

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук  
Образовательная программа «Прикладная математика и информатика»

УДК 51-76

Отчет об исследовательском проекте на тему:  
Автоматизированное определение стадий сна

**Выполнила:**

студентка группы БПМИ219  
Маркович Анна Александровна

\_\_\_\_\_

(подпись)

\_\_\_\_\_

(дата)

**Принял руководитель проекта:**

Чернышев Всеволод Леонидович  
Кандидат физико-математических наук  
Доцент департамента больших данных  
и информационного поиска  
факультета компьютерных наук НИУ ВШЭ

\_\_\_\_\_

(подпись)

\_\_\_\_\_

(дата)

Москва 2023

# Содержание

<b>Аннотация</b>	<b>3</b>
<b>1 Введение</b>	<b>4</b>
<b>2 Описание метода</b>	<b>5</b>
2.1 Предобработка данных . . . . .	5
2.2 Конструирование признаков . . . . .	5
2.2.1 Группы частот . . . . .	5
2.2.2 Классификация каналов . . . . .	6
2.2.3 PSD и PSD-индексы . . . . .	6
2.2.4 Когерентность, Phase-locking value и индексы . . . . .	7
2.3 Алгоритм . . . . .	7
2.3.1 Идея . . . . .	7
2.3.2 Метод Варда . . . . .	8
2.3.3 Оценка качества кластеризации (Stage Silhouette Coefficient, Stage Calinski– Harabasz Index and Stage Davies–Bouldin Index) . . . . .	9
2.3.4 Описание по шагам . . . . .	9
<b>3 Экспериментальное исследование</b>	<b>10</b>
3.1 Данные . . . . .	10
3.2 Регионы . . . . .	10
3.3 Гиперпараметры . . . . .	11
3.4 Качество кластеризации: подсчитанные признаки . . . . .	12
3.5 Анализ поведения признаков на различных стадиях . . . . .	13
3.6 Результат алгоритма yasa.SleepStaging . . . . .	14
3.7 Сравнение результатов . . . . .	15
<b>4 Выводы</b>	<b>16</b>
<b>5 Заключение</b>	<b>17</b>
<b>Список литературы</b>	<b>19</b>

## Аннотация

Анализ мозговой активности является важной и актуальной темой исследования в области нейронауки и психологии. В последние десятилетия наблюдается усиленный интерес к пониманию функционирования мозга и его взаимодействия с мышлением, поведением и психическими процессами. Одним из важнейших направлений исследований в этой области является изучение различных функциональных состояний по записи электроэнцефалограммы при отсутствии аннотации. В настоящей работе обращается внимание на сомнологию и новый метод обучения без учителя для определения стадий сна по записи ЭЭГ. Чтобы решить проблему определения стадий в непрерывной записи мозговой активности, т.е. нахождения оптимальных точек перехода между различными функциональными состояниями при отсутствии знаний о процессе, В. Л. Чернышев, Е.В. Михайлец и А. М. Разоренова предложили управляемый данными алгоритм, работающий без учителя и основанный на методах иерархической кластеризации (SDA – state detecting algorithm). Это автоматизированный метод, который находит кластеры, соответствующие естественным функциональным состояниям, непрерывным во времени, в данных временного ряда с максимальными различиями между смежными по времени кластерами в заданном пространстве признаков. В данной работе исследуется производительность данного алгоритма при определении стадий сна. Анализируются признаки, извлеченные из временных рядов, поведение алгоритма и чувствительность к схожим функциональным состояниям мозга. Результаты показали хорошую способность алгоритма адаптироваться к данным и выявлять основные ярко выделяющиеся процессы, проходящие в мозге в процессе записи ЭЭГ.

## Ключевые слова

Электроэнцефалограмма, анализ мозговой активности, непрерывность стадий сна во времени, агломеративная кластеризация, обучение без учителя.

# 1 Введение

Развитие алгоритмов определения функциональных состояний по электроэнцефалограмме (ЭЭГ) является активной исследовательской областью. ЭЭГ широко используется для изучения и анализа активности мозга. В прошлом алгоритмы определения стадий сна, медитации, пиков когнитивной деятельности и в целом любой активности по ЭЭГ основывались на ручной классификации, где эксперты самостоятельно анализировали и интерпретировали ЭЭГ-сигналы. Однако, такой подход требует большого объема времени и экспертных навыков, и может быть подвержен субъективности и человеческим ошибкам. С развитием машинного обучения и алгоритмов обработки сигналов ЭЭГ стали применяться автоматические методы. Это позволяет автоматизировать процесс классификации и снизить зависимость от человеческого фактора. В классических экспериментальных условиях записи ЭЭГ сопровождаются аннотациями событий, произошедших во время эксперимента (предъявление стимула, поведенческая реакция, обратная связь). Аннотации используются в качестве временных ориентиров, которые добавляют семантического понимания процессам, зарегистрированным во время записи. Анализ записи ЭЭГ без аннотации гораздо сложнее, но является необходимым во многих областях. Многие сферы неврологии, психологии, медицины постоянно находятся в поиске оптимальных, работающих без учителя методов для исследования самых разных вопросов и проблем. Например, диагностика эпилепсии – неврологического расстройства, которое может стать причиной преждевременной смерти. Недостаточная репрезентация данных об эпилептических припадках и быстро развивающиеся стадии заболевания натолкнули авторов [1] к созданию нового подхода. Также актуальной проблемой является социальная интеграция людей, страдающих такими неврологическими заболеваниями, как боковой амиотрофический склероз (БАС) или острые стадии болезни Альцгеймера. Классификация эмоций с использованием обработки сигналов ЭЭГ [5] потенциально может значительно продвинуть общество в этом вопросе.

