was selected with the aim of maximize the AUC. Let Se and Sp be the sensitivity and specificity, respectively. Each iteration selects the feature that minimizes $(1 - \mathrm{Se})^2 + (1 - \mathrm{Sp})^2$, that is, the minimum distance to point $(0, 1)$ in the ROC curve. In addition, we evaluated the use of a criterion related with the application. We selected the feature that minimize the error in estimating the total sleep time per patient. However, the results obtained with this last measure will not be reported due to poor performance.

3.4.2. Feature importance and RF

For comparison purposes, we consider RF [59] as classifier. In RF, we can incorporate the feature selection as part of the training process and that is much more efficient because there is no need to retrain several times. To do this, after each tree is trained with all features, the values of each feature are randomly permuted. The data with the permuted variable is run down in the tree and a measure of error is calculated. By doing this for all features and calculating the percentage increase in misclassification rate compared to the resulting error with all variables intact, we can estimate a measure of variable importance.

3.5. Final classifiers

We already described the feature extraction and feature selection. The final stage is train and test the classifiers that use the subsets of selected features. However, first we need to solve one more problem: the class imbalance.

Most of the total recording time corresponds to the asleep stage. There are different approaches to prevent the classifier from biasing towards the majority class, such as resample database, synthetic samples generation by convex combination, among others. In this work, we used two different methods to overcome the drawback of class imbalance. The first method is naive: we simply randomly remove samples from the majority class until the classes are balanced. This method was applied in FFS-SVM and RF with feature importance. In the second method, we use the SVM classifier but we impose an additional cost on the model for the minority class errors during training [58]. In this way, we can bias the classifier to pay more attention to the minority class. We penalized the classification in a ratio that takes into consideration the class imbalance.

These balanced strategies were only applied in the training to avoid the classifier to have a bias towards the majority class. In the test data, no processing is done to balance the classes. In this way the classifier will be tested under conditions similar to the real application.

In conclusion, taking into account the feature selection and classification routines along with the approaches to solve class imbalance, we have three algorithms that will be trained and tested: FFS-SVM with penalty errors in minority class (FS 1), FFS-SVM with artificial balance (FS 2), and feature importance and RF with artificial balance (FS 3).

4. Results

In this section, we present the results obtained with the three proposed methods, explaining the outcomes of each of the stages that compose it.

4.1. Parameters selection

In some features described in previous section was necessary to experimentally tune a set of hyperparameters. To find the best combination of hyperparameters a set of experiments was performed over a design database (500 patients, which was describen in 3.1). All hyperparameters combinations were explored using a grid search. First, we did a coarse search to get an idea of how the features behaves according to the hyperparameters and, then, we did a more detailed search to find the optimum hyperparameters. We use the AUC as an objective measure of the discriminating capacity of each single feature[38].

As mentioned above, these experiments were conducted for HR segments with a duration L of between 30 and 300 seconds, in steps of 30. The grid search was performed for the features ApEn, SampEn, FuzEn and DispEn. In the first three features, m was varied from 2 to 8 in steps of 1, r was varied from $\exp(-7)$ to $\exp(4)$ varying the exponent in steps of 0.55 and τ was varied from 1 to 4 in steps of 1. For FuzEn, the exponent q of the kernel function was varied from 2 to 4 in steps of 1. In DispEn, c was varied from 2 to 4 in steps of 1, m was varied from 2 to $m_{\max}$ in steps of 1, where $m_{\max} = \mathrm{floor}(\log(L)/c)$ and τ from 1 to 4 in steps of 1. The Gaussian kernel parameter used for SVM was selected varying γ with the equation $1/k^2$, with k from 1 to 16 in steps of 2. The optimum value was $k = 8$, and consequently $\gamma = 0.0156$.

The hyperparameters that maximize the AUC are very similar for different values of L, but the AUC always increases with L. The consequences of this will be discussed below.

4.2. Feature selection

In order to discard irrelevant features and generate a set of optimum features for a given classifier, the feature selection routines mentioned in Section 3 were applied. The methods FS 1, FS 2 and FS 3 were applied to the different databases obtained by varying L. The two methods of FFS-SVM were applied as explained above, obtaining 70 nested subsets of features. In the RF-based method, we obtain a feature ranking. In this case,

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

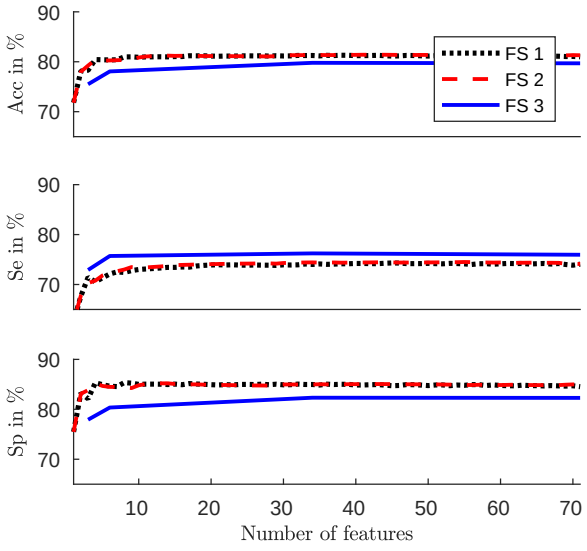

Figure 3: Performance (Acc, Se and Sp) versus number of features with FFS-SVM with penalty errors in minority class (black), FFS-SVM with artificial balance (red), and feature importance and RF with artificial balance (blue).

we did not consider increase the features of one by one, but we generated subsets considering groups whose ranking is similar.

The original feature set had high redundancy. Many features are from the same family. The experiment results showed that with only a few features (approximately 10) performance is close to the best result obtained. The figure 3 shows accuracy, sensitivity and specificity for database with $L = 150$ obtained with three methods. Results are very similar for the others L values.

Based on these results, we selected the first 40 features obtained in each method for each database to make the classifier faster and simpler. In the tables 2, 3 and 4 we show an exhaustive list of the selected features by each method. The selected features are generally very similar by each method, although the order in which they are selected are slightly different by each one. One of the most selected features is the mean value of the signal. The mean value of the HR decreases as the sleep stage is deeper, as it was established by Penzel [20]. The ApEn related features were also selected early. Being a measure of regularity, it can be able to differentiate the states of awake and asleep. During vigil there is associated a greater irregularity of the signal [20], as shown in the figure 1. There are other features in which it is more difficult to determine a physiological meaning directly. However, it is evident that the signal changes, and the features that we are using are able to reflect those changes. Although the interpretability of the results is always a benefit, it is not necessary that the selected features have a logical interpretation with the application, proof of this are works that use weak classifiers or the deep learning approaches [60].

4.3. Performance in unseen database

In order to evaluate the performance of the developed systems, we applied the three algorithms (FS1, FS2, and FS3) to the remaining unseen 4500 patients and we performed *k-fold* cross validation, with $k = 10$. We calculated all performances individually (per patient) and the reported results were obtained by averaging. In this way, the result obtained will be closer to the real application where the objective is to estimate the total sleep time per patient.

