Deep learning using EEG spectrograms for prognosis in idiopathic rapid eye movement behavior disorder (RBD)

Giulio Ruffini[a,b,*], David Ibañez[a], Eleni Kroupi[a], Jean-François Gagnon[c,e], Marta Castellano[a], Aureli Soria-Frisch[a]

[a]*Starlab Barcelona, Avda. Tibidabo, 47bis, 08035 Barcelona, Spain*
[b]*Neuroelectrics Corporation, 210 Broadway, Cambridge, MA 02139, USA*
[c]*Centre for Advanced Research in Sleep Medicine, Hôpital du Sacré-Cœur de Montréal, Montreal, Canada*
[d]*Department of Neurology, Montreal General Hospital, Montreal, Canada*
[e]*Department of Psychology, Université du Québec à Montréal, Montreal, Canada*
[f]*Department of Psychiatry, Université de Montréal*

Abstract

REM Behavior Disorder (RBD) is a serious risk factor for neurodegenerative diseases such as Parkinson's disease (PD). In this paper we describe deep learning methods for RBD prognosis classification from electroencephalography (EEG). We work using a few minutes of eyes-closed resting state EEG data collected from idiopathic RBD patients (121) and healthy controls (HC, 91). At follow-up after the EEG acquisition (mean of 4 $\pm$2 years), a subset of the RBD patients eventually developed either PD (19) or Dementia with Lewy bodies (DLB, 12), while the rest remained idiopathic RBD. We describe first a deep convolutional neural network (DCNN) trained with stacked multi-channel spectrograms, treating the data as in audio or image problems where deep classifiers have proven highly successful exploiting compositional and translationally invariant features in the data. Using a multi-layer architecture combining filtering and pooling, the performance of a small DCNN network typically reaches 80% classification accuracy. In particular, in the HC vs PD-outcome problem using a single channel, we can obtain an area under the curve (AUC) of 87%. The trained classifier can also be used to generate synthetic spectrograms to study what aspects of the spectrogram are relevant to classification, highlighting the presence of theta bursts and a decrease of power in the alpha band in future PD or DLB patients. For comparison, we study a deep recurrent neural network using stacked long-short term memory network (LSTM) cells [1, 2] or gated-recurrent unit (GRU) cells [3], with similar results. We conclude that, despite the limitations in scope of this first study, deep classifiers may provide a key technology to analyze the EEG dynamics from relatively small datasets and deliver new biomarkers.

Keywords: Biomarkers, EEG, spectrogram, time-frequency, DCNN, RNN, ConvNET, LSTM, GRU, time-frequency, PD, LBD

---

*Corresponding author, giulio.ruffini@neuroelectrics.com, +1 617-899 6911

1. Introduction

REM Behavior Disorder (RBD) is a serious risk factor for neurodegenerative diseases such as Parkinson's disease (PD). RBD is a parasomnia characterized by vivid dreaming and dream-enacting behaviors associated with REM sleep without muscle atonia [4]. Idiopathic RBD occurs in the absence of any neurological disease or other identified cause, is male-predominant and its clinical course is generally chronic progressive [5]. Several longitudinal studies conducted in sleep centers have shown that most patients diagnosed with the idiopathic form of RBD will eventually be diagnosed with a neurological disorder such as Parkinson disease (PD) or dementia with Lewy bodies (DLB) [6, 5, 4]. In essence, idiopathic RBD has been suggested as a prodromal factor of synucleinopathies (PD, DLB and less frequently multiple system atrophy (MSA) [4]).

RBD has an estimated prevalence of 15-60% in PD and has been proposed to define a subtype of PD with relatively poor prognosis, reflecting a brainstem-dominant route of pathology progression (see [7] and references therein) with a higher risk for dementia or hallucinations. PD with RBD is characterized by more profound and extensive pathology—not limited to the brainstem—,with higher synuclein deposition in both cortical and sub-cortical regions.

Electroencephalographic (EEG) and magnetoencephalographic (MEG) signals contain rich information associated with the computational processes in the brain. To a large extent, progress in their analysis has been driven by the study of spectral features in electrode space, which has indeed proven useful to study the human brain in both health and disease. For example, the "slowing down" of EEG is known to characterize neurodegenerative diseases [8, 9]. However, neuronal activity exhibits non-linear dynamics and non-stationarity across temporal scales that cannot be studied well using classical approaches. The field needs novel tools capable of capturing the rich spatio-temporal hierarchical structures hidden in these signals. Deep learning algorithms are designed for the task of exploiting compositional structure in data [10]. In past work, for example, we have used deep feed forward autoencoders for the analysis of EEG data to address the issue of feature selection, with promising results [11]. Interestingly, deep learning techniques, in particular, and artificial neural networks in general are themselves bio-inspired in the brain—the same biological system generating the electric signals we aim to decode. This suggests they should be well suited for the task.