Запись ЭЭГ представляет собой измерение активности мозга через распределенные по всей голове или специально выбранным областям электроды. Измерения проводятся с определенной частотой дискретизации (sample rate), и в итоге для каждого канала создается временной ряд. Одним из первых шагов в обработке записи является разделение этих рядов на эпохи фиксированной длины. Качественное выявление фаз сна требует учета связности эпох во времени. Важна роль мощности сигнала на определенных частотах, преобладание некоторых характеристик в группе каналов. Все инструменты для перечисленного выше SDA предоставляет. Задачи этого исследования – посмотреть, как алгоритм пользуется этими

инструментами и сравнить его работу с проверенными методами [9]. Цель, соответственно, – оценить оптимальность, результативность, эффективность данного подхода на данных по сну.

## 2 Описание метода

SDA разработан для поиска функциональных состояний в записях ЭЭГ, то есть для обнаружения точек изменения в пространстве признаков. Целью алгоритма является разделение временных рядов данных на непрерывные отрезки (стадии), которые отличаются по поведению характеристик ЭЭГ, таким образом основываясь на внутренней структуре данных. Рассматривается разнообразие общепринятых характеристик ЭЭГ (PSD, PLV, индекс когерентности) для предоставления подробного феноменологического описания обнаруженного функционального состояния.

### 2.1 Предобработка данных

Сигналы ЭЭГ были приведены к усредненному референсу. Полосовой фильтр 0,9–40 Гц был применен для удаления низкочастотных артефактов (дыхательные движения) и высокочастотной мышечной активности. Был проведен анализ независимых компонент [4], и компоненты ICA (independent component analysis), соответствующие таким артефактам, как моргание глаз, саккады, мышечная активность, сердечные сокращения, реограмма и т.д. были исключены, если таковые имелись.

### 2.2 Конструирование признаков

Все признаки, которые включены в рассмотрение, связаны с частотой сигналов. Они считаются усреднением по группам частот и областям мозга.

#### 2.2.1 Группы частот

Традиционно принято рассматривать мозговую активность, используя временно-частотный анализ. Сигнал таким образом раскладывается на частотный спектр, в котором принято выделять следующие 5 частотных диапазонов: Delta (0.9–4 Hz), Theta (4–8 Hz), Alpha (8–14 Hz), Beta (14–25 Hz) and Gamma (25–40 Hz).

### 2.2.2 Классификация каналов

Каналы объединяют в группы по расположению электродов на скальпе. Основные области, каждая из которых в свою очередь делится на левую, центральную и правую: Frontal, Central, Temporal, Parietal, Occipital. Подробно изображено на рис. 1

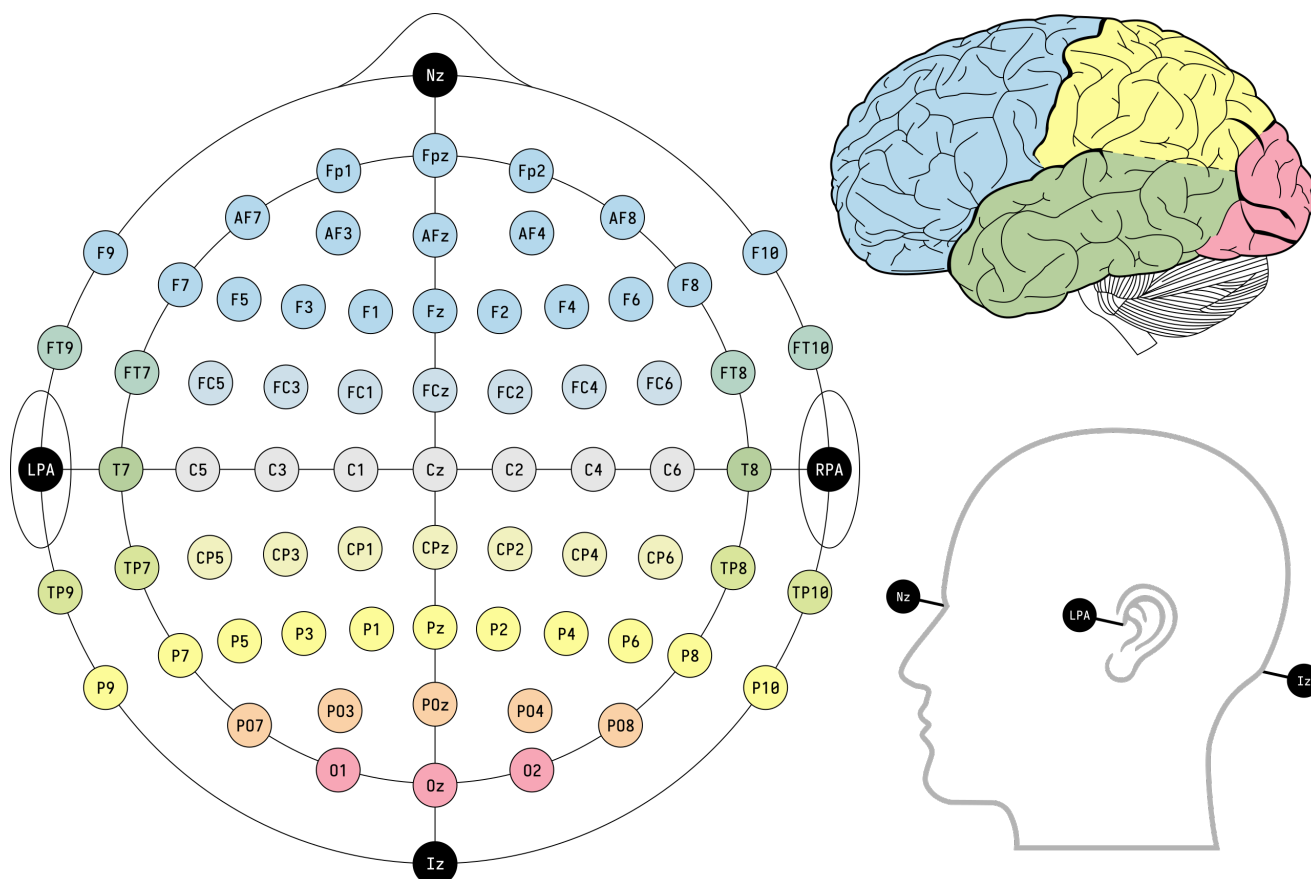


Рис. 1: Положение электродов ЭЭГ в системе 10-20 и связанных с ними долей головного мозга.

