Sleep and Wakefulness State Detection in Nocturnal Actigraphy Based on Movement Information

Alexandre Domingues, Teresa Paiva, and J. Miguel Sanches Senior Member, IEEE.

Abstract—Wrist actigraphy (ACT) is a low cost and well established technique for long term monitoring of human activity. It has a special relevance in sleep studies, where its non-invasive nature makes it a valuable tool for behavioural characterization and for the detection and diagnosis of some sleep disorders.

The traditional sleep/wakefulness state estimation algorithms from nocturnal ACT data are unbalanced from a sensitivity and specificity points of view since they tend to overestimate sleep state, with severe consequences from a diagnosis point of view. They usually maximize the overall accuracy that does not take into account the highly unbalanced state distribution.

In this paper, a method is proposed to appropriately deal with this unbalanced problem, achieving similar sensitivity and specificity scores in the state estimation process. The proposed method combines two linear discriminant classifiers, trained with two different criteria involving movement detection to generate a first state estimate. This result is then refined by a Hidden Markov Model based algorithm.

The global accuracy, the sensitivity and the specificity of the method are 77.8%, 75.6% and 81.6% respectively, performing better than the tested algorithms. If the performance is assessed only for movement periods this improvement is even higher.

Index Terms—Actigraphy, Sleep/Wake estimation, Hidden Markov Model, Movement detection, Linear Discriminant classifier

I. INTRODUCTION

Wrist ACT has received great attention since the publication, by the American Sleep Disorders Association, of the guidelines for its application in clinical environment [1]. Its relevance in the scope of sleep disorders is well documented in the reviews [2], [3] where it is shown that the number of publications including ACT is rapidly increasing despite its performance still being far from the Polysomnography (PSG), the golden standard for sleep disorder diagnosis [4].

The highly-portable and non-invasive nature of ACT sensors makes them ideal for long term monitoring applications. They are a valuable tool to gather behavioural information and to obtain estimates of some sleep parameters, such as sleep efficiency and fragmentation [5] and on the characterization of the circadian cycle [6], [7].

Figure 1 shows a typical segment of ACT data over a 24 hour period, the rest and activity periods along the circadian cycle are clearly visible.

A. Domingues and J. M. Sanches are with Institute for Systems and Robotics / Bioengineering Department - Instituto Superior Técnico / Technical University of Lisbon, Portugal. adomingues@gmail.com (A. Domingues) jmrs@ist.utl.pt (J. M. Sanches).

T. Paiva is with Centro de Electroencefalografia e Neurologia Clinica (CENC) / Faculdade de Medicina da Universidade de Lisboa (FMUL).

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

In the assessment of the performance of new algorithms

Fig. 1. ACT data acquired over a 24h period. Rest and activity periods are clearly distinguishable.

Certain disorders, such as circadian phase shifts, are accurately detected from ACT data [8] but sleep staging and accurate sleep/wakefulness (S/W) state discrimination are still open issues and active fields of research.

The different levels of agreement between PSG and ACT reported in the literature have raised some issues regarding the validity of ACT for S/W estimation [9], [10], [11], [12] and the metrics used to evaluate the suggested algorithms [13].

The validation of the ACT prediction rates is typically made from the hypnogram obtained from the PSG data. Although this information is accurate, it is also unbalanced from a state distribution point of view. In fact, in a healthy subject hypnogram, at least 85% of the epochs correspond to Sleep state [14], [15], [16]. Thus, the high accuracies and sensitivities reported in S/W estimation using nocturnal ACT data, often mask the low specificity associated to the poor wake detection ability, as reported in [9]. Table I shows some of the most relevant results obtained in S/W state estimation in adults1. It illustrates how diverse is the performance of the methods and mainly, how different are the sensitivity and specificity in most of them.