Deep recurrent neural networks (RNNs), are known to be potentially Turing complete [12], although, general RNN architectures are notoriously difficult to train [2]. In other work we studied a particular class of RNNs called Echo State Networks (ESNs) that combine the power of RNNs for classification of temporal patterns and ease of training [13]. The main idea behind ESNs and other "reservoir computation" approaches is to use semi-randomly connected, large, fixed recurrent neural networks where each node/neuron in the reservoir is activated in a non-linear fashion. The interior nodes with random weights constitute what is called the "dynamic reservoir" of the network. The dynamics of the reservoir provides a feature representation map of the input signals into a much larger dimensional space (in a sense much like a kernel method). Using such an ESN, we obtained an accuracy of 85% in a binary, class-balanced classification problem (healthy controls versus PD patients) using an earlier, smaller dataset [13]. The main limitations of this approach, in our view, were the computational cost of developing the dynamics of large random networks and the associated need for feature selection (e.g., which subset of frequency bands and channels to use as inputs to simplify the computational burden).

In [14], we explored algorithmic complexity metrics of EEG spectrograms, in a further attempt at deriving information from the dynamics of EEG signals in RBD patients, with good results, indicating that such metrics may be useful per se for classification or scoring. However, the estimation of such interesting but ad-hoc metrics may in general lack the power of machine learning methods where the relevant features are found directly by the algorithms.

1.1. Deep learning from spectrogram representation

Here we explore first a deep learning approach inspired by recent successes in image classification using deep convolutional neural networks (DCNNs), designed to exploit invariances and capture compositional features in the data (see e.g., [2, 10, 12]). These systems have been largely developed to deal with image data, i.e., 2D arrays, possibly from different channels, or audio data (as in [15]), and, more recently, with EEG data as well [16, 17]. Thus, inputs to such networks are data cubes (image per channel). In the same vein, we aimed to work here with the spectrograms of EEG channel data, i.e., 2D time-frequency maps. Such representations represent temporal dynamics as essentially images with the equivalent of image depth provided by multiple available EEG channels (or, e.g., current source density maps or cortically mapped quantities from different spatial locations). Using such representation, we avoid the need to select frequency bands or channels in the process of features selection. This approach essentially treats EEG channel data as an audio file, and our approach mimics similar uses of deep networks in that domain.

RNNs can also be used to classify images, e.g., using image pixel rows as time series. This is particularly appropriate in our case, given the good performance we obtained using ESNs on temporal spectral data. We study here also the use of stacked architectures of long-short term memory network (LSTM) or gated-recurrent unit (GRU) cells, which have shown good representational power and can be trained using backpropagation [1, 3].

Our general assumption is that some relevant aspects in EEG data from our datasets are contained in compositional features embedded in the time-frequency representation. This assumption is not unique to our particular classification domain, but should hold of EEG in general. In particular, we expect that deep networks may be able to efficiently learn to identify features in the time-frequency domain associated to bursting events across frequency bands that may help separate classes, as in "bump analysis" [18]. Bursting events are hypothesized to be representative of transient synchronies of neural populations, which are known to be affected in neurodegenerative diseases such as Parkinson's or Alzheimer's disease [19].

Finally, we note that we have made no attempt to optimize these architectures. In particular, no fine-tuning of hyper parameters has been carried out using a validation set approach, something we leave for further work. Our aim has been to implement a proof of concept of the idea that deep learning approaches can provide value for the analysis of time-frequency representations of EEG data.

2. EEG dataset

Idiopathic RBD patients (called henceforth RBD for data analysis labeling) and healthy controls were recruited at the Center for Advanced Research in Sleep Medicine of the Hôpital du Sacrè-Cœur de Montral. All patients with full EEG montage for resting-state EEG recording at baseline and with at least one follow-up examination after the baseline visit were included in the study. The first valid EEG for each patient enrolled in the study was considered baseline. Participants also underwent a complete neurological examination by a neurologist specialized in movement disorders and a cognitive assessment by a neuropsychologist. No controls reported abnormal motor activity during sleep or showed cognitive impairment on neuropsychological testing. The protocol was approved by the hospital's ethics committee, and all participants gave their written informed consent to participate. For more details on the protocol and on the patient population statistics (age and gender distribution, follow up time, etc.), see [9]. The cohort used here is the same as the one described in [14].

