Supplemental Files

An EEG-based sleep index and supervised machine learning as a suitable tool for automated sleep classification in children

Authors: drs. Eris van Twist[1], MSc., drs. Floor W. Hiemstra[2,3], MSc., drs. Arnout B.G. Cramer[1], MSc., dr. Sascha C.A.T. Verbruggen[1], MD PhD, dr.ir. David M.J. Tax[4], PhD, prof.dr Koen Joosten[1], MD PhD, dr. Maartje Louter[5], MD PhD, dr. Dirk C.G. Straver[5], MD PhD, prof.dr. Matthijs de Hoog[1], MD PhD, dr. Jan Willem Kuiper[1], MD PhD, dr. Rogier C.J. de Jonge[1], MD PhD.

[1] Department of Neonatal and Pediatric Intensive Care, Division of Pediatric Intensive Care, , Erasmus MC Sophia Children's Hospital, Rotterdam, The Netherlands

[2] Department of Intensive Care, Leiden University Medical Centre, Leiden, The Netherlands

[3] Laboratory for Neurophysiology, Department of Cellular and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands

[4] Pattern Recognition Laboratory, Delft University of Technology, Delft, The Netherlands

[5] Division of Clinical Neurophysiology, Department of Neurology, Erasmus MC, Rotterdam, The Netherlands

CONTENTS

METHODS…………………………………………………………………………………….............. 3
Supplemental Methods 1: Data Exploration……………………………………………………………… 3
Supplemental Methods 2: Decision Tree………………………………………………………………. 3
Supplemental Methods 3: Nested Cross Validation…………………………………………………… 3

RESULTS…………………………………………………………………………………................. 5
Supplemental Results 1: Four State Model………………………………………….............................. 5

TABLES………………………………………………………………………………….................... 6
Table S1. Definitions of sleep quality parameters used for model evaluation……...………................. 6
Table S2. Spearman correlation of spectral powers with sleep stages……...……….............................. 7
Table S3. Candidate index measures……...………... 7
Table S4. Sleep quality for visually scored hypnograms (true) and both unsmoothed and smoothed models for three state classification……...……….. 8

FIGURES……………………………………………………………………………….................. 9
Figure S1. Logarithmic histograms of the nine candidate index measures..…………………………... 9
Figure S2. Balanced accuracy per smoothing level, averaged over all channel-index combinations… 10
Figure S3. Decision tree plot for four state classification…………………………………………... 11
Figure S4. Balanced accuracy across EEG channel for two and three state classification……………. 12

METHODS

Supplemental Methods 1: Data exploration

To assess which spectral power ratios had potential to be used in an index measure, Spearman correlations of both the absolute and relative spectral powers with the ordinal sleep stage labels were obtained from a dataset consisting of 10,000 epochs from each channel. The three most correlated bandpowers were chosen as candidate index measures along with different arithmetic combinations of these bandpowers, to create nine potential index measures in total. Only arithmetic operations were performed that led to index measures that positively correlated with the sleep states.

Supplemental Methods 2: Decision Tree

A decision tree is a supervised machine learning method that categorizes data based on whether a variable in the data exceeds a given threshold or not. This is done one variable and threshold at a time. After the first split, two nodes are created, each containing the part of the data that either exceeded or did not exceed the threshold. These nodes can then be split further and further, depending on the data and the parameters that are set. Decision trees are relatively simple to understand and interpret, require little data preparation and limited computational power due to the logarithmic relation between input (i.e. trained) and output (i.e. predicted) data. All of the prior makes the decision tree suitable for real-time, bedside sleep monitoring.

Supplemental Methods 3: Nested Cross Validation

Nested cross validation is a machine learning strategy with an inner and an outer loop to combine model training and testing with hyperparameter optimalisation. It prevents data leakage, i.e. where information from the test set leaks into the model causing false high accuracy. In our analyses, we regarded smoothing as a hyperparameter, and performed the nested cross validation on every combination of EEG channel and index measure. For each channel-index combination, we performed five-fold outer cross validation. In this process data is split into five portions, four of which are the training set and the remaining one is the test set. In each fold, data is resampled to use different portions of the data until all combinations are exhausted. Through stratified cross validation we ensured each participant was only represented in either the training or test set and that distribution of sleep states was similar for all folds. The optimal level of smoothing for each fold was based on the highest score of the inner 5-fold cross validation. This process yielded the best performing channel and index combination, but did not yield the optimal level of smoothing, because this could have been different for each of the 5 outer folds. To acquire the optimal smoothing level, the inner cross validation was repeated on the entire dataset for the different levels of smoothing, and the smoothing with the highest cross validation score was selected. The decision tree was then fitted to the full dataset of the best performing channel-index combination and optimal smoothing level to create a final model.

