ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 33 за 2025 г.

ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Д.В. Фаевский, В.А. Викулов, В.А. Судаков

Анализ активности головного мозга по данным EEG и fNIRS с применением объяснимого искусственного интеллекта

Рекомендуемая форма библиографической ссылки: Фаевский Д.В., Викулов В.А., Судаков В.А. Анализ активности головного мозга по данным EEG и fNIRS с применением объяснимого искусственного интеллекта // Препринты ИПМ им. М.В.Келдыша. 2025. № 33. 23 с. EDN: CTJZSO https://library.keldysh.ru/preprint.asp?id=2025-33

О р д е н а Л е н и н а
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Р о с с и й с к о й а к а д е м и и н а у к

Д.В. Фаевский, В.А. Викулов, В.А. Судаков

Анализ активности головного мозга по данным EEG и fNIRS с применением объяснимого искусственного интеллекта

Москва – 2025

Фаевский Д.В., Викулов В.А., Судаков В.А.

Анализ активности головного мозга по данным EEG и fNIRS с применением объяснимого искусственного интеллекта

В работе рассматривается возможность комбинированного использования электроэнцефалографии (EEG) и функциональной спектроскопии в ближнем инфракрасном диапазоне (fNIRS) для анализа мозговой активности. Применение методов объяснимого ИИ (XAI SHAP-анализ) подтвердило биологическую интерпретируемость результатов: доминирование EEG-признаков согласуется с известными нейрофизиологическими маркерами, в то время как вклад fNIRS остаётся ограниченным из-за низкого временно́го разрешения. Ключевым ограничением является отсутствие учёта временно́го лага нейроваскулярной связи, что снижает полезность fNIRS-данных. Перспективным направлением дальнейших исследований является разработка асимметричных моделе́й, явно учитывающих временные задержки между модальностями (например, через кросс-модальное внимание или временно́е выравнивание).

Ключевые слова: EEG, fNIRS, мультимодальный анализ, объяснимый ИИ (XAI), SHAP-анализ, интерфейсы «мозг-компьютер».

Dmitry Vladimirovich Faevsky, Vladimir Alexandrovich Vikulov, Vladimir Anatolievich Sudakov

Analysis of brain activity based on EEG and fNIRS data using explainable artificial intelligence

The paper examines the possibility of combining electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS) for brain activity analysis. Application of explainable AI methods (XAI SHAP analysis) confirmed the biological interpretability of the results: the dominance of EEG features is consistent with known neurophysiological markers, while the contribution of fNIRS remains limited due to low temporal resolution. A key limitation is the lack of consideration of the time lag of neurovascular coupling, which reduces the usefulness of fNIRS data. A promising direction for further research is the development of asymmetric models that explicitly take into account time delays between modalities (e.g., through cross-modal attention or temporal alignment).

Key words: EEG, fNIRS, multimodal analysis, explainable AI (XAI), SHAP analysis, brain-computer interfaces (BCIs).

1. Введение

Современная наука о мозге, клиническая медицина и исследования в области технологий интерфейсов «мозг-компьютер» в значительной степени опираются на анализ активности мозга. Это позволяет глубже изучать когнитивные процессы, диагностировать неврологические расстройства и создавать инновационные системы взаимодействия человека с техникой. В настоящее время популярными технологиями мониторинга интерфейсов «мозг-компьютер» являются электроэнцефалография (ЭЭГ – англ. EEG) [1], функциональная магнитно-резонансная томография (фМРТ) [2], магнитоэнцефалография (МЭГ) [3] и функциональная спектроскопия в ближнем инфракрасном диапазоне (fNIRS) [4], каждая из которых имеет свои преимущества и ограничения. Объединение нескольких технологий мониторинга обеспечивает новый мультимодальный подход, который объединяет преимущества каждой технологии, а также преодолевает ограничения отдельных подходов. Гибридная система EEG-fNIRS рассматривается как один из наиболее доступных неинвазивных подходов к сбору и обработке данных, что позволяет расширить возможности существующих технологий в интерфейсах «мозг-компьютер».

Однако каждая из используемых модальностей (EEG/ fNIRS) имеет свои преимущества и недостатки: например, EEG обеспечивает высокое временно́е разрешение, вместе с тем имеет низкую пространственную точность, в то время как fNIRS фиксирует гемодинамические изменения с лучшим пространственным разрешением, но имеет недостатки в скорости регистрации. Комбинация этих методов открывает путь к мультимодальному анализу, объединяющему электрическую активность нейронов и связанные с метаболизмом гемодинамические процессы, что повышает точность и глубину интерпретации данных.

Несмотря на перспективность мультимодальных систем, их внедрение сталкивается с серьёзными проблемами. Различия в природе сигналов fNIRS и EEG, а также их временны́х и пространственных характеристиках усложняют совместную обработку данных. Кроме того, алгоритмы искусственного интеллекта, применяемые для анализа, часто остаются «чёрными ящиками», что снижает доверие к результатам в критически важных областях, таких как медицина. Это определяет необходимость внедрения методов объяснимого искусственного интеллекта (XAI), способных раскрыть логику принятия решений моделями и обеспечить валидацию выводов на основе нейрофизиологических знаний.

Целью данного исследования является разработка гибридной модели для совместного анализа fNIRS и EEG, направленной на преодоление ограничений отдельных модальностей. Особый акцент сделан на интеграции методов XAI для интерпретации решений модели, что позволит не только повысить точность предсказаний, но и установить соответствие между выявленными паттернами

активности мозга и известными нейрофизиологическими механизмами. Такой подход способствует созданию прозрачных и надёжных инструментов, востребованных как в фундаментальных исследованиях, так и в клинической практике.

2. Обзор литературы

Исследование работ, связанных с анализом данных на основе одной модальности (fNIRS и EEG по отдельности), показало наличие большого числа интересных работ. Выделим некоторые из них, в которых были достигнуты результаты высокой точности классификации, а также упомянуты подходы к интерпретируемости предлагаемых решений. В статье [5] предложен алгоритм ICGN, основанный на глубоких нейронных сетях, для улучшения точности классификации в системе BCI на основе fNIRS. Алгоритм ICGN использует контекстную информацию и гейтированные процессы для оптимизации классификации, что позволяет эффективно фильтровать и ранжировать релевантные данные. Результаты показали, что точность классификации с использованием ICGN составила 91,23 ± 1,60. Это значительно ($p < 0,025$) выше, чем у алгоритмов LSTM и Bi-LSTM, что подтверждает преимущество в улучшении точности классификации современных систем fNIRS-BCI.

В статье [6] представлен новый метод xMVPA для анализа работы мозга младенцев с помощью fNIRS-технологии. Этот подход сочетает искусственный интеллект с возможностью объяснения результатов и современные методы анализа данных. Главное преимущество xMVPA – относительно высокая точность и способность указывать на то, какие именно участки мозга активируются и как они связаны с развитием ребенка.

В исследовании [7] предложен новый метод анализа EEG-сигналов для отслеживания прогрессирования болезни Альцгеймера. Специально разработанная нейросеть смогла с точностью до 98,97% различить раннюю (MCI) и позднюю (AD) стадии заболевания на основе данных одного пациента. Для понимания работы алгоритма авторы использовали методы объяснимого искусственного интеллекта, которые показали, какие именно зоны мозга наиболее активны при развитии болезни.

