SETSCI conference proceedings

SETSCI Conference Proceedings
17, 1-6, 2024

Cognitive Models and Artificial Intelligence Conference

May 25-26, 2024, İstanbul-Türkiye

AICCONF 2024

https://doi.org/10.36287/setsci.17.1.001
2687-5527/

The Random Forest with Hybrid Principal Component Analysis for detecting Healthy and Insomnia individuals in three subtypes of Cyclic Alternating Pattern Sleep Study

Nadhim Azeez Sayel[1], Mohammad Fayaz[2]

[1] *University of UoITC, Iraq, nazim201369@uoitc.edu.iq*
[2] *Allameh Tabataba'i University (ATU), Tehran, Iran, mohammad.Fayaz.89@gmail.com*

Abstract – Sleep disorders such as insomnia, sleep bruxism, narcolepsy, and sleep-disordered breathing have different sleep quality parameters such as cyclic alternating pattern phases and A subtype phases. This study aims to analyze and predict the A phase subtypes from electroencephalography (EEG) data. The EEG dataset for 27 (8 control, 2 bruxism, 8 insomnia, 5 narcolepsy, and 4 sleep-disordered breathing) people were downloaded and processed. The re-reference method is a common average reference (CAR), and the filtering method is a basic FIR method with lower and higher edge of frequency pass bands of 0.5 and 40 Hz, respectively. The sinusoidal artifacts (line noise) were removed with a CleanLine plug-in. The remaining noises were execrated visually and then Independent Component Analysis (ICA) with ICLabel plug-in. The Power Spectral Density (PSD) Parameters such as 0.5 to 2 Hz for Delta, 4 to 6 Hz for Theta, 8 to 12 for Alpha, 18 to 22 Hz for Beta and 30 to 60 for Gamma brainwaves, alpha peak frequency, and alpha asymmetry were calculated with eegstats plugin. The hybrid principal component analysis (HPCA) was used to estimate the functional, longitudinal, and location effects, and functional principal components analysis (FPCA) was estimated to reach at least 90% fraction of variance (FVE). Finally, the random forest with HPCA and FPCA were compared. The CAP time, NREM time, and CAP Rate were calculated. The PSD difference between groups is not statistically significant for most of the frequency bands, meaning they did not show any differences between different groups. The HPCA is estimated for each dimension with 90% FVE. The result of random forest on training and testing dataset with first to nine HPCA and FPCA eigenfunctions of time domain showed that in all subtypes, the HPCA performs better than FPCA according to the accuracy, sensitivity, and specificity. The random forest with HPCA had the best performance while the random forest without FPCA had an accuracy of about 40% in training and testing.

Keywords – Sleep Disorders, Functional Data Analysis, Cyclic Alternating Pattern, Phase A subtypes, Random Forest

I. Introduction

The prevalence of sleep disorders such as insomnia is about 10% in adults and many people suffer from it (1). In this regard, promoting healthy sleep is essential, especially in developing countries, to improve physical and mental health and social wellbeing. (2) a recent study showed that the relationship between sleep disorders such as hypersomnolence and fatigue is high during COVID-19 (3). Among the sleep quality measures based on the electroencephalography (EEG) records, various sleep macrostructures and microstructures such as Full night polysomnography (PSG) and cyclic alternating pattern (CAP) are very famous (4). In non-rapid eye movement (non-REM) sleep, the CAP, as a marker of sleep instability, has two main phases: an activation phase, or A phase, and a quiescent phase, or B phase, with a duration between 2 to 60 s. (5) The A phase has three subtypes: A1, A2, and A3 (6).

The recent meta-analysis provides the references for CAP values and A subtypes for healthy people, and it shows that the A1 subtype is high in school-aged children, and old people have the highest number of subtypes A2 and A3. (7) motor sleep disorders such as sleep bruxism (SB) that are related to sleep arousal have higher A3 amount or higher sleep arousal strength against the control groups. (8) The differences between A subtypes are presented in the study with primary insomnia and normal sleepers, and A3 is higher in the insomnia group. (9) Another study comparing sleep-wake disorders such as narcolepsy–cataplexy (NC) with healthy controls showed that subtype A2 is not statistically different in the two groups. (10) Two studies discuss subtypes in sleep-disordered breathing (SBD). (11, 12)

EEGLAB is the most famous software for analyzing EEG datasets and experiments (13). It has various plug-ins for different analyses, such as Dusk2Dawn to automatic artifact removal in whole-night sleep with EEG (14) and sleep spindle detection (15). The automatic CAP detection was studied with many machine learning algorithms such as random forest (16) and studying A phase subtypes with one-dimensional symbolic aggregate approximation (1d-SAX) (17). In this study, we compare and predict A phase subtypes in different groups with hybrid principal components analysis (HPCA), functional principal component analysis (FPCA), and random forest.