As in previous work [9, 13, 14], the first level data in this study consisted of resting-state EEG collected from awake patients using 14 scalp electrodes. The recording protocol consisted of conditions with periods of "eyes open" of variable duration (~ 2.5 minutes) followed by periods of "eyes closed" in which patients were not asked to perform any particular task. EEG signals were digitized with 16 bit resolution at a sampling rate of 256 S/s. The amplification device implemented hardware band pass filtering between 0.3 and 100 Hz and notch filtering at 60 Hz to minimize the influence of power line noise. All recordings were referenced to linked ears. The dataset includes a total of 121 patients diagnosed of REM (random eye movement sleep) Behavioral Disorder (RBD) (of which 118 passed the quality tests) and 85 healthy controls (of which only 74 provided sufficient quality data) without sleep complaints in which RBD was excluded[1]. EEG data was collected in every patient at baseline, i.e., when they were still RBD. After 1-10 years of clinical follow-up, 14 patients developed Parkinson disease (PD), 13 Lewy body dementia (DLB) and the remaining 91 remained idiopathic. Our classification efforts here focuses on the previously studied HC *vs.* PD dual problem.

[1] We note, however, that several HC were not followed up.

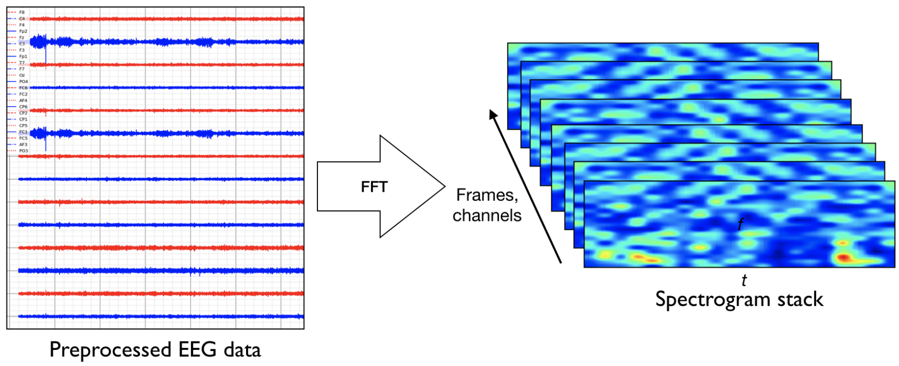

Figure 1: Generation of spectrogram stack for a subject from preprocessed (artifact rejection, referencing, detrending) EEG data.

2.1. Preprocessing and generation of spectrograms

To create the spectrogram dataset (the frames) was processed using Fourier analysis (FFT) after detrending data blocks of 1 second with a Hann window (FFT resolution is 2 Hz) (see Figure 1). To create the spectrogram frames, 20 second 14 channel artifact-free sequences were collected for each subject, using a sliding window of 1 second. FFT amplitude bins in the band 4-44 Hz were used. The data frames are thus multidimensional arrays of the form [channels (14)] x [FFTbins (21)] x [Epochs (20)]. The number of frames per subject was set to 148, representing about 2.5 minutes of data. We selected a minimal suitable number of frames per subject so that each subject provided the same number of frames. For training, datasets were balanced by random replication of subjects in the class with fewer subjects. For testing, we used a leave-pair-out strategy, with one subject from each class. Thus, both the training and test sets were balanced. Finally, the data was centered and normalized to unit variance for each frequency and channel.

3. Convolutional network architecture

The network (which we call SpectNet), implemented in *Tensorflow* (Python), is a relatively simple four hidden-layer convolutional net with pooling (see Figure 2). Dropout has been used as the only regularization. All EEG channels may be used in the input cube.

Our design philosophy has been to enable the network to find local features first and create larger views of data with some temporal (but not frequency) shift invariance via max-pooling.