В развитии идей и исследований сигналов мозга активно набирают популярность исследования, связанные с комбинированием разных сигналов, в частности двух модальностей – EEG и fNIRS.

Исследователи по-разному подходят к классификации методов и стратегий работы с комбинированными данными. Так, Li [8] в своей работе выделяет три основные стратегии анализа данных EEG и fNIRS одновременно: EEG с учётом сигналов fNIRS, fNIRS с учётом сигналов EEG и параллельный анализ.

С другой стороны, в зависимости от уровня интеграции сигналов современные методы мультимодального анализа данных fNIRS и EEG можно разделить на три основных подхода: раннее, промежуточное и позднее слияние.

Раннее слияние (early-stage fusion) предполагает объединение сырых данных и демонстрирует наилучшие результаты в задачах моторного воображения (motor imagery), обеспечивая точность до 76%, что подтверждается исследованиями [9]. Этот подход позволяет нейросетевым архитектурам непосредственно выявлять сложные кросс-модальные паттерны, но требует строгой синхронизации данных и значительных вычислительных ресурсов.

Промежуточное слияние (middle-stage fusion) основано на комбинации извлечённых признаков и представляет собой компромиссный вариант, особенно подходящий для мониторинга когнитивных состояний, где важны как точность, так и интерпретируемость результатов.

При позднем слиянии (late-stage fusion) проводится совместный анализ результатов независимой обработки каждой модальности. Данный подход оказывается наиболее устойчивым к шумам и асинхронности данных, что делает его предпочтительным для клинической диагностики.

Каждый из этих подходов сталкивается с уникальными проблемами интеграции и синхронизации данных, что ограничивает их универсальность. Выбор оптимального метода зависит от конкретной задачи и соблюдения баланса точности, интерпретируемости и устойчивости к шумам.

Многие авторы отмечают, что раскрытию потенциала мультимодальных исследований активности мозга может помочь создание открытых баз данных. В настоящее время открытые датасеты часто содержат записи только одной модальности, что осложняет разработку и тестирование гибридных алгоритмов.

В исследовании [10] отмечается, что несмотря на растущее количество приложений, сочетающих EEG и fNIRS, методологическая строгость предыдущих исследований остаётся неясной, что ограничивает точную интерпретацию полученных результатов и затрудняет трансляционное применение данного мультимодального подхода.

Отметим некоторые проблемы при работе с мультимодальными данными, которые нашли отражение в нескольких работах, а именно:

- синхронизация данных разных модальностей. Сигналы EEG отражают мгновенную электрическую активность нейронов (миллисекундный масштаб), тогда как fNIRS фиксирует медленные гемодинамические изменения (секундный масштаб);

- представление данных в общее пространство признаков и устранение их несбалансированной информативности. Это связано с тем, что EEG предоставляет многоканальные временны́е ряды с высокой частотой, а fNIRS – пространственно-временны́е карты оксигенации крови. По результатам исследований признаки EEG часто доминируют в задачах классификации

активности мозга из-за их высокой временно́й плотности, снижая вклад fNIRS [11].

Для успешной синхронизации данных в работе [12] предложили решить задачу разработки алгоритмов ресемплинга для сопоставления вре́менных рядов, так как EEG обычно записывается с частотой дискретизации 100–1000 Гц, а fNIRS – 10–50 Гц. Кроме того, имеет место задержка нейроваскулярной связи: гемодинамический ответ fNIRS запаздывает на 2–6 секунд относительно электрической активности, что требует введения динамических моделей для учёта этой задержки.

Для преобразования данных в общее пространство признаков предлагается использовать снижение размерности (PCA, ICA) для выделения латентных паттернов или использование графовых моделей для учёта функциональной связности между областями мозга. Несбалансированную информативность признаков авторы [13] предлагают устранять, используя взвешенные схемы (например, обучение с вниманием) или гибридные архитектуры нейросетей (CNN для fNIRS и RNN для EEG).

В статье [14] предложена новая методика мультимодальной фьюжн обработки EEG и fNIRS с использованием многоуровневого прогрессивного обучения и алгоритма атомарного поиска для выбора признаков. Этот метод эффективно извлекает и объединяет признаки мозговой активности, улучшая классификацию в задачах моторной визуализации и умственной арифметики. Эксперименты с набором данных EEG-fNIRS показали высокую точность классификации: 96,74% для задачи MI и 98,42% для задачи MA. Обеспечение совместимости данных из-за различий в количестве каналов между методами fNIRS и EEG предлагается решать с использованием алгоритма, основанного на рекуррентных графах (RP) и глубоких нейронных сетях (CNN-LSTM), который позволяет интегрировать данные из этих методов без необходимости уменьшения дискретизации. Авторы работ [15-16] утверждают, что алгоритм извлекает важные особенности мозговой активности, обеспечивая высокую точность классификации. Средняя точность составила 78,44% для fNIRS, 86,24% для EEG и 88,41% для гибридной системы EEG-fNIRS.

Отметим, что все попытки преодолеть сложности мультимодальных данных в большинстве своём используют ad-hoc методы, что затрудняет воспроизводимость и разработку стандартизированных протоколов синхронизации данных, которые могли бы помочь в раскрытии потенциала мультимодальных подходов.

С другой стороны, исследователи видят перспективу именно в комбинировании данных разных модальностей. В статье [17] оказалось, что сочетание EEG и fNIRS даёт более полную картину: EEG лучше фиксирует быстрые нейронные изменения в состоянии покоя, тогда как fNIRS эффективнее регистрирует медленные гемодинамические процессы при когнитивных нагрузках. Для обработки сигналов авторами были использованы нелинейные методы анализа (фрактальная размерность, энтропия),

генетический алгоритм для отбора признаков и ансамбль классификаторов (SVM, Random Forest и др.), достигший высокой точности распознавания (95,48%).

Открытой и актуальной остаётся задача обеспечить достаточную интерпретируемость алгоритмов. Даже при успешной интеграции данных модели редко объясняют, какие именно комбинированные паттерны EEG-fNIRS влияют на результат, что критично для медицинских приложений. Анализ работ по XAI проведён в разделе 3.

3. Объяснимый искусственный интеллект

Интерпретация моделей искусственного интеллекта в науках о мозге сталкивается с уникальными вызовами: необходимостью соотнесения алгоритмических решений с нейрофизиологическими механизмами, гетерогенностью данных (EEG, fNIRS) и высокими требованиями к надёжности в клинических приложениях. Методы объяснимого искусственного интеллекта (XAI), такие как LIME, SHAP, градиентные подходы и механизмы внимания, становятся ключевыми инструментами для преодоления этих проблем. Их применение позволяет не только валидировать модели, но и выявлять биологически значимые паттерны активности мозга. Рассмотрим наиболее частое применение XAI методов в науке о мозге.

LIME (Local Interpretable Model-agnostic Explanations). Метод генерирует локальные объяснения, аппроксимируя поведение сложной модели в окрестности конкретного примера данных с помощью интерпретируемых моделей (например, линейной регрессии). С помощью данного метода выявляют ключевы́е временные окна или каналы EEG, критичные для классификации состояний (например, эпилептических приступов или когнитивной нагрузки) [18]. Также имеются работы по интерпретации гемодинамических паттернов fNIRS: определение регионов мозга, чья оксигенация крови наиболее влияет на предсказание [19]. Несмотря на популярность метода, он имеет существенные ограничения – метод чувствителен к шуму в нейрофизиологических данных, а его объяснения носят локальный характер.

