Enabling Automated REM Sleep Behaviour Disorder Detection

Navin Cooray[1], Fernando Andreotti[1], Christine Lo[2], Mkael Symmonds[3], Michele T.M. Hu[2], and Maarten De Vos[1]

***Abstract*—This study aims to develop automated diagnostic tools to aid in the identification of rapid-eye-movement (REM) sleep behaviour disorder (RBD). Those diagnosed with RBD enact their dreams and therefore provide a unique characteristic of movement during REM sleep. To evaluate the characteristics of the disorder publicly available PSG recordings were used in combination with additional recordings sourced from multiple institutions. This combination of PSG recordings includes 44 healthy controls and 44 RBD subjects. This study aims to develop a robust diagnostic tool to analyse single channel electromyogram (EMG) recordings and manually annotated sleep stages to distinguish healthy controls from RBD subjects. Numerous techniques exist that can objectively quantify movement from EMG signals and are able to identify RBD subjects, which are further validated in this study. Using these established techniques in combination with additional features that the relationship of muscle movement between sleep stages and the general sleep architecture, the performance of RBD identification was shown to improve. This collection of objective EMG features used with a random forest classifier achieves RBD detection with 92% accuracy, 93% sensitivity, and 91% specificity.**

I. Introduction

Rapid eye movement (REM) sleep behaviour disorder (RBD) is a parasomnia whereby patients physically enact their dreams during REM sleep. Those suffering from RBD no longer exhibit muscle atonia, where muscle relax and are inactive, specifically during the REM stage of sleep [1] (as illustrated in Figure 1). As more patients are identified and studied over time, there is evidence suggesting RBD as a predictor for Parkinson's disease (PD) and may precede the disease by many years [2]–[4]. A predictive condition of PD may prove invaluable, not only in developing preventative medicine but to better understand the progression of this neurodegenerative disease.

Since RBDs acceptance as a distinct clinical disorder in 1986 [5], the diagnosis of RBD has evolved over time. The International Classification of Sleep Disorders (ICSD-3) [6] requires polysomnography (PSG) in order to diagnose RBD and specifically requires that the lack of atonia during REM to be observed. The PSG recording process requires equipment, bed availability, and specially trained staff to manually analyse hours worth of data. This expensive and time consuming process provides an increasing demand to develop automated diagnostic tools to better utilise the time of sleep clinicians and potentially improve their inter/intra-rate variability. One of the challenges faces clinicians is the fact that the ICSD-3 does not provide an exact definition for muscle atonia and what objectively constitutes muscle movement in an Electromyogram (EMG) signal. Literature contains numerous studies that have attempted to define and quantify muscle activity during REM in order to differentiate normal and abnormal movement.

Lapierre and Monplaisir were the first to specify a technique to quantify motor events during REM sleep, called the RBD PSG scoring method [7]. This visually manual technique aimed to distinguish EMG signals into tonic and phasic movement. Monplaisir *et al.* [7] was able to use 80 RBD subjects and 80 age and gender matched controls to achieve a classification accuracy of 85.6% (sensitivity and specificity of 88.9% and 82.5%, respectively). Burns *et al.* [8] used EMG variance to develop a metric called the Supra-Threshold REM EMG Activity Metric (STREAM) to identify possible RBD with 100% sensitivity and 71%. specificity. This metric is calculated by measuring the EMG variance during REM epochs and comparing it to a threshold calculated during NREM epochs. The atonia index, developed by Ferri *et al.* [9], is able to provide a score indicating the level of atonia during any given sleep stage. However recordings may be imbued with noise and artefacts that can distort the calculation, which can be mitigated using the corrected atonia index (AI). Frandsen *et al.* [10] were able to incorporate a sliding window and a threshold to quantify motor activity (QMA). This was then used to tune parameters to derive features that depict the number of motor activity events in duration and percentage of REM epochs to distinguish RBD from other subjects. While these established metrics prove to be able to identify RBD subjects, they were far from perfect and provided an opportunity for machine learning techniques to be used.