### 2.2.3 PSD и PSD-индексы

Power spectral densities для определенных диапазонов частот являются наиболее распространенными характеристиками, используемыми для описания длительных мозговых процессов, например, стадий сна. Power spectral density (спектральная плотность мощности) (по сути мера силы сигнала) вычисляется методом Multitaper [7] на 5 группах частот для каждой группы каналов. Так называемые PSD-индексы – соотношения мощности сигналов на следующих частотах: theta/delta, alpha/delta, alpha/theta, alpha/(delta+theta), beta/delta, beta/theta, beta/alpha, beta/(delta+theta) и некоторых других (всего 16).

### 2.2.4 Когерентность, Phase-locking value и индексы

Когерентность – это мера синхронизации между двумя сигналами ЭЭГ, которая оценивает сходство их спектральных характеристик по частоте. Для сигналов  $x$  и  $y$  для конкретной эпохи и частотного диапазона когерентность варьируется от 0 до 1 и считается как:

$$Coh_{xy} = \frac{|E[S_{xy}]|}{\sqrt{E[S_{xx}]E[S_{yy}]}} \quad (1)$$

где  $S_{xy}$  – CSD между сигналами,  $S_{xx}$  и  $S_{yy}$  – PSD, а  $E$  – среднее арифметическое для 5 соседних эпох с центром в данной.

Phase-locking value – мера фазовой синхронизации сигналов. Для сигналов  $x$  и  $y$  и конкретной эпохи и частотной группы считается следующим образом:

$$PLV_{xy} = \left| E \left[ \frac{S_{xy}}{|S_{xy}|} \right] \right| \quad (2)$$

где  $S_{xy}$  – CSD между сигналами,  $E$  – среднее арифметическое для 5 соседних эпох с центром в данной.

Индекс когерентности и индекс PLV канала  $x$  с порогом  $P$  – это количество каналов  $y, y \neq x$ , где  $Coh_{xy} \geq P$  и  $PLV_{xy} \geq P$  соответственно, для  $P \in \{0.6, 0.7, 0.8\}$ . В процессе кластеризации используются именно индексы.

## 2.3 Алгоритм

### 2.3.1 Идея

Классическая кластеризация не учитывает непрерывность стадий во времени, поэтому была введена дополнительная структура связного неориентированного графа в пространстве эпох (пары из эпох с расстоянием, не превышающим  $k$  секунд для заданного гиперпараметра  $k$ , соединены ребрами). Таким образом, формируются компактные во временном измерении кластеры и их границы являются потенциальными точками перехода между стадиями сна. Процесс кластеризации запускается с различными значениями гиперпараметров (количество кластеров, расстояние  $k$  секунд между эпохами, минимальная длина стадии и т.д.). На каждой итерации из полученных границ кластеров формируются новые путем склеивания слишком маленьких кластеров с ближайшим соседом, а затем объединяются стадии, кластерное расстояние между которыми меньше порога ( $0.3 \times W_{avg}$ , где  $W_{avg}$  – среднее расстояние Варда между всеми парами соседних стадий). В итоге для каждой тройки  $(N_{max}, K_{max}, L)$  получа-

ется разбиение пространства эпох на непрерывные сегменты, границы которых объединяются в один массив, сортируются и кластеризуются методом к-средних. Центры полученных кластеров являются точками с наилучшей способностью разделять потенциальные стадии в пространстве признаков, и можно принять их за итоговые границы стадий. Оптимальное количество стадий выбирается на основе показателей качества кластеризации, значения среднего расстояния между этапами и условий поставленной задачи.

### 2.3.2 Метод Варда

Метод Варда [6] – это метод агломеративной иерархической кластеризации, который минимизирует дисперсию внутри кластеров. Он работает по следующему принципу. Изначально каждый элемент образует отдельный кластер. На каждом шаге два кластера с наименьшим расстоянием между отделениями объединяются в один. Процесс продолжается до тех пор, пока не будет достигнуто указанное количество кластеров или не будет превышен порог расстояния между кластерами. Расстояние между двумя кластерами вычисляется как увеличение “суммы квадратов ошибок” ( $ESS$  – error sum of squares) после объединения двух кластеров в один кластер, где  $ESS$  кластера – это сумма квадратов отклонений от центра кластера (средний вектор). После математических преобразований расстояние между кластерами  $X$  и  $Y$  выражается формулой:

$$D(X, Y) = ESS(X \cup Y) - (ESS(X) + ESS(Y)) = \frac{N_X * N_Y}{N_X + N_Y} \|\bar{m}_X - \bar{m}_Y\|^2, \quad (3)$$

где  $N_X$  и  $N_Y$  – размеры,  $\bar{m}_X$  и  $\bar{m}_Y$  – центры кластеров. Центроидное расстояние между двумя кластерами – это евклидово расстояние между центрами этих кластеров. По сути, метрика Варда – это стоимость объединения двух кластеров, которая измеряет, насколько увеличится сумма квадратов при их объединении. Фактически, расстояние Варда – это взвешенное расстояние в квадрате от центра кластера, где весовой коэффициент зависит от размера кластера. Таким образом, расстояние Варда чувствительно к размерам кластеров, что полезно для получения кластеров, сбалансированных по размеру, в процессе кластеризации. Использование матрицы связности в методе Варда позволяет нам получать кластеры, содержащие близкие по времени эпохи, что требуется для нахождения стадий сна. На каждом шаге расстояние между кластерами вычисляется только в том случае, если между их элементами есть ребра.



