



ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ В АНАЛИЗЕ БОЛЬШИХ МЕДИЦИНСКИХ ДАННЫХ ДЛЯ РАННЕЙ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ

Александров Данила Сергеевич

аспирант факультета прикладной математики,

Национальный исследовательский Томский государственный университет, г.

Томск, Российская Федерация

Аннотация

В современных условиях здравоохранения анализ больших объёмов медицинской информации становится ключевым фактором в обеспечении точной и своевременной диагностики заболеваний. Машинное обучение, как одно из направлений искусственного интеллекта, активно применяется для обработки и интерпретации медицинских данных, включая изображения, геномные последовательности, записи ЭКГ и анамнез пациентов. В статье рассматриваются алгоритмы и методы машинного обучения, применяемые для раннего выявления онкологических, сердечно-сосудистых, неврологических и инфекционных заболеваний. Анализируются примеры реального внедрения, а также проблемы и перспективы дальнейшего развития технологий в странах СНГ. Делается вывод о необходимости создания национальных платформ больших медицинских данных с интеграцией ИИ-инструментов.

Ключевые слова: машинное обучение, большие медицинские данные, ранняя диагностика, искусственный интеллект, здравоохранение, предиктивная аналитика, биомедицинская информатика

1. Введение

С ростом цифровизации медицины в клиниках, лабораториях и научных центрах накапливаются огромные объёмы разнородной информации: изображения КТ и МРТ, биохимические анализы, генетические данные, электронные медицинские карты. Однако эффективное использование этих данных для диагностики и профилактики заболеваний требует новых подходов. Машинное обучение (МО) позволяет извлекать значимые закономерности, выявлять паттерны и строить прогнозные модели, обеспечивая переход от реактивной к превентивной медицине.

2. Типы данных и источники

2.1. Структурированные и неструктурированные данные

- **Структурированные:** показатели анализов, возраст, пол, индекс массы тела и др.
- **Неструктурированные:** текстовые записи, изображения, аудио, ЭКГ, врачебные заключения.

2.2. Основные источники

- Электронные медицинские карты (ЭМК)
- Диагностические изображения (рентген, КТ, МРТ, УЗИ)
- Сенсоры и устройства мониторинга (носимая электроника)
- Геномные и протеомные данные
- Социально-демографические сведения

3. Алгоритмы машинного обучения

3.1. Обучение с учителем

Применяется для задач классификации и регрессии. Примеры:

- Логистическая регрессия
- Метод опорных векторов (SVM)
- Градиентный бустинг (XGBoost, LightGBM)

3.2. Обучение без учителя

Используется для кластеризации и выявления аномалий. Примеры:

- Метод k-средних
- Иерархическая кластеризация
- Autoencoder для обнаружения редких патологий

3.3. Глубокое обучение

Алгоритмы, основанные на нейросетях, особенно эффективны в обработке изображений и сигналов:

- Convolutional Neural Networks (CNN) — для анализа МРТ, рентгена
- Recurrent Neural Networks (RNN) — для временных рядов, ЭКГ
- Transformer-модели — для текстовой медицинской документации

4. Примеры применения

4.1. Онкология

Машинное обучение позволяет выявлять опухоли на ранних стадиях по изображениям КТ и МРТ. Например, в проектах IBM Watson и Google Health достигнута точность свыше 90% в определении рака молочной железы.

4.2. Кардиология

Модели ИИ анализируют данные ЭКГ и выявляют признаки инфаркта, аритмий и сердечной недостаточности задолго до клинических проявлений. Используются в системах удалённого мониторинга.

4.3. Неврология

Глубокие нейросети применяются для ранней диагностики болезни Альцгеймера и рассеянного склероза по данным МРТ и нейропсихологических тестов.

4.4. Инфекционные заболевания

Анализ симптомов, историй болезни и эпидемиологических факторов позволяет моделировать вспышки заболеваний и выявлять риски распространения.

5. Проблемы внедрения в странах СНГ

5.1. Нехватка унифицированных медицинских данных

Медицинские учреждения используют разнородные ИТ-системы, отсутствует единый формат хранения и доступа к данным, что осложняет обучение моделей.

5.2. Этические и правовые аспекты

Возникают вопросы согласия пациентов, хранения персональных данных, ответственности за ошибки алгоритма.

5.3. Недостаточная подготовка кадров

Необходимы специалисты на стыке медицины и ИТ, способные разрабатывать, внедрять и верифицировать ИИ-модели.

6. Перспективы

- Создание национальных платформ больших медицинских данных
- Развитие биобанков и медицинских регистров с открытым доступом
- Интеграция ИИ в клиническую практику через телемедицину и мобильные приложения

- Сертификация ИИ-моделей в здравоохранении и обеспечение их верификации

Заключение

Машинное обучение становится неотъемлемой частью современной медицины, открывая путь к более точной, быстрой и индивидуализированной диагностике. Внедрение ИИ в анализ медицинских данных особенно актуально для стран СНГ, где цифровизация здравоохранения развивается активно, но требует структурных усилий для интеграции интеллектуальных решений. При грамотной организации и соблюдении этических норм использование МО повысит качество медицинских услуг и продвинет профилактику заболеваний на новый уровень.

Литература

1. Esteva A. et al. A guide to deep learning in healthcare. Nature Medicine. 2019.
2. Miotto R., Wang F., Wang S., Jiang X., Dudley J.T. Deep learning for healthcare: Review, opportunities and challenges. Briefings in Bioinformatics. 2018.
3. Shickel B. et al. Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. Journal of Biomedical Informatics. 2020.
4. Воронцов К.В. Машинное обучение и анализ медицинских данных. — М.: МФТИ, 2022.
5. Рахимов А.А., Мусаев Н.Б. Использование ИИ в диагностике хронических заболеваний. Ташкентский медицинский журнал, 2023.
6. IBM Watson Health. Artificial Intelligence in Oncology. 2023.
7. Минздрав РФ. Концепция развития ИИ в медицине до 2030 года. — М., 2023.
8. Zhang Y. et al. Explainable AI in healthcare: A survey. IEEE Access. 2021.