There is a small number of papers that approach the concept of automated RBD detection by applying machine learning algorithms [11]. Kempfner *et al.* (2013) [11] described an algorithm that measured and detected abnormal muscle activity during REM sleep. This technique required 3 EMG channels and from each channel a single feature was calculated by comparing the mean envelope of a mini-epoch to the minimum envelope of the entire epoch. A one-class support vector machine (OC-SVM) was then used to classify the mini-epochs as an inlier or outlier and this percentage of abnormal (outlier) and normal (inlier) epochs provided a muscle score to identify RBD subjects. This detection algorithm proved to be able to distinguish all Periodic Limb Movement (PLM) subjects (12), iRBD (12) and $9/12$ PD sufferers from healthy controls (12).

[1] University of Oxford, Institute of Biomedical Engineering, Dept. Engineering Sciences, Oxford, UK.

[2] Nuffield Department of Clinical Neurosciences, Oxford Parkinsons Disease Centre (OPDC), University of Oxford, UK

[3] Department of Clinical Neurophysiology, Oxford University Hospitals, John Radcliffe Hospital, University of Oxford, UK

Author to whom all correspondence should be directed is Navin Cooray (navin.cooray@eng.ox.ac.uk).

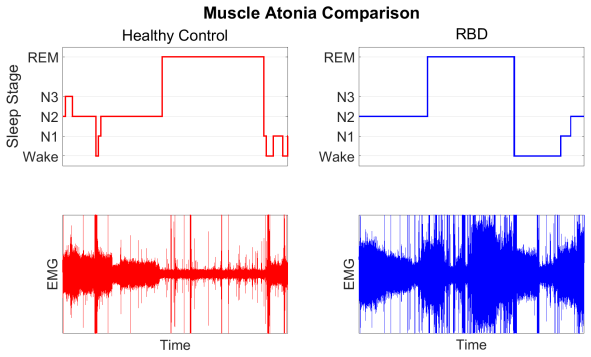

Fig. 1. Typically during REM for a healthy control (red) there is a clear muscle atonia, with a clear reduction in EMG amplitude during REM. A typical RBD subject (blue) that depicts the clear absence of muscle aotnia, with large EMG amplitudes and large spike artefacts.

As per the diagnosis criteria these techniques heavily focus on the the absence of atonia during REM, but RBD subjects don't necessarily display this symptom with every instance of REM and it may not even occur at all during a single night. Literature on RBD also describes other sleep characteristics that differentiate RBD subjects, specifically through sleep architecture [12], [13]. Once more EMG signals are often tainted with noise, such as heart beat artefacts, which can be difficult to remove. This study aims to develop a robust diagnostic tool to analyse Electromyogram (EMG) recordings together with sleep architecture to distinguish healthy controls from RBD subjects using recordings from multiple sources.

II. Data Material

Polysomnography (PSG) recordings were attained from several sources, including the publicly available CAP database and two datasets from the University of Oxford. The Cyclic Alternating Pattern (CAP HC) database includes 13 healthy subjects [14], [15]. Older healthy controls were acquired from the University of Oxford in a sleep and motor plasticity study (SMP) that contained 9 subjects. An additional 21 older healthy controls were attained from the Montreal Archive of SLeep Studies (MASS HC) database [16]. The CAP database also provided 22 RBD subjects (CAP RBD) which was combined with subjects from the John Radcliffe hospital that contained 22 RBD subjects (JR RBD). Because the recordings from the John Radcliffe Hospital were part of routine NHS care they included two nights for each subject, but only the 2^{nd} night was used for this study. These recordings were attained with approval from the Oxford University Hospital NHS Trust and comply with the requirements of the Department of Health Research Governance Framework for Health and Social Care 2005 (United Kingdom). The summary of the collated datasets are described in Table I. All PSG recordings include an EMG of the submentalis muscle and are annotated by sleep experts that detail the sleep stage for every 30 second epoch. Datasets that were annotated using the Rechtschaffen & Kales rules were converted to AASM sleep stages (S3 & S4 were combined and interpreted as N3). As a result these hypnograms included the following stages: wake, REM sleep, N1, N2, and N3.

