УДК 004.89:616.441-006.6 Научная статья

https://doi.org/10.32603/2071-8985-2026-19-9-34-43

Прогнозирование рецидива рака щитовидной железы с использованием методов машинного обучения

И. В. Тишко[1]✉, **А. А. Мелдо**[1,2,3], **С. В. Дора**[2], **Ю. Ш. Халимов**[2], **М. Е. Борискова**[2], **Л. А. Ахметшина**[3], **М. Д. Смирнов**[2]

[1] Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина), Санкт-Петербург, Россия

[2] Первый Санкт-Петербургский государственный медицинский университет им. академика И. П. Павлова, Санкт-Петербург, Россия

[3] Городской консультативно-диагностический центр № 1, Санкт-Петербург, Россия

✉ ivantishko29@gmail.com

Аннотация. Прогнозирование клинических событий для создания персонализированных лечебных траекторий в онкологии – одна из актуальных задач, решение которых может быть реализовано с помощью методов искусственного интеллекта. Стратификация риска рецидива рака щитовидной железы (РЩЖ) традиционно основывается на характеристиках самой опухоли, тогда как вклад внеопухолевых (метаболических, антропометрических) факторов остается малоизученным.

Цель работы. Разработка и валидация модели машинного обучения для прогнозирования рецидива РЩЖ с учетом как опухолевых, так и внеопухолевых характеристик пациента.

Материалы и методы. Работа выполнена в рамках междисциплинарного взаимодействия с ФГБОУ ВО ПСПбГМУ им. И. П. Павлова и основана на ретроспективных, валидированных, мультимодальных данных 263 пациентов. Методология включает предобработку данных (очистку, нормализацию, отбор признаков), обучение и сравнение моделей на основе логистической регрессии и ансамблевых методов машинного обучения (Gradient Boosting, XGBoost), а также интерпретацию вкладов признаков методом SHAP.

Результаты. Наилучшие результаты показала модель XGBClassifier (ROC AUC 0.9978; F1-score 0.9524). Установлено, что вклад внеопухолевых признаков (метаболических и антропометрических) в прогноз рецидива сопоставим с вкладом признаков, описывающих опухоль.

Заключение. Полученные результаты подтверждают целесообразность включения внеопухолевых характеристик пациента в модели прогнозирования рецидива РЩЖ и создают основу для их дальнейшей внешней валидации.

Ключевые слова: прогнозирование клинических событий, рак щитовидной железы, рецидив, машинное обучение, Gradient Boosting, XGBoost, SHAP-анализ, метаболические показатели

Для цитирования: Прогнозирование рецидива рака щитовидной железы с использованием методов машинного обучения / И. В. Тишко, А. А. Мелдо, С. В. Дора, Ю. Ш. Халимов, М. Е. Борискова, Л. А. Ахметшина, М. Д. Смирнов // Изв. СПбГЭТУ «ЛЭТИ». 2026. Т. 19, № 9. С. 34–43. doi: 10.32603/2071-8985-2026-19-9-34-43.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

© И. В. Тишко, А. А. Мелдо, С. В. Дора, Ю. Ш. Халимов, М. Е. Борискова, Л. А. Ахметшина, М. Д. Смирнов, 2026

Original article

Predicting Thyroid Cancer Recurrence Using Machine Learning Methods

I. V. Tishko[1]✉, **A. A. Meldo**[1,2,3], **S. V. Dora**[2], **Yu. Sh. Khalimov**[2], **M. E. Boriskova**[2], **L. A. Akhmetshina**[3], **M. D. Smirnov**[2]

[1] Saint Petersburg Electrotechnical University, St Petersburg, Russia

[2] Pavlov First Saint Petersburg State Medical University, St Petersburg, Russia

[3] City Consultative Diagnostic Center №1, St Petersburg, Russia

✉ ivantishko29@gmail.com

Abstract. The prediction of clinical events for designing personalized treatment trajectories in oncology is a relevant task that can be addressed using artificial intelligence methods. Risk stratification for thyroid cancer (TC) recurrence has traditionally relied on tumor-related characteristics, while the contribution of non-tumor (metabolic, anthropometric) factors remains understudied.

Purpose of the work. To develop and validate a machine learning model for predicting TC recurrence that accounts for both tumor and non-tumor patient characteristics.

Materials and methods. This paper presents the development and validation of a machine learning model for predicting thyroid cancer recurrence. The study is conducted as part of an interdisciplinary collaboration with Pavlov First Saint Petersburg State Medical University and is based on retrospective, validated, multimodal data of 263 patients. The methodology includes data preprocessing (cleaning, normalization, feature selection), training and comparison of models based on logistic regression and ensemble machine learning methods (Gradient Boosting, XGBoost), together with SHAP-based interpretation of feature contributions.

Results. The XGBClassifier model demonstrated the best results (ROC AUC 0.9978; F1-score 0.9524). It is shown that the contribution of non-tumor features (metabolic and anthropometric) to the recurrence prediction is comparable to that of tumor-related features. Conclusion. The findings support incorporating non-tumor patient characteristics into TC recurrence prediction models and provide a basis for further external validation of the approach.

