Neurophysiological Correlates of Sleep: Investigating Brain Activity Patterns and How it Impacts Health Outcomes

Team members: Lisong He, Bowen Deng, Iuliia Dmitrieva, Kaixuan Zhang

Subject Area: Brain Science/Neurophysiology
Team name: Sleeping Beauty
Data Update: We have a large amount of auxiliary data, apart from EEG, with variables like "time entered the hospital or multiple measurements of the same quantity by the same subject" which are uncorrelated to our interested target(sleep apnea and insomnia). For the convenience of subsequent research, we need to collect all useful variables in a whole data frame. All files have a common variable patient ID, which is the distinct information for each unique patient, So we choose to merge the variables we need by the patient ID into a single data frame for further analysis. As a result, we pick up BMI, BP, AGE, etc. from more than 70 indexes in total because we want to separate patients into groups to figure out their differences in EEG plots and construct a pattern among each group. One obstacle in joining the files was to pick one variable to match each entry; the solution was to take the entry that belongs to the closest interval between measurement time and study time. Group and merge functions from Pandas made such selection possible. The difference in measurement and study time was calculated in the group of each sleep study ID, among which only the smallest time interval remained.

STUDY_PAT_ID	SLEEP_STUDY_ID	AGE AT THE TIME	SLEEP_STUDY_DU...
1	4789	9.145205479452056	9:35:32
7	12595	2.1232876712328768	11:22:32
10	22339	13.101369863013698	11:14:13
16	24241	12.076712328767123	10:03:36
22	23233	4.739726027397261	11:15:36
25	18085	17.926027397260274	12:52:22
25	10579	18.90684931506849	10:22:20

Fig 1. The merged dataset; one patient could have done multiple sleep studies

STUDY_PAT_ID	STUDY_ENC_ID	MEAS_RECORDED_DATETIME	MEAS_TYPE	MEAS_VALUE_NUMBER	MEAS_VALUE_TEXT
1	55474909	7/13/17 11:35	BMI	18.10	NaN
1	55474909	7/13/17 11:35	BMIPCT	89.00	NaN
1	55527457	7/26/17 10:29	BMI	17.87	NaN
1	55527457	7/26/17 10:36	BMIPCT	87.19	NaN
1	55552468	2/24/20 9:17	BMIPCT	51.31	NaN

Fig 2. Part of redundant data in the same measurement; patient 1 has 37 measurements for Blood Pressure and BMI

Status Update: After establishing the merged data frame, we started to analyze the relationships between those variables that affected the grouping of ages; we first used the K-means analysis among those races, ages, and sleeping time to try to find the centroids of the BMI for multivariable purpose, but the result shows vague information (figure 3), then we tried to find the single variable's relationship among ages, and no clear correlation that was found (figure 4).