SHAP (SHapley Additive exPlanations). SHAP, основанный на теории игр, количественно оценивает вклад каждого признака в предсказание модели, обеспечивая глобальную и локальную интерпретацию. Данный метод можно использовать для ранжирования значимости временны́х, спектральных или пространственных признаков в мультимодальных данных.

В работе [20] для интерпретации результатов унимодальной архитектуры xAI-fNIRS было предложено использование алгоритма DeepShap. Алгоритм позволил определять, какие каналы влияют на результат классификации. В работе была рассмотрена система, которая демонстрирует точность классификации более 98% с использованием метода скользящего окна и моделей глубокого обучения (CNN, LSTM).

Отметим, что потенциал данного метода применительно к мультимодальным данным недостаточно раскрыт, так как применение этого подхода для сравнения вклада электрической активности (EEG) и гемодинамики (fNIRS) в общий результат может быть крайне перспективным. Так, при прогнозировании когнитивного утомления SHAP может выявить, что низкочастотные колебания EEG (тета-диапазон) и фронтальная оксигенация в fNIRS имеют наибольший вес в решении модели. При этом работ по использованию данного метода для интерпретации мультимодальных архитектур при классификации EEG-fNIRS сигналов практически нет.

Градиентные методы (Grad-CAM, Guided Backpropagation). Эти подходы визуализируют важность признаков через анализ градиентов модели по отношению к входным данным, что особенно эффективно для нейросетей. Методы применимы для картирования пространственно-временно́й активности мозга и позволяют, к примеру, выделять области коры, активирующиеся при обработке речи (fNIRS) или сенсомоторных ритмов (EEG). Так, например, в работе [21] этот метод используется для выделения каналов EEG, связанных с эмоциями. Относительными ограничениями данного подхода является наличие доступа к архитектуре модели и нестабильность для зашумленных сигналов.

Механизмы внимания. Слои внимания в нейросетях автоматически выделяют наиболее информативные части данных, обеспечивая встроенную интерпретацию. Метод может применяться для фокусировки на ключевых фазах сигнала (например, потенциалах, связанных с событиями в EEG) или помогать выделять релевантные области мозга в данных fNIRS или топографиях EEG. В мультимодальных исследованиях метод может помочь лучше анализировать комбинацию EEG и fNIRS с учётом их взаимного влияния (например, в трансформерных моделях для прогноза нейродегенеративных заболеваний attention-веса могут указывать на корреляцию между замедлением альфа-ритмов (EEG) и снижением оксигенации в теменной доле (fNIRS)).

Таким образом, LIME, SHAP, градиентные методы и механизмы внимания открывают новые возможности для изучения активности мозга, превращая «черный ящик» моделей в инструмент для генерации нейрофизиологических гипотез. Их интеграция в мультимодальный анализ EEG-fNIRS позволяет не только улучшать точность моделей, но и выявлять кросс-модальные паттерны, связанные с когнитивными функциями или заболеваниями.

4. Методика обработки данных

В качестве датасета в работе было использовано подмножество данных из открытого датасета Open access dataset for simultaneous EEG and NIRS brain-computer interface (BCI) [22], а именно часть датасета, в которой участникам была поставлена задача моторного воображения (Motor Imagery, MI). В выбранном нами подмножестве датасета (датасет A) содержится информация о

выполнении участниками тактильного моторного воображения (представление ощущения сжатия и разжатия руки, как будто участник держит мяч). На экране появлялась стрелка, указывающая на левую или правую сторону, что сигнализировало, какую руку нужно представлять. Задача выполнялась в течение 10 секунд, после чего следовал период отдыха (15-17 секунд). Каждая сессия включала 20 повторений (10 для левой и 10 для правой руки).

Для записи данных EEG использовался многоканальный усилитель BrainAmp EEG (Brain Products GmbH), использующий 30 активных электродов, размещенных по системе 10-5. Запись данных EEG осуществлялась с частотой дискретизации 200 Гц. Запись данных fNIRS осуществлялась с частотой дискретизации 12,5 Гц с использованием системы NIRScout (NIRx GmbH), создающей 36 физиологических каналов. Сигналы EEG и fNIRS записывались одновременно.

Краткие характеристики датасета:

- количество испытуемых: 29 здоровых взрослых добровольцев;
- количество временны́х сегментов на испытуемого: 37;
- общий объём данных: 1073 временны́х сегментов (29 × 37 = 1073);
- разбиение на обучающую и тестовую выборки:
  обучающая выборка: 800 временны́х сегментов (~74,6% данных);
  тестовая выборка: 273 временны́х сегмента (~25,4% данных).

Каждый временно́й сегмент содержит 7000 значений для каждого канала EEG-сигнала (соответствует 35 секундам при частоте дискретизации 200 Гц) и 438 значений для каждого канала fNIRS-сигнала (соответствует 35,04 секунды при частоте дискретизации 12,5 Гц). Метки (маркеры) синхронизировались с учётом временно́го смещения.

Отфильтрованные данные были разделены на сегменты длительностью 35 секунд (с интервалом [–10000 мс, +25000 мс] относительно триггеров). Сегментация также включала в себя проверку полноты сегментов: неполные эпохи удалялись. При различии интервалов для разных классов позиции маркеров были переопределены.

Базовый уровень рассчитан для интервала [–5000 мс, +2000 мс], далёкого от стимула, чтобы избежать влияния гемодинамического ответа. Коррекция выполнена отдельно для каждого класса (моторное воображение левой/правой руки) и независимо для каждого канала по формуле (1).

$$\Delta c_{corrected}(t) = \Delta c(t) - 1/N \sum_{t \in BL} \Delta c(t), \quad (1)$$

где $\Delta c(t)$ – исходное изменение концентрации (HbO_2 или HbR) в момент времени t, BL – интервал базового уровня, N – количество точек в BL. Среднее значение базового интервала вычиталось из данных сегментов.

Для расчёта относительных концентраций оксигенированного (HbO_2) и деоксигенированного (HHb) гемоглобина для данных fNIRS использован закон Бера–Ламберта:

$$A_\lambda = -log\left(I_\lambda / I_{baseline,\lambda}\right), \quad (2)$$

где A_λ – ослабление на длине волны λ, I_λ – текущая интенсивность, $I_{baseline,\lambda}$ – средняя интенсивность в базовой линии.

На основе измерений оптической плотности на двух длинах волн ($A_{\lambda 1}$, $A_{\lambda 2}$) была составлена и решена система линейных уравнений (3) относительно c_{oxy}, c_{deoxy}, учитывающая коэффициенты экстинкции HbO_2 и HHb(ε), дифференциальный фактор длины пути (DPF), расстояние между оптодами (l):

$$\begin{cases} A_{\lambda_1} = \left(\varepsilon_{oxy,\lambda_1} \cdot c_{oxy} + \varepsilon_{deoxy,\lambda_1} \cdot c_{deoxy}\right) \cdot l \cdot DPF_{\lambda_1}, \\ A_{\lambda_2} = \left(\varepsilon_{oxy,\lambda_2} \cdot c_{oxy} + \varepsilon_{deoxy,\lambda_2} \cdot c_{deoxy}\right) \cdot l \cdot DPF_{\lambda_2}. \end{cases} \quad (3)$$

Таким образом, исходные оптические плотности преобразованы в относительные концентрации оксигенированного (HbO) и деоксигенированного (HbR) гемоглобина.

