Validation of a new data-driven automated algorithm for muscular activity detection in REM sleep behavior disorder

Cesari, Matteo; Christensen, Julie Anja Engelhard; Sixel-Döring, Friederike; Trenkwalder, Claudia; Mayer, Geert; Oertel, Wolfgang H; Jennum, Poul; Sørensen, Helge Bjarup Dissing

Published in:
Journal of Neuroscience Methods

Link to article, DOI:
10.1016/j.jneumeth.2018.11.016

Publication date:
2019

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Validation of a new data-driven automated algorithm for muscular activity detection in REM sleep behavior disorder

Matteo Cesari¹, Julie A.E. Christensen¹,², Friederike Sixel-Döring³,⁴, Claudia Trenkwalder³, Geert Mayer⁴, Wolfgang H. Oertel⁴, Poul Jennum² and Helge B.D. Sorensen¹

¹Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark, ²Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Rigshospitalet Glostrup, Denmark, ³Paracelsus-Elena Klinik, Kassel, Germany, ⁴Department of Neurology, Philipps University, Marburg, Germany.

Corresponding author
Matteo Cesari
Department of Electrical Engineering
Technical University of Denmark
Ørsteds Plads, Building 349
2800 Kgs. Lyngby
Denmark
Email: maces@elektro.dtu.dk

Type of article: Research Paper

HIGHLIGHTS

- New data-driven method improves REM sleep behavior disorder (RBD) identification
- Muscular activity during non-REM sleep contributes to identify RBD patients
- The method is robust to apnea and arousal-related movements
ABSTRACT

**Background:** Documentation of REM sleep without atonia is fundamental for REM sleep behavior disorder (RBD) diagnosis. The automated REM atonia index (RAI), Frandsen index (FRI) and Kempfner index (KEI) were proposed for this, but achieved moderate performances.

**New Method:** Using sleep data from 27 healthy controls (C), 29 RBD patients and 36 patients with periodic limb movement disorder (PLMD), we developed and validated a new automated data-driven method for identifying movements in chin and tibialis electromyographic (EMG) signals. A probabilistic model of atonia from REM sleep of controls was defined and movements identified as EMG areas having low likelihood of being atonia. The percentages of movements and the median inter-movement distance during REM and non-REM (NREM) sleep were used for distinguishing C, RBD and PLMD by combining three optimized classifiers in a 5-fold cross-validation scheme.

**Results:** The proposed method achieved average overall validation accuracies of 70.8% and 61.9% when REM and NREM, and only REM features were used, respectively. After removing apnea and arousal-related movements, they were 64.2% and 59.8%, respectively.

**Comparison with Existing Method(s):** The proposed method outperformed RAI, FRI and KEI in identifying RBD patients and in particular achieved higher accuracy and specificity for classifying RBD.

**Conclusions:** The results show that i) the proposed method has higher performances than the previous ones in distinguishing C, RBD and PLMD patients, ii) removal of apnea and arousal-related movements is not required, and iii) RBD patients can be better identified when both REM and NREM muscular activities are considered.
Keywords: Data-driven method, Electromyography, Muscular activity, Periodic Limb Movement Disorder, Polysomnography, REM sleep behavior disorder, REM sleep without atonia.

INTRODUCTION

Rapid eye movement (REM) sleep behavior disorder (RBD) is a parasomnia characterized by loss of muscle atonia during REM sleep and a clinical history of dream enactment (American Academy of Sleep Medicine, 2014; Schenck et al., 1986; Schenck and Mahowald, 2002). Follow-up studies (Postuma et al., 2015; Schenck et al., 2013) and neurophysiological investigations (Boeve et al., 2013) support the hypothesis that RBD is an early stage of alpha-synucleinopathies, including Parkinson’s disease (PD), PD dementia, dementia with Lewy bodies and multiple system atrophy (Högl et al., 2018). Therefore, a correct diagnosis of RBD becomes of critical importance in order to identify alpha-synucleinopathies in their early stages, and RBD patients may become the target of neuroprotective treatments when they will become available.

In the diagnosis of RBD, the documentation of REM sleep without atonia (RSWA) in electromyographic (EMG) signals during polysomnography (PSG) is essential. Currently, the gold standard for RSWA scoring is the application of one of the following visual methods: i) The American Academy for Sleep Medicine (AASM) recommends RSWA identification when either tonic and phasic activity in the chin or phasic activity in the limbs is seen in more than 27% of 30-s REM sleep epochs (Berry et al., 2016); ii) RSWA is identified by the Montréal method when either in the chin signal more than 30% of 20-s REM sleep epochs show tonic activity, or when in the same signal more than 15% of these epochs contain phasic activity (Lapierre and Montplaisir, 1992; Montplaisir et al., 2010); iii) the SINBAR method proposes RSWA identification when >32% of 3-s REM sleep mini-epochs contain the combination of any (either tonic or phasic) chin EMG activity.
and phasic activity in the flexor digitorum superficialis muscles (Frauscher et al., 2012). These methods have some drawbacks, such as the facts that they are time consuming, require advanced expertise and are non-consistent, due to the non-uniquely interpretable rules for distinguishing between resting EMG, tonic and phasic activity.

To overcome these problems, several computerized methods have been developed and these include the REM atonia index (RAI) (Ferri et al., 2010, 2008), the supra-threshold REM activity metric (Burns et al., 2007), the short/long muscle activity index (Gutowski et al., 2018; Mayer et al., 2008), the Frandsen index (FRI) (Frandsen et al., 2015), the Kempfner index (KEI) (Kempfner et al., 2014; Kempfner and Nikolic, 2014) and the computerized version of the SINBAR method (Frauscher et al., 2014). In a previous study, we compared the performances of these methods (except the computerized SINBAR due to lack of important implementation details in the original published description) in identifying RBD in different scenarios and we concluded that none of the methods could be elected as the optimal one, due to the varying performances of the methods across the different scenarios (Cesari et al., 2018b). However, we found that RAI, FRI and KEI achieved generally higher performances than the others (Cesari et al., 2018b). A common characteristic of these methods is that they are influenced and biased by human definitions of thresholds, or definitions of tonic and phasic muscular activity. From this, we hypothesize that a data-driven method for muscular activity detection, independent from human definitions and rules for scoring, may improve the automatic analysis of muscular activity and give better outcomes in patient classification than previously proposed methods.

Recently, our group has focused on developing data-driven methods for sleep staging (Christensen et al., 2014; Koch et al., 2014). These methods revealed hidden differences between healthy controls and diseased patients, which manual methods could not identify. In another work (Cesari et al., 2018a), we have proposed the framework of a semi-supervised and data-driven
probabilistic method for limb movement detection and we found that leg movements detected with this data-driven approach led to high accuracy in distinguishing between healthy controls and patients suffering from periodic limb movement disorder (PLMD), which is a movement disorder mainly characterized by leg movements occurring in periodic series and is the cause of disturbed and fragmented sleep (American Academy of Sleep Medicine, 2014).

In this study we propose a modified and expanded version of the data-driven method for leg movement detection. In particular, the new method uses movement features from the chin and tibialis EMG signals, and achieves higher performances than previously proposed methods in distinguishing RBD patients from PLMD patients and healthy controls. Compared to the previous automated methods, where usually only healthy controls and RBD patients were included, we decided to include PLMD patients because it is highly relevant for clinics to distinguish RBD from a common movement disorder not involving neurodegeneration such as PLMD. Moreover, we investigate two other crucial aspects in RBD diagnosis. First, according to the standards (American Academy of Sleep Medicine, 2014), only the movements during REM sleep should be considered for RBD diagnosis. However, increased muscular activity during non-REM (NREM) sleep has been observed in RBD patients (Hanif et al., 2015; Mayer et al., 2008; Miguel and Arnulf, 2017), and therefore, we decided to analyze whether inclusion of NREM movements could improve the detection of RBD patients. Secondly, several studies recommend the removal of apneas and arousal-related movements prior to the application of automated methods for RSWA detection (Frauscher et al., 2014; McCarter et al., 2017). However, our previous comparative study of automated methods indicated that they seem to be robust to these events and their removal seems not to be necessary (Cesari et al., 2018b). Therefore, we evaluate the influence of movements related to apneas and arousals on the new data-driven method.
2. METHODS

2.1 Subjects and recordings

In the present study we included 27 healthy control subjects (C), 29 idiopathic RBD patients and 36 patients suffering from PLMD. The subjects underwent a comprehensive medication and medical evaluation and a full-night PSG study at the Danish Center for Sleep Medicine, Rigshospitalet, in the period 2009-2015. The use of medications interfering with PSG (i.e. hypnotics and anti-depressants) was advised to be discontinued two weeks before the sleep recording to minimize their effect on the study. RBD and PLMD diagnosis were made according to current guidelines (American Academy of Sleep Medicine, 2014). The study design was approved by the Danish Health Authorities and the Data Protection Agency. All recordings were anonymized and the work was carried out in accordance with the Declaration of Helsinki (The World Medical Association, 2013).