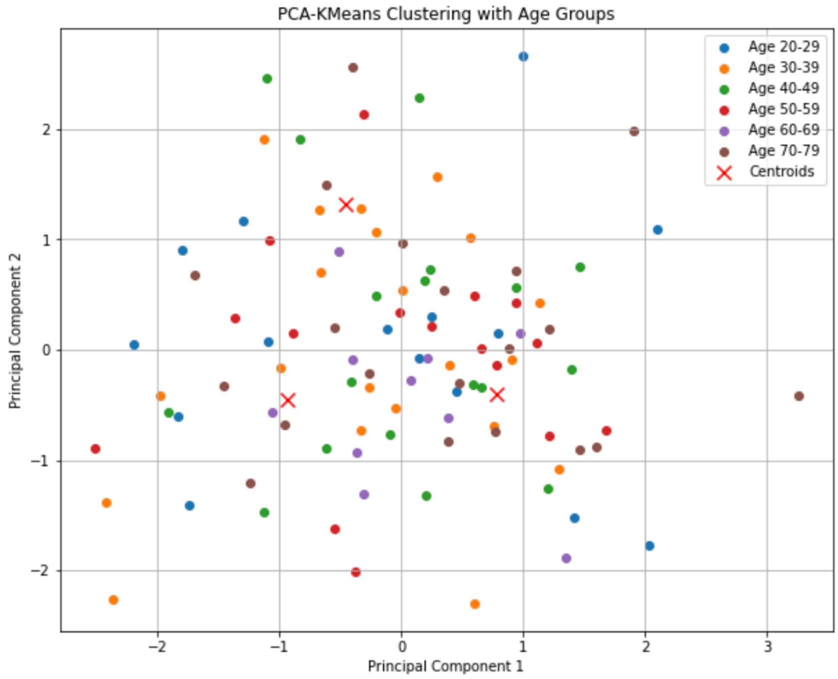

Fig 3. PCA-KMeans Clustering among age groups

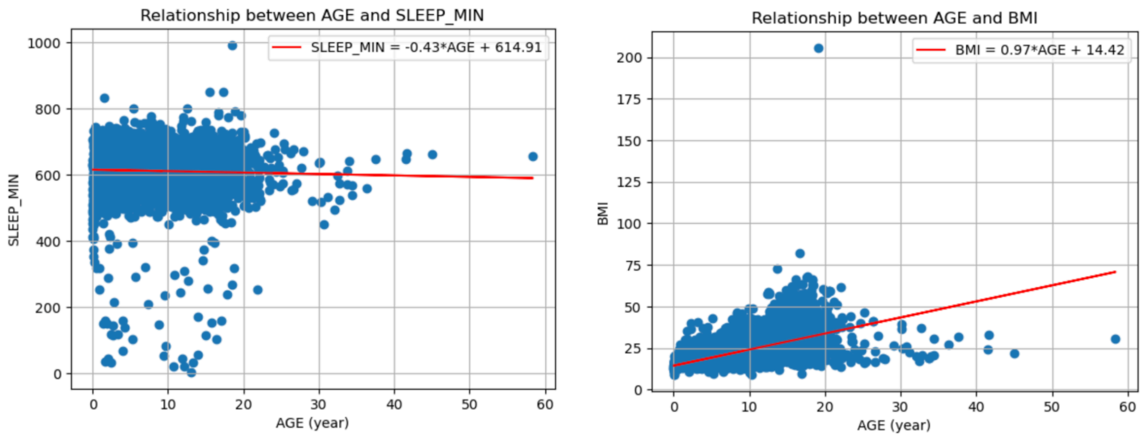

Fig 4. Relationship between Age and other variables

Therefore, we decide to use the official criteria of Age group by referencing the sleeping study paper to create age groups by sleep duration:

Age	Recommended Sleep Duration in hours	Number of subjects in dataset
0-3 months	14-17	101
4-11 months	12-15	142

1-2 years	11-14	502
3-5 years	10-13	775
6-13 years	9-11	1621
14-17 years	8-10	620
18-25 years	7-9	195
26-64 years	7-9	28
≥ 65 years	7-8	1

Table 1. The expert panel recommended sleep durations from paper

We aim to perform an in-depth analysis using the script we generated for the EEG data over the past few weeks. After many attempts to organize the workflow, we could process our first EEG file to get Global Field Power graph (the measure of overall brain activity patterns during different sleep stages) - which gives information about the depth of sleep and transitions between sleep stages and detects abnormalities related to sleep apnea and insomnia.