RESULTS

Supplemental Results 1: Four State Model

After the decision tree for four state classification classified two leaves as wake, we considered several options to improve the model post-hoc. First, we refit the decision tree using original class weights rather than balanced weights (i.e. every epoch has the same weight). We then explored the option of increasing the weight of REM compared to other epochs. While maintaining a class weight of 1 for other classes we incrementally increased the weight of REM by 0.1 until the model classified all four states. The desired tree structure was obtained at a class weight of 1.7, and this model achieved a balanced accuracy of 0.55. Reverting back to balanced class weights, we also tried increasing the number of leaves from four to 20 one leaf at a time until the model classified all states. However, this did not result in a model that classified all four states. Finally, we applied a custom-made algorithm that explored 10,000 potential cut-off values above and below which to classify different sleep states. The potential cut-off values were equally spaced over the full range of the gamma:delta-ratio. The algorithm consecutively optimized the cut-off point for the distinction between wake and sleep, then for SWS and other sleep states and lastly for the distinction between REM and NSWS. At each step, the cut-off was selected that maximized the balanced accuracy between two classes in question. This resulted in a model with a balanced accuracy of 0.58.

TABLES

Table S1. Definitions of sleep quality parameters used for model evaluation

Parameter	Definition
Wake time (hh:mm)	Total time spent in wake
NSWS time (hh:mm)	Total time spent in NSWS
SWS time (hh:mm)	Total time spent in in SWS
Total sleep time (hh:mm)	Total time spent asleep during recording time (sum of all sleep states combined)
Sleep period time (hh:mm)	Time between sleep onset and end of sleep
Sleep efficiency (%)	The percentage of time spent asleep while in bed
Number of awakenings	The number of times a participant reached wake during recording time
Sleep fragmentation index (hour^{-1})	Number of awakenings over total sleep time in hours

Note that total sleep time (TST) is defined as the total time spent asleep (i.e. sum of all sleep stages) between the onset and ending of PSG. Although PSGs follow a similar time pattern, the exact onset and ending times may vary across recordings. Daytime naps, which typically contribute to TST in infants and young children, are not included in TST with this approach.

Table S2. Spearman correlation of spectral powers with sleep stages*

	Two state	Three state	Four state
Absolute Alpha power	-0.02	-0.05	-0.13
Absolute Beta power	0.13	0.14	0.04
Absolute Delta power	-0.22	-0.39	-0.40
Absolute Gamma power	0.30	0.32	0.26
Absolute Theta power	-0.20	-0.22	-0.25
Relative Alpha power	0.16	0.35	0.29
Relative Beta power	0.30	0.47	0.41
Relative Delta power	-0.22	-0.45	-0.39
Relative Gamma power	0.41	0.57	0.54
Relative Theta power	-0.05	0.21	0.20
Total power	-0.17	-0.33	-0.35

* All p-values are statistically significant ($p < 0.001$). Correlations were determined for an exploration set of 10,000 epochs of each channel. The categorical sleep stage labels were converted to linear sleep stage labels as follows. Two states: Wake = 1, Sleep = 0; Three states: Wake = 2, NSWS = 1, SWS = 0; Four states: Wake = 3, REM = 2 NSWS = 1, SWS = 0. NSWS = non slow wave sleep, SWS = slow wave sleep.

