POLITECNICO DI TORINO
Repository ISTITUZIONALE

Single-channel EEG classification of sleep stages based on REM microstructure

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

(Article begins on next page)

20 April 2024

ORIGINAL RESEARCH PAPER

Single-channel EEG classification of sleep stages based on REM microstructure

Irene Rechichi[1] | **Maurizio Zibetti[2]** | **Luigi Borzì[1]** | **Gabriella Olmo[1]** |
Leonardo Lopiano[2]

[1] Department of Control and Computer Engineering, Politecnico di Torino, Torino, Italy

[2] Department of Neuroscience "Rita Levi Montalcini", Università degli Studi di Torino, Torino, Italy

Correspondence

Department of Control and Computer Engineering, Politecnico di Torino, Torino, 10129, Italy.
Email: gabriella.olmo@polito.it

Abstract

Rapid-eye movement (REM) sleep, or paradoxical sleep, accounts for 20–25% of total night-time sleep in healthy adults and may be related, in pathological cases, to parasomnias. A large percentage of Parkinson's disease patients suffer from sleep disorders, including REM sleep behaviour disorder and hypokinesia; monitoring their sleep cycle and related activities would help to improve their quality of life. There is a need to accurately classify REM and the other stages of sleep in order to properly identify and monitor parasomnias. This study proposes a method for the identification of REM sleep from raw single-channel electroencephalogram data, employing novel features based on REM microstructures. Sleep stage classification was performed by means of random forest (RF) classifier, K-nearest neighbour (K-NN) classifier and random Under sampling boosted trees (RUSBoost); the classifiers were trained using a set of published and novel features. REM detection accuracy ranges from 89% to 92.7%, and the classifiers achieved a F-1 score (REM class) of about 0.83 (RF), 0.80 (K-NN), and 0.70 (RUSBoost). These methods provide encouraging outcomes in automatic sleep scoring and REM detection based on raw single-channel electroencephalogram, assessing the feasibility of a home sleep monitoring device with fewer channels.

1 | INTRODUCTION

Human sleep, according to the Rechtschaffen and Kales (R&K) scoring criteria [1], is divided in stages $N1$ to $N4$ (from light to deep) and stage R (rapid eye movement – REM [2]). This latter exhibits mixed frequency, low voltage, skeletal muscle atonia and bursts of rapid eye movements [3]. It accounts for 20–25% of total night-time sleep in healthy subjects, and plays a fundamental role in the processing and consolidation of memories [4] and in the regulation of emotional states [5]. On the other hand, the REM behaviour disorder (RBD) [6] is a parasomnia featuring loss of physiological muscle atonia during REM sleep and abnormal sleep behaviour, manifested by dreams enactment and increased violence [7]. RBD is related to poor quality of life (QoL) and psychological disorders such as anxiety [8]. Evidence shows that RBD is a precursor to Parkinson's disease (PD) and other neurodegenerative diseases [9], and about 40% of PD patients complain of RBD and nocturnal hypokinesia, [10].

More in general, sleep disorders are increasing in the aging population worldwide, and simple sleep monitoring systems would help improving their QoL. The gold standard to diagnose and monitor sleep disorders is polysomnography (PSG), a collection of recordings that include electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), electrocardiogram, as well as pulse oximetry and photoplethysmography. However, PSG is costly, impractical and inconvenient for patients, because of the high number of electrodes employed and the non-familiar environment. Actually, the test should be performed over two nights, as the first one is affected by the }}first night effect″, *that is* very low sleep quality. Moreover, PSG recordings are scored manually by a sleep expert, and the rating process is subjective and time-consuming (the medical score is often available after many working days). Different studies in literature aim at providing automatic multi-stage sleep classification using several PSG signals [11–13], with results comparable to expert annotations, but considerable complexity. Other

studies address only EEG channels to perform multi-stage or stage-specific classification, achieving good performance [14, 15]. Some of these works adopt a feature-based approach [16, 17], while more recent ones address deep learning (DL) [18, 19] and attention mechanisms [20] on data from one or two EEG channels [21], with good performance. Hence, even though PSG and visual scoring of sleep epochs remain the clinical gold standard, automatic sleep scoring represents a promising approach for sleep disorder follow-up and longitudinal studies. The main contributions of this work are the implementation of algorithms for automatic sleep scoring (with a focus on REM sleep identification) from a single-channel raw EEG, and the definition of novel features based on REM sleep microstructures. It represents a feasibility test for sleep monitoring based on a single electrode, possibly to be performed at home, thus alleviating the inconvenience for the subject and the sleep expert.