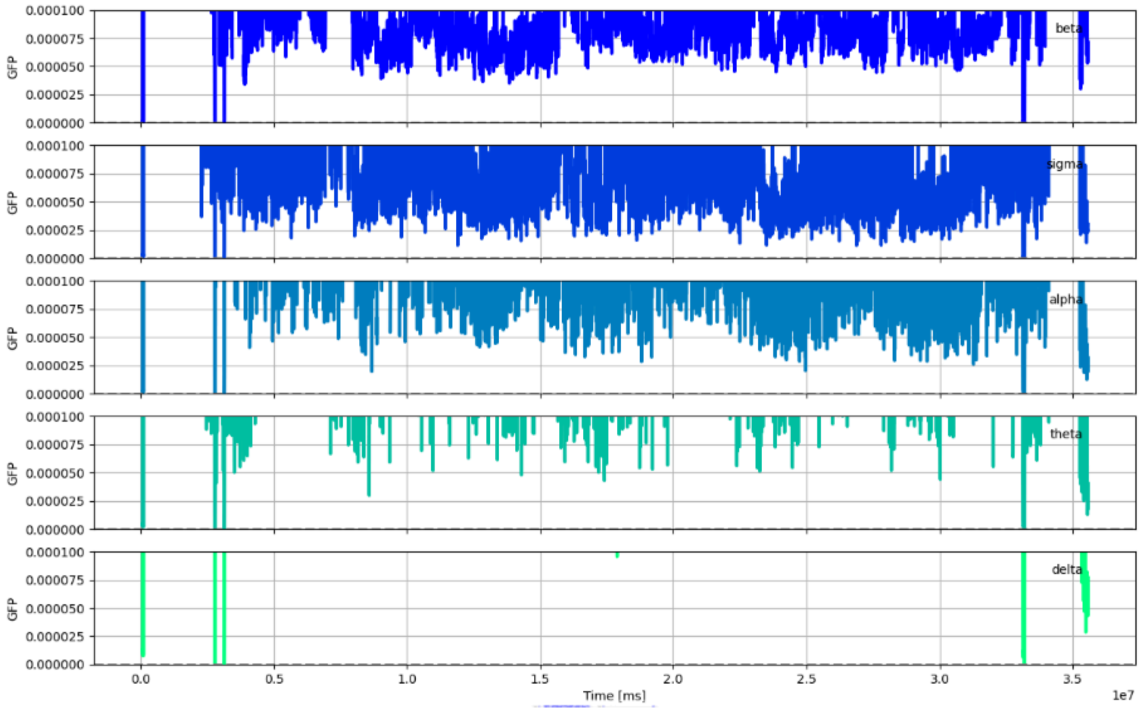

Fig 5. A sample of generated Global Field Power graph

Expectation Outcomes: We expect to do the baseline processing and data abstraction of sleep data and then match the results to every patient in the merged CSV. After that, we will use two techniques starting with decision tree for sleep apnea prediction from data available in the literature and conducting an assessment of available data to see if the criteria proposed by the scientific community show up in unsupervised-like techniques. If we do find a marker, we will explore it across different age groups. As to insomnia, we will not only use sleep apnea as a metric but also look at sleep duration, eye movement, leg movement, etc, to perform a multivariate analysis to identify factors that are prevalent in each age bracket.

References:

1. Zhang GQ, Cui L, Mueller R, Tao S, Kim M, Rueschman M, Mariani S, Mobley D, Redline S. The National Sleep Research Resource: towards a sleep data commons. J Am Med Inform Assoc. 2018 Oct 1;25(10):1351-1358. doi: 10.1093/jamia/ocy064. PMID: 29860441; PMCID: PMC6188513.
2. Lee H, Li B, DeForte S, Splaingard ML, Huang Y, Chi Y, Linwood SL. A large collection of real-world pediatric sleep studies. Sci Data. 2022 Jul 19;9(1):421. doi: 10.1038/s41597-022-01545-6. PMID: 35853958; PMCID: PMC9296671.
3. https://sleepdata.org/datasets/nchsdb
4. https://mne.tools/stable/index.html
5. Max Hirshkowitz, Kaitlyn Whiton, Steven M. Albert, Cathy Alessi, Oliviero Bruni, Lydia DonCarlos, Nancy Hazen, John Herman, Eliot S. Katz, Leila Kheirandish-Gozal, David N. Neubauer, Anne E. O'Donnell, Maurice Ohayon, John Peever, Robert Rawding, Ramesh C. Sachdeva, Belinda Setters, Michael V. Vitiello, J. Catesby Ware, Paula J. Adams Hillard,
National Sleep Foundation's sleep time duration recommendations: methodology and results summary, Sleep Health, Volume 1, Issue 1, 2015, Pages 40-43, ISSN 2352-7218