A Crowd-Trained Machine Learning approach to automatic analysis of EEG data

Quentin Geissmann, Giorgio Gilestro
quentin.geissmann13@imperial.ac.uk
Department of Life Sciences, Imperial College London

Background:

Electroencephalography (EEG) is a widespread technique, commonly used in research and as a diagnostic tool for brain dysfunctions (e.g. epilepsy and sleep disorders). Since it is non-invasive, it is widely used to study sleep physiology in humans and other mammals.
In rodents, EEG is used in conjunction with Electromyography (EMG) in order to classify sleep in three stages (fig.1). Such "sleep-scoring" is traditionally performed manually, by experts, which is very time consuming and subjective.
Several attempts to provide and promote automatic annotation have been carried. However, they remain poorly adopted for two reasons. Firstly, automatic algorithms are not very performantwide range of data. Then, no user interface make them available for practitioners.

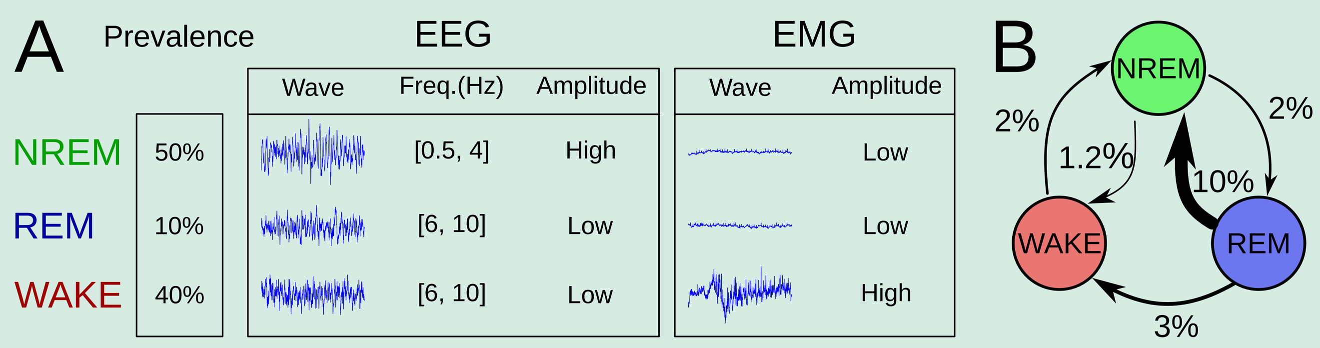

Figure 1. *The three vigilance states observed in rodents.*
A, characteristics of the EEG and EMG signals generally used by human experts.
B, proportions of transition between different stages for 5s epochs.

Aims:

1. Develop a robust algorithm for sleep-scoring from rodent EEG and EMG
2. Render the method versatile by using a wide range of experimental data
3. Implement a software tool allowing biologist to use the algorithm

Methods:

We based our work on a promising approach[2] and trained an algorithm to "understand" *the relationship between the features of the signals and sleep stages.* This was done by using expert-annotated data as a reference.
There are two critical steps:
1. Extract many features describing EEG and EMG signals
2. Use machine-learning techniques to associate features to annotations