2 | DATA

The data employed in this work belong to the DREAMS Subjects Database [22], available online. It is a collection of PSG recordings from 20 healthy individuals (four males) with neither underlying neurological pathology nor sleep disorders. At the time of recording they were taking no medication. Most subjects are aged 18–25, but 25% of the participants belong to the 45+ class (age 33.5 ± 14 years). The mean recording time is 8h 30m. All PSG recordings were annotated by an expert, according to both the R&K and the American Academy of Sleep Medicine (AASM) criteria [23]. The AASM annotation, used in this paper, scores each 30 s epoch in one out of five stages: awake (AWA), non-rapid eye movement sleep (arranged in stages $N1$, $N2$, $N3$, from light to deep) and REM. Ten subjects have been included in our study, for which regular sleep cycles can be identified, and at least 50 min REM periods are indeed present. The other PSGs have been discarded due to the irregularity of the sleep cycle or the absence of REM and $N3$ epochs, possibly due to the first-night effect [24]. According to the AASM criteria, non-classifiable sleep stages are labelled either 0, −1 or −2. These epochs are discarded too as they are not relevant. The final dataset is made of $N = 8.382$, 30 s epochs. As detailed in Table 1, the dataset is quite imbalanced towards the $N2$ class, a common situation because $N2$ accounts for 45–55% of the total sleep cycle [25]. On the other hand, $N1$ is very little represented, as it represents a transitional stage from AWA to $N2$. The EEG signals have been re-sampled at 256 Hz. All recordings have been pre-processed to reduce high-frequency noise and remove slow drifts. The signals were high-pass filtered with an IIR Chebyshev type I, order 1, cut-off frequency 0.5 Hz, and low-pass filtered with an IIR Chebyshev type I, order 11, cut-off frequency 40 Hz. In both cases an anti-causal filter (zero lag) has been used to avoid delay. Since the algorithm is based on raw EEG data, no further processing has been performed (e.g. no artefact removal, no spatial filtering).

3 | THE MICRO-STRUCTURE OF REM SLEEP

Even though it is commonly treated as a homogeneous state, the first evidence of the presence of two micro-states in paradoxical sleep dates back to the 1960s [26]. Such micro-states, denoted as phasic and tonic periods, represent markedly different brain states as regards cortical activity and information processing [3], and alternate during REM sleep. The tonic stage (TREM) is the longest and most quiescent one. It consists of segments with no significant ocular movements (EOG depolarizations lower than $25\ \mu V$ in a 4 s range) and features muscle atonia. On the other hand, the phasic stage (FREM) is characterized by bursts of rapid-eye movements (at least two consecutive depolarizations in a 4 s range), sawtooth waves, and irregular cardiac and respiratory activity. In automatic classification based on EEG signals only, REM sleep can be mistaken for AWA or $N1$ as it exhibits low-amplitude and mixed frequency components. Hence, given that TREM and FREM show very recognizable characteristics, a key idea of this paper is to define features typical of either micro-state, and use them, in addition to others, to feed the classification algorithms. In this way, the classification performance should improve, especially as regards REM stage. In more detail, it is known that the micro-states differ in terms of power spectral density (PSD): TREM (FREM) exhibits increased power in the alpha and beta (delta and theta) frequency ranges respectively [27]. Hence, we define two frequency bands significant for extracting features specific of either REM micro-structure: FREM and TREM bands. First of all, we estimate the PSD for each REM epoch. Then, we evaluate the median frequency (SEF50) and the spectral edge frequency at 95% (SEF95), that is the frequency below which 95% of total power lies. Finally, the FREM (TREM) frequency bounds are obtained by averaging SEF50 and SEF95 values belonging to the 25th and 75th percentiles of the related distributions, respectively. The obtained FREM and TREM bands turned out to be 2 – 8 Hz and 7 – 16 Hz, respectively. In Figure 1, sample SEF50 and SEF95 are shown as functions of the epoch number; the frequency bounds are reported as dotted lines. A similar approach is described in [28] and used to distinguish REM sleep from AWA and S1. Moreover, in [14] similar results are obtained based on the epoch PSD. As an example, in Figure 2 the power spectrum of two different REM epochs is depicted. The spectrum of Figure 2(A) is skewed towards the lower frequencies, suggesting a FREM behaviour whereas in Figure 2(B) the spectrum is more peaked and centred in the alpha and beta bands, suggesting a TREM micro-state. The values of SEF50 and SEF95 are reported as dotted lines and reveal a good accordance with the frequency bands defined in the present work.

TABLE 1 Number of epochs for each sleep stage

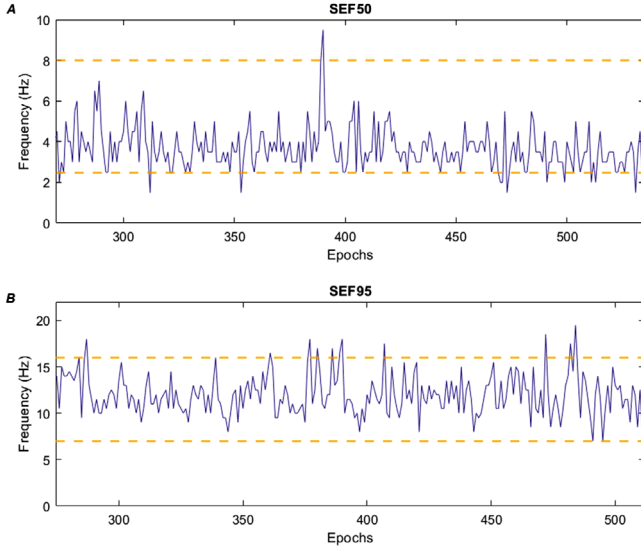| AWA	REM	N1	N2	N3
951	1347	555	3791	1738

