

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра физики открытых систем

**Анализ данных полисомнограмм с использованием языка
программирования Python**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студента 4 курса 4041 группы
направления 09.03.02 «Информационные системы и технологии»
код и наименование направления
института физики
наименование факультета, института, колледжа
Яковлева Вадима Денисовича
фамилия, имя, отчество

Научный руководитель

доцент, д.ф.-м.н.,

должность, ученая степень, уч. звание

подпись, дата

М.О. Журавлев

Инициалы Фамилия

Заведующий кафедрой физики открытых систем

полное наименование кафедры

профессор, д.ф.-м.н.,

должность, ученая степень, уч. звание

подпись, дата

А.А. Короновский

Инициалы Фамилия

Саратов 2025 г.

ВВЕДЕНИЕ

Работа посвящена актуальной проблеме автоматизации анализа данных полисомнографии (ПСГ) – "золотого стандарта" диагностики нарушений сна. Ручное стадирование сна (определение стадий: N1, N2, N3, REM, Wake) по данным электроэнцефалограммы (ЭЭГ), электроокулограммы (ЭОГ) и электромиограммы (ЭМГ) является крайне трудоемким и субъективным процессом. Цель работы: разработка и реализация на языке Python алгоритма автоматического стадирования сна, оценка его точности на данных здоровых пациентов и пациентов с синдромом обструктивного апноэ сна (СОАС), анализ ограничений метода. В работе использованы открытые базы данных ПСГ-записей и современные библиотеки Python для обработки сигналов и машинного обучения.

Целью данной выпускной квалификационной работы Целью данной выпускной квалификационной работы являлась разработка, реализация и оценка эффективности алгоритма автоматического стадирования сна на основе анализа данных полисомнограмм (ПСГ) с использованием языка программирования Python.

Материалы исследования. Основу исследовательской выборки составили 50 полных ночных записей полисомнограмм (ПСГ) с соответствующими им экспертными гипнограммами. Записи были получены от 25 уникальных пациентов, причем каждый пациент представлен двумя последовательными ночными записями. Такая структура выборки позволяет учитывать потенциальное влияние эффекта "первой ночи" (возможные искажения в данных из-за непривычной обстановки лаборатории сна) и повышает надежность анализа за счет большего объема данных на одного индивидуума

Выпускная квалификационная работа содержит введение, три раздела теоретической части ВКР (1. Формулировка проблемы стадирования сна; 2. Способы стадирования сна: ручной и автоматический; 3. Физиологические принципы стадирования и способы обработки данных: частотные ритмы,

методы анализа сигналов, статистические характеристики), четыре раздела практической части ВКР (4. Программная реализация алгоритма стадирования; 5. Сравнение полученных программой результатов для здоровых пациентов; 6. Сравнение результатов для записей с апноэ, 7. Ограничения методов автоматической разметки стадий сна), заключение и список использованных источников. Общий объем работы 42 стр.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

1. Формулировка проблемы стадирования сна. Сон, который занимает примерно треть жизни человека, является жизненно важным биологическим процессом. Сон состоит из циклически повторяющихся фаз — медленного сна (NREM), включающего стадии N1, N2, N3 и N4, и быстрого сна (REM), при этом один цикл длится около 90 минут [1]. Традиционно классификация стадий сна проводится вручную экспертами на основе данных полисомнографии (ПСГ), включающей электроэнцефалограмму (ЭЭГ), электроокулограмму (ЭОГ) и электромиограмму (ЭМГ), что является трудоёмким процессом и подвержено субъективным ошибкам. В связи с этим современные исследования сосредоточены на разработке методов автоматической классификации стадий сна с использованием машинного обучения: от традиционных методов, основанных на извлечении признаков, включая вейвлет-анализ, до глубокого обучения с применением сверточных нейронных сетей (CNN) для анализа сигналов и рекуррентных нейронных сетей для учёта временных зависимостей. Точная классификация стадий сна имеет важное клиническое значение для диагностики таких нарушений, как синдром обструктивного апноэ сна, характеризующийся остановками дыхания, храпом и фрагментацией сна, что повышает риски гипертонии, инсульта и когнитивных нарушений, а также бессонницы. Кроме того, она необходима для мониторинга терапии и проведения исследований, изучающих взаимосвязь сна с неврологическими заболеваниями.

