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: “Прогнозирование риска гестационного диабета с использованием нейронной сети”.

Направление подготовки 09.03.02 Информационные системы и технологии

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

Исполнитель: Протокол № от

Протокол № от Абрамова Алиса Андреевна

Ф.И.О, подпись

(гр.636-01, очная форма подготовки)

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

Лебедев Георгий Станиславович Лебедев Георгий Станиславович

д.т.н., доцент,

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

**«Прошел защиту»**

Оценка

**Москва, 2025**

ОГЛАВЛЕНИЕ

[Введение 4](#_Toc212127129)

[Глава 1. Теоретические основы прогнозирования гестационного диабета 5](#_Toc212127130)

[1.1. Физиология и факторы риска гестационного диабета 5](#_Toc212127131)

[1.2. Методы прогнозирования заболеваний с использованием машинного обучения 6](#_Toc212127132)

[Глава 2. Разработка и обучение модели нейронной сети 7](#_Toc212127136)

[2.1. Постановка задачи и гипотеза исследования 7](#_Toc212127137)

[2.2. Подготовка данных и предобработка 7](#_Toc212127138)

[2.3. Обучение и настройка модели 8](#_Toc212127142)

[Глава 3. Тестирование и анализ результатов 9](#_Toc212127144)

[3.1. Метрики и качество модели 9](#_Toc212127145)

[3.2. Интерпретация результатов и перспективы улучшения 10](#_Toc212127146)

[Заключение 11](#_Toc212127149)

[Список литературы 12](#_Toc212127150)

Введение

Гестационный диабет (ГД) является одним из наиболее распространенных осложнений беременности, представляя серьезную угрозу как для здоровья матери (риск преэклампсии, кесарева сечения и развития сахарного диабета 2 типа), так и для ребенка (макросомия, неонатальная гипогликемия). Своевременное выявление женщин с высоким риском развития ГД является критически важной задачей современного акушерства.

В этом контексте методы искусственного интеллекта, в частности нейронные сети, открывают новые возможности для предиктивной аналитики. Они позволяют выявлять сложные, неочевидные взаимосвязи в медицинских данных и строить высокоточные прогностические модели, что подтверждается последними исследованиями [3].

**Цель работы** — разработать и обучить нейронную сеть для прогнозирования вероятности развития гестационного диабета у женщин на основе физиологических показателей.

**Задачи исследования:**

1. Изучить особенности развития гестационного диабета и ключевые факторы риска.
2. Рассмотреть существующие подходы машинного обучения в прогнозировании заболеваний.
3. Реализовать модель нейронной сети на языке Python.
4. Провести обучение и оценку точности модели.
5. Проанализировать результаты и определить пути повышения эффективности.

**Гипотеза исследования:** нейронная сеть способна предсказать риск развития гестационного диабета с точностью не ниже 75%.

Глава 1. Теоретические основы прогнозирования гестационного диабета

* 1. **Физиология и факторы риска гестационного диабета.**

Гестационный диабет развивается вследствие нарушения метаболизма глюкозы в организме беременной женщины. В период беременности плацента вырабатывает гормоны (эстроген, кортизол, плацентарный лактоген), которые снижают чувствительность тканей к инсулину. Если поджелудочная железа не компенсирует этот эффект увеличением секреции инсулина, уровень глюкозы в крови повышается.

Согласно систематическому обзору Cochrane (Martis et al., 2018), изменение образа жизни остаётся единственным доказанно эффективным методом снижения риска осложнений при гестационном диабете [2]. Это подчёркивает необходимость раннего выявления пациенток из группы риска и разработки методов точного прогнозирования заболевания.

К основным факторам риска ГД относятся:

* избыточная масса тела до беременности (ИМТ > 25);
* возраст старше 25 лет;
* наследственная предрасположенность к диабету;
* наличие поликистоза яичников;
* гипертензия и нарушения липидного обмена;
* случаи гестационного диабета в предыдущих беременностях.

Ранняя идентификация этих факторов позволяет снизить вероятность осложнений, однако традиционные методы диагностики часто оказываются запоздалыми. Это делает актуальным применение алгоритмов машинного обучения для прогнозирования ГД уже на ранних сроках беременности.

1.2. Методы прогнозирования заболеваний с использованием машинного обучения.

Машинное обучение (ML) и искусственные нейронные сети (ANN) активно применяются в медицине для анализа больших объёмов данных и построения предиктивных моделей. Алгоритмы, такие как логистическая регрессия, деревья решений, случайный лес, градиентный бустинг и нейронные сети, позволяют классифицировать пациентов по вероятности наличия заболевания.

Преимущество нейронных сетей заключается в способности моделировать сложные нелинейные зависимости между параметрами, что особенно важно для медицинских данных, где влияние факторов взаимосвязано.

В контексте прогнозирования гестационного диабета значимыми признаками являются:

* количество предыдущих беременностей,
* уровень глюкозы и инсулина,
* индекс массы тела,
* возраст,
* артериальное давление.