The network has been trained using a cross-entropy loss function to classify frames (not subjects). It has been evaluated both on frames and, more impor-

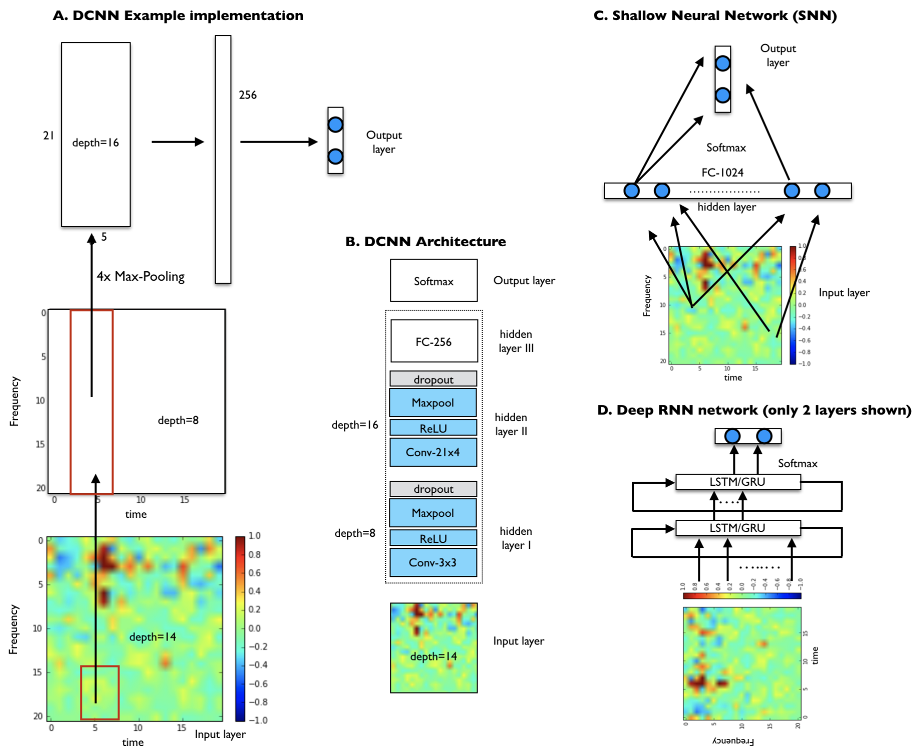

Figure 2: A. DCNN sample model displaying input, convolution with pooling layers, and some hidden-unit layers. The output, here for the binary classification problem using one-hot encoding, is a two node layer. B. Synthetic description of architecture. C. Shallow neural network architecture. D. Deep RNN using LSTM or GRU cells.

tantly, on subjects by averaging subject frame scores and choosing the maximal probability class, i.e., using a 50% threshold.

For development purposes, we have also tested the performance of this DCNN on a synthetic dataset consisting of gaussian radial functions randomly placed on the spectrogram time axis but with variable stability in frequency, width and amplitude (i.e, by adding some jitter top these parameters). Frame classification accuracy was high and relatively robust to jitter (∼95-100%, depending on parameters).

4. RNN network architecture

The architectures for the RNNs consisted of stacked LSTM [1, 2] or GRU cells [3]. The architecture chosen includes 3 stacked cells, where each cell uses as input the outputs of the previous one. Each cell typically used about 32 hidden units, and dropout was used to regularize it. The performance of LSTM and GRU variants was very similar.

Problem	**N (train/test group)**	**Frame train/Test ACC**	**Subject Test ACC (AUC)**
DCNN: HC vs PD	2x73 / 2x1	80% / 73%	80% (87%)
RNN: HC vs PD	2x73 / 2x1	77% / 74%	81% (87%)
DCNN: HC+RBD vs PD+DLB	2x159 / 2x1	73% / 68%	73% (78%)
RNN: HC+RBD vs PD+DLB	2x159 / 2x1	76% / 68%	72% (77%)

Table 1: Performance in different problems using 1 EEG channel (P4). From left to right: Neural Network used and problem addressed (group and whether binary or multiclass); Number of subjects in training and test sets per group (always balanced); train and test performance on frames; test accuracy and leave-pair-out (LPO) the area-under-the-curve metric (AUC) on subjects [20].

5. Classification performance assessment

Classification performance has been evaluated in two ways in a leave-pair out framework (LPO) with a subject from each class. First, by working with balanced dataset using the accuracy metric (probability of good a classification), and second, by using the area under the curve (AUC) [20]. To map out the classification performance of the DCNN for different parameter sets, we have implemented a set of algorithms in Python (using the Tensorflow module [21]) as described by the following pseudocode:

```
REPEAT N times (experiments):
  1- Choose (random, balanced) training and test subject sets (leave-pair-out)
  2- Augment smaller set by random replication of subjects
  3- Optimize the NN using stochastic gradient descent with frames as inputs
  4- Evaluate per-frame performance on training and test set
  5- Evaluate per-subject performance  averaging frame outputs
END
Compute mean and standard deviation of performances over the N experiments
```

For each frame, the classifier outputs (softmax) the probability of the frame belonging to each class and, as explained above, after averaging over frames per subject we obtain the probability of the subject belonging to each class. Classification is implemented by choosing the class with maximal probability.

