

МИНОБРНАУКИ РОССИИ

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

Кафедра физики открытых систем
наименование кафедры

**Автоматизированная разметка стадий сна по частотному анализу на
основе ИНС**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 4041 группы

направления 09.03.02 «Информационные системы и технологии»

код и наименование направления

Института физики

наименование факультета

Степанова Вячеслава Олеговича

фамилия, имя, отчество

Научный руководитель

доцент кафедры ФОС. к.ф.-м.н

должность, ученая степень, уч. звание

подпись, дата

А. О. Сельский

Инициалы Фамилия

Заведующий кафедрой физики
открытых систем

д. ф - м. н., профессор

должность, ученая степень, уч. звание

подпись, дата

А. А. Короновский

Инициалы Фамилия

Саратов 2026 г.

Введение

Сон является одним из важнейших физиологических процессов, связанных с восстановлением организма, работой центральной нервной системы и изменением биоэлектрической активности головного мозга. В течение ночи последовательно сменяются разные стадии сна, отличающиеся глубиной, частотным составом ЭЭГ-сигнала и характером активности мозга [1].

Электроэнцефалография позволяет регистрировать электрическую активность мозга и использовать полученные данные для анализа сна. ЭЭГ-сигнал отражает ритмическую активность коры головного мозга, поэтому его частотная структура может применяться для определения состояний бодрствования, медленного сна, глубокого сна и REM-фазы [2; 3]. В практической обработке особое значение имеют диапазоны Дельта, Тета, Альфа, Сигма и Бета, так как они связаны с различными стадиями сна и могут быть представлены в виде числовых признаков.

Актуальность темы определяется необходимостью автоматизации обработки больших массивов ЭЭГ-данных. Ручная разметка стадий сна требует значительных временных затрат и зависит от опыта специалиста. Использование методов частотного анализа и машинного обучения позволяет ускорить обработку ЭЭГ, повысить воспроизводимость результатов и получить инструмент для предварительной автоматической разметки стадий сна [4; 5].

В данной работе рассматривается автоматизированная разметка стадий сна по частотному анализу. В качестве практического алгоритма классификации используется метод Случайного Леса. Выбор такого подхода связан с тем, что данный метод работает с табличными признаками,

устойчив к шуму и позволяет оценивать важность отдельных параметров [5; 6].

Цель дипломной работы заключается в разработке и оценке метода автоматизированной разметки стадий сна по частотным признакам ЭЭГ-сигналов с использованием метода машинного обучения.

Для достижения поставленной цели в работе решаются следующие задачи: рассмотреть теоретические основы ЭЭГ-сигналов, стадий сна, частотных ритмов, спектрального и вейвлет-анализа; разработать практическую схему обработки ЭЭГ-записей, включающую сегментацию, расчёт энергетических долей ритмов и формирование датасета признаков; оценить результаты автоматической классификации стадий сна, проанализировать точность модели, важность признаков и ограничения предложенного подхода.

Объектом дипломной работы является процесс обработки ЭЭГ-сигналов при анализе стадий сна. Предметом дипломной работы выступают методы частотного анализа ЭЭГ-сигналов и метод Случайного Леса, применяемые для автоматизированной классификации стадий сна.

Практическая основа работы связана с обработкой ЭЭГ-записей 15 пациентов. Для каждого пациента использовались данные двух ночей, при этом каждая ночь включала 6 каналов. Первые 5 пациентов применялись для выявления закономерностей ритмов и стадий сна, а оставшиеся 10 пациентов использовались для проверки точности автоматической классификации.

Дипломная работа состоит из введения, трёх глав, заключения и списка литературы.

Основное содержание работы

В первой главе рассмотрены теоретические основы автоматизированной разметки стадий сна по ЭЭГ. Электроэнцефалография представляет собой метод регистрации электрической активности головного мозга с помощью электродов, размещённых на поверхности головы [2; 3]. Получаемый сигнал отражает суммарную активность больших групп нейронов коры головного мозга. При анализе сна ЭЭГ имеет особое значение, поскольку изменения функционального состояния мозга проявляются в форме, амплитуде и частотном составе сигнала.

Сон не является однородным состоянием. В течение ночи активность мозга изменяется, а сам сон проходит через несколько стадий. В современной разметке обычно выделяют бодрствование, стадии N1, N2, N3 и REM-сон [1]. Каждая из этих стадий связана с определёнными особенностями ЭЭГ, поэтому может рассматриваться как отдельный класс в задаче автоматической классификации.

