

УДК 004.046

СИСТЕМА АВТОМАТИЧЕСКОГО ПОИСКА УЧАСТКОВ ЭПИЛЕПТИЧЕСКОЙ АКТИВНОСТИ В СОСТАВЕ ОБЛАЧНОЙ ПЛАТФОРМЫ ХРАНЕНИЯ, СИСТЕМАТИЗАЦИИ И ОБРАБОТКИ МЕДИЦИНСКИХ ДАННЫХ

¹Алексеев Д.М., ¹Минюк А.Н., ²Понимаш З.А., ¹Шумилин А.С.

¹ФГАОУ ВО «Южный федеральный университет», Институт компьютерных технологий
и информационной безопасности, Таганрог, e-mail: minyuk@neurotech.ru;

²ФГАОУ ВО «Южный федеральный университет», Институт радиотехнических систем
и управления, Таганрог, e-mail: ponimash@neurotech.ru

Цель исследований заключается в разработке и оценке эффективности общей схемы облачной платформы, обеспечивающей выполнение определенного спектра задач, а также в выборе и анализе классификатора для решения задачи выделения зон эпилептической активности как части облачной платформы. В результате исследований разработана схема облачной платформы хранения, систематизации и обработки медицинских данных, которая включает в себя четыре основных уровня. Разделение потоков данных на уровни, стандартизация протоколов передачи информации и форматов её хранения позволяют создать гибкую, универсальную и высоконадёжную медицинскую информационную систему. Архитектура позволяет быстро интегрироваться в существующие медицинские системы и внедрять используемые в медицинской практике аппаратные средства, а также индивидуальные мобильные устройства. Единое пространство для хранения обезличенных данных дает возможность осуществлять исследование значительного массива классифицированной медицинской информации средствами машинного обучения. В рамках разработанной облачной платформы хранения, систематизации и обработки медицинских данных достигнута эффективная работа классификатора, основанного на выявлении потенциально опасных амплитуд. Точность (93 %) и высокая скорость работы данного классификатора являются оптимальными для разработанной облачной платформы.

Ключевые слова: облачные вычисления, классификатор, медицинские исследования, машинное обучение, хранение данных, обработка данных, эпилептическая активность, систематизация данных

AUTOMATIC SEARCH SYSTEM OF EPILEPTIC ACTIVITY SITES IN THE COMPOSITION OF CLOUD STORAGE, SYSTEMATIZATION AND MEDICAL DATA PROCESSING

¹Alekseev D.M., ¹Minyuk A.N., ²Ponimash Z.A., ¹Shumilin A.S.

¹Federal Autonomous Educational Institution of Higher Education Southern Federal University,
Institute of Computer Technologies and Information Security, Taganrog, e-mail: minyuk@neurotech.ru;

²Federal Autonomous Educational Institution of Higher Education Southern Federal University,
Institute of Radio Systems and Control, Taganrog, e-mail: ponimash@neurotech.ru

In this article the main purpose of the research is developing and evaluating the effectiveness of a general scheme of a cloud platform that ensures the performance of a certain range of tasks, as well as to select and analyze a classifier to solve the problem of identifying zones of epileptic activity as part of a cloud platform. Result of research is a scheme of cloud storage platform, systematization and processing of medical data has been developed, which includes four main levels. The division of data streams into levels, standardization of information transfer protocols and its storage formats, allows creating a flexible, universal and highly reliable medical information system. The architecture allows you to quickly integrate into existing medical systems and implement the hardware used in medical practice, as well as individual mobile devices. A single storage space for anonymous data makes it possible to carry out a study of a significant array of classified medical information using machine learning tools. Within the framework of the developed cloud storage platform, systematization and processing of medical data, an efficient operation of the classifier based on identifying potentially dangerous amplitudes was achieved. Accuracy (93 %) and high speed of this classifier are optimal for the developed cloud platform.