The results from classification are shown in Table 1. For comparison, using a shallow architecture neural network resulted in about 10% less ACC or AUC (in line with our results using SVM classifiers [22, 23]). Figure 6 provides the performance in the HC-PD problem using different EEG channels.

6. Interpretation

Once a network is trained, it can be used to explore which inputs optimally excite network nodes, including the outputs [24]. The algorithm for doing this is essentially maximizing a particular class score using gradient descent, and starting from a random noise image. An example of the resulting images using the DCNN above can be seen in Figure 4. This is particularly interesting technique, as it provides new insights on pathological features in EEG.

We also checked to see that the trained network was approximately invariant to translation of optimized input images, as expected.

7. Discussion

Our results using deep networks are complementary to earlier work with this type of data using SVMs and ESN/RNNs. However, we deem this approach superior for various reasons. First, it largely mitigates the need for feature selection. Secondly, results represent an improvement over our previous efforts, increasing performance by about 5–10% [22, 23].

We note in passing that one of the potential issues with our dataset is the presence of healthy controls without follow up, which may be a confound. We hope to remedy this by enlarging our database and by improving our diagnosis and follow up methodologies.

Future steps include the exploration of this approach with larger datasets as well as continuous refinement of the work done so far, including a more extensive exploration of network architecture and regularization schemes, including the use of deeper architectures, improved data augmentation, alternative data segmentation and normalization schemes. With regard to data preprocessing, we should consider improved spectral estimation using advanced techniques such as state-space estimation and multitapering—as in [25], and working with cortically or scalp-mapped EEG data prior creation of spectrograms.

Finally, we note that the techniques used here can be extended to other EEG related problems, such as brain computer interfaces or sleep scoring, epilepsy or data cleaning, where the advantages, deep learning approaches may prove useful as well.

Acknowledgments.

This work has been partially funded by The Michael J. Fox Foundation for Parkinson's Research under Rapid Response Innovation Awards 2013. The data was collected by Jacques Montplaisir's team at the Center for Advanced Research in Sleep Medicine affiliated to the University of Montrèal, Hopital du Sacre-Coeur, Montrèal and provided within the scope of the aforementioned project.

References

1. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation* 1997;9(8):1735–80.

2. Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016.

3. Cho K, van Merrienboer B, Gulcehre C, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), arXiv preprint arXiv:1406.1078*. 2014:.

4. Iranzo A, Fernández-Arcos A, Tolosa E, Serradell M, Molinuevo JL, Valldeoriola F, Gelpi E, Vilaseca I, Sánchez-Valle R, Lladó A, Gaig C, Santamaría J. Neurodegenerative disorder risk in idiopathic REM sleep behavior disorder: Study in 174 patients. *PLoS One* 2014;9(2).

5. Fulda S. Idiopathic REM sleep behavior disorder as a long-term predictor of neurodegenerative disorders. *EPMA J* 2011;2(4):451–8.

6. Postuma R, Gagnon J, Vendette M, Fantini M, Massicotte-Marquez J, et al . Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder. *Neurology* 2009;72:1296–300.

7. Kim Y, Kim YE, Park EO, Shin CW, Kim HJ, Jeon B. REM sleep behavior disorder portends poor prognosis in Parkinson's disease: A systematic review. *J Clin Neurosci* 2017;.

8. Fantini L, Gagnon J, Petit D, Rompre S, Decary A, Carrier J, Montplaisir J. Slowing of electroencephalogram in rapid eye movement sleep behavior disorder. *Annals of neurology* 2003;53(6):774–80.

9. Rodrigues-Brazète J, Gagnon J, Postuma R, Bertrand J, D Petit MJ. Electroencephalogram slowing predicts neurodegeneration in rapid eye movement sleep behavior disorder. *Neurobiol Aging* 2016;37(74-81).

10. Poggio T, Mhaskar H, Rosasco L, Miranda B, Liao Q. Why and when can deep but not shallow networks avoid the curse of dimensionality: a review. Tech. Rep. CBMM Memo 058; Center for Brains, minds and machines; 2016.