FIGURE 1 Sample (A) SEF50 and (B) SEF95 as functions of the epoch number

4 | FEATURE EXTRACTION

In this work, a total of 164 features have been extracted from the dataset and divided in four categories: time, frequency, time-frequency and non-linear. A combination of features already proposed in literature and novel ones is employed. Time features have been calculated for each 30 s epoch, whereas features belonging to the other three classes have been evaluated in each 1 s sub-epoch and then averaged across the 30 s epoch. This approach is widely adopted in sleep stage classification, with typical sub-epochs of 2, 5 or 10 s [12, 14]. In fact, given that the EEG signal is not stationary, shorter windows guarantee wide-sense stationarity. Furthermore, the method provides an adequate spectral resolution (1 Hz). A list of the extracted features is described in Table 2, possibly along with the reference of the paper(s) where they have been proposed. Many features are self-explaining. The discrete wavelet transform (DWT) was applied to the EOG signal in [13] and [31]. In this work it is applied to the EEG signal, and several related numerical and statistical measures are used as features. The Teager–Kaiser energy operator (TKEO) has been calculated for the whole spectrum (0 – 40 Hz) and its numerical and statistical measures adopted as features, whereas in [31] only two power bands were taken into account. All features have been subjected to min-max scaling, setting the normalized range in [−1,1]. As already discussed, the novel features proposed in this work are based on the REM sleep micro-structure and encompass the absolute and relative power in FREM and TREM bands, along with the energy density in these frequency bands, the spectral features SEF50, SEF95 and the differential frequency (SEFd), which consists in the difference between SEF95 and SEF50 [14]. As for feature selection, we have evaluated the variance of the extracted features and removed those with negligible variance; a threshold of 0.2 (heuristically selected) was applied and all the
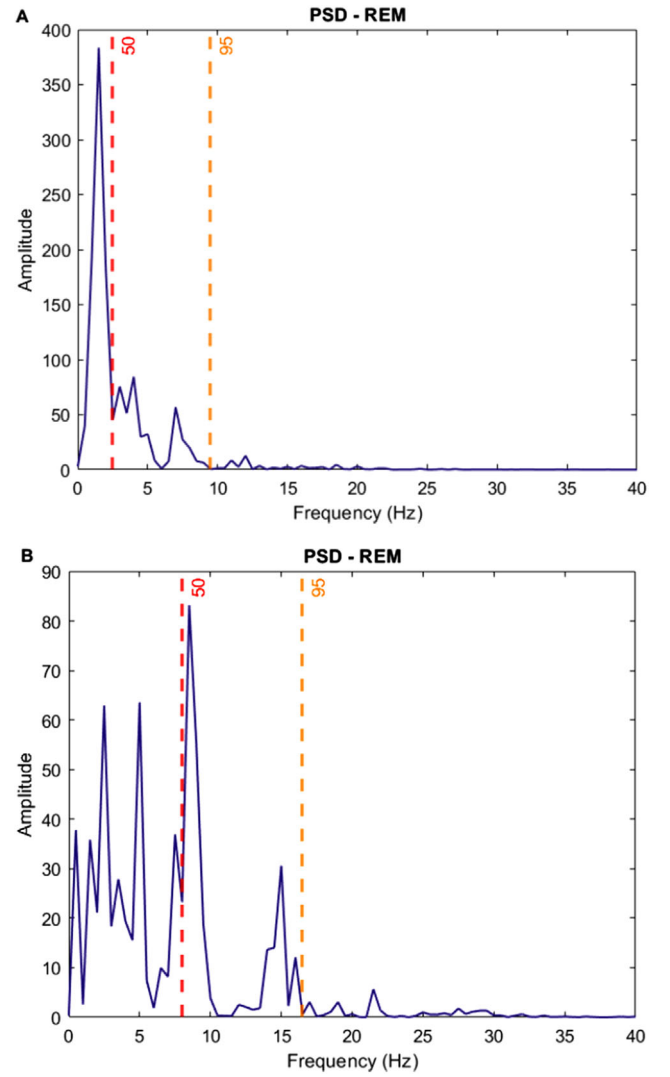

FIGURE 2 Power spectral densities of typical (A) FREM and (B) TREM micro-states. Median frequency (in red) and spectral edge frequency at 95% (orange) are shown

features not meeting this criterion removed. The 87 remaining features have been used to train supervised models, as described in the following section.

5 | AUTOMATIC SLEEP STAGE CLASSIFICATION

We applied a non-parametric classification method (K-NN) and two ensemble learning methods (RF, boosted trees), described in the following, along with the main relevant parameters.