В ходе расчетов использовались следующие параметры:

- длины волн λ_1, λ_2: 760 нм и 850 нм соответственно;
- дифференциальные факторы длины пути DPF_{λ_1}, DPF_{λ_2}, отражающие увеличение оптического пути света из-за рассеяния в биологических тканях: для 760 нм – 5,98, для 850 нм – 7,15;
- расстояние между оптодами l : 3,0 см;
- экстинкционные коэффициенты (молярные коэффициенты экстинкции), характеризующие способность вещества поглощать свет на определённой длине волны: для длины волны 760 нм (низкая): $\varepsilon_{oxy,\lambda_1}$= 0,974 (см$^{-1}$·М$^{-1}$), $\varepsilon_{deoxy,\lambda_1}$= 0,693 (см$^{-1}$·М$^{-1}$), для длины волны 850 нм (высокая): $\varepsilon_{oxy,\lambda_2}$ = 0,35 (см$^{-1}$·М$^{-1}$), $\varepsilon_{deoxy,\lambda_2}$= 2,1 (см$^{-1}$·М$^{-1}$).

Преобразованный сигнал был также разделён на сегменты длительностью 35 секунд (с интервалом [–10 с, +25 с] относительно маркеров. На этапе сегментации была проведена проверка полноты сегментов. В случае неполноты сегменты удалялись. При различии интервалов для разных классов позиции маркеров были переопределены. Базовый уровень скорректирован для интервала [–5 мс, +20 с] аналогично EEG, с учётом особенностей fNIRS-сигналов (классовая и канальная коррекция).

По итогу обработки данных описанный конвейер предобработки обеспечил:

1. Синхронизацию EEG и fNIRS;
2. Устранение артефактов (дрейф, высокочастотные шумы);
3. Нормализацию данных относительно базового уровня;
4. Подготовку структурированных сегментов для последующего анализа.

Все сигналы были стандартизированы (z-score нормализация). Данные содержат бинарные метки (например, «наличие/отсутствие когнитивной

нагрузки»), что позволяет применять модель машинного обучения для задач классификации.

5. Архитектура модели

Предложенная модель представляет собой двунаправленную LSTM-сеть с механизмом внимания, разработанную для совместного анализа данных EEG и fNIRS (см. рис 1). Архитектура специально оптимизирована для обработки временны́х рядов нейрофизиологических сигналов. Сеть состоит из трёх основных модулей: модуля обработки fNIRS данных, модуля обработки EEG данных и классификатора.

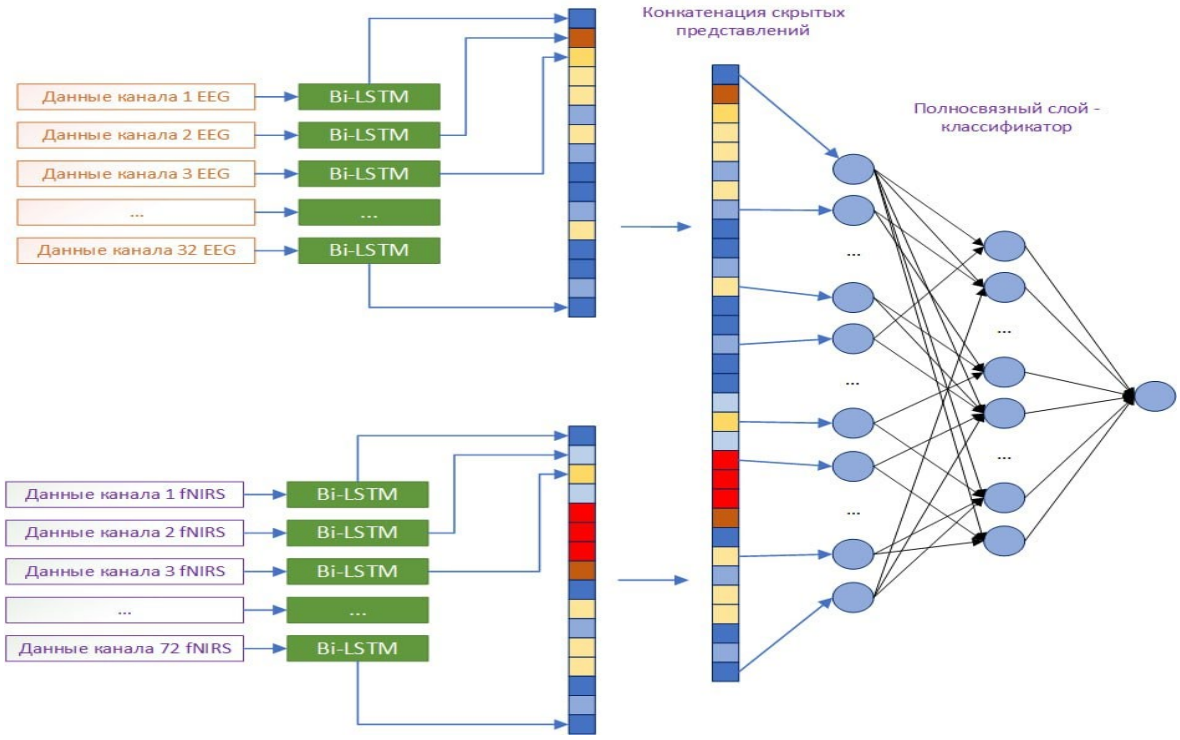


Рис 1. Мультимодальная архитектура нейронной сети для классификации сигналов EEG и fNIRS

Модуль обработки fNIRS данных представляет собой двунаправленный LSTM слой: 2 слоя, 64 скрытых нейрона в каждом направлении (общий выходной размер 128). В данном модуле осуществляется обработка 72 временны́х сегментов (для каждого из каналов fNIRS) размером 438 значений каждый. Для каждого сегмента сохраняется последнее скрытое состояние LSTM, которое выступает скрытым представлением признаков каждого канала.

Модуль обработки EEG данных также представляет собой двунаправленный LSTM слой: 2 слоя, 128 скрытых нейрона в каждом направлении (общий выходной размер 256). В это модуле обрабатываются 32 временны́х сегмента (для каждого из каналов EEG) размером 7000 значений

каждый. Аналогично fNIRS модулю, на выходе сохраняются последние скрытые состояния каждого из канала в качестве скрытого представления признаков.

Результат конкатенации скрытого представления признаков для каждого из каналов подаётся в классификатор, который представляет собой блок из полносвязного слоя, слоя нормализации L2 и выходного слоя с сигмоидной активацией для бинарной классификации. Для устойчивости обучения используется регуляризация Dropout с вероятностью 0,5.

Архитектура демонстрирует эффективное сочетание методов глубокого обучения для задач классификации мультимодальных нейрофизиологических данных, а также позволяет применять к ней методы XAI, так как сохраняется информация о вводимой информации в каждом из каналов.