11. Kroupi E, Soria-Frisch A, Castellano M, nez DI, Montplaisir J, Gagnon JF, Postuma R, Dunne S, Ruffini G. Deep networks using auto-encoders for PD prodromal analysis. In: *Proceedings of 1st HBP Student Conference, Vienna*. 2017:.

12. Ruffini G. Models, networks and algorithmic complexity. *Starlab Technical Note - arXiv:161205627* 2016;TN00339(DOI: 10.13140/RG.2.2.19510.50249).

13. Ruffini G, Ibañez D, Castellano M, Dunne S, Soria-Frisch A. EEG-driven RNN classification for prognosis of neurodegeneration in at-risk patients. *ICANN 2016* 2016;.

14. Ruffini G, Ibañez D, Kroupi E, Gagnon JF, Montplaisir J, Postuma RB, Castellano M, Soria-Frisch A. Algorithmic complexity of EEG for prognosis of neurodegeneration in idiopathic rapid eye movement behavior disorder (RBD). *bioRxiv* 2017;(200543).

15. van den Oord A, Dieleman S, Schrauwen B. Deep content-based music recommendation. In: *NIPS*. 2013:.

16. Tsinalis O, Matthews PM, Guo Y, Zafeiriou S. Automatic sleep stage scoring with single-channel EEG using convolutional neural networks. *arXiv:161001683v1* 2016;.

17. Vilamala A, Madsen KH, Hansen LK. Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring. In: *2017 IEEE International Workshop On Machine Learning For Signal Processing, Sept. 25–28, 2017, Tokyo, Japan*. 2017:.

18. Dauwels J, Vialatte F, Musha T, Cichocki A. A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. *NeuroImage* 2010;(49):668–93.

19. Uhlhaas PJ, Singer W. Neural synchrony in brain review disorders: Relevance for cognitive dysfunctions and pathophysiology. *Neuron* 2006;52:155–68.

20. Airola A, Pahikkala T, Waegeman W, Baets BD, Salakoski T. A comparison of AUC estimators in small-sample studies. In: *Journal of Machine Learning Research - Proceedings Track*. 2010:3–13.

21. Abadi M, et al. . Tensorflow: Large-scale machine learning on heterogeneous systems. tensorflow.org; 2015.

22. Soria-Frisch A, Marin J, Ibañez DI, Dunne S, Grau C, Ruffini G, Rodrigues-Brazète J, Postuma R, Gagnon JF, Montplaisir J, Pascual-Leone A. Machine learning for a Parkinson's prognosis and diagnosis system based on EEG. *Proc International Pharmaco-EEG Society Meeting PEG 2014, Leipzig, Germany* 2014;.

23. Soria-Frisch A, Marin J, Ibañez D, Dunne S, Grau C, Ruffini G, Rodrigues-Brazète J, Postuma R, Gagnon JF, Montplaisir J, Pascual-Leone A. Evaluation of EEG biomarkers for Parkinson disease and other Lewy Body diseases based on the use of machine learning techniques. *in preparation* 2018;.

24. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv:13126034 [csCV]* 2014;.

25. Kim SE, Behr MK, Ba D, Brown EN. State-space multitaper time-frequency analysis. *PNAS* 2017;www.pnas.org/cgi/doi/10.1073/pnas.1702877115(1702877115).