K-NN classifies observations based on their similarity to a given metric. It assigns a weight to each observation, depending on its distance to the other points in the dataset. Then, it selects the K-top observations, that is closest to the example, and chooses the most recurrent label. In this work, after

TABLE 2 Adopted features, along with proper references

Category	Name and description	Ref.
Time	Numerical and statistical measures (mean, standard deviation, skewness, kurtosis, range, max, min)	various
	Hjorth parameters (signal and its derivative)	[29]
	Zero crossing rate	[30]
	25th, 75th, 95th percentile and their differential	various
	Envelope: number of peaks, peak prominence, peak width	**
	Coastline (first and second derivative)	[13]
Frequency	Power percentage for each clinically relevant band	various
	SEF50, SEF95, SEFd, absolute power, relative power (TREM)	*[14]
	SEF50, SEF95, SEFd, absolute power, relative power (FREM)	**
	Entropy and approximate entropy	**
	Fast Fourier transform: numerical and statistical measures	various
	Relative power for each clinically relevant band	various
	Energy density in tonic and phasic REM	**
Time-frequency	Short time Fourier transform: magnitude and maximum value of its density (0 – 40 Hz)	**
	Discrete wavelet transform coefficients: Daubechies order ouri and Haar filter wavelet	[13]
Non-Linear	Teager-Kaiser energy operator: numerical and statistical measures	*[31]

*adapted from the indicated study
**novel features proposed in this study

heuristic optimization, the classification parameters are set as follows:

- Number of neighbours $K = 10$;
- Distance measure: Euclidean distance.

RF is an ensemble learning classification method, and consists of a high number of decision trees. Each tree is provided with a random subset of the available observations; each node of the tree uses a randomly selected subset of the provided features – thus reducing the risk for overfitting. The chosen parameters are:

- Number of learners: 30;
- Maximum number of nodes: $0.2 \cdot NF$, with NF being the number of features used to train the model.

RUSBoost is a particular division of boosted trees, in the form of random-under-sampling. This class of random forests has proved to perform well when learning from imbalanced datasets [32], a common situation in sleep stage classification patterns. The algorithm takes N as the basic unit for sampling. N is picked as the number of observations of the least repre-

TABLE 3 Performance of Random Forest classification

	AWA	REM	N1	N2	N3
Accuracy	0.977	0.927	0.940	0.869	0.954
Sensitivity	0.900	0.910	0.190	0.900	0.860
Specificity	0.980	0.950	0.960	0.840	0.976
Precision	0.880	0.760	0.540	0.840	0.900
F1-score	0.890	0.828	0.281	0.869	0.880

sented class – in our database, this corresponds to $N1$ class. Classes with more observations are under-sampled to N, in order to obtain a balanced dataset. The number of learners is again set at 30.

6 | RESULTS OF 5-STAGE CLASSIFICATION

We have employed the described algorithms to address the 5-stage sleep classification problem ($N1$–$N3$, AWA, REM) using data from healthy subjects. Performance is evaluated in terms of sensitivity, specificity, accuracy, precision, and F1-score. Table 3 reports the micro-averaged performance of the RF classifier; micro-averaging was chosen in order to take dataset imbalance into proper account, and k-fold cross validation ($k = 10$) is addressed. The performance is generally satisfactory, with the exception of $N1$ stage, which exhibits low sensitivity (hence, precision and F1-score). This behaviour is shared by all the tested algorithms. As already discussed, this is due to the peculiarity of $N1$, which is poorly recognizable and can be better described as a transition between AWA and $N2$ than an independent stage; this makes the opportunity of its inclusion in the classification task questionable. The impaired performance on $N1$ is also related to the dataset imbalance, as this class is very little represented (about 6% of the total sleep epochs), and under-sampling is not recommended, as it would waste most information. In any case, it can be appreciated that RF achieves an average accuracy of 83.3%, evaluated as in Equation (1) with TP, TN, FP, FN being true positives (negatives) and false positives (negatives) respectively:

$$accuracy = \frac{TP + TN}{TP + TN + FN + FP} \qquad (1)$$

Table 4 shows the micro-averaged performance of K-NN; again, k-fold cross validation ($k = 10$) is employed. It can be noticed that K-NN scored an overall accuracy of 83.5%, comparable to RF. Both models achieve very good performance on REM stage classification (F-1 score > 0.75), with high precision and recall. As for RUSBoost, the dataset has been divided in 80% training and 20% test subsets. The epochs are randomly selected. The algorithm performance, summarized in Table 5, yielded an overall accuracy of 70.1%, impaired with respect to RF and K-NN. This is not surprising, given the under-sampling approach followed by this method and the small available data
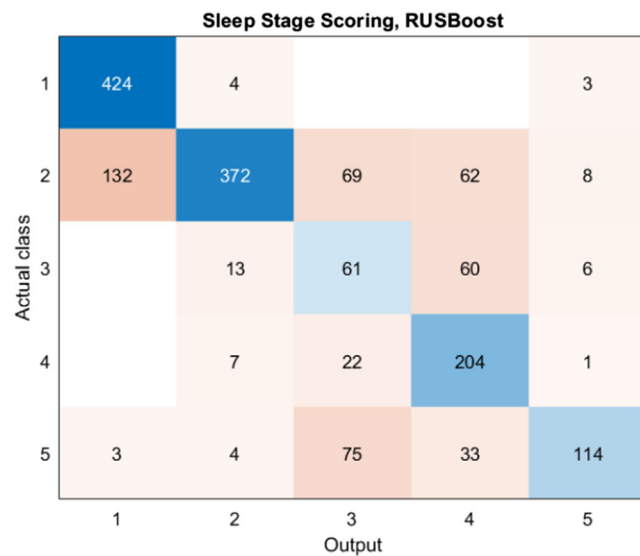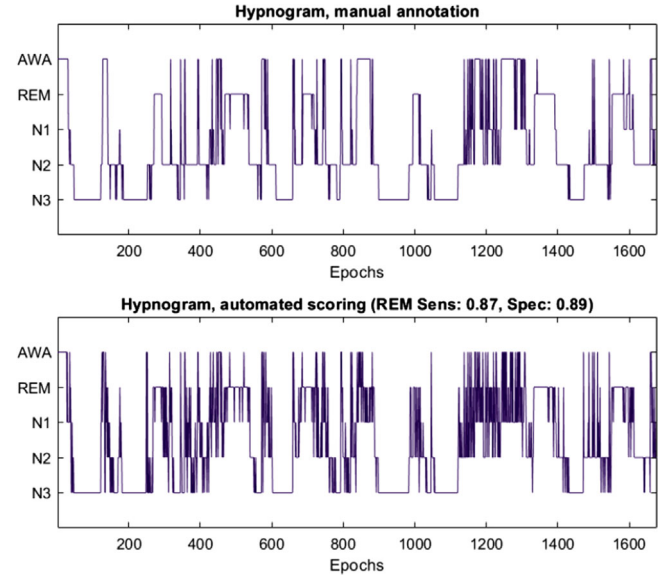
TABLE 4 Performance of K-NN classification

	AWA	REM	N1	N2	N3
Accuracy	0.979	0.925	0.943	0.871	0.953
Sensitivity	0.890	0.850	0.200	0.880	0.880
Specificity	0.989	0.940	0.990	0.861	0.971
Precision	0.910	0.740	0.570	0.850	0.880
F1-score	0.900	0.791	0.296	0.865	0.880

TABLE 5 Performance of RUSBoost classification

	AWA	REM	N1	N2	N3
Accuracy	0.921	0.890	0.854	0.822	0.897
Sensitivity	0.498	0.872	0.436	0.579	0.984
Specificity	0.988	0.899	0.892	0.973	0.892
Precision	0.864	0.568	0.269	0.930	0.758
F1-score	0.632	0.688	0.333	0.714	0.856

FIGURE 4 An example of comparison between hypnogram manual annotation and RUSBoost classification results

set. However, RUSBoost provides encouraging performance in terms of accuracy (89%), specificity (89.9%) and sensitivity (87.2%) on REM class, as can also be inferred by the confusion matrix reported in Figure 3. The REM stage is mistaken with $N2$ (3.7%), $N1$ (3.6%) and AWA (2.0%). A comparison between manual scoring and automated RUSBoost scoring is reported in Figure 4. It is worth noticing that the automatic scoring exhibits a more fluctuating trend if compared to manual annotation. This is quite reasonable, as manual annotation of hypnograms is driven by the human interpretation, which, for example, tends to rule out short AWA periods embedded in REM stages. In most papers addressing automatic hypnogram scoring, this phenomenon is by-passed by adding a smoothing stage after the automatic scoring [33]. Finally, the performance

TABLE 6 Performance comparison: RF, K-NN, and RUSBoost (proposed) and already published methods

Classifier	Study 1[*]	Study 2	This study		
	Threshold	RF	RF	K-NN	RUS
Accuracy	0.885	0.96	0.927	0.925	0.89
Sensitivity	0.823	0.83	0.91	0.85	0.872
Specificity	0.893	0.98	0.95	0.94	0.899
Precision	N/A	0.84	0.76	0.74	0.568
F-1 score	N/A	0.81	0.828	0.791	0.688

[*]same dataset as our study

yielded by our classifiers have been compared to those of already published methods. Study 1 [14] employs the same dataset as the present work (DREAMS Subjects Database), whereas Study 2 [12] was trained and tested on combined datasets of healthy controls (HC) and RBD patients (MASS, CAP and proprietary data collected at John Radcliffe Hospital, not publicly available). This dataset showed male predominance, while 80% of the participants in the DREAMS Subjects Database were female. Moreover, it employs features extracted from EEG, EOG and EMG signals; for all these reasons, only indirect comparisons can be done in this case.