### Table I

**PERFORMANCES REPORTED IN THE LITERATURE FOR S/W ESTIMATION USING ACT DATA IN ADULT POPULATIONS.**

<table>
<thead>
<tr>
<th>Author</th>
<th>Sens (%)</th>
<th>Spec (%)</th>
<th>Acc (%)</th>
<th>G-mean (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cole et al. [6]</td>
<td>95</td>
<td>64</td>
<td>88</td>
<td>78</td>
</tr>
<tr>
<td>Sadeh et al. [7]</td>
<td>96</td>
<td>75</td>
<td>92</td>
<td>841</td>
</tr>
<tr>
<td>Kushida et al. [17]</td>
<td>90</td>
<td>45</td>
<td>89</td>
<td>64</td>
</tr>
<tr>
<td>Ancoli-Israel et al. [18]</td>
<td>89</td>
<td>69</td>
<td>86</td>
<td>781</td>
</tr>
<tr>
<td>Hedner et al. [19]</td>
<td>95</td>
<td>36</td>
<td>83</td>
<td>59</td>
</tr>
<tr>
<td>Sivertsen et al. [11]</td>
<td>95</td>
<td>64</td>
<td>90</td>
<td>72</td>
</tr>
</tbody>
</table>

1The methods from Sadeh et al. and Hedner et al. were implemented to compare the performance of the proposed algorithm.
and comparison with existing methods, common datasets and figures of merit must be used. In [14], for instance, the authors implemented and compared the algorithms described in [6] and [7] obtaining results significantly different than the originally reported.

In order to overcome the intrinsic limitations of ACT, new approaches are being explored. In [20] cardio-respiratory signals are combined with ACT data yielding promising results and in [21] the authors show that it is possible to accurately characterize human activity from accelerometer data in a non sleep scope. In [13] the authors employ artificial neural networks and decision trees to score infant ACT data obtaining relevant results and stressing the importance of using common metrics.

In [22], [23] the authors hypothesized that movements during wakefulness and sleep states are intrinsically different, not only in terms of magnitude but from a spectral and statistical distribution perspective. While movements during sleep state are typically random and without purpose, i.e. purposeless, movements during wakefulness state are more coherent and correlated, usually with a defined purpose.

The method proposed in this paper for S/W estimation, is based on a movement detector (MD) designed to discriminate and detect movement and quietness events from the nocturnal ACT data.

The classification/testing procedure is performed in three stages. After the feature extraction, an estimation of the S/W state is obtained for each epoch with a Linear Discriminant Classifier (LDC) [24], with parameters $\theta^*$ and $\theta_M^*$, during the quietness and movement epochs respectively, as represented in Fig. 3.

The previous result is then refined with a Hidden Markov Model (HMM) that incorporates statistical information computed from the training data.

The features used in the LDC’s are the coefficients of a Rayleigh mixture distribution, to describe the first order statistics of the ACT data, the residues of an Auto-regressive (AR) model fitted to the data, describing its high order statistics and finally the signal energy that takes into account the intensity of the signal, the most important feature used in the traditional approaches for ACT data processing.

ACT based S/W estimation algorithms typically combine several features extracted from the intensity and frequency counts of the recorded signal. In these algorithms, periods of strong activity are normally scored as wakefulness and long periods quietness as sleep.

This strategy leads to acceptable and relevant accuracies but also to the well documented limitations of ACT such as the poor ability to detect wakening episodes during quietness periods, very typical in insomnia, and the generalized tendency to overestimate sleep [8], [13].

Since sleeping is the natural state during the night, when estimating S/W states, for sleep disorders diagnosis purposes, it is generally more important to accurately estimate wakefulness than sleep.

The traditional state estimation algorithms usually maximize the overall accuracy that does not take into account this unbalanced state distribution, which leads to poor wakefulness state detection rate.

The proposed method is designed to achieve a similar performance during movement and quietness periods and it is
tuned, using the Geometric mean (G-mean) [25] as the optimization criteria, for a balance between sleep and wakefulness detection ability. Thus, minimizing the tendency of ACT to underestimate the wakefulness periods [8], [13].

The presented algorithm is optimized for proper S/W estimation in the scope of sleep disorders diagnosis. Other applications may have different requirements not fulfilled by this method. For example, the detection of rest and activity periods for ambulatory blood pressure monitoring, such as the work described in [26], requires an algorithm less sensitive to micro-wakening episodes.

II. METHODS

This section describes the method used to estimate the S/W states from the nocturnal ACT data. These data are acquired with an actigraph sensor located at the non-dominant wrist of the subject. The sensor, which is basically a 3D accelerometer, provides the magnitude of the acceleration vector.