II. Materials and Method

A. Downloading Dataset

The EEG dataset came from a study about Cyclic Alternating Patterns (CAP) in sleep (18). It is downloaded

from PhysioNet at this address (https://archive.physionet.org/physiobank/database/capslpdb/#database-files) (19). The complete dataset includes 108 polysomnographic recordings from the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy. It includes recording with EEG, EMG, etc., in 10-20 international systems. (20) In this study, only control, bruxim, insomnia, narcolepsy and sleep-disordered breathing groups were selected. Among them, those patients who has at least five channels including F2-F4 (Frontal Right), F4-C4 (Central Right), C4-P4 (Parietal Right), P4-O2 (Occipital Right) and C4-A1 (Central Left) are selected. The age, gender and CAP time, and rate of each participant were presented in a descriptive way. The CAP rate is CAPtime/NREMtime (18)

In this study, the following events are selected. Three phases of A periods from the CAP, including subtype A1, A2, and A3, are related to the synchronized events with low impact on autonomic and somatomotor activities, mixed synchronized–desynchronized EEG events with an intermediate influence on the autonomic and somatomotor activities, and predominantly desynchronized EEG events with heavy effects on the autonomic and somatomotor activities, respectively. (21, 22).

B. Data Preprocessing

The dataset was preprocessed with the EEGLAB (23) in MATLAB R2021b. The epoch of each event is from 800 ms before the event to 1500 ms after the events. The sampling rate is 512 Hz, the re-reference method is a common to average reference (CAR), and the filtering method is a basic FIR method with the lower and higher edge of frequency pass bands are 0.5 and 40 Hz, respectively. The sinusoidal artifacts (line noise) were removed with a CleanLine plug-in. The remaining noises were execrated visually, and then Independent Component Analysis (ICA) was estimated (23). Only estimated independent components with more than 50% probability with brain sources were selected, and others were removed with ICLabel plug-in (24). The comparison between groups was done with one way analysis of variance (ANOVA) and nonparametric Kruskal-Wallis test.

Finally, the eegstats plugin was used to estimate the Power Spectral Density (PSD) Parameters using the Welch function of size frequencies of channels, including 0.5 to 2 Hz for Delta, 4 to 6 Hz for Theta, 8 to 12 for Alpha, 18 to 22 Hz for Beta and 30 to 60 for Gamma brainwaves, alpha peak frequency, and alpha asymmetry.

C. Statistical Analysis

The descriptive statistics of participants were reported. Then, the dataset is split into three groups for each subtype: 1) Dataset for subtype A1, 2) Dataset for subtype A2, and 3) Dataset for subtype A3. The hybrid principal component analysis (HPCA) was used in each dataset to estimate the functional, longitudinal, and location effects. (25)

In this regard, the time domain of each evoked related potential (ERP) is considered a functional domain for a functional dimension, from -700 ms to 1440 ms after each observed subtype A. The number of each subtype is considered as a functional domain for longitudinal dimension because the number of subtypes is not the same for each participant, only the minimum number of them that exist in all participants were used. For example, subtype A1 is 13 subtypes A1 events, subtype A2 is 10 subtype A2 events and subtype A3 is 13 subtype A3 events. The location dimension is the channel location with five channels, including F2-F4 (Frontal Right), F4-C4 (Central Right), C4-P4 (Parietal Right), P4-O2 (Occipital Right), and C4-A1 (Central Left) are selected. The HPCA were used to estimate the eigenfunctions and eigenscore for each group. Previously, the HPCA were estimated for ERP in healthy patients with visual and auditory oddball task and P300 components (26).

In this study, the penalized functional principal components analyse (FPCA) for only time domain of ERP were also estimated for comparison purposes with HPCA. In this regard, the penalization is the square of the slope or velocity and the fdata2pc() function from fda.usc package were used. (27) The difference between HPCA and FPCA is that in the first one, it considers three dimensions, including time, repeats, and location of ERPs, but in the second one, it only considers the time of ERP and does not consider the other two. Therefore, the second one is simpler than the first one.

Finally, the machine learning method previously used for Diffusion Tensor Imaging (DTI) is used (28). The categorical response is the multinomial five response groups (1=Control, 2=Bruxism, 3=Insomnia, 4=Narcolepsy, 5=SDB) in each subtype. The first to ninth eigenfunctions of the time domain in HPCA and the first to ninth eigenfunctions of the time domain in FPCA were covariates in two scenarios for each subtype. With the first to ninth eigenfunctions, more than 90% of the fraction of variance explained (FVE) was extracted. The training and testing datasets were randomly selected with 70% and 30%, respectively. The accuracy, sensitivity, and specificity for five categories for both training and testing were reported. All calculations are done with R version 4.3.1 and RStudio 2023.06.2. Previously, a study showed that using HPCA in the EEG-fMRI dataset improved the speed of calculation and accuracy of the predictions (29).

III. RESULTS

A. Descriptive Statistics

The list of participant codes along with gender, age in years, CAP time, NREM time, and CAP Rate (18) that has at least 5 mentioned EEG electrodes are presented in Table 1-1. The N16 was excluded in most analyses due to some missing values in time domains.

Table 1: The List of Participant.

