

ФИЗИОЛОГИЯ ВЫСШЕЙ НЕРВНОЙ (КОГНИТИВНОЙ) ДЕЯТЕЛЬНОСТИ ЧЕЛОВЕКА

УДК 612.821.6

ПОИСК НАИБОЛЕЕ ИНФОРМАТИВНЫХ РЕГИОНОВ ДЛЯ БИНАРНОЙ КЛАССИФИКАЦИИ ШИЗОФРЕНИИ ПО ДАННЫМ ФМРТ СОСТОЯНИЯ ПОКОЯ НА ОСНОВЕ МЕТОДА ВЫДЕЛЕНИЯ ФУНКЦИОНАЛЬНО ОДНОРОДНЫХ РЕГИОНОВ

© 2024 г. А. Д. Жемчужников¹, С. И. Карташов¹, С. О. Козлов¹, В. А. Орлов¹,
А. А. Пойда^{1, *}, Н. В. Захарова^{2, 3}, Л. В. Бравве⁴, Г. Ш. Мамедова⁴, М. А. Кайдан⁴

¹Национальный исследовательский центр «Курчатовский институт», Москва, Россия

²Международный научно-образовательный центр нейropsychиатрии

Самарского государственного медицинского университета, Самара, Россия

³Национальный медицинский исследовательский центр психиатрии и неврологии им. В.М. Бехтерева,
Санкт-Петербург, Россия

⁴ГБУЗ «Психиатрическая клиническая больница № 1 им. Н.А. Алексеева
Департамента здравоохранения города Москвы», Москва, Россия

*e-mail: Poyda_AA@nrcki.ru

Поступила в редакцию 16.01.2024 г.

После доработки 22.04.2024 г.

Принята к публикации 22.07.2024 г.

В работе решается задача автоматической бинарной классификации участников эксперимента с диагнозом шизофрении и группы контроля на наборе данных, полученном на томографе *Siemens Magnetom Verio* 3Тл. Набор включал данные 36 участников эксперимента, проходящих лечение в «ГБУЗ ПКБ №1 ДЗМ», и 36 участников эксперимента из группы контроля. Для решения поставленной задачи были применены методы машинного обучения. В результате была достигнута точность разделения 76%, что соответствует результатам, полученным в других научных исследованиях. Наиболее высокая точность получена для уже известного в литературе параметра локальной однородности (*regional homogeneity* – *ReHo*). Разработанный авторами набор признаков на основе метода выделения функционально однородных регионов (*Functionally Homogeneous Regions* – *FHR*) обеспечил достижение максимальной точности классификации 74%. Но при этом набор признаков *FHR* обеспечивает более высокую точность классификации при использовании малого числа регионов мозга. Так, например, уже на 8 регионах набор *FHR* обеспечил почти максимальную точность классификации – 72.5% (против 65% у набора *ReHo*), что позволяет предположить, что именно выделенные 8 регионов дают наиболее высокий уровень разделения.

Ключевые слова: автоматическая классификация шизофрении, фМРТ, состояние покоя, методы машинного обучения

DOI: 10.31857/S0044467724040035

ВВЕДЕНИЕ

В настоящее время в психиатрии существует несколько актуальных задач по разработке алгоритмов и методов для объективизации поставленного диагноза у пациентов с шизофренией и прогнозирования эффективности применяемой терапии. Согласно международной классификации болезней (МКБ-11), для этого расстройства упразднены категориальные подтипы, что может говорить о возможности существования характерных

нейрофизиологических маркеров, определяющих особенности патогенеза. В связи с этим авторами статьи была поставлена цель — создание классификатора для автоматического выявления признаков шизофрении по данным функциональной МРТ состояния покоя с использованием методов машинного обучения.

Исследования структурно-функциональных особенностей головного мозга методами машинного обучения перспективны в нескольких аспектах. Во-первых, при установлении прогностически

значимых биомаркеров психических расстройств. Применение нейронных сетей помогло выявить аномалии головного мозга при шизофрении с точностью от 81.5% (Cui et al., 2022) и до 84% при оценке признаков снижения плотности серого вещества в островковой доле, лобной и верхней височной долях и уменьшения белого вещества в поясной извилине и гиппокампе (Chang et al., 2020). Точность прогноза эффективности терапии у нелеченных ранее больных достигает 82.5% (Cao et al., 2020).

С точностью 82% можно оценить прогноз выздоровления для каждого пациента по результатам машинного обучения в рамках анализа структурной МРТ с учетом уменьшения площади верхней височной, нижней лобной и нижней теменной областей (Kambeitz-Pankovic et al., 2016).

Высокая ценность относительно отдаленного (1 год) прогноза функционирования показана как для линейного (метод опорных векторов — 75%), так и нелинейного (дерево решений — 90%) алгоритмов машинного обучения, применимых для анализа динамики функциональной связности сети режима по умолчанию в состоянии покоя (Kottaram et al., 2020).

Во-вторых, методы машинного обучения используются для классификации участников эксперимента по признаку шизофрении. Например, в работе (Antonucci et al., 2020) использовался метод опорных векторов для разделения пациентов с шизофренией и группы контроля по данным фМРТ, полученным при выполнении задачи на контроль внимания. В результате была достигнута точность 66.9%.

В работе (Kalmady et al., 2019) использовался метод опорных векторов для разделения больных шизофренией и группы контроля по данным фМРТ состояния покоя. Ключевой особенностью работы является одновременное использование нескольких атласов для уменьшения размерности анализируемых данных. По каждому атласу получался набор признаков, к совокупности которых применялся метод опорных векторов. В результате была получена точность 87%.

Аналогичные показатели точности (80%) были достигнуты при анализе методами машинного обучения объединенных результатов структурной и функциональной МРТ (Gutiérrez-Gómez et al., 2020). Однако при объединении маркеров полигенного риска и структурной МРТ точность получилась ниже (71.6%) (Hu et al., 2021).

Как видно из обзора, разные исследователи рапортуя о разной точности в зависимости как от выбранных методов, так и от набора данных. Поэтому в данной работе мы поставили задачу бинарной классификации участников эксперимента с диагнозом шизофрении и группы контроля на наборе данных, который мы сами получили

на томографе *Siemens 3 Тл* в НИЦ «Курчатовский институт» на участниках эксперимента, проходящих лечение в «ГБУЗ ПКБ № 1 ДЗМ». Решение анализировать собственные данные (а, не, например, выложенные в международных базах данных) было принято в связи с возможностью проведения повторных исследований на тех же пациентах в процессе прохождения терапии. Такой подход в перспективе может стать основой для разработки математического аппарата для прогнозирования вероятности успеха терапии и оценки ее эффективности. В дополнение к существующим исследованиям, где в основном используются методы нейронных сетей и опорных векторов, мы используем и другие методы машинного обучения, такие как *K-ближайших соседей (K-Neighbors)*, чрезвычайно рандомизированные леса (*Extremely Randomized Trees*), деревья решений (*Decision Tree*) и другие (всего 38 методов классификации).

Для формирования наборов признаков для классификации мы использовали как известные в литературе методы (ALFF, fALFF, ReHo), так и собственный ранее разработанный метод выделения функционально однородных регионов (FHR) (Kozlov et al., 2020, 2021), который учитывает размер регионов скоррелированности каждого вокселя.

МЕТОДИКА

Пациенты

Обследованы 36 пациентов (20 мужчин и 16 женщин, средний возраст 28.9 ± 7.3 лет) из числа госпитализированных в острые отделения «ГБУЗ ПКБ № 1 ДЗМ» с диагнозами расстройств шизофренического спектра (F20 по МКБ-10 и 6A20 по МКБ-11) в 2018–2019 гг. Диагностика состояния проводилась в течение двух дней двумя опытными психиатрами с привлечением всех необходимых данных (опрос родственников, анализ медицинских карт, результаты физических и лабораторных тестов и т.д.).