TABLE I
Combined Datasets

Dataset Name	Age	#Subjects	#Male	#Female
CAP HC	63.6 ± 9.0	13	7	6
SMP HC	31.3 ± 4.2	9	5	4
MASS HC	64.6 ± 4.8	21	12	9
Combined HC	**54.3 ± 16.2**	**43**	**24**	**19**
CAP RBD	63.6 ± 9.0	22	19	3
JR RBD	62.1 ± 6.9	21	19	2
Combined RBD	**70.7 ± 6.2**	**43**	**38**	**5**

III. Methods

A. Signal Preprocessing

EMG signals from the subjects were re-sampled at 256Hz (due to EMG's inherent broad spectrum [17]) and filtered between 10 and 100Hz , using an 8^{th} ordered bandpass filter. Finally a 10^{th} ordered 50Hz notch filter was also used to suppress noise from mains supply. Subjects from the dataset had to be excluded because EMG recordings were unusable, as electrodes has either fallen or simply failed (this included one healthy control from the SMP dataset and one RBD subject from the JR dataset).

B. Feature Extraction

Once the EMG was preprocessed, features for each epoch were calculated, including the atonia index, STREAN, motor activity duration, and motor activity percentage. For every recording the atonia index was calculated for each sleep stage (REM, N3, and N2, providing three features). These were also used to calculate the atonia index ratio between N2/REM and N3/REM, in order to depict the absence of atonia relative to other stages (two features). The motor activity duration and percentage, as detailed by QMA, was calculated for each epoch and averaged across all REM epochs (two features). Additionally the STREAM score for an entire night were calculated by comparing REM variance to NREM variance for every three second mini-epoch (one feature). Features that described the sleep architecture were also calculated and included the percentage of REM, N2, and N3 epochs in total (three features). This provides a total of 11 features, comprising of six established EMG metrics and five features that describe sleep architecture and atonia index relative to sleep stages.

C. Classification

Features detailing EMG movement were combined with simple features that describe sleep architecture could be used to analyse and predict RBD. This study focused on a random forest classifier (RF) to identify RBD. The classifier was trained and evaluated using a leave-one-subject-out methodology. For the purpose of this study two random forest classifiers were evaluated, one using only objective EMG metrics (6 features) while the other combined features detailing sleep architecture and atonia ratios (11 features in total).

1) Random Forest Classifier: A random forest classifier was ideal candidate for this study because it is relatively parameter-free, robust to outliers, relatively fast to train and resistant to over-fitting [18]. Once more this algorithm is able

TABLE II
PERFORMANCE OF RBD DETECTION

Classifier	Accuracy	Sensitivity	Specificity
Atonia Index Threshold	0.80	0.79	0.81
STREAM Threshold	0.78	0.81	0.74
QMA Threshold	0.79	0.65	0.93
RF (EMG Metrics)	0.86	**0.93**	0.79
RF (All Features)	**0.92**	**0.93**	**0.91**

to quantify the importance of individual features, providing a way to rank and assess their value in RBD identification. This study designed the classifier with 500 trees, where each tree was designed using a randomly selected set of mtry features (mtry = $\sqrt{M}$, rounded up), where M is the total number of features used. These features will be used to design the decision nodes of each tree, which classify test subjects as either healthy or RBD. Each tree will then contribute a vote to the final classification. The performance of the classifier was evaluated by using a leave-one-subject-out methodology. Therefore random forests were trained on all subjects excluding one, which was then used to test the trained classifier. This test was done for every subject within the dataset. Feature importance was calculated by averaging the mean decrease in gini-index for every node a feature is used (across all folds of the validation process). The performance of the classifier was assessed using accuracy, sensitivity and specificity.