Groups	Sample*	Gender**	Age (years)	CAP time	NREM time	CAP Rate***
Control	N1	F	37	12139	26040	0.4662
	N3	F	35	7167	20310	0.3529
	N4	F	25	6165	17370	0.3549
	N5	F	35	11601	22980	0.5048
	N11	F	28	8554	18480	0.4629
	N16	F	41	12765	22260	0.5735
	N2	M	34	7642	21180	0.3608
	N10	M	23	5000	17280	0.2894
Bruxism	BRUX1	M	34	8653	22380	0.3866
	BRUX2	M	23	12236	21030	0.5818
Insomnia	INS2	F	58	4605	18150	0.2537
	INS4	F	58	8185	15480	0.5287
	INS5	F	59	10621	18870	0.5629
	INS6	F	54	6156	14250	0.4320
	INS7	F	47	10830	19590	0.5528
	INS1	M	54	7471	19830	0.3768
	INS3	M	82	7086	15090	0.4696
	INS8	M	64	10262	15900	0.6454
Narcolepsy	NARCO1	F	29	3404	19080	0.1784
	NARCO2	F	44	17474	24060	0.7263
	NARCO3	F	18	10822	18630	0.5809
	NARCO4	M	43	11649	21600	0.5393

	NARCO5	M	24	5029	8220	0.6118
Sleep-disordered breathing	SBD1	M	65	5744	13260	0.4332
	SBD2	M	77	18799	24630	0.7633
	SDB3	M	78	13249	16950	0.7817
	SDB4	M	65	23306	27180	0.8575

*The code in the dataset, **F: Female/M: Male , *** CAP calculated with Terzano's rules.

B. Power Spectral Density (PSD) Parameters Comparisons

The comparison between PSD parameters group by Frequency Bands and Groups are reported in tables 2 to 4 for subtypes A1, A2, and A3, respectively. The difference between groups is not statistically significant for most of the frequency bands and it means that they did not show any differences between different groups.

Table 2: The Comparison between Power Spectral Density (PSD) Parameters between Groups group by Frequency Bands for All Channels in Subtype A1

Frequency Bands	**Group**	**N Subject**	**AVG**	**SD**	**P-Value**	
					ANOVA	**Kruskal-Walis**
0.5-2.0 Hz (Delta)	Normal	8	13.55	5.57	*<0.05*	*<0.05*
	Bruxism	2	9.67	4.04		
	Insomnia	8	6.02	2.20		
	Narcolepsy	5	11.77	3.53		
	SDB	4	4.95	5.07		
4.0-6.0 Hz (Theta)	Normal	8	3.75	4.25	*0.08*	*0.03*
	Bruxism	2	0.78	2.12		
	Insomnia	8	-0.05	2.12		
	Narcolepsy	5	4.39	4.03		
	SDB	4	-1.43	5.20		
8.0-12.0 Hz (Alpha)	Normal	8	0.22	4.64	*0.51*	*0.34*
	Bruxism	2	-1.61	1.24		
	Insomnia	8	-2.27	1.46		
	Narcolepsy	5	0.09	2.83		
	SDB	4	-2.67	5.08		
18.0-22.0 Hz (Beta)	Normal	8	-7.02	5.11	*0.75*	*0.43*
	Bruxism	2	-10.20	2.32		
	Insomnia	8	-6.33	2.74		
	Narcolepsy	5	-7.84	2.85		
	SDB	4	-8.36	5.09		
30.0-60.0 Hz (Gamma)	Normal	8	-13.22	5.94	*0.33*	*0.32*
	Bruxism	2	-20.56	7.45		
	Insomnia	8	-14.89	2.03		
	Narcolepsy	5	-14.44	3.30		
	SDB	4	-15.37	3.14		
Peak alpha frequency	Normal	8	9.73	1.38	*0.61*	*0.65*
	Bruxism	2	9.90	0.36		
	Insomnia	8	8.98	0.92		
	Narcolepsy	5	9.39	1.08		
	SDB	4	8.86	1.23		
Alpha center of gravity	Normal	8	9.93	1.14	*0.99*	*0.94*
	Bruxism	2	10.11	0.12		
	Insomnia	8	9.91	1.05		
	Narcolepsy	5	9.76	1.52		
	SDB	4	9.77	0.42		

Calculated with eegstats plug-in in EEGLAB

Table 3 : The Comparison between Power Spectral Density (PSD) Parameters between Groups group by Frequency Bands for All Channels in Subtype A2

