

¹Волков К.А., ^{2,3}Полиданов М.А., ¹Капралов С.В., ⁴Барулина М.А.
СОЗДАНИЕ ПРОГРАММЫ ДЛЯ ПРЕДСКАЗАНИЯ ВЕРОЯТНОСТИ РЕЦИДИВА РАКА ЩИТОВИДНОЙ ЖЕЛЕЗЫ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

¹ФГБОУ ВО Саратовский ГМУ им. В. И. Разумовского Минздрава России,
г. Саратов, Российская Федерация

²ЧУОО ВО «Университет «Реавиз», г. Санкт-Петербург, Российская Федерация

³ЧУОО ВО «Медицинский университет «Реавиз», г. Самара, Российская Федерация

⁴ФГАОУ ВО «Пермский государственный национальный исследовательский
университет»; г. Пермь, Российская Федерация

Аннотация. Несмотря на высокие показатели выживаемости при раннем обнаружении и адекватном лечении, проблема рецидива рака щитовидной железы остается актуальной и требует особого внимания. Рецидив заболевания может произойти даже после успешного лечения, что делает необходимым регулярное наблюдение за пациентами. Однако предсказание рецидива на основе клинических показателей представляет собой сложную задачу для медицины в целом. В связи с вышеописанными трудностями возникает необходимость в разработке автоматизированной модели предсказания рецидива рака щитовидной железы. Использование методов машинного обучения позволяет анализировать большие объемы данных и выявлять скрытые закономерности между различными клиническими показателями и вероятностью рецидива.

Ключевые слова: рак щитовидной железы, щитовидная железа, машинное обучение, прогнозирование рецидивов, случайный лес.

Volkov K.A, Polidanov M.A., Kapralov S.V., Barulina M.A.
CREATING A PROGRAM TO PREDICT THE PROBABILITY OF THYROID CANCER RECURRENCE USING MACHINE LEARNING METHODS

Abstract. Despite high survival rates with early detection and adequate treatment, the problem of recurrence of thyroid cancer remains relevant and requires special attention. Recurrence can occur even after successful treatment, making regular follow-up of patients essential. However, predicting relapse based on clinical parameters is challenging for medicine in general. Due to the above difficulties, there is a need to develop an automated model for predicting thyroid cancer recurrence. The use of machine learning methods allows analyzing large amounts of data and identifying hidden patterns between various clinical parameters and the probability of recurrence.

Keywords: thyroid cancer, thyroid, machine learning, recurrence prediction, random forest, python

Актуальность. Рак щитовидной железы (РЩЖ) является одним из наиболее распространенных видов рака среди эндокринных заболеваний [1-3]. Несмотря на высокие показатели выживаемости при раннем обнаружении и адекватном лечении, проблема рецидива остается актуальной и требует особого внимания.

Цель. Разработка модели машинного обучения по предсказанию факта рецидива у пациентов с раком щитовидной железы после проведенного оперативного вмешательства.

Материалы и методы исследования. В соответствии с целью исследования были проанализированы истории болезни 300 пациентов, с

выполненным оперативным вмешательством по поводу рака щитовидной железы. Средний возраст – 43,54 года. Всем включенным в исследование больным было проведено комплексное обследование согласно клиническим рекомендациям по диагностике и лечению больных РЦЖ.

Выбор наиболее подходящей модели в машинном обучении критически важен, так как он напрямую влияет на точность и эффективность предсказания. Каждая модель оценивалась по метрикам, таким как средняя точность и стандартное отклонение, что позволяет определить, какая из них демонстрирует наилучшие результаты. Лучшей по показателю средней точности оказалась модель случайного леса, она же в дальнейшем и использовалась. Для подбора гиперпараметров использовался метод RandomizedSearchCV. В процессе поиска гиперпараметров модель обучалась на тренировочных данных, отобранных как 70% от исходного датасета.

Результаты исследования. Процесс разработки начался с тщательной предобработки данных, что является критически важным этапом в построении надежных моделей. В ходе предобработки были удалены выбросы и столбцы, содержащие однообразные значения, что позволило улучшить качество данных и избежать искажений в обучении модели.