Table S3. Candidate index measures

Calculation	**Index**	**Greek symbols**
Gamma	Relative gamma power	γ
Beta	Relative beta power	β
1/Delta	Inverse relative delta power	$1 / \delta$
Gamma/Delta	Gamma:delta-ratio	γ / δ
Beta/Delta	Beta:delta-ratio	β / δ
(Gamma + Beta)/Delta	(Gamma+Beta):delta-ratio	$(\gamma + \beta) / \delta$
Gamma – Delta	Difference between relative gamma and delta power	$\gamma - \delta$
Beta - Delta	Difference between relative beta and delta power	$\beta - \delta$

Gamma = relative Gamma power, Beta = relative Beta power, Delta = relative Delta power

Table S4. Sleep quality for visually scored hypnograms (true) and both unsmoothed and smoothed models for three state classification

Parameter	True	Unsmoothed model		Smoothed model	
	MEAN (SD)	MEAN (SD)	RMSE	MEAN (SD)	RMSE
Wake time (hh:mm)	1:15 (0:53)	2:00 (1:23)	1:16	1:57 (1:27)	1:16
NSWS time (hh:mm)	4:46 (1:30)	3:44 (1:27)	1:38	3:56 (1:37)	1:34
SWS time (hh:mm)	1:52 (1:01)	2:09 (0:58)	1:17	1:59 (57)	1:16
Total sleep time (hh:mm)	6:38 (1:39)	5:53 (1:44)	1:16	5:56 (1:50)	1:16
Sleep period time (hh:mm)	9:28 (1:18)	9:24 (1:17)	0:15	9:14 (1:22)	0:30
Sleep efficiency (%)	69.8 (14.4)	62.1 (15.6)	12.9	63.8 (16.3)	12.7
Number of awakenings	11.3 (8.0)	51.5 (26.1)	46.6	9.7 (6.0)	7.2
Sleep fragmentation index (hour^{-1})	1.8 (1.6)	9.5 (5.6)	9.1	1.9 (1.4)	1.4

SD = standard deviation, RMSE = root mean squared error, with the error being the difference between model prediction and true value for each patient.