A. Algorithm

The complete state estimation method, displayed in Fig. 3, is composed by: a pre-processing step, feature extraction, training, state estimation and a final classification refinement.

Two pre-processing operations are performed on the data, 1) magnitude normalization and DC component removal, and 2) movement segmentation.

Magnitude normalization and DC component removal is required to minimize the inter-patient and inter-device variability. This procedure, performed in a sliding window basis, is done according to

$$\tilde{a}(n) = \frac{a(n) - \mu(n)}{\sigma(n)}$$  (1)

where $a(n)$ is the actigraphy sample, $\mu(n)$ and $\sigma(n)$ are the mean and standard deviation of the data within the 5 minute window centred at the $n^{th}$ sample and $\tilde{a}(n)$ is the normalized actigraphy sample.

In a second pre-processing step, movement events are identified on the normalized data with the MD displayed in Fig. 4. This detector is composed by i) a non-causal low-pass stretching filter and a ii) threshold based binarization block. The smoothing filter computes the movement envelope, its stretching filter and a ii) threshold based binarization block.

Fig. 4. Structure of the movement detector.

$$t^* = 0.68 \pm 0.22.$$ 

The output of the detector is a binary function $\tau(n) \in \{m, q\}$, where $m$ corresponds to movement and $q$ to quietness.

Figure 2 displays an example of pre-processed data. Figure 2.a) shows the normalized ACT signal and the movement indicator and Fig. 2.b) the corresponding hypnogram segment. The hypnogram discriminates 5 different states, namely wakefulness, Rem sleep, and 3 non-Rem sleep states. All epochs marked as Rem and non-Rem were translated into a single sleep label.

The next three sections describe 1) the features, 2) the main classification stage (MCS) and 3) the refinement algorithm, which compose the proposed method, here called Movement
based State Detection (MSD).

1) Features: In this work an extended set of features is used, one related with ACT intensity and two describing first and higher order statistics, used to discriminate the intrinsic characteristics of the ACT data. After pre-processing, each ACT time course is divided in contiguous epochs of \( T = 30 \) seconds.

Let \( w_j \) represent a \( L \) dimensional window, 210 seconds long, centred on the \( j^{th} \) epoch, where \( j \in \{1, \ldots, M\} \) with \( M \) the total number of epochs. Features are extracted from each window, \( w_j \), as follows:

- Energy, \( E_j \): The energy of the epoch is \( E_j = \sum_k h(k)w_j^2(k) \) where \( h = \{h(k)\} \) is a Hanning window.
- Residue, \( r_j \): Residue of the 8-order AR model \( \{27\} \) estimated for each \( w_j \) based on a \( L \) dimension window, centred on it, as proposed by the authors in \( \{23\} \).

By using a forward search feature selection algorithm \( \{24\} \), it was concluded that the residue is more discriminative than the coefficients of the AR model in the estimation of the S/W state. Figure 6-B) shows the normalized histograms of the residues obtained for sleep and wakefulness movements. The residues roughly follow a Gaussian distribution with different means and standard deviations for the two states.

- Coefficients \( \alpha_j(k) \): Coefficients of a three component Rayleigh Mixture Model (RMM) \( \{22\} \), fitted to each \( w_j \). It was found that only the coefficients, \( \alpha_j(k) \) are discriminative for S/W state estimation and not the parameters of the Rayleigh components in the mixture. The number of components in the mixture, \( L \), was selected by fitting several mixture distributions, with different values of \( L \), to the complete set of data and measuring the goodness-of-fit. The used optimality criterion was the Kullback-Leibler (KL) divergence.

Table II shows that the value of the KL divergence decreases until \( L = 3 \) and then remains approximately constant for higher values of \( L \). The selected value was thus \( L = 3 \).

\[ J = \sqrt{\text{spec} \times \text{sens}} \]  \hspace{1cm} (2)

where \( \text{sens} \) and \( \text{spec} \) are the sensitivity, the ability of the method to correctly detect sleep, and specificity, the ability of the method to correctly detect wakefulness, respectively. This criterion is adopted for simultaneous maximization of the sensitivity and specificity, which is not guaranteed when only the overall accuracy (acc) is considered.