The results related to REM class are shown in Table 6. It can be noticed that all the three methods proposed in this paper outperform Study 1 [14], which is based on the same dataset, as for accuracy, sensitivity and specificity. Data regarding precision and F1-score of Study 1 are not provided; nevertheless, RF and K-NN exhibit reasonable precision and a good F1-score. The performance of Study 2 refers to the HC group. It can be appreciated that, even though the proposed methods do not outperform [12] in overall accuracy of REM class, both RF and K-NN exhibit higher sensitivity and comparable

FIGURE 3 Confusion matrix yielded by the RUSBoost classifier. The class labels (1–5) represent, in order, $N3$, $N2$, $N1$, REM, AWA

TABLE 7 Performance of RF classification on RBD subjects

	AWA	REM	N1	N2	N3
Accuracy	0.881	0.895	0.948	0.826	0.866
Sensitivity	0.852	0.581	0.411	0.650	0.761
Specificity	0.893	0.955	0.986	0.881	0.903
Precision	0.757	0.715	0.683	0.620	0.733
F1-score	0.802	0.641	0.513	0.635	0.747

specificity and F1-score. We deem these results quite remarkable, as our method is very simplified with respect to Study 2, and employs a single EEG channel for classification. Moreover, the combined dataset used in [12] encompasses 6360 observations for REM class, against only 1347 ones available in our data.

7 | RESULTS ON RBD DATASET

The final objective of our work is to study pathological subjects, in particular those affected by RBD, with a single channel EEG. Hence, besides the preliminary test on the capability of our algorithms to correctly classify the five sleep stages in healthy subjects, we have considered the PSG of 22 subjects affected by RBD (19 males, aged 71 ± 6 years) enclosed in the CAP Sleep Dataset [34] available on PhysioNet [35]. EEG recordings related to the C3-A2 channel (C4-A1, if not available) were segmented into 30-s epochs for feature extraction (cf. 4). The total number of available epochs was 14583; however, data exhibited a vast prevalence of the $N2$ class, with more than 5000 observations versus 688 of the least represented class, $N1$. As this could cause classification bias, our choice was to undersample the $N2$ observations to $N = 2965$, that is the mean number of observations in the other three classes. The total number of available epochs for feature extraction was 12277. The extracted features match those formerly implemented (cf. Table 2). Dimensionality reduction was performed, excluding NaNs and low-variance features (cf. Section 4). Sleep stage classification is addressed by means of the three already introduced classifiers, trained on available data and validated through k-fold cross-validation, ($k = 5$). For the sake of brevity, Table 7 reports the results of the RF classifier only. The results of this model are rather satisfactory, with an overall accuracy of 87.11%.

8 | RESULTS OF BINARY CLASSIFICATION

As discussed in Section 3, we propose to exploit the dual nature of the REM stage (TREM and FREM micro-structure) to enhance classification. To this end, a set of novel features describing the two micro-states has been implemented in our model (cf. Table 2). To acknowledge the contribution of such features in the classification task, a binary classification problem has been set up, in order to distinguish between REM

TABLE 8 Performance of binary classification on healthy controls from the CAP Sleep Database

	Accuracy	Sens	Spec	Prec	MSE
K-NN	0,926	0,963	0,864	0,923	0,08
DT	0,949	0,979	0,897	0,941	0,05
SVM	0,863	0,935	0,741	0,859	0,13

TABLE 9 Performance of binary classification on RBD subjects from the CAP Sleep Database

	Accuracy	Sens	Spec	Prec	MSE
K-NN	0,759	0,779	0,740	0,749	0,24
DT	0,765	0,767	0,763	0,764	0,23
SVM	0,643	0,702	0,582	0,627	0,36

and NREM sleep. The wake stage has been discarded, and stages $N1$, $N2$ and $N3$ included in the NREM class. As we are dealing with a binary problem, we have tested K-NN along with two further classifiers suitable for the problem at hand, namely Decision Tree (DT) and SVM. Each sample set was validated through Leave-One-Out cross-validation (LOO-CV), leaving out one observation at a time. The classification has been tested on both healthy and RBD subjects from the CAP Sleep Database. Results are displayed in Tables 8 and 9, in terms of accuracy, sensitivity, specificity, precision and mean squared error (MSE), that is test error. Sensitivity measures the performance on the REM class. It can be appreciated that REM detection efficiency is high in the control group, with DT yielding 97.9% sensitivity and the lowest mean squared error (cf. Table 8). Likewise, the performance on the RBD subjects is quite promising. Indeed, all three models yielded sensitivity in excess of 70%. The performance of K-NN when the novel features are not implemented, are shown in Table 10. A slight yet measurable impairment in all the performance metrics can be appreciated in this case. Finally, feature correlation with target was computed, by means of Pearson correlation coefficient. Setting the target to REM class, all implemented features displayed good correlation. A set of these is displayed in Figure 5.

