



Московский государственный университет имени М.В. Ломоносова

Факультет вычислительной математики и кибернетики

Интеллектуальные технологии в медицине 633

Аладина Екатерина Владимировна

Онежко Надежда Николаевна

**Использование FLAML и AutoGluon для диагностики
болезни Альцгеймера при помощи данных томографа и
данных когнитивных тестов пациентов**

Отчет по проекту Data science в обработке медицинских данных

Преподаватель:

Д.А.Андриков

Москва, 2025

Содержание

1 Введение	3
2 Постановка задачи и цели	4
3 Материалы и методы	5
3.1 Данные	5
3.2 Методы	6
3.2.1 FreeSurfer	6
3.2.2 FLAML и AutoGluon	6
3.2.3 Особенности работы с AutoML-фреймворками	7
4 Этапы работы	8
5 Результаты	10
5.1 Результаты прогнозирования когнитивного статуса	10
5.2 Результаты прогнозирование клинической прогрессии	10
5.3 Результаты прогнозирования скорости атрофии гиппокампа	11
5.4 Результаты прогнозирования функционального статуса	11
6 Вывод	13
7 Список литературы	14
8 Приложение 1. Подробный список используемых признаков.	15

1 Введение

Болезнь Альцгеймера (БА) является наиболее распространенной причиной деменции в мире и представляет собой одну из ключевых медико-социальных проблем этого века. Это хроническое нейродегенеративное заболевание характеризуется прогрессирующим ухудшением когнитивных функций, включая память, мышление и способность к выполнению повседневных задач, что в итоге приводит к полной зависимости пациента от посторонней помощи. По данным Всемирной организации здравоохранения, численность людей с деменцией превышает 55 миллионов и, как ожидается, будет удваиваться каждые 20 лет. Это создает колоссальное экономическое бремя для систем здравоохранения и оказывает тяжелейшее психологическое воздействие на семьи пациентов.

На сегодняшний день БА остается неизлечимым заболеванием. Существующие методы терапии в основном носят симптоматический характер и не способны остановить или существенно замедлить патологический процесс. В этой связи критически важным направлением современных исследований становится ранняя и точная диагностика, а также прогнозирование индивидуальной траектории развития болезни. Чем раньше выявлены признаки нейродегенерации, тем больше возможностей для своевременного вмешательства, подбора терапии, планирования ухода и участия в клинических испытаниях новых препаратов.

Современная диагностика БА основана на комплексном подходе, который включает клиническую оценку, нейровизуализацию, анализ биомаркеров, нейропсихологическое тестирование.

Однако интеграция столь разнородных и многомерных данных (клинических, нейропсихологических, генетических и нейровизуализационных) для построения индивидуальных прогнозов представляет собой сложную вычислительную задачу. Традиционный подход к машинному обучению требует от исследователя глубоких экспертных знаний для выполнения трудоемких этапов: отбора признаков, выбора модели, настройки гиперпараметров и валидации. Поэтому решено использовать Automated Machine Learning (AutoML).

2 Постановка задачи и цели

Целью данной работы является оценка возможностей и эффективности современных AutoML-фреймворков FLAML и AutoGluon для решения ключевых прогнозных задач в области исследования болезни Альцгеймера на основе реальных клинико-нейровизуализационных данных. Мы ставим перед собой задачу не только сравнить качество предсказаний, генерируемых этими инструментами, но и оценить их практическую применимость, удобство использования и способность выявлять значимые предикторы в сложной предметной области.

Для этого были сформулированы данные этапы работы:

- Предобработка данных для последующей работы с ними
- Выделение основных направлений работы и постановка подзадач
- Проверка гипотезы о применении фреймворков в рамках подзадач
- Сравнение результатов для AutoGluon и FLAML

В рамках работы было выделено несколько подзадач, чтобы оценить возможности AutoML-фреймворков для различных прогностических параметров. Для этого были выбраны такие подзадачи, как прогнозирование когнитивного спада, скорости атрофии ключевых структур мозга и риска прогрессии заболевания, что в перспективе может способствовать развитию методов персонифицированной медицины и улучшению качества помощи пациентам.

По итогу был получен такой список подзадач:

1. Прогноз когнитивных тестов. Предсказать MMSE.
2. Прогноз диагностической конверсии. Прогнозирование переходов от CN к MCI и от MCI к AD
3. Прогноз изменения объёмов мозга. Предсказание изменения объемов гиппокампа.
4. Прогноз функционального статуса. Предсказание изменения FAQ.