Критерии включения: возраст от 21 до 35 лет (ср. возраст 28.2 ± 7.4 лет), соответствие состояния критериям шизофрении по МКБ-10, праворукость, критика к собственному состоянию, сохранение памяти о психотических симптомах, информированное согласие на участие в исследовании.

Критерии исключения: органические заболевания головного мозга, тяжелые соматические и/или неврологические состояния, потенциально влияющие на физиологию или структуру мозга, признаки злоупотребления психоактивными веществами (ПАВ) и общие противопоказания для получения данных МРТ, отказ от участия.

Группу контроля составили 36 добровольцев (19 мужчин, 17 женщин, ср. возраст 28.9 ± 6.2 лет),

не состоящих в родстве с пациентами, обследованные по единому протоколу с пациентами.

МРТ-сканирование

Получение данных проводилось на базе ресурсного центра ядерно-физических методов исследований НИЦ «Курчатовский институт» на магнитно-резонансном томографе (МРТ) *Magnetom Verio* (*Siemens*, Германия) напряженностью магнитного поля в 3 Тесла, с использованием 32-канальной головной катушки.

В результате сканирования на МРТ записаны:

– анатомическая 3D-T1-взвешенная последовательность со следующими параметрами: 176 срезов, $TR = 2530$ мс, $TE = 3.31$ мс, толщина срезов = 1 мм, угол отклонения = 7° , время инверсии = 1200 мс и $FOV = 256 \times 256$ мм²;

– фМРТ-последовательность со следующими параметрами: 56 срезов, $TR = 750$ мс, $TE = 33$ мс и изотропным вокселем $2 \times 2 \times 2$ мм³, 900 временных точек, что эквивалентно 648 секундам.

Последовательность фМРТ использована для регистрации состояния покоя, во время которого участнику эксперимента необходимо было лежать неподвижно, с закрытыми глазами, стараться не заснуть и стараться не фокусироваться на одной и той же мысли, избегая мысленного счета и фокусирования внимания на шуме работающего томографа. В отсутствие внешних стимулов при изучении состояния покоя исследователи фокусируются на спонтанной нейрональной активности, отражаемой BOLD-сигналом (blood

oxygen level dependent – зависящий от уровня оксигенации), регистрируемым с помощью фМРТ.

За последние два десятилетия было разработано большое количество методов для анализа данных фМРТ состояния покоя. Наиболее часто используемыми являются: анализ функциональной связности (как правило, на основе расчета коэффициента корреляции Пирсона) (Biswal et al., 1995), метод независимых компонент (Calhoun et al., 2009), теория графов (где каждый регион мозга представляется в виде узла, а связи с другими регионами – в виде ребер графа) (Bassett, Sporns, 2017) и методы на основе анализа сетей состояния покоя (Rosazza, Minati, 2011).

Схема анализа данных

Процесс анализа состоит из трех основных шагов и схематически представлен на рис. 1.

На первом шаге выполняется предобработка данных с целью уменьшить их искажения (например, возникшие из-за движений участников эксперимента во время сканирования) и вклад шумов различной природы (например, физиологических).

На втором шаге из очищенных данных формируются различные наборы признаков, т.е. параметров, на основании значений которых будет осуществляться разделение участников эксперимента на группы. Для формирования наборов признаков в данной работе были использованы известные в литературе методы: амплитуда низкочастотных флуктуаций (amplitude of low-frequency fluctuation – ALFF) (Zou et al., 2008), дробная

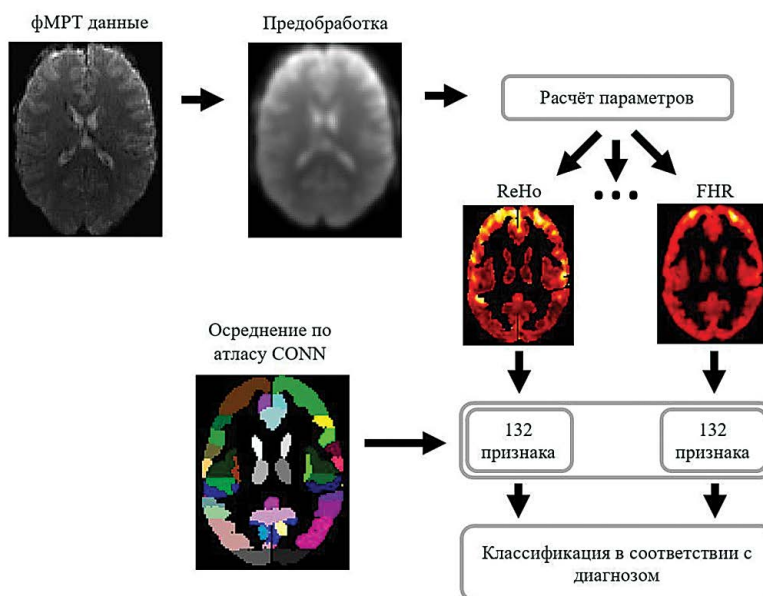


Рис. 1. Схема анализа фМРТ-данных для классификации.

Fig. 1. A scheme of fMRI data analysis for classification.

амплитуда низкочастотных флуктуаций (fractional ALFF – fALFF) (Xu et al., 2018), локальная однородность (ReHo) (Zang et al., 2004), а также разработанный ранее авторами метод функциональной однородности (FHR) (Kozlov et al., 2020, 2021).

На третьем шаге применяются методы машинного обучения для разделения участников эксперимента на группы. Мы пробовали различные методы, включая нейронные сети и метод опорных векторов (SVM). Но на нейронных сетях мы на текущий момент не смогли получить стабильных результатов, что связываем в первую очередь с малым числом участников эксперимента. Поэтому в данную статью описание исследований с использованием нейронных сетей не включено. Так как набор данных для обучения мал (всего 72 участника эксперимента), то для валидации обученных классификаторов использовался подход перекрестной проверки (cross-validation).

Предобработка данных

Предобработка фМРТ-данных проведена на основе свободно распространяемого программного пакета SPM12 в среде Matlab (R2023b Update 4) и специально разработанных терминальных скриптов системы MacOS. Структурные и функциональные данные приводились к центру в передней комиссуре вручную. Далее производился расчет и коррекция артефактов движения. С помощью записанных во время исследования карт неоднородности магнитного поля проведена коррекция функциональных данных с целью удаления артефактов магнитной восприимчивости. Структурные и функциональные МРТ-объемы были нормализованы на шаблонные изображения в MNI-пространстве. В работе для регистрации фМРТ использованы мультисрезовые (ультрабыстрые) последовательности (Larkman et al., 2001), полученные в НИЦ «КИ» на МР-томографе Siemens Magnetom Verio 3 Тл. Ультрабыстрые последовательности для получения фМРТ-данных имеют более высокое временное разрешение, а потому более чувствительны для обнаружения физиологических источников шумов в процессе временной фильтрации (Todd et al., 2017).

Формирование наборов признаков

Для расчета параметров *ReHo*, *ALFF*, *fALFF* мы использовали программный пакет *BRANT* (Xu et al., 2018), написанный на языке *MATLAB*. В качестве атласа был использован атлас программного пакета *CONN* (Nieto-Castanon, Whitfield-Gabrieli, 2022), включающий 132 региона и комбинирующий кортикальный и субкортикальный атласы *Harvard-Oxford* пакета *FSL* и атлас *AAL* для мозжечка. Далее в статье будем называть этот атлас «*CONN*».

Для формирования набора признаков на основе рассчитанных значений параметров *ReHo*, *ALFF*, *fALFF* мы использовали результат работы *BRANT* без усреднения (*raw*-версия). Мы усреднили значения рассчитанных параметров по регионам атласа *CONN*, получив тем самым для каждого участника эксперимента и каждого параметра числовой вектор, состоящий из 132 значений (признаков), где каждое значение соответствует одному региону атласа.