9 | CONCLUSIONS

This work proposes an automatic sleep stage classification based on a single EEG channel. Three classification algorithms are addressed, namely RF, K-NN and RUSBoost. Novel fea-

TABLE 10 Performance of binary classification on RBD subjects from the CAP Sleep Database w/o implementation of novel features

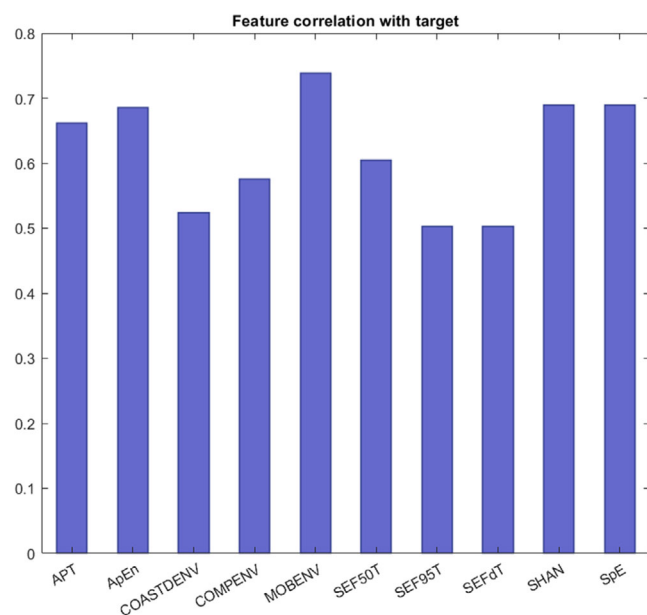|	Accuracy	Sens	Spec	Prec	MSE
K-NN	0,742	0,776	0,708	0,727	0,26

FIGURE 5 A set of novel features implemented and their correlation (Pearson) with REM class

tures are proposed, based on the micro-structure of REM sleep. This represents a novelty, as REM sleep has always been considered a single, homogeneous stage. The achieved results reveal that all three methods achieve very good ability in detecting REM stage, with micro-averaged accuracy of 92.7%, 92.5% and 89.9%, respectively. High sensitivity and specificity – with a satisfactory trade-off between the two – underline good detection and a low number of false positives. This is demonstrated again (in RF and K-NN) by the precision value, which is reasonable for experimental raw data (≈75%). The results outperform those published in [14] using the same dataset and are comparable with [12], where features from EEG, EOG and EMG signals are used. Finally, we have explored the capability of classifying REM versus NREM sleep stages, using data from both healthy controls and patients affected by RBD. The results are quite encouraging, and confirm the usefulness of the proposed features, based on fine REM sleep classification. In conclusion, the proposed methods is able to perform 5-stage sleep classification in healthy controls using only one EEG channel; this a step forward towards the implementation of sleep measures at home, with a simplified sensor configuration. In fact, PSG is stressful for patients due to both the high number of electrodes and the diverse environment; it is costly, being performed in hospital and implying a time-consuming manual annotation. Moreover, the performance on REM stage detection are promising in view of future studies on RBD in PD patients. Future developments will be in the direction of addressing a larger dataset, in order to validate the classification performance (in particular, those of the RUSBoost method, which suffers from data scarcity). Moreover, it is possible that the results are biased by the unbalanced sex ratio of the sample, as it was suggested that sleep is different in adults male and female subjects who do not have neurological disorders [36]; the verification of this point is left to future developments. Finally, we are setting up a protocol to train and validate our algorithm on PD subjects, using PSG signals as a term of comparison.

ORCID
Irene Rechichi https://orcid.org/0000-0002-9420-4922
Gabriella Olmo https://orcid.org/0000-0002-3670-9412