Для сравнения с унимодальными моделями для классификации сигналов fNIRS и EEG была использована аналогичная архитектура сети с незначительными изменениями: скрытые представления не были объединены между собой и классификатор имел меньший размер для соответствия конкатенированных скрытых представлений каналов для каждой из модальностей.

Для объяснения решений, принимаемых глубокой нейронной сетью, был применён метод SHAP (SHapley Additive exPlanations) [23]. Данный подход обеспечивает количественную оценку вклада каждого входного признака в итоговый выход модели. В исследовании использовался **shap.DeepExplainer** – специализированный инструмент для интерпретации глубоких нейронных сетей, который вычисляет приближённые значения Шепли с учетом иерархической структуры многослойных архитектур.

Для оценки значимости отдельных каналов данных SHAP-значения подвергались двухэтапной агрегации: сначала усреднялись по всем сэмплам, затем – по каждому из каналов. SHAP-значения для каждого отдельного сэмпла вычислялись по формуле Шепли для каждого из признаков наблюдения:

$$\phi_i = \sum_{S \subseteq P \setminus \{j\}} \frac{|S|!(|P|-|S|-1)!}{|P|!} * \left(f(x_s \cup \{j\}) - f(x_s)\right), \quad (4)$$

где ϕ_i – SHAP значение для признака *i*, *f* – модель машинного обучения, *x* – вектор признаков, *P* – общее количество признаков (во всех 104 каналах наблюдений 72 канала fNIRS и 32 канала EEG), *S* – подмножество признаков без *i*-го признака, x_s – вектор признаков, где только признаки из S имеют свои оригинальные значения, а остальные заменены на значения из референсного датасета (фоновых данных).

Для алгоритма DeepSHAP использовалась упрощённая аппроксимация SHAP через градиенты со следующей формулой:

$$\phi_i \approx \frac{1}{M} \sum_{k=1}^{M} \frac{\partial f(x^{(k)})}{\partial x_i} \cdot (x_i - \overline{x_i}),$$

где M – количество сэмплов из референсного датасета, $x^{(k)}$ – точка между референсным $\overline{x_i}$ значением и текущим значением x, $\frac{\partial f(x^{(k)})}{\partial x_i}$ – градиент модели по *i*-му признаку.

Агрегация SHAP-значений происходила в два этапа:

Этап 1. Усреднение по всем сэмплам. Для канала j усреднение по всем N сэмплам:

$$\overline{\phi}_j = \frac{1}{N}\sum_{i=1}^{N}\left|\phi_{i,j}\right|, \quad (5)$$

Этап 2. Усреднение по каждому из каналов:

$$Shap_j = \frac{1}{T}\sum_{i=1}^{T}\overline{\phi}_j, \quad (6)$$

где T – это количество наблюдений, $\overline{\phi}_j$ – усреднённое значение Shap для каждого канала в j-наблюдении.

Такой подход позволяет абстрагироваться от локальных вариаций временны́х рядов внутри каналов и получить обобщённую оценку важности каждого канала в целом. Двойное усреднение способствует снижению влияния случайных флуктуаций в индивидуальных предсказаниях, повышая устойчивость интерпретации.

6. Эксперименты и результаты

В работе использовался комплекс стандартных и специализированных метрик для оценки качества бинарной классификации. Основные метрики: точность, precision, recall, f1-мера и ROC-AUC. Все метрики вычислялись отдельно для тестового набора данных после завершения обучения. Для расчёта использовалась библиотека scikit-learn.

Модель была обучена с использованием библиотеки PyTorch в течение 300 эпох с размером батча 16 и lrate 0.001 и использованием оптимизатора Adam и классической loss-функцией для задачи классификации – Binary Cross Entropy Loss.

При обучении унимодальных моделей для классификации сигналов fNIRS и EEG использованы аналогичные значения гиперпараметров и loss-функция.

Экспериментальная оценка классификационных моделей – одной, использующей исключительно данные EEG, и другой, сочетающей EEG с fNIRS, – выявила существенные различия в прогностических показателях.

Модель, основанная только на EEG, достигла accuracy 0,8861 при precision (0,8924) и recall (0,8722), в результате чего показатель F1-score составил 0,8822. Показатель ROC-AUC, равный 0,8858, ещё раз подтверждает высокую способность модели решать поставленную задачу. Макро- и

средневзвешенные значения точности, recall и F1-score стабилизировались на уровне 0,89, демонстрируя неизменную эффективность классификации.

Мультимодальная модель продемонстрировала незначительно меньшую точность (0,8824), хотя и с сопоставимой recall (0,8358), что дало оценку F1 в 0,8798 балла. Соотношение ROC-AUC (0,8838) было незначительно снижено по сравнению с моделью, основанной только на EEG. Макро- и средневзвешенные значения точности (0,88) незначительно уступают унимодальной модели.

Матрица ошибок приведена на рисунке 2.

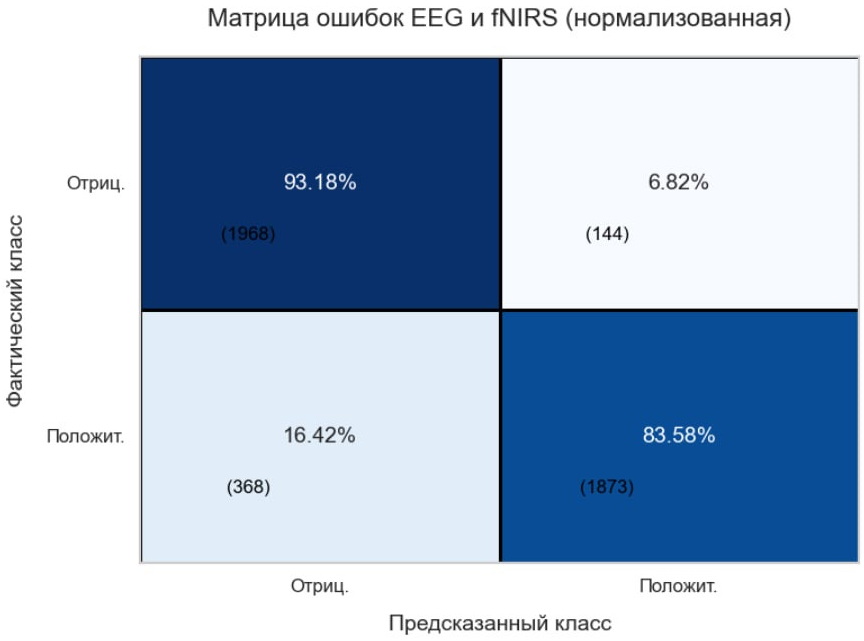


Рис 2. Матрица ошибок мультимодальной модели

Обе модели продемонстрировали высокие значения AUC (>0,87), при этом более высокая оценка по EEG-модели указывает на лучшую прогностическую способность классов. Хотя модель EEG+fNIRS сохранила высокую точность, можно отметить, что интеграция fNIRS не улучшила классификацию в рамках текущей архитектуры. Это подтверждается также и нашим исследованием SHAP значений.

