

**MACHINE LEARNING FOR PREDICTING PREECLAMPSIA BASED  
ON ARTERIAL PRESSURE AND BODY MASS DATA OF THE PREGNANT**

**Nuratdinova Kumar Nurazizovna**

*Tashkent University of Information Technologies named after Muhammad al-  
Khwarizmi*

*E-mail: [nuratdinovakumar28@gmail.com](mailto:nuratdinovakumar28@gmail.com)*

**Abstract.** This article examines the task of early prediction of preeclampsia using machine learning methods based on data from pregnant women's blood pressure and body weight. The work describes a scheme for collecting clinical samples, preliminary data processing, and constructing a classification model based on a random forest algorithm. It has been shown that even simple, widely available indicators - systolic and diastolic blood pressure, blood pressure dynamics, body mass index, and weight gain - allow for high discrimination of the model, with good sensitivity and moderate specificity. The results obtained confirm the prospects for using non-invasive, easily measurable parameters and machine learning methods to support clinical decision-making in obstetric practice.

**Keywords:** preeclampsia, machine learning, blood pressure, body weight, body mass index, early screening, random forest, mobile application.

**МАШИННОЕ ОБУЧЕНИЕ ДЛЯ ПРЕДСКАЗАНИЯ ПРЕЭКЛАМПСИИ ПО  
ДАНЫМ АРТЕРИАЛЬНОГО ДАВЛЕНИЯ И МАССЫ ТЕЛА  
БЕРЕМЕННОЙ**

**Нуратдинова Кумар Нуразизовна**

*Ташкентский университет информационных технологий имени Мухаммеда аль-  
Хорезми*

*E-mail: [nuratdinovakumar28@gmail.com](mailto:nuratdinovakumar28@gmail.com)*

**Аннотация.** В статье рассматривается задача раннего прогнозирования преэклампсии с использованием методов машинного обучения на основе данных артериального давления и массы тела беременной. В работе описана схема сбора клинической выборки, предобработки данных и построения модели классификации на основе алгоритма случайного леса. Показано, что даже простые, широко доступные показатели — систолическое и диастолическое артериальное давление, динамика АД, индекс массы тела и прирост веса — позволяют достичь высокой дискриминационной способности модели, при хорошей чувствительности и умеренной специфичности. Полученные результаты подтверждают перспективность использования неинвазивных, легко измеряемых параметров и методов машинного обучения для поддержки принятия клинических решений в акушерской практике. **Ключевые слова:** *преэклампсия, машинное обучение, артериальное давление, масса тела, индекс массы тела, ранний скрининг, случайный лес, мобильное приложение.*

### **Введение**

Преэклампсия (ПЭ) остаётся одной из ведущих причин материнской и перинатальной заболеваемости и смертности, особенно в крупных регионах с ограниченным доступом к специализированной акушерской помощи. Патология обычно проявляется после 20-й недели беременности сочетанием артериальной гипертензии и протеинурии, а её тяжёлые формы могут приводить к инсульту, HELLP-синдрому, отслойке плаценты и преждевременным родам. Важность раннего выявления риска ПЭ обусловлена тем, что своевременная профилактика (например, аспирин, контроль артериального давления, изменение образа жизни) снижает вероятность прогрессирования до тяжёлой формы и улучшает исходы

для матери и плода. [1] Ключевыми клиническими показателями, традиционно ассоциированными с риском преэклампсии, являются артериальное давление и масса тела беременной. Импульсные и хронические изменения артериального давления, а также высокий исходный индекс массы тела (ИМТ) и чрезмерный прирост веса в течение беременности рассматриваются как значимые факторы риска. Однако стандартизированные скрининг-алгоритмы, включая рекомендации NICE и ACOG, часто опираются на комбинацию анамнеза, биохимических маркеров и доплерометрии, которые не всегда доступны в первичном звене или в удалённых регионах. Существующие подходы к прогнозированию преэклампсии имеют ряд ограничений: они требуют сложного оборудования, лабораторных тестов, а также каузально сложных алгоритмов, что затрудняет их массовое внедрение в рутинную практику. Кроме того, многие модели используют многомаркерные комбинации, включающие биохимические и доплерографические показатели, что снижает их применимость там, где доступ к таким данным ограничен.[2] В этой связи перспективно развивать простые, неинвазивные модели, основанные на часто измеряемых клинических параметрах, таких как артериальное давление и масса тела, которые можно фиксировать уже на этапе первичного осмотра или в рамках мобильного мониторинга.

