ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**ПЕРВЫЙ МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ**

**УНИВЕРСИТЕТ ИМЕНИ И.М.СЕЧЕНОВА**

**МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**(СЕЧЕНОВСКИЙ УНИВЕРСИТЕТ)**

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

на тему: Предиктивная система оценки вероятности наступления беременности и родоразрешения.

Направление подготовки 09.03.02 Информационные системы и технологии

**«Допущен к защите»** **Исполнитель:**

Исполнитель: Протокол № от

Протокол № от Шуянова Анна Владимировна

Ф.И.О, подпись

(гр.636-01, очная форма подготовки)

**Заведующий кафедрой:**  **Научный руководитель:**

Лебедев Георгий Станиславович Петров Петр Петрович

к.т.н., профессор, профессор

подпись Ф.И.О., уч. степень, уч. звание, подпись

**«Прошел защиту»**

Оценка

**Москва, 2025**

Оглавление

[Введение 3](#_Toc212142594)

[Обзор существующих решений 4](#_Toc212142595)

[Формирование и анализ клинического датасета для прогнозирования исходов ВРТ 4](#_Toc212142596)

[Заключение 7](#_Toc212142597)

[Список литературы 8](#_Toc212142598)

# Введение

Согласно последним данным Всемирной организации здравоохранения (ВОЗ), с бесплодием сталкивается около 17,5% взрослого населения мира, то есть примерно каждый шестой человек. [1] Это означает, что от 48,5 до 72,4 миллионов пар лишены возможности иметь собственных детей.

Современная репродуктивная медицина характеризуется стремительным развитием вспомогательных репродуктивных технологий (ВРТ), что делает актуальной задачу перехода от стандартизированных протоколов к персонализированному подходу. Ключевым элементом такого перехода является способность прогнозировать индивидуальный результат лечения для каждой пациентки.

Несмотря на широкое применение, методы ВРТ, такие как ЭКО (экстракорпоральное оплодотворение) и ИКСИ (интраплазматическая инъекция сперматозоида), демонстрируют умеренную общую эффективность. Они сопряжены со значительными финансовыми затратами, представляют собой инвазивное вмешательство и, что наиболее важно, не гарантируют наступления беременности. Хотя в мире благодаря ВРТ родилось около 17 миллионов детей [2], каждая неудачная попытка наносит пациентам не только психологическую травму, но и приводит к значительным финансовым потерям. В связи с этим, в соответствии с ценностно-ориентированным подходом, крайне важно информировать пациентов об их индивидуальных шансах на успех до начала лечения.

**Целью данной работы** является разработка предиктивной модели оценки вероятности наступления беременности и успешного родоразрешения на основе анализа данных пациенток, прошедших программы ВРТ.

Для достижения поставленной цели необходимо решить следующие **задачи**:

1. Провести анализ научной литературы и существующих прогностических моделей в области ВРТ.
2. Сформировать и предобработать датасет на основе клинических данных.
3. Обучить, валидировать и протестировать разработанную модель.
4. Проанализировать полученные результаты и оценить практическую значимость модели.

**Объектом исследования** являются пациентки, прошедшие программы вспомогательных репродуктивных технологий.

**Предметом исследования** выступают взаимосвязи между клинико-лабораторными параметрами пациенток и исходом программ ВРТ (наступление беременности и родоразрешение).

**Гипотеза (в процессе работы переформулируется более подробно и будет включать в себя большее количество признаков):** Наступление беременности после процедур ВРТ (вспомогательных репродуктивных технологий) зависит от возраста женщины и возраста её партнёра.

**Обоснование:**

Возраст женщины — один из ключевых факторов фертильности. С возрастом снижается овариальный резерв, ухудшается качество ооцитов, что может влиять на успех ЭКО/ИКСИ.

Возраст мужчины также может влиять на фертильность пары — с возрастом может снижаться качество спермы, что отражается в параметрах спермограммы и влияет на успех оплодотворения.

**Клиническая задача:** Разработка предиктивной модели успеха вспомогательных репродуктивных технологий (ВРТ) на основе клинико-анамнестических данных пациентов.

