

**ПРИМЕНЕНИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В МЕДИЦИНЕ**

*Абдуллахужаев Саидисломбек Немамиллаевич*

*Ташкентский университет информационных технологий имени Мухаммада аль-Хорезми*

**АННОТАЦИЯ:** В данной статье рассматриваются подходы машинного обучения, используемые в медицинской практике для предсказания острых заболеваний, включая онкологические болезни. Оценены различные модели, такие как сверточные нейронные сети (CNN), генеративные состязательные сети (GAN), искусственные нейронные сети (ANN) и другие алгоритмы. Результаты сравнительного анализа показали, что CNN обладают наивысшей точностью, что делает их наиболее подходящими для диагностики и предсказания рака. Акцентируется внимание на значимости комбинирования различных методов, что способствует повышению точности медицинских прогнозов.

**КЛЮЧЕВЫЕ СЛОВА:** машинное обучение, предсказание, острые заболевания, диагностика, рак, сверточные нейронные сети, генеративные состязательные сети, искусственные нейронные сети, медицинские изображения, геномные данные, логистическая регрессия, анализ данных, сегментация, градиентный бустинг, байесовские сети, машины опорных векторов, деревья решений, многомерные данные, прогнозирование рецидивов, глубокие нейронные сети, методы уменьшения размерности.

**ВВЕДЕНИЕ:**

Машинное обучение (МО) представляет собой одну из самых быстро развивающихся направлений науки и технологий, активно используемое в разных сферах жизни, включая медицину. С учётом роста объемов медицинских данных, усложненности их обработки и необходимости оперативного принятия решений, методики МО стали важнейшими ресурсами для диагностики, прогноза и терапии заболеваний. Это особенно критично в таких областях, как онкология, где высокая точность и прогноза имеют жизненно важное значение для спасения людей.

Одним из многообещающих направлений является использование моделей машинного обучения для предсказания фенотипов острых заболеваний, таких как рак. Современные технологии, включая сверточные нейронные сети (CNN), генеративные состязательные сети (GAN) и искусственные нейронные сети (ANN), эффективно анализируют как медицинские изображения, так и многомерные клинические и генетические данные, что значительно увеличивает точность диагностики. Например, CNN показали выдающиеся результаты в

интерпретации медицинских изображений, таких как рентгеновские и МРТ, обеспечивая высокую точность в классификации опухолей.

Ещё одной важной областью является анализ генетической информации для предсказания заболеваний, где методы МО помогают идентифицировать ключевые маркеры и оценивать риск развития болезней, включая рак. Генеративные состязательные сети помогают справляться с проблемой нехватки данных, создавая синтетические медицинские изображения, что способствует улучшению обучения моделей.

В данной статье представлен обзор использования различных методов машинного обучения в диагностике и прогнозировании острых заболеваний, акцентируя внимание на их эффективности в анализе медицинских изображений и клинических данных. Также произведён анализ различных моделей, их применимости и результатов, что помогает выявить наиболее перспективные подходы для практического применения в медицине.

**Выживаемость** в процентах — это процент людей, которые остаются живыми после диагноза рака через определенный промежуток времени. Например, если выживаемость составляет 67%, это означает, что 67% пациентов, у которых был диагностирован рак, выживают через 5 лет после диагностики.

В данной таблице, содержащая данные о заболеваемости, смертности и выживаемости от рака в 2023 году для десяти стран. Эти данные были собраны из Всемирной организации здравоохранения (ВОЗ) и World Population Review:

№	Страна	Заболеваемость (на 100,000)	Смертность (на 100,000)	Выживаемость (5-летняя, %)
1	Новая Зеландия	679	100.2	70
2	Бельгия	695.3	92.4	67
3	Норвегия	731.3	91.8	70
4	Швейцария	664.8	78.8	68
5	Австралия	527.8	84.6	69
6	Канада	473.5	96.4	67
7	США	324.7	82.3	67
8	Франция	350.2	106.7	68
9	Германия	432.5	99.7	66
10	Италия	358.9	94.2	66

---

## ОСНОВНАЯ ЧАСТЬ:

### **Применения для анализа сверточные нейронные сети (CNN)**

Сверточные нейронные сети находят широкое применение в области анализа медицинских изображений, таких как рентгеновские снимки, магнитно-резонансная томография и ультразвуковые исследования. Они оказываются полезными при диагностике заболеваний, классификации опухолей и сегментации изображений. К примеру, CNN, такие как GoogLeNet, демонстрируют точность свыше 90% в идентификации злокачественных опухолей, что делает их очень эффективными для оценки риска рака. Кроме того, сегментация изображений помогает точно локализовать опухоли, что в свою очередь улучшает результаты лучевой терапии и уменьшает возможность повреждения здоровых тканей.