Frequency Bands	**Group**	**N Subject**	**AVG**	**SD**	**P-Value**	
					ANOVA	**Kruskal-Walis**
0.5-2.0 Hz (Delta)	Normal	8	12.07	5.85	*0.03*	*0.02*
	Bruxism	2	7.30	0.71		
	Insomnia	8	5.13	3.30		
	Narcolepsy	5	9.42	1.88		
	SDB	4	7.55	3.29		
4.0-6.0 Hz (Theta)	Normal	8	3.33	4.47	*0.15*	*0.13*
	Bruxism	2	0.67	1.32		
	Insomnia	8	-0.69	3.03		
	Narcolepsy	5	3.85	3.56		
	SDB	4	0.10	3.93		
8.0-12.0 Hz (Alpha)	Normal	8	-0.47	5.10	*0.76*	*0.55*
	Bruxism	2	-2.92	0.28		
	Insomnia	8	-2.74	2.34		
	Narcolepsy	5	-0.91	2.72		
	SDB	4	-0.65	5.74		
18.0-22.0 Hz (Beta)	Normal	7	-7.97	5.83	*0.93*	*0.86*
	Bruxism	2	-8.84	1.15		
	Insomnia	8	-6.82	3.55		
	Narcolepsy	5	-8.61	3.36		
	SDB	4	-6.85	5.37		
30.0-60.0 Hz (Gamma)	Normal	8	-13.28	5.21	*0.73*	*0.43*
	Bruxism	2	-16.76	1.54		
	Insomnia	8	-14.89	2.90		
	Narcolepsy	5	-15.01	1.56		
	SDB	4	-13.89	2.53		
Peak alpha frequency	Normal	8	10.16	1.45	*0.82*	*0.86*
	Bruxism	2	9.63	0.52		
	Insomnia	8	10.06	1.26		
	Narcolepsy	5	9.66	1.04		
	SDB	4	9.32	1.28		
Alpha center of gravity	Normal	8	10.15	1.14	*0.61*	*0.67*
	Bruxism	2	9.93	0.04		
	Insomnia	8	10.19	1.04		
	Narcolepsy	5	9.96	0.36		
	SDB	4	9.32	0.67		

Calculated with eegstats plug-in in EEGLAB

Table 4: The Comparison between Power Spectral Density (PSD) Parameters between Groups group by Frequency Bands for All Channels in Subtype A3

Frequency Bands	**Group**	**N Subject**	**AVG**	**SD**	**P-Value**	
					ANOVA	**Kruskal-Walis**
0.5-2.0 Hz (Delta)	Normal	8	10.07	5.10	*0.28*	*0.12*
	Bruxism	2	7.97	2.08		
	Insomnia	8	6.83	2.93		
	Narcolepsy	5	9.64	2.69		
	SDB	4	5.93	2.89		
4.0-6.0 Hz (Theta)	Normal	8	1.62	4.07	*0.21*	*0.14*
	Bruxism	2	-1.19	2.98		
	Insomnia	8	0.26	2.64		
	Narcolepsy	5	3.85	3.68		
	SDB	4	-0.79	2.32		
8.0-12.0 Hz (Alpha)	Normal	8	-1.29	5.14	*0.68*	*0.50*
	Bruxism	2	-3.76	0.96		
	Insomnia	8	-1.85	3.05		
	Narcolepsy	5	0.47	2.38		
	SDB	4	-2.04	3.47		
18.0-22.0 Hz (Beta)	Normal	8	-8.62	4.92	*0.53*	*0.57*
	Bruxism	2	-10.05	2.42		
	Insomnia	8	-5.84	3.19		
	Narcolepsy	5	-6.99	3.01		
	SDB	4	-8.23	3.84		
30.0-60.0 Hz (Gamma)	Normal	8	-15.28	5.50	*0.97*	*0.85*
	Bruxism	2	-14.17	1.23		
	Insomnia	8	-15.36	3.09		
	Narcolepsy	5	-14.77	2.34		
	SDB	4	-16.12	2.74		
Peak alpha frequency	Normal	8	10.25	0.68	*0.14*	*0.09*
	Bruxism	2	11.22	1.53		
	Insomnia	8	9.35	1.64		
	Narcolepsy	5	8.71	1.64		
	SDB	4	8.81	1.46		
Alpha center of gravity	Normal	8	10.11	0.77	*0.54*	*0.49*
	Bruxism	2	10.72	1.00		
	Insomnia	8	9.86	0.89		
	Narcolepsy	5	9.57	1.74		
	SDB	4	9.22	1.14		

Calculated with eegstats plug-in in EEGLAB

C. The HPCA result

The number and fraction of variance explained (FVE) of functional principal components (FPC) for longitudinal and functional dimensions and principal component analysis (PC) for regional dimension were reported group by Group and sub Type in Table 5. For example, in subtype A1 and group 1 (control), with 1 PC in the regional dimension, 94.1% FVE of the regional dimension was captured. The first two FPCs of longitudinal dimensions captured 88.4% + 11.3% FVE of longitudinal dimension. And with first seven FPCs of functional dimensions, 25.4% + 18.8% + 14.7% + 9.7% + 8.7% + 8.2% + 6.1% FVE of functional dimensions were captured. The last column in the table shows the minimum number of PC that has at least 90% FVE in each dimension.

Table 5: The PCs and FVE group by Sub Type, Group and dimensions.