Table 5 shows the percentage confusion matrix obtained by applying the algorithms to the unseen database. The number of false positives (FP), true positives (TP), false negatives (FN) and true negatives (TN) was calculated for each patient and the averaged values are reported. In all algorithms a better performance was obtained in the classification of the majority class (when the patient is asleep). The performance also increases with the length of the segment L.

In table 6, we summarized some common performance measures: accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). These measures was reported for all three algorithms applied to databases obtained varying L. SVM-based algorithms are more sensitive and RF is more specific. It is worth noting that in this work we take the awake state as a positive class. The better accuracy of the classifiers that use SVM is explained by the fact that they classify the majority class better, which is the one that ultimately has the greater importance in this measure.

5. Discussion

The use of pulse oximetry as the only signal to diagnose OSAHS is still controversial. Different devices present a great variability for different situations and patients. The dispersion of obtained values is high. The interpretation of the measured desaturations is ambiguous without prior knowledge of the device used [61]. In addition, pulse oximetry is highly sensitive to motion artifacts [33]. However, significant results have been obtained through the use of advancing signal processing tools. As mentioned earlier, this makes the use of PPG very attractive because of its simplicity and low cost, but the algorithms designed have great difficulty in meeting its objectives due to the low quality of the signal.

Sleep classification is performed primarily using the EEG signal, although information is also extracted from other signals such as EOG, EMG to detect movement, among others. Thus, detecting whether a signal segment corresponds to awake or asleep from the HR is a very complicated task. Although there are changes in the temporal dynamics of the HR signal [20], the information we have is much more difficult to interpret. If we add to this the low accuracy of the HR reported by the pulse oximeter, we can get an idea of the difficulty in reaching an acceptable performance. Finally, in addition to all these drawbacks, we also have a low sampling frequency of HR, which makes it difficult to obtain an adequate feature extraction.

In this work we have developed an algorithm to classify the sleep stages in awake and asleep using only HR signals obtained by pulse oximetry. To summarize, the main methodological steps that must be applied to reuse the proposed scheme with other databases is as follows:

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

L	FS 1
30	mean, ApEn, TSMRE_5, LF, LF/HF, $\max \mathbf{S}$, $r_{\max}$, AR coefficient 4, TSMRE_{26}, TSMRE_8, $\text{mean}_{(t,s)}$, HF, LZ of AC, FuzEn, std, TP, TSMRE_7, SBE_2, $\text{var}\,\mathbf{S}$, VLF, DispEn, $\text{RE}_{(t,s)}$, SBE_6, TSMRE_{27}, TSMRE_{32}, SBE_1, TSMRE_{29}, TSMRE_{31}, TSMRE_1, TSMRE_{25}, first AC min, E_{SVD}, LZ, TSMRE_{30}, AR coefficient 1, TSMRE_6, TSMRE_{23}, TSMRE_{13}, TSMRE_{15}, TSMRE_{14}.
60	mean, ApEn, ApEn_{max}, first ZC, normalize HF, std, TSMRE_5, TSMRE_17, $r_{\max}$, SBE_1, SBE_8, TSMRE_{32}, TSMRE_{24}, E_{SVD}, SBE_6, TSMRE_{16}, $\text{std}_{(t,s)}$, TSMRE_{25}, $\text{var}\,\mathbf{S}$, FuzEn, LF/HF, $\text{mean}_{(t,s)}$, TSMRE_{18}, SBE_3, VLF, TP, LZ, TSMRE_{19}, SampEn, HF, TSMRE_{14}, TSMRE_{23}, TSMRE_{21}, LZ of AC, SBE_7, TSMRE_{12}, TSMRE_{22}, TSMRE_{15}, TSMRE_{13}, TSMRE_{27}.
90	mean, ApEn, ApEn_{max}, first ZC, std, TSMRE_3, TSMRE_1, SBE_8, E_{SVD}, SBE_4, TSMRE_{27}, SampEn, TSMRE_4, $r_{\max}$, normalize LF, LF/HF, SBE_5, LF, SBE_6, TSMRE_2, TSMRE_{27}, AR coefficient 3, VLF, TP, LZ of AC, TSMRE_{13}, FuzEn, normalize HF, TSMRE_{14}, HF, $\text{mean}_{(t,s)}$, TSMRE_9, TSMRE_{12}, SBE_3, SBE_2, TSMRE_5, SBE_1, SBE_7, TSMRE_{19}, TSMRE_6.
120	ApEn, mean, first ZC, ApEn_{max}, std, TSMRE_1, SBE_5, TSMRE_2, SampEn, E_{SVD}, ECM, SBE_4, TSMRE_{16}, SBE_1, LZ of AC, TSMRE_{15}, first AC min, LF, TSMRE_{20}, TSMRE_{21}, HF, $\text{std}_{(t,s)}$, TSMRE_{22}, TP, VLF, TSMRE_{24}, TSMRE_{17}, FuzEn, TSMRE_{18}, TSMRE_{19}, TSMRE_{27}, $\text{mean}_{(t,s)}$, TSMRE_{26}, TSMRE_{23}, TSMRE_{32}, SBE_6, AR coefficient 4, TSMRE_{14}, SBE_3, $\max \mathbf{S}$.
150	TSMRE_3, mean, TSMRE_5, ApEn_{max}, ApEn, E_{SVD}, $\text{var}\,\mathbf{S}$, first ZC, SBE_6, TSMRE_1, TSMRE_2, ECM, SBE_2, HF, SBE_8, $r_{\max}$, TSMRE_9, SBE_4, LZ, TSMRE_7, LZ of AC, TSMRE_{12}, FuzEn, TSMRE_{22}, TSMRE_8, SampEn, TSMRE_6, SBE_3, normalize LF, std, TSMRE_{14}, TSMRE_{17}, TSMRE_{15}, TSMRE_{18}, VLF, TSMRE_5, TSMRE_{10}, normalize HF, $\max \mathbf{S}$, TSMRE_{31}.
180	TSMRE_3, mean, first ZC, SBE_1, LF, SBE_4, ApEn_{max}, TSMRE_1, E_{SVD}, ECM, ApEn, TSMRE_{13}, SBE_8, TSMRE_{15}, std, SBE_2, FuzEn, SBE_7, SBE_5, VLF, TSMRE_2, LZ of AC, TP, $\text{std}_{(t,s)}$, TSMRE_{18}, SBE_6, SBE_3, TSMRE_{29}, HF, normalize HF, $\text{mean}_{(t,s)}$, $\max \mathbf{S}$, $\text{var}\,\mathbf{S}$, AR coefficient 2, TSMRE_{31}, TSMRE_{27}, normalize LF, LZ, TSMRE_{14}, TSMRE_{24}.
210	TSMRE_3, mean, LZ of AC, ApEn_{max}, SBE_5, TSMRE_1, std, AR coefficient 4, TSMRE_{11}, E_{SVD}, ECM, FuzEn, LF/HF, SBE_4, TSMRE_{13}, LZ, first ZC, TSMRE_2, SBE_3, TSMRE_{12}, SBE_7, SBE_2, TSMRE_{15}, $\text{var}\,\mathbf{S}$, TSMRE_8, SampEn, first AC min, SBE_6, $\max \mathbf{S}$, SBE_8, TSMRE_{14}, TSMRE_5, TSMRE_{10}, $r_{\max}$, TSMRE_{19}, TSMRE_6, TSMRE_{16}, $\text{RE}_{(t,s)}$, LF, $\text{std}_{(t,s)}$.
240	TSMRE_3, mean, first ZC, SBE_8, ApEn_{max}, TSMRE_1, E_{SVD}, TSMRE_{17}, LF/HF, std, TSMRE_{31}, TSMRE_2, AR coefficient 4, TSMRE_{28}, SBE_3, LZ of AC, TSMRE_{13}, $\text{RE}_{(t,s)}$, SBE_6, TSMRE_{15}, SBE_2, $\text{std}_{(t,s)}$, SampEn, SBE_7, VLF, TSMRE_30, ECM, normalize LF, ApEn, $\text{mean}_{(t,s)}$, TP, TSMRE_4, SBE_4, SBE_5, HF, FuzEn, SBE_1, TSMRE_{18}, TSMRE_{16}, normalize HF.
270	TSMRE_3, mean, LZ of AC, std, SBE_5, ApEn_{max}, first ZC, TSMRE_1, TSMRE_{10}, LZ, AR coefficient 2, E_{SVD}, ECM, AR coefficient 1, SBE_4, AR coefficient 4, TSMRE_4, TSMRE_2, SBE_2, SBE_8, TSMRE_{32}, TSMRE_{30}, SBE_7, $\text{var}\,\mathbf{S}$, LF/HF, SBE_6, SBE_3, TSMRE_{22}, VLF, DispEn, TSMRE_{27}, TSMRE_{25}, TSMRE_{31}, SBE_1, first AC min, TP, TSMRE_{19}, AR coefficient 3, TSMRE_{29}, TSMRE_{20}.
300	TSMRE_3, mean, LZ of AC, ApEn_{max}, TSMRE_1, SBE_5, AR coefficient 4, std, TSMRE_2, LF/HF, E_{SVD}, ECM, TSMRE_{10}, DispEn, normalize LF, SampEn, TSMRE_{16}, TSMRE_{12}, TSMRE_{30}, HF, AR coefficient 2, TSMRE_{32}, SBE_6, TSMRE_{25}, TSMRE_{15}, AR coefficient 3, $\text{mean}_{(t,s)}$, TP, TSMRE_4, TSMRE_8, $\text{std}_{(t,s)}$, TSMRE_9, AR coefficient 1, SBE_3, VLF, TSMRE_{18}, TSMRE_{11}, $\text{var}\,\mathbf{S}$, LF, SBE_2.

