

DOI: 10.25205/978-5-4437-1843-9-21

**ИННОВАЦИОННЫЕ ПОДХОДЫ К РАННЕЙ ДИАГНОСТИКЕ РАКА ШЕЙКИ МАТКИ:  
ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ  
НА ОСНОВЕ РУТИННЫХ ГЕМАТОЛОГИЧЕСКИХ ПОКАЗАТЕЛЕЙ**

**INNOVATIVE APPROACHES TO EARLY DIAGNOSIS OF CERVICAL CANCER:  
APPLICATION OF MACHINE LEARNING ALGORITHMS  
BASED ON ROUTINE HEMATOLOGICAL INDICATORS**

А. И. Кузнецов

*Московский авиационный институт (национальный исследовательский университет)*

A. I. Kuznetsov

*Moscow Aviation Institute (National Research University)*

✉ drednout5786@yandex.ru

**Аннотация**

Статья посвящена разработке системы поддержки принятия врачебных решений для раннего выявления рака шейки матки (РШМ) на основе рутинных гематологических показателей. РШМ является серьезной проблемой здравоохранения, с высокой смертностью и сложностями в ранней диагностике. В исследовании, проведенном на 452 пациентках, была построена модель с использованием алгоритма XGB, показавшая ROC-AUC 98,2 % и высокую чувствительность и специфичность. Результаты подчеркивают значимость рутинных анализов в диагностике РШМ и потенциал интеграции модели в клиническую практику для улучшения исходов лечения.

**Abstract**

The article focuses on the development of a decision support system for the early detection of cervical cancer (CC) based on routine hematological indicators. CC poses a significant healthcare challenge, with high mortality rates and difficulties in early diagnosis. In a study involving 452 patients, a model was built using the XGB algorithm, achieving a ROC-AUC of 98.2 % along with high sensitivity and specificity. The results highlight the importance of routine tests in diagnosing CC and the potential for integrating the model into clinical practice to improve treatment outcomes.

**Обоснование**

Рак шейки матки (РШМ) занимает четвертое место по распространенности и третье по смертности. В 2022 г. в мире зарегистрирован 662 301 новый случай заболевания и 348 874 смерти от РШМ. В России за тот же год было 18 369 заболевших и 7903 умерших [1]. РШМ часто проявляется симптомами, которые могут быть связаны и с другими заболеваниями, что усложняет его раннее выявление. Традиционные диагностические методы, такие как биопсия, проводятся уже при видимом поражении, являются инвазивными и могут вызывать дискомфорт, кровотечение [2, 3]. Современные алгоритмы машинного обучения демонстрируют значительные успехи в предсказании исходов в различных областях онкологии [4–7]. Анализ показателей периферической крови давно считается важным инструментом для оценки состояния здоровья человека. Тем не менее мы не обнаружили исследований, посвященных созданию прогностических моделей для раннего выявления РШМ с применением алгоритмов машинного обучения, основанных на рутинных гематологических показателях, которые входят в стандартные и обязательные обследования при злокачественных заболеваниях.

**Цель** — разработка системы поддержки принятия врачебных решений (СППВР) для выявления пациенток с высоким риском наличия РШМ на основе рутинных показателей крови.

**Методы**

Одноцентровое когортное ретроспективное исследование проводилось на базе МНИОИ им. П. А. Герцена в 2000–2024 гг. В исследование включены 452 пациентки в возрасте 42,0 [33,75; 50,0] лет с подозрением на РШМ. Верификация исхода (выявление РШМ) была выполнена на основании гистологического исследования. Выборка была разделена на обучающую и тестовую в соотношении 70 : 30. Построение модели производилось с использованием Python 3.11.

### Результаты

Модель, построенная с помощью алгоритма Extra Tree, показала ROC-AUC = 97,9 %, Random Forest = 97,8 %, Catboost = 97,3 %, Lasso = 97,2 %, Ridge = 97,0 %, SVM = 96,9 %, LightGBM = 94,3 %, k-Nearest Neighbors = 93,9 %, Naive Bayes = 90,2 %. Наилучшие показатели были достигнуты при использовании XGB. На тестовой выборке модель продемонстрировала ROC-AUC = 98,2 % [96,0; 99,7], чувствительность = 95,0 % [91,2; 98,3], специфичность = 91,7 % [81,0; 100,0]. Список наиболее значимых факторов риска включает агрегацию тромбоцитов с АДФ, СОЭ, Д-димер, трансферин, фибриноген, СРБ, всего 15 факторов из рутинных гематологических показателей, входящих в ОМС.

### Заключение

Разработанная модель на основе алгоритма XGB может быть интегрирована в СППВР для ранней диагностики РШМ. Это позволит улучшить результаты лечения и снизить затраты на здравоохранение, оптимизируя профилактические стратегии для групп высокого риска.

### Литература

1. International Agency for Research on Cancer. Cancer Today. URL: <https://gco.iarc.fr/today/en> (date of access: 29.04.2025).
2. Keskin N., Biyik I., Ince O. et al. Evaluation of the consistency ratios of cervical smear, cervical biopsy and conization results // *Ginekologia Polska*. 2021. Vol. 92, No. 11. P. 778–783.
3. Сушинская Т. В., Волченко Н. Н., Доброхотова Ю. Э. и др. Эффективность цитологической диагностики цервикальной интраэпителиальной неоплазии и рака шейки матки в зависимости от способа взятия материала // *Онкогинекология*. 2017. Т. 3, № 23. С. 51–59.
4. Akazawa M., Hashimoto K. Artificial Intelligence in Ovarian Cancer Diagnosis // *Anticancer Research*. 2022. Vol. 40, No. 8. P. 4795–4800.
5. Li H., Lin J., Xiao Y. et al. Colorectal Cancer Detected by Machine Learning Models Using Conventional Laboratory Test Data // *Technology in Cancer Research & Treatment*. 2021. Vol. 20.
6. Bentick K., Runevic J., Akula S. et al. Machine learning models based on routinely sampled blood tests can predict the presence of malignancy amongst patients with suspected musculoskeletal malignancy // *Methods*. 2023. Vol. 220. P. 55–60.
7. Каприн А. Д., Сушинская Т. В., Кузнецов А. И. и др. Способ прогнозирования вероятности выявления рака шейки матки на основе рутинных показателей крови: патент на изобретение № 2802395; заявл. 16.05.2023; опубл. 28.08.2023. URL: <https://www.elibrary.ru/item.asp?id=54658031> (date of access: 29.04.2025).