Sleep recordings were analyzed by expert technicians according to AASM guidelines (Berry et al., 2016). In particular, the entire night was divided into 30-s epochs which were manually scored as wakefulness, REM sleep or one of the three stages of NREM sleep (N1, N2 or N3). Moreover, sleep apneas, periodic leg movements and arousals were manually scored. Table 1 shows demographic and sleep information for the subjects included in the current study.

The EMG signals from the chin, tibialis left (TIBL) and tibialis right (TIBR) muscles were analyzed at a sampling frequency of 256 Hz between lights off and on. Before lights off, clinical staff checked that the electrode impedances were lower than recommended values (Berry et al., 2016). Because of different amplifiers used for the recordings, of which one had anti-aliasing cut-off frequency at 70 Hz, we decided to pre-process all the signals with a zero-phase 4th order Butterworth bandpass filter with 3dB cut-off at 10 and 70 Hz. In addition, we applied a zero-phase 4th order Butterworth notch filter at 50 Hz (3dB cut-off at 48 and 52 Hz) to remove power-line
interference. Saturation artefacts due to electrode pop-ups were removed by searching for areas in the EMG signals exceeding the empirically defined threshold of 4000 µV. To ensure a complete removal of such artefacts, we removed also the 5 seconds preceding and following them. The pre-processing and algorithms were implemented in MATLAB R2017b, The MathWorks, Natick, MA, USA.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>C</th>
<th>RBD</th>
<th>PLMD</th>
<th>Kruskal-Wallis</th>
<th>C vs RBD</th>
<th>C vs PLMD</th>
<th>RBD vs PLMD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total count</td>
<td>27</td>
<td>29</td>
<td>36</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fraction of men</td>
<td>0.48</td>
<td>0.72</td>
<td>0.61</td>
<td>3.46</td>
<td>2</td>
<td>0.18</td>
<td>- n.s.</td>
</tr>
<tr>
<td>Age [years, µ±σ]</td>
<td>56.6±9.2</td>
<td>57.7±17.2</td>
<td>58.8±14.8</td>
<td>1.69</td>
<td>2</td>
<td>0.43</td>
<td>- n.s.</td>
</tr>
<tr>
<td>#arousals/hsleep [µ±σ]</td>
<td>7.6±4.6</td>
<td>11.2±12.4</td>
<td>9.3±8.3</td>
<td>0.59</td>
<td>2</td>
<td>0.75</td>
<td>- n.s.</td>
</tr>
<tr>
<td>AHI [#apneas/hsleep, µ±σ]</td>
<td>3.3±4.9</td>
<td>11.2±16.7</td>
<td>7.6±8.9</td>
<td>5.56</td>
<td>2</td>
<td>0.06</td>
<td>- n.s.</td>
</tr>
<tr>
<td>PLMS index [µ±σ]</td>
<td>6.3±11.2</td>
<td>26.7±34.4</td>
<td>50.9±42.5</td>
<td>26.78</td>
<td>2</td>
<td>&lt;0.001</td>
<td>1.84 0.09</td>
</tr>
<tr>
<td>#nPLMS/hsleep [µ±σ]</td>
<td>5.5±5.7</td>
<td>6.7±6.1</td>
<td>7.2±6.6</td>
<td>1.34</td>
<td>2</td>
<td>0.51</td>
<td>- n.s.</td>
</tr>
<tr>
<td>#arousals/hREM [µ±σ]</td>
<td>11.0±8.6</td>
<td>16.7±17.2</td>
<td>8.5±5.9</td>
<td>4.33</td>
<td>2</td>
<td>0.11</td>
<td>- n.s.</td>
</tr>
<tr>
<td>#apneas/hREM [µ±σ]</td>
<td>7.9±13.3</td>
<td>14.7±22.7</td>
<td>10.7±14.7</td>
<td>2.29</td>
<td>2</td>
<td>0.32</td>
<td>- n.s.</td>
</tr>
<tr>
<td>PLM/hREM [µ±σ]</td>
<td>4.9±9.3</td>
<td>37.1±46.4</td>
<td>25.5±33.1</td>
<td>10.06</td>
<td>2</td>
<td>&lt;0.01</td>
<td>2.60 0.01</td>
</tr>
<tr>
<td>#nPLM/hREM [µ±σ]</td>
<td>7.5±8.9</td>
<td>9.1±8.6</td>
<td>10.2±9.3</td>
<td>1.79</td>
<td>2</td>
<td>0.41</td>
<td>- n.s.</td>
</tr>
<tr>
<td>#arousals/hNREM [µ±σ]</td>
<td>9.8±5.6</td>
<td>14.7±15.2</td>
<td>12.8±9.4</td>
<td>1.47</td>
<td>2</td>
<td>0.48</td>
<td>- n.s.</td>
</tr>
<tr>
<td>#apneas/hNREM [µ±σ]</td>
<td>2.2±3.4</td>
<td>11.6±16.6</td>
<td>7.0±8.8</td>
<td>7.69</td>
<td>2</td>
<td>0.02</td>
<td>2.11 0.05</td>
</tr>
<tr>
<td>#PLM/hNREM [µ±σ]</td>
<td>6.3±12.4</td>
<td>26.6±36.2</td>
<td>55.12±46.9</td>
<td>25.72</td>
<td>2</td>
<td>&lt;0.001</td>
<td>1.85 0.15</td>
</tr>
<tr>
<td>#nPLM/hNREM [µ±σ]</td>
<td>4.9±5.3</td>
<td>5.4±5.5</td>
<td>5.4±5.8</td>
<td>0.60</td>
<td>2</td>
<td>0.74</td>
<td>- n.s.</td>
</tr>
<tr>
<td>Sleep efficiency [% µ±σ]</td>
<td>86.9±9.3</td>
<td>75.5±24.0</td>
<td>74.4±25.3</td>
<td>7.42</td>
<td>2</td>
<td>0.02</td>
<td>2.16 0.07</td>
</tr>
<tr>
<td>Time in bed [min, µ±σ]</td>
<td>500.8±71.8</td>
<td>449.3±86.8</td>
<td>447.7±76.9</td>
<td>8.28</td>
<td>2</td>
<td>0.02</td>
<td>2.59 0.02</td>
</tr>
</tbody>
</table>
REM latency [min, μ±σ] | 94.0±42.9 | 154.7±99.0 | 114.7±73.8 | 7.56 | 2 | 0.02 | 2.62 | 0.02 | 0.94 | 0.63 | 1.95 | 0.13
---|---|---|---|---|---|---|---|---|---|---|---|---
W [%, μ±σ] | 13.0±9.2 | 20.1±15.8 | 19.7±13.4 | 6.38 | 2 | 0.04 | 1.77 | 0.13 | 2.52 | 0.04 | 0.31 | 0.91
REM [%, μ±σ] | 20.1±5.9 | 14.1±7.9 | 15.6±6.6 | 9.78 | 2 | 0.01 | 2.85 | <0.01 | 2.49 | 0.04 | 0.83 | 0.74
N1 [%, μ±σ] | 8.0±4.4 | 11.3±9.3 | 10.2±8.7 | 1.84 | 2 | 0.40 | - | n.s. | - | n.s. | - | n.s.
N2 [%, μ±σ] | 44.8±8.8 | 35.3±16.1 | 37.8±17.1 | 5.11 | 2 | 0.08 | - | n.s. | - | n.s. | - | n.s.
N3 [%, μ±σ] | 14.0±7.6 | 14.8±15.5 | 10.8±10.1 | 2.49 | 2 | 0.29 | - | n.s. | - | n.s. | - | n.s.