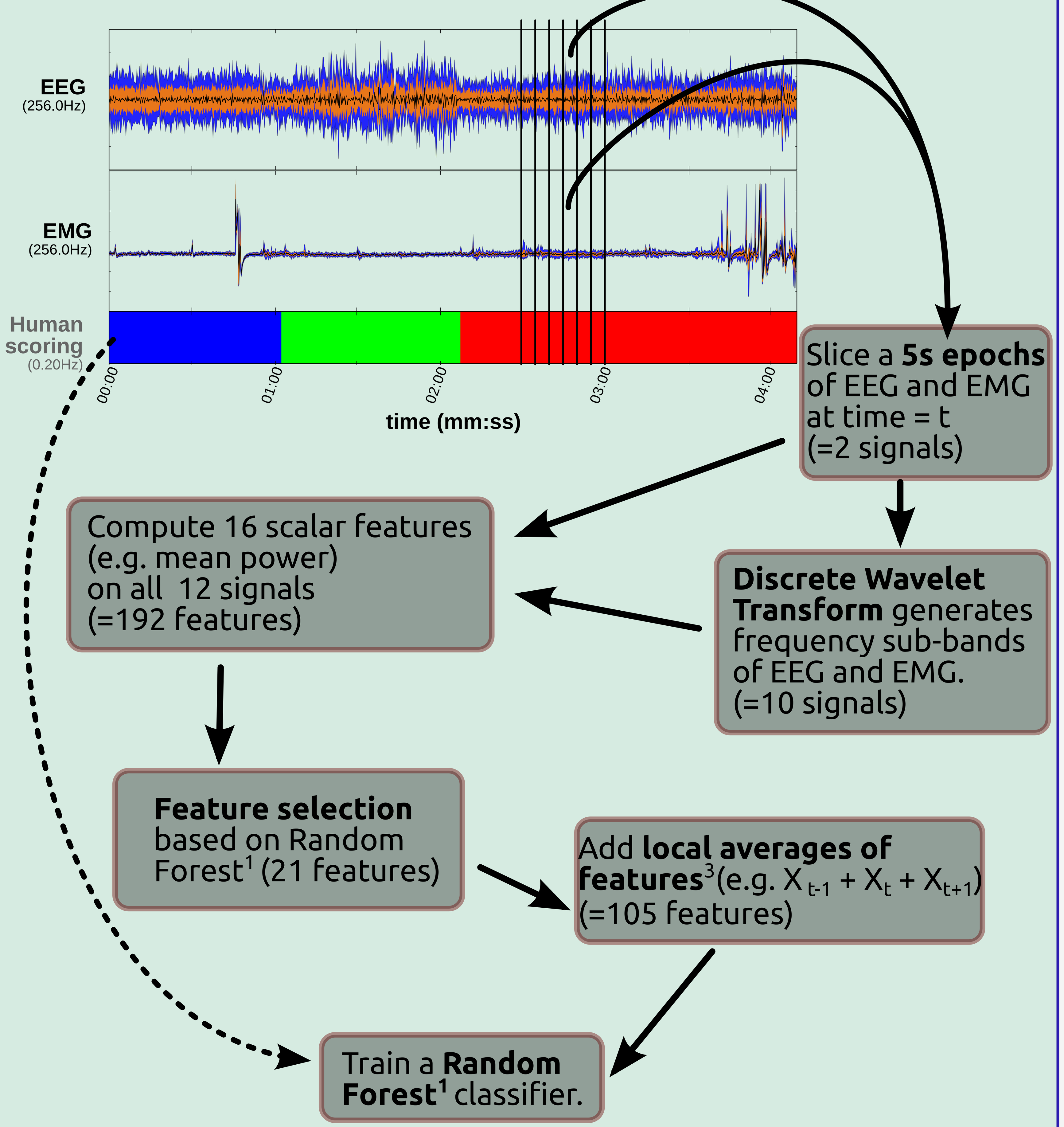

Figure 3. *Summary methodological flowchart.*
Features are extracted from every 5s of EEG and EMG. Afterwards, feature selection allow to eliminate irrelevant variables. Then, additional variables are created to account for temporal consistency.
Finally, random forest analysis is used in order to model relation between the features and the annotation.

Results:

The random forest classifier was assessed by *cross-validation*. In summary, the model is trained with a subset of the data and tested against the remaining data. The resulting confusion matrix (table 1) shows how much error is made for each vigilance state. Differences in sleep structure were also investigated (fig. 3). Finally, the algorithm can be used to generate a values of confidence along with each prediction (fig. 4).

Table 1. *Confusion matrix.*
Agreement between human(rows) and our classifier (columns). Each cell in the 3x3 confusion matrix is a percentage of the total number of epochs. The total number of compared epochs was $\simeq 2.10^5$.
1, PPV is positive predictive value (=100% - 'false discovery rate'). **2**, OA is the *overall accuracy.*

		Algorithm			PPV[1]
		NREM	REM	WAKE	(%)
Human	NREM	47.9	1.36	3.03	92
	REM	0.63	5.15	0.25	85
	WAKE	2.24	0.42	39.3	94
	Sensitivity(%)	94	74	92	OA[2]=92