The MCS is composed by the following three steps.

i) Classification of the test data with LDC(\( \theta^* \)). The parameters \( \theta^* \) were obtained through the maximization of the cost function (2), taking into account the whole training data set.

\[ \text{TP} = \frac{\text{TP} \times \text{TN}}{\text{TP} + \text{TN}}, \quad \text{FP} = \frac{\text{FP} \times \text{FN}}{\text{FP} + \text{FN}}, \quad \text{TN} = \frac{\text{TN} \times \text{FP}}{\text{TN} + \text{FP}}, \quad \text{FN} = \frac{\text{FN} \times \text{TP}}{\text{FN} + \text{TP}} \]

TP, TN, FP, and FN are the true positives, true negatives, false positives and false negatives, respectively.

\[ \text{TP} = \frac{\text{TP} \times \text{TN}}{\text{TP} + \text{TN}}, \quad \text{FP} = \frac{\text{FP} \times \text{FN}}{\text{FP} + \text{FN}}, \quad \text{TN} = \frac{\text{TN} \times \text{FP}}{\text{TN} + \text{FP}}, \quad \text{FN} = \frac{\text{FN} \times \text{TP}}{\text{FN} + \text{TP}} \]

The physical reasoning to use a Rayleigh mixture model is related with the model adopted for the actigraph sensor, as described in \( \{28\} \). If the acceleration along each axis is described by a zero mean Gaussian distribution, the acceleration magnitude follows a Rayleigh and a Maxwell distribution in 2D and 3D respectively. Figure 6-A) shows the normalized histograms of Sleep and Wakefulness movement data and the two Rayleigh distributions fitted to the data.
ii) Classification of the test data with LDC($\theta_k^m$). LDC($\theta_k^m$) is trained using the whole data but the cost function (2) is optimized taking into account only the movement data ($\tau(n) = m$).

iii) Combination of the previous results, where quietness epochs are scored from LDC($\theta^*$) and movement epochs are scored using LDC($\theta_k^M$).

3) HMM regularization algorithm: This final procedure refines the results obtained in the MCS leading to the final estimation, $\hat{x} \in \{s, w\}$.

A Hidden Markov Model was chosen for this task since it models processes which have a temporal relation between states, which is the case in the sleep/wake cycle.

Two hidden states are considered, $x \in \{s, w\}$, where $s$ and $w$ refer to sleep and wakefulness states, respectively. Let us consider $x^* \in \{s^*, w^*\}$, the output of the MCS, $\tau \in \{m, q\}$ the output of the MD and $t \in \mathbb{N}$ the time, in seconds, since the last movement (quietness periods) or since the patient started to move (movement periods).

The observation model takes into account the following information, extracted in the training step:

- The accuracy rate of the MCS, given by $P(x^*|x)$.
- The conditional distribution of the activity given the state, expressed as $P(\tau|x)$ and shown in Table III.
- The duration of the quietness and movement periods during sleep and wakefulness. Expressed as

$$P(t|x, \tau) = \begin{cases} N(\sigma_x) & \text{if } \tau = m \\ E(\lambda_x) & \text{if } \tau = q \end{cases}$$

(3)

where $N(\sigma_x)$ is a Gaussian probability distribution with zero mean and $E(\lambda_x)$ an exponential probability distribution.

$E(\lambda_x)$ is an exponential probability distribution, $P(t|\lambda_s, w)$, giving the probability of a quietness period of length $t$ being observed during sleep and wakefulness. This distribution arises naturally if movement events are assumed to be a stochastic Poisson process. The parameters $\lambda_s$ and $\lambda_w$ are computed as

$$\hat{\lambda}_x = \frac{1}{\overline{T}_x}$$

(4)

where $\overline{T}_x$ is the mean duration of all quietness intervals for sleep ($x = s$) and wakefulness ($x = w$) states.

$N(\sigma_s)$ is a zero mean Gaussian probability distribution, $P(t|\sigma_s, w)$, giving the probability of a movement of length $t$ being observed during sleep and wakefulness states. The parameter $\sigma_s$ and $\sigma_w$ is the standard deviation of the duration of all movements recorded during sleep and wakefulness states, respectively.

