End-to-end Deep Learning of Polysomnograms for Classification of REM Sleep Behavior Disorder

Brink-Kjær, Andreas; Gunter, Katarina Mary; Mignot, Emmanuel; During, Emmanuel; Jennum, Poul; Sørensen, Helge B.D.

Published in:
Proceedings of 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society

Link to article, DOI:
10.1109/EMBC48229.2022.9871576

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871576

End-to-end Deep Learning of Polysomnograms for Classification of REM Sleep Behavior Disorder

Andreas Brink-Kjaer[+,1,2,3], *Member, IEEE*, Katarina Mary Gunter[1,2,3], Emmanuel Mignot[3], Emmanuel During[4,5,*], Poul Jennum[2,*], and Helge B.D. Sorensen[1,*], *Senior Member, IEEE*

***Abstract*— Rapid eye movement (REM) sleep behavior disorder (RBD) is parasomnia and a prodromal manifestation of Parkinson's disease. The current diagnostic method relies on manual scoring of polysomnograms (PSGs), a procedure that is time and effort intensive, subject to interscorer variability, and requires high level of expertise. Here, we present an automatic and interpretable diagnostic tool for RBD that analyzes PSGs using end-to-end deep neural networks. We optimized hierarchical attention networks in a 5-fold cross validation directly to classify RBD from PSG data recorded in 143 participants with RBD and 147 age- and sex-matched controls. An ensemble model using logistic regression was implemented to fuse decisions from networks trained in various signal combinations. We interpreted the networks using gradient SHAP that attribute relevance of input signals to model decisions. The ensemble model achieved a sensitivity of 91.4 % and a specificity of 86.3 %. Interpretation showed that electroencephalography (EEG) and leg electromyography (EMG) exhibited most patterns with high relevance. This study validates a robust diagnostic tool for RBD and proposes an interpretable and fully automatic framework for end-to-end modeling of other sleep disorders from PSG data.**

***Clinical relevance*— This study presents a novel diagnostic tool for RBD that considers neurophysiologic biomarkers in multiple modalities.**

I. INTRODUCTION

Rapid eye movement (REM) sleep behavior disorder (RBD) is a parasomnia where abnormal movements, vocalizations, and dream enactment occurs during REM sleep [1]. These abnormal behaviors are secondary to a lack of paralysis of skeletal muscles during REM sleep, a phenomenon called REM sleep without atonia (RSWA). Idiopathic RBD (iRBD) is often prodromal to α-synucleinopathies such as Parkinson's disease (PD), dementia with Lewy bodies (DLB), and multiple system atrophy (MSA) [2].

Current clinical guidelines [1], as described by the American Academy of Sleep Medicine (AASM), rely on video-polysomnography (v-PSG) to demonstrate RSWA and document vocalizations or complex motor behaviors during REM sleep. Quantifying RSWA is accomplished by manual scoring [3] although automatic methods using electromyography (EMG) signals have been developed [4].

One limitation of these guidelines is that they do not consider other PSG biomarkers, such as slowing of the electroencephalography (EEG) in wake, REM, and during arousal [5], [6], abnormal electrooculography (EOG) [7], and abnormal heart responses in the electrocardiography (ECG) [8]. Recently, machine learning driven methods that integrates multiple biomarkers of RBD have been proposed [9]–[11]. However, these only rely on *hand-crafted* features, which may not all be optimal for classification of RBD.

We hypothesize that using information from the entire PSG would discriminate better between controls and patients with RBD than using preselected features. To test this hypothesis, we used a deep neural network we recently developed for end-to-end modeling of the full PSG [12] to optimize diagnosis of RBD. To explain features derived from the model, we experimented with selecting individual channels of the PSG and use relevance attribution to find relevant patterns.