Table 2: Feature ranking. Top 40 features for different L using the algorithm FS 1.

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

L	FS 2
30	mean, ApEn, first ZC, LF, TSMRE_{10}, AR coefficient 4, TSMRE_{30}, $r_{\max}$, normalize LF, SBE_7, SBE_1, TSMRE_4, $\text{std}_{(t,s)}$, AR coefficient 1, std, SBE_2, $\text{RE}_{(t,s)}$, TSMRE_{29}, TSMRE_8, VLF, TSMRE_3, LZ, max $\mathbf{S}$, TP, TSMRE_{27}, AR coefficient 3, SBE_6, DispEn, TSMRE_{23}, TSMRE_2, TSMRE_{21}, TSMRE_{12}, SampEn, FuzEn, TSMRE_{16}, TSMRE_{17}, var $\mathbf{S}$, first AC min, $\text{ApEn}_{\max}$, E_{SVD}.
60	mean, ApEn, $\text{ApEn}_{\max}$, TSMRE_3, SBE_5, std, $r_{\max}$, first ZC, normalize HF, TSMRE_8, TSMRE_{15}, TSMRE_{18}, SBE_4, TSMRE_{23}, DispEn, TSMRE_{20}, AR coefficient 1, $\text{mean}_{(t,s)}$, TSMRE_{28}, max $\mathbf{S}$, TSMRE_7, VLF, SBE_2, first AC min, TSMRE_{11}, SampEn, TSMRE_{12}, TSMRE_{14}, $\text{RE}_{(t,s)}$, TSMRE_{31}, LF/HF, SBE_6,TSMRE_9, TSMRE_{32}, LZ of AC, TSMRE_{16}, E_{SVD}, TSMRE_1, FuzEn, LF.
90	mean, ApEn, $\text{ApEn}_{\max}$, first ZC, TSMRE_6, std, SBE_4, E_{SVD}, TSMRE_{11}, TSMRE_3, SBE_6, SBE_2, DispEn, TSMRE_{20}, LZ of AC, TSMRE_2, var $\mathbf{S}$, AR coefficients 1, TSMRE_5, SBE_3, max $\mathbf{S}$, $\text{mean}_{(t,s)}$, TSMRE_{24}, TSMRE_{15}, TSMRE_{17}, $\text{RE}_{(t,s)}$, TSMRE_{16}, LZ, TSMRE_{23}, TSMRE_{13}, TSMRE_{30}, TSMRE_{19}, normalize HF, first AC min, AR coefficients 4, $r_{\max}$, AR coefficients 2, ECM, TSMRE_8, TSMRE_{14}.
120	TSMRE_3, mean, first ZC, SBE_1, normalize LF, SBE_7, TSMRE_{18}, $\text{ApEn}_{\max}$, E_{SVD}, TSMRE_{28}, std, TSMRE_1, FuzEn, SBE_5, TSMRE_2, first AC min, $r_{\max}$, TSMRE_{27}, SBE_6, $\text{mean}_{(t,s)}$, SBE_4, TSMRE_{26}, TSMRE_{21}, SBE_8, TSMRE_{31}, TSMRE_{29}, TSMRE_{22}, LF, LZ, $\text{std}_{(t,s)}$, var $\mathbf{S}$, TSMRE_{25}, TSMRE_{15}, TSMRE_8, LZ of AC, TSMRE_{20}, ECM, AR coefficients 2, AR coefficients 1, TSMRE_{24}.
150	TSMRE_3, mean, first ZC, $\text{ApEn}_{\max}$, SampEn, SBE_5, E_{SVD}, TSMRE_1, TSMRE_{29}, SBE_6, SBE_1, normalize HF, HF, LZ, LZ of AC, std, TSMRE_2, ApEn, TSMRE_{16}, first AC min, LF, var $\mathbf{S}$, TSMRE_{13}, TSMRE_{20}, TSMRE_{31}, TSMRE_{17}, LF/HF, TSMRE_{24}, SBE_4, ECM, SBE_2, normalize LF, FuzEn, TSMRE_6, SBE_3, $\text{RE}_{(t,s)}$, TSMRE_{15}, $r_{\max}$, SBE_8, TSMRE_{17}.
180	TSMRE_3, mean, first ZC, SBE_1, VLF, SBE_5, $\text{ApEn}_{\max}$, FuzEn, SBE_4, E_{SVD}, ECM, SBE_2, TSMRE_1, std, TSMRE_{29}, TSMRE_{11}, normalize HF, DispEn, SBE_3, LZ of AC, TSMRE_{30}, TSMRE_{19}, SBE_8, TSMRE_{17}, SampEn, TSMRE_2, SBE_6, first AC min, LF, SBE_7, $\text{std}_{(t,s)}$, var $\mathbf{S}$, LZ, AR coefficients 1, TSMRE_{13}, TSMRE_{24}, max $\mathbf{S}$, TSMRE_{10}, TSMRE_{23}, HF.
210	TSMRE_3, mean, LZ of AC, $\text{ApEn}_{\max}$, LF/HF, TSMRE_1, SBE_6, $\text{mean}_{(t,s)}$, ECM, AR coefficients 4, FuzEn, TSMRE_{12},TSMRE_{23}, SBE_7, E_{SVD}, TSMRE_2, TSMRE_{28}, std, TSMRE_{17}, TSMRE_{17}, SampEn, TSMRE_{18}, $r_{\max}$, SBE_8, LF, VLF, TSMRE_{27}, SBE_5, TSMRE_4, TSMRE_{30}, first ZC, TSMRE_8, TSMRE_{15}, TP, SBE_4, $\text{std}_{(t,s)}$, LZ, SBE_3, TSMRE_{16}, TSMRE_5, ApEn.
240	TSMRE_3, mean, LZ of AC, SBE_5, std, $\text{ApEn}_{\max}$, LF/HF, TSMRE_1, E_{SVD}, TSMRE_{26}, TSMRE_2, LF, TSMRE_{30}, first ZC, TSMRE_{13}, TSMRE_{19}, ApEn, SBE_6, TSMRE_{31}, SBE_2, TSMRE_{24}, TSMRE_{10}, LZ, SBE_8, TSMRE_{11}, ECM, SampEn, AR coefficients 1, TSMRE_{15}, TSMRE_{28}, TSMRE_{29}, var $\mathbf{S}$, FuzEn, AR coefficients 2, max $\mathbf{S}$, SBE_7, TP, TSMRE_{14}, TSMRE_{20}, VLF.
270	TSMRE_3, mean, LZ of AC, SBE_5, std, $\text{ApEn}_{\max}$, E_{SVD}, TSMRE_{11}, $\text{RE}_{(t,s)}$, first ZC, ECM, SBE_4, TSMRE_2, ApEn, TSMRE_{14}, TSMRE_9, LZ, AR coefficients 2, LF/HR, AR coefficients 1, SBE_2, TSMRE_{20}, TP, std, SampEn, TSMRE_{24}, TSMRE_{25}, TSMRE_{17}, TSMRE_{31}, SBE_3, SBE_6, max $\mathbf{S}$, TSMRE_7, TSMRE_{16}, SBE_7, TSMRE_{12}, TSMRE_{11}, TSMRE_{13}, AR coefficients 3, first AC min.
300	TSMRE_3, mean, LZ of AC, TSMRE_1, $\text{ApEn}_{\max}$, SBE_5, E_{SVD}, TSMRE_2, LF/HF, ECM, FuzEn, AR coefficients 4, TSMRE_{14}, LF, first ZC, TSMRE_{16}, SBE_4, TSMRE_{26}, SBE_7, TSMRE_{29}, TSMRE_{19}, normalize LF, TSMRE_{27}, SBE_2, AR coefficients 2, TSMRE_8, DispEn, SBE_6, $\text{std}_{(t,s)}$, AR coefficients 1, AR coefficients 3, TSMRE_{13}, LZ, var $\mathbf{S}$, TSMRE_{32}, VLF, SBE_8, TSMRE_6, SBE_1, ApEn.