3 Материалы и методы

3.1 Данные

В работе использовались данные, предоставленные Инициативой по нейропроекции болезни Альцгеймера (Alzheimer's Disease Neuroimaging Initiative — ADNI). ADNI — это масштабная многолетняя лонгитюдная программа, целью которой является разработка клинических, визуализационных, генетических и биохимических биомаркеров для раннего обнаружения и мониторинга болезни Альцгеймера. Данные ADNI представляют собой стандартизованный и хорошо аннотированный набор, включающий информацию о тысячах пациентов с различными когнитивными статусами: от когнитивной нормы (CN) через легкие когнитивные нарушения (MCI) до деменции вследствие БА (AD).

Работа велась с двумя основными типами данных, интегрированными в единый табличный набор: клинико-демографические и нейропсихологические данные - таблицы с результатами когнитивных тестов, демографией и информацией о визитах пациентов и количественные нейропроекционные данные - морфометрические параметры структур головного мозга, полученные в результате автоматической обработки МРТ-снимков. Автоматическая обработка МРТ-снимков проводилась при помощи инструмента автоматической обработки МРТ FreeSurfer, который позволяет количественно оценить сотни морфометрических параметров мозга (объемы, толщину коры, площадь поверхности отдельных регионов), трансформируя визуальные данные в табличные, пригодные для машинного анализа.

Предобработка данных включала несколько этапов. Для начала проводилась обработка временных рядов, таким образом для каждого пациента данные упорядочивались по дате визита. Это позволило создавать прогнозные признаки с временным лагом (например, прогноз когнитивной оценки на следующем визите). Далее конструировались целевые переменные: для регрессии производился расчет изменений и скоростей изменения ключевых показателей (MMSE, FAQ, объем гиппокампа), а для классификации создавались новые признаки прогрессии (например, “HAS_P ROGRESSEDj⁻). – (,).(,)^{-200”}.

Подробный перечень используемых признаков смотреть в Приложении 1.

3.2 Методы

Для решения поставленных задач использовался комплекс методов, объединяющий подходы обработки нейровизуализационных данных, а также современные фреймворки автоматизированного машинного обучения.

3.2.1 FreeSurfer

Морфометрическая характеристика структур головного мозга выполнялась с использованием конвейера FreeSurfer (версия 7.x). Данный инструмент обеспечивает автоматическую сегментацию МРТ-изображений, позволяя получить количественные показатели для корковых и подкорковых областей по атласу Desikan-Killiany. В качестве ключевых признаков использовались: объемы подкорковых структур (Subcortical Volume, SV), объемы и площадь поверхности корковых областей (Cortical Volume, CV и Surface Area, SA), толщина коры (Thickness, TS), данные показатели были нормализованы относительно оценочного общего внутричерепного объема (eTIV) для минимизации влияния индивидуальных анатомических вариаций.

3.2.2 FLAML и AutoGluon

Для автоматического построения, настройки и сравнения прогнозных моделей были использованы FLAML (Fast and Lightweight AutoML) и AutoGluon.

FLAML использовался в конфигурации, ориентированной на быстрый поиск оптимальной модели с ограниченными вычислительными ресурсами. Фреймворк автоматически выполнял выбор алгоритма (включая линейные модели, случайный лес, градиентный бустинг LightGBM и XGBoost) и оптимизацию его гиперпараметров с помощью бюджетного поиска. Основной акцент делался на эффективность и простоту использования.

AutoGluon применялся в режиме, направленном на максимизацию прогнозной точности за счет построения сложных многоуровневых ансамблей (stacking,

blending). Фреймворк автоматически генерировал новые признаки, выполнял оптимизацию гиперпараметров для широкого спектра моделей (KNN, Random Forest, LightGBM, CatBoost, нейронные сети) и комбинировал их прогнозы.

Для интерпретации результатов использовались встроенные в фреймворки методы определения важности признаков: В FLAML применялся метод, основанный на анализе вклада признаков в алгоритмах градиентного бустинга (LightGBM, XGBoost), а AutoGluon предоставляет оценку важности на основе пермутационного теста (permutation importance), который оценивает падение точности модели при случайному перемешивании значений признака.

3.2.3 Особенности работы с AutoML-фреймворками

FLAML и AutoGluon следуют разным подходам к автоматизации машинного обучения, что определяет их сильные стороны и сферы применения.

FLAML создан для быстрого получения рабочей модели с минимальными вычислительными затратами. Его алгоритм — адаптивный поиск с контролем ресурсов, который целенаправленно оценивает наиболее перспективные модели и их настройки. Это позволяет FLAML быстро находить не просто хорошую, а оптимальную для заданного бюджета времени модель.