Keywords: cloud computing, classifier, medical research, machine learning, data storage, data processing, epileptic activity, data systematization

В век всеобщей информатизации и активного развития информационных технологий медицинские учреждения в ходе выполнения диагностических исследований обрабатывают и систематизируют значительные объемы данных для последующей реабилитации и лечения пациентов в целом. Эффективность оказываемой

медицинской помощи прямо пропорциональна оперативности и удобству использования данной информации специалистами медицинских организаций. Наличие задач, связанных с хранением, систематизацией и обработкой увеличивающихся объемов данных, обуславливает актуальность разработки и интеграции в медицинские уч-

реждения медицинских информационных систем (МИС).

Возможность оперирования данными в электронном виде обеспечивает оперативность получения врачом необходимой информации о пациенте, что увеличивает скорость принятия решения о постановке диагноза и методах лечения [1].

В настоящий момент наиболее актуальными проблемами при реализации медицинских систем являются проблемы, связанные с идентификацией, классификацией и систематизацией нервно-психических заболеваний. К классу данных патологий относят эпилепсию как одно из наиболее распространенных нервно-психических заболеваний, которому подвержено около 1 % человечества.

Электроэнцефалография (ЭЭГ), которая была предложена австрийским психиатром и психофизиологом Хансом Бергером в 1924 г., на сегодняшний день рассматривается как основной метод исследования функциональной активности головного мозга. ЭЭГ – это запись суммарной электрической активности мозга, которая позволяет судить о его физиологической зрелости, функциональном состоянии, общемозговых расстройствах и их характере. Активность нейронов мозга при эпилептической активности отличается от нормального состояния, что демонстрируется в сигналах ЭЭГ. Обнаружение эпилептической активности в сигналах ЭЭГ представляет собой достаточно сложную задачу, в связи с тем, что до настоящего времени неизвестен точный механизм возникновения эпилептических припадков. Разработано множество методов для изучения и анализа сигналов ЭЭГ с целью выявления патологических изменений мозга во время эпилептических приступов. Для автоматического обнаружения эпилептической активности с использованием сигналов ЭЭГ в основном используются линейные (частотно-временные, математические и статистические) методы, где не учитывается нелинейность анализируемого сигнала [2].

Одним из актуальных направлений в области разработки и реализации систем хранения, систематизации и обработки медицинских данных является использование возможностей облачных сервисов.

Основной целью реализации облачной платформы является создание единого информационного пространства для сбора, хранения и предоставления результатов медицинских исследований, с использованием распределенной команды квалифицированных медицинских специалистов. К категории медицинских исследований

относятся результаты медицинских исследований, проведенных с использованием диагностического оборудования различных производителей.

Полученные данные могут использоваться как медицинскими учреждениями, так и научно-исследовательскими организациями. Пациент может предоставлять результаты собственных медицинских исследований другим пользователям облачной платформы или группам квалифицированных медицинских специалистов. Данные могут быть использованы медицинским персоналом, который оказывает комплекс услуг по их исследованию, анализу или экспертизе, после чего предоставляет результаты исследований.

Одной из ключевых задач при реализации облачной системы хранения, систематизации и обработки медицинских данных является разработка системы автоматического поиска участков эпилептической активности. В рамках работы цель исследований заключается в разработке и оценке эффективности общей схемы облачной платформы, обеспечивающей выполнение определенного спектра задач, а также в выборе и анализе классификатора для решения задачи выделения зон эпилептической активности как части облачной платформы.

Основными задачами при реализации облачной платформы являются:

- предоставление эффективных и удобных механизмов для обмена данными исследований между различными пользователями платформы;
- реализация интерфейсов для интеграции в существующие медицинские информационные системы (Hospital Information Systems);
- создание многофункционального пользовательского интерфейса и базы данных для системы анализа с использованием алгоритмов машинного обучения;
- разработка облачного сервиса (SaaS) для хранения, обработки и классификации медицинских данных.

Материалы и методы исследования

При решении задачи выбора и анализа классификатора для выделения зон эпилептической активности предполагается оценка трех ее типов:

- абсансная;
- пик – медленные волны;
- острые волны.

В качестве материалов исследования при построении системы автоматизированного поиска зон ЭПИ-активностей используются четыре типа классификаторов:

1. Классификатор, основанный на построении нормированной взаимокорреляционной функции (далее ВКФ).

2. Классификатор, базирующийся на обучении нейрона с сигмоидальной функцией активации и с MSE (Mean Square Error – средний квадрат ошибки) как функцией потерь.