Figure 7 - A) shows the histograms of the length of the quietness periods during sleep and wakefulness states. As expected the mean value is larger during sleep and the periods tend to be longer. Figure 7 - B) shows the histogram of the length of the recorded movements, as expected, during wakefulness, movement duration is typically longer and has a higher standard deviation.

$$P(x^*, \tau, t|x) = P(x^*|x)P(t, \tau|x) = P(x^*|x)P(\tau|x)P(t|x, \tau)$$

(5)

The transition matrix is computed from the training data as

$$P = \begin{bmatrix} N(s\!s) & N(s\!w) & N(s\!w) + N(w\!w) \\ N(s\!s) + N(s\!w) & N(w\!s) & N(w\!s) + N(w\!w) \end{bmatrix}$$

(6)

where $N(.)$ is a counting operator for ss, sw, ws and ww, corresponding to sleep-sleep, sleep-wakefulness, wakefulness-sleep and wakefulness-wakefulness transitions, respectively.

The hidden state, $x(t)$, is estimated along the time from the observations, $y(t)$, and the model parameters. The initial probabilities are set to 0.0 and 1.0 for Sleep and Wakefulness states respectively, because all patients were awake in the beginning of the exam, and the optimal solution, the most probable state sequence, is computed using the Viterbi Algorithm [29].

B. Comparative methods

In this section, two of the methods for S/W discrimination listed in Table I, are described, as well as the adaptations required for their use with the existing datasets. The two methods, here called Sadeh’s [7] and Hedner’s [19], were selected by their reported performance, particularly its Sens and Spec balance.

1) Sadeh’s algorithm: In [7] Sadeh et al. propose a scoring algorithm, using 60 seconds epochs, that linearly combines several features in the following discriminative function,

$$PS = \alpha \theta^T$$

(7)

where $\alpha$ is a vector of adjustable parameters defined in [7] as

$$\alpha = [7.601; -0.065; -1.08; -0.056; -0.703].$$

(8)

$\theta$ is a vector of features extracted from the data expressed as

$$\theta = [1; \mu; Nat; \sigma; LogAct]$$

(9)
where $\mu$ is the mean number of activity counts on a 11 minute window centred in the current epoch, $Nat$ is the number of epochs with activity level equal to or higher than 50 but lower than 100 activity counts in a window of 11 minutes, $\sigma$ is the standard deviation of the activity on the last 6 minutes and $\log Act$ is the natural logarithm of the number of activity counts during the scored epoch plus 1. A given epoch is scored as sleep if $PS = 0$ and wakefulness otherwise.

Since the initial algorithm was developed for a different Actigraph device and database, the 5 parameters from the discriminative function were optimized for the current data. The optimal parameters were found maximizing the cost function given by (2) leading to,

$$\alpha^* = [4.097; -0.528; -0.51; -0.259; -1.65].$$

2) Hedner’s algorithm: In [19] Hedner et al. present a S/W state estimation algorithm with focus on sleep apnea patients. The algorithm is divided in 4 distinct steps with the last one aiming at the detection of periodic movements, typical from apnea patients. Here, the last step is discarded and the algorithm is as follows:

i) Determination of the background movement activity of the patient throughout the night, $\sigma$.

ii) Bandpass filter between 2 and 2.5 Hz, leading to a signal regarded as the energy of the activity.

iii) For each 30 second epoch, values of energy below $\sigma$ are discarded and the remaining energy is integrated using a 5-minute Hanning window. Values below a fixed threshold, $\theta$ are scored as sleep and values above are scored as wakefulness.

The two parameters, $\sigma$ and $\theta$ were computed maximizing the cost function given by (2).

III. RESULTS

This section describes the data used in this work, presents the experimental results obtained with the MSD and compares them with the two methods described in the previous section. The performance is assessed with several Figures of Merit (FOM). These FOMs are computed in a leave-one-patient-out cross validation basis, where each patient dataset is tested after training the classifier with the remaining data. All the classification routines are implemented using PRTools [24] for Matlab.

A. Data

The nocturnal ACT data was acquired with a Somnowatch™ device, from Somnomedics, placed in the non-dominant wrist of the subjects, acquiring with a sampling rate of 1Hz. The core of these devices is a 3D accelerometer that measures the acceleration along 3 orthogonal axis with a configurable output format.

Here, the output of the actigraph is the acceleration magnitude. Some authors suggest that this configuration, also known as digital integration, is the most reliable to measure activity levels [8], [30].