II. METHODS

A. Data Description

We included PSG data from 82 iRBD patients, 61 patients with PD and RBD (PD+RBD), and 147 sleep clinic controls with various sleep complaints. These came from two sleep centers, the Danish Center for Sleep Medicine (DCSM) and the Stanford Sleep Medicine Center (STNF). Patients from the DCSM were recruited between 2009 and 2015 and evaluated using v-PSG that was scored according to the AASM guidelines [1]. These were asked to discontinue medication that affect sleep (antidepressants, antipsychotics, hypnotics), except for dopaminergic medications, two weeks prior to evaluation. STNF patients were recruited from between 2016 and 2021 and were evaluated similarly to participants from DCSM, although RSWA was established using the automatic Sleep Innsbruck Barcelona (SINBAR) criteria [13]. Participants from the STNF were not required to hold their sleep treatments prior to PSG examination. Demographics, apnea-hypopnea index (AHI), diagnosis of periodic leg movement disorder (PLMD), and groupings of participants are shown in Table I. Notably, from the DCSM, the patients from DCSM have a larger proportion of PD+RBD and the controls are characterized by a higher AHI and more PLMD.

[+]Corresponding author available at andbri@dtu.dk.
[*]Shared last authorship.
[1]Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark.
[2]Danish Center for Sleep Medicine, Glostrup University Hospital, Glostrup, Denmark.
[3]Center for Sleep Sciences and Medicine, Stanford University, CA, USA.
[4]Department of Psychiatry and Behavioral Sciences, Stanford University, CA, USA.
[5]Department of Neurology and Neurological Sciences, Stanford University, CA, USA.

TABLE I

SUMMARY OF DEMOGRAPHICS AND CLINICAL CHARACTERISTICS OF PATIENTS.

	STNF (n = 154)	DCSM (n = 136)
Age	67.58 ± 6.27	65.59 ± 7.90
Sex (male)	102 (66.23 %)	88 (64.71 %)
AHI	18.01 ± 17.49	8.85 ± 13.76
PLMD	11 (7.14 %)	35 (25.74 %)
iRBD	52 (33.77 %)	30 (22.06 %)
PD+RBD	10 (6.49 %)	51 (37.50 %)

B. Preprocessing of Polysomnographic Signals

Signals were preprocessed to streamline data for further analysis. To do so, we implemented a previously used pipeline for end-to-end modeling of PSG data [12], which includes: i) channel selection [EEG (C_3-A_2 and C_4-A_1), EOG (left and right), EMG (chin and the difference between left and right leg), and ECG (lead II)]; ii) signal resampling to 128 Hz; iii) bandpass filter to AASM recommendations [3]; and iv) scaling signal amplitudes to their 5th and 95th percentiles.

C. End-to-end Deep Learning from Polysomnograms

A hierarchical attention network previously used for age estimation was implemented [12]. This network was optimized to predict clinical variables based on a whole night of PSG data. First, we trained the network to classify RBD based on 5-minute epochs of data, and secondly, on whole nights using a latent space of processed 5-minute epochs of data. To increase complexity of the network, we substituted the gated recurrent units (GRUs) for long short-term memory (LSTM) layers and added its output, averaged across time-instances, to the latent space. Moreover, one additional LSTM layer was used in phase (2) of the network. The overall structure of the network is shown in Fig. 1.

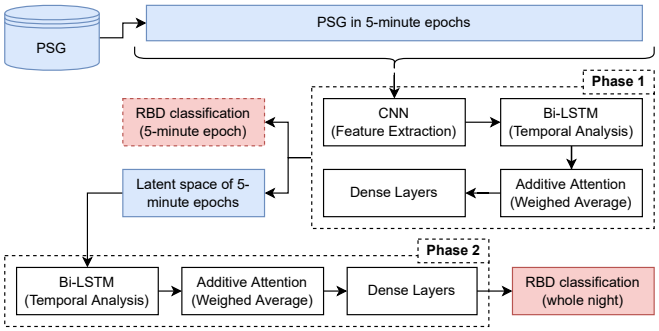

Fig. 1. High-level overview of the network architecture used in this study. The PSG input is a combination of EEG, EOG, EMG, and ECG. The network can be optimized on 5-minute epoch (Phase 1) of data and whole nights of data (Phase 2). Implementation details are available in the original publication [12].