Table 1: Demographic and sleep data for the cohort studied. Statistical comparison of the fraction of men was made with chi-square test; all other statistical comparisons were made with Kruskal-Wallis tests. Chi-square statistics ($\chi^2$) and degrees of freedom (df) are shown for each test. In case of $p$-values < 0.05, post-hoc comparisons with Wilcoxon rank sum tests were performed and the correspondent z-statistics and $p$-values corrected with Tukey-Kramer procedure are reported. C: healthy controls, RBD: patients suffering from idiopathic REM sleep behavior disorder and PLMD: patients suffering from periodic limb movement disorder. PLMS: periodic limb movement (PLM) series; nPLM: limb movements not included in PLM series; AHI: apnea/hypopnea index, hrREM/hNREM/hsleep: hours of REM/NREM/sleep; n.s.: not significant. *data available only for 18 C, 26 RBD, and 32 PLMD patients.

2.2 Data-driven method for muscular movement probability estimation

The data-driven method was built on the assumption that healthy controls show atonia during REM sleep (Berry et al., 2016). Briefly, a set of probabilistic models of atonia was built from REM sleep of healthy controls and muscular activity was defined as segments of EMG signals having low likelihood of being atonia. Fig. 1 shows a schematic overview of the data-driven method applied to the chin EMG signals of the entire cohort included in this study. The different steps are described in more details in the following subsections and the whole process has been repeated independently for TIBL and TIBR signals as well.
Fig. 1: Schematic overview of the steps of the data-driven method for muscular activity probability estimation. From each subject, mean absolute amplitude values (MAAV) of 1-s windows (with 50% overlap) from the chin EMG signal were extracted and then normalized (MAAV') with respect to the surrounding windows. The healthy control subjects (C) were randomly divided into 3 subgroups (K={1,2,3}) of 9 subjects and for each subgroup the MAAV' values during REM sleep were used to train 5 different probabilistic models $M_{K,Dth}$ with different levels of compressions ($Dth$={0.01, 0.02, ..., 0.05}). For each model, the likelihood values of the training data of belonging to the model were calculated (likelihood vector $A$) and their $p$-th percentiles (with $p$={1, 2, ..., 10}) were used to define 10 different thresholds ($T_p$). At the same time, for each model, the likelihood of MAAV' values through the entire sleep recording of each of the remaining 18 C, 29 REM sleep behavior disorder (RBD) and 36 periodic limb movement disorder (PLMD) patients were calculated (likelihood vector $L$). For each value $i$ of $L$, the respective probability of muscular activity $p(MA)_i$ was computed proportionally to the difference $T_p - L_i$. Thus, for each of the 18 C, 29 RBD and 36 PLMD a muscular activity probability profile $p(MA)$ was obtained for each combination of $Dth$ and $p$. The same procedure was repeated independently for tibialis left and right EMG signals. In the graph, the red color represents the three different random splits of C subjects, the blue color the training of the models with different $Dth$ values and the green color the different thresholds $T_p$ in correspondence of different $p$ values to calculate movement probability profiles. [FULL PAGE]

2.2.1 Feature extraction from EMG signals

After the preprocessing described in section 2.1, each EMG signal was divided into 1-s windows with 50% overlap and for each of them, the mean absolute amplitude value (MAAV) was calculated as (Phinyomark et al., 2012):

$$MAAV = \frac{1}{N} \sum_{i=1}^{W} |x_i|$$

(1)

where $x_i$ ($i=1, ..., N$) are the EMG samples in each window.

We decided to use 1-s windows as RAI (the index showing the best performances for RBD identification in our previous work (Cesari et al., 2018b)) is calculated using 1-s windows and because it has been shown that phasic activity can be successfully identified with this window length (Fairley et al., 2012). Moreover, we visually compared the MAAV values obtained with 1-s windows and 3-s windows, and noticed that short bursts of phasic activity were excessively
smoothed when 3-s windows were used. We decided to use 50% overlap to ensure that short movements located at the edge of the windows were not smoothed down by the MAAV calculation.

2.2.2 MAAV normalization

Interpretation of raw EMG values is difficult due to the many factors that influence electrophysiological signals, therefore their normalization is fundamental for inter- and intra-subject studies (Halaki and Ginn, 2012). Thus, for each window \( n \) we normalized the feature \( \text{MAAV}_n \) with respect to the previous and following 900 windows (i.e. 7.5 minutes of recording), according to (2):

\[
\text{MAAV'}_n = \frac{\text{MAAV}_n}{\min\{\text{MAAV}_{n-900}, \text{MAAV}_{n-899}, \ldots, \text{MAAV}_{n+899}, \text{MAAV}_{n+900}\} + \epsilon}
\]  

where \( \epsilon = 10^{-5} \mu V \) to ensure \( \text{MAAV'}_n < \infty \) (Kempfner and Nikolic, 2014). The normalization of the window features at the beginning and end of the recording was allowed by repeating 900 times the first and last window features in the two extremities of the feature vector. Fig. 2 shows an example of EMG signal, MAAV features extracted and respective normalized values.

---

**Fig. 2:** Example of pre-processed electromyographic (EMG) signal, mean absolute amplitude value (MAAV) features calculated for 1-s windows with 50% overlap, and normalized features (MAAV') for the correspondent windows. [ONE COLUMN]

2.2.3 Training of the data-driven models
For training a model representing atonia, we used the assumption that healthy controls during REM sleep show a resting EMG pattern. Thus, we arbitrarily decided to randomly divide the 27 control subjects into 3 subgroups of 9 subjects. From each subgroup, we fitted a probability density with an online Parzen window kernel estimator (Kristan et al., 2011) to the MAAV’ features during REM sleep of the 9 subjects, which were considered training data. In short, this method takes the training values one at a time and treats each of them as a Dirac-delta function. After each training point is taken into account, the overall probabilistic model is refined and compressed by replacing clusters of Dirac-delta functions with Gaussian univariate distributions with optimized values of mean and variance. At the end, the final model consists of a mixture of univariate Gaussian distributions. Fig. 3 shows an example of the training process of a probabilistic model of atonia. The level of compression can be controlled by varying the parameter $D_{th}$ and for each of the 3 subgroups, we trained 5 different models by changing the $D_{th}$ value in the range 0.01, 0.02, ..., 0.05. We thus obtained in total 15 probabilistic models $M_{K,D_{th}}$ where $K=$\{1,2,3\} is the training subgroup and $D_{th}=$\{0.01, 0.02, ..., 0.05\} is the parameter controlling the compression level. A more thorough explanation of the Parzen window kernel estimator used is provided in (Kristan et al., 2011) and the respective Matlab code can be found in the authors’ webpage (Kristan, 2016). Each of the three training datasets contains approximately 45000-60000 seconds of REM sleep (a more detailed overview is reported in Table S1 in Supplemental Material) and with this amount of training data we assume the trained probabilistic models to be robust to outliers and good representations of the normal REM sleep atonia.
Fig. 3: Illustration of how a probabilistic model of atonia was trained. The normalized mean absolute amplitude values ($MAAV'$) training points obtained from REM sleep of healthy controls (blue dots) were used to define Dirac-delta functions that were then compressed into Gaussian components (black lines). The overall model (red line) is the mixture of such Gaussian distributions.