### 2.3.3 Оценка качества кластеризации (Stage Silhouette Coefficient, Stage Calinski–Harabasz Index and Stage Davies–Bouldin Index)

В данном алгоритме используются три основные меры качества кластеризации, которые не требуют заранее известной истинной разметки данных. Это, во-первых, коэффициент Силуэта [8] (Silhouette Coefficient), измеряющий, насколько объекты похожи на свой собственный кластер (когезия) по сравнению с другими кластерами в среднем. Он варьируется от -1 для неправильной кластеризации до 1 для кластеризации с высокой плотностью. Во-вторых, индекс Калински–Харабаша [2] (Calinski–Harabasz Index), также известный как критерий отношения дисперсии, характеризующий отношение дисперсии между кластерами к дисперсии внутри кластеров. Он принимает положительные значения и выше, когда кластеры плотные и хорошо разделены. И, наконец, индекс Дэвиса–Боулдина [3] (Davies–Bouldin Index), который измеряет разделение между кластерами, сравнивая расстояние между кластерами с размером самих кластеров. Ноль – это наименьшее возможное значение, чем оно меньше, тем лучше распределение. Качественное разбиение пространства эпох на функциональные состояния означает хорошо отделенные друг от друга пары соседних стадий. Поэтому значения описанных метрик считаются на всех парах смежных кластеров, как на наборах данных, состоящих из двух кластеров. Затем среднее значение берется за результат.

### 2.3.4 Описание по шагам

Повтор следующих шагов для каждой тройки параметров ( $N$ ,  $K$ ,  $L$ ):

- 1 Итеративный поиск потенциального разделения на  $N$  кластеров с матрицей смежности с расстоянием в  $K$  эпох.
- 2 Массив границ кластеров сортируется по возрастанию номеров эпох, формирует разделение пространства эпох на непрерывные сегменты (стадии).
- 3 Каждый сегмент, по длительности не превышающий  $L$ , объединяется с одним из соседей по метрике Варда.
- 4 Процесс объединения соседних сегментов продолжается до тех пор, пока минимальное расстояние не превысит долю  $W$  от его среднего значения для всех пар соседних сегментов.

Далее на основе полученных результатов выбирается финальное разделение на фазы сна. Следующие шаги повторяются для каждого подмножества  $N_{max} \times K_{max} \times L$ :

- 1 Для каждой тройки ( $N$ ,  $K$ ,  $L$ ) соответствующие массивы границ кластеров объединяются в один и итоговый массив сортируется по возрастанию.
- 2 Данный массив кластеризуется с помощью метода K-Means для всех значений  $N_{KM}$ .
- 3 Для каждой итерации предыдущего шага выбираются 3 вида центров кластеров (среднее арифметическое, медиана, мода) в качестве претендентов для итогового разграничения фаз.

Потом на основе метрики Силуэта выбирается наилучший массив границ стадий. Выбор метрики связан с тем, что она учитывает все элементы всех пар кластеров, причем неправильно выбранные элементы вносят отрицательный вклад. В итоге получены  $N_{KM}$  вариантов разметки стадий, а итоговое разделение выбирается на основе трех упомянутых выше метрик качества кластеризации, усредненного расстояния Варда и центроидного расстояния между кластерами.

## 3 Экспериментальное исследование

### 3.1 Данные

В данном исследовании алгоритм применяется к записи ЭЭГ длительностью 3,5 часа. Испытуемый – мужчина, возраст 52 года, без психических отклонений и нарушений сна. Запись производилась на аппарате Nihon Kohden с аппаратным фильтром на 50 Гц с частотой дискретизации 1000 Гц, монтаж 20-10. Доступная аннотация: "Первые 8 минут записи – это бодрствование с закрытыми глазами. Затем начался сон, засыпание (1 стадия) произошло почти сразу, но потом в середине записи был длинный промежуток без сна, но с закрытыми глазами. Ближе к последней трети записи наступил глубокий сон (3 стадия), это видно по медленным волнам на ЭЭГ."

### 3.2 Регионы

В изучаемой записи обнаружено 70 ЭЭГ-каналов, которые были сгруппированы по 16 областям мозга. На рисунке 2 изображено их распределение по скальпу.