Sub Type	Group*	Dim	Principal Components (FVE %)								Final
			PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PCs
A1	1	REG	94.1%	-	-	-	-	-	-	-	1
		LONG	88.4%	11.3%	-	-	-	-	-	-	2
		FUNC	25.4%	18.8%	14.7%	9.7%	8.7%	8.2%	6.1%	-	7
	2	REG	100%	-	-	-	-	-	-	-	1
		LONG	86.1%	13.4%	-	-	-	-	-	-	2
		FUNC	27.3%	23.2%	14.6%	11.1%	7.9%	6.4%	-	-	6
	3	REG	100%	-	-	-	-	-	-	-	1
		LONG	92.5%	-	-	-	-	-	-	-	1
		FUNC	25.3%	20.4%	14.3%	11.4%	7.7%	6.4%	5.7%	-	7
	4	REG	100%	-	-	-	-	-	-	-	1
		LONG	50.6%	49.4%	-	-	-	-	-	-	2
		FUNC	27.4%	20.7%	13.6%	9.5%	7.3%	7.1%	5.3%	-	7
	5	REG	96.2%	-	-	-	-	-	-	-	1
		LONG	87.1%	11.4%	-	-	-	-	-	-	2
		FUNC	27.0%	21.5%	15.7%	9.9%	8.3%	5.8%	5.0%	-	7
A2	1	REG	100%	-	-	-	-	-	-	-	1
		LONG	100%	-	-	-	-	-	-	-	1
		FUNC	21.3%	20.1%	14.3%	12.4%	9.7%	7.0%	5.4%	-	7
	2	REG	100%	-	-	-	-	-	-	-	1
		LONG	61.7%	38.0%	-	-	-	-	-	-	2
		FUNC	35.3%	22.0%	13.7%	8.1%	7.4%	4.9%	-	-	6
	3	REG	100%	-	-	-	-	-	-	-	1
		LONG	44.1%	23.4%	18.0%	7.9%	-	-	-	-	4
		FUNC	22.1%	15.8%	15.3%	12.2%	9.5%	8.3%	6.7%	-	7
	4	REG	100%	-	-	-	-	-	-	-	1
		LONG	78.3%	21.7%	-	-	-	-	-	-	2
		FUNC	24.9%	17.8%	15.3%	10.7%	9.1%	7.3%	4.7%	4.5%	8
	5	REG	100%	-	-	-	-	-	-	-	1
		LONG	100%	-	-	-	-	-	-	-	1
		FUNC	32.9%	18.3%	13.9%	10.5%	7.7%	6.7%	-	-	6
A3	1	REG	99.1%	-	-	-	-	-	-	-	1
		LONG	56.4%	33.6%	-	-	-	-	-	-	2
		FUNC	26.5%	18.4%	14.9%	12.8%	9.0%	7.9%	4.6%	-	7
	2	REG	99.6%	-	-	-	-	-	-	-	1
		LONG	100%	-	-	-	-	-	-	-	1
		FUNC	27.8%	19.4%	15.4%	11.1%	7.9%	6.5%	5.0%	-	7
	3	REG	100%	-	-	-	-	-	-	-	1
		LONG	66.4%	33.4%	-	-	-	-	-	-	2
		FUNC	24.0%	18.2%	16.1%	12.1%	8.9%	6.2%	5.1%	-	7
	4	REG	100%	-	-	-	-	-	-	-	1
		LONG	99.0%	-	-	-	-	-	-	-	1
		FUNC	21.6%	20.2%	12.5%	10.8%	9.4%	7.5%	6.4%	4.7%	8
	5	REG	100%	-	-	-	-	-	-	-	1
		LONG	78.0%	21.7%	-	-	-	-	-	-	2
		FUNC	21.2%	19.8%	13.4%	11.6%	11.2%	9.3%	5.8%	-	7

**Groups: (1=Control), (2=Bruxism), (3=Insomnia),(4=Narcolepsy),(5=SDB)*

The figure 1 has 5 columns for each group and 3 rows for functional domain FPCs 1 to 3. For example, the top-left corner of figure 1 consists first estimated eigenfunctions for each subtype: Subtype A1 (Black), Subtype A2 (Red) and Subtype A3 (Blue) in group 1 (Control). The subtypes behaviors are not the same: Subtype A1 has two jumps around 0 and 500 ms, subtype A2 has one peak at about -100 ms and one valley at about 800 ms and subtype A3 has two peaks at around 0 and 1100 ms and one valley at -500 ms. Different pattern are existed for different groups and FPCs.