2. Способы стадирования сна: ручной и автоматический.

Классификация стадий сна традиционно выполняется экспертами посредством визуального анализа 30-секундных эпох полисомнограмм по критериям AASM. Специалисты определяют ключевые паттерны, такие как сонные веретена и К-комплексы для стадии N2, быстрые движения глаз и низкий мышечный тонус для REM и другие. Этот метод учитывает контекст, артефакты и индивидуальные особенности, а также позволяет выявлять атипичные и патологические паттерны. Однако он отличается высокой трудоёмкостью, подвержен субъективности и вариабельности между экспертами, а также зависит от опыта специалиста. В ответ на эти ограничения разрабатываются методы автоматического стадирования с использованием машинного обучения, которые делятся на два основных подхода [2]. Первый — основанный на извлечении признаков, включая спектральные, временные и статистические характеристики, с последующей классификацией традиционными алгоритмами, такими как SVM и Random Forest; эффективность этого подхода зависит от качества выбранных признаков. Второй — глубокое обучение, при котором модели, например сверточные (CNN) и рекуррентные сети (LSTM, GRU), обучаются напрямую на сырых сигналах полисомнографии, достигая высокой точности, близкой к экспертной. Автоматизация позволяет значительно ускорить обработку больших объёмов данных, обеспечивает воспроизводимость и согласованность результатов, а также масштабируемость, что особенно важно для клиник с ограниченными ресурсами.

3. Физиологические принципы стадирования и способы обработки данных: частотные ритмы, методы анализа сигналов, статистические характеристики. Классификация стадий сна основывается на характерных паттернах электроэнцефалограммы. Во время бодрствования доминируют альфа-ритм (8-13 Гц) и бета-ритм (13-30 Гц), в стадии легкого сна N1 появляется тета-ритм (4-7 Гц) и исчезает альфа-активность. Стадия N2 характеризуется наличием сонных веретён (12-14 Гц) и К-комплексов, а

глубокий сон N3 сопровождается высокоамплитудным дельта-ритмом (0,5-4 Гц). REM-сон отличается низкоамплитудной смешанной активностью и «пилообразными» волнами. Архитектура сна циклична с периодом около 90 минут, при этом интенсивность стадий меняется: NREM-сон экспоненциально снижается, а REM-сон линейно возрастает в последующих циклах [3]. Для количественного анализа ЭЭГ-паттернов применяются методы спектрального анализа, включая классическую периодограмму, которая оценивает спектральную плотность мощности (СПМ) через дискретное преобразование Фурье, но имеет недостаток высокой дисперсии из-за отсутствия усреднения. Метод Бартлетта улучшает оценку, разделяя сигнал на непересекающиеся сегменты и усредняя их периодограммы, что снижает дисперсию, однако ухудшает частотное разрешение. Метод Уэлча, являющийся усовершенствованием метода Бартлетта, использует перекрывающиеся сегменты и весовые окна, обеспечивая оптимальный компромисс между снижением дисперсии и сохранением разрешения. Для комплексного анализа полисомнографических данных применяются статистические характеристики сигналов, такие как среднее, дисперсия, стандартное отклонение и амплитуда, а также спектральные метрики — относительная мощность ритмов, доминирующая частота и спектральная энтропия [4]. Для задач машинного обучения формируются признаки, представляющие собой комбинации статистических и спектральных параметров, которые используются в классификаторах, таких как SVM и Random Forest.

4. Программная реализация алгоритма стадирования.