Для расчета параметра функциональной однородности использован разработанный авторами метод выделения функционально однородных регионов (сокращенно *FHR*), который позволяет выделить по данным фМРТ индивидуальные непересекающиеся пространственно-связанные регионы с высокой внутренней корреляцией (Kozlov et al., 2020, 2021). Метод *FHR* состоит из 3 основных шагов.

На первом шаге выделяются зоны однородности для каждого вокселя. Зоной однородности заданного вокселя мы называем максимальную по размеру пространственно-связную область (включающую заданный воксель), в которой каждый воксель имеет высокий уровень корреляции с заданным. Под пространственно-связной областью мы понимаем такой набор вокселей, между любой парой которых можно построить путь по соседним вокселям из этого набора. Соседние воксели – это воксели, имеющие общую грань. Для расчета уровня корреляции между двумя вокселями мы рассчитываем коэффициент корреляции Пирсона между динамиками этих вокселей. Мы считаем, что уровень корреляции высокий, если коэффициент корреляции Пирсона превышает значение 0.5. В работе мы пробовали и другие значения: 0.4, 0.6, 0.7, но при значении 0.5 точность классификации выше, чем при других вариантах.

Пример зоны однородности для уровня корреляции 0.5 приведен на рис. 2. На рисунке показана зона однородности вокселя А. Все воксели внутри этой зоны имеют корреляцию с вокселем А более 0.5. Зона имеет максимальный размер, потому что в нее не может быть добавлен ни один новый воксель: соседние воксели не удовлетворяют условию высокой корреляции, а добавление вокселей с высокой корреляцией (например, вокселя С) нарушит условие пространственной связности.

На втором и третьем шагах происходит фильтрация выделенных зон однородности по признаку взаимного перекрытия, а также разделение вокселей между оставшимися после фильтрации зонами с целью обеспечения их непересекаемости.

Но для расчета параметра функциональной однородности мы используем результаты первого шага. Для этого мы каждому вокселю ставим в соответствие размер его зоны однородности, после чего усредняем полученные значения по регионам

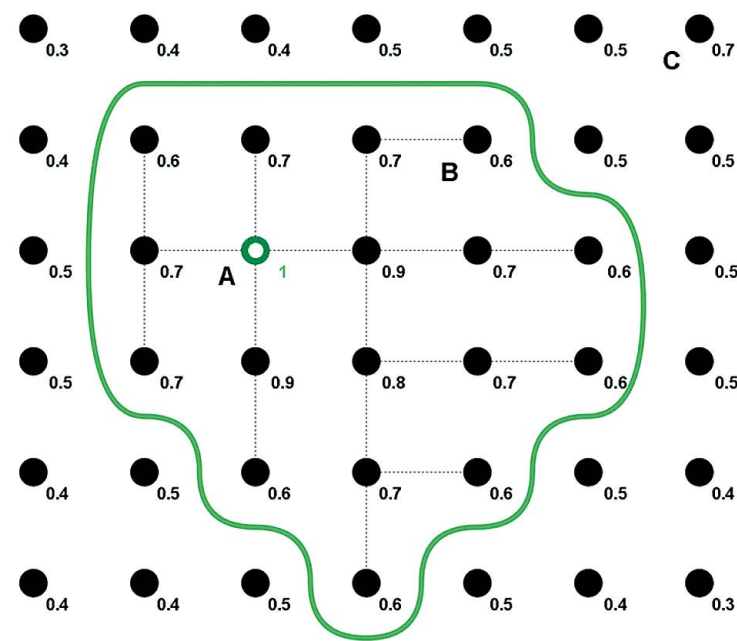


Рис. 2. Пример зоны однородности (двумерный срез) для вокселя А. Черные кружки обозначают воксели, цифры под ними — уровень корреляции с вокселем А, пунктирные линии между вокселями соответствуют пространственной связности между вокселями с высокой корреляцией, линия обозначает границу зоны однородности.

Fig. 2. An example of a homogeneity zone (two-dimensional slice) for voxel A. Black circles indicate voxels, numbers below them indicate the level of correlation with voxel A, dotted lines between voxels correspond to spatial connectivity between voxels with high correlation, and a line indicates the boundary of the homogeneity zone.

атласа *CONN* аналогично тому, как это сделано для параметра *ReHo*. В результате получаем для каждого участника эксперимента числовой вектор из 132 значений. Далее в статье будем называть этот набор признаков по названию метода: *FHR*.

Известно (Tibshirani, 2001), что большинство использованных для анализа классификаторов (см. следующий подраздел) чувствительны к различиям средних значений и дисперсий признаков. Для нивелирования данного фактора мы использовали два алгоритма предобработки данных: стандартизацию и нормализацию. Стандартизация изменяет масштаб набора данных, чтобы среднее значение было равно 0, а стандартное отклонение 1. Нормализация масштабирует набор данных таким образом, чтобы каждое значение находилось в диапазоне от 0 до 1. Преобразование линейно, минимальное значение становится 0, максимальное 1.

Для каждого из 4 наборов признаков (*ReHo*, *ALFF*, *fALFF*, *FHR*) мы сделали по 2 дополнительные модификации, применив к ним стандартизацию или нормализацию. Таким образом, для последующей классификации было сформировано 12 наборов признаков: *ReHo*, *stand_ReHo*, *norm_ReHo*, *FHR*, *stand_FHR*, *norm_FHR*, *ALFF*, *norm_ALFF*, *stand_ALFF*, *fALFF*, *norm_fALFF*, *stand_fALFF* (префиксы *stand_* и *norm_* обозначают примененную предобработку стандартизации или нормализации).

Разделение участников эксперимента на группы с использованием различных классификаторов

Для анализа полученных наборов признаков было использовано 38 методов классификации (Wolf, 2022), реализованных в программной библиотеке *scikit-learn* (Pedregosa et al., 2011) на языке программирования Python 3.

Одной из особенностей решаемой задачи является то, что число участников эксперимента (72) меньше размерности пространства признаков (132). Выбранные 38 классификаторов не имеют ограничений на соотношение числа признаков и количества измерений, но от этого может зависеть их эффективность, что для ряда классификаторов выразилось в том, что их точность оказалась близка к уровню случайной оценки, а иногда метод вообще не сходил к решению, и вычислительный процесс завершался с ошибкой (см. раздел «Результаты исследования»).

Для оценки точности выбранного классификатора применялся метод перекрестной проверки (*k-fold cross-validation*) (Arlot, Celisse, 2010). Данный метод используется при небольшом количестве исходных данных с целью получения более точного результата. На каждом цикле перекрестной проверки 72 участника эксперимента делились случайным образом на обучающую выборку (64 набора) и проверочную

(8 наборов). Важное условие – в проверочной выборке должны быть наборы из обоих классов. На обучающей выборке проводилось обучение классификатора, точность которого затем вычислялась на проверочной выборке. Мы использовали 1000 циклов перекрестной проверки, рассчитывая результирующую точность, специфичность и чувствительность по формулам:

$$\begin{aligned}\text{Чувствительность} &= \frac{\sum_{i=1}^n \frac{\text{ИП}_i}{\text{ИП}_i + \text{ЛО}_i}}{n}, \\ \text{Специфичность} &= \frac{\sum_{i=1}^n \frac{\text{ИП}_i}{\text{ИП}_i + \text{ЛП}_i}}{n}, \\ \text{Точность} &= \frac{\sum_{i=1}^n \frac{\text{ИП}_i + \text{ИО}_i}{\text{ИО}_i + \text{ИП}_i + \text{ЛП}_i + \text{ЛО}_i}}{n}.\end{aligned}$$

Здесь n – число итераций (в нашем случае 1000), ИП и ИО – число истинно положительных и истинно отрицательных результатов в каждом цикле перекрестной проверки соответственно, ЛО и ЛП – число ложноотрицательных и ложноположительных результатов в каждом цикле перекрестной проверки соответственно. Индекс i соответствует номеру итерации.

