

UDC 004.021:004.75

ОБРАБОТКА ДАННЫХ В СЕТИ ИТ-ДИАГНОСТИКИ БОЛЕЗНИ АЛЬЦГ ЕЙМЕРА



В.А. Вишняков
Профессор кафедры ИКТ БГУИР,
д.т.н., профессор
vish@bsuir.by



Юй Чуюэ
Аспирантка кафедры ИКТ
БГУИР
усу18779415340@gmail.com

В.А. Вишняков

Окончил Минский радиотехнический институт. Область научных интересов связана с разработкой методов и алгоритмов инфокоммуникационных систем, организацией учебного и научно-исследовательского процессов в области сетей IoT и блокчейн.

Ю. Чуюэ

Учится в аспирантуре Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с областью обработки информации, ИТ-диагностики и интернета вещей (IoT).

Аннотация. Доклад посвящен процессу представления и обработки звуковых данных для диагностики болезни Альцгеймера (БА). Представлена структура сети Интернета вещей (IoT) для ИТ-диагностики болезни Альцгеймера, ее основные компоненты и этапы работы. Основываясь по результатам экспериментов данные описаны, представлены и обработаны.

Ключевые слова: машинное обучение, обработка данных БА, сеть IoT ИТ-диагностики.

Введение. На пути к интеллектуальным медицинским услугам, с увеличением объема информации, собираемой с помощью систем мониторинга пациентов, в области здравоохранения крайне важно принимать сложные аналитические решения, используя различные типы данных, собранных от каждого пациента [1]. Проблема больших данных в этой области важна не только из-за огромного объема, но и из-за разнообразия типов данных и скорости управления ими [2]. Научный мир называет это «проблемой больших данных», в то время как сети обработки данных, построенные с использованием новых технологий, таких как интернет вещей, облачные вычисления. Сети Интернет вещей и больших данных могут помочь усовершенствовать механизмы обработки информации в медицинских приложениях, включая: сбор и передачу данных, анализ данных, удаленную диагностику пациентов, мониторинг и обратная связь в режиме реального времени.

Акцент в службах здравоохранения смещается с модели реагирования на раннее прогнозирование. В случае (болезни Альцгеймера) на ранних стадиях не наблюдается дефицита движений, чувствительности или координации [3], поэтому пациентам часто ставят диагноз после того, как заболевание прогрессировало в течение нескольких лет. Это не только наносит ущерб пациенту, но и влечет за собой экономические издержки для общества (по оценкам, только в США они составляют более 100 миллиардов долларов

ежегодно) [4]. Потребность в точной и неинвазивной диагностической сетевой системе, интегрирующей ИТ-медицину для работников в клиниках, становится все более актуальной проблемой здравоохранения.

Распространенные формы данных, используемых для диагностики болезни Альцгеймера на основе информационных технологий, включают данные медицинской визуализации, медицинские карты пациентов, результаты когнитивных тестов, биомаркеры и мультимодальные данные. Представление этих данных включает:

1 Инструменты визуализации данных: программное обеспечение используется для создания визуальных представлений сложных данных (таких как визуализация головного мозга), чтобы обеспечить удобные интерфейсы, позволяющие клиницистам эффективно получать доступ к данным пациента и интерпретировать их.

2 Модели машинного обучения: анализ больших наборов данных для идентификации алгоритмов, указывающих на болезнь Альцгеймера, что позволяет использовать существующие данные для определения наличия у испытуемых болезни Альцгеймера.

3 Платформы интернета вещей: использование существующих платформ Интернета вещей для всестороннего представления данных о пациентах.

Алгоритм диагностики БА. В области ИТ-медицины аудиоданные могут использоваться в качестве оценочного показателя для диагностического исследования болезни Альцгеймера. Этот тип данных часто используется с помощью таких методов, как анализ голоса, обработка языка, автоматическое распознавание речи (*ASR*) – преобразование речи в текст, которое затем позволяет анализировать и интерпретировать лингвистический контент с использованием технологий обработки естественного языка (*NLP*). Передовые методы обработки данных и машинного обучения обычно используются при обработке и анализе этих данных для извлечения ключевой информации о когнитивном состоянии пациента.