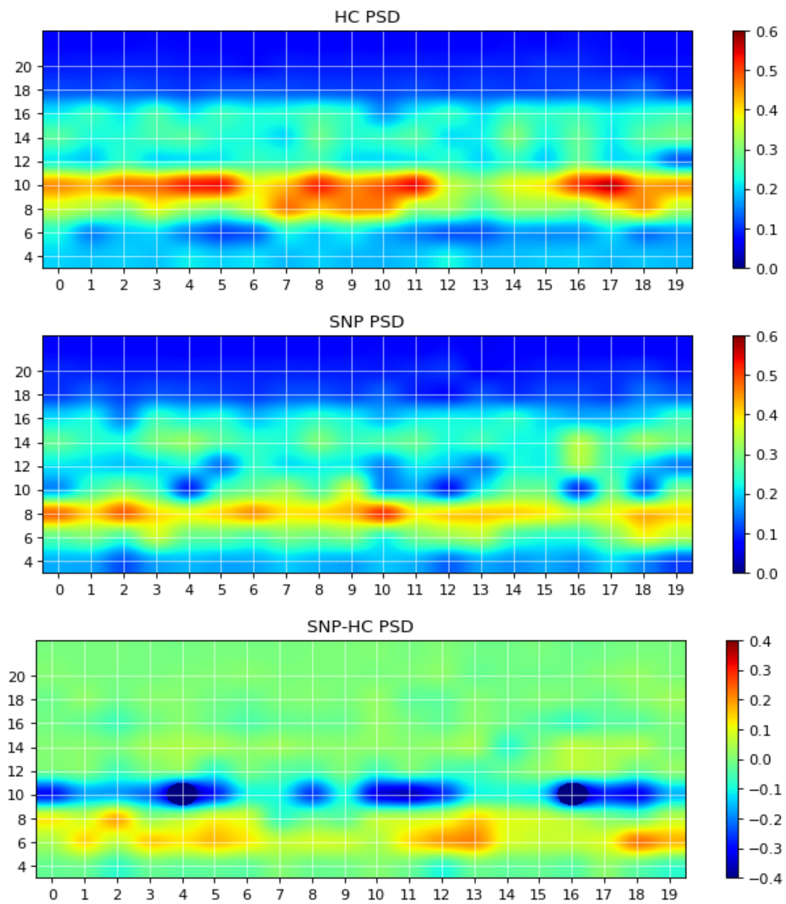

Figure 3: Sample images produced by maximizing network outputs for a given class. This particular network was trained using P4 electrode data on the problem of HC vs PD+DLB (i.e., HC vs RBDs that will develop synucleinopathy or SNP). The main features are the presence of 10 Hz bursts in the HC image (top) compared to more persistent 6–8 Hz power in the pathological spectrogram (bottom). The difference is displayed at the bottom.

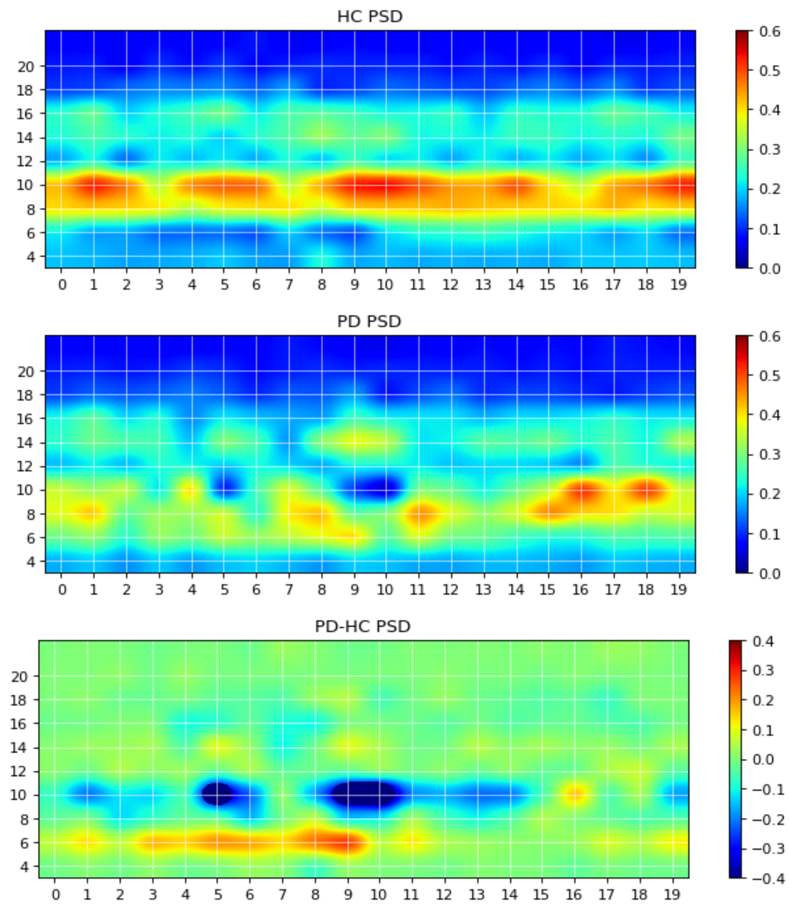

Figure 4: Sample images produced by maximizing network outputs for a given class. This particular network was trained using P4 electrode channel data on the problem of HC vs PD. The main features are the presence of 10 Hz bursts in the HC image (top) compared to more persistent 6 Hz power in the pathological spectrogram (bottom). The difference is displayed at the bottom.