Keywords: clinical event prediction, thyroid cancer, recurrence, machine learning, Gradient Boosting, XGBoost, SHAP analysis, metabolic parameters

For citation: Predicting Thyroid Cancer Recurrence Using Machine Learning Methods / I. V. Tishko, A. A. Meldo, S. V. Dora, Yu. Sh. Khalimov, M. E. Boriskova, L. A. Akhmetshina, M. D. Smirnov // LETI Transactions on Electrical Engineering & Computer Science. 2026. Vol. 19, no. 9. P. 34–43. doi: 10.32603/2071-8985-2026-19-9-34-43.

Conflict of interests. The authors declare no conflict of interests.

Введение. Рак щитовидной железы (РЩЖ) составляет 0.4–3 % в структуре заболеваемости злокачественных опухолей различных локализаций и является самой распространенной злокачественной опухолью эндокринных желез. Среди лиц с изменениями щитовидной железы рак выявляется в 20 % случаев [1]. В России ежегодно регистрируется более 2000 случаев РЩЖ [2]. Своевременное хирургическое лечение обеспечивает достаточно высокую 5-летнюю выживаемость большинства больных, однако в 20 % случаев развивается местный рецидив опухоли [3].

Клиническая значимость задачи выходит за рамки исключительно статистической точности прогноза. Своевременное выявление пациентов группы высокого риска рецидива позволяет дифференцировать частоту и объем динамического наблюдения (ультразвуковое исследование, определение уровня тиреоглобулина), избегая как избыточного обследования пациентов с низким риском, так и запоздалого выявления прогрессирования у пациентов с высоким риском. В этом контексте включение в прогностическую модель не только опухолевых, но и общих метаболических и антропометрических характеристик пациента представляется перспективным направлением, поскольку соответствующие показатели рутинно доступны в клинической практике и могут быть получены без дополнительных инвазивных процедур.

Традиционные медицинские исследования по задачам стратификации риска основаны на статистических методах, которые составляют основу так называемой доказательной медицины. Однако появляется все больше работ, демонстрирующих модели прогноза различных клинических событий на основе методов машинного обучения (МО). Так, в [4]–[6] модели МО (Random Forest, XGBoost, CatBoost и др.) использованы для построения прогнозов рецидива РЩЖ на основе клинико-патологических данных пациентов и продемонстрировали улучшенную дискриминацию по сравнению с традиционными методами.

Вместе с тем следует отметить, что риск рецидива в большинстве случаев оценивается, исходя из параметров самой опухоли – ее размеров, распространенности, метастатического поражения лимфатических узлов, молекулярно-генетического и патоморфологического статуса [7], [8].

В целом, количество таких исследований ограничено, и в большинстве из них акцент делается на признаки, связанные непосредственно с опухолью. Например, авторы многоцентрового исследования [9], включавшего сведения о 2388 пациентах, использовали случайный лес для прогнозирования рецидива РЩЖ, однако ограничились опухолевыми и стандартными клиническими параметрами. В других исследованиях сравнивались несколько алгоритмов МО (включая XGBoost и Logistic Regression), но также в рамках набора показателей, относящихся непосредственно к опухоли [10], [11].

Более точная стратификация риска рецидива РЩЖ позволит создавать персонализированные лечебно-диагностические траектории для пациентов, а также планировать мероприятия по мониторингу состояния оперированных пациентов.

В России отсутствуют крупные публикации, посвященные прогнозированию рецидива РЩЖ на основе современных методов машинного обучения на комплексных наборах реальных клинических данных (с учетом внеопухолевых биохимических и антропометрических параметров). Анализ литературы по сравнительному анализу моделей машинного обучения для прогнозирования выживаемости в смежных областях онкологии подтверждает актуальность применения ансамблевых методов [12].

В настоящей статье впервые определена значимость внеопухолевых факторов (метаболических показателей, антропометрии и др.) в развитии рецидива РЩЖ. Согласно оценкам методов объяснимости (SHAP), их вклад сопоставим с влиянием опухолевых признаков [5], [13]. Это углубляет понимание биологии рецидива и расширяет потенциал персонализированного подхода к лечебным тактикам и траекториям мониторинга состояния пациентов с РЩЖ. Предлагаемый подход выходит за рамки типичных задач, основанных только на параметрах онкопроцесса.

Цель исследования. Разработка и валидация модели машинного обучения для высокоточного прогнозирования рецидива рака щитовидной железы для создания персонализированных лечебных траекторий.

Задачи исследования:

1. Провести предобработку и очистку ретроспективных мультимодальных данных пациентов, включая устранение утечек, нормализацию, обработку пропусков и аномалий.

2. Разработать, обучить и сравнить несколько моделей машинного обучения (Logistic Regression, Gradient Boosting, XGBoost) для прогнозирования рецидива РЩЖ.

3. Оценить прогностическую точность разработанных моделей с использованием метрик ROC AUC, F1-score, Sensitivity, Specificity, Precision и Accuracy.