### **Применения для создания синтетических данных генеративные состязательные сети (GAN)**

Генеративные состязательные сети (GAN) используются для генерации синтетических медицинских данных, что помогает решить проблему недостатка реальных данных для обучения моделей. Например, GAN применяются для создания изображений МРТ пациентов с опухолями. Внедрение метрики Вассерштейна в модели 3D-alpha-WGAN позволило значительно улучшить качество генерируемых изображений, обеспечивая более реалистичное изображение патологии и точное выделение контуров объектов. Эти изображения активно используются для обучения сверточным нейронным сетям.

### **Методы уменьшения размерности (MeLiF)**

Метод MeLiF используется для анализа геномной экспрессии, что позволяет выделить наиболее значимые признаки для создания моделей. Это особенно актуально в контексте предсказания рака, поскольку геномные данные могут содержать десятки тысяч различных признаков. Сокращение размерности данных способствует более быстрому обучению моделей и повышает их точность. Например, использование MeLiF продемонстрировало высокую эффективность в прогнозировании онкологических заболеваний, основываясь на данных транскриптомов.

### **Применения для прогнозирования рецидивов искусственные нейронные сети (ANN)**

Искусственные нейронные сети применяются для распознавания образов и прогнозирования результатов на основе многомерных данных. В исследовании они использовались для прогнозирования рецидива рака молочной железы. ANN обучались на данных 2441 пациента, что обеспечило высокую точность прогноза (ROC-кривая – 0,726), превзойдя традиционную

---

систему стадирования TNM (ROC-кривая – 0,677). Эти результаты подтверждают, что ANN являются эффективным инструментом для анализа больших и сложных наборов данных, позволяя улучшить качество прогнозов в медицине.[2]

### **Применения для выполнения нелинейной классификации машины опорных векторов (SVM)**

Методы SVM предназначены для проведения нелинейной классификации и построения гиперплоскостей, которые разделяют различные классы, что делает их особенно актуальными для определения опухолей как злокачественных или доброкачественных. Они предоставляют вероятностные результаты, что способствует лучшему пониманию интерпретации данных. SVM показали свою высокую эффективность в анализе белковых характеристик и классификации опухолевых образований, хотя конкретные числовые параметры в данном исследовании не приводятся. Данный метод считается мощным инструментом для детального анализа медицинских данных и диагностики заболеваний.[2]

### **Применения для классификации дерева решений (DT)**

Деревья решений предоставляют четкую визуализацию процессов классификации, организуя переменные в виде структуры, напоминающей дерево. В различных исследованиях эти методы применялись для оценки вероятности развития рака, используя клинические данные. Деревья решений обладают высокой интерпретируемостью и простотой в использовании, однако их точность может быть не столь высока, как у более сложных подходов, таких как искусственные нейронные сети или методы опорных векторов. Тем не менее, они остаются ценным инструментом для анализа данных, особенно в условиях ограниченных вычислительных мощностей.[2]

### **Применения для прогнозирования вероятности возникновения рака Байесовские сети (BN)**

Байесовские сети анализируют вероятностные отношения между разными переменными, включая информацию о генетических, клинических и эпидемиологических факторах. Эти модели использовались для определения риска и предсказания шансов на возникновение рака. Их умение интегрировать различные категории данных делает их ценным инструментом для разработки прогностических моделей. Однако для достижения высокой точности необходима качественная обработка данных и учет сложных взаимодействий между переменными.[2]

### **Использования глубоких нейронных сетей (DNN)**

Используются для анализа многомерных данных (клинические данные, экспрессия генов, копий числа аллелей). Например, Multi-modal DNN достигла AUC 0,845 при интеграции различных типов данных.[3]

### **Использования конволюционных нейронных сетей (CNN)**

Используются для анализа цифровых маммограмм. В исследовании Li и др., CNN показали AUC 0,86, демонстрируя высокую точность в определении высокого и низкого риска рака.[3]

### **Использования GoogLeNet-LDA**

Эта модель была применена для анализа плотности груди и прогнозирования краткосрочного риска. Достигнута AUC 0,73, что показывает её эффективность в работе с изображениями.[3]

### **Использования модели MTP-BCR (Multi-Time Point Breast Cancer Risk)**

Использует продольные данные маммографии для прогнозирования долгосрочного риска рака. Модель достигла AUC 0,80 для 10-летнего риска и 0,81 для 5-летнего риска.[3]

### **Использования модели на основе трансформеров (Transformer-based models)**

Трансформеры применялись для анализа цифровых маммограмм. Например, модель с использованием ResNet-18 и трансформера достигла AUC 0,72 для краткосрочного прогнозирования.[3]