Бодрствование характеризуется более активной работой мозга и наличием быстрых компонентов ЭЭГ. При спокойном состоянии с закрытыми глазами может проявляться Альфа-ритм. Стадия N1 является переходом от бодрствования ко сну и характеризуется усилением Тета-активности. Стадия N2 относится к более устойчивому медленному сну, одним из признаков которого является Сигма-активность. Стадия N3 соответствует глубокому медленному сну и связана с выраженной Дельта-активностью. REM-сон отличается смешанной частотной картиной и требует анализа совокупности ритмов и их соотношений.

Для автоматизированной разметки важно, что стадии сна отличаются не только наличием отдельных ритмов, но и их соотношением. Один ритм не всегда однозначно указывает на конкретную стадию. Поэтому в задаче классификации целесообразно использовать набор признаков, отражающих

энергетические доли нескольких диапазонов. В рамках работы такими признаками являются Delta_ratio, Theta_ratio, Alpha_ratio, Sigma_ratio и Beta_ratio.

В первой главе также рассмотрены методы частотной обработки ЭЭГ-сигналов. Для автоматизированной разметки стадий сна исходный ЭЭГ-сигнал необходимо преобразовать в набор признаков. Временная запись сама по себе содержит большое количество отсчётов, но для алгоритма классификации удобнее использовать обобщённые характеристики, отражающие частотную структуру сигнала.

Одним из базовых методов перехода от временного представления к частотному является преобразование Фурье [7]. Оно позволяет разложить сигнал на частотные составляющие и определить, какие диапазоны выражены сильнее. Для сравнения разных сегментов ЭЭГ удобнее использовать не абсолютную энергию, а относительную долю энергии ритма. Такой подход позволяет сравнивать сегменты не по общей амплитуде, а по распределению энергии между ритмами.

Преобразование Фурье хорошо подходит для оценки общего частотного состава сегмента, но ЭЭГ является нестационарным сигналом. Его частотная структура может меняться во времени, особенно при переходах между стадиями сна, появлении сонных веретён, кратковременных всплесков активности и артефактов. Для анализа таких сигналов используется вейвлет-преобразование, позволяющее рассматривать сигнал одновременно во временной и частотной областях [8; 9].

После выделения частотных признаков ЭЭГ возникает задача классификации. Каждый сегмент сигнала может быть представлен набором числовых характеристик, например долями энергии Дельта-, Тета-, Альфа-, Сигма- и Бета-ритмов. Далее необходимо определить, к какой стадии сна

относится данный сегмент. Для решения этой задачи в работе используется метод Случайного Леса.

Метод Случайного Леса относится к ансамблевым методам машинного обучения [5; 6]. Его работа основана на построении множества деревьев решений. Каждое дерево обучается на части данных и формирует собственное решение о принадлежности объекта к определённому классу. Итоговый ответ модели определяется голосованием деревьев. Для задачи разметки стадий сна метод Случайного Леса удобен тем, что работает с табличными признаками, способен учитывать нелинейные зависимости между признаками и позволяет оценивать важность параметров.

Во второй главе описана практическая реализация метода автоматизированной разметки стадий сна. Практическая часть дипломной работы выполнялась на основе ЭЭГ-записей 15 пациентов. Для каждого пациента использовались данные двух ночей, при этом каждая ночь включала 6 каналов. Такая структура позволила работать не с единичным сигналом, а с массивом ночных данных, в котором учитывались различия между пациентами, отдельными ночами и каналами внутри одной записи.

Первые 5 пациентов использовались для выявления закономерностей ритмов и стадий сна. Эта часть данных была нужна для понимания того, как распределяются энергетические доли Дельта, Тета, Альфа, Сигма и Бета в сегментах ЭЭГ и как эти признаки могут быть связаны со стадиями сна. Оставшиеся 10 пациентов применялись для проверки точности модели.

Обработка данных выполнялась с использованием библиотек Python. NumPy применялась для работы с массивами, арифметических операций, сегментации сигналов и расчёта энергетических характеристик. PyWavelets использовалась для непрерывного вейвлет-преобразования [9]. Matplotlib применялась для построения графиков сигналов и спектров CWT. CSV использовалась для сохранения признаков в табличном виде. Scikit-learn

применялась для обучения модели Случайного Леса, разбиения выборки, оценки точности и анализа важности признаков [6].

Общая логика обработки начиналась с подготовки ЭЭГ-сигнала к разбиению на фрагменты. Для автоматической классификации исходная ночная запись не использовалась целиком, поскольку метод Случайного Леса работает с табличными признаками. Поэтому запись делилась на сегменты, после чего каждый сегмент рассматривался как отдельный объект классификации.