The network was optimized and tested using 5-fold cross validation with 70 % for training, 10 % for validation, and 20 % for testing. It was optimized first on 5-minute epochs of data and second on whole nights of data. Network hyperparameters were selected based on experiments for age estimation [12]. In both phases we used a cross entropy loss, Adam optimization, a learning rate of 10^{-3}, a factor of 10^{-5} for L2 regularization, and early stopping. In the 5-minute epoch optimization phase, we used a batch size of 32, a patience of 3 for early stopping, and a dropout rate of 0.1 and 0.75 for each input signal and at the last dense layer, respectively. For whole night optimization, we used a batch size of 64, a patience of 10 for early stopping, and a dropout of 0.2 at the last dense layer.

We also experimented with various combinations of PSG signals as input, specifically, each modality separate, EMG+ECG, EEG+EOG+EMG, and all modalities together. Moreover, a logistic regression ensemble model was fitted based on probability output of the models with various input signals. It was optimized with Broyden-Fletcher-Goldfarb-Shanno quasi-Newton algorithm [14] as a two-loop leave-one-out cross validation, with the inner loop optimized a factor for L2 regularization in a range $(10^{-6}, 10^{-0.5})$ while the outer loop was used to calculate performance.

The effect of age, sex, diagnoses, and sleep clinic was investigated using traditional logistic regression with accuracy of each prediction as a dependent variable. The odds ratio (OR) of each variable was used to interpret associations.

Saliency of PSG inputs were generated using gradient SHAP [15], [16] on the EEG+EOG+EMG network after the 5-minute epoch optimization. To remove noise, sample relevance scores were filtered by a gaussian window with a length of 10 seconds and standard deviation of 0.234 seconds. The distribution of absolute values of the relevance scores were computed across input signals and sleep stages manually scored according to the AASM [3].

III. RESULTS

Median fit across cross validation folds for the logistic regression ensemble model used a factor of 0.0108 for L2 regularization and was

$$\begin{aligned}\text{logit}\left(P(RBD)\right) = &-3.41 \\ &+ 3.36 \cdot P(RBD)_{EEG} \\ &+ 2.3 \cdot P(RBD)_{EEG+EOG+EMG} \\ &+ 1.02 \cdot P(RBD)_{ECG}. \quad (1)\end{aligned}$$

The ensemble included the EEG, EEG+EOG+EMG, and ECG models in all folds and no other models were included in 89.31 % of folds. The performance of each model is shown in Table II.

Associations between demographics and clinical variables to performance is shown in Table III.

An example of relevance attribution scores for EEG+EOG+EMG model is shown in Fig. 2. The distribution of the absolute value of relevance scores across signals and manually scored sleep stages is shown in Table IV.

TABLE II

PERFORMANCE OF END-TO-END DEEP LEARNING MODELS OF RBD USING VARIOUS CONFIGURATIONS OF POLYSOMNOGRAPHIC SIGNALS.

		Accuracy	Sensitivity	Specificity
EEG	Overall	84.5	86.8	82.0
	STNF	92.2	94.3	90.5
	DCSM	75.7	80.2	69.1
EOG	Overall	77.9	81.5	74.1
	STNF	96.1	98.6	94.0
	DCSM	57.4	66.7	43.6
EMG	Overall	79.7	91.4	66.9
	STNF	94.8	94.3	95.2
	DCSM	62.5	88.9	23.6
ECG	Overall	72.1	76.2	67.6
	STNF	83.8	77.1	89.3
	DCSM	58.8	75.3	34.5
EMG + ECG	Overall	72.4	78.8	65.5
	STNF	87.7	94.3	82.1
	DCSM	55.1	65.4	40.0
EEG + EOG + EMG	Overall	85.2	86.8	83.5
	STNF	94.2	90.0	97.6
	DCSM	75.0	84.0	61.8
EEG + EOG + EMG + ECG	Overall	76.9	76.8	77.0
	STNF	94.8	92.9	96.4
	DCSM	56.6	63.0	47.3
Ensemble - LR	Overall	89.0	91.4	86.3
	STNF	97.4	97.1	97.6
	DCSM	79.4	86.4	69.1

IV. DISCUSSION

Our ensemble model achieved an accuracy of 97.4 % and 79.4 % in classifying RBD in the STNF and DCSM cohort, respectively. The discrepancy could be a result of different criteria used for RSWA, AASM [3] for DCSM and SINBAR for STNF [13], as the SINBAR criteria is more stringent. Some of this discrepancy is also explained by an increased fraction of controls with PLMD in the DCSM.