В исследовании Li и соавт. (2025) предложена интерпретируемая модель прогнозирования риска ГДМ на основе многослойного перцептрона (MLP), показавшая точность 88,4% и AUC 0.943 [3]. Эти результаты подтверждают эффективность нейронных сетей в медицинской диагностике и легли в основу выбора архитектуры настоящего исследования.

Глава 2. Разработка и обучение модели нейронной сети

2.1. Постановка задачи и гипотеза исследования.

Задача формулируется как бинарная классификация: определить, имеется ли у женщины высокий риск развития гестационного диабета (1) или нет (0).

Гипотеза исследования утверждает, что нейронная сеть способна прогнозировать наличие заболевания по медицинским показателям с точностью выше 75%.

2.2. Подготовка данных и предобработка.

Для обучения использован открытый набор данных *Pima Indians Diabetes Dataset* с платформы Kaggle [1].  
Датасет содержит 768 записей с параметрами: количество беременностей, уровень глюкозы, артериальное давление, толщина кожной складки, уровень инсулина, индекс массы тела, коэффициент наследственности и возраст.

Данные были очищены от пропусков, нормализованы с использованием StandardScaler, затем разделены на обучающую (80%) и тестовую (20%) выборки.

2.3. Обучение и настройка модели.

Для построения модели использовалась библиотека **TensorFlow/Keras** [5].  
Архитектура нейронной сети:

* входной слой (8 признаков);
* два скрытых слоя (16 и 8 нейронов, активация ReLU);
* выходной слой (1 нейрон, активация sigmoid).

Функция потерь: binary\_crossentropy;  
оптимизатор: Adam (learning\_rate = 0.001);  
метрика: accuracy.

Модель обучалась 50 эпох при batch\_size = 8 с валидационным разбиением 0.2. После обучения проведена оценка на тестовой выборке и анализ матрицы ошибок.

Глава 3. Тестирование и оценка результатов

3.1. Метрики и качество модели.

Результаты тестирования показали:

* **Общая точность (accuracy):** 0.71 (71%)
* **Матрица ошибок:**

|  | **Предсказано 0** | **Предсказано 1** |
| --- | --- | --- |
| **Факт 0** | 80 | 20 |
| **Факт 1** | 25 | 29 |

**Класс 0 (“нет диабета”)** — точность 0.76, полнота 0.80, f1-score 0.78.

**Класс 1 (“диабет”)** — точность 0.59, полнота 0.54, f1-score 0.56.

Данные результаты демонстрируют, что модель лучше различает здоровых женщин, чем пациенток с диабетом, что связано с дисбалансом классов в исходных данных.

3.2. Интерпретация результатов и перспективы улучшения.

Полученная точность 71% частично подтверждает гипотезу исследования.

Несмотря на результат ниже ожидаемых 75%, нейронная сеть успешно выявила ключевые зависимости между физиологическими параметрами и риском заболевания.

Для улучшения модели возможно:

* балансировка классов (oversampling/undersampling);
* регуляризация и добавление Dropout-слоёв;
* оптимизация гиперпараметров (GridSearchCV);
* расширение обучающей выборки за счёт клинических данных.

Таким образом, нейронные сети демонстрируют высокий потенциал для задач медицинской диагностики, требуя, однако, дальнейшей настройки и валидации.

Заключение

В ходе исследования построена и протестирована нейронная сеть для прогнозирования риска гестационного диабета.  
Модель показала точность 71%, что подтверждает применимость методов искусственного интеллекта для анализа медицинских данных.

Результаты согласуются с зарубежными исследованиями (Li et al., 2025), где аналогичные модели на основе MLP достигали точности 88% [3]. Разработанная методика может служить основой для внедрения в клинические системы поддержки принятия решений.

Перспективы дальнейших исследований включают расширение базы данных, использование глубоких архитектур (например, LSTM) и развитие объяснимых моделей (Explainable AI).

Список литературы

1. Kaggle. *Pima Indians Diabetes Database.* URL: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database
2. Martis R, Crowther CA, Shepherd E, Alsweiler J, Downie MR, Brown J.Treatments for women with gestational diabetes mellitus: an overview of Cochrane systematic reviews.Cochrane Database of Systematic Reviews 2018, Issue 8. Art. No.: CD012327. DOI: 10.1002/14651858.CD012327.pub2.
3. Li, X., Zhang, Y., Chen, L., et al.  
   Development of an early gestational diabetes mellitus risk prediction model using neural networks.  
   *Gynecological Endocrinology*, 2025, Vol. 41, No. 4, pp. 312–320.  
   DOI: 10.1080/09513590.2025.2470317
4. Chollet F. *Deep Learning with Python.* 2nd ed. Manning Publications, 2021.
5. TensorFlow Documentation. https://www.tensorflow.org/
6. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/index.php