Для каждого сегмента рассчитывались энергетические доли ритмов Дельта, Тета, Альфа, Сигма и Бета. Энергия нормировалась по суммарной энергии в диапазоне 0,5–30 Гц. Благодаря этому каждый сегмент описывался не абсолютными значениями мощности, а относительным распределением энергии между основными ритмами. Такой формат удобен при обработке данных разных пациентов, поскольку снижает влияние различий в общей амплитуде сигнала.

После расчёта энергетических долей определялся доминирующий ритм. Доминирующим считался тот ритм, на который приходилась наибольшая доля энергии в сегменте. Далее на основе соотношения ритмов формировалась метка стадии сна и показатель уверенности. Уверенность определялась через разницу между ведущей и второй по величине энергетической долей.

Сформированные признаки сохранялись в CSV-файлы. После сохранения каждый сегмент представлялся строкой таблицы, а столбцы содержали энергетические доли ритмов, доминирующий ритм, метку стадии и уверенность. В таком виде датасет использовался как входная база для обучения модели Случайного Леса.

После формирования датасета признаков была выполнена настройка и

обучение модели. Входными признаками выступали энергетические доли ритмов Дельта, Тета, Альфа, Сигма и Бета, а целевой переменной была метка стадии сна. В практической реализации использовался классификатор с 400 деревьями. Количество деревьев задавалось параметром `n_estimators = 400`. На тестовую часть выделялось 25 % данных. Также применялись параметры `random_state = 42`, `stratify = y` и `class_weight = "balanced_subsample"`.

В третьей главе проведена оценка результатов автоматизированной классификации. После обучения модели Случайного Леса была проведена оценка качества классификации на тестовой выборке. Общий датасет включал 45012 строк, каждая из которых соответствовала отдельному сегменту ЭЭГ с рассчитанными частотными признаками. После разбиения данных обучающая выборка составила 33759 сегментов, а тестовая выборка включала 11253 сегмента.

Полученная точность составила 0,9988, или около 99,88 %. Такой результат показывает, что модель Случайного Леса практически полностью воспроизвела сформированную разметку стадий сна на основе частотных признаков ЭЭГ. При интерпретации результата важно учитывать способ формирования исходных меток. В данной работе стадии сна определялись на основе частотных признаков и физиологических правил, после чего модель Случайного Леса обучалась воспроизводить эту разметку.

По результатам классификационного отчёта все стадии сна были распознаны с высокими значениями `precision`, `recall` и `F1-score`. Наилучший результат был получен для стадии N2, где все показатели составили 1,0000. Стадия N3 имела наибольшую поддержку, равную 9765 сегментам, и также была классифицирована с очень высокой точностью. Для стадий N1, REM и Wake показатели также оказались высокими, хотя количество сегментов этих классов было заметно меньше.

После обучения модели был выполнен анализ важности признаков.

Наибольшую важность получил Sigma_ratio со значением 0,264. Далее следовали Theta_ratio со значением 0,255, Beta_ratio со значением 0,213, Delta_ratio со значением 0,201 и Alpha_ratio со значением 0,067. Такое распределение признаков является физиологически объяснимым. Сигма-ритм связан со стадией N2 и сонными веретёнами [10], Тета-активность характерна для медленных стадий сна и может проявляться в REM-фазе, Бета-активность связана с быстрыми компонентами ЭЭГ, а Дельта-ритм отражает глубокий медленный сон.

Полученные результаты показывают, что автоматизированная разметка стадий сна по частотным признакам ЭЭГ может использоваться как рабочий подход для предварительной классификации. При этом такой результат не означает, что метод полностью заменяет экспертную разметку сна. У предложенного подхода есть ограничения, которые необходимо учитывать при интерпретации результатов.

Первое ограничение связано с тем, что классификация строилась на частотных признаках ЭЭГ. Такой подход хорошо отражает распределение энергии между ритмами, но не всегда позволяет точно отделить стадии со сходной частотной структурой. Второе ограничение связано с артефактами. ЭЭГ-сигналы могут содержать помехи, вызванные движениями пациента, мышечным напряжением, изменением положения электродов или техническими факторами. Третье ограничение связано с тем, что метод Случайного Леса классифицирует каждый сегмент как отдельный объект и не учитывает временную последовательность стадий сна, если такая информация специально не добавлена в признаки.