IV. RESULTS & DISCUSSION

It is clear from Figure 2 the distinguishing characteristics of the EMG metrics calculated from healthy controls and RBD subjects. These numerically capture the absence of atonia during REM by measuring and quantifying muscle movement during the PSG recordings. However, there is also a noticeable overlap between these two cohorts. RBD subjects can be observed to have very good sleep, with very low muscle movement during REM. Healthy control subjects that provide features depicting high muscle activity during REM are due to excessive noise, particularly from very strong heart beat artefacts. Potentially the features describing sleep architecture and atonia ratios between stages can be used to aid in differentiating RBD from healthy controls as shown in Figure 3. For RBD subjects the atonia index during N3 is higher than REM, whereas healthy controls are nearly equal (this is similar when comparing the ratio of N2 and REM). It would also appear that RBD subjects within this study have longer periods of N3 sleep compared to healthy controls.

By including only the atonia index during REM, motor activity duration, motor activity percentage and the STREAM features the RF achieved an accuracy of 86%, sensitivity of 93%, and specificity of 91% using a leave-one-subject-out methodology. This was only marginally better than using individual metrics with a threshold to identify RBD subjects (thresholds for AI, STREAM and QMA were 0.9, 15% and 7%, respectively). By incorporating the atonia index ratios and the sleep architecture features with the EMG metrics, the RF (all features) achieved an accuracy of 92%, sensitivity of 93%, and specificity of 91%, which outperformed all other methods except QMA, with respect to specificity. These results

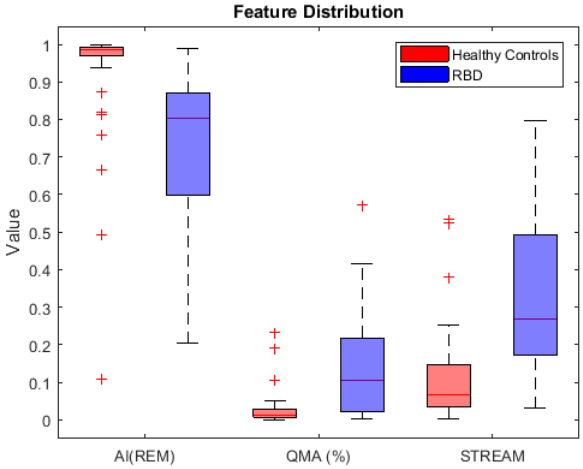

Fig. 2. Feature distribution for the calculated atonia index, quantified motor activity (percentage) and STREAM in healthy controls (red) and RBD subjects (blue). Note the increased median and variation within the RBD cohort when compared to healthy subjects for QMA and STREAM, but the inverse for AI.

are summarised in Table II . The relatively lower specificity score, may not necessarily be a disadvantage, because a highly sensitive classifier would potentially enable the discovery of subjects on the cusp of converting to RBD or at the very least developing issues with sleep.

Individually, the EMG metrics are able to distinguish between healthy and RBD subjects with some success. But by combining these established EMG metrics with a RF classifier the results are improved. Furthermore by training an RF with additional features that detail sleep architecture and the atonia index ratio between stages the performance was greatly improved. These results improve upon individual metrics by combining well established scores together with sleep architecture to provide a more robust differentiator. These features take advantage of the fractured sleep observed

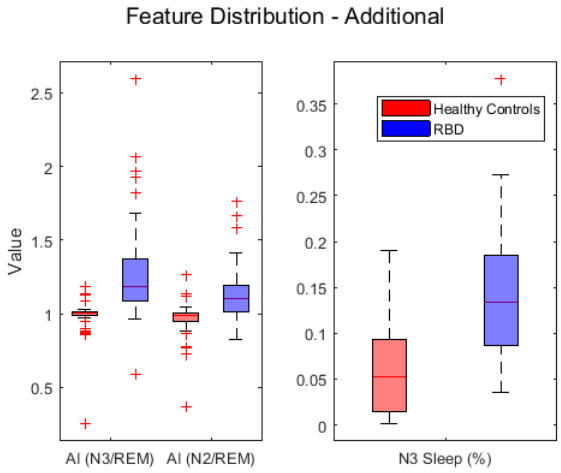

Fig. 3. The sleep architecture of RBD (blue) and healthy controls (red) are also clearly differentiated by the atonia index ratios between sleep stages and the sleep architecture, measured by the percentage of REM and N3 sleep.