Table 3: Feature ranking. Top 40 features for different L using the algorithm FS 2.

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

L	FS 3
30	mean, HF, LF, ApEn_{max}, TP, std, $\text{mean}_{(t,s)}$, VLF, TSMRE_6, TSMRE_5, TSMRE_7, TSMRE_8, TSMRE_4, TSMRE_9, TSMRE_{10}, TSMRE_3, TSMRE_2, ApEn, ECM, normalize LF, E_{SVD}, var $\mathbf{S}$, SBE_1, TSMRE_1, LF/HF, max $\mathbf{S}$, SBE_4, FuzEn, TSMRE_{12}, normalize HF, SBE_5, SBE_8, TSMRE_{27}, TSMRE_{11}, SampEn, AR coefficients 1, TSMRE_{26}, TSMRE_{28}, SBE_7, TSMRE_{25}.
60	mean, ApEn_{max}, TSMRE_1, HF, E_{SVD}, ECM, AR coefficients 1, TSMRE_2, TSMRE_{19}, TSMRE_{16}, $\text{mean}_{(t,s)}$, TSMRE_{18}, SBE_1, TSMRE_{17}, LF, $\text{RE}_{(t,s)}$, var $\mathbf{S}$, max $\mathbf{S}$, TP, TSMRE_{20}, normalize LF, TSMRE_{15}, VLF, SBE_5, std, TSMRE_{14}, TSMRE_{21}, normalize HF, $\text{std}_{(t,s)}$, LF/HF, TSMRE_{22}, TSMRE_{25}, TSMRE_4, TSMRE_{24}, SBE_2, TSMRE_{23}, SampEn, r_{max}, SBE_8, TSMRE_{13}.
90	mean, ApEn_{max}, TSMRE_2, TSMRE_3, AR coefficients 1, ECM, first ZC, E_{SVD}, HF, SBE_1, normalize LF, LF/HF, SBE_5, normalize HF, $\text{mean}_{(t,s)}$, TSMRE_{26}, TSMRE_{30}, TSMRE_{25}, TSMRE_{27}, TSMRE_{28}, TSMRE_{29}, TSMRE_{31}, VLF, TSMRE_{23}, LF, TSMRE_{24}, TSMRE_{32}, TSMRE_{21}, $\text{std}_{(t,s)}$, TSMRE_{22}, TSMRE_{20}, TP, max $\mathbf{S}$, TSMRE_4, first AC min, std, AR coefficients 2, AR coefficients 4, TSMRE_3, $\text{RE}_{(t,s)}$, DispEn.
120	mean, ApEn_{max}, TSMRE_1, TSMRE_2, first ZC, AR coefficients 1, ECM, SBE_1, E_{SVD}, SBE_5, HF, first AC min, normalize LF, AR coefficients 4, $\text{mean}_{(t,s)}$, TSMRE_{32}, normalize HF, LF/HF, AR coefficients 2, TSMRE_{30}, TSMRE_3, SBE_8, TSMRE_{31}, VLF, TSMRE_4, DispEn, LF, TSMRE_{22}, TSMRE_5, SBE_2, TSMRE_{28}, TSMRE_{29}, TP, TSMRE_{26}, std, r_{max}, TSMRE_{20}, ApEn, TSMRE_{24}, TSMRE_{18}.
150	mean, ApEn_{max}, TSMRE_1, first ZC, TSMRE_2, AR coefficients 1, AR coefficients 4, first AC min, normalize LF, LF/HF, E_{SVD}, SBE_1, SBE_5, AR coefficients 2, normalize HF, ECM, TSMRE_3, HF, LF, $\text{mean}_{(t,s)}$, TSMRE_5, TSMRE_4, SBE_8, TSMRE_{32}, SBE_2, TSMRE_{31}, r_{max}, VLF, TSMRE_{28}, TSMRE_{29}, TSMRE_{30}, TSMRE_{26}, SampEn, DispEn, TSMRE_{22}, LZ of AC, TSMRE_{24}, TSMRE_{20}, TP, TSMRE_{19}.
180	mean, ApEn_{max}, TSMRE_1, first ZC, TSMRE_2, AR coefficients 1, AR coefficients 4, first AC min, SBE_5, AR coefficients 2, normalize LF, E_{SVD}, SBE_1, ECM, normalize HF, LF/HF, TSMRE_3, HF, SBE_8, $\text{mean}_{(t,s)}$, TSMRE_4, VLF, LF, DispEn, TSMRE_{30}, TSMRE_{32}, SBE_2, TSMRE_5, TSMRE_{31}, r_{max}, SBE_4, ApEn, TSMRE_{28}, TSMRE_{27}, TSMRE_{29}, LZ of AC, TSMRE_7, TSMRE_6, TSMRE_{25}, SampEn, TSMRE_{26}.
210	mean, ApEn_{max}, TSMRE_1, first ZC, TSMRE_2, AR coefficients 1, AR coefficients 4, first AC min, AR coefficients 2, SBE_5, normalize LF, SBE_1, LF/HF, normalize HF, E_{SVD}, LZ of AC, ECM, TSMRE_3, SBE_8, HF, TSMRE_{32}, TSMRE_4, TSMRE_5, LF, SBE_4, SampEn, DispEn, $\text{mean}_{(t,s)}$, VLF, LZ, TSMRE_6, TSMRE_7, SBE_6, TSMRE_{30}, TSMRE_{31}, SBE_2, r_{max}, ApEn, SBE_7, TSMRE_{29}.