Большое число признаков для классификации может приводить к ухудшению общей точности классификатора, из-за потери полезных данных в «шуме», приводящему к снижению точности. В этом случае может помочь подход, при котором в анализ передаются не все признаки, а только наиболее значимые. Кроме того, подобный подход может улучшить интерпретируемость результата и показать признаки, наиболее значимые для классификации. Поэтому в анализ была включена опция выбора и использования наиболее значимых признаков.

Существует множество алгоритмов для выбора наиболее значимых признаков. В данной работе применялось 3 алгоритма (Pudjihartono et al., 2022): AdaBoost (Freund, Schapire, 1999), ExtraTreesClassifier (Sharaff, Gupta, 2019) и алгоритм, основанный на L2-норме. Данные алгоритмы позволяют определить вес каждого признака, после чего можно отфильтровать малозначимые (обладающие низкими весами) значения.

Таким образом, в процессе анализа мы перебирали вариации по следующим параметрам классификации.

- Набор признаков (*ReHo*, *ALFF*, *fALFF*, *FHR*).
- Алгоритм предобработки данных (нормализация, стандартизация, без предобработки).
- Метод классификации (38 классификаторов перечислены в таблице с результатами). Каждый классификатор может иметь собственные

конфигурационные параметры. В рамках данного исследования перебор параметров классификаторов не проводился. Использовались базовые параметры, предоставляемые программной реализацией по умолчанию.

– Алгоритм выбора наиболее значимых признаков (*AdaBoost*, *ExtraTrees*, L2-норма, без фильтрации признаков).

Чтобы сократить вычислительную сложность, мы разбили анализ на два шага. На первом шаге перебирали комбинации («набор признаков» + «алгоритм предобработки данных» + «метод классификации»). Целью первого шага было зафиксировать алгоритм предобработки данных для всего последующего анализа. Первый шаг показал, что: 1) предобработка при помощи алгоритма нормализации, а также вариант без предобработки позволили получить почти одинаковую максимальную точность классификации для всех исследуемых наборов признаков (различия в десятых долях процента); 2) предобработка при помощи алгоритма стандартизации в общем случае снизила максимальную точность классификации (для набора FHR – на 1%, для набора fALFF – на 2%, для наборов ReHo и ALFF точность не изменилась). Таким образом, из двух алгоритмов предобработки данных мы выбрали алгоритм нормализации. А при выборе между вариантом с предобработкой и без предобработки мы учли, что, несмотря на одинаковую достигнутую точность, проведение предобработки данных перед классификацией – методически рекомендуемый подход (Jo, 2019). Поэтому на втором шаге анализа использовались только 4 из 12 наборов признаков: norm_ReHo, norm_FHR, norm_ALFF, norm_fALFF. Чтобы упростить сокращения в статье, далее по тексту под обозначениями ReHo, FHR, ALFF, fALFF будут подразумеваться именно их нормированные варианты.

На втором шаге анализа для каждого из 4 наборов признаков перебирались комбинации («метод классификации» + «алгоритм выбора наиболее значимых признаков»).

Для перебора возможных комбинаций и расчета результирующей точности использовалась разработанная авторами программная платформа на языке программирования Python 3, позволяющая настроить параметры перекрестной проверки, а также автоматически перебрать все выбранные классификаторы. Данная программная платформа позволяет работать с выбранным набором данных и/или их объединением, перебирать все выбранные классификаторы, алгоритмы выбора признаков, а также число используемых признаков.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

В таблице 1 представлена точность, достигнутая на анализируемых наборах признаков.

Таблица 1. Точность, чувствительность, специфичность классификации участников эксперимента по наборам признаков norm_ReHo, norm_ALFF, norm_fALFF, norm_FHR. Классификаторы упорядочены по убыванию значения точности, усредненного между 4 наборами признаков. Значения округлены до сотых. Максимальная точность для каждого набора выделена жирным шрифтом

Table 1. Accuracy, sensitivity, specificity of the classification of subjects according to the sets of features norm_ReHo, norm_ALFF, norm_fALFF, norm_FHR. The classifiers are ordered in descending order of the accuracy value averaged between 4 sets of features. The values are rounded to hundredths. The maximum accuracy for each set is highlighted in bold

Model	ALFF	fALFF	FHR	ReHo
LogisticRegression	T: 0.71	T: 0.69	T: 0.70	T: 0.76
	Ч: 0.68	Ч: 0.69	Ч: 0.72	Ч: 0.78
	C: 0.74	C: 0.71	C: 0.69	C: 0.76
RidgeClassifierCV	T: 0.68	T: 0.70	T: 0.70	T: 0.76
	Ч: 0.67	Ч: 0.70	Ч: 0.72	Ч: 0.80
	C: 0.70	C: 0.71	C: 0.68	C: 0.73
LinearSVC	T: 0.73	T: 0.67	T: 0.67	T: 0.75
	Ч: 0.69	Ч: 0.66	Ч: 0.66	Ч: 0.77
	C: 0.77	C: 0.68	C: 0.69	C: 0.75
PassiveAggressiveClassifier	T: 0.72	T: 0.66	T: 0.68	T: 0.75
	Ч: 0.69	Ч: 0.66	Ч: 0.67	Ч: 0.76
	C: 0.75	C: 0.68	C: 0.69	C: 0.76
LogisticRegressionCV	T: 0.69	T: 0.68	T: 0.69	T: 0.75
	Ч: 0.67	Ч: 0.69	Ч: 0.70	Ч: 0.77
	C: 0.72	C: 0.68	C: 0.70	C: 0.74
MLPClassifier	T: 0.72	T: 0.65	T: 0.69	T: 0.73
	Ч: 0.71	Ч: 0.67	Ч: 0.71	Ч: 0.76
	C: 0.74	C: 0.63	C: 0.70	C: 0.72
CalibratedClassifierCV	T: 0.69	T: 0.67	T: 0.68	T: 0.74
	Ч: 0.67	Ч: 0.69	Ч: 0.70	Ч: 0.78
	C: 0.74	C: 0.67	C: 0.68	C: 0.74
NuSVC	T: 0.73	T: 0.65	T: 0.68	T: 0.69
	Ч: 0.77	Ч: 0.65	Ч: 0.71	Ч: 0.76
	C: 0.71	C: 0.67	C: 0.65	C: 0.66
Perceptron	T: 0.65	T: 0.67	T: 0.69	T: 0.74
	Ч: 0.65	Ч: 0.69	Ч: 0.70	Ч: 0.76
	C: 0.66	C: 0.66	C: 0.69	C: 0.73
SGDClassifier	T: 0.65	T: 0.67	T: 0.69	T: 0.74
	Ч: 0.66	Ч: 0.70	Ч: 0.66	Ч: 0.76
	C: 0.65	C: 0.65	C: 0.74	C: 0.73
RidgeClassifier	T: 0.68	T: 0.67	T: 0.68	T: 0.70
	Ч: 0.70	Ч: 0.67	Ч: 0.68	Ч: 0.74
	C: 0.67	C: 0.67	C: 0.69	C: 0.68
AdaBoostRegressor	T: 0.70	T: 0.66	T: 0.67	T: 0.68
	Ч: 0.73	Ч: 0.78	Ч: 0.63	Ч: 0.68
	C: 0.69	C: 0.56	C: 0.72	C: 0.70
ExtraTreesClassifier	T: 0.70	T: 0.61	T: 0.69	T: 0.67
	Ч: 0.70	Ч: 0.59	Ч: 0.68	Ч: 0.72
	C: 0.73	C: 0.65	C: 0.73	C: 0.65
SVC	T: 0.71	T: 0.62	T: 0.68	T: 0.66
	Ч: 0.75	Ч: 0.62	Ч: 0.72	Ч: 0.79
	C: 0.70	C: 0.64	C: 0.66	C: 0.57
LinearDiscriminantAnalysis	T: 0.65	T: 0.70	T: 0.62	T: 0.68
	Ч: 0.65	Ч: 0.71	Ч: 0.61	Ч: 0.70
	C: 0.66	C: 0.69	C: 0.64	C: 0.66
GaussianNB	T: 0.64	T: 0.60	T: 0.66	T: 0.70
	Ч: 0.46	Ч: 0.50	Ч: 0.66	Ч: 0.81
	C: 0.83	C: 0.70	C: 0.68	C: 0.60

