

**Фомин Андрей Валерьевич,**  
студент магистратуры 1 курса гр. Ф2-ИСТ-51м,  
ФГБОУ ВО «Поволжский государственный университет  
телекоммуникаций и информатики»  
Fomin Andrey Valerevich,  
1st-year Master's student, Group F2-IST-51m,  
Volga State University of Telecommunications and Informatics

**Краснов Сергей Викторович, д.т.н., проф.,**  
ФГБОУ ВО «Поволжский государственный университет  
телекоммуникаций и информатики»  
Krasnov Sergey Viktorovich, Dr. Sci. (Tech.), Professor,  
Povolzhskiy State University of Telecommunications and Informatics

**СОВРЕМЕННЫЕ МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ  
ДЛЯ АНАЛИЗА СТРУКТУРИРОВАННЫХ МЕДИЦИНСКИХ ДАННЫХ  
MODERN MACHINE LEARNING METHODS  
FOR ANALYZING STRUCTURED MEDICAL DATA**

**Аннотация.** В статье рассматриваются современные подходы машинного обучения для анализа структурированных медицинских данных, построению предсказательных моделей и обеспечению их интерпретируемости. Особое внимание уделено методам обучения представлений и построению ансамблей моделей. Исследованы перспективы применения таких систем для прогнозирования рисков заболеваний.

**Abstract.** The article examines modern machine learning approaches for analyzing structured medical data, building predictive models, and ensuring their interpretability. Special attention is paid to representation learning methods and the construction of model ensembles. The prospects for applying such systems to disease risk forecasting are investigated.

**Ключевые слова:** Машинное обучение, представление признаков, ансамблевые методы, обучение представлений, предиктивное моделирование.

**Keywords:** Machine learning, feature representation, ensemble methods, representation learning, predictive modeling.

Современная медицина характеризуется стремительным увеличением массива данных, связанных с диагностикой, лечением, наблюдением и мониторингом пациентов. Электронные медицинские записи, лабораторные показатели, сведения о приёмах лекарственных препаратов и результаты обследований представляют собой ценный информационный ресурс, однако их анализ традиционными статистическими методами часто оказывается затруднительным из-за высокой вариативности и гетерогенности данных. Машинное обучение предлагает инструменты, способные выявлять скрытые зависимости, учитывать сложные взаимодействия признаков и обеспечивать адаптивность моделей к различным структурам данных. Это делает ML одним из наиболее перспективных направлений в аналитике медицинских данных, особенно при решении задач ранней диагностики и прогнозирования риска заболеваний [1].

Одним из ключевых направлений развития является создание методов представления медицинской информации в форме, удобной для работы алгоритмов. Начнём с многоуровневых моделей представления, они позволяют преобразовывать диагнозы, лечебные процедуры и структуры визитов в компактные векторные представления. Эти



подходы лежат в основе архитектуры Med2Vec, демонстрирующей способность отражать скрытые связи между медицинскими концептами и создавать эмбединги, увеличивающие эффективность последующих задач классификации и прогнозирования [3].

Значительный интерес для анализа структурированных медицинских данных представляют и методы глубокого обучения, ориентированные на моделирование последовательности событий в истории пациента. В систематическом обзоре исследований по обучению представлений показано, что рекуррентные нейронные сети и более современные трансформерные архитектуры способны учитывать временную структуру данных и динамику изменения показателей. Такие модели обрабатывают не отдельные признаки изолированно, а всю последовательность визитов, что позволяет фиксировать развитие заболевания во времени и выявлять закономерности, недоступные традиционным методам. Благодаря этому подходы, основанные на анализе последовательностей, обеспечивают более точный прогноз хронических и длительно развивающихся состояний и повышают качество долгосрочных медицинских прогнозов [2].

Не менее важным направлением является выбор алгоритмов классификации и построение ансамблей. Сегодня большинство применяемых решений используют модели на основе деревьев решений, случайных лесов и градиентного бустинга. Однако стоит отметить ограниченность комплексной методологии, включающей сравнение нескольких алгоритмов, формирование ансамблей и обязательную внешнюю валидацию. Такой подход значительно повышает устойчивость и надёжность моделей, что особенно важно в медицинском контексте, где ошибки классификации могут иметь критические последствия. Использование ансамблевых методов позволяет компенсировать недостатки отдельных алгоритмов и улучшить качество предсказаний на разнообразных выборках [1].

Теперь поговорим, про интерпретацию результатов машинного обучения, что имеет ключевое значение в клинической практике. Врачи должны понимать, какие факторы повлияли на прогноз, и иметь возможность оценить обоснованность решения модели. Современные подходы включают методы explainable ai, позволяющие отображать индивидуальные влияния параметров пациента на итоговое решение. Это обеспечивает прозрачность и повышает доверие со стороны клиницистов. Дополнительным фактором является необходимость калибровки вероятностей, поскольку медицинские модели нередко используются для определения индивидуальных рисков, выраженных в процентах. Калиброванные вероятности делают выводы более интерпретируемыми и применимыми в реальной клинической практике [1].

Таким образом, современные методы машинного обучения формируют фундамент для построения эффективных систем анализа медицинских данных. Они обеспечивают гибкость, адаптивность и возможность интеграции различных источников информации. Обоснованный выбор архитектуры, корректная обработка данных, интерпретируемость и строгая валидация позволяют создавать модели, пригодные как для научных исследований, так и для реального применения в клинической практике. Машинное обучение демонстрирует значительный потенциал в прогнозировании риска заболеваний и поддержке принятия решений на уровне «пациент-врач», а развитие методов представления данных и объяснимого ИИ делает такие системы всё более надёжными и прозрачными.

#### *Список литературы:*

1. Brnabic A., Hess L. Систематический обзор методов машинного обучения, используемых для анализа данных реальной клинической практики при поддержке принятия решений между пациентом и врачом [Электронный ресурс] - Электронные текстовые данные. - Режим доступа: <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-021-01403-2>, свободный. - Загл. с экрана.



2. Si Y., Du J., Li Z., Jiang X., Miller T., Wang F., Zheng W.J., Roberts K. Глубокое обучение представлений данных пациентов из электронных медицинских записей (EHR): систематический обзор [Электронный ресурс] -Электронные текстовые данные. - Режим доступа: <https://arxiv.org/pdf/2010.02809>, свободный. - Загл. с экрана.

3. Choi E., Bahadori M.T., Searles E., Coffey C., Sun J. Многоуровневое обучение представлений медицинских концептов [Электронный ресурс] - Электронные текстовые данные. - Режим доступа: <https://arxiv.org/pdf/1602.05568>, свободный. - Загл. с экрана.