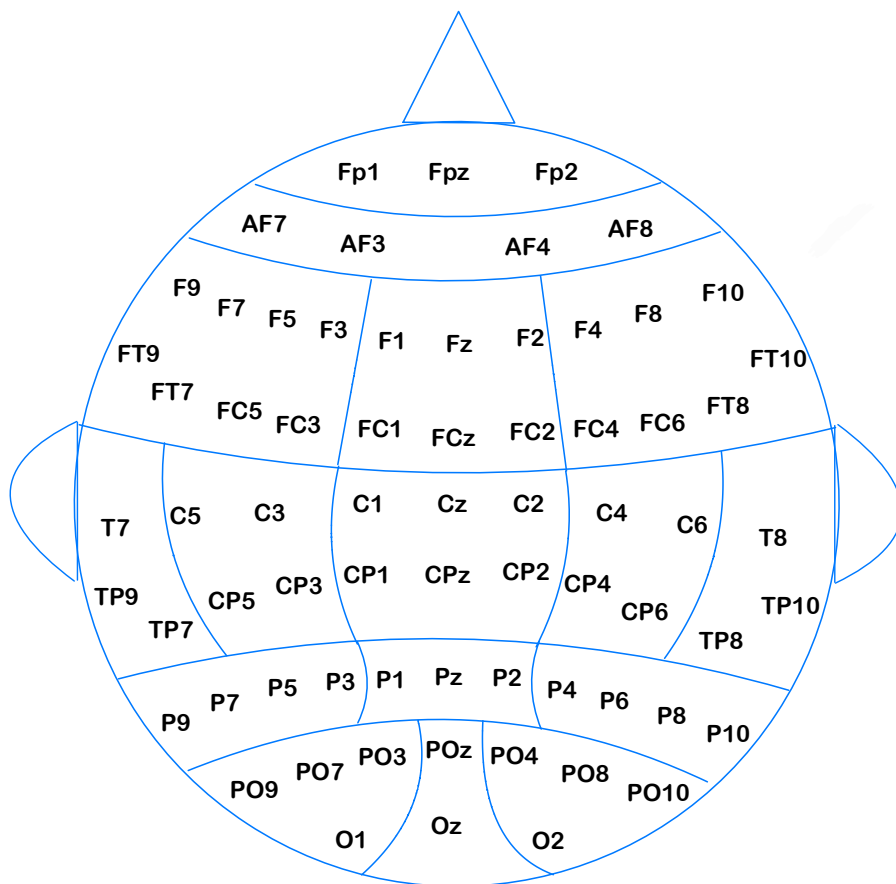


Рис. 2: Расположение сенсоров и разделение на группы.

### 3.3 Гиперпараметры

Из-за большого объема данных было решено разбить запись на эпохи длительностью 5 секунд. Для кластеризации были эмпирически выделены граничные значения параметров (таб. 1).

Таблица 1: Парамерты в алгоритме кластеризации

Имя	Значение	Описание
$N$	[2, 20]	Количество кластеров в методе Варда
$K$	[4, 10]	Максимальное расстояние между эпохами для матрицы смежности (в эпохах)
$L$	{25, 30, 35}	Минимальная длина стадии (в эпохах) при объединении смежных сегментов
$W$	0.3	Коэффициент Варда при объединении смежных сегментов 0.3
$N_{max}$	{10, 15, 20}	Максимальное количество кластеров в массиве границ стадий
$K_{max}$	{7, 8, 9, 10}	Максимальное расстояние между эпохами для массива границ стадий
$N_{KM}$	[2, 12]	Количество кластеров в массиве границ стадий для метода к-средних
Тип центра	среднее, мода, медиана	Типы центров кластеров

### 3.4 Качество кластеризации: подсчитанные признаки

Для оценки качества разбиения на этапы всего набора данных используются показатели, усредненные по всем парам соседних стадий – средние расстояния между фазами и показатели качества фазы. В этом исследовании это Silhouette, Calinski–Harabasz и Davies–Bouldin coefficients. Среди подсчитанных значений для вариантов разбиения на 3-13 кластеров, показатели для 6 и 9 стадий сильно выделяются.

Таблица 2: Показатели метрик качества кластеризации

N	center	Ward dist.	Centroid dist.	Silhouette coef.	Calinski–Harabasz	Davies–Bouldin
6	median	116309.14	24.79	0.26	261.59	1.75
9	median	66475.69	29.14	0.32	147.02	1.57

### 3.5 Анализ поведения признаков на различных стадиях

Рис. 3 иллюстрирует различия в поведении признаков между 6 выявленными стадиями сна. Для каждой пары соседних стадий мы можем наблюдать различную динамику признаков, предоставляемых алгоритму (PSD, индекс когерентности, индекс PLV), в разных диапазонах частот. На рисунке видно, как эти различия в поведении признаков были зафиксированы алгоритмом.

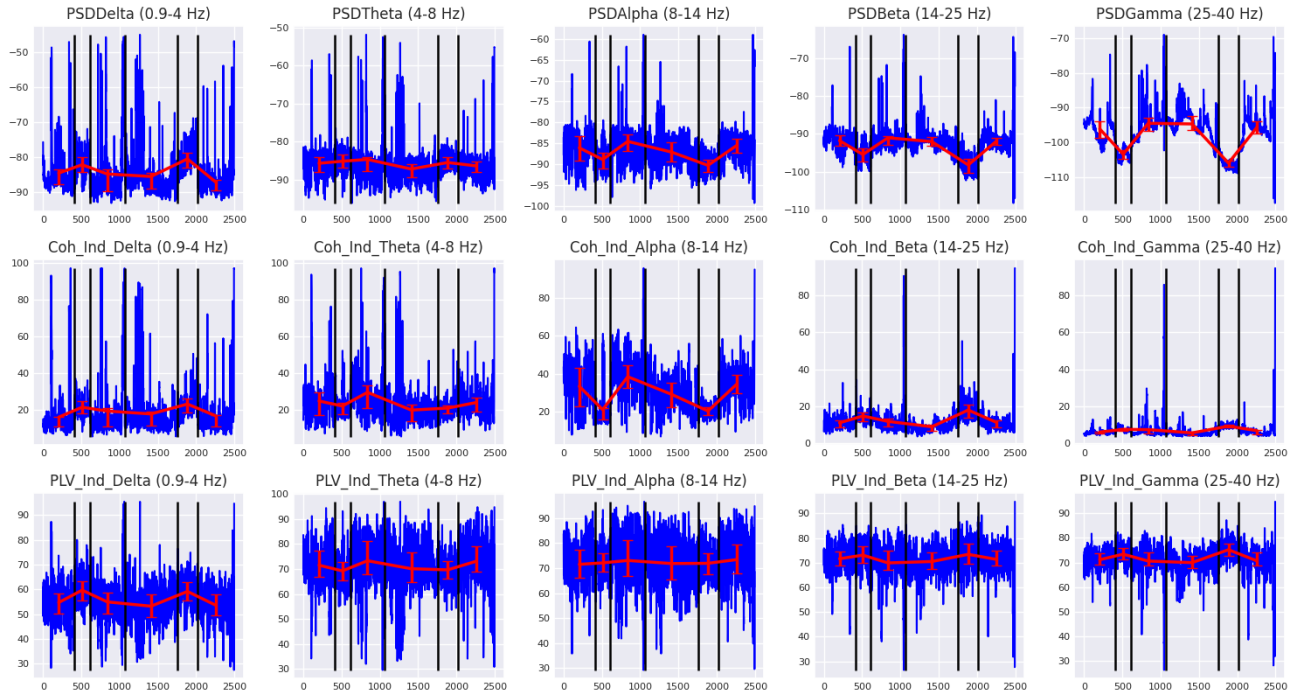


Рис. 3: Горизонтальная ось обозначает индекс эпохи вдоль записи. Вертикальные черные линии – это границы стадий. Синими линиями обозначена динамика характеристик ЭЭГ на протяжении записи; это показатели PSDs (1-я строка), когерентности и PLV с пороговым значением 0,7 (2-я и 3-я строки соответственно) в 5 частотных диапазонах, усредненные по всем 70 каналам. Красные линии представляют медиану  $\pm$  интерквартильный размах для выбранного объекта на каждом этапе.