REFERENCES
1. Rechtschaffen, A., Kales, A.: A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. US Department of Health, Education and Welfare, Public Health Service, National Institutes of Health, National Institute of Neurological Diseases and Blindness, Neurological Information Network, (1968)
2. Aserinsky, E., Kleitman, N.: Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. Science 118(3062), 273–4 (1953)
3. Simor, P. et al.: The microstructure of REM sleep: Why phasic and tonic? Sleep Medicine Reviews 52, 101305 (2020)
4. Perogamvros, L.L., Schwartz, S.: The roles of the reward system in sleep and dreaming. Neurosci. Biobehav. Rev. 36(8), 1934–51 (2012)
5. van der Helm, E., Walker, M.P.: Overnight therapy? The role of sleep in emotional brain processing. Psychol. Bull. 135(5), 731–748 (2009)
6. Schenck, C., et al.: Chronic behavioral disorders of human REM sleep: a new category of parasomnia. Sleep 9(2), 239–308 (1986)
7. Boeve, B.: REM sleep behavior disorder: Updated review of the core features, the RBD-neurodegenerative disease association, evolving concepts, controversies, and future directions. Ann. N. Y. Acad. Sci. 1184, 15–54 (2010)
8. Rolinski, M., et al.: REM sleep behaviour disorder is associated with worse quality of life and other non-motor features in early Parkinson's disease. J. Neurol. Neurosur. Ps 85(5), 560–566 (2014)
9. St Louis, E., Boeve, B.: REM sleep behavior disorder: Diagnosis, clinical implications, and future directions. Mayo Clinic Proc. 92(11), 1723–1736 (2017)
10. Schenck, C.H., Boeve, B.F., Mahowald, M.W.: Delayed emergence of a parkinsonian disorder or dementia in 81% of older men initially diagnosed with idiopathic rapid eye movement sleep behavior disorder: A 16-year update on a previously reported series. Sleep Med. 14(8), 744–8 (2013)
11. Yildirim, O., Baloglu, U.B., Acharya, U.R.: A deep learning model for automated sleep stages classification using PSG signals. Int. J. Environ. Res. Public Health 16(4), 599 (2019)
12. Cooray, N., et al.: Detection of REM sleep behaviour disorder by automated polysomnography analysis. Clin. Neurophysiol. 130(4), 505–514 (2019)
13. Yetton, B.D., et al.: Automatic detection of rapid eye movements (REMs): A machine learning approach. J. Neurosci. Meth. 259, 72–82 (2016)
14. Imtiaz, S., Rodriguez-Villegas, E.: A low computational cost algorithm for REM sleep detection using single channel EEG. Ann. Biomed. Eng. 42(11), 2344–2359 (2014)
15. Berthomier, C., et al.: Automatic analysis of single-channel sleep EEG: validation in healthy individuals. Sleep 30(11), 1587–1595 (2007)
16. Basha, A.J. et al.: Support vector machine and simple recurrent network based automatic sleep stage classification of fuzzy kernel. J. Amb. Intel. Hum. Comp. 1–9 (2020)
17. Kim, S.W., et al.: Wearable multi-biosignal analysis integrated interface with direct sleep-stage classification. IEEE Access 8, 46131–40 (2020)
18. Zhang, J., et al.: Orthogonal convolutional neural networks for automatic sleep stage classification based on single-channel EEG. Comput. Meth. Prog. Bio. 183, 105089 (2020)

19. Jadhav, P., et al.: Automatic sleep stage classification using time–frequency images of CWT and transfer learning using convolution neural network. Biocybern. Biomed. Eng. 40(1), 494–504 (2020)

20. Zhu, T., Luo, W., Yu, F.: Convolution-and attention-based neural network for automated sleep stage classification. Int. J. Environ. Res. Pub. 17(11), 4152 (2020)

21. Fernandez-Blanco, E., Rivero, D., Pazos, A.: Convolutional neural networks for sleep stage scoring on a two-channel EEG signal. Soft Comput. 24(6), 4067–4079 (2020)

22. Devuyst, S., Dutoit, T., Kerkhofs, M.: The DREAMS databases and assessment algorithm. https://zenodo.org/record/2650142#.XwiStygzZPZ. Accessed 1 Jan 2005

23. Jorgensen, G., et al.: An australasian commentary on the AASM manual for the scoring of sleep and associated events. Sleep Biol. Rhy. 18, 163–185 (2020)

24. Byun, J., et al.: The first night effect during polysomnography, and patients' estimates of sleep quality. Psychiat. Res. 274, 27–29 (2019)

25. Rahman, M., Bhuiyan, M.I., Hassan, A.: Sleep stage classification using single-channel EOG. Comput. Biol. Med. 102, 211–220 (2018)

26. Moruzzi, G.: Active processes in the brain stem during sleep. Harvey Lect. 58, 233–97 (1963)

27. Simor, P. et al.: EEG spectral power in phasic and tonic REM sleep: different patterns in young adults and children. J. Sleep Res. 25(3) 267–77 (2016)

28. Corsi-Cabrera, M., et al.: Power and coherent oscillations distinguish REM sleep, stage 1 and wakefulness. Int. J. Psychophysiol. 60(1), 59–66 (2006)

29. Motamedi, F., et al.: Signal processing techniques to human sleep EEG signals - a review. Biom. Sig. Proc. 10, 21–33 (2014)

30. Susmáková, K., Krakovská, A.: Discrimination ability of individual measures used in sleep stages classification. Artif. Intell. Med. 44(3), 261–77 (2008)

31. Lajnef, T., et al.: Learning machines and sleeping brains: Automatic sleep stage classification using decision-tree multi-class support vector machines. J. Neurosci. Meth. 250, 94–105 (2015)

32. Seiffert, C., et al.: A hybrid approach to alleviating class imbalance. IEEE Trans. Syst. Man. Cyb. 40(1), 185–197 (2010)

33. Virkkala, J., et al.: Automatic sleep stage classification using two-channel electrooculography. J. Neur. Meth. 166(1), 109–115 (2007)

34. Terzano, M.G., et al.: Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. Sleep Med. 2(6), 537–553 (2001)

35. Goldberger, A., et al.: PhysioBank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. Circulation. 101(23), e215–e220 (2000)

36. Luca, G. et al.: Age and gender variations of sleep in subjects without sleep disorders. Ann. Med. 47(6) (2015)