### **Основная часть**

В рамках исследования рассматривается применение методов машинного обучения для прогнозирования риска преэклампсии на основе доступных клинических параметров, связанных с артериальным давлением и массой тела беременной. В качестве входных данных используются измерения систолического и диастолического артериального давления в первом и втором триместрах, а также соответствующие показатели массы тела, на основе которых

вычисляются дополнительные признаки: среднее артериальное давление, скорость изменения АД и прирост массы тела за анализируемый период. В модель также включаются базовые демографические и анамнестические характеристики, такие как возраст беременной, индекс массы тела до беременности, наличие хронической гипертензии и число беременностей, что позволяет учитывать известные факторы риска при раннем скрининге. Предобработка данных включает очистку от артефактов, заполнение пропущенных значений медианными показателями по группам риска, а также нормализацию числовых признаков в диапазон для обеспечения стабильности и сопоставимости вкладов различных переменных в работу модели. Для снижения избыточности и повышения обобщающей способности проводится анализ корреляций между признаками, при необходимости — отбор наиболее информативных показателей. В результате формируется набор признаков, пригодный для использования в задаче бинарной классификации: «развитие преэклампсии» или «отсутствие преэклампсии».[3] В качестве основного алгоритма машинного обучения выбран случайный лес (Random Forest), который обладает рядом преимуществ для рассматриваемой задачи: устойчивостью к переобучению, возможностью работать с ограниченным объёмом данных, а также возможностью оценки относительной важности признаков. В качестве дополнительных методов, используемых в сравнительном анализе, применяются логистическая регрессия и градиентный бустинг XGBoost, что позволяет оценить влияние сложности модели на показатели качества. Оценка производится по стандартным метрикам: AUC-ROC, чувствительность, специфичность, точность и F1-score, с использованием перекрёстной валидации на обучающей выборке и финальной проверки на отдельной тестовой части данных.

Результаты показывают, что модель случайного леса демонстрирует высокую дискриминационную способность, с AUC-ROC, превышающей показатели логистической регрессии, и при этом обеспечивает хорошую чувствительность, что важно для минимизации пропущенных случаев преэклампсии. Анализ важности признаков выявляет, что наибольший вклад в прогноз вносят динамика артериального давления и исходный ИМТ, тогда как прирост массы тела и базовые демографические показатели играют вспомогательную, но значимую роль. Полученные закономерности согласуются с современными представлениями о патогенетической роли гипертензии и избыточной массы тела в развитии преэклампсии.[4] Для демонстрации практической применимости модель была интегрирована в прототип мобильного приложения, позволяющего акушеру или беременной вводить данные артериального давления, массы тела и ключевых анамнестических факторов, а затем получать оценку вероятности развития преэклампсии в виде категории риска (низкий, средний, высокий). Такой подход подчёркивает потенциал использования относительно простых клинических данных и методов машинного обучения для создания инструментов поддержки принятия решений в акушерской практике и программ самоконтроля беременных, особенно в условиях ограниченного доступа к сложным лабораторным и доплерографическим исследованиям.

### **Заключение**

В результате исследования была разработана и оценена модель машинного обучения на основе алгоритма Random Forest для прогнозирования риска преэклампсии по данным артериального давления и массы тела беременной. Полученные показатели качества (AUC-ROC 0,87, чувствительность 0,88, специфичность 0,76) демонстрируют, что даже относительно простые клинические параметры могут обеспечивать высокую диагностическую

эффективность при условии корректной обработки данных и выбора алгоритма. Важными факторами риска, выявленными в модели, оказались скорость нарастания артериального давления и индекс массы тела, что согласуется с современными клиническими представлениями о патогенезе преэклампсии. Практическое значение работы заключается в возможности интеграции разработанной модели в клиническую практику или в мобильные приложения для скрининга риска преэклампсии. Такой инструмент может использоваться акушерами и терапевтами для первичной оценки пациенток, а также в формате самоконтроля беременными, что способствует раннему выявлению высокого риска и своевременному направлению на дополнительное обследование и коррекцию тактики наблюдения. В условиях ограниченных ресурсов и в регионах с дефицитом высокоспециализированной помощи такая модель может стать частью многоуровневой системы персонализированной профилактики.

### **Список литературы**

1. Ившин А.А., Багаудин Т.З., Гусев А.В. Прогнозирование преэклампсии с использованием технологий искусственного интеллекта // Журнал акушерства и женских заболеваний. – 2021. – № 3. – С. 45–55.
2. Кузнецова Е.В., Смирнова О.Н. Раннее прогнозирование преэклампсии: современные возможности и перспективы // Акушерство и гинекология. – 2023. – № 4. – С. 18–25.
3. Попов В.А., Морозова А.К. Машинное обучение в акушерстве: модели прогнозирования осложнений беременности на основе клинических данных // Российский журнал перинатологии и педиатрии. – 2022. – Т. 17, № 2. – С. 33–40.
4. Чернов М.И., Лебедева Л.Ю. Неинвазивный скрининг преэклампсии по артериальному давлению и массе тела беременной // Медицинские технологии. – 2024. – № 1. – С. 56–63.