Besides our ensemble model, the EEG model performed best with an overall accuracy of 84.5 %. Interestingly, the model that included all signals (EEG+EOG+EMG+ECG) did not perform as well as other models due to overfitting.

TABLE IV

MEDIAN OF ABSOLUTE RELEVANCE SCORES OF SIGNALS IN EEG+EOG+EMG MODEL ACROSS MANUALLY SCORED SLEEP STAGES.

	W	N1	N2	N3	REM	Overall
C_3-A_2	2.57	1.84	2.22	2.59	2.47	2.40
C_4-A_1	3.26	2.44	2.81	3.54	3.23	3.09
EOG_L	2.33	1.37	1.35	1.45	1.96	1.63
EOG_R	2.05	1.24	1.26	1.44	1.82	1.67
Leg EMG	2.58	2.09	1.82	1.49	2.70	2.28
Chin EMG	1.04	0.62	0.59	0.54	1.14	0.73
Overall	2.21	1.52	1.56	1.56	2.14	1.84

Although our ensemble model did not perform significantly better for PD+RBD versus iRBD patients alone, controls with PLMD had significantly worse predictions (OR = 0.28, p = 0.016). Interestingly, the ECG model had the lowest OR for PLMD (OR = 0.20, p = 0.00012) and highest for PD+RBD (OR = 5.13, p = 0.0004) relative to the iRBD group, i.e., the model's accuracy was much higher for PD+RBD and controls without PLMD. It is likely that PD+RBD are differentiated better as they exhibit lesser autonomic response to leg movements and cortical arousal [8]. None of the models performed worse in participants with higher AHI.

In a previous study, Cooray et al. classified RBD with 92 % accuracy using a random forest with hand-crafted features derived from automatic sleep stage scoring [10]. These results were achieved in a different dataset, which makes them difficult to compare. There is a lack of direct comparisons for fully automatic methods that consider several modalities, however, this is out of the scope of this study.

Relevance attribution with gradient SHAP showed that the EEG+EOG+EMG model relies on all modalities but mostly EEG and leg EMG. The attribution scores for EMG were highest in REM sleep but was still important in NREM sleep, which agrees with previous studies that found abnormal

TABLE III

MULTINOMIAL LOGISTIC REGRESSION REGRESSION ANALYSIS OF ACCURACY SHOWING ASSOCIATIONS BETWEEN DEMOGRAPHICS, DIAGNOSES, AND PERFORMANCE. THE ORS FOR AGE AND AHI ARE LISTED FOR AN INCREASE OF 10. NOTE THAT THE CONTROL AND PD+RBD VARIABLE IS IN RELATION TO THE IRBD GROUP.

	Age		Sex		AHI		PLMD		Control		PD+RBD		DCSM	
	OR	p-val	OR	p-val	OR	p-val	OR	p-val	OR	p-val	OR	p-val	OR	p-val
EEG	0.71	0.16	0.78	0.51	1.02	0.87	0.47	0.098	0.60	0.29	0.97	0.95	**0.26**	**0.0022**
EOG	1.28	0.31	1.07	0.84	0.98	0.88	0.30	0.01	0.41	0.078	0.45	0.093	**0.06**	**6.6e-08**
EMG	1.28	0.33	1.26	0.55	1.16	0.35	**0.29**	**0.0081**	**0.39**	**0.048**	2.35	0.11	**0.07**	**2.9e-08**
ECG	1.04	0.86	0.80	0.51	0.85	0.081	**0.20**	**0.00012**	1.65	0.18	**5.13**	**0.0004**	**0.20**	**6.7e-06**
EMG + ECG	0.73	0.15	1.05	0.88	1.08	0.5	0.89	0.78	0.57	0.15	**3.77**	**0.0029**	**0.10**	**1.5e-09**
EEG + EOG + EMG	0.66	0.093	0.90	0.78	0.96	0.73	0.73	0.54	1.21	0.68	**4.66**	**0.0046**	**0.12**	**5.6e-06**
EEG + EOG + EMG + ECG	0.83	0.4	0.94	0.86	1.02	0.84	0.45	0.077	1.26	0.6	2.07	0.093	**0.07**	**2.3e-09**
Ensemble - LR	1.09	0.74	0.54	0.17	0.95	0.71	**0.28**	**0.016**	1.44	0.52	2.31	0.13	**0.09**	**6.6e-05**

