

**Моделирование многоэтапных процессов лечения с поддержкой согласованных  
врачебных решений**

**Писарькова В.П. (ИТМО), Жданова Е.А. (ИТМО)**

**Научный руководитель – кандидат технических наук, доцент Ковальчук С.В.  
(ИТМО)**

**Введение.** Искусственный интеллект активно используется в медицине, и одной из перспективных областей является репродуктивная медицина. В связи с ростом случаев репродуктивных расстройств и необходимостью повышения эффективности лечения бесплодия, разработка прогностических моделей для предсказания наступления беременности становится особенно актуальной. [1] При обучении нейронных сетей используется три типа подхода: обучение с учителем, обучение без учителя и обучение с подкреплением. Каждый из этих подходов обладает потенциалом для решения задач в области лечения бесплодия, однако они существенно различаются по методологии и применимости. [2]

Проблема данной задачи заключается в персонализации решения, где необходимо учитывать особенности, которые были собраны на разных этапах, таких, как диагностическая стадия (выявление причин бесплодия), лабораторная стадия (использование вспомогательных репродуктивных технологий, таких как ЭКО) и стадия переноса эмбрионов. Моделирование многоэтапных процессов лечения позволяет выявить наиболее значимые признаки и получить прогноз на разных этапах лечения. [3]

**Основная часть.** В ходе работы была разработана предсказательная модель для прогнозирования результатов лечения бесплодия, основанная на анамнезах пациентов и характеристиках диагностической и лабораторной стадий. Для имитации многоступенчатой модели были выделены группы признаков, характеризующие этапы процесса лечения, и последовательно добавлены в модель.

После очистки датасета от нерелевантных данных (информации для внутреннего учета данных, ссылок на записи) было отобрано 38 признаков диагностического этапа, включая анамнез, и 24 лабораторных. Текстовые данные анамнеза были предварительно обработаны с помощью библиотек nltk, rymorphy2, созданного вручную словаря и регулярных выражений: выполнены очистка данных (удаление знаков препинания, приведение к нижнему регистру), удаление стоп-слов, токенизация и лемматизация. На основе диагноза пациента была создана новая характеристика – степень бесплодия, принимающая значения 1 (первичное) или 2 (вторичное). К полученным леммам была применена модель word2vec для преобразования в эмбединги с контекстом в пять слов и создающая эмбединги размерностью 32.

Из численных данных было выделено 63 признака, которые были стандартизованы с использованием StandardScaler. Для моделирования применялись два алгоритма: гибридная нейронная сеть с разделением входных слоев для текстовых и табличных данных (LSTM и полносвязные слои), и алгоритм XGBoost. Подбор гиперпараметров осуществлялся на основе F1-меры, оценивающей качество прогнозирования исхода лечения. Для нейронной сети использовалась сигмоидная функция активации и коэффициент скорости обучения  $1e-4$ , для алгоритма XGBoost использовалась функция потерь logloss и максимальная глубина каждого дерева, равная 8. Количество эпох обучения варьировалось от 10 до 100.

Были исследованы три варианта модели: использующая только данные диагностического этапа, только данные лабораторного этапа, и использующая полный набор данных. Это позволило оценить вклад информации, полученной на каждой стадии лечения, и влияние различных факторов на эффективность модели.

**Выводы.** В результате выполнения работы были проанализированы существующие методы использования искусственного интеллекта в репродуктивной медицине и разработана прогностическая модель для предсказания результатов лечения бесплодия.

Первичная модель, построенная на данных диагностического этапа, продемонстрировала точность 0.46. В практическом контексте данный результат позволяет утверждать, что разработанный алгоритм почти в половине случаев способен предсказать исход лечения бесплодия. Применение такой прогностической модели может существенно оптимизировать клинический процесс для врачей и предотвратить необоснованные траты у пациентов (для примера, средняя стоимость цикла ЭКО в Москве и Санкт-Петербурге составляет от 90 тысяч рублей).

Модель, обученная исключительно на данных лабораторного этапа, показала незначительно улучшенную точность - 0.53. Ожидаемо, модель, использующая данные обоих этапов, достигла максимальной точности – 0.6. Несмотря на более высокую точность комбинированной модели, анализ результатов каждого этапа имеет важное значение для выявления важных признаков на ранних стадиях лечения.

Анализ важности признаков для диагностического этапа выявил следующие ключевые характеристики: общее количество ооцитов и зрелых ооцитов, толщина эндометрия, продолжительность и степень бесплодия, концентрация и подвижность сперматозоидов, рост и вес женщины, а также объем эякулята. Примечательно, что фактор "Доктор" также вошел в число наиболее значимых, что подчеркивает важность персонализации в процессе лечения бесплодия.

На лабораторном этапе наибольшее значение имели следующие характеристики: количество перенесённых эмбрионов в полость матки, количество фолликулов, которые были пропунктированы и показатель GR-1 (наряду с другими высокими и средними оценками качества переносимого эмбриона). При полном анализе было выявлено около 15 наиболее важных признаков, включая перечисленные выше.

Наилучший показатель качества модели при полных данных был получен с использованием алгоритма нейронной сети – 0.6, в сравнении с XGBoost, достигнувшей точности 0.51. Обе модели позволили определить ключевые признаки для принятия решений о результатах лечения, визуализация была сделана с помощью графика SHAP.

Эти результаты указывают на необходимость многоэтапного моделирования в задаче лечения бесплодия. Предсказание исхода лечения на ранних этапах, несмотря на меньшую точность, предоставляет ценную информацию для первоначальной оценки и планирования лечения. Однако более точная многоэтапная модель позволяет уточнить прогноз, обеспечивая персонализированный подход к лечению.

#### **Список использованных источников:**

1. Carson, S. A., & Kallen, A. N. (2021). Diagnosis and management of infertility: a review. *Jama*, 326(1), 65-76
2. Wang, R., Pan, W., Jin, L., Li, Y., Geng, Y., Gao, C., Chen, G., Wang, H., Ma, D., & Liao, S. (2019). Artificial intelligence in reproductive medicine. *Reproduction*, 158(4), R139–R154. <https://doi.org/10.1530/REP-18-0523>
3. Zhdanova, E., Korneev, I., & Kovalchuk, S. (2023). Predictive modeling of multistep clinical pathways: application to infertility treatment process. *Procedia Computer Science*, 229, 272–283. <https://doi.org/10.1016/j.procs.2023.12.029>