The nocturnal ACT data were jointly acquired with PSG data for validation purposes and the hypnogram obtained from the PSG by trained technicians, is used as a ground truth to identify sleep and wakefulness states in epochs of 30 seconds. Twenty nine adult subjects (age $48 \pm 13$ years, 13 Males, 16 Females), with no particular pre-diagnosed sleep disorder, participated in this study.

The Sleep Efficiency (SE), computed as the ratio between total sleep time and total bed time was obtained for each patient. All the values of SE fell within the range 75% – 85%. These values are below the typical values found in healthy subjects, usually above 85% [31], which indicated sleep disturbances, although not necessarily pathological.

The normalization step applied to the data reduces the variability observed in the datasets recorded with distinct devices. This step also contributes to the generalization of the described algorithm to data acquired with different models/brands of actigraph devices.

B. ACT data characterization

Nocturnal ACT data is highly unbalanced from a state distribution point of view. This can be confirmed in Table III where experimental conditional distribution means and standard deviation values, computed from the relative frequencies observed in the real hypnograms are displayed.

| $P(x|\tau)$ | $P(x|m)$ | $P(x|s)$ |
|-------------|----------|----------|
| $P(w|m)$   | $0.58 \pm 0.19$ | $0.42 \pm 0.18$ |
| $P(w|q)$   | $0.15 \pm 0.1$  | $0.85 \pm 0.09$  |
| $P(m|w)$   | $0.18 \pm 0.07$ | $0.82 \pm 0.07$  |
| $P(m|s)$   | $0.06 \pm 0.03$ | $0.94 \pm 0.04$  |

As expected, during movement periods, the most frequent state is wakefulness, $P(w|m) = 0.58$, although closely followed by sleep, $P(s|m) = 0.42$. During quietness periods the gap between the two states is larger, $P(w|q) = 0.85$ against $P(w|q) = 0.15$.

This observation suggests high correlations between movement and wakefulness state and quietness and sleep state respectively, but in fact, the probability of a patient moving during wakefulness is much smaller than the probability of not moving, as can be seen from $P(m|w)$ and $P(q|w)$.
This fact illustrates the main limitation of nocturnal ACT for S/W state estimation: although the methods are based on the recorded movements, they only occur during 6 ± 3% of the time in the whole register.

The information from Table III clarifies why simple empirical classification rules can actually lead to apparently impressive performances. The traditional FOMs, accuracy, sensitivity and specificity, are not able to cope with the type of unbalanced data present on nocturnal ACT.

The classification results displayed in Table IV, obtained with two naive methods, are used to illustrate the previous point.

**TABLE IV**

<table>
<thead>
<tr>
<th>Sens(%)</th>
<th>Spec(%)</th>
<th>Acc(%)</th>
<th>G-mean(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>100</td>
<td>0</td>
<td>80.4 ± 5.9</td>
</tr>
<tr>
<td>M2</td>
<td>97.2 ± 2.5</td>
<td>22.9 ± 7</td>
<td>84.8 ± 5.4</td>
</tr>
</tbody>
</table>

Method M1 classifies all data as *sleep* resulting in a surprising global accuracy of 80.4%. This result is relevant because it shows that a good accuracy is not a good indicator of the performance of the method since it has no ability to detect *wakefulness* state. The second method, M2, classifies all the quietness periods as *sleep* and all the movement periods as *wakefulness*. Even by misclassifying all the *sleep* epochs during movement periods, the method is able to achieve a sensitivity of 97.2% and global accuracy of 84.8%.

The limitation of both methods is revealed by the low specificity and G-mean.

**C. S/W classification**

Table V summarizes the results obtained with the individual LDC classifiers, \(LDC(\theta^*)\) and \(LDC(\theta_{M}^*)\), the results of the intermediate MCS (which is the combination of both LDC’s) and the final MSD. Some results are presented for two distinct scenarios; i) when all the data is considered in the classification and ii) when only movement data is considered \((\tau(n) = m)\).

The \(LDC(\theta^*)\) classifier achieves a global accuracy of 75.9% with sensitivity, specificity and G-mean in the same range. The limitation of this classifier arises when only movement data is considered, with the sensitivity falling to 15.8%. This means that during movement periods the classifier tends to classify all epochs as *wakefulness*.