FIGURES

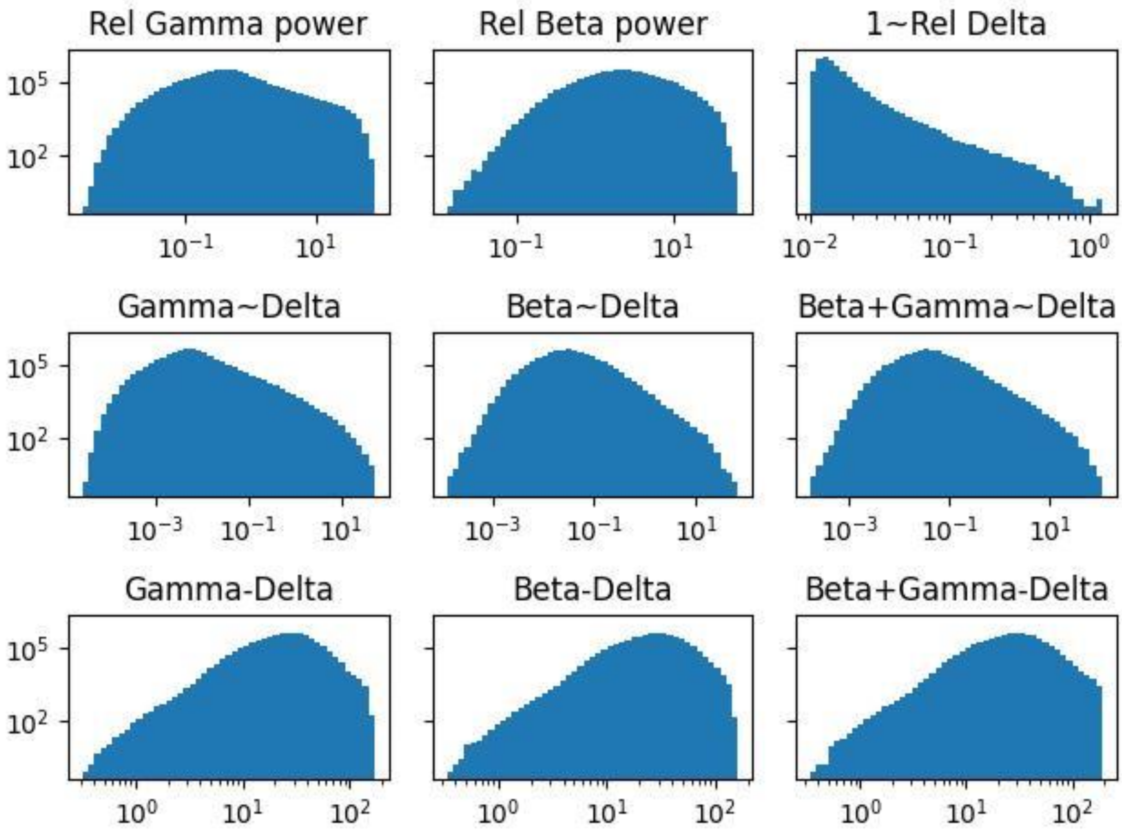


Figure S1. Logarithmic histograms of the nine candidate index measures. The candidate index measures are spectral powers or power ratios derived from EEG. Note that all index measures show right skew distributions.

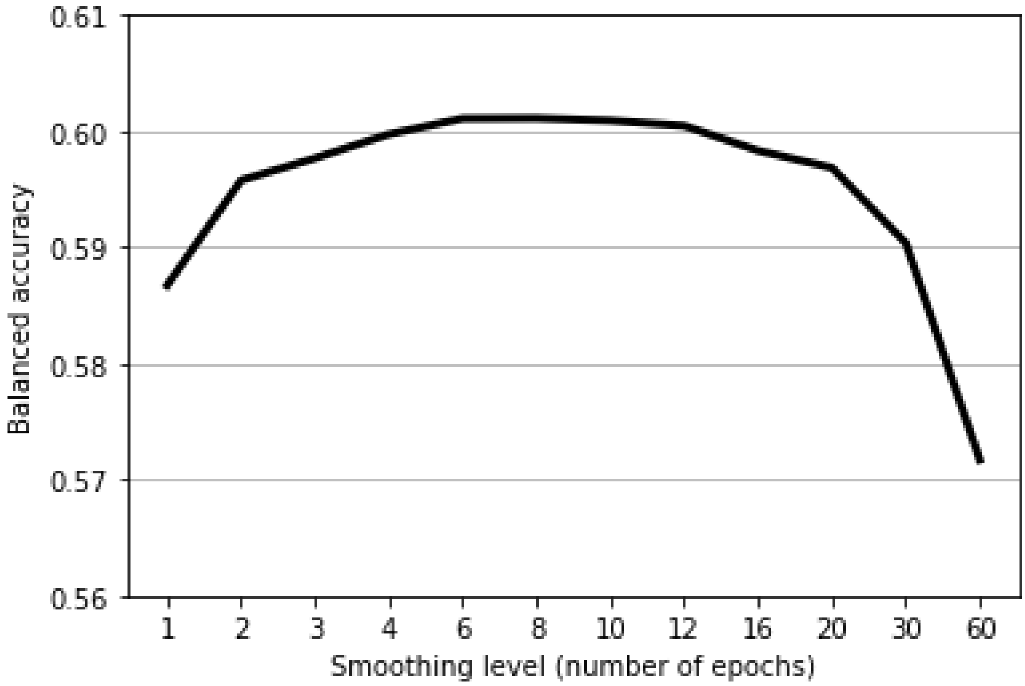


Figure S2. Balanced accuracy per smoothing level, averaged over all channel-index combinations. This figure shows the balanced accuracy, averaged over all channel-index combinations in three state classification, for each level of smoothing. The x-axis shows the smoothing level in number of epochs, ranging from 0 to 60. The y-axis shows the mean balanced accuracy over all inner cross validation scores, ranging from 0.56 to 0.61. The maximum balanced accuracy is achieved with smoothing of 8 epochs, but even at a smoothing of 60 epochs performance is on average only 0.03 lower than at the optimal level.