2.2.4 Muscular activity probability estimation

With this step we aimed to estimate the muscular activity probability for each 1-s window in the EMG signals of the test subjects not included in the training phase (i.e. 18 C, 29 RBD and 36 PLMD). For each trained model $M_{K,Dth}$ we considered the respective $MAAV'$ training values and calculated their likelihood of belonging to the model, thus obtaining a vector of likelihood values $A$ (Fig. 1). From this, we defined 10 thresholds $T_p$ corresponding to the $p$-th percentiles of $A$ with $p=\{1, 2, \ldots, 10\}$. Also, for each model $M_{K,Dth}$ and for each test subject, we calculated the likelihood of the respective $MAAV'$ test values throughout the whole sleep recording of belonging to the model, thus obtaining a vector of likelihoods $L$ (Fig. 1). For the $i$-th value of $L$ we calculated the muscular activity probability $p(\text{MA})_i$ of the $i$-th window as:

$$p(\text{MA})_i = \begin{cases} 0, & (T_p-L_i)<0 \\ \frac{(T_p-L_i)}{T_p}, & (T_p-L_i)\geq0 \end{cases}$$

(3)

Based on the trained model illustrated in Fig. 3, Fig. 4 presents muscular activity probabilities of nine $MAAV'$ test values. It can be seen that when $MAAV'$ test values are located in areas with high
probability density, the muscular activity probability is low and vice-versa. Fig. 5 shows the
muscular activity probabilities for each window of a pre-processed EMG signal.

In summary, for each of the RBD and PLMD patients we obtained 150 muscular activity
probability profiles, coming from the 15 models $M_{K,Dh}$ and the 10 thresholds values correspondent
to the respective $p$ values. For the control subjects instead we obtained 100 muscular activity
probability profiles, since each control subject was used in the training phase once and therefore
only 10 models $M_{K,Dh}$ were available.

![Graph](image)

**Fig. 4:** Illustration of nine $MAAV'$ values and their muscular activity probability values ($p(\text{MA})$) computed using the
trained model illustrated in Fig. 3 with threshold $T_3$. For each $MAAV'$ value (green dot), the muscular activity
probability (green bar) is computed using the trained probability model of atonia (red curve). High values of muscular
activity probability mean low probability of being atonia. [ONE COLUMN]
2.3 Subjects classification

The muscular activity probability profiles as illustrated in Fig. 5 and calculated with previous steps were used to calculate movement features from chin, tibialis left and right muscles, which were used to differentiate healthy controls, RBD and PLMD patients.

2.3.1 Movement features calculation

To calculate movement features from each of the muscular activity probability profiles (i.e. 150 for each RBD and PLMD patients and 100 for each healthy control), we decided to test three scenarios for defining a movement. In particular, a single 1-s window was considered to contain a movement if its muscular activity probability was exceeding the threshold of $H\%$ (with $H = \{30, 60, 90\}$). Clusters of movements closer than 1-s were merged as previously done in literature (Mayer et al., 2008) and an example of areas labeled as movements with the three different thresholds can be seen in Fig. 6.
Fig. 6: Identification of movements by applying different thresholds $H$ to a muscular activity probability profile ($p(\text{MA})$): a) $H = 30\%$; b) $H = 60\%$; and c) $H = 90\%$. For each 1-s window, a movement is detected if its correspondent $p(\text{MA})$ value exceeds the threshold $H$. Clusters of movements closer than 1-s are merged. [ONE COLUMN]
After detecting movements in each muscular activity probability profile, the following movement features were calculated for each of them by using manually labeled sleep stages: 1) the percentage of 1-s windows during REM sleep containing movements (%REM), 2) the percentage of 1-s windows during NREM sleep containing movements (%NREM), 3) the median inter-movement distance during REM sleep (IMREM), and 4) the median inter-movement distance during NREM sleep (IMNREM). The inter-movement distance was calculated as the distance (in seconds) between the onsets of two consecutive movements and in the case of absence of movements it was set to infinite.

For each RBD and PLMD patients we therefore obtained 3 estimations of each movement feature for a fixed combination of $D_{th}$, $p$, and $H$ values, corresponding to the three different training subgroups used for modelling atonia. On the other side for each control subject, we obtained 2 estimations of each movement feature for a fixed combination of $D_{th}$, $p$, and $H$ values, since each control was used once in the training. In order to reduce the influence of a particular training set, we decided to calculate the average value across the training cases for each feature. Therefore, for each EMG signal of each subject in the cohort we obtained 150 values for each feature, corresponding to all possible combinations of 5 values of $D_{th}$, 10 values of $p$ and 3 values of $H$. A schematic overview of the process for calculating movement features is shown in Fig. 7 for an EMG signal (either chin or tibialis left or right) of one of the RBD or PLMD patients. For the control subjects a similar overview can be drawn, but only two probability profiles were available for each combination of $D_{th}$ and $p$, since each control subject was used for training once.
Fig. 7: Schematic overview of the steps implemented for calculating the movement features from the muscular activity probability profiles \(p(\text{MA})\) of an EMG signal of a patient (i.e. RBD or PLMD) in the cohort. For each combination of the parameters \(D_{th}\) and \(p\), three muscular activity probability profiles were available corresponding to the three training subgroups \(K=\{1,2,3\}\). From each of them, movements were defined by identifying areas having movement probability exceeding one of the three selected threshold \(H\) and clusters of movement distant less than 1 s were merged. From the defined movements, the following features were calculated: \(\%_{\text{REM}}, \%_{\text{NREM}}\) (i.e. the percentages of 1-s windows in REM and NREM sleep identified as movements), \(\text{IM}_{\text{REM}}\) and \(\text{IM}_{\text{NREM}}\) (i.e. the median inter-movement distance during REM and NREM sleep calculated as distance between movement onsets). For each subject, these features were averaged across the three training cases, thus obtaining the final movement features for each combination of \(D_{th}, p\) and \(H\) values. For the control subjects a similar implementation was carried out, but with only two probability profiles available for each combination of the parameters \(D_{th}\) and \(p\), since each control was used for one of the training cases. The red color corresponds to the training cases, the blue to the models with different compression \(D_{th}\) values, the green to the different threshold likelihood values \(T_p\) and the orange one to the probability thresholds \(H\) [FULL PAGE].

2.3.2 Classification

For a fixed combination of \(D_{th}, p\) and \(H\) values, we obtained 12 features from each subject in the cohort corresponding to the 4 features \(\%_{\text{REM}}, \%_{\text{NREM}}, \text{IM}_{\text{REM}}, \text{IM}_{\text{NREM}}\) calculated for chin, tibialis left and right muscles. These features were used in a 5-fold cross validation classification to distinguish C, RBD and PLMD with 3 different optimized classifiers: a random forest (RF), a support vector machine (SVM) with linear kernel and a SVM with radial basis function (RBF) kernel. The RF classifier has been implemented in Matlab with the built-in function fitcensemble.
(Mathworks, 2018a). To solve the multi-class problem, the SVMs were implemented with a “one vs one” approach with the Matlab built-in error-correcting output codes model fitcecoc (Mathworks, 2018b) and the IM$_{REM}$ and IM$_{NREM}$ features previously set to infinity were treated as the maximum real number with double precision during the classification process. In the training set of each of the 5 folds, we optimized with Bayesian optimization (Mathworks, 2018c) and 5-fold cross-validation the number of trees and their depth for the RF classifier, and the box constraint and kernel scale for the SVM classifiers. Moreover, to avoid the influence of the larger number of PLMD patients compared to RBD patients and controls on the final performances, we set uniform prior probabilities of each class in all the classifiers. Thus, for each fold, each of the three optimized classifiers returned in output a classification probability mixture (summing to 1) of the three classes for the respective training and validation sets. For each subject in the validation set of each fold, the three classification probability mixtures were merged with the Dempster-Shafer fusion method described in (Polikar, 2006) and implemented in (Bagheri, 2016), thus obtaining a final classification probability mixture and the class with the highest probability value was chosen as the final prediction. For each fold it was possible to calculate the overall validation accuracy as:

\[
AC_{overall} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}
\]

where TP, TN, FP and FN are the number of true positives, true negatives, false positives and false negatives defined in the number of subjects classified by the algorithm in the validation group of each fold. Moreover, we calculated for each class \( i \) with \( i = \{C, RBD, PLMD\} \) the validation accuracy, sensitivity and specificity as:

\[
AC_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}, \quad SE_i = \frac{TP_i}{TP_i + FN_i}, \quad SP_i = \frac{TN_i}{TN_i + FP_i} \tag{5}
\]

where TP$_i$, TN$_i$, FP$_i$ and FN$_i$ are the true positives, true negatives, false positives and false negatives for each class \( i \).
To compare our method with other methods that analyze only muscular activity during REM sleep, we performed the same classification scheme by considering only $\%_{\text{REM}}$ and $\text{IM}_{\text{REM}}$ from the chin, tibialis left and right muscles as features. A schematic overview of the classifications we have implemented is shown in Fig. 8.