Также была проведена кодировка категориальных переменных, что обеспечило возможность их корректного использования в алгоритмах машинного обучения, и исключены коррелирующие признаки, чтобы минимизировать мультиколлинеарность и повысить интерпретируемость модели. Для выбора наиболее подходящей модели был проведен сравнительный анализ нескольких алгоритмов классификации. В результате был выбран метод случайного леса, который продемонстрировал высокую эффективность в решении задачи классификации.

С использованием метода случайного поиска гиперпараметров была осуществлена оптимизация модели, что позволило определить наилучшие параметры для повышения ее производительности. Полученная модель достигла точности предсказаний 96%, что свидетельствует о ее высокой надежности и способности к правильной классификации.

Результаты исследования подчеркивают потенциал применения машинного обучения в области медицины, особенно в контексте ранней диагностики и мониторинга заболеваний. Высокая точность модели может значительно улучшить процесс принятия клинических решений и повысить качество медицинской помощи.

Заключение. Результаты эксперимента свидетельствуют о том, что разработанная модель [4-7] эффективно справляется с задачей классификации на основе медицинских параметров, что может быть особенно важно для принятия решений в клинической практике. Высокая точность указывает на надежность модели и её способность правильно идентифицировать случаи рецидива, что может способствовать улучшению диагностики и лечения.

Литература

1. Берштейн, Л. М. Рак щитовидной железы: эпидемиология, эндокринология, факторы и механизмы канцерогене / Л. М. Берштейн // Практ. онкология. – 2007. – Т.8, №1. – С. 1-8.
2. Лушников, Е. Ф. Рак щитовидной железы в России после Чернобыля / Е. Ф. Лушников, А. Ф. Цыб, С. Ямасита // Проблемы эндокринологии. – М.: Медицина, 2006. – №4. – 128 с.
3. Bentz, B. G. B-RAF V600E mutational analysis of fine needle aspirates correlates with diagnosis of thyroid nodules / B. G. Bentz [et al.] // Otolaryngol. Head Neck Surg. – 2009. – Vol. 140, № 5. – P. 709-714.
4. Свидетельство о регистрации программы для ЭВМ № 2024689824. Система для прогнозирования возникновения рецидивов после оперативного вмешательства при раке щитовидной железы : заявлено 28.11.2024 : опубл. 11.12.2024. / Полиданов М. А., Петрунькин Р. П., Кудашкин В. Н., Волков К. А., Кравченя А. Р., Рафеева П. Д., Трухина М. К., Капралов С. В., Амиров Э. В., Масляков В. В. с
5. Свидетельство о регистрации программы для ЭВМ № 2025663285. Программа для предсказания вероятности рецидивов в послеоперационном периоде при раке щитовидной железы : № 2025661995 : заявлено 15.05.2025 : опубл. 27.05.2025 / Волков К.А., Долгова Е.М., Полиданов М.А., Петрунькин Р.П., Капралов С.В., Кравченя А.Р., Рафеева П.Д.
6. Свидетельство о регистрации базы данных RU № 2025622798 База данных клинико-лабораторных и инструментальных исследований для предсказания вероятности рецидивов в послеоперационном периоде при раке щитовидной железы : № 2025622318 : заявлено 04.06.2025 : опубл. 27.06.2025 / Петрунькин Р.П., Полиданов М.А., Волков К.А., Долгова Е.М., Капралов С.В., Масляков В.В., Долесов К.П., Емельянова И.Р.
7. Свидетельство о регистрации базы данных RU 2025621872. База данных клинико-лабораторных исследований для возможности предсказания вероятности рецидива рака щитовидной железы методами машинного обучения : № 2025621436 : заявлено 13.04.2025 : опубл. 24.04.2025 / Полиданов М.А., Петрунькин Р.П., Капралов С.В., Масляков В.В., Кудашкин В.Н., Волков К.А., Кашихин А.А., Цуканова П.Б., Амиров Э.В., Аскеров М.Р.О.