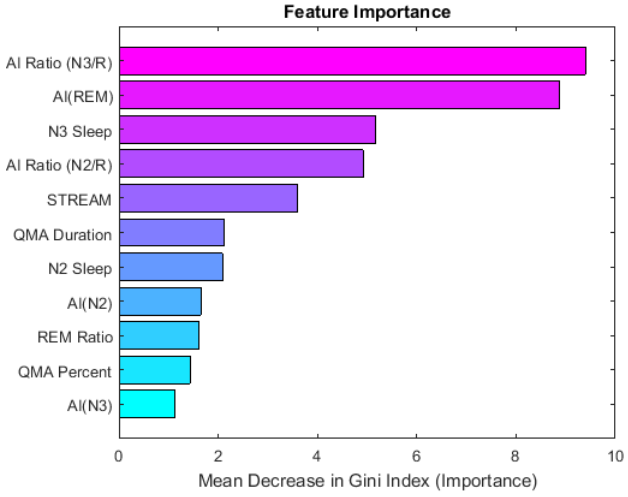

Fig. 4. The mean decrease in gini index from the leave-one-subject-out validation was used to rank features based on importance. The atonia index ratio features proved most important, followed closely by the atonia index during REM and the percent of N3 sleep.

in RBD subjects compared to healthy controls. This is most evident in features that describe N3 and REM as a percentage of total sleep. Also using atonia index ratios allows the RF to explore relationships between sleep stages which compensates for EMG signals tarnished with recurring artefacts and noise.

Healthy controls incorrectly classified as RBD provide features that are indicative of both healthy and RBD subjects. The atonia index ratio between stages and the percent of N3 sleep helped mitigate these issues as indicated by the improved RBD detection performance and the high feature importance ranking shown in Figure 4. Furthermore some RBD subjects have a very good night's sleep, with very low muscle movement during REM, therefore are sometimes misclassified as healthy. But by incorporating sleep features describing the percentage of N3 or N2 sleep, the RF has the opportunity to correctly classify these subjects based on their unusual sleep architecture. Features that depict sleep architecture would also be highly sensitive to other sleeping disorders, which can be defined by fractured and erratic sleep patterns. This study could be further validated by incorporating additional datasets of healthy controls and numerous other sleep disorders other than RBD. A possible extension of this study would be to remove the need for manual sleep annotation by incorporating automatic sleep staging algorithms, especially in the context of take-home recordings that would enable a viable screening tool [19].

V. Conclusion

Using a combination of objective EMG scores and metrics for sleep architecture an SVM can achieve RBD detection with 92% accuracy, 95% sensitivity and 83% specificity. In essence this study has shown a robust and effective method in automatically analysing PSG recordings to identify RBD subjects.

Acknowledgments

NC acknowledges the support of the RCUK Digital Economy Programme grant number EP/G036861/1 (Oxford Centre for Doctoral Training in Healthcare Innovation), Sleep and Circadian Neuroscience Institute (SCNi) and the Rotary Foundation. The content of this article is solely the responsibility of the author and does not necessarily represent the official views of the RCUK, SCNi or the Rotary Foundation.

References

[1] M. H. Kryger, T. Roth, and W. C. Dement, *Principles and Practice of Sleep Medicine*, 5th ed. Elsevier, 2011.

[2] R. B. Postuma and J. Y. Montplaisir, "Parkinson risk in idiopathic REM sleep behavior disorder Preparing for neuroprotective trials," 2015.