For each of the two classifications (i.e. including the features from both REM and NREM sleep, or only REM), we selected the optimal combination of $D_{\text{th}}$, $p$ and $H$ values as the one leading to the highest average overall validation accuracy across the folds.

![Fig. 8: Schematic overview of the two classification schemes implemented. For each combination of $D_{\text{th}}$, $p$ and $H$ we used the percentages of muscular activity and the median inter-movement distance during REM and NREM sleep ($\%_{\text{REM}}$, $\%_{\text{NREM}}$, $\text{IM}_{\text{REM}}$, $\text{IM}_{\text{NREM}}$) for the chin, tibialis left (TIBL) and right (TIBR) for the 27 healthy controls (C), 29 REM sleep behavior disorder (RBD) patients and the 36 patients suffering from periodic limb movement disorder (PLMD). In classification 1, features extracted from both REM and NREM sleep were considered, while only features derived from REM sleep were considered for classification 2. In both classifications, a 5-fold cross-validation technique was used and a support vector machine (SVM) with linear kernel, a SVM with radial basis function (RBF) kernel and random forest (RF) classifiers were optimized during training in each fold. In the validation, their classification outputs were merged with a Dempster-Shafer decision fusion algorithm to obtain the final validation classification output. The]
blue color corresponds to the models with different compression $D_{th}$ values, the green to the different threshold likelihood values $T_p$ and the orange to the probability thresholds $H$ [FULL PAGE].

2.4 Evaluation of the influence of apneas and arousals

In order to evaluate the influence of movements related to apneas and arousals, we decided to apply all the steps explained in sections 2.2 and 2.3 to the EMG signals after having removed movements related to manually labeled apneas and arousals. In particular, we removed from analysis the areas of signals from 5 s before an apnea onset until 5 s after their end, and from 3 seconds before an arousal onset until 12 s after the same onset. These segment lengths were chosen empirically based on visual analysis of some recordings to ensure the complete removal of movements related to apneas and arousals. This is the same approach as in our previous work (Cesari et al., 2018b).

2.5 Comparison with other methods for RSWA detection

To further evaluate our method, we decided to compare it with previously published methods designed for identifying RSWA. Our previous study has shown that among other methods, RAI (Ferri et al., 2010, 2008), FRI (Frandsen et al., 2015) and KEI (Kempfner et al., 2014; Kempfner and Nikolic, 2014) are the most powerful methods for distinguishing patients groups based on the automatically identified level of RSWA in manually labeled REM epochs (Cesari et al., 2018b). We therefore decided to apply these methods on our cohort, and use their estimates of movements in REM sleep in the same classification scheme. In short, RAI calculates the percentage of 1-s mini-epoch during REM with atonia in the chin EMG signal, FRI is an index of the percentage of 3-s mini-epochs with movements during REM sleep in the chin signal and finally KEI returns as well a percentage of 3-s mini-epochs during REM sleep with movements in chin, tibialis left and right EMG signals. More details on the methods implementation can be found in the Supplemental Material of our previous work (Cesari et al., 2018b).
We applied these three methods to the pre-processed EMG signals (Section 2.1), thus we obtained for each subject their respective values of RAI, FRI and KEI. For each index, a similar approach to the one presented in Section 2.3.2 was applied, where an optimized SVM with linear kernel, an optimized SVM with RBF kernel and an optimized RF were trained in a 5-fold cross-validation scheme and then merged with the Dempster-Shafer method to obtain the final output prediction (same random partitions of subjects and optimization settings were used). The described classification was performed in order to allow a performance comparison between the here proposed data-driven approach and other known computerized methods. However, in the original works (Ferri et al., 2010, 2008; Frandsen et al., 2015; Kempfner et al., 2014; Kempfner and Nikolic, 2014), such a classification approach was not proposed, as they aimed at providing the movement features and not the final classification. Therefore we decided to label the classification performances as RAI with classification (RAI\textsubscript{c}), FRI with classification (FRI\textsubscript{c}) and KEI with classification (KEI\textsubscript{c}).
3. RESULTS

Table 2 shows the optimal combination of $D_{th}$, $p$ and $H$ that led to the highest average overall validation accuracy across the 5 folds for the classification of C, PLMD and RBD subjects in the 4 configurations presented in this work: i) considering all movements during REM and NREM sleep; ii) considering all movements during only REM sleep; iii) considering the movements during REM and NREM sleep after removing the ones related to apneas and arousals; and iv) considering the movements during only REM sleep after removing the ones related to apneas and arousals.

<table>
<thead>
<tr>
<th>Sleep stages</th>
<th>Movements</th>
<th>Optimal $D_{th}$</th>
<th>Optimal $p$</th>
<th>Optimal $H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>REM + NREM</td>
<td>All movements</td>
<td>0.01</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>REM</td>
<td>All movements</td>
<td>0.01</td>
<td>8</td>
<td>60</td>
</tr>
<tr>
<td>REM + NREM</td>
<td>Removed apnea and arousal-related movements</td>
<td>0.03</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>REM</td>
<td>Removed apnea and arousal-related movements</td>
<td>0.03</td>
<td>8</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 2: Combinations of the parameters $D_{th}$, $p$ and $H$ that led to the highest average validation overall accuracy across the 5 folds used in the 4 classifications analyzed in this study.

Fig. 9 shows the distribution of a) $\%_{REM}$, b) $\%_{NREM}$, c) IM$_{REM}$ and d) IM$_{NREM}$ for the combination of parameters $D_{th}$, $p$ and $H$ presented in the first row of Table 2. Table 3 shows the numerical values of the features, together with detailed results of statistical analysis for group comparison. During REM sleep, RBD patients are characterized by a significantly higher percentage of movements in the chin compared to healthy controls and PLMD patients. Significantly increased muscular tone compared to controls is seen in the tibialis muscles for PLMD patients during both REM and NREM sleep and for RBD patients only during REM sleep. During REM sleep, RBD patients show the lowest average values of the median inter-movement distance in all three muscles, but significant differences are seen mainly compared to controls and not to PLMD patients. During NREM sleep, PLMD patients show the lowest values of median inter-movement distance in the tibialis muscles, which are significantly different from controls but not from RBD patients. Low values of inter-movement distance mean that the movements are close to each other, thus more
frequent. In summary, the features distributions in Fig. 9 and Table 3 show that RBD patients are characterized by significantly increased and significantly more frequent muscular activity in the chin, tibialis left and right muscles during REM than controls. Also, the data-driven method for muscular activity detection finds significantly increased and more frequent limb activity during NREM for PLMD patients compared to controls. The main significant difference between PLMD and RBD patients is the amount of muscular activity in the chin during REM sleep. Similar patterns were observed for the movement features obtained for the other optimal combinations of \(D_{th}, p\) and \(H\) and their values and statistical comparisons are shown in Tables S2, S3 and S4 in the Supplementary Material.

**Fig. 9:** Visualization of the movement features a) percentage of muscular activity during REM sleep (%\(_{\text{REM}}\)), b) percentage of muscular activity during NREM sleep (%\(_{\text{NREM}}\)), c) median inter-movement distance during REM sleep (\(\text{IM}_{\text{REM}}\)) and d) median inter-movement distance during NREM sleep (\(\text{IM}_{\text{NREM}}\)). The features were calculated for the combination of parameters \(D_{th} = 0.01, p = 3\) and \(H = 90\) and without removing any movement related to apneas and arousals. Each feature value from a subject is represented as a dot, together with bars representing the average group values and whiskers representing one standard deviation value for each group. Kruskal-Wallis tests were applied to analyze group differences for each feature. In case the p-value was lower than 0.05, Wilcoxon rank sum tests were
applied for pairwise comparisons and p-values corrected with Tukey-Kramer procedure. Statistical differences are reported as *(p<0.05), **(p<0.01) and ***(p<0.001). C: healthy controls, RBD: patients suffering from REM sleep behavior disorder, PLMD: patients suffering from periodic limb movement disorder, TIBL: tibialis left muscle, TIBR: tibialis right muscle [FULL PAGE].