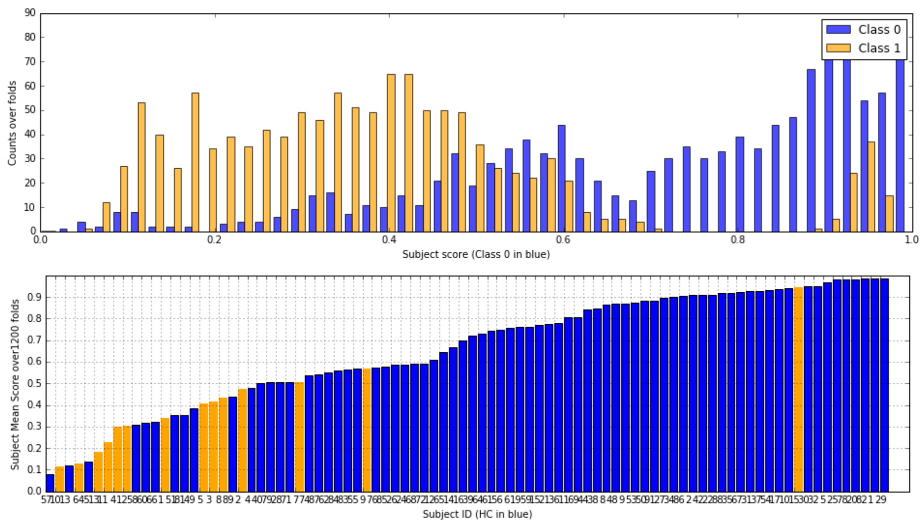

Figure 5: A. RNN Frame score histogram per class (HC in blue, PD in orange) using a single channel. B. Subject mean score. In this particular run, ACC=80%, AUC=87% (both ±1%). There are clearly some subjects that do not classify correctly (this is consistent with DCNN approach results). The PD outlier is unusual in terms of other metrics, such as slow2fast ratio (EEG slowing) or complexity (see [14]).

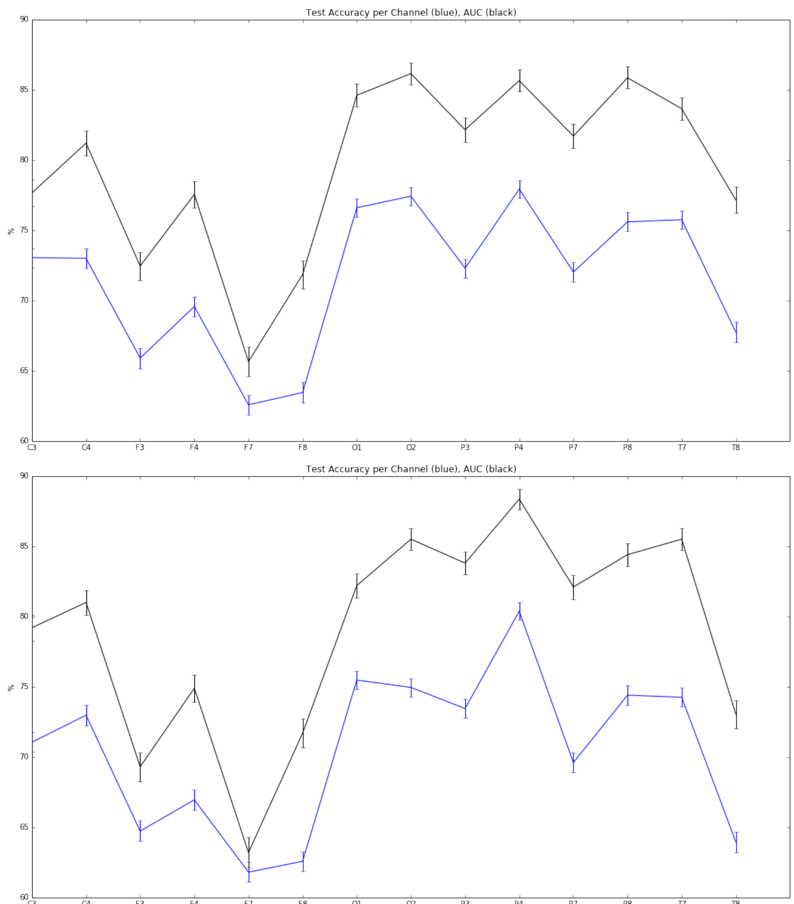

Figure 6: Top: Accuracy and AUC per channel (averages and SEM over 2000 folds). Occipital and parietal electrodes provide better discrimination. Top: DCNN architecture. Bottom: RNN.