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
Proposed Research: This research plans to examine the neurophysiological sleep patterns across different age groups, focusing on medication and specific conditions like insomnia and apnea. Mining sleep data plus demographic matching will facilitate a baseline physiological profile for each age group. The study anticipates employing PCA, AdaBoost, and time series analysis to advance diagnostics and treatment of related health issues. If successful, this initiative could significantly improve patient care, offering a framework for researchers and physicians to approach sleep health.

Data Explanation: We aim to use a dataset from the National Sleep Research Source (https://www.sleepdata.org/), specifically, the Nationwide Children's Hospital - Sleep DataBank dataset. This dataset has over 4000 sleep recordings and supporting data, so it would make a perfect source for a data mining project. Notably, we have two groups of files divided into sleep and health data: first includes the sleep recording (edf format), sleep stage information (tsv format), and the recording annotations (annot format). Health data includes a rich set of data for every patient in the format of CSVs; notably, from the website, there are 5.6 million records of clinical data extracted from the EHR, and are separated into encounters, medications, measurements (e.g., body mass index), diagnoses, and procedures. We also note that the data is for a broad range (0-68) group, allowing for a more in-depth exploration and many potential findings.
Each EDF contains a time series recording brain activity from each channel of the EEG cap divided by anatomical regions of the head. As an example, the following figure includes traces that have letter references. F# corresponds to the channel location on the scalp, and M# is a supporting electrode with a brain spiking activity as a time series graph. Then figure 2 provides a map to interpret the channel placement and match it in the code.

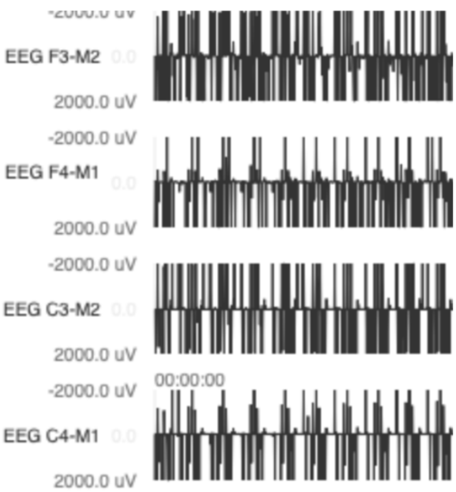

figure 1

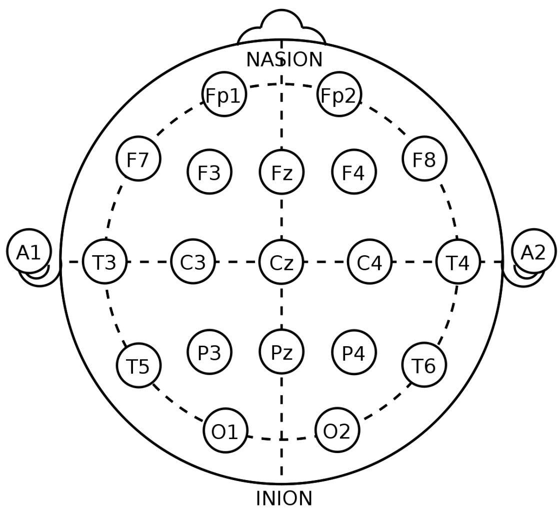

figure 2

The health data contains numerous categories representing a combination of patient ID (for matching between files and EEG recordings) and other supporting information (such as DOB, sex, and others). Below are a few examples of the Demographic and Medication Data:

	STUDY_PAT_ID	BIRTH_DATE	PCORI_GENDER_CD	PCORI_RACE_CD	PCORI_HISPANIC_CD	GENDER_DESCR	RACE_DESCR	ETHNICITY_DESCR	LANGUAGE_DESCR	PEDS_GEST_AGE_NUM_WEEKS	PEDS_GEST_AGE_NUM_DAYS
0	1	02/10/2010	M	05	N	Male	White	Not Hispanic or Latino	English	NaN	NaN
1	7	10/21/2016	F	05	N	Female	White	Not Hispanic or Latino	English	38.0	270.0
2	10	12/04/2004	F	03	N	Female	Black or African American	Not Hispanic or Latino	English	40.0	280.0
3	16	01/11/2007	M	05	N	Male	White	Not Hispanic or Latino	English	NaN	NaN
4	22	11/02/2014	M	03	N	Male	Black or African American	Not Hispanic or Latino	English	25.0	177.0