4. Провести анализ значимости признаков с помощью метода SHAP (SHapley Additive exPlanations) для интерпретации вклада различных факторов, в том числе внеопухолевых (метаболических и антропометрических), в прогноз модели.

5. Определить наиболее эффективную модель и подтвердить гипотезу о значимом вкладе общего метаболического статуса пациента в вероятность рецидива РЩЖ.

Методология исследования и подготовка данных. Исходный набор данных включал 134 признака от 263 пациентов и содержал демографические, клинические и лабораторные показатели. В набор данных вошли как характеристики опухоли (размеры, результаты цитологического и гистологического исследования), так и параметры общего состояния организма (результаты анализов крови, эндокринологические тесты и т. п.). Целевой переменной было наличие рецидива. Исходный набор характеризовался высокой гетерогенностью и качественными ограничениями, типичными для медицинских ретроспективных исследований [14]: значительным количеством пропусков, неоднородными форматами представления данных, смешением числовых и текстовых значений, а также наличием выбросов и потенци-

альных источников утечки целевой переменной. В связи с этим процесс подготовки данных был разделен на несколько последовательных и взаимосвязанных технологических этапов, направленных на глубокую очистку, стандартизацию и структурирование данных перед этапом моделирования.

Предобработка и очистка данных. На этом этапе проведены основные процедуры по повышению качества данных и их подготовке к последующему использованию в задачах машинного обучения.

Устранение утечек данных и нерелевантных признаков. Для предотвращения искусственного завышения точности модели и соблюдения принципов конфиденциальности удалены служебные поля, не несущие прогностической ценности или потенциально приводящие к утечке целевой переменной. К таким признакам относились идентификаторы пациентов, административные данные, а также исходные колонки, описывающие рецидивы в различные временные интервалы.

Нормализация текстовых данных. Строковые значения в ряде признаков содержали нерелевантные символы, опечатки и артефакты ввода, что могло препятствовать корректной обработке. Выполнена очистка таких полей удалением лишних пробелов и приведением к единому формату. Также выполнены преобразование текстовых значений признаков в единый числовой формат и обработка некорректных, неоднозначных и нерелевантных записей: подобные значения приводились к пропущенным, что обеспечило корректную последующую обработку признаков и повысило качество исходного набора данных.

Обработка пропусков и аномалий. После первичной очистки выполнены шаги по работе с пропущенными значениями и потенциальными аномалиями.

One-Hot Encoding категориальных признаков. Все категориальные признаки, представленные объектным типом (object type), преобразованы в числовой формат с использованием One-Hot Encoding. При этом для каждой категории, включая пропущенные значения (NaN), создавалась отдельная бинарная колонка (dummy_na = True), что позволило модели учитывать сам факт отсутствия данных как потенциально информативный признак. Признаки с большим числом уникальных категорий (более 50) удалены во избежание значительного увеличения размерности данных.

Для борьбы с выбросами и снижения размерности применен *метод на основе межквартильного размаха* (IQR – Interquartile Range). Рассчитанные квартильные границы использовались для выявления аномальных значений. Вместо удаления отдельных строк, которые в медицинских данных могут содержать важную клиническую информацию, выбран более осторожный подход: столбцы удалялись только в случае, если доля выбросов превышала 30 %. Такой порог позволил сохранить максимальное количество записей (263) и минимизировать потерю данных, одновременно повышая надежность признаков, содержащих значительную часть аномальных значений.

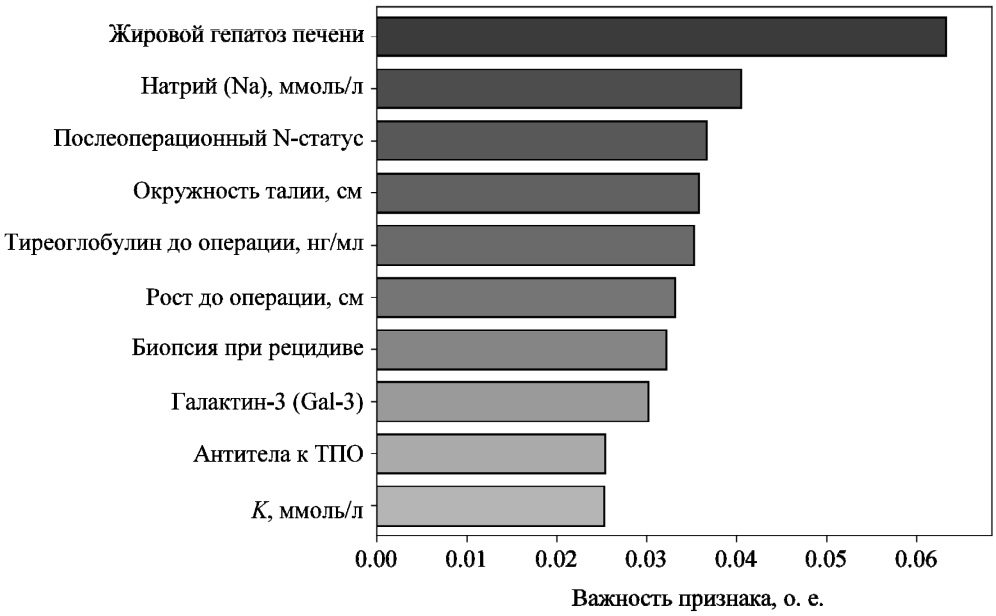


Рис. 1. Топ 10 важных признаков модели XGBoost
Fig. 1. Top 10 important features of the XGBoost model