Для работы алгоритма использовались данные в формате EDF(+), представляющем собой открытый стандарт для хранения биомедицинских сигналов, включающий заголовок с метаданными (идентификация пациента, технические параметры) и данные сигналов в виде 16-битных целых чисел [5]. В исследовании анализировались 50 записей полисомнографии от 25 пациентов (15 здоровых и 10 с обструктивным апноэ сна), которые были разделены в пропорции 70/30 на обучающую и тестовую выборку. Каждая

запись содержала многоканальные данные ЭЭГ, ЭОГ и ЭМГ. Реализация алгоритма выполнена с использованием научного стека Python, включающего специализированную библиотеку MNE для работы с нейрофизиологическими данными, NumPy для операций с многомерными массивами, SciPy для цифровой обработки сигналов и математических функций, scikit-learn для машинного обучения, Matplotlib для визуализации результатов и random для генерации псевдослучайных чисел. В алгоритме реализованы ключевые функции: `hjorth_parameters`, вычисляющая параметры активности, мобильности и сложности сигнала на основе статистики временного ряда и его производных, что особенно полезно для анализа ЭЭГ; `spectral_entropy`, рассчитывающая энтропию Шеннона спектра сигнала по методу Уэлча, отражающую степень хаотичности частотного распределения и важную для различения стадий сна; `get_epoch_features`, извлекающая 22 признака из 30-секундных эпох, включая статистические характеристики, параметры Хьёрта, спектральные мощности в различных частотных диапазонах, спектральные соотношения и энтропийные метрики; `gather_data_for_training`, формирующая матрицу признаков и вектор меток путём чтения EDF-файлов и гипнограмм, сегментации сигналов на эпохи, извлечения признаков и синхронизации с экспертными метками стадий; `train_sleep_staging_model`, обучающая классификатор на основе Random Forest с использованием случайных подвыборок данных, случайного выбора признаков при разбиении узлов, агрегации предсказаний деревьев голосованием и оценкой точности на обучающих данных; `smooth_stages`, применяющая скользящее окно из трёх эпох для постобработки гипнограмм с целью устранения кратковременных артефактов классификации через мажоритарное голосование; и `predict_hypnogram_for_file`, интегрирующая функция, управляющая путями к данным, координирующая вызов всех модулей, визуализирующая результаты и генерирующая отчёт с метриками точности.

5. Сравнение полученных программой результатов для здоровых пациентов. Анализ эффективности разработанного алгоритма на данных

здоровых пациентов продемонстрировал среднюю точность классификации в 87.23%. Этот показатель свидетельствует о высокой надежности и стабильности работы модели при распознавании различных стадий сна у лиц без патологий. Достигнутая точность отражает способность алгоритма корректно интерпретировать сложные биофизические сигналы. Кроме того, высокая точность указывает на успешность выбранных методов обработки данных, включая эффективное извлечение признаков, применение современных методов машинного обучения и адекватную предобработку сигналов.