Точечный график на рис. 4 иллюстрирует стабильность полученных границ 6 стадий (плотные и хорошо разделенные кластеры потенциальных границ, в основном сосредоточенные в одной точке). Подсчитанные оценки качества разделения стадий, проиллюстрированные на графике, показывают свою информативную ценность.

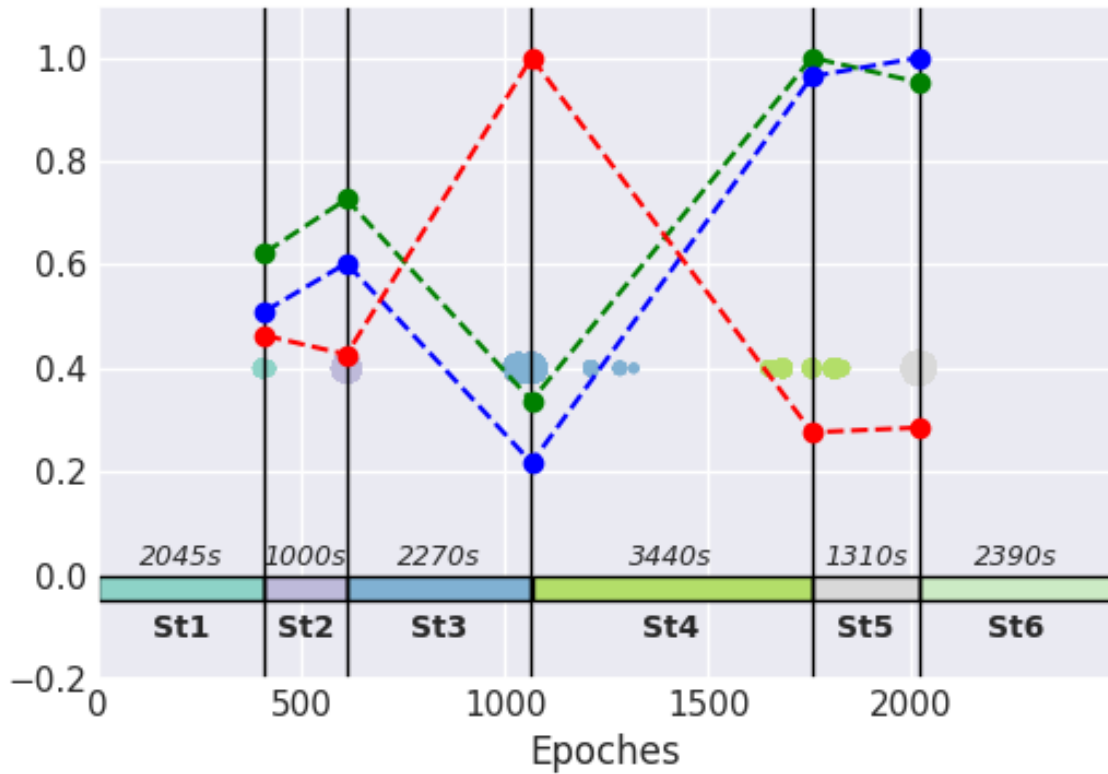


Рис. 4: Горизонтальная ось обозначает индекс эпохи вдоль записи. Вертикальные черные линии – это границы фазы, полученные в виде центров соответствующих кластеров. Точечная диаграмма (на уровне 0.4) представляет собой объединенный массив потенциальных границ стадий; Отображается идентификатор стадии и общая продолжительность этапа (в секундах). Вертикальная ось обозначает стандартизованные путём деления на максимальное значение показатели качества кластеризации (значения в диапазоне  $[0, 1]$ ). Пунктирные линии показывают показатели качества кластеризации, рассчитанные для соседних этапов (центроидное расстояние – зеленая линия, оценка Silhouette – синяя линия, оценка Calinski–Harabasz – оранжевая линия, оценка Davies–Bouldin – красная линия).

### 3.6 Результат алгоритма `yasa.SleepStaging`

Библиотека `yasa` [9] предоставляет мощный инструмент предсказания фаз сна в записи ЭЭГ. Основные стадии, циклически повторяющиеся на протяжении всего сна: N1, N2, N3 – засыпание, медленный и глубокий сон, R (REM – rapid eye movement) – быстрый сон, W – бодрствование. Библиотека `yasa` характеризует каждую эпоху набором вероятностей принадлежности к определенной стадии. На рисунке 5 можно наблюдать распределение фаз по всей записи. Цветами выделены стадии, высота «пику» отвечает за вероятность.