Окончание таблицы 1

Model	ALFF	fALFF	FHR	ReHo
RandomForestClassifier	T: 0.70	T: 0.60	T: 0.65	T: 0.64
	Ч: 0.72	Ч: 0.61	Ч: 0.61	Ч: 0.70
	C: 0.70	C: 0.62	C: 0.72	C: 0.60
NearestCentroid	T: 0.67	T: 0.60	T: 0.65	T: 0.64
	Ч: 0.62	Ч: 0.60	Ч: 0.49	Ч: 0.72
	C: 0.72	C: 0.61	C: 0.81	C: 0.56
HistGradientBoostingClassifier	T: 0.64	T: 0.61	T: 0.65	T: 0.64
	Ч: 0.64	Ч: 0.65	Ч: 0.64	Ч: 0.65
	C: 0.65	C: 0.59	C: 0.69	C: 0.67
BernoulliNB	T: 0.68	T: 0.59	T: 0.64	T: 0.62
	Ч: 0.67	Ч: 0.58	Ч: 0.53	Ч: 0.67
	C: 0.70	C: 0.60	C: 0.76	C: 0.58
KNeighborsClassifier	T: 0.64	T: 0.61	T: 0.65	T: 0.63
	Ч: 0.60	Ч: 0.51	Ч: 0.58	Ч: 0.64
	C: 0.69	C: 0.73	C: 0.74	C: 0.65
AdaBoostClassifier	T: 0.68	T: 0.60	T: 0.62	T: 0.61
	Ч: 0.71	Ч: 0.59	Ч: 1.00	Ч: 0.63
	C: 0.66	C: 0.62	C: 0.00	C: 0.62
GaussianProcessClassifier	T: 0.65	T: 0.63	T: 0.64	T: 0.59
	Ч: 0.63	Ч: 0.64	Ч: 0.56	Ч: 0.56
	C: 0.69	C: 0.63	C: 0.74	C: 0.65
GradientBoostingClassifier	T: 0.66	T: 0.59	T: 0.63	T: 0.63
	Ч: 0.69	Ч: 0.59	Ч: 0.62	Ч: 0.64
	C: 0.65	C: 0.60	C: 0.66	C: 0.63
BaggingClassifier	T: 0.63	T: 0.59	T: 0.62	T: 0.59
	Ч: 0.58	Ч: 0.55	Ч: 0.53	Ч: 0.55
	C: 0.70	C: 0.65	C: 0.72	C: 0.67
ExtraTreeRegressor	T: 0.60	T: 0.57	T: 0.60	T: 0.61
	Ч: 0.61	Ч: 0.57	Ч: 0.59	Ч: 0.58
	C: 0.60	C: 0.58	C: 0.62	C: 0.64
LabelSpreading	T: 0.63	T: 0.59	T: 0.59	T: 0.56
	Ч: 0.56	Ч: 0.60	Ч: 0.50	Ч: 0.44
	C: 0.70	C: 0.60	C: 0.69	C: 0.70
LabelPropagation	T: 0.63	T: 0.59	T: 0.59	T: 0.56
	Ч: 0.56	Ч: 0.60	Ч: 0.50	Ч: 0.52
	C: 0.70	C: 0.60	C: 0.69	C: 0.61
ExtraTreeClassifier	T: 0.59	T: 0.57	T: 0.59	T: 0.59
	Ч: 0.60	Ч: 0.58	Ч: 0.59	Ч: 0.59
	C: 0.60	C: 0.56	C: 0.6	C: 0.59
DecisionTreeClassifier	T: 0.62	T: 0.54	T: 0.58	T: 0.56
	Ч: 0.64	Ч: 0.56	Ч: 0.46	Ч: 0.56
	C: 0.61	C: 0.53	C: 0.70	C: 0.57
DecisionTreeRegressor	T: 0.62	T: 0.54	T: 0.57	T: 0.56
	Ч: 0.64	Ч: 0.56	Ч: 0.58	Ч: 0.56
	C: 0.61	C: 0.53	C: 0.58	C: 0.57

Из таблицы видно, что максимальная точность классификации 76% достигается для набора *ReHo* при использовании классификаторов *RidgeClassifierCV* и *LogisticRegression*. Максимальная точность для набора признаков FHR 70% достигается при использовании этих же двух классификаторов.

Для набора *fALFF* максимальная точность 70% достигается при использовании классификатора *RidgeClassifierCV*, классификатор *LogisticRegression* — на втором месте с отставанием 1%. Для набора признаков *ALFF* максимальная точность 73% достигается при использовании классификатора *NuSVC*.

Таблица 2. Максимальная точность классификации (по всем классификаторам) участников эксперимента по наборам признаков norm_ReHo, norm_ALFF, norm_fALFF, norm_FHR в зависимости от алгоритма выбора наиболее значимых признаков. В скобках приведено число использованных признаков

Table 2. The maximum classification accuracy (for all classifiers) of the subjects according to the sets of signs norm_ReHo, norm_ALFF, norm_fALFF, norm_FHR, depending on the algorithm for selecting the most significant signs. The number of features used is given in parentheses

Алгоритмы выбора наиболее значимых признаков	ReHo	ALFF	fALFF	FHR
Все 132 признака	0.76	0.73	0.68	0.69
ExtraTrees	0.76 (55)	0.71 (54)	0.70 (53)	0.70 (57)
AdaBoost	0.71 (34)	0.73 (32)	0.68 (32)	0.70 (33)
L2	0.74 (56)	0.73 (55)	0.68 (55)	0.68 (52)

Классификаторы, для которых более, чем в 5% случаев алгоритм не сходил к решению или возникала ошибка: QuadraticDiscriminantAnalysis, GaussianMixture, BayesianGaussianMixture, IsolationForest, MiniBatchKMeans, DummyClassifier, OneClassSVM. Возможно, это связано с особенностью решаемой задачи, где число участников эксперимента (72) меньше размерности пространства признаков (132).

Интересно сравнить полученные результаты с результатами работы (Kalmady, 2019), в которой тоже использовался набор признаков *ReHo*, давший при использовании метода *SVM* 74–76%. В нашем случае *ReHo* для классического линейного *SVM* показал такую же точность (76%).

В таблице 2 представлена максимальная точность (среди классификаторов), достигнутая при использовании разных алгоритмов выбора наиболее значимых признаков, а также число признаков, которое эти методы выбрали как значимые.

Из таблицы 2 видно, что наибольшая точность на 3 из 4 наборов признаков достигается при использовании алгоритма *ExtraTreesClassifier* для выбора наиболее значимых признаков.

Но при этом было замечено, что методы выбора признаков не всегда оптимально определяют число признаков для классификаторов. Поэтому мы провели дополнительный тест, в котором вручную ограничивали число признаков для классификации. Для выбора наиболее значимых признаков в рамках лимита был использован алгоритм *ExtraTrees*, так как при использовании этого алгоритма была получена максимальная точность на 3 из 4 наборов признаков (см. таблицу 2). Полученная при этом максимальная точность классификации представлена на графике на рис. 3.

Примечательно, что точность классификации растет при увеличении числа признаков только до определенного предела, после которого может уменьшаться. Падение максимальной точности при увеличении числа признаков мы связываем с «разбавлением» полезных данных

незначимыми компонентами, содержащимися в других признаках.

Можно заметить, что при ручном ограничении числа признаков значительно увеличилась максимальная точность для набора *FHR*, и только для него. Это логично связать с тем, что для набора *FHR* максимальная точность достигается на малом числе признаков, в то время как алгоритмы выбора оптимального числа признаков предлагали использовать от 33 признаков и более.