Table 3: Movement feature values obtained for the optimal combination of parameters $D_{th}=0.01$, $p=3$ and $H=90$ and considering all movements shown as mean and standard deviation. Kruskal-Wallis tests were applied to analyze group differences for each feature and the respective p-values and $\chi^2$ statistics are reported. In case the p-value was lower than 0.05, Wilcoxon rank sum tests were applied for pairwise comparisons and the respective z-values and p-values corrected with Tukey-Kramer procedure are reported. C: healthy controls, RBD: patients suffering from REM sleep behavior disorder, PLMD: patients suffering from periodic limb movement disorder, %REM/%NREM percentage of muscular activity during REM/NREM sleep, IMREM/IMNREM: median inter-movement distance during REM/NREM sleep; n.s.: not significant.

Table 4 shows the values of RAI, FRI and KEI obtained for the different subject groups in the two cases of including all movements and excluding the ones related to apneas and arousals. The values of RAI and FRI showed significantly increased muscular tone in RBD patients compared to controls and PLMD patients, while KEI values showed significantly increased tone in RBD and PLMD compared to controls when all movements were considered and only significantly increased tone in RBD patients compared to controls when apnea and arousal-related movements were removed.

\[
\begin{array}{|c|c|c|c|c|c|c|c|c|}
\hline
\text{Feature} & \text{C} & \text{RBD} & \text{PLMD} & \text{Kruskal-Wallis} & \text{C vs RBD} & \text{C vs PLMD} & \text{RBD vs PLMD} \\
\hline
\%\text{REM, CHIN} & 1.12\pm1.25 & 3.75\pm3.14 & 2.04\pm2.23 & 16.44 & 2 & <0.001 & 3.94 & <0.001 \text{ p} & 1.76 & 0.20 & 2.59 & 0.03 \\
\%\text{REM, TIBL} & 1.68\pm1.69 & 10.61\pm13.03 & 7.21\pm9.59 & 21.17 & 2 & <0.001 & 4.46 & <0.001 \text{ p} & 3.27 & <0.01 & 1.46 & 0.31 \\
\%\text{REM, TIBR} & 1.93\pm4.15 & 8.53\pm10.91 & 3.66\pm3.39 & 19.49 & 2 & <0.001 & 4.08 & <0.001 \text{ p} & 3.19 & <0.01 & 1.79 & 0.24 \\
\%\text{NREM, CHIN} & 1.48\pm2.06 & 1.5\pm1.27 & 2.47\pm3.17 & 2.01 & 2 & 0.37 & - & n.s. & - & n.s. & - & n.s. \\
\%\text{NREM, TIBL} & 1.29\pm1.03 & 6.17\pm10.41 & 5.42\pm6.56 & 13.63 & 2 & <0.01 & 1.82 & 0.08 & 3.91 & <0.001 & 1.26 & 0.30 \\
\%\text{NREM, TIBR} & 1.02\pm4.9 & 5.76\pm11.19 & 4.19\pm3.09 & 17.07 & 2 & <0.001 & 2.03 & 0.06 & 4.33 & <0.001 & 1.54 & 0.19 \\
IM\text{REM, CHIN} & 180.06\pm280.53 & 45.59\pm32.96 & 61.65\pm60.08 & 8.52 & 2 & 0.01 & 2.87 & 0.01 & 2.09 & 0.08 & 0.76 & 0.69 \\
IM\text{REM, TIBL} & 113.67\pm49.52 & 32.27\pm36.21 & 48.61\pm43.1 & 25.27 & 2 & <0.001 & 4.80 & <0.001 & 3.11 & <0.01 & 2.50 & 0.05 \\
IM\text{REM, TIBR} & 124.66\pm117.9 & 36.17\pm35.71 & 45.98\pm30.87 & 12.61 & 2 & <0.01 & 2.86 & <0.01 & 3.17 & 0.02 & 1.18 & 0.71 \\
IM\text{NREM, CHIN} & 85.3\pm133.83 & 44.52\pm32.94 & 50.34\pm45.14 & 0.85 & 2 & 0.65 & - & n.s. & - & n.s. & - & n.s. \\
IM\text{NREM, TIBL} & 64.79\pm48.75 & 44.38\pm35.74 & 31.05\pm14.74 & 6.57 & 2 & 0.04 & 1.29 & 0.17 & 2.92 & 0.03 & 0.20 & 0.81 \\
IM\text{NREM, TIBR} & 88.19\pm100.09 & 36.74\pm33.76 & 28.85\pm13.15 & 12.84 & 2 & <0.01 & 2.71 & <0.01 & 3.49 & <0.01 & 0.09 & 0.99 \\
\hline
\end{array}
\]
Table 4: Values of REM atonia index (RAI), Frandsen index (FRI) and Kempfner index (KEI) shown as mean and standard deviation for the different subject groups in the two conditions of including all movements and removing the ones related to apneas and arousals. Kruskal-Wallis tests were applied to analyze group differences for each feature and the respective p-values and $\chi^2$ statistics are reported. In case the p-value was lower than 0.05, Wilcoxon rank sum tests were applied for pairwise comparisons and the respective z values and p-values corrected with Tukey-Kramer procedure are reported. C: healthy controls, RBD: patients suffering from REM sleep behavior disorder, PLMD: patients suffering from periodic limb movement disorder.

Table 5 holds the overall validation accuracies and the class validation accuracies, sensitivities and specificities as mean and standard deviation across the 5 folds when all movements were considered. The first column presents the results when movement features during both REM and NREM sleep were considered, while the second one when only REM features were used by the classifier. Generally, the proposed method achieved either similar or better performance values than the previously published methods in both cases of including and excluding NREM movement features. More specifically, when NREM features were considered, the average performances were always either similar or higher than when they were not included, with the only exception of sensitivity for healthy controls.
Table 5: Overall validation accuracies (ACC\textsubscript{overall}) and class-wise validation accuracies (ACC), sensitivities (SE) and specificities (SP) (C: healthy controls, RBD: REM sleep behavior disorder, PLMD: periodic limb movement disorder) expressed as mean and standard deviation across the 5 folds used in the classification schemes when all movements were considered. For the proposed method, the results are shown when both REM and NREM movement features were considered (first column, with the optimal combination of parameters $D_\text{th} = 0.01$, $p = 3$ and $H = 90$) and when only REM features were considered (second column, with the optimal combination of parameters $D_\text{th} = 0.01$, $p = 8$ and $H = 60$). RAI\textsubscript{c}: REM atonia index with classification, FRI\textsubscript{c}: Frandsen index with classification, KEI\textsubscript{c}: Kempfner index with classification [ONE COLUMN]

Table 6 shows the results of the classification in case of removing the movements related to apneas and arousals, with the same performance measures and structure as in Table 5. Also in this case the average performances of proposed method are generally similar or higher than other automatic methods. Again, it is seen that NREM features increased the performances compared to using only REM features. The comparison of Tables 5 and 6 reveals that the inclusion of all movements, and not only the ones not related to apneas and arousals, in general led to increased performances when both REM and NREM sleep features were used. On the other hand, when only REM sleep features were used, the performances do not differ much, with the only exception of the sensitivity for RBD, which shows a drop of around 10% in the case where movements related to apneas and arousals were not considered.
For completeness of the results, Tables S5 and S6 in the Supplementary Material show the average and standard deviation of all the performances across the 5 folds during training and validation of each single classifier and the final validation results obtained by applying Dempster-Shafer method.

Since the previous methods (RAI, FRI and KEI) were developed with the main aim of identifying RBD, it is particularly relevant to notice that our proposed method (in all the four cases here presented) achieves sensitivity for RBD identification in the same range of the previous methods and overcomes them in accuracy and specificity. These performances are boosted when all movements and NREM features are included.