Результаты визуализации значимости признаков (рисунок 3) демонстрируют, что вклад каналов fNIRS в принятие решений моделью существенно ниже по сравнению с каналами EEG. Это может говорить о недостаточно эффективном использовании fNIRS-сигналов в процессе классификации. На основании полученных данных можно заключить, что архитектура LSTM демонстрирует более высокую точность при обработке EEG-сигнала, тогда как ее эффективность для анализа fNIRS-данных оказывается ограниченной.

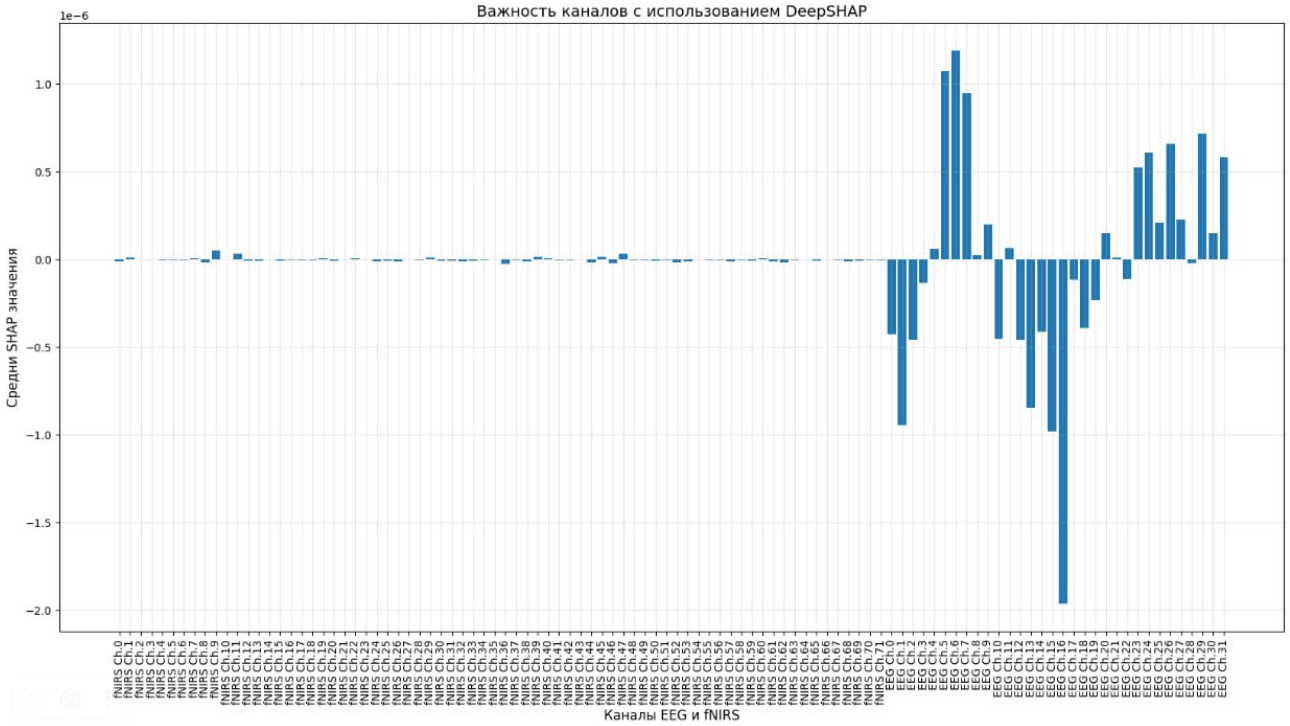
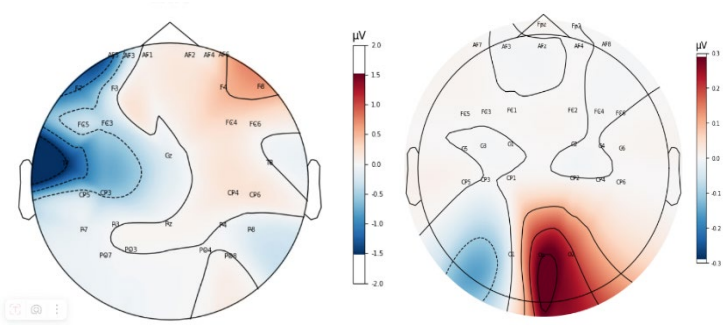


Рис 3. Важность признаков по каналам, определённая с использованием алгоритма DeepSHAP

Это объясняется также тем, что LSTM способны автоматически выделять информативные временны́е зависимости в данных, подавляя шумовые и нестабильные компоненты за счёт механизмов долгосрочного запоминания и сглаживания временны́х флуктуаций. Данное свойство является ключевым фактором, определяющим интерпретируемость классификационных результатов модели.

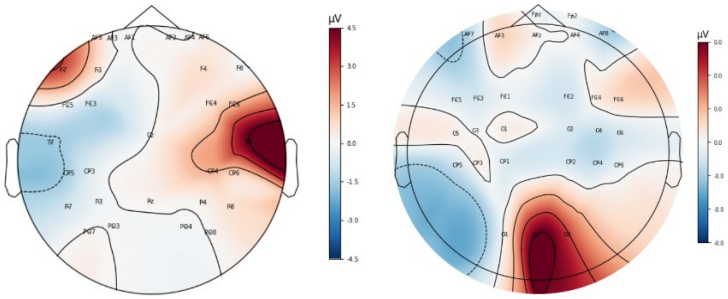
Различия в физиологических характеристиках EEG и fNIRS обуславливают необходимость применения специализированных подходов к обработке каждого типа сигналов. В связи с этим целесообразна разработка асимметричной архитектуры, в которой для каждого модального потока (EEG и fNIRS) будут использоваться специализированные подмодели, учитывающие их специфические особенности.

На рисунках 4 и 5 приведены примеры визуализаций важности каналов, нанесённых на карту размещения каналов (датчиков) на поверхности головы.



Важность каналов EEG Важность каналов fNIRS

Рис 4. Визуализация важности каналов EEG и fNIRS для задачи представления сжимания левой руки (метка 0 в датасете)



Важность каналов EEG Важность каналов fNIRS

Рис 5. Визуализация важности каналов EEG и fNIRS для задачи представления сжимания правой руки (метка 1 в датасете)

Результаты исследования хорошо согласуются с нейрофизиологическими особенностями EEG и fNIRS, а также с их ролью в решении когнитивных задач. EEG регистрирует быстрые изменения электрической активности мозга с миллисекундным разрешением, что критично для анализа динамических когнитивных процессов (например, принятия решений, внимания, рабочей памяти). LSTM, оптимизированные для временны́х зависимостей, эффективно выделяют паттерны в EEG, так как сигнал содержит чёткие временны́е маркеры (например, ERPs – связанные с событиями потенциалы). Это объясняет высокую значимость EEG-каналов в модели. fNIRS измеряет медленные гемодинамические изменения (задержка 2–6 сек.), отражающие косвенную метаболическую активность. Его временно́е разрешение недостаточно для быстрых когнитивных процессов. LSTM, несмотря на способность работать с долгосрочными зависимостями, могут "терять" полезные сигналы fNIRS из-за их низкой частотной характеристики и высокой зашумленности. Это приводит к меньшему вкладу fNIRS в классификацию.