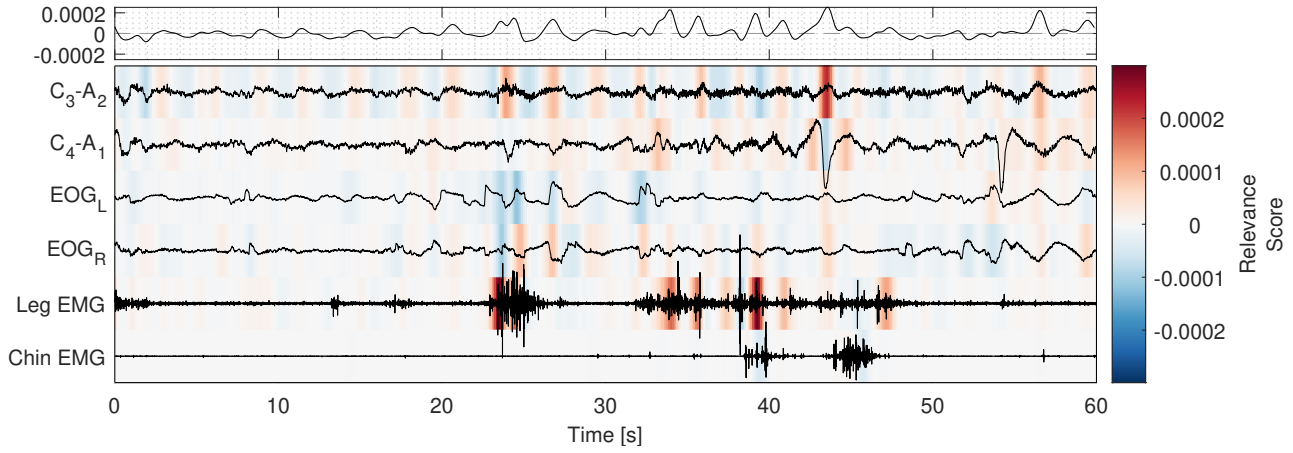

Fig. 2. Example of interpretation of EEG+EOG+EMG model through relevance attribution of samples. Relevance attribution was computed using gradient SHAP. Red and blue indicates positive and negative attribution to the RBD classification, respectively. The top plot shows relevance scores averaged across channels.

motor patterns in NREM sleep [6], [17]. Interestingly, EEG alone seems to be a very strong predictor of RBD without any information about RSWA.

In future research, we could include other EEG derivations, which have been shown to better capture EEG slowing [5]. Moreover, including EMG from the upper extremities could potentially better capture RSWA [1]. Including additional PSG could also help the EEG+EOG+EMG+ECG model generalize better.

V. CONCLUSIONS

We optimized a deep neural network directly from EEG, EOG, EMG, and ECG to distinguish RBD from sleep clinic controls with a variety of other sleep disorders and achieved an accuracy of 89.0 %. This proves the feasibility in end-to-end modeling of sleep disorders from PSG signals without hand-crafting features or using automatic sleep staging, thereby, we avoid restricting modeling to clinical guidelines. We further interpreted the relevance of each signal and found the highest attribution to patterns in the EEG and leg EMG.

REFERENCES

[1] American Academy of Sleep Medicine., *International classification of sleep disorders*, 3rd ed. Darien, IL: American Academy of Sleep Medicine, 2014, p. 383, ISBN: 9780991543410.

[2] B. Högl, A. Stefani, and A. Videnovic, "Idiopathic REM sleep behaviour disorder and neurodegeneration — an update," *Nature Reviews Neurology*, vol. 14, pp. 40–55, Jan. 2018, ISSN: 1759-4758. DOI: 10.1038/nrneurol.2017.157.