240	mean, ApEn_{max}, TSMRE_1, first ZC, AR coefficients 4, first AC min, TSMRE_2, AR coefficients 1, AR coefficients 2, SBE_5, normalize LF, E_{SVD}, SBE_1, LF/HF, HF, normalize HF, ECM, TSMRE_3, LZ of AC, DispEn, SBE_8, TSMRE_{32}, $\text{mean}_{(t,s)}$, TSMRE_4, VLF, SBE_2, TSMRE_{31}, LF, ApEn, TSMRE_5, SBE_4, r_{max}, LZ, TSMRE_{29}, SBE_6, TSMRE_{30}, TSMRE_7, SampEn, TSMRE_6, TSMRE_{19}.
270	mean, ApEn_{max}, TSMRE_1, first ZC, AR coefficients 4, AR coefficients 1, TSMRE_2, first AC min, AR coefficients 2, SBE_5, LZ of AC, normalize LF, LF/HF, normalize HF, SBE_8, SBE_1, E_{SVD}, TSMRE_3, HF, DispEn, ECM, TSMRE_4, SBE_4, TSMRE_{32}, TSMRE_5, LZ, SBE_6, $\text{mean}_{(t,s)}$, SBE_2, TSMRE_{31}, LF, TSMRE_7, ApEn, VLF, SampEn, TSMRE_{30}, r_{max}, TSMRE_6, SBE_7, TSMRE_8.
300	mean, TSMRE_1, ApEn_{max}, first ZC, AR coefficients 4, AR coefficients 1, first AC min, AR coefficients 2, TSMRE_2, SBE_5, LZ of AC, normalize LF, LF/HF, SBE_1, ECM, SBE_8, normalize HF, E_{SVD}, TSMRE_3, DispEn, SBE_4, HF, $\text{mean}_{(t,s)}$, TSMRE_5, r_{max}, TSMRE_7, TSMRE_{32}, SBE_2, TSMRE_4, SBE_6, SampEn, VLF, SBE_7, ApEn, LF, TSMRE_{30}, TSMRE_{31}, LZ, TSMRE_9, SBE_3.

Table 4: Feature ranking. Top 40 features for different L using the algorithm FS 3.

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

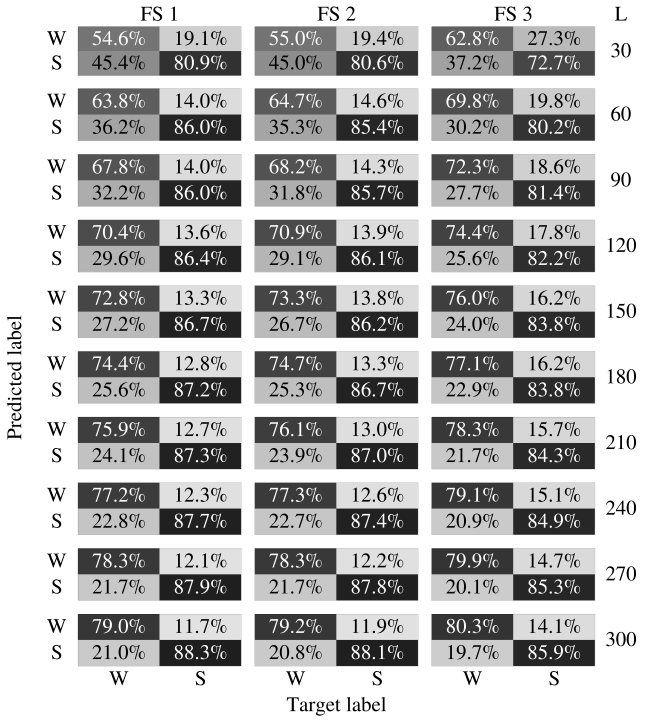

Table 5: Confusion matrix of the final results in database of 4500 remaining patients for different values of L. The performance obtained can be seen in shades of grey. FS 1: FFS-SVM with penalty error in minority class. FS 2: FFS-SVM with artificial balance. FS 3: variable selection with RF. Waking and sleeping states are labeled as W and S.

1. Preprocessing HR signals to have zero mean and unit variance and segmentation into non-overlap windows of length L.
2. Feature extraction and hyperparameters optimization using a subset of the database to design the system.
3. Feature selection to obtain the optimum features for the chosen classifier.
4. Extraction of the optimum selected features for the chosen classifier using a subset of the database to test the system and train the classifier.

In order to reuse the method described in this article with new patients, it is necessary to apply the preprocessing step so as to standardize and segment the signal Then, the selected features with the founded optimal hyperparameters should be extracted and used with the trained classifier. Although the design described in this work is computationally complex, its use is fast once the system has been designed.

We have been anticipating in previous sections the discussion about the length L of the segment to be classified. Remember that the length of the segments used in a hypnogram is 30 seconds. It is easy to note that the smaller L makes it easier to translate this scheme to the hypnogram.