<table>
<thead>
<tr>
<th></th>
<th>This work (REM+NREM)</th>
<th>This work (REM)</th>
<th>RAIC</th>
<th>FRIC</th>
<th>KEIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC_{overall} [%]</td>
<td>64.21±13.47</td>
<td>59.82±7.95</td>
<td>41.17±14.95</td>
<td>35.85±8.23</td>
<td>39.12±4.36</td>
</tr>
<tr>
<td>ACC_C [%]</td>
<td>76.08±9.26</td>
<td>78.13±10.33</td>
<td>56.37±11.39</td>
<td>62.87±9.76</td>
<td>67.19±11.60</td>
</tr>
<tr>
<td>SE_C [%]</td>
<td>72.43±31.86</td>
<td>82.29±26.68</td>
<td>30.43±23.24</td>
<td>57.38±30.51</td>
<td>51.43±7.76</td>
</tr>
<tr>
<td>SP_C [%]</td>
<td>76.49±14.54</td>
<td>76.23±10.43</td>
<td>66.23±16.26</td>
<td>65.82±21.27</td>
<td>73.56±17.26</td>
</tr>
<tr>
<td>ACC_{RBD} [%]</td>
<td>78.42±13.58</td>
<td>75.09±5.69</td>
<td>68.59±9.77</td>
<td>67.54±7.74</td>
<td>63.22±13.48</td>
</tr>
<tr>
<td>SE_{RBD} [%]</td>
<td>61.33±28.08</td>
<td>59.19±16.51</td>
<td>66.29±20.82</td>
<td>46.62±23.75</td>
<td>45.10±21.22</td>
</tr>
<tr>
<td>SP_{RBD} [%]</td>
<td>87.44±16.02</td>
<td>83.11±11.70</td>
<td>68.47±11.99</td>
<td>77.98±14.42</td>
<td>72.92±13.03</td>
</tr>
<tr>
<td>ACC_{PLMD} [%]</td>
<td>73.92±9.56</td>
<td>66.43±8.28</td>
<td>57.37±14.80</td>
<td>41.28±12.03</td>
<td>47.84±9.57</td>
</tr>
<tr>
<td>SE_{PLMD} [%]</td>
<td>61.87±11.78</td>
<td>43.87±8.70</td>
<td>25.11±18.39</td>
<td>17.08±24.84</td>
<td>29.27±10.94</td>
</tr>
<tr>
<td>SP_{PLMD} [%]</td>
<td>82.51±1.51</td>
<td>80.93±11.21</td>
<td>76.59±21.72</td>
<td>61.73±24.89</td>
<td>61.52±20.85</td>
</tr>
</tbody>
</table>

Table 6: Overall validation accuracies (ACC_{overall}) and class-wise validation accuracies (ACC), sensitivities (SE) and specificities (SP) (C: healthy controls, RBD: REM sleep behavior disorder, PLMD: periodic limb movement disorder) expressed as mean and standard deviation across the 5 folds used in the classification schemes when the movements related to apneas and arousals were excluded from the analysis. For the proposed method, the results are shown when both REM and NREM movement features were considered (first column, with the optimal combination of parameters $D_{th} = 0.03$, $p = 1$ and $H = 30$) and when only REM features were included (second column, with the optimal combination of parameters $D_{th} = 0.03$, $p = 8$ and $H = 30$). RAIC: REM atonia index with classification, FRIC: Frandsen index with classification, KEIC: Kempfner index with classification. [ONE COLUMN]
4. DISCUSSION

The current paper presents a new data-driven method for estimating muscular activity probabilities in sleep chin, tibialis left and right EMG signals. From such muscular activity probabilities, movement features expressing the coverage and frequency of movements in REM and NREM were extracted and used to successfully distinguish healthy controls, RBD and PLMD patients. There are three main outcomes of our study: i) our data-driven method achieved higher performances in the overall classification accuracy of C, RBD and PLMD groups compared to previously proposed methods for RSWA detection and in particular it overcomes them in the accuracy and specificity for RBD detection; ii) including movement features during NREM sleep leads to increased performances in the classification when compared to considering only REM sleep features; and iii) our method performs best when movements related to apneas and arousals during REM and NREM sleep are included in the analysis.

When the proposed method is compared to the three other methods implemented in this work, some important methodological differences rise. First, our method does not quantify movements based on empirically human-defined amplitude thresholds as RAI does. We decided to avoid that approach because empirically human-defined thresholds may be highly affected by noise, electrodes placement and hardware used for the recordings. Second, compared to RAI and FRI, the proposed method quantifies muscular activity in the limbs and not just in the chin. This was done mainly because limb movements in REM sleep in RBD patients were reported to increase the accuracy of RBD diagnosis (Frauscher et al., 2012). KEI also includes the analysis of tibialis muscles in the quantification of RSWA, but chin and limb movements are compressed in one index and not kept separate as in our case. Third, KEI is based on the same rationale of our method since it defines a cluster of atonia with a one-class SVM. Despite the good results achieved in the original published work (Kempfner et al., 2014; Kempfner and Nikolic, 2014), we noticed that the training of KEI
required several days due to the complex mathematical transformations performed in the one-class SVM, while our method used only few hours thanks to the relatively easy math involved in the probabilistic approach. Therefore, the probabilistic method we proposed is faster and less computationally expensive. Finally, the proposed method has another important methodological difference compared to the other methods previously proposed. This consists in the fact that our method returns a muscular activity probability profile for each muscle analyzed, from which movements are then defined. We think that a muscular activity probability profile is more informative than a simple movement/no movement detector, since it is possible to distinguish between major and minor movements that have higher and lower movement probability respectively (see for example different movements in Fig. 6). A future study should investigate the distribution of movement probabilities in different patient groups in order to analyze whether they hold important information for diagnosis.

We used three different classifiers with expected complementary performance and combined their outputs with the Dempster-Shafer decision fusion technique to obtain the final classification of groups. The Dempster-Shafer decision fusion method has been proven to work effectively in several scenarios (Polikar, 2006), and we made this choice for mainly two reasons: First, we observed that if we were using only one of the three classifiers, it was not guaranteed that it was performing well for all the methods from literature (RAI, FRI and KEI), and it could therefore introduce a bias beneficial only for our method in the analysis of the final performances. Secondly, we noticed that the combination of the outputs of the three classifiers generally increased the overall performance of the classification. These observations were supported by the analysis of the performances of the single classifiers in reported Table S5 and S6 in the Supplementary Material. However, we cannot exclude that using another classification technique would have led to different results from the ones here presented.
The analysis of Tables 5 and 6 reveals that our method outperforms RAI, FRI and KEI in group classification. However, it should be kept in mind that these three indices have been originally implemented for identifying RBD subjects from the level of automatically detected RSWA level and not strictly for distinguishing controls, RBD and PLMD subjects. Therefore, a fair comparison of our proposed method with RAI, FRI and KEI can be carried on only by considering the accuracy, sensitivity and specificity of RBD patients when only features extracted during REM sleep were taken into account. Such analysis reveals that our method achieved always higher average specificity and accuracy for RBD identification and sensitivity in the same ranges compared to the other methods. The only exception is the index RAI that achieved a higher average sensitivity of about 7% when apnea and arousal related movements were excluded. However, we can conclude that in general our method is as sensitive as previous methods for RBD identification, but it is more accurate and specific. This improvement may be due to the fact that our method includes not just the percentage of muscular activity during REM, but also the median inter-movement distance, which can be considered a measure of the frequency of movements. Therefore, a more accurate and specific identification of RBD patients can be achieved if both quantity and frequency of movements are taken into account.

The accuracy, sensitivity and specificity in identifying RBD patients were further increased when the percentage of movements and the median inter-movement distance during NREM were considered, both when all movements were considered and when apnea and arousal-related muscular activities were removed. This means that RBD patients are characterized not only by increased and more frequent movements during REM sleep, but also by increased movement patterns in NREM sleep. This is in line with the study from Mayer et al., where an increased mean specific muscle tone was observed in RBD patients compared to controls (Mayer et al., 2008), with the results previously presented by our group on increased muscular tone in RBD patients in NREM.
sleep (Hanif et al., 2015) and with a case study of an RBD patient, where abnormal behaviors and abnormal phasic muscular activity has been reported during NREM sleep (Miguel and Arnulf, 2017). Also, our finding agrees with the significantly increased activity score, activity index and short burst inactivity index in RBD patients compared to controls in actigraphy analysis (Stefani et al., 2018). No distinction between REM and NREM sleep are made from actigraphy analysis, and it is assumable that RBD patients are characterized by increased muscular activity during both REM and NREM sleep. We think that this is an important observation that may potentially lead to revise the definition of RBD and the guidelines for its diagnosis.