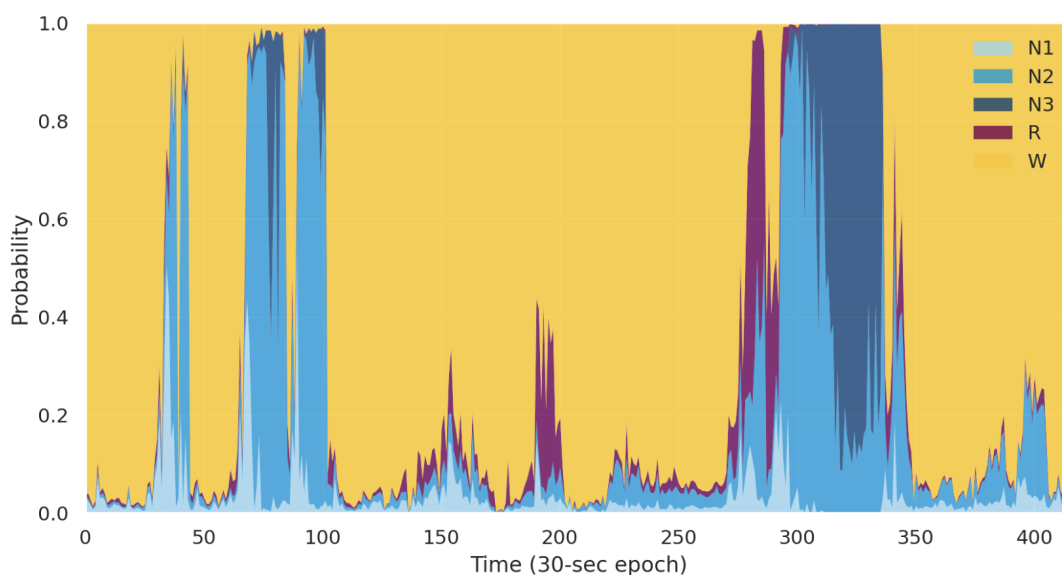


Рис. 5: График предсказаний стадий сна от yasa

### 3.7 Сравнение результатов

На графике 6 проиллюстрировано «наложение» разметок по версии yasa и SDA.

В первом случае алгоритм нашел 6 стадий. Глобально можно наблюдать, как выделены в отдельные стадии сон (синий цвет) и бодрствование (желтый цвет). В 5 фазе из 6 объединены медленный и глубокий сон. Они сильно похожи по своим параметрам, в сравнении с остальными фазами, и при этом один переходит в другой. Важно отметить, что в разметке yasa ближе к границе N2 и N3 вероятности понижаются, аналогично ситуации в 1 стадии, где «вкрапление» медленного сна присутствует с меньшей вероятностью, чем бодрствование. То есть получается, что SDA плохо определяет стадии там, где вероятность yasa снижается. В разбиении на 6 стадий SDA объединяет длительную фазу бодрствования с быстрым сном, который соответствует общей возбужденности нервной системы в этой фазе, то есть по уровню сопоставим с бодрствованием.

В обоих случаях (6 и 9 стадий) бодрствование разваливается на две большие стадии. Если обратить внимание на рисунок 5, можно заметить, что примерно в той области, где происходит разделение (4-ая стадия в разметке SDA для 9 стадий), у yasa имеется «всплеск» быстрого сна. SDA посчитал это важным.

При увеличении количества стадий до 9, разметка yasa совпадает с SDA для N2, N3 и даже быстрого сна. Да, границы чуть смещены, но глобально разметки yasa и sda совпадают.

Стоит отметить, что чем больше стадий, тем более тонкие различия выявляет алгоритм.

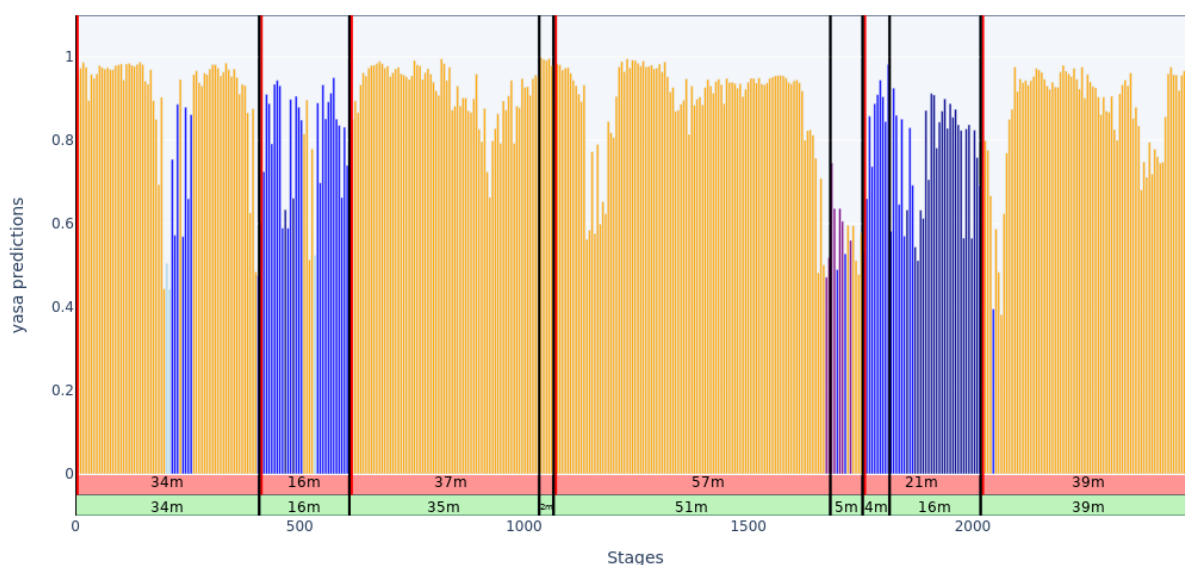


Рис. 6: 7 стадий SDA. Гистограмма отвечает за предсказания стадий сна от yasa. Желтый цвет – бодрствование, светло-голубой – засыпание, синий – медленный сон, темно-синий – глубокий сон, фиолетовый – быстрый сон. Вертикальные линии – границы стадий по SDA. Красные линии и полоса внизу – разметка 6 стадий, черные линии и зеленая полоса – разметка 9 стадий. В прямоугольниках внизу указана длительность стадии в минутах.