В статье авторов [5], алгоритм классификатора случайного леса был применен для анализа полных текстовых данных, извлеченных из расшифрованных аудиоданных интервьюируемых. Набор данных *ADReSS 2020 Challenge* [6] был использован в этом исследовании для предоставления расшифровок аудиоданных.

Классификатор случайного леса – это алгоритм контролируемого обучения, обычно используемый для задач классификации и регрессии. При идентификации болезни Альцгеймера алгоритм случайного леса используется для различения пациентов с болезнью Альцгеймера и здоровых людей. Он работает путем построения нескольких деревьев решений и агрегирования их прогнозов для повышения общей точности и надежности. Каждое дерево решений обучается независимо, для окончательного решения делается «голосование» или усредненный прогноз. В исследовании использовались числовые векторы в качестве признаков для обучения модели случайного леса и использовался метод *GridSearchCV* для оптимизации гиперпараметров случайного леса, чтобы найти оптимальные настройки модели. Производительность модели в первую очередь оценивается точностью классификации.

Структура сети для ИТ-диагностики БА. Структура сети Интернета вещей для ИТ-диагностики болезни Альцгеймера представляет собой систему, предназначенную для мониторинга, анализа и прогнозирования показателей БА путем интеграции различных технологических компонентов и процессов: от датчиков до обработки на стороне сервера и выдачи результатов в приложения смартфона. При построении сети обработки данных для диагностики болезни Альцгеймера необходимо учитывать ее масштабируемость. Во многих приложениях, требующих автоматизированного принятия решений, нередко приходится получать данные из различных источников, которые могут предоставить дополнительную информацию.

Основные компоненты системы. Разработка сети Интернета вещей облегчает персонализированные медицинские услуги за счет обеспечения взаимодействия между устройствами в режиме реального времени, обработки и анализа данных и использования различных протоколов для передачи данных и обмена ими. Система представляет собой интеллектуальную медицинскую платформу на основе Интернета вещей, которая обеспечивает мониторинг и диагностику пациентов с болезнью Альцгеймера в режиме реального времени с использованием сенсорных устройств, сетевых протоколов и технологий обработки данных. В ней используется модель клиент-сервер для сбора, передачи, обработки и отображения данных в режиме реального времени. Рассмотрим основные этапы работы такой системы.

1 Сбор и ввод данных:

а) Источники данных: данные могут поступать из смартфонов, на который поступают звуковая информация голоса пациентов.

б) Форматирование данных: собранные необработанные данные преобразуются в формат, подходящий для дальнейшей обработки. При сборе голоса пациента с болезнью Альцгеймера он может быть преобразован в текст или расшифровку.

2 Передача данных:

а) Загрузка данных: данные из смартфона пересылаются на сервер или в облачную платформу, где может выполняться их предварительная обработка и извлечение необходимых функций.

б) Сетевые протоколы: используются сетевые протоколы (такие как *HTTP*, *MQTT*, *TCP/IP* и т.д.) для обеспечения безопасной и эффективной передачи данных. Например, клиенты могут отправлять данные на локальный сервер с помощью *HTTP POST*-запроса по указанному *URL*-адресу или осуществлять подключение между устройствами и своевременную связь по протоколу *MQTT*.

3 Обработка и анализ данных:

а) Очистка данных: удаление недопустимых, неполных или нерелевантных данных.

б) Интеллектуальный анализ данных: после отправки данных агенту прогнозирования БА на базе обученной нейронной сети, происходит обработка и интерпретация данных от новых пациентов. С помощью голосовых данных платформа *Flask Framework* может использоваться для приема *HTTP*-запросов, функции просмотра могут использоваться для анализа и обработки данных, которые затем передаются в предварительно обученную нейронную сеть для прогнозирования диагноза болезни Альцгеймера, генерируя диагностические результаты.