To explain this, let the hypnogram H be a sequence as $H = \{t_{W_1}, t_{S_1}, t_{W_2}, t_{S_2}, \ldots, t_{S_{N-1}}, t_{W_N}\}$, where t_{W_i} and t_{S_i} are the lengths of i-th awake and asleep segments. In case of awake,

	Acc	Sp	Se	PPV	NPV	L
FS 1	**73.7**	**80.9**	54.6	**48.6**	83.1	
FS 2	73.6	80.6	55.0	48.4	83.6	30
FS 3	69.7	72.7	**62.8**	43.1	**84.65**	
FS 1	**79.4**	**86.0**	63.8	**58.3**	86.5	
FS 2	79.2	85.4	64.7	57.6	86.7	60
FS 3	76.6	80.2	**69.8**	52.5	**87.6**	
FS 1	**80.6**	**86.0**	67.8	**59.3**	87.7	
FS 2	80.4	85.7	68.2	59.0	87.8	90
FS 3	78.2	81.4	**72.3**	54.4	**88.6**	
FS 1	**81.4**	**86.4**	70.4	**60.5**	88.5	
FS 2	**81.4**	86.1	70.9	60.3	88.7	120
FS 3	79.4	82.2	**74.4**	56.0	**89.4**	
FS 1	**82.3**	**86.7**	72.8	**61.8**	89.3	
FS 2	82.0	86.2	73.3	61.1	89.4	150
FS 3	80.3	83.0	**76.0**	57.3	**90.0**	
FS 1	**83.1**	**87.2**	74.4	**63.2**	89.8	
FS 2	82.8	86.7	74.7	62.4	89.9	180
FS 3	81.2	83.8	**77.1**	58.6	**90.4**	
FS 1	**83.6**	**87.3**	75.9	**64.0**	90.3	
FS 2	83.4	87.0	76.1	63.4	90.3	210
FS 3	81.9	84.3	**78.3**	59.8	**90.8**	
FS 1	**84.2**	**87.7**	77.2	**65.2**	90.7	
FS 2	84.0	87.4	77.3	64.6	90.7	240
FS 3	82.6	84.9	**79.1**	61.0	**91.1**	
FS 1	**84.7**	**87.9**	78.3	**66.0**	91.1	
FS 2	84.6	87.8	78.3	65.7	91.1	270
FS 3	83.1	85.3	**79.9**	62.0	**91.4**	
FS 1	**85.2**	**88.3**	79.0	**67.0**	91.3	
FS 2	85.0	88.1	79.2	66.7	91.4	300
FS 3	83.6	85.9	**80.3**	63.1	**91.5**	

Table 6: Performance of the algorithms in database of 4500 remaining patients for different values of L. FS 1: forward feature selection and SVM with penalty error in minority class. FS 2: forward feature selection and SVM with artificial balance. FS 3: variable selection with random forest. The best results are highlighted in bold type.

the figure 4 shows the percentage P of the total sleep time only considering segments of length greater than L, that is

$$P(\%) = \frac{\sum_{i \in \mathcal{I}} t_{W_i}}{\sum_{\forall i} t_{W_i}} \cdot 100, \tag{13}$$

where $\mathcal{I}$ is the set of $t_{W_i} > L$. For the case of asleep is analogous.

In the real application, the HR signal is segmented into non-overlap windows of length L, and then the segments are classified obtaining the hypnogram. If there are asleep/awake segments of length less than L, some windows will not belong to a single state. That is, when L increases, the segments with mix awake and asleep stages also increase. The classifier was not designed for these mixtures. Figure 4 shows the magnitude of this mixture increment with L. On the other hand, as we reported in the results, as we increased L, the performance of the algorithm improves, because there is more information to detect dynamic changes and the features can be calculated more

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

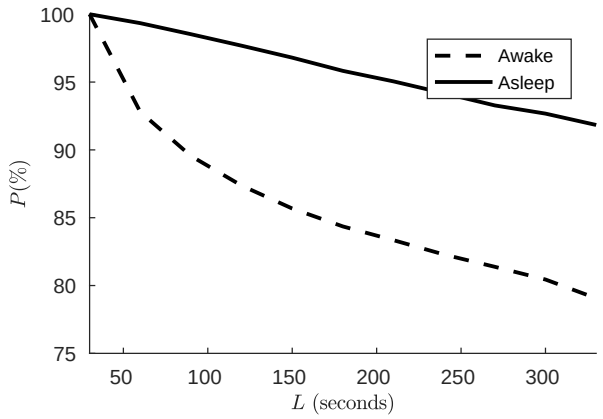

Figure 4: The percentage $P(\%)$ of the total asleep/awake time only considering segments of length greater than L.

exactly. There is a trade-off between the performance and the ability to apply the algorithm in a real situation mediated by L. When choosing a larger L to obtain better classification results, we sacrifice accuracy in the boundary of the asleep/awake transition and probably will miss some short transitions.

The development of this algorithm is at an early stage. Before to be applied in clinical practice, it is necessary to perform a large scale assessment of the method to validate its use.

As mentioned previously, in this work we used the heart rate signal with a sampling frequency of 1 Hz. The AASM stipulates the use of 25 Hz in oximetry as desirable, and 10 Hz as minimum recommended, but there are no specific recommendations about the sampling frequency of the HR [1]. Considering the range of possibles HR frequencies, there is no need for a high sampling frequency.

The use of a different sampling frequency to that we use in this work could change the performance obtained, since many features vary with the sampling frequency. In the case of applying the designed system to signals with a sampling frequency greater than 1 Hz, it will be necessary to subsample the signal. Conversely, the use of signals with a sampling frequency smaller than 1 Hz is not possible without re-performing all experiments to find new optimal parameters.

There are several previous works that perform automatic sleep stage classification. EEG is generally used, but several alternatives have been proposed for ECG. However, as far as we know, there are not studies using only HR from PPG. The comparisons between studies are not simple. Different signals, databases and number of classes are used. In order to compare with our work, in cases where it was necessary, the different sleep stages were considered as unique. However, to be fair, we will report which works discriminated sleep stages in more detail.

Beattie et al. [25] used PPG signals and accelerometer. In that work the authors considered 5 classes. The database used by Beattie was composed of 60 participants were self-reported normal sleepers. We can not make a direct comparison because they have additional information. They use the PPG signal (not only the HR calculated from it), in addition to the accelerometer signals. The best accuracy, sensitivity and specificity obtained in this work were 90.6%, 69.3% and 94.6% respectively.

Uçar et al. [26] used PPG and HRV from PPG, and the combination of these two to classify in awake and asleep. The signals used in this work contain more information than those of our work. The confusion matrix was not reported in this paper, it was deduced from the data reported by the authors. The best accuracy, sensitivity and specificity were 76%, 74% and 80% respectively. The database used contains registers of 10 patients.

Adnane et al. [22] used ECG signal from 18 patients (4 normal sleepers, 6 mixed normal and insomniac sleepers, and 8 insomniac sleepers). They only considered two classes. The best accuracy, sensitivity and specificity were 80%, 69.1% and 84.5% respectively.

Xiao et al. [23] extracted 41 features from ECG and used RF to differentiate among wake, REM and non-REM stages. The results reported were 83.94%, 51.15% and 90.15% of accuracy, sensitivity and specificity, respectively. The authors only analyzed data labeled with “stationary”, that is, they classified 5-minute windows corresponding to a single class.

Yücelbaş et al. [24] used ECG signals and classify the sleep stage in wake, REM and non-REM. They used two different databases. In total, 28 patients were considered. The reported results were discriminated by healthy subjects and patients. For the first database, the accuracies was 87.11% for the healthy and 78.08% for the patient. For the second database, 77.02% and 76.79%. The reported data do not allow to compare all performance measures.