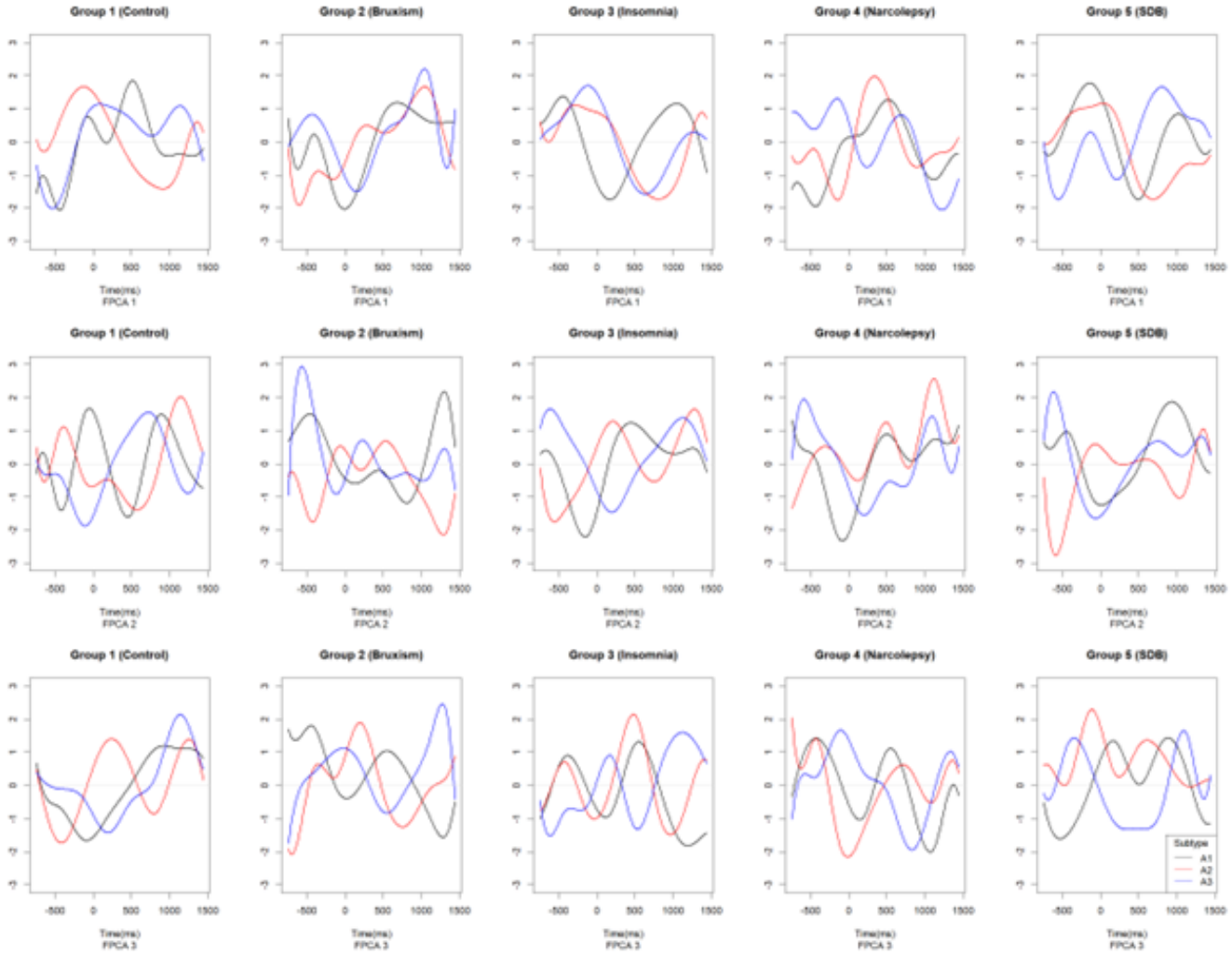

Fig. 1. First three Eigenfunctions for functional dimensions group by five groups and subtype (colors)

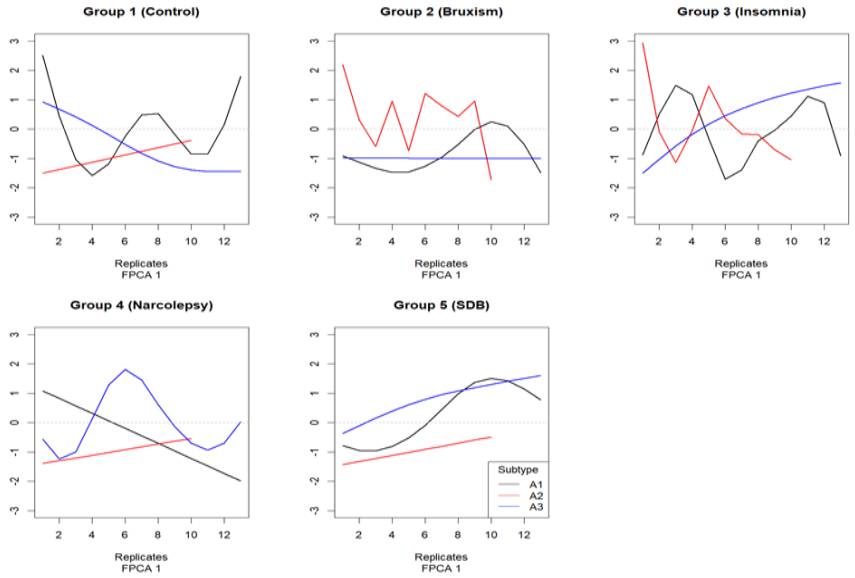

Fig. 2. First Eigenfunctions for longitudinal dimensions group by five groups and subtype (colors)

Figure 2 shows the estimated eigenfunction for longitudinal dimensions for each subtype A1, A2, and A3. The patterns are not the same.

Figure 3 shows the estimated eigenscores for regional dimensions for each subtype A1, A2, and A3. The patterns are not the same between groups and subtypes.