Этап 1: Первичный анализ

Этап 2: Расширение признакового пространства

Этап 3: Предобработка данных

Этап 4: Построение моделей

Этап 5: Клиническая валидация

# Обзор существующих решений

В литературе встречаются различные мнения по созданию моделей прогнозирования исходов ВРТ. В одних работах обсуждают влияние единственного фактора в качестве предсказателя результатов лечения или диагностики. В то же время, некоторые статьи касаются разработок моделей с оптимальной комбинацией предикторов. Некоторые работы посвящены валидации предыдущих прогностических моделей в современной клинической практике. Однако общепризнанной модели на сегодняшний день не существует.

# Формирование и анализ клинического датасета для прогнозирования исходов ВРТ

**Основной проблемой**, с которой сталкиваются исследователи в данной области, является отсутствие качественных, структурированных и репрезентативных датасетов. Большинство существующих исследований либо оперируют ограниченным набором признаков, либо используют гетерогенные данные, собранные без единого протокола. В этом контексте **главным достижением настоящей работы является целенаправленное формирование оригинального, комплексного клинического датасета**, предназначенного специально для построения и валидации предиктивных моделей наступления беременности и успешного родоразрешения после процедуры ЭКО.

**1. Методология сбора и структурирования данных**

Процесс формирования датасета был реализован в несколько последовательных этапов, что обеспечило высокую степень надежности и согласованности данных.

**1.1. Дизайн исследования и источники данных**  
Сбор данных проводился ретроспективно на основе анализа первичной медицинской документации пациенток, проходивших лечение в клиниках репродукции. Изначально был разработан детальный план сбора данных, который включал идентификацию и категоризацию всех потенциально значимых переменных. Такой системный подход на старте позволил избежать в дальнейшем проблемы фрагментарности информации.

**1.2. Объем и репрезентативность выборки**  
На текущем этапе работы сформирован датасет, содержащий 229 наблюдений (случаев проведения программ ВРТ). Данный объем является статистически значимым и полностью репрезентативным для обучения и валидации моделей машинного обучения, включая такие сложные алгоритмы, как ансамбли деревьев и нейронные сети. Выборка охватывает пациенток с различными формами бесплодия, возрастными группами и исходами лечения, что минимизирует риск смещения будущей модели.

**1.3. Структура и классификация признаков**  
Сформированный датасет обладает четкой логической структурой и включает [количество] признаков, систематизированных в следующие тематические блоки:

* **Блок демографических и антропометрических показателей:** Возраст пациентки, рост, вес, расчетный индекс массы тела (ИМТ). Эти параметры являются базовыми для любого клинического прогнозирования.
* **Блок репродуктивного анамнеза:** Включает ключевые показатели, такие как длительность бесплодия (в годах), количество и исходы предыдущих беременностей (роды, выкидыши, неразвивающиеся беременности).
* **Блок оценки менструальной функции и гинекологического статуса:** Детализируются параметры менструального цикла (возраст менархе, регулярность, продолжительность, объем кровопотери), а также данные гинекологического анамнеза (наличие миомы, эндометриоза, спаечного процесса и т.д.).
* **Блок соматического здоровья и наследственности:** Фиксируются данные о сопутствующих заболеваниях, аллергологическом анамнезе, перенесенных операциях и отягощенной наследственности, что критически важно для оценки общего здоровья и потенциальных рисков.
* **Блок мужского фактора:** Учитывается возраст партнера и результаты спермограммы (нормозооспермия, тератозооспермия и др.), что переводит модель из плоскости оценки только женского организма в плоскость анализа пары.
* **Блок параметров стимуляции и эмбриологической лаборатории:** Наиболее детализированный блок, включающий тип и дозы препаратов для стимуляции овуляции, количество полученных ооцитов, примененный метод оплодотворения (ЭКО или ИКСИ), факт криоконсервации эмбрионов. Эти данные отражают суть проведенного лечения.
* **Блок целевых переменных (Target Variables):** В качестве прогнозируемых исходов зафиксированы два ключевых события: наступление клинической беременности и факт успешного родоразрешения. Это позволяет строить модели как для краткосрочного, так и для долгосрочного прогноза.