В работе использовалось подмножество данных для решения сенсомоторных заданий, которые требовали быстрых реакций в условиях

дефицита времени. В этом случае EEG оказывается более информативным из-за связи с мгновенной нейронной активностью. fNIRS мог бы лучше отражать кумулятивные изменения в префронтальной коре при долговременных когнитивных усилиях (например, требующих устойчивого внимания или сложного обучения).

Хотя гибридные системы (EEG + fNIRS) часто рассматриваются как перспективное направление в нейроинтерфейсах, их реальная эффективность остаётся предметом дискуссий. Результаты нашего исследования показывают, что в теории комбинация EEG и fNIRS должна компенсировать недостатки каждой из модальностей и на практике требует сложной интеграции данных. При этом симметричная архитектура может ухудшить качество классификации из-за несоответствия вре́менных масштабов, а также в некоторых задачах (например требующих быстрого принятия решений) fNIRS может не давать значимого вклада, так как запаздывает относительно нейронных процессов.

Другой проблемой архитектуры с использованием LSTM для обеих модальностей является то, что LSTM хорошо работает с EEG, но может «подавлять» fNIRS из-за его низкой динамики. Решением этого может быть разработка независимых моделей для каждой модальности и объединение их в единую мультимодальную архитектуру. Однако усложнение архитектуры может привести к переобучению, особенно при малых объёмах данных.

В некоторых случаях EEG может быть избыточным: если ключевые паттерны лучше отражаются в fNIRS (например, префронтальная активность при когнитивной нагрузке), fNIRS может быть информативнее EEG. Также можно отметить, что одномодальные системы проще в развёртывании и интерпретации. Мультимодальность оправдана, только если прирост точности значим для приложения (например, медицинская диагностика).

Таким образом, мультимодальный подход (EEG + fNIRS) не всегда даёт автоматическое преимущество. Его эффективность зависит от соответствия задачи физиологическим особенностям сигналов, тонкой настройки асимметричной архитектуры, а также баланса между сложностью модели и качеством данных.

Преимущество EEG-модели в точности классификации относительно мультимодальной указывает на то, что EEG-сигналы несут более чёткие паттерны активности, релевантные для решаемой задачи. Низкий вклад fNIRS-каналов, выявленный SHAP-анализом, согласуется с их меньшей информативностью в данном контексте, что может отражать специфику вовлечённых корковых зон – вероятно, исследуемый когнитивный процесс сильнее проявляется в электрической, чем в гемодинамической активности.

Несмотря на добавление fNIRS, комбинирование сигналов не привело к синергетическому эффекту. Как уже было отмечено выше, это может объяснено разной природой временны́х зависимостей, архитектурными ограничениями и наличием скрытых корреляций между модальностями, которые не были учтены в текущей архитектуре.

В будущих исследованиях важно учитывать ряд ограничений, связанных со сложностью синхронизации данных. Мультимодальный анализ EEG и fNIRS требует точной временно́й синхронизации сигналов из-за разной природы сигналов: EEG отражает мгновенную электрическую активность (миллисекундный масштаб), тогда как fNIRS фиксирует медленные гемодинамические изменения (секундный масштаб). Это усложняет согласование времен́ных меток и может приводить к артефактам при совместном анализе. Нейроваскулярная связь подразумевает запаздывание fNIRS-сигнала относительно нейронной активности, что требует дополнительных методов временного выравнивания (например, динамического программирования или кросс-корреляции). Также необходимо учитывать технические ограничения оборудования: разные частоты дискретизации и аппаратные задержки между системами EEG и fNIRS вносят дополнительные погрешности.

Возможными вариантами преодоления ограничений может быть использование алгоритмов temporal alignment (например, Dynamic Time Warping) для коррекции вр́еменных несоответствий, использование более ярких и однозначных маркерных событий (стимулы, триггеры) для синхронизации потоков данных, а также явный учёт в архитектуре лага между нейронной и гемодинамической активностью.

С другой стороны, результаты SHAP-анализа и классификации критически зависят от этапа предобработки сигналов. Шумы и артефакты негативно влияют на качество извлекаемых моделью признаков: EEG подвержена влиянию мышечных артефактов (ЭМГ), движения глаз (ЭОГ), а fNIRS – дрейфу базовой линии и движениям головы. Неполная их коррекция может искажать значимость признаков в XAI. Стандартизация конвейеров предобработки (например, автоматическое удаление артефактов с помощью ICA для EEG и PCA для fNIRS) может помочь в преодолении этих проблем.

В то же время подавление низкочастотных компонентов в EEG может удалить полезные паттерны, релевантные для fNIRS в случае анализа в будущих асимметричных архитектурах. Это может быть решено использованием робастных алгоритмов (например, wavelet-denoising), минимизирующих потерю сигнала.

Таким образом, необходимо рассмотреть возможность разработки специализированных протоколов синхронизации для мультимодальных данных, решить вопросы автоматизации и стандартизации предобработки с открытым кодом для воспроизводимости экспериментов. Интеграция инструментов XAI в этап предобработки – например, SHAP – может помочь идентифицировать, какие этапы очистки данных сильнее влияют на итоговые предсказания. Эти шаги позволят повысить надёжность интерпретируемости моделей и избежать ложных выводов о важности признаков.

Несмотря на все ограничения и описанные выше сложности, разработанная мультимодальная архитектура и методы XAI-интерпретации

открывают новые возможности для медицинской диагностики. Так, комбинация EEG (для выявления нарушений электрической активности) и fNIRS (для мониторинга церебральной гемодинамики) может повысить точность обнаружения ишемических очагов, особенно в острых состояниях. SHAP-анализ может позволить идентифицировать критические биомаркеры (например, асимметрию кровотока или патологические паттерны), что упростит дифференциальную диагностику. fNIRS эффективен для оценки префронтальной коры, связанной с исполнительными функциями, а EEG – для анализа внимания и импульсивности. Их совместное использование может улучшить классификацию подтипов СДВГ.

Основными преимуществами данной технологии, в сравнении с МРТ, выступают три важнейших фактора: неинвазивность процедуры, мобильность аппаратуры и безопасность для пациента, дополненные возможностью непрерывного мониторинга в реальном времени. Использование методов объяснимого искусственного интеллекта (XAI) способно повысить надёжность диагностики за счёт снижения субъективности клинической интерпретации данных.

Перспективным направлением развития данного подхода является расширение функциональности системы за счёт поддержки дополнительных модальностей. В частности, включение данных фМРТ с их высоким пространственным разрешением позволит компенсировать присущие fNIRS ограничения, обеспечив трёхмерную визуализацию мозговой активности. Однако при этом возникает необходимость разработки новых алгоритмов синхронизации с EEG-данными, учитывая относительно низкое временно́е разрешение фМРТ (около 1-2 секунд).

Дальнейшее расширение диагностических возможностей может быть достигнуто путём интеграции других физиологических сигналов, таких как ЭКГ для оценки вегетативной регуляции и ЭМГ для контроля двигательных артефактов. Это откроет новые перспективы применения системы, включая области нейрореабилитации и исследований стрессовых состояний.

Особый интерес представляет создание унифицированных fusion-архитектур, способных эффективно обрабатывать данные более чем двух модальностей. Реализация этой задачи требует развития и стандартизации мультимодальных протоколов, включая формирование открытых наборов данных с синхронизированными записями EEG, fNIRS, МРТ и соответствующими метаданными, что позволит преодолеть существующие технологические и методологические ограничения.