3. Классификатор, построенный на базе одномерной сверточной сети.

4. Классификатор, основанный на выявлении потенциально опасных амплитуд.

Для решения задачи хранения, систематизации и обработки медицинских данных разработана облачная платформа, общая схема которой представлена на рис. 1.

Разработанная облачная система включает 4 основных уровня:

1. Data Storage layer: представляет собой хранилище данных «Global Storage», которое включает базу данных для хранения обследований и обезличенных отчетов, а также различную диагностическую, антропометрическую и демографическую информацию о пациентах, не используя привязку к персональным данным. Можно сделать вывод, что хранилище данных содержит полную информацию для обучения машинных алгоритмов. Безопасность данных достигается за счет идентификации пациента по защищенному уникальному идентификатору.

2. Data Consumer layer (уровень потребителей данных): включает в себя системы, осуществляющие прием и обработку данных из «Global Storage», а также передачу в «Global Storage» новых данных. Данный уровень связан с уровнем «Global Storage» посредством стандартизированного протокола «Storage API». Потребителями данных могут являться:

– AI/ML Server (Artificial Intelligence/Machine Learning): реализует алгоритмы машинного обучения для анализа и классификации накопленных медицинских данных.

– HIS (hospital information system) – медицинская информационная система, разработанная сторонними

производителями (Third-party HIS software), которая имеет инструменты для взаимодействия с «Global Storage» посредством «Storage API».

– CloudIn Workflow Manager – разрабатываемая система, которая содержит базу персональных данных пациентов. Реализация системы планируется в соответствии с требованиями HIPAA Compliance [3].

3. Software layer: содержит программные средства конечных клиентов, по результатам обследования которых формируются и отображаются медицинские данные (обследования в виде сигналов, графиков и диаграмм, а также отчетные и персональные данные пациентов).

4. Hardware layer: на данном уровне рассматривается аппаратное диагностическое оборудование различных производителей для проведения обследований: например, электроэнцефалографы для анализа функциональной активности головного мозга (наличия ЭПИ-активности).

Результаты исследования и их обсуждение

Рассмотрим предложенные типы классификаторов, используемых для построения системы автоматизированного поиска зон ЭПИ-активностей.

1. Классификатор, основанный на построении нормированной ВКФ: в качестве паттерна использовались сигналы с известной меткой класса. Если значение ВКФ превышает 0,8, то считается, что найдена ЭПИ-активность нужного класса. Такой метод обеспечивает среднюю точность около 60%, однако вероятность ложной тревоги составляет 0,2, что является недопустимым для решения поставленной задачи.

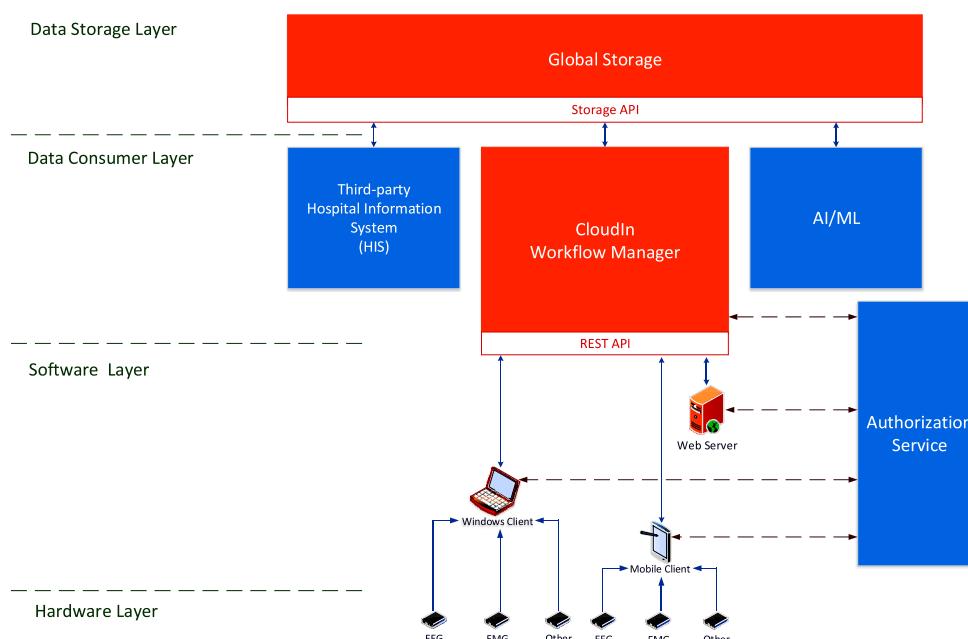


Рис. 1. Общая схема облачной платформы хранения, систематизации и обработки медицинских данных