Импутация числовых пропусков и стандартизация. После удаления столбцов с высокой долей выбросов оставшиеся числовые пропуски заполнены средними значениями соответствующих признаков с использованием SimpleImputer (strategy = 'mean'). Этот метод обеспечивает сохранение общей статистической структуры данных. Завершающим этапом предобработки числовых признаков стала стандартизация с помощью StandardScaler, что привело все признаки к нулевому среднему и единичному стандартному отклонению. Это необходимо для корректной работы многих алгоритмов машинного обучения, особенно тех, которые чувствительны к масштабу признаков (например, логистической регрессии и методов, основанных на расстоянии).

Интерпретируемость и SHAP-анализ. Для обеспечения доверия клиницистов использован метод SHAP. Анализ вкладов признаков в индивидуальные прогнозы выявил иерархию факторов:

1. Метаболические показатели: окружность талии, уровень натрия (Na) и калия (K), а также галектин-3 (Gal-3).

2. Клинические данные: послеоперационный N-статус, наличие жирового гепатоза, результаты биопсии при рецидиве и уровень тиреоглобулина до операции.

3. Антропометрия: рост пациента (косвенно, через связь с метаболическими и клиническими маркерами).

Вклад каждого признака в формирование индивидуального прогноза вероятности рецидива представлен на рис. 1.

Результаты исследования. Для оценки прогностической способности моделей использованы стандартные метрики классификации. Сравнительный анализ характеристик моделей машинного обучения на основе примененных алгоритмов представлен в таблице.

Для глубокого анализа прогностической способности моделей использована матрица ошибок (Confusion Matrix), позволяющая детально оценить распределение верных и ошибочных предсказаний в разрезе каждого класса. Визуализация матрицы ошибок для модели XGBoost, показавшей наилучшие результаты по совокупности метрик (F1-score и ROC AUC), приведена на рис. 2.

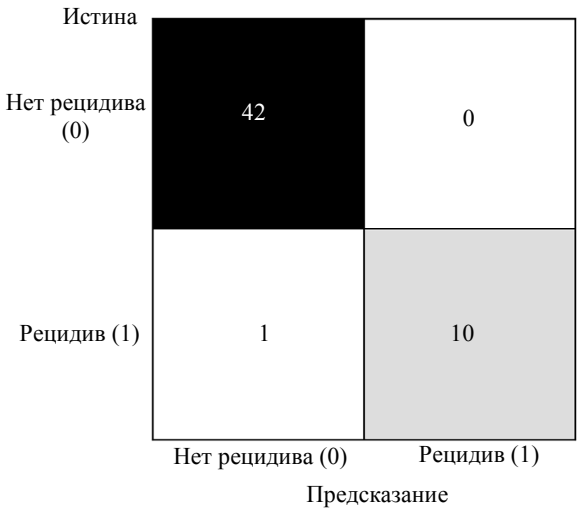


Рис. 2. Матрица ошибок для модели XGBoost
Fig. 2. Confusion matrix for the XGBoost model

Выводы и заключение. В ходе проведенного исследования разработана и валидирована система предиктивного анализа рецидива рака щитовидной железы на основе современных алгоритмов машинного обучения. Сравнительный анализ показал, что ансамблевые методы (XGBoost, Gradient Boosting) обладают значительным преимуществом перед традиционными статистическими подходами, обеспечивая высокую дискриминационную способность (ROC AUC до 0.9978).

Основные выводы и достоинства предложенного подхода заключаются в следующем.

1. Высокая прогностическая точность. Использование XGBoost позволило минимизировать количество ложноотрицательных прогнозов, что критично в онкологии. Модель успешно идентифицирует пациентов группы высокого риска, обеспечивая при этом нулевой уровень ложноположительных срабатываний на тестовой выборке.

Сравнение характеристик моделей машинного обучения
Comparison of machine learning model characteristics

Модель	Прогностическая точность метрики					
	Sensitivity	Specificity	Precision	Accuracy	F1-score	ROC AUC
Gradien Boosting lassifier	0.9091	0.9762	0.9091	0.9623	0.9091	0.9957
XGBClassifier	0.9091	1.0000	1.0000	0.9811	0.9524	0.9978
LogisticRegression	0.7273	1.0000	1.0000	0.9434	0.8421	0.9567

2. Роль внеопухолевых факторов. Важным научным результатом работы стало подтверждение гипотезы о значимом вкладе общего метаболического статуса пациента (антропометрические данные, уровень электролитов, показатели липидного обмена) в вероятность рецидива. SHAP-анализ показал, что эти признаки сопоставимы по важности с классическими патоморфологическими характеристиками опухоли.