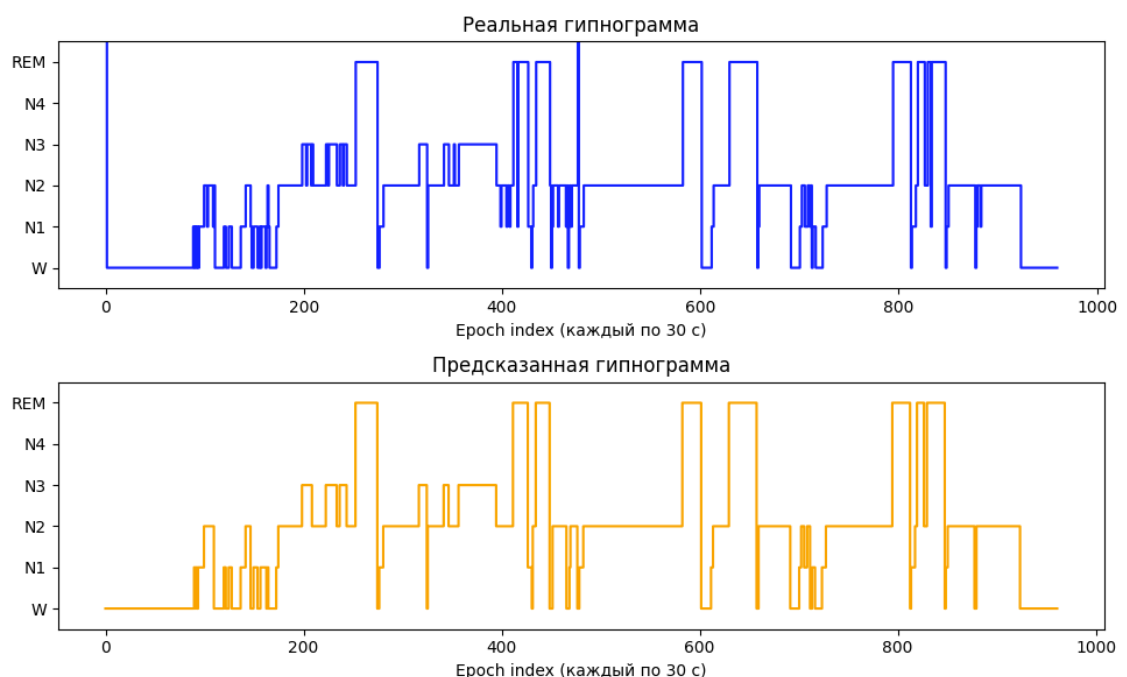


Рис. 1 Сравнение предсказанной гипнограммы с ручной экспертной разметкой для здорового пациента

6. Сравнение результатов для записей с апноэ. Анализ эффективности алгоритма на данных пациентов с синдромом обструктивного апноэ сна (СОАС) показал среднюю точность классификации 88.65%, что свидетельствует о высокой надежности и адаптивности разработанного решения даже при наличии серьезных нарушений сна. Данный результат подтверждает способность модели эффективно распознавать и классифицировать стадии сна в условиях, когда физиологические сигналы характеризуются значительными артефактами и патологическими

изменениями, связанными с остановками дыхания, фрагментацией сна и изменениями мышечного тонуса.

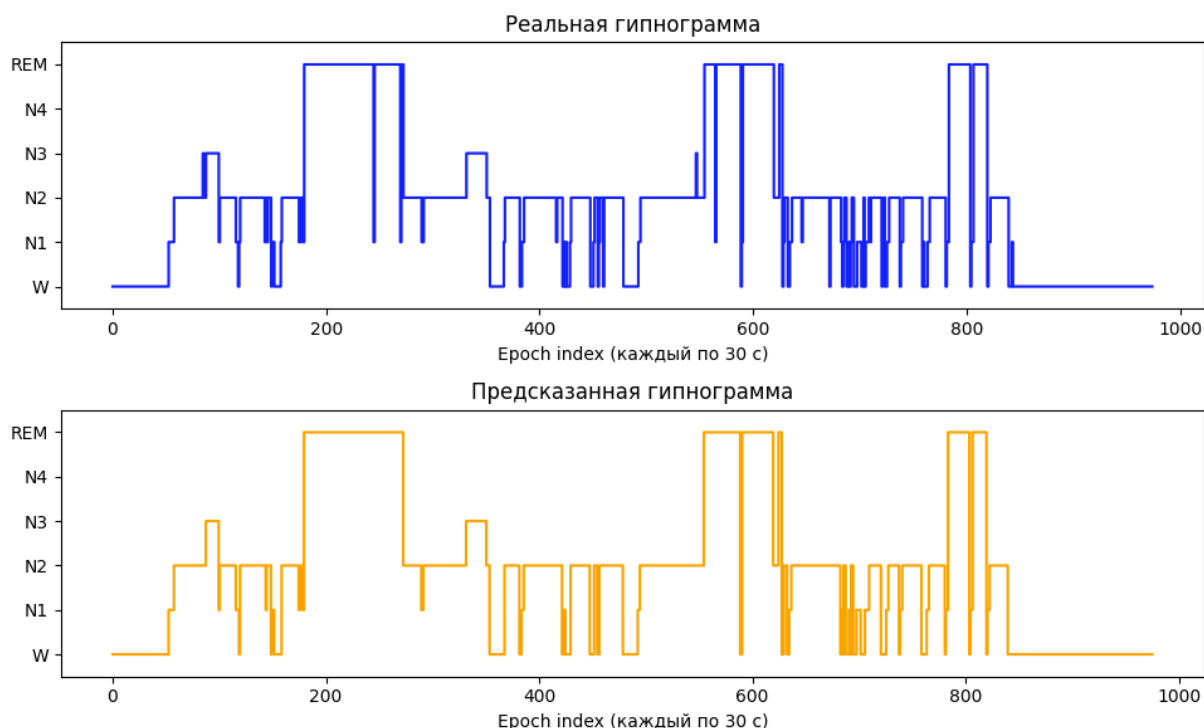


Рис. 2 Сравнение предсказанной гипнограммы с ручной экспертной разметкой для здорового пациента

7. Ограничения методов автоматической разметки стадий сна.