2. Классификатор, базирующийся на обучении нейрона с сигмоидальной функцией активации и с функцией потерь Mean Square Error (MSE) – средний квадрат ошибки [4]. Имеется выборка из 10 сигналов на каждый класс, которая искусственно расширена с помощью добавления шума с нормальным и равномерным распределением, добавления тренда и низкочастотных составляющих, соответствующих дельта-волнам. После расширения выборка составляет 300 сигналов на каждый класс. Средняя точность на тестовой выборке увеличилась до 83 %, вероятность ложной тревоги составила 0,05. Ложное срабатывание наблюдается только на артефактах, схожих с ЭПИ. Наибольшее количество ложных срабатываний характерно для шага > 1 в связи с повышением порога. Результат работы данной системы с учетом указанных условий представлен на рис. 2.

Недостатком рассмотренной системы является «сканирование» сигнала окна с шагом равным 1. Результат прохождения одного слоя нейронной сети записывается в виде

$$Y(x) = f_a(W \cdot x);$$

$$f_a(x_i) = \frac{1}{1 + e^{-x_i}}.$$

где Y – вектор выхода, x – вектор входа, W – матрица весовых коэффициентов.

Проблема заключается в низкой скорости работы данной системы в связи с необходимостью постоянного умножения матрицы на вектор (для средней записи требуется около 500000 умножений). Проход

с шагом > 1 неосуществим, так как классификатор не инвариантен к смене фазы, из-за чего уровень сигнала снижается с 0,9 до 0,7.

3. Классификатор, построенный на базе одномерной сверточной сети: сверточная сеть инвариантна к сдвигу фазы.

Свёрточная нейронная сеть – специальная архитектура искусственных нейронных сетей, предложенная в 1988 г. Яном Лекуном и нацеленная на эффективное распознавание изображений, входит в состав технологий глубокого обучения [5].

Архитектура системы следующая:

1DConv(3) → Prelu → Pooling →
1DConv(2) → 1DConv(3) → Prelu →
Pooling → 1DConv(2) → 1DConv(3) →
Prelu → Pooling → Full → Prelu → Softmax,

где 1DConv(x) – одномерный сверточный слой, x – размер ядра свертки, Prelu – функция активации P-Relu, Pooling – слой Max Pooling (обеспечивает выбор из окрестности максимального значения, что позволяет системе быть инвариантной к сдвигам сигнала), Full – полносвязный слой, Softmax – полносвязный softmax-слой. В качестве функции ошибки используется перекрестная энтропия.

При расширении выборки наряду с методами, использующимися для второго классификатора, были применены сдвиги сигнала на случайное число отсчетов (равное 9).

Установлено, что точность составляет 95 % при условии сканирования сигнала с шагом 27. Это позволяет повысить скорость в 3 раза, однако подобный прирост недостаточен. Пример работы данного классификатора представлен на рис. 3.