### **Использования внимание-мультимодальные модели (Attention-based Multimodal Models)**

Используют интеграцию данных (геномика, клинические данные, изображения). Эти модели демонстрируют высокий потенциал, достигая AUC 0,95 на комбинированных наборах данных, таких как METABRIC.[3]

### **Применения многофакторной логистической регрессии (ЛР)**

Многомерная логистическая регрессия была использована для построения прогностической модели, учитывающей различные факторы риска, включая возраст, стадию, локализацию опухоли, клинические симптомы и лабораторные показатели, такие как СОЭ и уровень глюкозы в крови. Результаты показали, что модель LR обладает высокой точностью: AUC=0,893, чувствительность 0,849 и специфичность 0,889, что делает ее наиболее точной для прогнозирования рецидива базально-клеточного рака кожи после фотодинамической терапии (ФДТ).[4]

### **Стохастический градиентный бустинг (СГБ)**

Стохастический градиентный бустинг применялся для прогнозирования рецидивов с использованием тех же данных, что и для ЛР, но с учетом нелинейных взаимосвязей между факторами. Эта модель показала несколько

худшие результаты по сравнению с ЛР, с AUC = 0,692, чувствительностью 0,592 и специфичностью 0,764. Несмотря на это, СГБ все же показал свою применимость, но его точность оказалась ниже, что может быть связано с особенностями настройки модели и выборки.[4]

### Однофакторная логистическая регрессия (один фактор за раз)

Однофакторная логистическая регрессия использовалась для оценки влияния каждого фактора в отдельности, включая размер опухоли, возраст, стадию, локализацию и форму опухоли. Результаты показали, что наибольшее влияние на возникновение рецидива оказывают размер опухоли, возраст, стадия II и локализация опухоли в грудной клетке. Однако точность этих моделей была низкой по сравнению с многофакторной регрессией, поскольку учет только одного фактора за раз не учитывал взаимодействие между факторами.[4]

Сравнительная таблица самых основных алгоритмов, методов машинного обучения из статьи:

№	Метод, алгоритм, модель	Где применялся	Результат
1	Сверточные нейронные сети (CNN)	Анализ медицинских изображений (рентген, МРТ, УЗИ, маммограммы)	Высокая точность (90%+) в классификации опухолей и сегментации изображений.
2	Генеративные состязательные сети (GAN)	Генерация синтетических медицинских данных (МРТ)	Создание реалистичных изображений опухолей для обучения моделей.
3	Искусственные нейронные сети (ANN)	Прогнозирование рецидивов рака молочной железы	Точность прогноза (ROC = 0.726), улучшение традиционного метода TNM.
4	Глубокие нейронные сети (DNN)	Анализ многомерных данных (клинические данные, экспрессия генов)	AUC = 0.845 при интеграции различных типов данных.

5	Многофакторная логистическая регрессия (ЛР)	Прогнозирование рецидивов рака кожи (ФДТ)	AUC = 0.893, чувствительность = 0.849, специфичность = 0.889.
6	Внимание-мультимодальные модели	Интеграция геномных, клинических данных и изображений	AUC = 0.95 на комбинированных данных (METABRIC).

### ЗАКЛЮЧЕНИЕ:

В ходе исследования различных методов, алгоритмов и моделей, используемых в машинном обучении для предсказания рака, было установлено, что сверточные нейронные сети (CNN) демонстрируют наилучшие результаты в задачах классификации опухолей и сегментации медицинских изображений. Эффективность этого метода превышает 90 %, что делает его очень полезным для диагностики рака. Генеративные состязательные сети (GAN), генерирующие синтетические данные для целей обучения, и глубокие нейронные сети (DNN), способные интегрировать различные типы информации, также показали замечательные результаты в точности прогнозирования. В отличие от них, многомерная логистическая регрессия (LR) показала надежность в прогнозировании рецидивов рака, но ее производительность не сравнима с CNN, особенно при анализе изображений.

### СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ:

1. Машинное обучение в медицине — Викиконспекты. URL: <https://neerc.ifmo.ru/wiki/index.php?title=%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5%D0%B2%D0%BC%D0%B5%D0%B4%D0%B8%D1%86%D0%B8%D0%BD%D0%B5> (дата обращения: 10.11.2024).
2. Application of machine learning methods in cancer prediction and early. URL: <https://zenodo.org/records/7523833> (дата обращения: 18.11.2024).
3. Breast cancer risk prediction using machine learning: a systematic review. URL: <https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2024.1343627/full> (дата обращения: 29.11.2024).
4. Методы машинного обучения в прогнозировании рецидивов базальноклеточного рака кожи после фотодинамической терапии. URL: <https://www.tmj-vgmu.ru/jour/article/download/2338/2188> (дата обращения: 10.12.2024).