figure 3: table of basic patient information

	STUDY_MED_ID	STUDY_ENC_ID	STUDY_PAT_ID	MED_START_DATETIME	MED_END_DATETIME	MED_ORDER_DATETIME	MED_TAKEN_DATETIME	MED_SOURCE_TYPE	QUANTITY	DAYS_SUPPLY	...	GENERIC_DRUG_DESCR	DRUG_ORDER_STATUS	DRUG_ACTION	ROUTE	ROUTE_SOURCE_VALUE
0	7000001	55775794	6811	06/22/2015 08:00:00	06/24/2015 07:30:00	06/21/2015 13:56:00	06/23/2015 08:26:00	IP Meds	25.00	NaN	...	clobazam	Discontinued	Given	Oral	Oral
1	7000002	56611300	14179	06/11/2019 08:00:00	06/15/2019 00:24:00	06/11/2019 05:11:00	06/13/2019 08:37:00	IP Meds	10.00	NaN	...	cetirizine HCl	Discontinued	Given	Oral	Jejunal Tube (JT-NJ-OJ)
2	7000003	56611300	14179	06/11/2019 08:00:00	06/15/2019 00:24:00	06/11/2019 05:11:00	06/13/2019 08:37:00	IP Meds	10.00	NaN	...	cetirizine HCl	Discontinued	Given	Oral	Jejunal Tube (JT-NJ-OJ)
3	7000004	55795549	16363	12/23/2016 14:00:00	12/29/2016 11:42:00	12/23/2016 13:05:00	12/23/2016 14:27:00	IP Meds	NaN	NaN	...	NaN	Discontinued	New Bag	NaN	Intravenous
4	7000005	55795549	16363	08/08/2017 08:00:00	09/07/2017 11:18:00	08/07/2017 23:06:00	08/21/2017 20:44:00	IP Meds	2.50	NaN	...	NaN	Discontinued	Given	Oral	Feeding Tube
5	7000006	55795549	16363	03/03/2017 17:00:00	03/09/2017 20:42:00	03/03/2017 16:52:00	03/05/2017 19:00:00	IP Meds	0.08	NaN	...	NaN	Discontinued	Rate Verify	NaN	Intravenous

figure 4: table of medication information

Other datasets for use: Diagnosis.csv for the diagnosis information upon a patient; Encounter.csv for the occurrence recording of a patient's encounter hospital information, including encounter type (surgery, care, etc.); Measurement.csv, Procedure.csv, SLEEP_ENC_ID.csv, SLEEP_STUDY.csv including other helpful information upon patient for further analysis.

Technical Plan: We will first establish a data frame for each group that includes valuable variables. Identify age groups since the extensive dataset ranges from 0-68. As we know, the physiology of a child is different from that of an adult; hence, separating the data into age buckets is necessary to accommodate for anatomical and health differences. After identifying the age brackets, we will start with processing the sleep recordings using the MNE package (4) and Fourier Analysis to extract information about sleep stages and frequency bands. Then, we will match demographic files using patient ID to see how many recordings each age group has. Further, we will match all the CSV health data similarly by patient ID and identify the kinds of data types we have for each group. After matching available categories, we aim to survey them and identify a baseline physiological profile for the bracket.

Once we have established age profiles, we will perform individual analysis in each category:

- Using computational algorithms, find signals and variables to detect sleep apnea in patients who suffer from long-time sleep disorder.
- Predict the probability of patients who will be diagnosed with a sleep disorder and provide feasible suggestions to both doctors and patients.
- Perform time series analysis on the medication data to observe and record changes in patient's behavior over time.
- Build a unique understanding of sleeping patterns to promote personalized and effective treatment.

Potential Impact: This research aims to accelerate diagnosis through automated prediction, enhance workflow efficiency, foster a novel comprehension of sleep patterns, and facilitate personalized treatments. By utilizing computational algorithms, early detection of chronic sleep disorders such as sleep apnea can be achieved, increasing the accuracy of prognoses and providing practical recommendations to healthcare physicians and patients. Furthermore, integrating time series analysis on varied patient data aims to create a cohesive predictive model by tracking behavioral and laboratory changes over time, thus optimizing the diagnostic process.

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