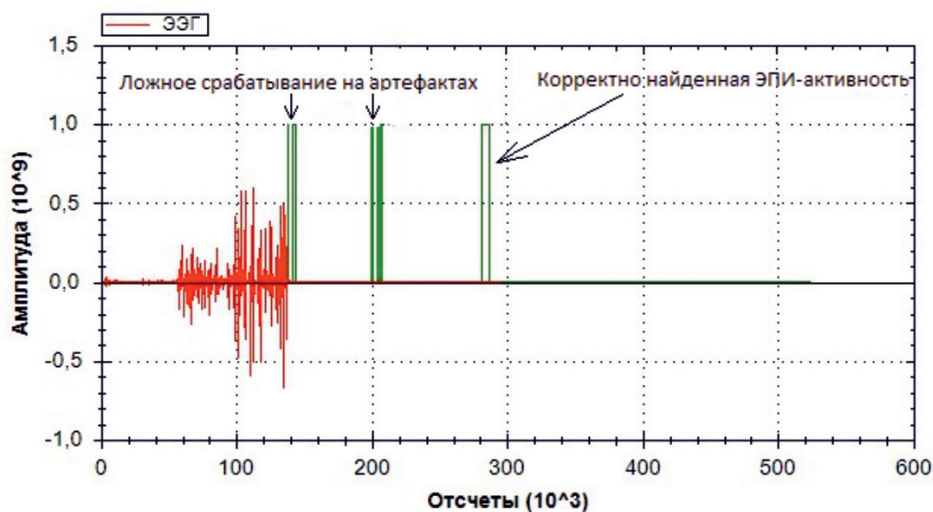


Рис. 2. Результат работы системы (шаг = 3)