3. Интерпретируемость. Применение методов объяснимого ИИ (XAI) позволяет не только получать точный прогноз, но и визуализировать логику принятия решения моделью, что способствует росту доверия со стороны практикующих врачей.

Несмотря на высокие метрики, работа имела ряд ограничений. Во-первых, объем выборки (n = = 263 пациента) относительно невелик, что может ограничивать обобщающую способность модели. Во-вторых, исследование носило ретроспективный характер и базировалось на данных одного медицинского центра, что также может ограничивать переносимость результатов на другие популяции пациентов.

Дальнейшее развитие темы планируется по следующим направлениям: масштабирование данных за счет многоцентрового сбора для повышения робастности алгоритмов; интеграция мультимодальных данных, включая результаты молекулярно-генетических исследований (статус мутаций BRAF, TERT и др.) и данные лучевой диагностики (DICOM-снимки), для создания комплексного цифрового фенотипа пациента; внешняя валидация – проспективное тестирование модели на независимых наборах данных («слепое» тестирование) для подтверждения ее клинической эффективности; программная реализация – разработка системы поддержки принятия врачебных решений (СППВР) в виде веб-интерфейса или модуля медицинской информационной системы (МИС), предоставляющего количественную оценку индивидуального риска рецидива в режиме реального времени.

Реализация перечисленных этапов позволит перейти от исследовательского прототипа к практическому инструменту персонализированной медицины, способному оптимизировать тактику наблюдения и лечения пациентов с РЩЖ.

Отдельного внимания в дальнейшем требуют вопросы, связанные с переходом от ретроспективной модели к ее применению в реальной клинической практике. Прежде чем модуль поддержки принятия решений сможет использоваться врачом, необходима проспективная валидация на независимой выборке пациентов, оценка устойчивости модели к изменению состава и объема поступающих данных, а также соблюдение требований регуляторных органов, предъявляемых к программному обеспечению для медицинских целей. Указанные шаги определяют дальнейшую траекторию развития предложенного подхода.

Список литературы

1. Яйцев С. В., Привалов В. А. Распространенность рака щитовидной железы // Проблемы эндокринологии. 2002. Т. 48, № 4. С. 13–16. doi: 10.14341/probl11558.

2. Мерабишвили В. М. Злокачественные новообразования щитовидной железы (С73) в Санкт-Петербурге. Заболеваемость, смертность, достоверность учета, погодичная летальность и выживаемость больных // Вопр. онкологии. 2020. Т. 66, № 5. С. 479–488.

3. Стратификация риска рецидива папиллярного рака щитовидной железы на основании результатов молекулярно-генетических исследований / С. А. Лукьянов, С. В. Сергийко, С. Е. Титов, И. В. Решетов, Ю. А. Веряскина, А. В. Важенин, А. В. Гостимский, Л. И. Ипполитов, М. О. Рогова // Опухоли головы и шеи. 2020. Т. 10, № 1. С. 93–100. doi: 10.17650/2222-1468-2020-10-1-93-100.

4. Development and validation of prediction models for papillary thyroid cancer structural recurrence using machine learning approaches / H. Wang, C. Zhang, Q. Li, T. Tian, R. Huang, J. Qiu, R. Tian // BMC Cancer. 2024. Vol. 24, no. 1. P. 427. doi: 10.1186/s12885-024-12146-4.

5. Predicting thyroid cancer recurrence using supervised CatBoost: A SHAP-based explainable AI approach / A. A. Hanani, T. B. Donmez, M. Kutlu, M. Mansour // Medicine (Baltimore). 2025. Vol. 104, no. 22. P. e42667. doi: 10.1097/MD.0000000000042667.

6. Interpretable machine learning for thyroid cancer recurrence prediction: leveraging XGBoost and SHAP analysis / A. Schindele, A. Krebold, U. Heiß, K. Nimptsch, E. Pfaehler, C. Berr, R. A Bundschuh, T. Wendler, O. Kertels, J. Tran-Gia // Europ. J. of Radiol. 2025. Vol. 186. P. 112049. doi: 10.1016/j.ejrad.2025.112049.

7. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer / B. R. Haugen, E. K. Alexander, K. C. Bible, G. M. Doherty, S. J. Mandel, Yu. E. Nikiforov, F. Pacini, G. W. Randolph, A. M. Sawka, M. Schlumberger, K. G. Schuff, S. I. Sherman, Ju. A. Sosa, D. L. Steward, R. M. Tuttle, L. Wartofsky // Thyroid. 2016. Vol. 26, no. 1. P. 1–133. doi: 10.1089/thy.2015.0020.

8. Overview of the 2022 WHO classification of thyroid neoplasms / Z. W. Baloch, S. L. Asa, J. A. Barletta,

R. A. Ghossein, C. C. Juhlin, C. K. Jung, V. A. LiVolsi, M. G. Papotti, M. Sobrinho-Simões, G. Tallini, O. Mete // Endocrine Pathol. 2022. Vol. 33, № 1. P. 27–63. doi: 10.1007/s12022-022-09707-3.