## 4 Выводы

SDA разработан таким образом, чтобы легко подстраиваться структуру данных, на которых нужно найти условно стационарные участки записи. В данном случае на вход подаются нейрофизиологические признаки. Также учитывается пространственное распределение электрической активности, регистрируемой на поверхности скальпа. Это включает оценку коэффициента когерентности или корреляции между различными электродами, а также учет амплитуды и фазы сигналов в различных областях головы. Алгоритм способствует оценке связей и функциональной связности между различными областями мозга. Это включает анализ фазовой синхронизации или взаимного влияния между различными частотными компонентами и мозговыми структурами. Метод чувствителен к наличию артефактов, таких как электромиографические сигналы (связанные с мышечной активностью) или электрокардиографические сигналы (связанные с сердечной активностью). Необходимо учитывать возможность идентификации и удаления таких артефактов, чтобы получить данные, связанные исключительно с активностью мозга.

На опытных данных по сну алгоритм показал вполне хорошие результаты. С тем



небольшим описанием, которое было предоставлено вместе с записью, результаты полностью согласуются. Ясно, что выбранные характеристики сигналов и связей между ними хорошо передают временную динамику зарегистрированной мозговой активности в течение сна и бодрствования. SDA обладает высокой разделяющей способностью – используются адаптированные к задаче лучшие метрики качества кластеризации.

Хотя как с засыпанием, так и с быстрым сном возникли несовпадения. Причин у ошибок с определением четких границ или с выявлением коротких, но важных стадий, может быть много. Например, неконсистентность данных или контекстуальные факторы. Алгоритм может быть чувствителен к особенностям конкретного набора данных или контекста, в котором был разработан. При применении на других наборах данных или в различных условиях его эффективность может быть снижена. Может быть, ключевую роль играют возраст, пол, здоровье или индивидуальные особенности, которые могут влиять на стадии сна. Это может приводить к неправильной классификации для определенных групп или условий.

## 5 Заключение

Исследования активности мозга и анализ ЭЭГ позволяют получить информацию о мозговых процессах, связях между различными областями мозга и взаимодействии мозга с окружающей средой. Это открывает широкий спектр возможностей в таких областях, как нейронаука, медицина, нейротехнологии и психология.

В будущем, с постоянным развитием технологий, мы можем ожидать появления новых методов и приборов для более точного и детализированного изучения активности мозга. Это может включать разработку более чувствительных и точных устройств для записи ЭЭГ, улучшение методов анализа данных и разработку алгоритмов машинного обучения для более эффективной классификации и интерпретации ЭЭГ-сигналов.

Потенциальные полезные применения исследования активности мозга и анализа ЭЭГ включают: диагностику и мониторинг неврологических и психических расстройств, таких как эпилепсия, бессонница, шизофрения и депрессия; разработку интерфейсов мозг-компьютер для помощи людям с ограниченными возможностями и улучшения взаимодействия между человеком и компьютером; улучшение понимания мозговых механизмов восприятия, памяти, внимания и принятия решений. Оптимизацию обучения и тренировки, включая изучение эффективности различных методов обучения и тренировки мозга. Исследование эмоциональных и когнитивных состояний человека и их влияния на поведение и здоровье.

В итоге, будущее исследований активности мозга и анализа ЭЭГ представляет огром-

ный потенциал для расширения наших знаний о мозге, создания новых технологий и приложений, а также улучшения нашего понимания и взаимодействия с миром. Это захватывающее путешествие, которое продолжает привлекать ученых и исследователей, и мы можем ожидать, что в будущем мы узнаем еще больше полезного и удивительного об этом сложном и загадочном органе – нашем мозге.

## Список литературы

- [1] Javad Birjandtalab, Mazyar Baran Pouyan и Mehrdad Nourani. *Unsupervised EEG analysis for automated epileptic seizure detection*. Под ред. Xudong Jiang, Guojian Chen, Genci Capi и Chiharu Ishii. Июль 2016. DOI: [10.1117/12.2243622](https://doi.org/10.1117/12.2243622).
- [2] Tadeusz Caliński и Joachim Harabasz. “A dendrite method for cluster analysis”. В: *Communications in Statistics-theory and Methods* 3 (1974), с. 1—27.
- [3] David L. Davies и Donald W. Bouldin. “A Cluster Separation Measure”. В: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1.2 (1979), с. 224—227. DOI: [10.1109/TPAMI.1979.4766909](https://doi.org/10.1109/TPAMI.1979.4766909).
- [4] Amira Ectiouei, Wassim Zouch, Mohamed Ghorbel, Mohamed Ben Slima, Ahmed Ben Hamida и Chokri Mhiri. “Automated EEG Artifact Detection Using Independent Component Analysis”. В: *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*. 2020, с. 1—5. DOI: [10.1109/ATSIP49331.2020.9231574](https://doi.org/10.1109/ATSIP49331.2020.9231574).
- [5] Elfadel I.M. Gonzalez H.A. Yoo J. *EEG-based Emotion Detection Using Unsupervised Transfer Learning*. Июль 2019. DOI: [10.1109/EMBC.2019.8857248](https://doi.org/10.1109/EMBC.2019.8857248).
- [6] Fionn Murtagh и Pierre Legendre. “Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion?” В: *Journal of Classification* 31.3 (окт. 2014), с. 274—295. DOI: [10.1007/s00357-014-9161-z](https://doi.org/10.1007/s00357-014-9161-z). URL: <https://doi.org/10.1007/s00357-014-9161-z>.
- [7] D. B. Percival и A. T. Walden. “Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques.” В: Cambridge University Press., 1993.
- [8] Peter J. Rousseeuw. “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. В: *Journal of Computational and Applied Mathematics* 20 (1987), с. 53—65. ISSN: 0377-0427. DOI: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7). URL: <https://www.sciencedirect.com/science/article/pii/0377042787901257>.
- [9] YASA. *An open-source, high-performance tool for automated sleep staging*. URL: <https://raphaelvallat.com/yasa/build/html/generated/yasa.SleepStaging.html> (дата обр. 21.05.2023).