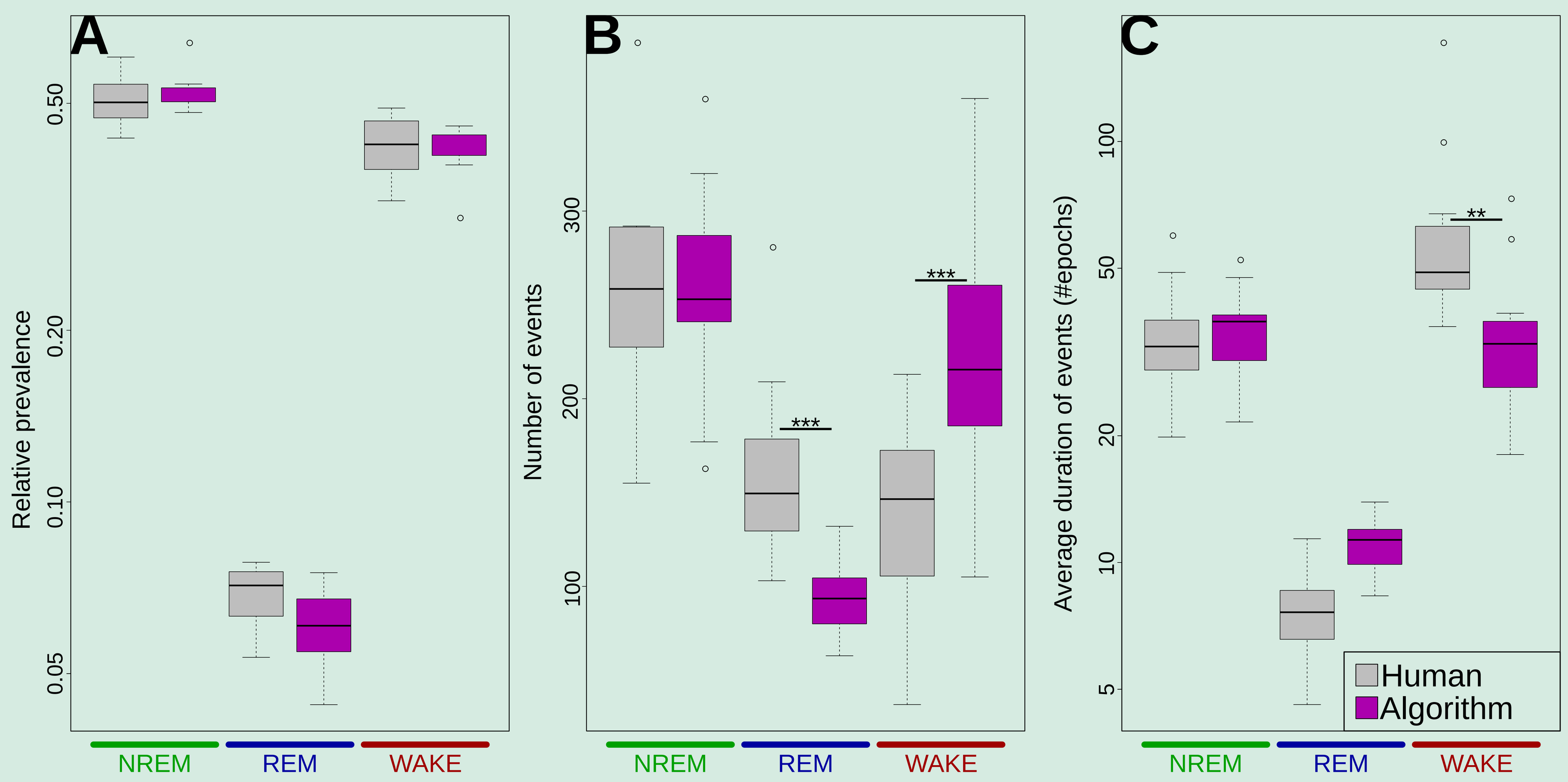

Figure 3. *Structural differences between human and classifier.*
Three metrics describing structure of sleep were computed for both ground truth and predicted time series.
A, No significant difference in state prevalence was found. **B**, The number of events was significantly over-estimated by the classifier for wake state and under-estimated for REM state. **C**, The average duration of wake and REM episodes were under-estimated and marginally over-estimated, respectively. Log scales were used for the response variables in A and C. n = 12 per combination of factors.