9. A machine learning-based model for predicting recurrence in intermediate- and high-risk differentiated thyroid cancer: insights from a retrospective single-center study of 2388 patients / Y. Li, Z. Tang, A. Ren, G. Tian, J. Zhang, Y. Wang, J. Liu, J. Ming // Frontiers in Endocrinol. (Lausanne). 2025. Vol. 16. P. 1552479. doi: 10.3389/fendo.2025.1552479.

10. Yapici I. S., Arslan R. U. Predictive analytics for thyroid cancer recurrence: a feature selection and data balancing approach // The Europ. Phys. J. Spec. Topics. 2025. Vol. 234. P. 4751–4771. doi: 10.1140/epjs/s11734-025-01720-x.

11. Machine learning prediction of thyroid cancer recurrence for early screening and clinical decision pathways: a retrospective cohort study / G. Hu, B. Huang, L. Cai, Y. Zhang, Y. Zhang, Y. Liu, G. Wu // Discover Oncol. 2026. Vol. 17, no. 1. P. 239. doi: 10.1007/s12672-025-04293-2.

12. Сравнительный анализ моделей машинного обучения для прогнозирования выживаемости при раке молочной железы / Е. А. Чернов, А. А. Мелдо, А. В. Агаева, М. А. Богданов, М. Ю. Вальков // Сб. материалов XIII науч.-практ. конф. «Наука настоящего и будущего». СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2025. С. 80–82.

13. Machine learning prediction of recurrence in pediatric thyroid cancer: malignant endocrine tumors cohort analysis using XGBoost and SHAP / A. Redlich, E. Pfaehler, M. Kunstreich, M. Schmutz, C. Lapa, M. Kuhlen // J. of Clin. Endocrinol. & Metabolism. 2026. Vol. 111, no. 3. P. e844–e852. doi: 10.1210/clinem/dgaf487.

14. Бобровская Т. М. Методология формирования наборов данных и их использование для оценки диагностической точности систем искусственного интеллекта в лучевой диагностике: автореф. дис. ... канд. мед. наук: 3.3.9. М., 2025. 138 с.

Информация об авторах

Тишко Иван Владимирович – студент 4-го курса факультета компьютерных технологий и информатики, кафедра вычислительной техники СПбГЭТУ «ЛЭТИ».
E-mail: ivantishko29@gmail.com
https://orcid.org/0009-0003-8583-0387

Мелдо Анна Александровна – профессор кафедры вычислительной техники СПбГЭТУ «ЛЭТИ»; профессор кафедры лучевой диагностики, лучевой терапии и онкологии ФГБОУ ВО «Северный государственный медицинский университет» Минздрава России, Троицкий пр., д. 51, г. Архангельск, 163000, Россия; доцент кафедры лучевой диагностики Медицинского института ФГБОУ ВО «Санкт-Петербургский государственный университет», Северный пр., д. 1, Санкт-Петербург, 194354, Россия.
E-mail: anna.meldo@yandex.ru
https://orcid.org/0000-0002-4906-9901

Дора Светлана Владимировна – д-р мед. наук, профессор кафедры терапии факультетской с курсом эндокринологии, кардиологии с клиникой им. акад. Г. Ф. Ланга, Первый Санкт-Петербургский государственный медицинский университет имени академика И. П. Павлова, ул. Льва Толстого, д. 6–8, Санкт-Петербург, 197022, Россия.
E-mail: doras2001@mail.ru
https://orcid.org/0000-0002-8249-6075

Халимов Юрий Шавкатович – д-р мед. наук, профессор, заведующий кафедрой терапии факультетской с курсом эндокринологии, кардиологии с клиникой им. акад. Г. Ф. Ланга, Первый Санкт-Петербургский государственный медицинский университет имени академика И. П. Павлова, ул. Льва Толстого, д. 6–8, Санкт-Петербург, 197022, Россия.
E-mail: yushkha@gmail.com
https://orcid.org/0000-0002-7755-7275

Борискова Марина Евгеньевна – д-р мед. наук, профессор, руководитель отделения хирургической эндокринологии и пластической хирургии, Первый Санкт-Петербургский государственный медицинский университет имени академика И. П. Павлова, ул. Льва Толстого, д. 6–8, Санкт-Петербург, 197022, Россия.
E-mail: boriskovam@gmail.com
https://orcid.org/0000-0002-0037-6222

Ахметшина Лена Ахатовна – заведующий эндокринологическим отделением № 2 для взрослых, Городской консультативно-диагностический центр № 1, ул. Сикейроса, д. 10, лит. Д, Санкт-Петербург, 194354, Россия.
E-mail: doctor-akhmet@mail.ru
https://orcid.org/0009-0006-2451-9335

Смирнов Матвей Дмитриевич – студент 5-го курса, группа 521, лечебный факультет Первого Санкт-Петербургского государственного медицинского университета имени академика И. П. Павлова, ул. Льва Толстого, д. 6–8, Санкт-Петербург, 197022, Россия.
E-mail: matveyysmirn@rambler.ru
https://orcid.org/0009-0000-9756-0378

Вклад авторов:

Тишко И. В. – разработка идеи и дизайна исследования, предобработка данных, разработка и обучение моделей машинного обучения, анализ и интерпретация результатов, подготовка текста статьи.
Мелдо А. А. – научное руководство исследованием, консультирование по методологии применения методов машинного обучения и интерпретируемого ИИ, критический пересмотр текста статьи.
Дора С. В. – клиническая консультация по вопросам эндокринологии и кардиологии, медицинская интерпретация результатов исследования.
Халимов Ю. Ш. – клиническая консультация по вопросам эндокринологии, участие в формировании клинической концепции исследования.
Борискова М. Е. – клиническая консультация по вопросам хирургической эндокринологии, оценка клинической значимости полученных результатов.
Ахметшина Л. А. – клиническая консультация по вопросам ведения пациентов с заболеваниями щитовидной железы, участие в формировании набора клинических признаков.
Смирнов М. Д. – участие в сборе и клинической разметке данных, консультация по медицинским аспектам исследования.

References

1. Jajcev S. V., Privalov V. A. Rasprostranennost' raka shhitovidnoj zhelezy // Problemy jendokrinologii. 2002. T. 48, № 4. S. 13–16. doi: 10.14341/probl11558. (In Russ.).

2. Merabishvili V. M. Zlokachestvennye novoobrazovanija shhitovidnoj zhelezy (S73) v Sankt-Peterburge. Zabolevaemost', smertnost', dostovernost' ucheta, pogodichnaja letal'nost' i vyzhivaemost' bol'nyh // Vopr. onkologii. 2020. T. 66, № 5. S. 479–488. (In Russ.).

3. Stratifikacija riska recidiva papilljarnogo raka shhitovidnoj zhelezy na osnovanii rezul'tatov molekuljarno-geneticheskih issledovanij / S. A. Luk'janov, S. V. Sergijko, S. E. Titov, I. V. Reshetov, Ju. A. Verjaskina, A. V. Vazhenin, A. V. Gostimskij, L. I. Ippolitov, M. O. Rogova // Opuholi golovy i shei. 2020. T. 10, № 1. S. 93–100. doi: 10.17650/2222-1468-2020-10-1-93-100. (In Russ.).

4. Development and validation of prediction models for papillary thyroid cancer structural recurrence using machine learning approaches / H. Wang, C. Zhang, Q. Li, T. Tian, R. Huang, J. Qiu, R. Tian // BMC Cancer. 2024. Vol. 24, no. 1. P. 427. doi: 10.1186/s12885-024-12146-4.

5. Predicting thyroid cancer recurrence using supervised CatBoost: A SHAP-based explainable AI approach / A. A. Hanani, T. B. Donmez, M. Kutlu, M. Mansour // Medicine (Baltimore). 2025. Vol. 104, no. 22. P. e42667. doi: 10.1097/MD.0000000000042667.

6. Interpretable machine learning for thyroid cancer recurrence prediction: leveraging XGBoost and SHAP analysis / A. Schindele, A. Krebold, U. Heiß, K. Nimptsch, E. Pfaehler, C. Berr, R. A Bundschuh, T. Wendler, O. Kertels, J. Tran-Gia // Europ. J. of Radiology. 2025. Vol. 186. P. 112049. doi: 10.1016/j.ejrad.2025.112049.

7. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer / B. R. Haugen, E. K. Alexander, K. C. Bible, G. M. Doherty, S. J. Mandel, Yu. E. Nikiforov, F. Pacini, G. W. Randolph, A. M. Sawka, M. Schlumberger, K. G. Schuff, S. I. Sherman, Ju. A. Sosa, D. L. Steward, R. M. Tuttle, L. Wartofsky // Thyroid. 2016. Vol. 26, no. 1. P. 1–133. doi: 10.1089/thy.2015.0020.

8. Overview of the 2022 WHO Classification of Thyroid Neoplasms / Z. W. Baloch, S. L. Asa, J. A. Barletta, R. A. Ghossein, C. C. Juhlin, C. K. Jung, V. A. LiVolsi, M. G. Papotti, M. Sobrinho-Simões, G. Tallini, O. Mete // Endocrine Pathol. 2022. Vol. 33, № 1. P. 27–63. doi: 10.1007/s12022-022-09707-3.

9. A machine learning-based model for predicting recurrence in intermediate- and high-risk differentiated thyroid cancer: insights from a retrospective single-center study of 2388 patients / Y. Li, Z. Tang, A. Ren, G. Tian, J. Zhang, Y. Wang, J. Liu, J. Ming // Frontiers in Endocrinol. (Lausanne). 2025. Vol. 16. P. 1552479. doi: 10.3389/fendo.2025.1552479.

10. Yapici I. S., Arslan R. U. Predictive analytics for thyroid cancer recurrence: a feature selection and data balancing approach // The Europ. Phys. J. Spec. Topics. 2025. Vol. 234. P. 4751–4771. doi: 10.1140/epjs/s11734-025-01720-x.

11. Machine learning prediction of thyroid cancer recurrence for early screening and clinical decision pathways: a retrospective cohort study / G. Hu, B. Huang, L. Cai, Y. Zhang, Y. Zhang, Y. Liu, G. Wu // Discover Oncol. 2026. Vol. 17, no. 1. P. 239. doi: 10.1007/s12672-025-04293-2.