7. Заключение

В рамках текущей архитектуры (на основе BiLSTM) модель на чистом EEG показала более высокую точность, чем гибридная EEG+fNIRS. SHAP-анализ подтвердил низкий вклад fNIRS-каналов в классификацию, что связано с разной природой сигналов (EEG (миллисекундная динамика) лучше

соответствует возможностям LSTM, чем медленные гемодинамические изменения fNIRS) и отсутствием явной синергии (симметричность архитектуры не учитывает временно́й лаг нейроваскулярной связи, что снижает полезность fNIRS).

Перспективным направлением дальнейших исследований может стать разработка механизма асимметричной обработки модальностей, а также явное моделирование задержек (например, через временно́е выравнивание или кросс-модальный attention).

XAI обеспечивает прозрачность и биологическую интерпретируемость, так как в результате SHAP-анализа выявлено, что ключевыми для сенсомоторных задач являются признаки, извлекаемые из сигналов в EEG-каналах, а не fNIRS, что соответствует известным нейрофизиологическим маркерам. fNIRS при этом имеет ограниченную значимость в данного типа задачах и согласуется с его низким временны́м разрешением.

Таким образом, можно сказать, что XAI-методы полезны для понимания ограничений модели (например, доминирование признаков EEG) и обоснования решений для клинических приложений.

Несмотря на выявленные ограничения и сложности, предложенный подход обладает потенциалом для формирования нового поколения интерпретируемых интерфейсов «мозг-компьютер». Его адаптация к задачам клинической практики и расширение за счёт интеграции дополнительных модальностей требует тесного междисциплинарного взаимодействия между нейрофизиологами, специалистами по машинному обучению и клиницистами. Такой подход может стать основой для трансформации персонализированной медицины.

Библиографический список

1. Foong R., et al., Assessment of the efficacy of EEG-based MI-BCI with visual feedback and EEG correlates of mental fatigue for upper-limb stroke rehabilitation, IEEE Trans. Biomed. Eng. 67 (3). 2019. pp. 786–795.
2. Valente G., et al., Optimizing fMRI experimental design for MVPA-based BCI control: combining the strengths of block and event-related designs, NeuroImage 186. 2019. pp. 369–381.
3. Roy S., et al., Channel selection improves meg-based brain-computer interface, in: 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE, 2019.
4. Li C., et al., A between-subject fNIRS-BCI study on detecting self-regulated intention during walking, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (2). 2020. pp. 531–540.
5. Aslam M., Rajbdad F., Azmat S., Perveen K., Naraghi-Pour M., Xu J. Electroencephalograph (EEG) based classification of mental arithmetic using explainable machine learning. Biocybernetics and Biomedical Engineering. 2025. Apr 1;45(2). pp. 154-69.

6. Andreu-Perez J., Emberson L.L., Kiani M., Filippetti M.L., Hagras H., Rigato S. Explainable artificial intelligence based analysis for interpreting infant fNIRS data in developmental cognitive neuroscience. Communications biology. 2021. Sep. 15; 4(1):1077.
7. Morabito F.C., Ieracitano C., Mammone N. An explainable Artificial Intelligence approach to study MCI to AD conversion via HD-EEG processing. Clinical EEG and neuroscience. 2023. Jan; 54(1). pp. 51-60.
8. Li R., Yang D., Fang F., Hong K.S., Reiss A.L., Zhang Y. Concurrent fNIRS and EEG for brain function investigation: a systematic, methodology-focused review. Sensors. 2022. Aug 5;22(15):5865.
9. Yang L.I., Zhang X., Dong M.I. Early-stage fusion of EEG and fNIRS improves classification of motor imagery. Frontiers in Neuroscience. 2023. 16:1062889.
10. Yeung M.K., Chu V.W. Viewing neurovascular coupling through the lens of combined EEG–fNIRS: A systematic review of current methods. Psychophysiology. 2022. Jun;59(6): e14054.
11. Liu Z., Shore J., Wang M., Yuan F., Buss A., Zhao X. A systematic review on hybrid EEG/fNIRS in brain-computer interface. Biomedical Signal Processing and Control. 2021. Jul 1; 68:102595.
12. Zhou X., Sobczak G., McKay C.M., Litovsky R.Y. Comparing fNIRS signal qualities between approaches with and without short channels. PloS one. 2020. Dec 23;15(12): e0244186.
13. Hossain K.M., Islam M.A., Hossain S., Nijholt A., Ahad M.A. Status of deep learning for EEG-based brain–computer interface applications. Frontiers in computational neuroscience. 2023. Jan 16;16:1006763.
14. Qiu L., Zhong Y., He Z., Pan J. Improved classification performance of EEG-fNIRS multimodal brain-computer interface based on multi-domain features and multi-level progressive learning. Frontiers in Human Neuroscience. 2022. Aug 4;16:973959.
15. Mughal N.E., Khan M.J., Khalil K., Javed K., Sajid H., Naseer N., Ghafoor U., Hong K.S. EEG-fNIRS-based hybrid image construction and classification using CNN-LSTM. Frontiers in Neurorobotics. 2022. Aug 31;16:873239.
16. Kumar C., Rahimi N., Gonjari R., McLinden J., Hosni S.I., Shahriari Y., Shao M. Context-aware multimodal auditory bci classification through graph neural networks. In: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) 2023. Jul 24. pp. 1-4.
17. Maher A., Qaisar S.M., Salankar N., Jiang F., Tadeusiewicz R., Pławiak P., Abd El-Latif A.A., Hammad M. Hybrid EEG-fNIRS brain-computer interface based on the non-linear features extraction and stacking ensemble learning. biocybernetics and biomedical engineering. 2023. Apr 1;43(2):463-75.
18. Hussain I., Jany R., Boyer R., Azad A.K., Alyami S.A., Park S.J., Hasan M.M., Hossain M.A. An explainable EEG-based human activity recognition model using machine-learning approach and LIME. Sensors. 2023& Aug 27;23(17):7452.

19. Shibu C.J., Sreedharan S., Arun K.M., Kesavadas C., Sitaram R. Explainable artificial intelligence model to predict brain states from fNIRS signals. Frontiers in Human Neuroscience. 2023. Jan 19;16:1029784.
20. Shibu C.J., Sreedharan S., Arun K.M., Kesavadas C., Sitaram R. Explainable artificial intelligence model to predict brain states from fNIRS signals. Frontiers in Human Neuroscience. 2023. Jan 19;16:1029784.
21. Liu B., Guo J., Chen C.P., Wu X., Zhang T. Fine-grained interpretability for EEG emotion recognition: Concat-aided grad-CAM and systematic brain functional network. IEEE Transactions on Affective Computing. 2023. Jun 23;15(2):671-84.
22. Open access dataset for simultaneous EEG and NIRS Brain-Computer Interfaces (BCIs) [Электронный ресурс] // Technische Universität Berlin. Режим доступа: https://doc.ml.tu-berlin.de/hBCI/contactthanks.php (дата обращения: 21.05.2025).
23. Lundberg S.M., Lee S.I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). 2017. Curran Associates Inc., Red Hook, NY, USA, pp. 4768–4777.

Оглавление