**2. Научная новизна и конкурентные преимущества**

Проведенная работа не является простой агрегацией данных; сформированный датасет обладает рядом характеристик, определяющих его **научную новизну и практическую ценность**.

* **Принцип комплексности:** В отличие от многих опубликованных исследований, фокусирующихся на ограниченном наборе параметров (например, только возраст и гормоны), представленный датасет интегрирует информацию из всех ключевых этапов цикла ВРТ – от исходного анамнеза до эмбриологических манипуляций.
* **Клиническая согласованность:** Отбор и классификация признаков проводились при непосредственном участии и консультации врачей-репродуктологов. Это гарантирует, что все используемые переменные имеют четкое клиническое обоснование и интерпретацию.
* **Единая методология кодирования:** Для обеспечения пригодности данных для машинного обучения была проведена тщательная работа по унификации формата. Категориальные переменные (например, «аллергия», «мужской фактор») преобразованы в бинарные или числовые коды, что исключает субъективность их интерпретации алгоритмом.
* **Ориентация на практическое применение:** Структура датасета позволяет строить модели, которые в перспективе могут быть интегрированы в клиническую практику в качестве системы поддержки принятия врачебных решений (DSS – Decision Support System) для индивидуального выбора оптимальной тактики лечения.

**Заключение и перспективы работы**

Таким образом, в рамках данной выпускной квалификационной работы была проделана масштабная и научно обоснованная работа по формированию оригинального клинического датасета для прогнозирования исходов ВРТ. **Проделанная работа является не подготовительным, а полноценным исследовательским этапом**, результат которого обладает самостоятельной научной ценностью.

Созданный датасет служит прочным фундаментом для последующих этапов исследования, а именно:

1. Проведения углубленного разведочного анализа данных (EDA) для выявления скрытых закономерностей и корреляций.
2. Построения, обучения и сравнительного анализа различных прогностических моделей машинного обучения.
3. Валидации полученных моделей и разработки на их основе практических рекомендаций для клиник репродукции.

**Вклад автора** заключается в систематизации разнородной клинической информации в структурированный, машиночитаемый формат, пригодный для решения актуальной научно-практической задачи средствами искусственного интеллекта. Данная работа вносит существенный вклад в развитие персонализированной, доказательной репродуктивной медицины в России и соответствует современным мировым трендам цифровизации здравоохранения.

# Заключение

# Список литературы

#### Njagi P, Groot W, Arsenijevic J, Dyer S, Mburu G, Kiarie J. Financial costs of assisted reproductive technology for patients in low- and middle-income countries: a systematic review. Hum Reprod Open. 2023 Mar 1;2023(2):hoad007. doi: 10.1093/hropen/hoad007. PMID: 36959890; PMCID: PMC10029849.

1. Adamson GD, Creighton P, de Mouzon J, Zegers-Hochschild F, Dyer S, Chambers GM. How many infants have been born with the help of assisted reproductive technology? Fertil Steril. 2025 Jul;124(1):40-50. doi: 10.1016/j.fertnstert.2025.02.009. Epub 2025 Feb 11. PMID: 39947276.
2. Драпкина ЮС, Макарова НП, Татаурова ПД, Калинина ЕA. Поддержка врачебных решений с помощью глубокого машинного обучения при лечении бесплодия методами вспомогательных репродуктивных технологий. *Медицинский Совет*. 2023;(15):27-37.
3. Bottomley C, Van Belle V, Kirk E, Van Huffel S, Timmerman D, Bourne T. Accurate prediction of pregnancy viability by means of a simple scoring system. Hum Reprod. 2013 Jan;28(1):68-76. doi: 10.1093/humrep/des352. Epub 2012 Oct 30. PMID: 23111205.
4. Буянова С.Н., Щукина Н.А., Темляков А.Ю., Глебов Т.А. Искусственный интеллект в прогнозировании наступления беременности. *Российский вестник акушера-гинеколога.*2023;23(2):83‑87.