Из графика на рисунке 3 можно сделать вывод, что на наборах *FHR*, *ReHo*, *ALFF* достигается сопоставимая между ними точность классификации

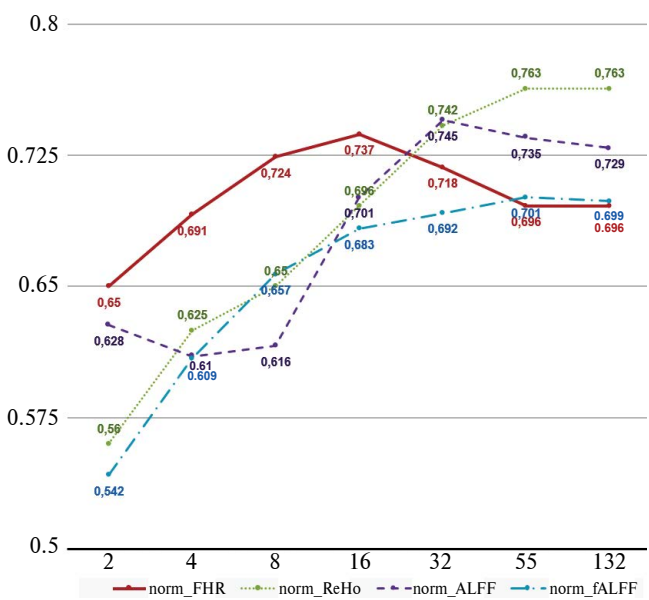


Рис. 3. Зависимость точности классификации от выбранного числа признаков. По горизонтали – число признаков (логарифмическая шкала), по вертикали – максимальная точность среди всех классификаторов.
Fig. 3. Dependence of classification accuracy on the selected number of features. Horizontally – the number of features (logarithmic scale), vertically – maximum accuracy among all classifiers.

Таблица 3. Наиболее значимые регионы по набору признаков FHR
Table 3. The most significant regions based on a set of characteristics FHR

Сокращенное название по атласу	Полное название по атласу	Соответствующий анатомический регион	Частота попадания в 8 лучших при перезапусках, %
Putamen l	Putamen l	Скорлупа слева	100
Putamen r	Putamen r	Скорлупа справа	98
Thalamus l	Thalamus l	Таламус слева	84
Thalamus r	Thalamus r	Таламус справа	73
PostCG r	Postcentral Gyrus Right	Постцентральная извилина справа	55
iLOC l	Lateral Occipital Cortex, inferior division Left	Нижний отдел латеральной затылочной коры слева	52
FO l	Frontal Operculum Cortex Left	Лобная доля слева	44
Accumbens r	Accumbens r	Прилежащее ядро справа	32

74–76%. На наборе *fALFF* – немного меньше (70%). Но при использовании малого числа признаков (до 16) набор *FHR* обеспечивает более высокую точность классификации по сравнению с другими наборами признаков. Это может быть важно в задаче выделения и анализа областей мозга, функционирование которых изменяется при патологии шизофрении.

Мы взяли 8 признаков, определенных алгоритмом *ExtraTrees*, на которых достигается максимальная точность для набора данных *FHR*, и посмотрели, каким регионам эти признаки соответствуют. Результат приведен в таблице 3. Мы сделали 5000 повторений. В таблице приведены регионы, которые попали в 8 лучших большее число раз. Последний столбец в таблице показывает, как часто данный регион попал в 8 лучших.

Таким образом, можно предположить, что именно выделенные 8 регионов дают наиболее высокий уровень разделения, что открывает перспективы для их дальнейшего углубленного анализа и определения значимости при постановке диагноза шизофрении.

Для сравнения мы также выбрали 8 лучших признаков, определенных алгоритмом *AdaBoost* и *L2* на наборе данных *FHR*, и сравнили соответствующие им регионы с регионами, выбранными алгоритмом *ExtraTrees*. Пересечение между 8 лучшими регионами, выбранными по алгоритмам *AdaBoost* и *ExtraTrees*, составило 4 региона (*Putamen l*, *Putamen r*, *Thalamus r*, *PostCG r*). Но если взять 16 лучших регионов по алгоритму *AdaBoost* и 8 лучших регионов по *ExtraTrees*, то пересечение составляет уже 7 регионов из 8 возможных (только регион *iLOC l* не вошел в 16 лучших по версии алгоритма *AdaBoost*).

Пересечение между 8 лучшими регионами, выбранными по алгоритмам *L2* и *ExtraTrees*, составило 6 регионов (не вошли в пересечение регионы *iLOC l* и *Accumbens r*).

ЗАКЛЮЧЕНИЕ

Проведенное исследование показало, что мы смогли на полученных данных достичь точности бинарной классификации патологии шизофрении 76%, используя набор, построенный с применением метода локальной однородности (*ReHo*), усредненный по регионам атласа CONN, и классификаторы *RidgeClassifierCV* и *LogisticRegression*. Полученная точность соответствует результатам, полученным в других научных исследованиях.

Другие методы формирования наборов признаков для классификации обеспечили максимальную точность классификации: 74% – метод выделения функционально однородных регионов и метод расчета амплитуды низкочастотных флуктуаций, 70% – метод расчета дробной амплитуды низкочастотных флуктуаций.

Предложенный авторами набор признаков на основе метода *FHR* показал интересные результаты. Во-первых, полученная точность не намного ниже: 74% при использовании заранее заданного числа признаков и алгоритма *ExtraTrees* для выбора наиболее значимых признаков.

Во-вторых, при использовании малого числа признаков (до 16) набор *FHR* обеспечивает более высокую точность классификации по сравнению с другими наборами признаков. Это может быть важно в задаче выделения и анализа областей мозга, функционирование которых изменяется при патологии шизофрении. Так, уже при использовании 8 признаков достигается точность 72.5%, в то время как остальные наборы при использовании 8 признаков обеспечивают максимальную точность 66% и менее. Это позволяет предположить, что именно выделенные 8 регионов дают наиболее высокий уровень разделения, что открывает перспективы для их дальнейшего углубленного

анализа и определения значимости при постановке диагноза шизофрении.

Скорлупа связана с дофамином и ассоциируется с фенотипами, включая многие нервно-психические и нейродегенеративные расстройства, которые сопровождаются нарушением двигательного контроля, импульсивным поведением и другими когнитивными нарушениями (Luo et al., 2019).

У пациентов с шизофренией наблюдается уменьшение объема таламуса справа и слева в сравнении со здоровыми добровольцами (Gilbert et al., 2001). Эта область играет ключевую роль в фильтрации информации и обладает обширными взаимосвязями с другими регионами мозга.

Прилежащее ядро — неотъемлемая часть лимбической и префронтальной кортико-полосатопаллидально-таламических цепей, участвует в нескольких когнитивных, эмоциональных и психомоторных функциях, которые нарушаются при шизофрении. В моделях на животных нарушения развития в энторинальной коре и гиппокампе вызывают нарушение регуляции входных сигналов в прилежащее ядро, что приводит к поведенческим отклонениям, указывающим на психотическую патологию (Lauer et al., 2001).

Согласно проведенным исследованиям, у большинства пациентов с шизофренией наблюдается нормальная анатомия и физиология затылочной коры, у меньшинства — сниженные значения, а совсем у немногих — увеличенные активность и структура (Tohid et al., 2015). Наблюдаемые изменения в объеме серого и белого вещества затылочных долей совершенно очевидны. Однако механизм, лежащий в основе этого поражения, еще до конца не изучен.

Также у пациентов с шизофренией была обнаружена положительная корреляция между оценкой шизотипии и средним коэффициентом кластеризации сенсомоторных подсетей и подсетей в режиме по умолчанию. В числе главных регионов, определяющих эти корреляции, была правая постцентральная извилина, которая демонстрирует нарушенную функциональную связность при шизофрении (Messaritaki et al., 2022).