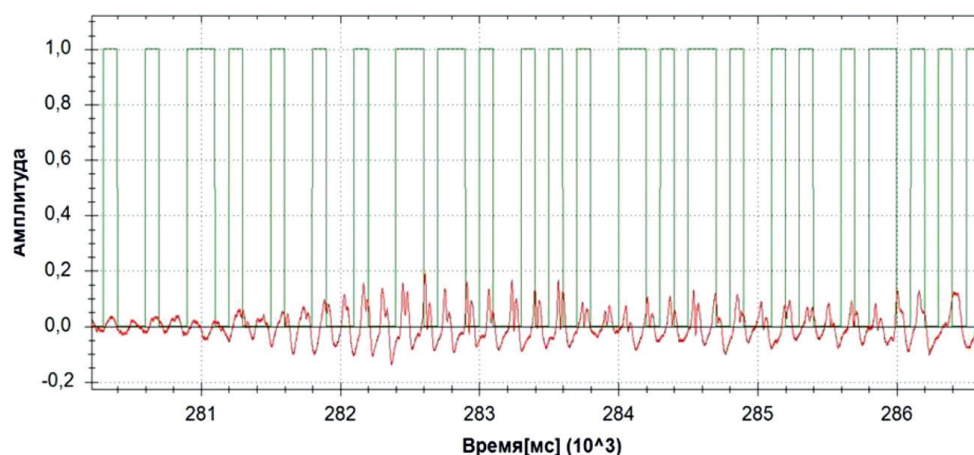


Рис. 3. Пример работы классификатора на базе одномерной сверточной сети

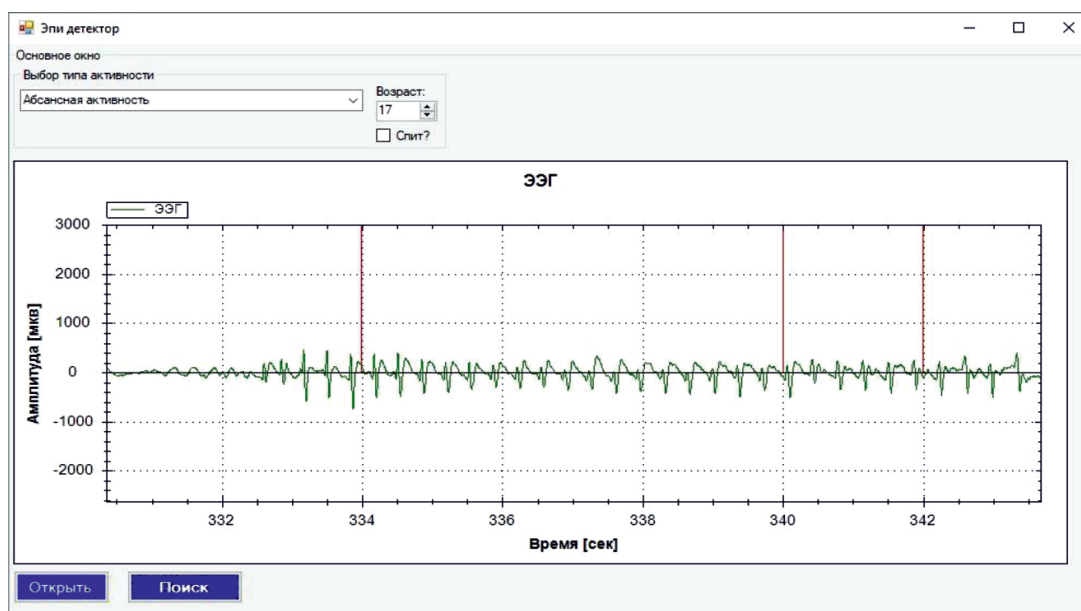


Рис. 4. Результат работы классификатора в «ЭПИ-детектор»

4. Классификатор, основанный на выявлении потенциально опасных амплитуд: для установления порога учитываются следующие данные – возраст человека (взрослый человек или ребенок) и текущее состояние (сон или бодрствование). Сигналы, у которых амплитуда превышала порог, классифицируются с помощью следующего алгоритма: вычислялся модуль спектра сигнала, который классифицируется с помощью метода ближайшего соседа, где функция расстояния представляют собой взвешенную сумму обратной величины к расстоянию в пространстве L2 и коэффициента корреляции.

$$r(x^{(1)}, x^{(2)}) = \frac{w_1}{1 + \sqrt{\sum_{i=0}^N (x_i^{(2)} - x_i^{(1)})^2}} + \frac{w_2 \cdot \text{cov}(x^{(1)}, x^{(2)})}{\sqrt{D(x^{(1)}) \cdot D(x^{(2)})}}.$$

Точность данной системы на тестовой выборке составляет 93 %. Результат работы выбранного классификатора, реализованного в рамках тестового прикладного программного обеспечения «ЭПИ-детектор», представлен на рис. 4.

Выводы

Оценка эффективности разработанной облачной платформы хранения, систематизации и обработки медицинских данных.

Иерархичное разделение потоков данных на уровни, стандартизация протоколов передачи данных и форматов их хранения обеспечивают создание универсальной, гибкой и надежной медицинской информационной системы. Разработанная архитектура позволяет быстро интегрироваться в существующие медицинские системы. Единое пространство для хранения обезличенных данных дает возможность осуществлять исследование значительного массива классифицированной медицинской информации средствами машинного обучения.

В рамках разработанной облачной платформы хранения, систематизации и обработки медицинских данных достигнута эффективная работа классификатора, основанного на выявлении потенциально опасных амплитуд. Точность (93%) и высокая скорость работы данного классификатора являются оптимальными для разработанной облачной платформы.

Стоит отметить, что данные результаты получены с использованием небольшой тестовой выборки, т.е. количество сигналов

на каждый класс измерялось десятками. Для повышения репрезентативности выборки необходимо ее увеличение. В связи с этим направления дальнейшей работы в области исследований по данной тематике связаны с увеличением тестовой выборки при проведении экспериментов по анализу классификаторов для автоматического поиска ЭПИ-активности.

Работа выполнена при поддержке гранта РФФИ № 18-07 00 50.

Список литературы

1. Митькина П.А. Особенности хранения медицинской информации // Современные научные исследования и инновации. 2017. № 5 [Электронный ресурс]. URL: <http://web.snauka.ru/issues/2017/05/82546> (дата обращения: 27.11.2018).
2. Артеменко С.В. Нейросетевые алгоритмы обработки электроэнцефалограмм для диагностики эпилепсии: автореф. ... канд. техн. наук [Электронный ресурс]. URL: <https://www.bsu.by/ru/sm.aspx?guid=814633> (дата обращения: 26.11.2018).
3. Health Insurance Portability and Accountability Act [Электронный ресурс]. URL: https://en.wikipedia.org/wiki/Health_Insurance_Portability_and_Accountability_Act (дата обращения: 28.11.2018).
4. MSE [Электронный ресурс]. URL: <https://ru.wikipedia.org/wiki/MSE> (дата обращения: 28.11.2018).
5. Свёрточная нейронная сеть [Электронный ресурс]. URL: https://ru.wikipedia.org/wiki/Свёрточная_нейронная_сеть (дата обращения: 29.11.2018).