The results here obtained are encouraging, because it addresses a limitation of all apnea diagnosis methods based only on desaturation. The risk of overfitting in our algorithm is minimal, because we use a large number of records registered in real conditions. Although we have applied strategies to prevent the unbalance of classes, it can be seen that in all the developed methods there is still a bias towards the majority class that should be addressed in future work.

The table 7 summarizes similar works. Other related results, even obtained with EEG, can be found in [24].

In future work, the algorithm developed for patients will be adapted and applied to obtain simplified hypnograms and total sleep time estimation. Finally, the apnea detection and sleep estimation systems for OSAHS diagnosis will be evaluated together.

6. Conclusions

In this work, we developed an automatic system to classify the sleep stage in awake or asleep using machine learning techniques from photoplethysmographic HR signals. It was shown that information theory-related features and complexity measures, and their extensions to time-frequency domain, are useful to differentiate between awake and asleep stages. It was shown that very simple and inexpensive signals such as those obtained by pulse oximetry can achieve performances comparable to those obtained with signals that contain more information. The use of a large database allows a good generalization

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

Method	Signal	N. of classes	N. of patients	Epoch time	Acc	Se	Sp	Prec	NPV
Beattie et al. [25]	PPG + accelerometer	5	60	30 s	90.6	69.3	94.6	70.5	94.3
Uçar et al. [26]	PPG	2	10	30 s	76.8	76	77	41.2	93.8
	HRV	2	10	30 s	72.4	74	72	35.9	92.9
	PPG + HRV	2	10	30 s	76.7	80	76	41.4	94.7
Adnane et al. [22]	ECG	2	18	30 s	80	69.1	84.5	64.5	87
Xiao et al. [23]	ECG	3	45	5 min	83.9	51.1	90.2	49.58	90.7

Table 7: Comparison with the literature.

capacity of the developed method. We developed three alternatives to eliminate the possible redundancy of the extracted features, FFS schemes based on SVM and variable selection with RF with artificial balance of training data, and FFS-SVM with penalty errors in minority class without balanced class. The FFS-SVM with penalty error in minority class has slightly better performance than the other methods. However, all performances are very similar. As future work, we will adapt the algorithm to obtain simplified hypnograms and total sleep time estimation. The ultimate aim is to use this algorithm and an apnea event detector for OSAHS diagnosis.

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.

[1] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, B. Vaughn, The AASM manual for the scoring of sleep and associated events, Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine.

[2] R. G. Norman, I. Pal, C. Stewart, J. A. Walsleben, D. M. Rapoport, Interobserver agreement among sleep scorers from different centers in a large dataset., Sleep 23 (7) (2000) 901–908.

[3] K. P. Pang, D. J. Terris, Screening for obstructive sleep apnea: an evidence-based analysis, American Journal of Otolaryngology 27 (2) (2006) 112–118.

[4] A. Yadollahi, Z. Moussavi, Apnea detection by acoustical means, in: Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE, IEEE, 2006, pp. 4623–4626.

[5] F. Roche, E. Sforza, D. Duverney, J.-R. Borderies, V. Pichot, O. Bigaignon, G. Ascher, J.-C. Barthélémy, Heart rate increment: an electrocardiological approach for the early detection of obstructive sleep apnoea/hypopnoea syndrome, Clinical Science 107 (1) (2004) 105–110.

[6] J. I. Salisbury, Y. Sun, Rapid screening test for sleep apnea using a nonlinear and nonstationary signal processing technique, Medical Engineering & Physics 29 (3) (2007) 336–343.

[7] B. Raymond, R. Cayton, M. Chappell, Combined index of heart rate variability and oximetry in screening for the sleep apnoea/hypopnoea syndrome, Journal of Sleep Research 12 (1) (2003) 53–61.

[8] M. J. Sateia, International classification of sleep disorders: highlights and modifications, Chest Journal 146 (5) (2014) 1387–1394.

[9] P. J. Strollo Jr, R. M. Rogers, Obstructive sleep apnea, New England Journal of Medicine 334 (2) (1996) 99–104.

[10] E. Shahar, C. W. Whitney, S. Redline, E. T. Lee, A. B. Newman, F. Javier Nieto, G. T. O'connor, L. L. Boland, J. E. Schwartz, J. M. Samet, Sleep-disordered breathing and cardiovascular disease: cross-sectional results of the sleep heart health study, American Journal of Respiratory and Critical Care Medicine 163 (1) (2001) 19–25.

[11] D. Leger, V. Bayon, J. P. Laaban, P. Philip, Impact of sleep apnea on economics, Sleep medicine reviews 16 (5) (2012) 455–462.

[12] N. AlGhanim, V. R. Comondore, J. Fleetham, C. A. Marra, N. T. Ayas, The economic impact of obstructive sleep apnea, Lung 186 (1) (2008) 7–12.

[13] L. J. Epstein, D. Kristo, P. J. Strollo, N. Friedman, A. Malhotra, S. P. Patil, K. Ramar, R. Rogers, R. J. Schwab, E. M. Weaver, et al., Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults, Journal of Clinical Sleep Medicine 5 (03) (2009) 263–276.

[14] N. Netzer, A. H. Eliasson, C. Netzer, D. A. Kristo, Overnight pulse oximetry for sleep-disordered breathing in adults: a review, Chest 120 (2) (2001) 625–633.

[15] G. Schlotthauer, L. E. Di Persia, L. D. Larrateguy, D. H. Milone, Screening of obstructive sleep apnea with empirical mode decomposition of pulse oximetry, Medical Engineering & Physics 36 (8) (2014) 1074–1080.

[16] L.-W. Hang, H.-L. Wang, J.-H. Chen, J.-C. Hsu, H.-H. Lin, W.-S. Chung, Y.-F. Chen, Validation of overnight oximetry to diagnose patients with moderate to severe obstructive sleep apnea, BMC Pulmonary Medicine 15 (1) (2015) 24.

[17] A. Sabil, J. Vanbuis, G. Baffet, M. Feuilloy, M. Le Vaillant, N. Meslier, F. Gagnadoux, Automatic identification of sleep and wakefulness using single-channel eeg and respiratory polygraphy signals for the diagnosis of obstructive sleep apnea, Journal of sleep research (2018) e12795.

[18] Y. Dong, Z. Hu, K. Uchimura, N. Murayama, Driver inattention monitoring system for intelligent vehicles: A review, IEEE transactions on intelligent transportation systems 12 (2) (2011) 596–614.

[19] J. Mantua, N. Gravel, R. Spencer, Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography, Sensors 16 (5) (2016) 646.

[20] T. Penzel, J. W. Kantelhardt, L. Chung-Chang, K. Voigt, C. Vogelmeier, Dynamics of heart rate and sleep stages in normals and patients with sleep apnea, Neuropsychopharmacology 28 (S1) (2003) S48.

[21] S. Aeschbacher, M. Bossard, T. Schoen, D. Schmidlin, C. Muff, A. Maseli, J. D. Leuppi, D. Miedinger, N. M. Probst-Hensch, A. Schmidt-Trucksäss, et al., Heart rate variability and sleep-related breathing disorders in the general population, The American Journal of Cardiology 118 (6) (2016) 912–917.