Эффективность алгоритмов автоматического стадирования сна существенно зависит от качества исходных данных. Высокий уровень шума и артефактов, вызванных движениями пациента или плохим контактом электродов, критически снижает точность классификации. Кроме того, индивидуальные особенности ЭЭГ-паттернов у разных пациентов могут значительно отличаться от усреднённых моделей, используемых алгоритмами, что усложняет их универсальное применение. Возрастные различия также играют важную роль, поскольку паттерны сна у детей и пожилых людей требуют разработки специализированных алгоритмов. Клинические ограничения включают влияние различных нарушений сна, таких как апноэ, бессонница и синдром беспокойных ног, которые искажают нормальную архитектуру сна. Фармакологическое воздействие лекарственных препаратов также изменяет характерные паттерны стадий сна, усложняя их распознавание.

Дополнительной сложностью является динамика сна: экспоненциальное ослабление признаков глубокой стадии N3 в поздних циклах и нелинейное изменение интенсивности REM-сна, который растёт при сокращении длительности к утру. Со стороны алгоритмической реализации существуют проблемы переобучения, когда модели плохо обобщают знания на новых данных, а также трудности с точной детекцией микроструктур сна, таких как сонные веретёна, K-комплексы и микропробуждения, что остаётся нерешённой задачей. Кроме того, сложные модели глубокого обучения часто функционируют как «чёрные ящики», затрудняя интерпретацию принимаемых решений. Важным фактором является также вариативность трактовки стандартов AASM разными экспертами, что влияет на качество разметки и обучение моделей. Ресурсные ограничения включают необходимость больших размеченных датасетов, сбор и разметка которых являются дорогостоящими и трудоёмкими процессами, а также высокие вычислительные требования для обучения сложных моделей. Наконец, изменение протоколов записи данных требует регулярного переобучения алгоритмов, что усложняет их адаптацию и внедрение в клиническую практику.

ЗАКЛЮЧЕНИЕ

Разработанный на базе Python алгоритм автоматической классификации стадий сна демонстрирует высокую эффективность, достигая средней точности 87.23% на данных здоровых пациентов, что сопоставимо с современными аналогами в данной области. Основным достижением является автоматизация трудоёмкого процесса ручного анализа, заменённого объективным и воспроизводимым вычислительным методом. Алгоритм реализует комплексный подход, объединяя методы предобработки сигналов, извлечения спектральных и временных признаков, а также классификацию с использованием машинного обучения. Это делает его ценным инструментом

для клинической диагностики, научных исследований и персонализированной медицины сна.

В перспективе планируется повысить обобщаемость модели за счёт расширения обучающей выборки данными пациентов различных возрастных групп и с разнообразными нарушениями сна, а также улучшить интерпретируемость решений алгоритма, что повысит уровень доверия со стороны клиницистов. Кроме того, важным направлением развития является интеграция алгоритма с портативными устройствами, что позволит адаптировать его для домашнего мониторинга с использованием носимых ЭЭГ-гаджетов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Полуэктов М. Г. и др. Сомнология и медицина сна. – 2020.
2. Boostani R., Karimzadeh F., Nami M. A comparative review on sleep stage classification methods in patients and healthy individuals //Computer methods and programs in biomedicine. – 2017. – Т. 140. – С. 77-91.
3. Kharchenko V., Zhdanova I. V. The Wave Model of Sleep Dynamics and an Invariant Relationship between NonREM and REM Sleep //Clocks & Sleep. – 2023. – Т. 5. – №. 4. – С. 686-716.
4. Гольденберг Л. М., Матюшкин Б. Д., Поляк М. Н. Цифровая обработка сигналов. – 1990.
5. Kemp B., Olivan J. European data format ‘plus’(EDF+), an EDF alike standard format for the exchange of physiological data //Clinical neurophysiology. – 2003. – Т. 114. – №. 9. – С. 1755-1761.