[3] R. Berry *et al.*, *The AASM Manual for the scoring of sleep and associated events : rules, terminology and technical specifications*, Version 2. Darien, IL: American Academy of Sleep Medicine: American Academy of Sleep Medicine, 2018.

[4] M. Cesari *et al.*, "Comparison of computerized methods for rapid eye movement sleep without atonia detection," *Sleep*, vol. 41, Oct. 2018, ISSN: 0161-8105. DOI: 10.1093/sleep/zsy133.

[5] M. Livia Fantini *et al.*, "Slowing of electroencephalogram in rapid eye movement sleep behavior disorder," *Annals of Neurology*, vol. 53, pp. 774–780, Jun. 2003, ISSN: 03645134. DOI: 10.1002/ana.10547.

[6] A. Brink-Kjær *et al.*, "Arousal characteristics in patients with Parkinson's disease and isolated rapid eye movement sleep behavior disorder," *Sleep*, vol. 44, Dec. 2021, ISSN: 0161-8105. DOI: 10.1093/sleep/zsab167.

[7] J. A. E. Christensen *et al.*, "Nocturnal eye movements in patients with idiopathic rapid eye movement sleep behaviour disorder and patients with Parkinson's disease," *Journal of Sleep Research*, e13125, Aug. 2020, ISSN: 0962-1105. DOI: 10.1111/jsr.13125.

[8] G. L. Sorensen, J. Kempfner, M. Zoetmulder, H. B. D. Sorensen, and P. Jennum, "Attenuated heart rate response in REM sleep behavior disorder and Parkinson's disease," *Movement Disorders*, vol. 27, pp. 888–894, Jun. 2012, ISSN: 08853185. DOI: 10.1002/mds.25012.

[9] M. Cesari *et al.*, "A data-driven system to identify REM sleep behavior disorder and to predict its progression from the prodromal stage in Parkinson's disease," *Sleep Medicine*, Apr. 2020, ISSN: 13899457. DOI: 10.1016/j.sleep.2020.04.010.

[10] N. Cooray, F. Andreotti, C. Lo, M. Symmonds, M. T. Hu, and M. De Vos, "Detection of REM sleep behaviour disorder by automated polysomnography analysis," *Clinical Neurophysiology*, vol. 130, pp. 505–514, Apr. 2019, ISSN: 13882457. DOI: 10.1016/j.clinph.2019.01.011.

[11] M. Cesari *et al.*, "Video-polysomnography procedures for diagnosis of rapid eye movement sleep behavior disorder (RBD) and the identification of its prodromal stages: guidelines from the International RBD Study Group," *Sleep*, Oct. 2021, ISSN: 0161-8105. DOI: 10.1093/sleep/zsab257.

[12] A. Brink-Kjaer, E. Mignot, H. B. Sorensen, and P. Jennum, "Predicting Age with Deep Neural Networks from Polysomnograms," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, vol. 2020-July, IEEE, Jul. 2020, pp. 146–149, ISBN: 978-1-7281-1990-8. DOI: 10.1109/EMBC44109.2020.9176254.

[13] B. Frauscher *et al.*, "Normative EMG Values during REM Sleep for the Diagnosis of REM Sleep Behavior Disorder," *Sleep*, vol. 35, pp. 835–847, Jun. 2012, ISSN: 0161-8105. DOI: 10.5665/sleep.1886.

[14] J. Nocedal and S. J. Wright, *Numerical Optimization* (Springer Series in Operations Research and Financial Engineering), 2nd ed. Springer New York, 2006, ISBN: 978-0-387-30303-1. DOI: 10.1007/978-0-387-40065-5.

[15] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Online]. Available: https://github.com/slundberg/shap.

[16] N. Kokhlikyan *et al.*, "Captum: A unified and generic model interpretability library for PyTorch," *arXiv*, Sep. 2020. arXiv: 2009.07896.

[17] U. Hanif, L. Trap, P. Jennum, M. Zoetmulder, and H. B. D. Sorensen, "Quantification of muscle activity during sleep for patients with neurodegenerative diseases," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, vol. 2015-Novem, IEEE, Aug. 2015, pp. 6002–6005, ISBN: 978-1-4244-9271-8. DOI: 10.1109/EMBC.2015.7319759.