[22] M. Adnane, Z. Jiang, Z. Yan, Sleep–wake stages classification and sleep efficiency estimation using single-lead electrocardiogram, Expert Systems with Applications 39 (1) (2012) 1401–1413.

[23] M. Xiao, H. Yan, J. Song, Y. Yang, X. Yang, Sleep stages classification based on heart rate variability and random forest, Biomedical Signal Processing and Control 8 (6) (2013) 624–633.

[24] Ş. Yücelbaş, C. Yücelbaş, G. Tezel, S. Özşen, Ş. Yosunkaya, Automatic sleep staging based on SVD, VMD, HHT and morphological features of single-lead ECG signal, Expert Systems with Applications 102 (2018) 193–206.

[25] Z. Beattie, Y. Oyang, A. Statan, A. Ghoreyshi, A. Pantelopoulos, A. Russell, C. Heneghan, Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals, Physiological Measurement 38 (11) (2017) 1968.

[26] M. K. Uçar, M. R. Bozkurt, C. Bilgin, K. Polat, Automatic sleep staging in obstructive sleep apnea patients using photoplethysmography, heart rate variability signal and machine learning techniques, Neural Computing and Applications (2016) 1–16.

[27] M. Rostaghi, H. Azami, Dispersion entropy: A measure for time-series analysis, IEEE Signal Processing Letters 23 (5) (2016) 610–614.

[28] S. M. Pincus, Approximate entropy as a measure of system complexity., Proceedings of the National Academy of Sciences 88 (6) (1991) 2297–

2301.
[29] J. S. Richman, J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, American Journal of Physiology-Heart and Circulatory Physiology 278 (6) (2000) H2039–H2049.
[30] W. Chen, Z. Wang, H. Xie, W. Yu, Characterization of surface EMG signal based on fuzzy entropy, IEEE Transactions on neural systems and rehabilitation engineering 15 (2) (2007) 266–272.
[31] A. Rényi, et al., On measures of entropy and information, in: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, Vol. 1, 1961, pp. 547–561.
[32] S. Mallat, A wavelet tour of signal processing, Academic Press, 1999.
[33] J. Allen, Photoplethysmography and its application in clinical physiological measurement, Physiological Measurement 28 (3) (2007) R1.
[34] P. A. Kyriacou, Pulse oximetry in the oesophagus, Physiological Measurement 27 (1) (2005) R1.
[35] Z. Zhang, Z. Pi, B. Liu, TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic (PPG) signals during intensive physical exercise, Biomedical Engineering, IEEE Transactions 62 (2) (2014) 522 – 531.
[36] S. Redline, M. H. Sanders, B. K. Lind, S. F. Quan, C. Iber, D. J. Gottlieb, W. H. Bonekat, D. M. Rapoport, P. L. Smith, J. P. Kiley, et al., Methods for obtaining and analyzing unattended polysomnography data for a multicenter study, Sleep 21 (7) (1998) 759–768.
[37] E. J. Nieto, G. T. O'Connor, D. M. Rapoport, S. Redline, The sleep heart health study: design, rationale, and methods, Sleep 20 (12) (1997) 1077–1085.
[38] T. Fawcett, An introduction to ROC analysis, Pattern recognition letters 27 (8) (2006) 861–874.
[39] R. O. Duda, P. E. Hart, D. G. Stork, Pattern classification, John Wiley & Sons, 2012.
[40] M. Kuhn, K. Johnson, Applied predictive modeling, Vol. 26, Springer, 2013.
[41] F. Takens, Detecting strange attractors in turbulence, in: Dynamical Systems and Turbulence, Warwick 1980, Springer, 1981, pp. 366–381.
[42] L.-S. Xu, K.-Q. Wang, L. Wang, Gaussian kernel approximate entropy algorithm for analyzing irregularity of time-series, in: Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on, Vol. 9, IEEE, 2005, pp. 5605–5608.
[43] J. F. Restrepo, G. Schlotthauer, M. E. Torres, Maximum approximate entropy and r threshold: A new approach for regularity changes detection, Physica A: Statistical Mechanics and its Applications 409 (2014) 97–109.
[44] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Physical Review Letters 88 (17) (2002) 174102.
[45] B. Boashash, N. A. Khan, T. Ben-Jabeur, Time–frequency features for pattern recognition using high-resolution TFDs: A tutorial review, Digital Signal Processing 40 (2015) 1–30.
[46] B. Boashash, Time-frequency signal analysis and processing: a comprehensive reference, Academic Press, 2015.
[47] M. E. Torres, M. M. Anino, G. Schlotthauer, Automatic detection of slight parameter changes associated to complex biomedical signals using multiresolution q-entropy, Medical Engineering & Physics 25 (10) (2003) 859–867.
[48] A. Lempel, J. Ziv, On the complexity of finite sequences, IEEE Transactions on Information Theory 22 (1) (1976) 75–81.
[49] M. Aboy, R. Hornero, D. Abásolo, D. Álvarez, Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis, IEEE Transactions on Biomedical Engineering 53 (11) (2006) 2282–2288.
[50] G. Schlotthauer, A. Humeau-Heurtier, J. Escudero, H. L. Rufiner, Measuring complexity of biomedical signals, Complexity 2018.
[51] R. Casal, G. Schlotthauer, Sleep detection in heart rate signals from photoplethysmography, in: 2017 XVII workshop on information processing and control (RPIC), IEEE, 2017, pp. 1–6.
[52] T. F. of the European Society of Cardiology, et al., Heart rate variability: standards of measurement, physiological interpretation, and clinical use, Circulation 93 (1996) 1043–1065.
[53] M. Bonnet, D. Arand, Heart rate variability: sleep stage, time of night, and arousal influences, Electroencephalography and Clinical Neurophysiology 102 (5) (1997) 390–396.
[54] B. Boashash, L. Boubchir, G. Azemi, Time-frequency signal and image processing of non-stationary signals with application to the classification of newborn EEG abnormalities, in: Signal Processing and Information Technology (ISSPIT), 2011 IEEE International Symposium on, IEEE, 2011, pp. 120–129.
[55] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research 3 (Mar) (2003) 1157–1182.
[56] R. Kohavi, G. H. John, Wrappers for feature subset selection, Artificial Intelligence 97 (1-2) (1997) 273–324.
[57] V. Vapnik, The nature of statistical learning theory, Springer Science & Business Media, 2013.
[58] S. Abe, Support vector machines for pattern classification, Vol. 2, Springer, 2005.
[59] L. Breiman, Random forests, Machine Learning 45 (1) (2001) 5–32.
[60] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553) (2015) 436.
[61] N. Böhning, B. Schultheiss, S. Eilers, T. Penzel, W. Böhning, E. Schmittendorf, Comparability of pulse oximeters used in sleep medicine for the screening of OSA, Physiological Measurement 31 (7) (2010) 875.

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
R. Casal, L. Di Persia & G. Schlotthauer; "Sleep-wake stages classification using heart rate signals from pulse oximetry"
Heliyon, Vol. 5, No. 10, pp. e02529, 2019.