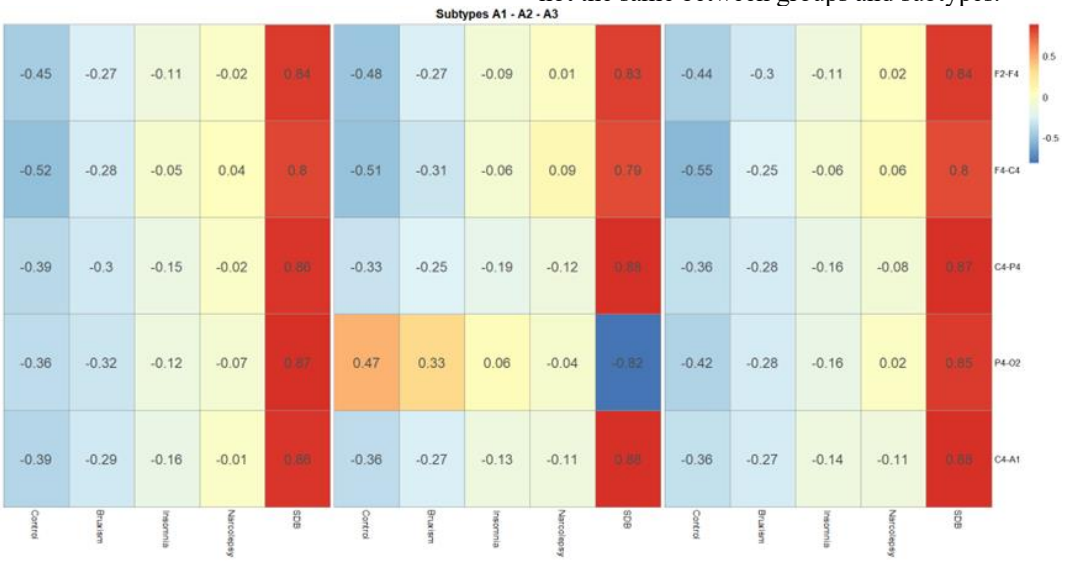

Fig. 3. First Eigenscores for Regional dimensions group by five groups and subtype (Left= Subtype A1, middle = Subtype A2 and right = Subtype A3)

Table 6: The Random Forest with Eigenfunctions results group by subtype, method and train/test.

Sub Type	Method	Dataset	n	Accuracy	Sensitivity					Specificity				
					Groups					Groups				
					1	2	3	4	5	1	2	3	4	5
A1	FPCA	Train	1535	0.895	0.93	0.89	0.87	0.83	0.93	0.98	0.97	0.97	0.96	0.97
		Test	655	0.878	0.95	0.84	0.77	0.87	0.93	0.98	0.97	0.98	0.93	0.97
	HPCA	Train	1535	0.998	1	0.99	1	1	0.99	1	1	0.99	0.99	1
		Test	655	0.996	1	0.98	1	1	1	1	1	1	0.99	1
A2	FPCA	Train	1535	0.872	0.93	0.88	0.77	0.94	0.84	0.96	0.96	0.96	0.97	0.96
		Test	655	0.911	0.92	0.89	0.83	0.96	0.93	0.97	0.97	0.97	0.99	0.97
	HPCA	Train	1535	1	1	1	1	1	1	1	1	1	1	1
		Test	655	1	1	1	1	1	1	1	1	1	1	1
A3	FPCA	Train	1535	0.856	0.91	0.78	0.82	0.91	0.84	0.95	0.96	0.97	0.97	0.95
		Test	655	0.865	0.90	0.83	0.83	0.90	0.83	0.95	0.96	0.99	0.96	0.95
	HPCA	Train	1535	0.997	0.99	1	0.99	0.99	1	0.99	0.99	1	0.99	0.99
		Test	655	0.998	1	0.99	1	1	1	1	0.99	1	1	1

D. The Random Forest with HPCA and FPCA

The result of random forest on training and testing dataset with first to nine HPCA and FPCA eigenfunctions of time domain is presented in the table 6. In all subtypes, the HPCA performs better than FPCA according to the accuracy, sensitivity and specificity.

*Groups: (1=Control), (2=Bruxism), (3=Insomnia), (4=Narcolepsy), (5=SDB)

IV. Conclusion

The random forest with HPCA has the best performance while the random forest without FPCA reached an accuracy of about 40% in both training and testing. The sleep-related EEG dataset is complex and, in this study, only predefined sleep stages and A phase subtypes are considered.