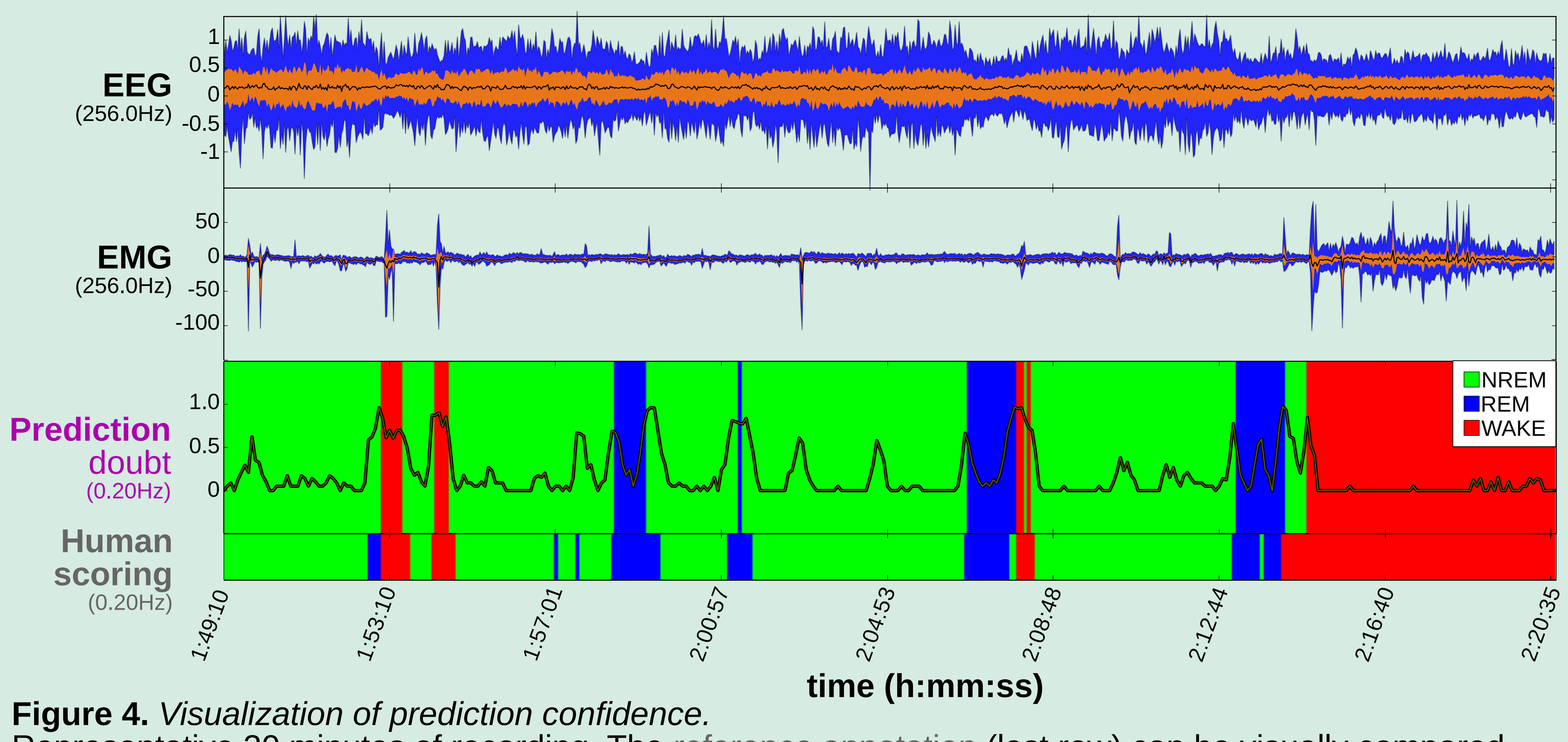

Figure 4. *Visualization of prediction confidence.*
Representative 30 minutes of recording. The reference annotation (last row) can be visually compared to the prediction (third row). The algorithm can also be used to generate a value of confidence or doubt. A doubt close to zero indicates that an epoch is unambiguously classified, whist, for values close to one, the algorithm is hardly better than a random prediction.

Take Home Message:

1. The presented approach can produce accurate (92%) predictions
2. Predicted prevalences are similar to ground truth
3. There are however structural differences in sleep fragmentation
4. A value of confidence can be generated to assess quality of predictions

References
1. L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001.
2. Şen, M. Peker, A. Çavuşoğlu, and F. V. Çelebi, "A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms," J Med Syst, vol. 38, no. 3, pp. 1–21, Mar. 2014.
3. J. J. Rodríguez, C. J. Alonso, and J. A. Maestro, "Support vector machines of interval-based features for time series classification," Knowledge-Based Systems, vol. 18, no. 4–5, pp. 171–178, Aug. 2005.

Acknowledgments
Thanks to Valentina Ferretti and Eleonora Steinberg who acquired and annotated mice recordings which were essential for this research