The \(LDC(\theta_{M}^*)\) is only evaluated for movement segments. It achieves a G-mean of 68.7%, approximately 30% higher than the \(LDC(\theta^*)\) during the same periods. On the other hand, the specificity drops to 63.4% due to the *wakefulness* periods misclassified as *sleep*.

The MCS combines the scores from the two LDC’s. It achieves a G-mean of 72.8% when the complete data is considered (the decrease in specificity is due to reason explained before) and the performance during movements is similar to the \(LDC(\theta_{M}^*)\).

Finally, the performance of the MSD clearly reflects the improvement obtained with the HMM. It achieves a G-mean of 78.5% when all the data is considered and 73.7% when limited to movement data.

**TABLE V**

<table>
<thead>
<tr>
<th></th>
<th>Sens (%)</th>
<th>Spec (%)</th>
<th>Acc (%)</th>
<th>G-mean (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDC((\theta^*))</td>
<td>i) 75.9 ± 9.8</td>
<td>74.8 ± 9.6</td>
<td>75.9 ± 8.7</td>
<td>75.3</td>
</tr>
<tr>
<td></td>
<td>ii) 15.8 ± 11.6</td>
<td>96.6 ± 2.9</td>
<td>65.7 ± 12.2</td>
<td>39.1</td>
</tr>
<tr>
<td>LDC((\theta_{M}^*))</td>
<td>ii) 74.5 ± 16.8</td>
<td>63.4 ± 11.6</td>
<td>66.5 ± 11.3</td>
<td>68.7</td>
</tr>
<tr>
<td>MCS</td>
<td>i) 72.1 ± 3.2</td>
<td>73.6 ± 11.1</td>
<td>71.3 ± 8.2</td>
<td>72.8</td>
</tr>
<tr>
<td>MSD</td>
<td>i) 75.6 ± 8.3</td>
<td>81.6 ± 7.5</td>
<td>77.8 ± 8.1</td>
<td>78.5</td>
</tr>
<tr>
<td></td>
<td>ii) 73.8 ± 14.2</td>
<td>73.5 ± 9.5</td>
<td>75.5 ± 9.2</td>
<td>73.7</td>
</tr>
</tbody>
</table>

It is important to stress the balance of Sens and Spec achieved with the proposed method in global terms but specially during movement periods. This results can be very useful when ACT is used together with other physiological data, e.g. *electrocardiography* (ECG), whose sensors are typically sensitive to movement artefacts.

In order to assess the generalization capability of the algorithm the following procedure was performed:

i) Ten datasets were randomly selected from the pool of 29 available datasets.

ii) From these 10 datasets, 5 were randomly selected to train the algorithm.

iii) The remaining 5 datasets were used to test the algorithm and the average G-mean was computed.

This procedure was repeated 15 times resulting in an average G-mean of 76.3 ± 2%. This value is only 2% smaller than the G-mean reported in Table V and, together with the low standard deviation, suggests that the reported results should be extensible to other datasets.

The sensibility of the method, to small variations on the MD threshold, was assessed by forcing random variations of ±20% on each dataset threshold. The variation in the G-mean’s for i) All the data and for ii) Movement data (see Table V) was less that 1% in average.

Table VI compares the results obtained with the MSD with the two comparative methods. Using the complete datasets MSD achieves higher sensitivity, specificity, global accuracy and G-mean than the considered methods. While the difference in global Acc is relatively small (≈ 3%), the increase in the G-mean is 10.5% and 7.4%. This result clearly illustrates the limitation of using the global accuracy as the only performance
A new database of ACT data was built specifically for this project. With these data, the MSD yields a global accuracy of 77.8%, a sensitivity of 75.6% and a specificity of 81.6%, revealing a balance in the detection of both sleep and wakefulness states, a key issue of this work. Additionally, under the G-mean metrics the proposed method clearly outperforms the other tested methods. During the movement periods the method achieves an accuracy of 75.5%, sensitivity of 73.8%, specificity of 73.5% and G-mean 73.7%.

ACKNOWLEDGEMENT

This work was supported by project the FCT (ISR/IST plurianual funding) through the PIDDAC Program funds and FCT project Detection of Brain Micro-states in Fibromyalgia (PTDC/SAU-BEB/104948/2008).