12. Sravnitel'nyj analiz modelej mashinnogo obuchenija dlja prognozirovanija vyzhivaemosti pri rake molochnoj zhelezy / E. A. Chernov, A. A. Meldo, A. V. Agaeva, M. A. Bogdanov, M. Ju. Val'kov // Sb. materialov XIII nauch.-prakt. konf. «Nauka nastojashhego i budushhego». SPb.: Izd-vo SPbGJeTU «LJeTI», 2025. S. 80–82. (In Russ.).

13. Machine learning prediction of recurrence in pediatric thyroid cancer: malignant endocrine tumors cohort analysis using XGBoost and SHAP / A. Redlich, E. Pfaehler, M. Kunstreich, M. Schmutz, C. Lapa, M. Kuhlen // J. of Clin. Endocrinol. & Metabolism. 2026. Vol. 111, no. 3. P. e844–e852. doi: 10.1210/clinem/dgaf487.

14. Bobrovskaja T. M. Metodologija formirovanija naborov dannyh i ih ispol'zovanie dlja ocenki diagnosticheskoj tochnosti sistem iskusstvennogo intellekta v luchevoj diagnostike: avtoref. dis. … kand. med. nauk: 3.3.9. M., 2025. 138 s. (In Russ.).

Information about the authors

Ivan V. Tishko – 4th-year student of the Faculty of Computer Technologies and Informatics, Department of Computer Engineering, Saint Petersburg Electrotechnical University.
E-mail: ivantishko29@gmail.com
https://orcid.org/0009-0003-8583-0387

Anna A. Meldo – Professor of the Department of Computer Engineering, Saint Petersburg Electrotechnical University; Professor of the Department of Radiation Diagnostics, Radiation Therapy and Oncology, Northern State Medical University, Tikhoretsky Pr., 51, Arkhangelsk, 163000, Russia; Associate Professor of the Department of Radiation Diagnostics, School of Medicine, Saint Petersburg State University, Severny Pr., 1, St. Petersburg, 194354, Russia.
E-mail: anna.meldo@yandex.ru
https://orcid.org/0000-0002-4906-9901

Svetlana V. Dora – Dr Sci. (Med.), Professor of the Department of Faculty Therapy with courses in Endocrinology and Cardiology (Academician G. F. Lang Clinic), Pavlov First Saint Petersburg State Medical University, L'va Tolstogo St., 6–8, St. Petersburg, 197022, Russia.
E-mail: doras2001@mail.ru
https://orcid.org/0000-0002-8249-6075

Yuriy Sh. Khalimov – Dr Sci. (Med.), Professor, Head of the Department of Faculty Therapy with courses in Endocrinology and Cardiology (Academician G. F. Lang Clinic), Pavlov First Saint Petersburg State Medical University, L'va Tolstogo St., 6–8, St. Petersburg, 197022, Russia.
E-mail: yushkha@gmail.com
https://orcid.org/0000-0002-7755-7275

Marina E. Boriskova – Dr Sci. (Med.), Professor, Head of the Department of Surgical Endocrinology and Plastic Surgery, Pavlov First Saint Petersburg State Medical University, L'va Tolstogo St., 6–8, St Petersburg, 197022, Russia.
E-mail: boriskovam@gmail.com
https://orcid.org/0000-0002-0037-6222

Lena A. Akhmetshina – Head of the Adult Endocrinology Department No. 2, City Consultative Diagnostic Center No. 1, Sikeyrosa St., 10, lit. D, St. Petersburg, 194354, Russia.
E-mail: doctor-akhmet@mail.ru
https://orcid.org/0009-0006-2451-9335

Matvey D. Smirnov – 5th-year student of Pavlov First Saint Petersburg State Medical University, L'va Tolstogo St. 6–8, St. Petersburg, 197022, Russia.
E-mail: matveyysmirn@rambler.ru
https://orcid.org/0009-0000-9756-0378

Author contribution statement:

Tishko I. V. – conceived the study design, performed data preprocessing, developed and trained the machine learning models, analyzed and interpreted the results, drafted the manuscript.

Meldo A. A. – supervised the study, provided consultation on machine learning and explainable AI methodology, critically revised the manuscript.

Dora S. V. – provided clinical consultation on endocrinology and cardiology, contributed to the medical interpretation of the results.

Khalimov Yu. Sh. – provided clinical consultation on endocrinology, contributed to the clinical concept of the study.

Boriskova M. E. – provided clinical consultation on surgical endocrinology, assessed the clinical relevance of the findings.

Akhmetshina L. A. – provided clinical consultation on the management of thyroid cancer patients, contributed to the formation of the clinical feature set.

Smirnov M. D. – participated in data collection and clinical annotation, provided consultation on medical aspects of the study.

Статья поступила в редакцию 28.07.2026; принята к публикации после рецензирования 22.09.2026; опубликована онлайн 27.11.2026.

Submitted 28.07.2026; accepted 22.09.2026; published online 27.11.2026.