The inclusion of features during NREM sleep boosts also the overall accuracy, and the three performance measures for PLMD detection in both cases of considering all movements and removing the ones related to apneas and arousals. This was expected, since limb movements in PLMD patients are known to occur mostly during NREM sleep (American Academy of Sleep Medicine, 2014). For healthy controls, the inclusion of NREM sleep features led to a drop in sensitivity, while accuracy and specificity either increased or remained at the same level (Tables 5 and 6).

In our previous analysis on RAI, FRI, KEI and other automated methods, we showed that there are no significant changes in the performance measures whether or not movements related to apneas and arousals are included (Cesari et al., 2018b). However, in the current study, some changes can be seen for some of the performance measures of RAI, FRI and KEI (Table 5 and 6). Due to the classification approach used in this work, a statistical analysis is not comprehensive, but based on our previous results we assume that the changes we found are not significant.

It is interesting to analyze how our method’s performance changes when considering all movements or removing the ones related to apneas and arousals. For identification of RBD patients, the average sensitivity is higher when all movements are considered both when including and
excluding NREM sleep features. Here, all other performances remain stable. For PLMD identification, the same pattern is observed, but only when both REM and NREM sleep features are considered. For healthy controls, the performances are not changing when movements related to apneas or arousals are included or not. From this we can conclude that including all movements during REM and NREM sleep increases the capability of the classifier of distinguishing RBD and PLMD patients.

From this analysis, we recommend to apply our method to both REM and NREM sleep and to include all movements. Thus, this study shows that movements traditionally considered as noise, indeed hold information that can help in distinguishing groups when using data-driven methods.

This study has some limitations. First, we cannot exclude that other window lengths than 1-s and overlaps of 50% could lead to even better final classification results. Second, in the pre-processing step we removed only the artefacts caused by electrode pop-up and we did not consider any other artefact, such as snoring which might affect the chin signal and ECG interference in all three muscular signals. Third, our method is not fully automated since it still requires manual scoring of sleep stages and therefore it does not completely solve the problem of time consumption of manual analysis of sleep recordings. To our knowledge, only one attempt has been made until now for a fully automated RSWA detector (Kempfner et al., 2014), but this method is computationally expensive and has been validated only on a small cohort. To provide a fully automated system for RSWA and RBD identification, the here proposed method could e.g. be combined with a REM/NREM/wake detector trained on EEG and EOG, in order not to have the confounding influence of EMG in the classification. This approach is however challenging due to the poor performances that automated sleep staging detectors have been shown until now in RBD patients (Boostani et al., 2017). Alternatively, in order to have a fully data-driven approach, our muscular activity detector can be combined with a previously proposed data-driven method for sleep staging.
Fourth, in this study we did not include patients suffering from severe sleep apnea, which might be a confounder in RBD detection (Iranzo and Santamaría, 2005). The inclusion of patients with severe sleep apnea might change our considerations concerning the removal of movements related to apneas and arousals. Fifth, the healthy controls that we included in this study did not show abnormal RSWA level and thanks to this, it was possible to define a probabilistic model of atonia. However, it is known that the use of anti-depressants is associated with higher prevalence of RSWA but not RBD (Lee et al., 2016) and that some healthy controls actually show RSWA (Ferri et al., 2018). In our study, we have not included subjects with high level of RSWA, but not diagnosed with RBD, therefore we cannot make any conclusion whether this method is able to differentiate subjects with high RSWA level from RBD patients and future studies should investigate that. Sixth, recent studies carried out by the SINBAR group suggest the inclusion of the flexor digitalorum superficialis EMG signal for RBD diagnosis (Frauscher et al., 2012). We could not include these signals in our study, but we think that our method can be easily adapted to these signals in order to identify muscular activity from the arms. Finally, we have included healthy controls, RBD and PLMD patients on a ratio 1:1 as done in the previous studies (Ferri et al., 2010, 2008; Frandsen et al., 2015; Kempfner et al., 2014; Kempfner and Nikolic, 2014), which is not representative of the real distribution in the general population (Haba-Rubio et al., 2018; Scofield et al., 2008). A future study should investigate the performance of the proposed data-driven method on a larger database representing the distributions of patients in the general population.
CONCLUSIONS

We have proposed a new data-driven method for 1) identifying muscular movements in chin and tibialis muscles during sleep and 2) extracting features representing degree and frequency of movement in REM and NREM sleep in order to classify healthy controls, RBD and PLMD patients. The features used to classify the groups were obtained using REM sleep in healthy controls to build models of atonia from which muscular activity probability profiles could be computed and used to define movements. Our results show that the proposed method overcomes other literature methods (RAI, FRI, KEI) in particular concerning the accuracy and specificity for RBD identification. Moreover, it was observed that analysis of NREM muscular activity increased the performances for RBD identification, suggesting that movements in NREM sleep hold diagnostic information in RBD. Finally, the results suggest that our method achieves the best performances when movements related to apneas and arousals are not removed. More specifically, the overall average validation accuracy obtained when NREM movement features and arousal and apnea-related movements were included was 70.8%, which is higher compared to 61.9% (obtained when only REM features and arousal and apnea-related movements were included), 64.2% (obtained when REM and NREM features were included and arousal and apnea-related movements excluded) and 59.8% (obtained when only REM features were included and arousal and apnea-related movements excluded).

We think that this method opens a new way of detecting muscular movement during sleep with a data-driven approach and supports the hypothesis that data-driven algorithms can improve the quality of analyzing sleep studies. In future, the proposed method should be tested in databases from other clinics to analyze its performances and the robustness to inter-clinical variability.
FUNDING SOURCES

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
REFERENCES


Cesari, M., Christensen, J.A.E., Jennum, P., Sorensen, H.B.D., 2018a. Probabilistic Data-Driven


Högl, B., Stefani, A., Videnovic, A., 2018. Idiopathic REM sleep behaviour disorder and


https://doi.org/10.1097/WNP.0000000000000021

https://doi.org/10.1097/WNP.0000000000000065

https://doi.org/10.1016/j.jneumeth.2014.07.002

Kristan, M., 2016. oKDE and odKDE [WWW Document]. URL
http://www.vicos.si/File:Maggot_v3.5.zip

https://doi.org/10.1016/j.patcog.2011.03.019

https://doi.org/10.1212/WNL.42.7.1371

https://doi.org/10.5664/jcsm.5582
Mathworks, 2018a. Fitcensemble [WWW Document]. URL
https://se.mathworks.com/help/stats/fitcensemble.html

Mathworks, 2018b. Fitcecoc [WWW Document]. URL
https://se.mathworks.com/help/stats/fitcecoc.html

Mathworks, 2018c. Bayesian Optimization Algorithm [WWW Document]. URL
https://se.mathworks.com/help/stats/bayesian-optimization-algorithm.html#bvaz8tr-1

Quantification of tonic and phasic muscle activity in REM sleep behavior disorder. J. Clin.
Neurophysiol. 25, 48–55. https://doi.org/10.1097/WNP.0b013e318162acd7

Diagnostic REM sleep muscle activity thresholds in patients with idiopathic REM sleep behavior disorder with and without obstructive sleep apnea. Sleep Med. 33, 23–29.
https://doi.org/10.1016/j.sleep.2016.03.013

https://doi.org/10.1016/j.sleep.2016.09.006


https://doi.org/10.1016/j.eswa.2012.01.102

https://doi.org/10.1109/MCAS.2006.1688199

Postuma, R.B., Iranzo, A., Hogl, B., Arnulf, I., Ferini-Strambi, L., Manni, R., Miyamoto, T., Oertel,