Региональное истончение коры в лобной и латерально-височной (связанной с языком) областях коры у пациентов с шизофренией по сравнению со здоровыми субъектами согласуется с результатами структурного мета- и мегаанализа МРТ (Fusar-Poli et al., 2012).

Указанные выше работы и заключения говорят о том, что выделенные в ходе проведенного исследования области головного мозга играют важную роль при патогенезе шизофрении и нуждаются в тщательном изучении их структуры, функций, а также связанности с другими регионами. Разработанные подходы могут лечь

в основу создаваемых методов ранней диагностики и оценки эффективности терапии. Тем не менее, следует учесть, что результаты получены на малой выборке участников эксперимента, поэтому требуют дополнительной валидации.

ВКЛАД АВТОРОВ

А.Д. Жемчужников — анализ данных с использованием методов классификации; С.И. Карташов — постановка МРТ-эксперимента и проведение сканирования; С.О. Козлов — формирование наборов признаков для классификации; В.А. Орлов — предобработка фМРТ-данных; А.А. Пойда — написание текста, руководство работой; Н.В. Захарова, Л.В. Бравве, Г.Ш. Мамедова, М.А. Кайдан — клинический анализ и подготовка пациентов к исследованию, обсуждение результатов исследования, редактирование текста статьи.

ФИНАНСИРОВАНИЕ

Получение и обработка экспериментальных данных, разработка и тестирование метода классификации выполнены в рамках государственного задания НИЦ «Курчатовский институт». Клинический анализ и подготовка пациентов к исследованию проведены в рамках выполнения государственного задания Департамента здравоохранения г. Москвы №123031600072-3 «Молекулярно-генетические и нейрофизиологические предикторы первично психотических расстройств» от 16.03.2023 г.

БЛАГОДАРНОСТИ

Представленное исследование было проведено с использованием вычислительных ресурсов федерального центра коллективного пользования «Комплекс моделирования и обработки данных исследовательских установок мега-класса» НИЦ «Курчатовский институт».

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Все участники эксперимента дали письменное информированное согласие после полного описания процедур исследования согласно Хельсинкской декларации. Перед сканированием все

участники исследования заполняли опросный лист на наличие противопоказаний, лист добровольного информированного согласия и согласие на обработку персональных данных. Методика проведения исследования одобрена локальным этическим комитетом НИЦ «Курчатовский институт» (№ 5 от 05.04.2017 г.). Результаты настоящей работы являются частью исследовательской программы «Молекулярные и нейрофизиологические маркеры эндогенных заболеваний», проводимой в ГБУЗ «ПКБ № 1 ДЗМ», одобренной Независимым междисциплинарным комитетом по этической экспертизе клинических исследований от 14 июля 2017 года (протокол № 12). Некоторые результаты этого проекта опубликованы ранее (Corponi et al., 2021; Kartashov et al., 2021; Morozova et al., 2022; Zakharova et al., 2023).

СПИСОК ЛИТЕРАТУРЫ

- Zakharova N.V., Mamedova G., Bravve L.V., Kaydan A., Kartashov S., Orlov V.A., Ushakov V. Differential diagnosis of delusional symptoms in schizophrenia: Brain tractography data. *Cognitive Systems Research*. 2023. 77. 217–225.
<https://doi.org/10.1016/j.cogsys.2022.10.008>
- Antonucci L.A., Pergola G., Piloni A., Dwyer D., Kambeitz-Ilankovic L., Penzel N., Romano R., Gelao B., Torretta S., Rampino A., Trojano M., Caforio G., Falkai P., Blasi G., Koutsouleris N., Bertolino A. A Pattern of Cognitive Deficits Stratified for Genetic and Environmental Risk Reliably Classifies Patients With Schizophrenia From Healthy Control Subjects. *Biol. Psychiatry*. 2020. 87, 8. 697–707.
- Arlot S., Celisse A. A survey of cross-validation procedures for model selection. *Statistics Surveys*, 2010, P: 40–79.
- Bassett D.S., Sporns O. Network neuroscience. *Nature Neuroscience*. 2017, 20(3), 353–364.
- Biswal B., Zerrin Yetkin F., Haughton V.M., Hyde J.S. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*. 1995. 34 (4): 537–541.
- Calhoun V.D., Liu J., Adalı T. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage*. 2009. 45(1), S163–S172,
<https://doi.org/10.1016/j.neuroimage.2008.10.057>
- Cao B., Cho R.Y., Chen D., Xiu M., Wang L., Soares J.C., Zhang X.Y. Treatment response prediction and individualized identification of first-episode drug-naïve schizophrenia using brain functional connectivity. *Mol. Psychiatry*. 2020. 25(4): 906–913.
- Chang Y.-W., Tsai S.J., Wu Y.-F., Yang A.C. Development of an AI-Based Web Diagnostic System for Phenotyping Psychiatric Disorders. *Front. Psychiatry*. 2020. 11. 542394.
- Corponi F., Zorkina Y., Stahl D., Murru A., Vieta E., Serretti A., Morozova A., Reznik A., Kostyuk G.P., Chekhonin V. Disfunción de los lóbulos frontales en grupos clínicos de esquizofrenia aguda. *Revista de Psiquiatría y Salud Mental*, 2021,
<https://doi.org/10.1016/j.rpsm.2021.12.002>
- Cui Y., Li C., Liu B., Sui J., Song M., Chen J., Chen Y., Guo H., Li P., Lu L., Lv L., Ning Y., Wan P., Wang H., Wang H., Wu H., Yan H., Yan J., Yang Y., Zhang H., Zhang D., Jiang T. Consistent brain structural abnormalities and multisite individualised classification of schizophrenia using deep neural networks. *Br J Psychiatry*. 2022. 221 (6): 732–739.
<https://doi.org/10.1192/bjp.2022.22>
- Freund Y., Schapire R.E. A Short Introduction to Boosting. 1999.
- Fusar-Poli P., Radua J., McGuire P., Borgwardt S. Neuroanatomical maps of psychosis onset: voxel-wise meta-analysis of antipsychotic-naïve VBM studies. *Schizophr Bull*. 2012 Nov; 38 (6): 1297–307. Epub 2011 Nov 10. PMID: 22080494; PMCID: PMC3494061.
<https://doi.org/10.1093/schbul/sbr134>
- Gilbert A.R., Rosenberg D.R., Harenski K., Spencer S., Sweeney J.A., Keshavan M.S. Thalamic volumes in patients with first-episode schizophrenia. *Am J Psychiatry*. 2001 Apr; 158 (4): 618–24. PMID: 11282698.
<https://doi.org/10.1176/appi.ajp.158.4.618>
- Gutiérrez-Gómez L., Vohryzek J., Chiêm B., Baumann P.S., Conus P., Cuenod K.D., Hagmann P., Delvenne J.C. Stable biomarker identification for predicting schizophrenia in the human connectome. *Neuroimage Clin*. 2020. Vol. 27. P. 102316.
- Hu K., Wang M., Liu Y., Yan H., Song M., Chen J., Chen Y., Wang H., Guo H., Wan P., Lv L., Yang Y., Li P., Lu L., Yan J., Wang H., Zhang H., Zhang D., Wu H., Ning Y., Jiang T., Liu B. Multisite schizophrenia classification by integrating structural magnetic resonance imaging data with polygenic risk score. *Neuroimage Clin*. 2021. 32. 102860.
- Jo Jun-Mo. Effectiveness of Normalization Pre-Processing of Big Data to the Machine Learning Performance. *The Journal of the Korea institute of electronic communication sciences*. 2019. 14. 547–552.
<https://doi.org/10.13067/JKIECS.2019.14.3.547>
- Kalmady S.V., Greiner R., Agrawal R., Shivakumar V., Narayanaswamy J.C., Brown M.R.G., Greenshaw A.J., Dursun S.M., Venkatasubramanian G. Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. *NPJ Schizophr*. 2019. 5 (1): 2.
- Kartashov S., Vartanov A., Захарова H., Arkhipov A., Strelets V., Maslennikova A., Ushakov V. Tractographic studies in the human brain at the norm and the paranoid schizophrenia. *Procedia Computer Science*. 2021. 190. 434–438.
<https://doi.org/10.1016/j.procs.2021.06.051>
- Kambeitz-Ilankovic L., Meisenzahl E.M., Cabral C., von Salder S., Kambeitz J., Falkai P., Möller H.J., Reiser M., Koutsouleris N. Prediction of outcome in the