4 Машинное обучение и прогнозирование модели. Используемая модель машинного обучения распознает болезнь Альцгеймера на основе собранных характеристик данных, выводя результаты прогнозирования. Эта модель проверяется с использованием накопления данных и может корректироваться в процессе обучения, повышая точность с течением времени.

5 Взаимодействие с пользовательским интерфейсом. Этот шаг имеет значение для визуализации данных с использованием таких инструментов, как диаграммы, информационные панели и т.д., для представления результатов пользователям и создания отчетов об анализе данных или сводок. Медицинские работники и пациенты взаимодействуют с системой через интерфейсную страницу, написанную на *JAVA*, просматривая результаты анализа и получая прогнозы, поддерживая запись, преобразование голоса в текст и отображение результатов прогнозирования.

Для связи в режиме реального времени протокол *MQTT* может использоваться для передачи результатов прогнозирования в режиме реального времени, отправляя результаты обратно исходному клиенту или другим клиентам, подписанным на соответствующую тему. Платформа *OSTIS* может быть внедрена для реализации клиентом,

где ее база знаний и средство решения проблем позволяют легко интегрировать любую модель решения проблем или знания в систему *OSTIS* без затрат на управление (по принципу «подключи и работай»), облегчая последующее обслуживание или расширение модели.

6 Хранение данных. После того, как агент прогнозирования обработает данные и выдаст результаты прогнозирования БА, необходимо рассмотреть возможность резервного копирования и восстановления данных, чтобы обеспечить безопасность данных и возможность восстановления для последующего анализа или отчетности. Один из методов заключается в хранении данных о пациентах, результатов анализа и исторических записей в структурированной базе данных в рамках одной и той же записи данных. Тип и структура базы данных разработаны для удовлетворения конкретных потребностей диагностики БА, обеспечивая эффективный поиск данных и управление ими, например, база данных *MongoDB*. Данные также могут быть сохранены и постоянно храниться с помощью конфигурации *EMQX* broker, гарантируя, что сообщения *MQTT* (такие как результаты прогнозирования) могут быть немедленно переданы новым клиентам при необходимости.

7 Безопасность данных и конфиденциальность.

а) Шифрование: данные шифруются для защиты во время передачи и хранения. После того, как пользователи отправляют личные данные, система использует технологию децентрализованного хранения *IPFS* для генерации хэш-значения для каждого файла, обеспечивая целостность и безопасность данных. Пользователи могут не только генерировать и публиковать хэш-значения для проверки подлинности файлов, когда это необходимо, но также могут загружать соответствующие данные и результаты. Во время передачи данных также могут использоваться зашифрованные версии протоколов *HTTP* и *MQTT* (такие как *HTTPS* и *MQTTS*) для обеспечения безопасности передачи данных.

б) Контроль доступа: убедитесь, что только авторизованные пользователи могут получить доступ к конфиденциальным данным. В сети *Ethereum* этого шага можно достичь путем написания смарт-контрактов для ограничения и просмотра разрешений пользователей. Экспортируйте все файлы из базы данных, используя *MongoDB* в качестве примера, получите файлы *Patient_info.bson* и *Patient_info.metadata.json*, повторно зашифруйте эти два файла с помощью технологии *IPFS*, получите зашифрованные хэш-значения, напишите смарт-контракты и опубликуйте хэш-значения в сети *Ethereum* или загрузите данные в базу данных хэш-значений для управления данными и торговли ими. При использовании функции аутентификации *EMQX* broker это также может гарантировать, что только авторизованные пользователи смогут получать доступ к сообщениям и публиковать их.

в) Соответствие требованиям: использование правил защиты данных и конфиденциальности.