Для повышения качества автоматической разметки можно выделить несколько направлений улучшения. К ним относятся расширение набора признаков, улучшение обработки артефактов, учёт соседних сегментов и увеличение объёма обучающей выборки. Дополнительный учёт временной

структуры сна может сделать автоматическую разметку более устойчивой и физиологически правдоподобной.

Заключение

В ходе выполнения дипломной работы была рассмотрена задача автоматизированной разметки стадий сна по частотным признакам ЭЭГ-сигналов с использованием метода машинного обучения. Было показано, что стадии сна имеют различные проявления в частотной структуре ЭЭГ. Бодрствование связано с Альфа- и Бета-компонентами, медленные стадии сна характеризуются участием Тета-активности, стадия N2 связана с Сигма-ритмом, а глубокий сон N3 проявляется через Дельта-активность.

В работе были рассмотрены методы частотной обработки ЭЭГ. Преобразование Фурье позволяет оценивать распределение энергии сигнала по частотным диапазонам, а вейвлет-преобразование даёт возможность учитывать изменение частотной активности во времени. Эти методы создают основу для расчёта энергетических долей Дельта, Тета, Альфа, Сигма и Бета, которые используются как признаки для модели машинного обучения.

Практическая обработка включала сегментацию ЭЭГ-сигналов, расчёт энергетических долей ритмов, определение доминирующего ритма, формирование метки стадии сна и оценку уверенности классификации. Каждый сегмент был представлен в виде строки датасета с признаками `Delta_ratio`, `Theta_ratio`, `Alpha_ratio`, `Sigma_ratio` и `Beta_ratio`. Такой формат позволил перейти от исходной ЭЭГ-записи к таблице признаков, пригодной для обучения модели Случайного Леса.

В результате обучения модели Случайного Леса была получена точность 99,88 % на тестовой выборке. Данный результат отражает качество воспроизведения сформированной автоматической разметки по частотным

признакам. Для клинической проверки метода в дальнейшем требуется сопоставление с экспертной разметкой сна.

Анализ важности признаков показал, что наибольший вклад в работу модели внёс Sigma_ratio . Далее следовали Theta_ratio , Beta_ratio , Delta_ratio и Alpha_ratio . Полученное распределение признаков является физиологически объяснимым и подтверждает связь выбранных частотных характеристик со стадиями сна.

Практическая значимость работы состоит в том, что предложенная схема позволяет автоматизировать обработку ЭЭГ-сегментов и получить предварительную разметку стадий сна без ручного анализа каждого фрагмента. Такой подход может использоваться как вспомогательный инструмент при обработке больших массивов ночных данных, формировании обучающих датасетов и первичной оценке структуры сна.

В итоге цель дипломной работы была достигнута. Была разработана и оценена схема автоматизированной разметки стадий сна по частотным признакам ЭЭГ-сигналов с использованием алгоритма машинного обучения на основе метода Случайного Леса. Полученные результаты подтверждают, что сочетание FFT, SWT, расчёта энергетических долей ритмов и машинной классификации может применяться для предварительного определения стадий сна по ЭЭГ.

Список литературы

1. Ковальзон В. М. Основы сомнологии. Физиология и нейрохимия цикла «бодрствование-сон». М.: БИНОМ. Лаборатория знаний, 2012. 239 с.
2. Гнездицкий В. В. Обратная задача ЭЭГ и клиническая электроэнцефалография. М.: МЕДпресс-информ, 2004. 624 с.
3. Зенков Л. Р. Клиническая электроэнцефалография. М.: МЕДпресс-информ, 1996.

4. Ходашинский И. А. Биометрические данные и методы машинного обучения в диагностике и мониторинге нейродегенеративных заболеваний. Обзор // Научная статья. 2022.
5. Воронцов К. В. Машинное обучение. Курс лекций. Электронный ресурс. Дата обращения: 28.05.2026.
6. Scikit-learn. RandomForestClassifier. Документация библиотеки. Электронный ресурс. Дата обращения: 28.05.2026.
7. Иванов А. А. Обзор методов математического анализа ЭЭГ. Количественная ЭЭГ // Научная статья. 2023.
8. Комаров П. В. Исследование динамически меняющегося сигнала с применением вейвлет-преобразований // Научная статья. 2024.
9. PyWavelets. Continuous Wavelet Transform. Документация библиотеки. Электронный ресурс. Дата обращения: 28.05.2026.
10. Грубов В. В. Вейвлетный анализ сонных веретен на ЭЭГ и разработка метода их автоматической диагностики // Известия вузов. Прикладная нелинейная динамика. 2011.