В отличие от многих фреймворков, AutoGluon достигает высокой точности не за счет тонкой настройки гиперпараметров отдельных моделей, а с помощью сложного стек-ансамблирования. Его принцип работы построен на обучении множества разнообразных моделей (случайный лес, градиентный бустинг, линейные модели и др.), которые объединяются в «слои».

4 Этапы работы

Исследование было организовано как последовательный конвейер обработки данных и анализа, направленный на сравнительную оценку эффективности AutoML-фреймворков для прогнозных задач в области нейродегенеративных заболеваний.

Начальный этап включал консолидацию и предобработку гетерогенных данных ADNI. Клинические и демографические показатели были интегрированы с количественными морфометрическими дескрипторами мозга, полученными с помощью FreeSurfer, в единые временные ряды по пациентам. Для реализации прогнозного моделирования были сконструированы целевые переменные с временным лагом, а также рассчитаны показатели динамики (абсолютные изменения и годовые скорости). Объемы структур мозга нормализовались относительно общего внутричерепного объема. Стандартные процедуры обработки пропусков и отбора признаков завершили формирование финального датасета.

Далее была проведена формализация прогнозных задач, сформулированных в контексте клинических потребностей. Выделены четыре подзадачи: 1) регрессия для предсказания когнитивного статуса (MMSE); 2) классификация и регрессия для оценки риска прогрессии диагноза ($CN \rightarrow MCI \rightarrow AD$); 3) регрессия для прогнозирования скорости атрофии гиппокампа; 4) регрессия и классификация для прогноза изменений функционального статуса (FAQ). Для каждой задачи выполнено стратифицированное разбиение на обучающую и тестовую выборки.

Процедура автоматизированного моделирования осуществлялась параллельно с использованием FLAML и AutoGluon на идентичных датасетах. FLAML применялся в режиме быстрого поиска оптимальной одиночной модели, AutoGluon — в режиме построения комплексных ансамблей с автоматическим feature engineering.

На этапе сравнительного анализа проводилась оценка моделей по ключевым метрикам (R^2 , Accuracy, AUC-ROC), анализировалась важность признаков и практические аспекты использования фреймворков (время обучения, сложность настройки, интерпретируемость). Это позволило сделать содержательные

выводы об областях эффективного применения AutoML в нейронаучных исследованиях и выявить существующие ограничения метода.

5 Результаты

Проведенное исследование позволило получить количественные оценки эффективности AutoML-фреймворков FLAML и AutoGluon для решения ключевых прогнозных задач в области болезни Альцгеймера. Результаты представлены в разрезе четырех сформулированных подзадач.

5.1 Результаты прогнозирования когнитивного статуса

Задача регрессии по прогнозу будущего значения теста MMSE на 12 месяцев вперед продемонстрировала наивысшую эффективность среди всех рассмотренных сценариев. FLAML достиг коэффициента детерминации $R^2 = 0.698$ ($MAE = 1.544$), автоматически выбрав в качестве оптимальной модель градиентного бустинга LightGBM. Топ-3 наиболее важных предикторов составили другие когнитивные тесты: mPACCtrailsB (внимание и исполнительные функции), ADAS13 (общая когнитивная оценка) и mPACCdigit (работа с цифрами). Модель объяснила около 70% дисперсии целевой переменной. AutoGluon показал несколько более низкий результат $R^2 = 0.551$ ($MAE = 1.460$), сформировав сложный ансамбль моделей. Ключевым отличием в интерпретации стало то, что наиболее важным признаком для AutoGluon оказалось текущее значение MMSE, за которым следовали морфометрические параметры мозга (ST7SV, ST129SA). Это указывает на различную стратегию feature selection, примененную фреймворками.

Оба инструмента справились с задачей удовлетворительно, при этом FLAML продемонстрировал более высокое значение R^2 . Результат подтверждает высокую корреляцию между различными когнитивными тестами и возможность построения точных краткосрочных прогнозов на их основе.

5.2 Результаты прогнозирование клинической прогрессии

Результаты по данной задаче оказались наиболее контрастными и выявили критическую зависимость итогов от корректности подготовки данных. FLAML дал практически идеальное качество моделей классификации и регрессии: Accurасу

$= 0.993$ – 1.000 для бинарных задач и $R^2 = 0.855$ для скорости прогрессии. Во всех случаях лучшей моделью также был LightGBM. AutoGluon, решавший задачу бинарной классификации (CN vs MCI), показал правдоподобные, но скромные результаты: Accuracy = 0.609, Balanced Accuracy = 0.619. Это соответствует уровню, типичному для сложной задачи раннего прогноза на реальных данных. Экстремально высокие метрики, полученные FLAML, с высокой вероятностью свидетельствуют о переобучении. Результаты AutoGluon, напротив, отражают объективную сложность прогнозирования смены диагностических стадий.