- psychosis prodrome using neuroanatomical pattern classification. *Schizophr. Res.* 2016. 173 (3): 159–165.
- Kottaram A., Johnston L.A., Tian Y., Ganella E.P., Laskaris L., Cocchi L., McGorry P., Pantelis C., Kotagiri R., Cropley V., Zalesky A. Predicting individual improvement in schizophrenia symptom severity at 1-year follow-up: Comparison of connectomic, structural, and clinical predictors. *Hum. Brain Mapp.* 2020. 41 (12): 3342–3357.
- Kozlov S., Poyda A., Orlov V., Malakhov D., Ushakov V., Sharaev M. Selection of functionally homogeneous brain regions based on correlation-clustering analysis. *Procedia Comput. Sci.* Elsevier BV, 2020. 169. 519–526.
- Kozlov S., Poyda A., Orlov V., Sharaev M., Ushakov V. Selection of functionally homogeneous human brain regions for functional connectomes building based on fMRI data. *Advances in Cognitive Research, Artificial Intelligence and Neuroinformatics*. Cham: Springer International Publishing. 2021. 709–719.
- Larkman D.J., Hajnal J.V., Herlihy A.H., Coutts G.A., Young I.R., Ehnholm G. Use of multicoil arrays for separation of signal from multiple slices simultaneously excited. *J. Magn. Reson. Imaging.* 2001. 13 (2): 313–317.
- Lauer M., Senitz D., Beckmann H. Increased volume of the nucleus accumbens in schizophrenia. *J Neural Transm (Vienna)*. 2001;108 (6):645–60. PMID: 11478417. <https://doi.org/10.1007/s007020170042>
- Luo X., Mao Q., Shi J., Wang X., Li C.R. Putamen Gray Matter Volumes in Neuropsychiatric and Neurodegenerative Disorders. *World J Psychiatry Ment Health Res.* 2019. Vol. 3(1): 1020. Messaritaki E., Foley S., Barawi K., Ettinger U., Jones D. Increased sensorimotor structural connectivity in high schizotypy. *bioRxiv*, 2022. <https://doi.org/10.1101/2022.05.12.491533>
- Morozova A., Zorkina Y., Abramova O., Pavlova O., Pavlov K., Soloveva K., Volkova M., Alekseeva P., Andryshchenko A., Kostyuk G., Gurina O., Chekhonin V. Neurobiological Highlights of Cognitive Impairment in Psychiatric Disorders. *Int J Mol Sci.* 2022 Jan 22;23(3):1217. PMID: 35163141; PMCID: PMC8835608. <https://doi.org/10.3390/ijms23031217>.
- Nieto-Castanon A., Whitfield-Gabrieli S. CONN functional connectivity toolbox: RRID SCR_009550, release 22. Hilbert Press. 2022. <https://doi.org/10.56441/hilbertpress.2246.5840>
- Pedregosa F., Varoquaux Gael, Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M. Duchesnay E., Louppe G. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 2011. 12(Oct), 2825–2830.
- Pudjihartono N., Fadason T., Kempa-Liehr A.W., O'Sullivan J.M. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Front. Bioinform.* 2022.2:927312. <https://doi.org/10.3389/fbinf.2022.927312>
- Rosazza C., Minati L. Resting-state brain networks: Literature review and clinical applications. *Neurological Sciences*. 2011; 32(5),773–785
- Sharaff A., Gupta H. Extra-Tree Classifier with Metaheuristics Approach for Email Classification. 2019. https://doi.org/10.1007/978-981-13-6861-5_17.
- Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2001. 546.
- Todd N., Josephs O., Zeidman P., Flandin G., Moeller S., Weiskopf N. Functional Sensitivity of 2D Simultaneous Multi-Slice Echo-Planar Imaging: Effects of Acceleration on g-factor and Physiological Noise. *Front. Neurosci.* 2017. 11: 158. <https://doi.org/10.3389/fnins.2017.00158>
- Tohid H., Faizan M., Faizan U. Alterations of the occipital lobe in schizophrenia. *Neurosciences (Riyadh)*. 2015 Jul; 20 (3): 213–24. PMID: 26166588; PMCID: PMC4710336. <https://doi.org/10.17712/nsj.2015.3.20140757>
- Wolf A. The Machine Learning Simplified: A Gentle Introduction to Supervised Learning. Leanpub, 2022.
- Xu K., Liu Y., Zhan Y., Ren J., Jiang T. BRANT: A Versatile and Extendable Resting-State fMRI Toolkit. *Front Neuroinform.* 2018. 12: 52. <https://doi.org/10.3389/fninf.2018.00052>
- Zang Y., Jiang T., Lu Y., He Y., Tian L. Regional homogeneity approach to fMRI data analysis. *Neuroimage*. 2004. 22 (1): 394–400. PMID: 15110032
- Zou Q.H., Zhu C.Z., Yang Y., Zuo X.N., Long X.Y., Cao Q.J., Wang Y.F., Zang Y.F. An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF. *J Neurosci Methods*. 2008 Jul 15; 172 (1): 137–41. Epub 2008 Apr 22. PMID: 18501969; PMCID: PMC3902859. <https://doi.org/10.1016/j.jneumeth.2008.04.012>

ON MOST INFORMATIVE REGIONS FOR BINARY CLASSIFICATION OF SCHIZOPHRENIA BASED ON RESTING STATE FMRI DATA DONE BY SELECTION OF FUNCTIONALLY HOMOGENEOUS REGIONS METHOD

A. D. Zhemchuzhnikov^a, S. I. Kartashov^a, S. O. Kozlov^a, V. A. Orlov^a,
A. A. Poyda^{a, #}, N. V. Zakharova^{b, c}, L. V. Bravve^d, G. Sh. Mamedova^d, M. A. Kaydan^d

^aNational Research Center "Kurchatov Institute", Moscow, Russia

^bInternational Centre for Education and Research in Neuropsychiatry, Samara State Medical University,
Samara, Russia

^cFederal State Budgetary Institution V.M. Bekhterev National Research Medical Center For Psychiatry And Neurology,
Saint Peterburg, Russia

^dPsychiatric Hospital no. 1 Named after N.A. Alexeev of the Department of Health of Moscow, Moscow, Russia
[#]e-mail: Poyda_AA@nrcki.ru

In this work we solve the problem of automatic binary classification of subjects with a diagnosis of schizophrenia and control groups on a data set obtained on a Siemens 3T tomograph. The data set included 36 subjects undergoing treatment at Psychiatric Hospital no. 1 Named after N.A. Alexeev of the Department of Health of Moscow (GBUZ PKB No. 1 DZM) and 36 subjects from the control group. Machine learning methods were used to solve this problem. As a result, an accuracy of 76% was achieved, which corresponds to the results obtained in other scientific studies. The highest accuracy was obtained for the local homogeneity parameter (regional homogeneity – ReHo), already known in the literature. At the same time, the set of features developed by the authors based on the method for identifying functionally homogeneous regions (FHR) gave a classification accuracy of 74%. But at the same time, the set of FHR features provides higher classification accuracy when using a small number of brain regions. For example, already in 8 regions, the FHR set provided an almost maximum classification accuracy of 72.5% (versus 65% for the ReHo set), which suggests that it is the selected 8 regions that give the highest level of separation.

Keywords: automatic classification of schizophrenia, fMRI, resting state, machine learning methods.