REFERENCES


TABLE VI

<table>
<thead>
<tr>
<th></th>
<th>Sens(%)</th>
<th>Spec(%)</th>
<th>Acc(%)</th>
<th>G-mean(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MSD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i)</td>
<td>75.6 ± 8.3</td>
<td>81.6 ± 7.5</td>
<td>77.8 ± 8.1</td>
<td>78.5</td>
</tr>
<tr>
<td>ii)</td>
<td>73.8 ± 14.2</td>
<td>73.5 ± 9.5</td>
<td>75.5 ± 9.2</td>
<td>73.7</td>
</tr>
<tr>
<td><strong>Sadeh</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i)</td>
<td>75.1 ± 6.5</td>
<td>61.6 ± 15.2</td>
<td>73.9 ± 3.7</td>
<td>68.0</td>
</tr>
<tr>
<td>ii)</td>
<td>47.3 ± 12.8</td>
<td>75.8 ± 9.2</td>
<td>62.4 ± 12.8</td>
<td>59.9</td>
</tr>
<tr>
<td><strong>Heidrich</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i)</td>
<td>73.6 ± 10.7</td>
<td>68.6 ± 11</td>
<td>74.1 ± 9.3</td>
<td>71.1</td>
</tr>
<tr>
<td>ii)</td>
<td>14.5 ± 14.1</td>
<td>88.7 ± 26.1</td>
<td>64.2 ± 16.3</td>
<td>35.9</td>
</tr>
</tbody>
</table>


Alexandre Domingues Alexandre Domingues graduated in Electronics and Telecommunications Engineering in 2006, after a period working as a software engineer he joined a bioinformatics group, where he worked in optimization strategies for metabolic networks. In 2009 he received the master of Science degree in Electrical and Computer Engineering and in 2010 he entered the PhD School in Biomedical Engineering at Instituto Superior Tecnico, Lisbon - Portugal, working on statistical algorithms for the diagnosis of sleep disorders.

Teresa Paiva Medical Doctor since 1969, specialized in Neurology and Sleep Medicine, Professor of the Medical Faculty of Lisbon. She developed a longstanding clinical work, while implementing national teaching programs, namely the first worldwide Master Degree in Sleep Sciences, and orienting national awareness campaigns for the general public.

J. Miguel Sanches J. Miguel Sanches (JS) received the E.E., M.Sc. and Ph.D. degrees from the Lisbon Institute of Technology (Instituto Superior Tecnico (IST)), Technical University of Lisbon (UTL), Portugal, in 1991, 1996 and 2003 respectively and the habilitation (agregao) in 2013 by the University of Lisbon (UL) in Biomedical Engineering. JS is from the recently created Department of Bioengineering (DBE) at the IST and before he was at the Department of Electrical and Computer Engineering (DEEC) where has taught in the area of signal processing, systems and control. JS has been actively involved in the course of Biomedical Engineering advising master thesis of the Biomedical Engineering course and PhD students of the doctoral program in Biomedical Engineering.

JS is researcher at the Institute for Systems and Robotics (ISR) and his work has been focused in Biomedical Engineering (BME), namely, in biomedical image processing, physiologically based modeling of biological systems and statistical signal processing of physiological data. Today, his research activity is focused in the morphological and textural characterization of tissues from Ultrasound (US) images, functional Magnetic Resonance Imaging (fMRI) and biological quantification from fluorescence images of microscopy. JS is also involved in the development of signal processing algorithms for polysomnography data and smartphones applications for long term monitoring for sleep disorders diagnosis purposes. In this scope, Heart Rate Variability analysis is today one of his main interests.

Almost all of his research work is in collaboration with medical and biological institutions, specially the Medical School of Lisbon (FMLS), the Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP) and the Electrocencephalography and Clinical Neurophysiology Center.

JS is senior member of the IEEE Engineering in Medicine and Biology Society since 2011 and Member of the Bio Imaging and Signal Processing Technical Committee (BISP-TC) of the IEEE Signal Processing Society. He is also president of the Portuguese Association of Pattern Recognition (APPR), an affiliated of the International Association of Pattern Recognition (IAPR).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.