References

1. Morin CM, Jarrin DC. Epidemiology of insomnia: prevalence, course, risk factors, and public health burden. Sleep medicine clinics. 2022;17(2):173-91.
2. Lim DC, Najafi A, Afifi L, Bassetti CL, Buysse DJ, Han F, et al. The need to promote sleep health in public health agendas across the globe. The Lancet Public Health. 2023;8(10):e820-e6.
3. Sarkanen T, Partinen M, Bjorvatn B, Merikanto I, Benedict C, Nadorff MR, et al. Association between hypersomnolence and the COVID-19 pandemic: The International COVID-19 Sleep Study (ICOSS). Sleep medicine. 2023;107:108-15.
4. Mendonça F, Mostafa SS, Morgado-Dias F, Ravelo-Garcia AG, Penzel T. A review of approaches for sleep quality analysis. Ieee Access. 2019;7:24527-46.
5. Mendonça F, Fred A, Mostafa SS, Morgado-Dias F, Ravelo-García AG. Automatic detection of cyclic alternating pattern. Neural Computing and Applications. 2022;34(13):11097-107.
6. Terzano MG, Parrino L. Origin and significance of the cyclic alternating pattern (CAP). Sleep medicine reviews. 2000;4(1):101-23.
7. Migueis D, Lopes M, Ignacio P, Thuler L, Araujo-Melo M, Spruyt K, et al. A systematic review and meta-analysis of the cyclic alternating pattern across the lifespan. Sleep Medicine. 2021;85:25-37.
8. Carra M, Rompré P, Kato T, Parrino L, Terzano MG, Lavigne G, et al. Sleep bruxism and sleep arousal: an experimental challenge to assess the role of cyclic alternating pattern. Journal of oral rehabilitation. 2011;38(9):635-42.
9. Chouvarda I, Mendez MO, Rosso V, Bianchi AM, Parrino L, Grassi A, et al. Cyclic alternating patterns in normal sleep and insomnia: structure and content differences. IEEE transactions on neural systems and rehabilitation engineering. 2012;20(5):642-52.
10. Poryazova R, Werth E, Parrino L, Terzano MG, Bassetti CL. Cyclic alternating pattern in narcolepsy patients and healthy controls after partial and total sleep deprivation. Clinical neurophysiology. 2011;122(9):1788-93.
11. Thomas RJ, Terzano MG, Parrino L, Weiss JW. Obstructive sleep-disordered breathing with a dominant cyclic alternating pattern—a recognizable polysomnographic variant with practical clinical implications. Sleep. 2004;27(2):229-34.
12. Gnoni V, Drakatos P, Higgins S, Duncan I, Wasserman D, Kabiljo R, et al. Cyclic alternating pattern in obstructive sleep apnea: a preliminary study. Journal of Sleep Research. 2021;30(6):e13350.
13. Fayaz M. The Bibliometric Analysis of EEGLAB software in the Web of Science indexed Articles. Neuroscience Informatics. 2023:100154.
14. Somervail R, Cataldi J, Stephan AM, Siclari F, Iannetti GD. Dusk2Dawn: an EEGLAB plugin for automatic cleaning of whole-night sleep electroencephalogram using Artifact Subspace Reconstruction. Sleep. 2023;46(12):zsad208.
15. Ray LB, Sockeel S, Soon M, Bore A, Myhr A, Stojanoski B, et al. Expert and crowd-sourced validation of an individualized sleep spindle detection method employing complex demodulation and individualized normalization. Frontiers in human neuroscience. 2015;9:507.
16. Cui J, Huang Z, Wu J. Automatic detection of the cyclic alternating pattern of sleep and diagnosis of sleep-related pathologies based on cardiopulmonary resonance indices. Sensors. 2022;22(6):2225.
17. Vergara-Sánchez DL, Calvo H, Moreno-Armendáriz MA. Distributional Representation of Cyclic Alternating Patterns for A-Phase Classification in Sleep EEG. Applied Sciences. 2023;13(18):10299.
18. Terzano MG, Parrino L, Sherieri A, Chervin R, Chokroverty S, Guilleminault C, et al. Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep. Sleep medicine. 2001;2(6):537-54.
19. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. circulation. 2000;101(23):e215-e20.
20. Huang H, Zhang J, Zhu L, Tang J, Lin G, Kong W, et al. EEG-based sleep staging analysis with functional connectivity. Sensors. 2021;21(6):1988.
21. Terzano M, Mancia D, Salati M, Costani G, Decembrino A, Parrino L. The cyclic alternating pattern as a physiologic component of normal NREM sleep. Sleep. 1985;8(2):137-45.
22. Parrino L, Ferri R, Bruni O, Terzano MG. Cyclic alternating pattern (CAP): the marker of sleep instability. Sleep medicine reviews. 2012;16(1):27-45.
23. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods. 2004;134(1):9-21.
24. Pion-Tonachini L, Kreutz-Delgado K, Makeig S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. NeuroImage. 2019;198:181-97.
25. Scheffler A, Telesca D, Li Q, Sugar CA, Distefano C, Jeste S, et al. Hybrid principal components analysis for region-referenced longitudinal functional EEG data. Biostatistics. 2020;21(1):139-57.
26. Fayaz M, Abadi A, Khodakarim S. The Comparison between Visually and Auditory Oddball Tasks in the EEG Experiment with Healthy Subjects. Frontiers in Biomedical Technologies. 2020.
27. Febrero-Bande M, De La Fuente MO. Statistical computing in functional data analysis: The R package fda. usc. Journal of statistical Software. 2012;51:1-28.
28. Fayaz M, Abadi A, Khodakarimd S. Random-Splitting Random Forest with Multiple Mixed-Data Covariates. Journal of Biostatistics and Epidemiology. 2023;9(1):20-34.
29. Fayaz M, Abadi A, Khodakarim S. The Functional Regression With Reconstructed Functions From Hybrid Principal Components Analysis: With EEG-fMRI Application. Statistics, Optimization & Information Computing. 2022;10(3):890-903.