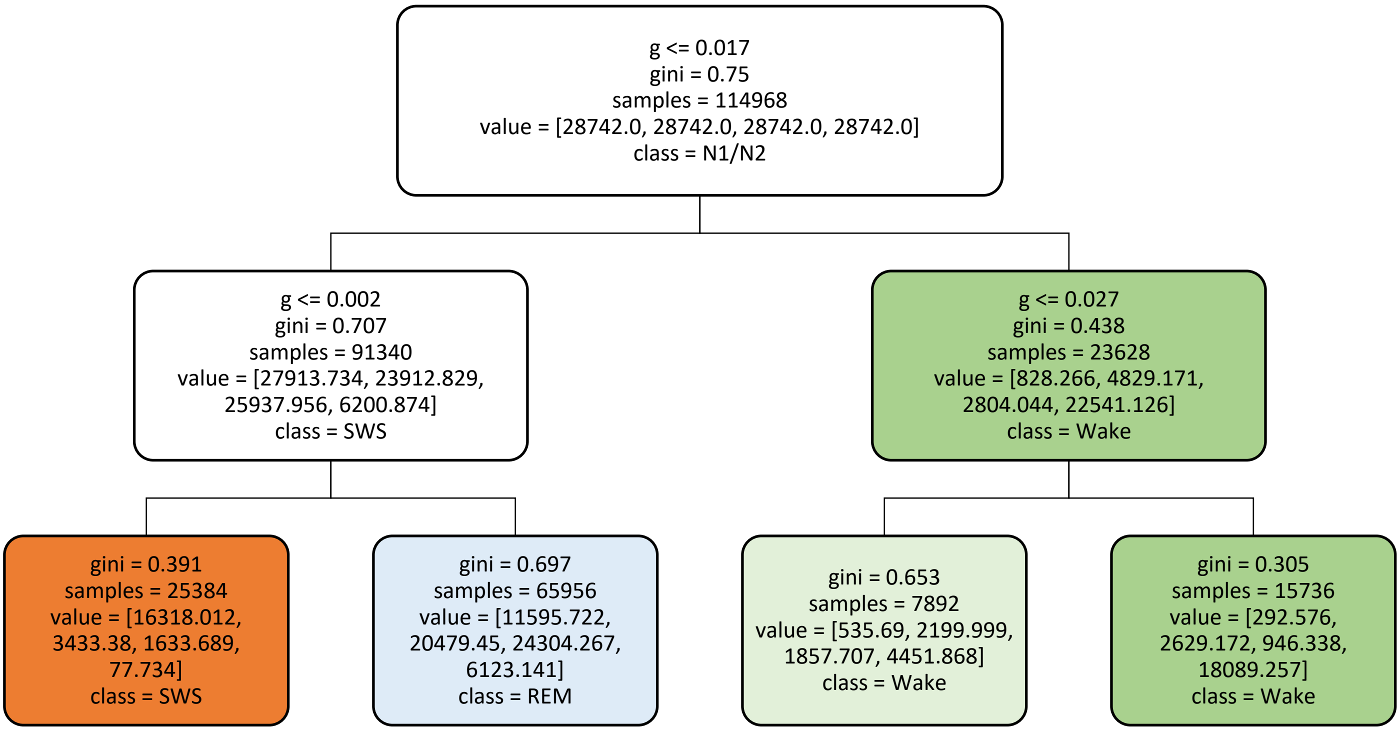
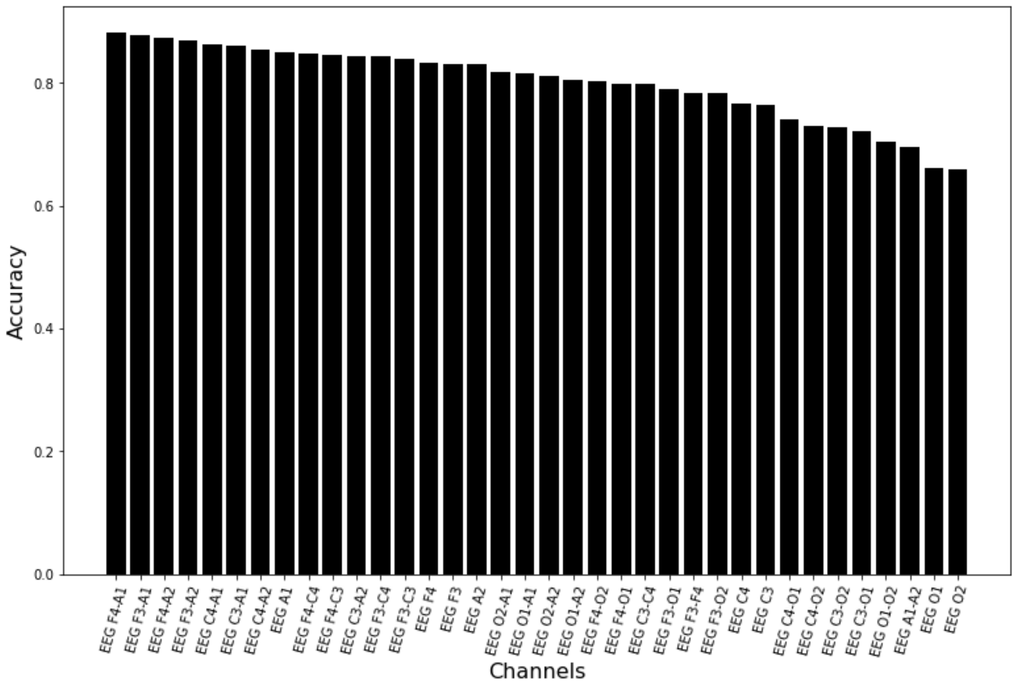