5.3 Результаты прогнозирования скорости атрофии гиппокампа

Данная регрессионная задача оказалась наименее решаемой с использованием доступных данных и выбранного подхода. FLAML показал близкие к нулю значения R^2 : от 0.0016 для левого гиппокампа до 0.0433 для правого. Лучшими моделями выступили Extra Trees и LightGBM, но их прогнозная сила была минимальна. AutoGluon подтвердил эту тенденцию, получив исходный $R^2 = 0.0022$. Эксперимент с принудительным сокращением числа признаков до 10 улучшил метрику лишь до $R^2 = 0.0391$. Наибольший вклад, согласно пермутационной важности, вносили клинические шкалы (CDRSB, MMSE) и генетический маркер APOE4. Таким образом, прямое прогнозирование изменений объемов гиппокампа на основе одноточечных данных и когнитивных тестов не представляется возможным с удовлетворительной точностью. Полученные модели не превосходят простое среднее значение, что указывает на стохастичность процесса или недостаточность признаков для его описания.

5.4 Результаты прогнозирования функционального статуса

Результаты варьировались в зависимости от конкретной формулировки задачи (регрессия на значение и на изменение, классификация).

При прогнозе будущего значения FAQ получились следующие результаты:

AutoGluon продемонстрировал высокое качество ($R^2 = 0.649$). Наиболее важным предиктором стала шкала CDRSB, что клинически обосновано. Для прогноза изменения FAQ оба фреймворка показали полную неэффективность, получив отрицательные значения R^2 (около -0.117), что хуже константного прогноза. Для задачи классификации факта ухудшения FAQ FLAML достиг сбалансированного результата: Accuracy = 0.778, ROC-AUC = 0.697 при использовании модели Extra Trees. Это указывает на принципиальную решаемость задачи бинарной классификации ухудшения состояния. Задача прогноза будущего значения функционального статуса решается хорошо, так как оно сильно зависит от текущего клинического состояния. Однако прогноз величины изменения (dFAQ) является статистически недостоверным.

6 Вывод

Сравнительный анализ FLAML и AutoGluon показал, что в задачах, не связанных с прямым прогнозом изменений, оба фреймворка показали работоспособность. FLAML проявил себя как более быстрый и склонный к выбору простых, эффективных моделей (LightGBM). AutoGluon предоставил более глубокий анализ важности признаков (пермутационный метод) и в одной из задач (прогноз значения FAQ) достиг более высокого качества за счет построения ансамблей.

В задачах, где не было артефактов подготовки данных (прогноз MMSE, классификация ухудшения FAQ), оба фреймворка показали сопоставимое качество. Однако интерпретация лучших моделей различалась. FLAML, опираясь на встроенную важность признаков в деревьях, чаще указывал на нейропсихологические тесты как на ключевые предикторы (например, mPACCtrailsB для MMSE). AutoGluon, использующий пермутационную важность (model-agnostic метод), в той же задаче выделил текущее значение MMSE и морфометрические параметры.

Таким образом, гипотеза о применимости AutoML-фреймворков для диагностики болезни Альцгеймера на медицинских данных подтвердилась.

7 Список литературы

- [1] Ballard C. et al. Alzheimer's disease //the Lancet. – 2011. – Т. 377. – №. 9770. – С. 1019-1031.
- [2] Данные используемые в задаче - датасет ADNI
- [3] Инструмент для работы со снимками МРТ FreeSurfer
- [4] Документация AutoGluon
- [5] Документация FLAML

8 Приложение 1. Подробный список используемых признаков.

Данные о пациентах

PTID - Participant ID - уникальный идентификатор пациента

IMAGEUID - Image Unique ID - уникальный идентификатор изображения
MRI

EXAMDATE - Examination Date - дата проведения обследования

SITE - Site ID - идентификатор исследовательского центра

AGE - Age at baseline - возраст на момент включения в исследование

PTGENDER - Participant Gender - пол пациента

PTEDUCAT - Participant Education - годы образования

PTETHCAT - Participant Ethnic Category - этническая принадлежность

PTRACCAT - Participant Race Category - расовая принадлежность

PTMARRY - Participant Marital Status - семейное положение

APOE4 - APOE 4 allele carrier status - статус носительства аллеля APOE4
(фактор риска болезни Альцгеймера)

Month и M - Months from baseline - месяцы от момента включения в исследование

Клинические и нейропсихологические тесты:

CDRSB - общий балл по шкале оценки деменции

ADAS11 - шкала оценки болезни Альцгеймера (11 пунктов)

ADAS13 - шкала оценки болезни Альцгеймера (13 пунктов)

ADASQ4 - тест на запоминание слов из ADAS

MMSE - краткая шкала оценки психического статуса

RAVLT_immediate - немедленное воспроизведение

RAVLT_learning - обучение/запоминание

RAVLT_forgetting - забывание

RAVLT_perc_forgetting - процент забывания

LDELTOTAL - отсроченное воспроизведение логической памяти
LDELTOTAL_BL - базовый уровень отсроченного воспроизведения
TRABSCOR - тест следования связям (часть B)
FAQ - опросник функциональной активности
mPACCdigit - модифицированный доклинический когнитивный композит
- mPACCtrailsB - модифицированный доклинический когнитивный композит (следование связям B)

Метрики по областям мозга

Столбцы начинающиеся с ST - это объемы конкретных регионов мозга после сегментации:

ST...SV - Subcortical Volume (подкорковый объем)
ST...CV - Cortical Volume (корковый объем)
ST...SA - Surface Area (площадь поверхности)
ST...TA - Total Area (общая площадь)
ST...TS - Thickness (толщина коры)

Где ... - номер региона по FreeSurfer. Например:

ST101SV - Left Thalamus volume (левый таламус)

Основные подкорковые структуры (+ SV - Subcortical Volume)

ST101SV - Left Thalamus (левый таламус)
ST102SV - Left Caudate (левое хвостатое ядро)
ST103SV - Left Putamen (левая скорлупа)
ST104SV - Left Pallidum (левый бледный шар)
ST105SV - Left Hippocampus (левый гиппокамп)
ST107SV - Left Accumbens area (левая прилежащая область)
ST108SV - Left Ventral DC (левый вентральный диэнцефалон)
ST109SV - Right Thalamus (правый таламус)
ST110SV - Right Caudate (правое хвостатое ядро)
ST111SV - Right Putamen (правая скорлупа)
ST112SV - Right Pallidum (правый бледный шар)
ST113SV - Right Hippocampus (правый гиппокамп)

ST114SV - Right Amygdala (правое миндалевидное тело)

ST115SV - Right Accumbens area (правая прилежащая область)

ST116SV - Right Ventral DC (правый вентральный диэнцефалон)

Корковые регионы (+ CV - Cortical Volume, SA - Surface Area, TA

- Thickness)

Лобная доля (Frontal Lobe)

ST23 - Left Superior Frontal (левая верхняя лобная извилина)

ST24 - Left Rostral Middle Frontal (левая ростральная средняя лобная)

ST25 - Left Caudal Middle Frontal (левая каудальная средняя лобная)

ST26 - Left Precentral (левая прецентральная извилина)

ST31 - Left Pars Opercularis (левая покрышечная часть)

ST32 - Left Pars Triangularis (левая треугольная часть)

ST34 - Left Lateral Orbitofrontal (левая латеральная орбитофронтальная)

ST35 - Left Medial Orbitofrontal (левая медиальная орбитофронтальная)

ST38 - Left Frontal Pole (левый лобный полюс)

Височная доля (Temporal Lobe)

ST13 - Left Entorhinal (левая энторинальная кора)

ST14 - Left Fusiform (левая веретенообразная извилина)

ST15 - Left Inferior Temporal (левая нижняя височная извилина)

ST39 - Left Middle Temporal (левая средняя височная извилина)

ST40 - Left Superior Temporal (левая верхняя височная извилина)

ST43 - Left Banks STS (левый банк верхней височной борозды)

ST44 - Left Transverse Temporal (левая поперечная височная извилина)

ST45 - Left Temporal Pole (левый височный полюс)

Теменная доля (Parietal Lobe)

ST46 - Left Inferior Parietal (левая нижняя теменная доля)

ST47 - Left Superior Parietal (левая верхняя теменная доля)

ST48 - Left Postcentral (левая постцентральная извилина)

ST49 - Left Supramarginal (левая надкраевая извилина)

ST54 - Left Precuneus (левая предклиньевая извилина)

Затылочная доля (Occipital Lobe)

ST55 - Left Lateral Occipital (левая латеральная затылочная)

ST56 - Left Lingual (левая язычна извилина)

ST57 - Left Cuneus (левая клиновидная извилина)

ST58 - Left Pericalcarine (левая околошпорная кора)

Другие важные регионы

ST59 - Left Insula (левый островок)

ST60 - Left Cingulate (левая поясная извилина)