г) Экспорт данных: На данном этапе в системе может быть задействовано несколько пользователей с разными ролями (например, клиенты, поставщики данных). Данные (такие как информация о пациенте и метаданные) могут быть экспортированы в определенные форматы файлов (такие как: *bson* и *json*) для дальнейшего использования.

8 Обратная связь и оптимизация. Основное внимание на этом этапе уделяется мониторингу производительности, оптимизации и корректировке для постоянного улучшения системы на основе результатов экспериментов и отзывов пользователей. Инструменты отладки *Flask framework* и возможности мониторинга *EMQX* могут использоваться для отслеживания производительности системы и проблем, мониторинга производительности потока обработки данных, предоставления обратной связи в режиме реального времени и позволяют разработчикам корректировать и оптимизировать поток обработки данных на основе обратной связи.

Приложение. Результаты диагностики могут быть представлены в различных формах, включая, но не ограничиваясь ими, исчерпывающие аналитические отчеты, диаграммы, визуальные данные или легко понятные текстовые описания. Для голосовых данных авторы внедрили интуитивно понятное представление результатов диагностики, а модель машинного обучения выводит процентную вероятность наличия у участника болезни Альцгеймера. Чтобы сделать результаты системы более полными, можно добавить объяснение модели, указав, какие результаты тестов или поведенческие индикаторы указывают на возможное снижение когнитивных способностей, или включив исторические результаты тестирования участника, чтобы исключить субъективное влияние отдельных случаев.

Заключение. В докладе описаны данные, используемые для обучения нейронной сети распознавания БА, представлена сеть *IoT* обработки данных, которая используется в ИТ-диагностике болезни Альцгеймера. Она охватывает этапы от сбора данных до обработки данных, генерации результатов, защиты данных и обмена ими, а также конечного вывода данных. Эта сеть может быть применена к различным приложениям, таким как анализ медицинских данных, удаленный мониторинг или платформы обмена данными.

Список литературы

- [1] Sakr S, Elgammal A. Towards a comprehensive data analytics framework for smart healthcare services[J]. Big Data Research, 2016, 4: 44-58.
- [2] Dragomir A, Vrahatis A G, Bezerianos A. A network-based perspective in Alzheimer's disease: Current state and an integrative framework[J]. IEEE journal of biomedical and health informatics, 2018, 23(1): 14-25.
- [3] McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease[J]. Neurology, 1984, 34(7): 939-939.
- [4] Polikar R, Topalis A, Parikh D, et al. An ensemble based data fusion approach for early diagnosis of Alzheimer's disease[J]. Information Fusion, 2008, 9(1): 83-95.
- [5] Vishniakou U.A., Yu Ch. Using Machine Learning for Recognition of Alzheimer's Disease Based on Transcription Information. Doklady BGUIR. 2023;21(6):106-112.
- [6] Luz S, Haider F, de la Fuente S, et al. Alzheimer's dementia recognition through spontaneous speech: The ADReSS challenge[J]. arXiv preprint arXiv:2004.06833, 2020.

Авторский вклад

Вишняков Владимир Анатольевич – разработал концепцию, предоставил структуру доклада, проверил результаты эксперимента и оформил текст.

Чуйэ Юй – выполнила программную разработку системы, провела эксперименты в соответствии с структурой алгоритма.

DATA PROCESSING IN THE NETWORK FOR THE DIAGNOSIS OF ALZHEIMER'S DISEASE

Vishniakou U.A.

*Doctor of Technical Sciences,
Professor of the Department of
Infocommunication Technologies,
BSUIR.*

Yu C.Y.

*PhD student in the Department
of infocommunication
technologies at BSUIR.*

Abstract. The report is devoted to the process of presenting and processing audio data for the diagnosis of Alzheimer's disease (AD). The structure of the Internet of Things network for the diagnosis of Alzheimer's disease, its main components and stages of work are presented. Based on the results of experiments, the data are described, presented and processed.

Keywords: machine learning, AD data processing, IoT network of IT-diagnostics.