Figure S3. Decision tree plot for four state classification. Boxes correspond with nodes. Samples correspond to numbers of epochs assigned to a node, values to the respective numbers of epochs of SWS, NSWS, REM and wake assigned to a node (decimals are introduced as a consequence of using balanced class weights). Splits were made based on the gamma:delta-ratio (g) using the Gini impurity criterion calculated as 1 – gini. In this decision tree starting with N1/N2 sleep, high gamma:delta-ratios follow the right branch and low gamma:delta-ratios follow the left branch. The leave nodes are classified as SWS/REM and wake/wake. None of the leave nodes classify as NSWS**.**



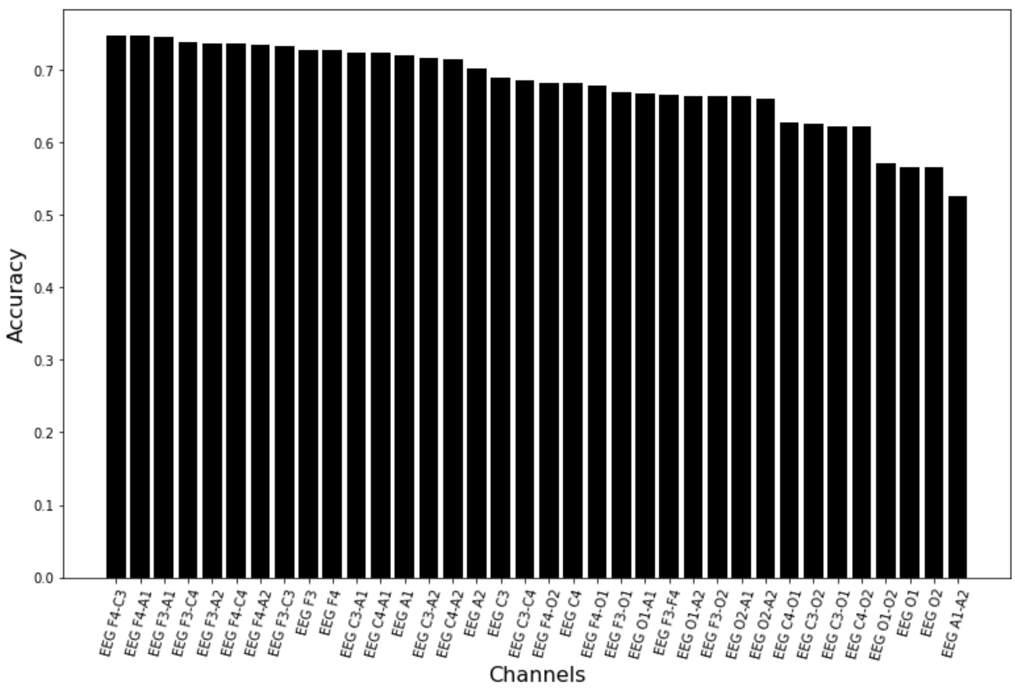


Figure S4AB. Balanced accuracy across EEG channel for two and three state classification. Figure S4A (top) depicts the balanced accuracy across EEG channels for two state classification, and figure S4B (bottom) depicts the balanced accuracy across EEG channels for three state classification. Note the variation in y-axis limits between the figures. The optimal EEG channels are F4-A1 for two state classification and F4-C3, F3-A1 and F4-A1 for three stat classification. Both models generalize well to multiple electrode pairs. F: frontal channel; A: auricular channel; O: occipital channel; C: central channel