[3] E. Tolosa, L. Molinuevo, A. Iranzo, A. Ferna, A. Llado, F. Valldeoriola, E. Gelpi, I. Vilaseca, and R. Sa, "Neurodegenerative Disorder Risk in Idiopathic REM Sleep Behavior Disorder : Study in 174 Patients," vol. 9, no. 2, 2014.

[4] C. H. Schenck, S. R. Bundlie, and M. W. Mahowald, "Delayed emergence of a parkinsonian disorder in 38% of 29 older men initially diagnosed with idiopathic rapid eye movement sleep behaviour disorder." *Neurology*, vol. 46, no. 2, pp. 388–93, feb 1996.

[5] C. H. Schenck, S. R. Bundlie, M. G. Ettinger, and M. W. Mahowald, "Chronic behavioral disorders of human REM sleep: a new category of parasomnia." *Sleep*, vol. 9, no. 2, pp. 293–308, jun 1986.

[6] M. J. Sateia, "International classification of sleep disorders-third edition highlights and modifications," *Chest*, vol. 146, no. 5, pp. 1387–1394, 2014.

[7] O. Lapierre and J. Montplaisir, "Polysomnographic features of REM sleep behavior disorder: development of a scoring method." *Neurology*, vol. 42, no. 7, pp. 1371–4, jul 1992. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/1620348

[8] J. W. Burns, F. B. Consens, R. J. Little, K. J. Angell, S. Gilman, and R. D. Chervin, "EMG variance during polysomnography as an assessment for REM sleep behavior disorder." *Sleep*, 2007.

[9] R. Ferri, C. Franceschini, M. Zucconi, S. Vandi, F. Poli, O. Bruni, C. Cipolli, P. Montagna, and G. Plazzi, "Searching for a marker of REM sleep behavior disorder: submentalis muscle EMG amplitude analysis during sleep in patients with narcolepsy/cataplexy." *Sleep*, vol. 31, no. 10, pp. 1409–1417, 2008.

[10] R. Frandsen, M. Nikolic, and M. Zoetmulder, "Motor activity in REM sleep Analysis of automated quanti fi cation of motor activity in REM sleep behaviour disorder," pp. 583–590, 2015.

[11] J. Kempfner, P. Jennum, M. Nikolic, J. A. E. Christensen, and H. B. D. Sorensen, "SLEEP phenomena as an early biomarker for Parkinsonism," 2013.

[12] C. H. Schenck, T. D. Hurwitz, and M. W. Mahowald, "REM sleep behaviour disorder: an update on a series of 96 patients and a review of the world literature," *Journal of Sleep Research*, vol. 2, no. 4, pp. 224–231, 1993.

[13] J. Massicotte-marquez, J. Carrier, A. De, A. Mathieu, D. Petit, and J. Montplaisir, "Slow-wave Sleep and Delta Power in Rapid Eye Movement Sleep Behavior Disorder," pp. 1–3, 2005.

[14] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, jun 2000.

[15] M. G. Terzano, L. Parrino, A. Sherieri, R. Chervin, S. Chokroverty, C. Guilleminault, M. Hirshkowitz, M. Mahowald, H. Moldofsky, A. Rosa, R. Thomas, and A. Walters, "Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep." *Sleep medicine*, vol. 2, no. 6, pp. 537–53, nov 2001.

[16] C. O'Reilly, N. Gosselin, J. Carrier, and T. Nielsen, "Montreal Archive of Sleep Studies : an open-access resource for instrument benchmarking and exploratory research," *Journal of Sleep Research*, pp. 628–635, 2014.

[17] A. Boxtel, P. Goudswaard, and L. Schomaker, "Amplitude and Bandwidth of the Frontalis Surface EMG: Effects of Electrode Parameters," *Psychophysiology*, vol. 21, no. 6, pp. 699–707, nov 1984.

[18] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.

[19] K. B. Mikkelsen, D. B. Villadsen, M. Otto, and P. Kidmose, "Automatic sleep staging using ear-EEG," *BioMedical Engineering OnLine*, pp. 1–15, 2017.