

DOI: <https://doi.org/10.17816/DD626145>

Разработка прогностической модели для диагностики рака предстательной железы на основе радиомики бипараметрической магнитно-резонансной томографии карт измеряемого коэффициента диффузии и стекинга алгоритмов машинного обучения

А.И. Кузнецов

Московский авиационный институт (национальный исследовательский университет), Москва, Россия

АННОТАЦИЯ

Обоснование. Рак предстательной железы является одним из наиболее распространённых онкологических заболеваний среди мужчин [1, 2]. За последние годы в мире создан ряд прогностических моделей на основе текстурного анализа изображений бипараметрической магнитно-резонансной томографии. Исследования показали, что радиомические признаки, извлечённые из карт измеряемого коэффициента диффузии, являются наиболее воспроизводимыми [3]. Однако эти модели имеют ограничения по точности, поскольку построены с помощью одного алгоритма машинного обучения, который учитывает только линейную зависимость [4–6].

Цель — повышение точности прогностической модели, диагностирующей рак предстательной железы, за счёт использования стекинга алгоритмов машинного обучения, которые учитывают не только линейные, но и нелинейные зависимости на основе радиомики бипараметрической магнитно-резонансной томографии карт измеряемого коэффициента диффузии.

Материалы и методы. Проведено одноцентровое когортное ретроспективное исследование пациентов с подозрением на рак предстательной железы в отделении рентгеновской диагностики и томографии Объединённой больницы с поликлиникой (Москва) с 2017 по 2023 год. Наличие рака предстательной железы подтверждено биопсией или радикальной простатэктомией. Статистический анализ проводился с использованием Python 3.11.

Результаты. В исследовании приняло участие 67 мужчин в возрасте 60 [54; 66] лет, из них у 57 — рак предстательной железы, а у 10 — доброкачественное образование простаты. Программное обеспечение LIFEx позволило выделить 96 радиомических параметров.

Статистически значимые различия выявлены по: PARAMS_ZSpatialResampling (размер вокселя по оси Z) ($p=0,001$), SHAPE_Sphericity[onlyFor3DROI] (сферичность формы) ($p=0,006$), SHAPE_Compacity[onlyFor3DROI] (компактность формы) ($p=0,004$), GLRLM_HGRE ($p=0,039$), GLRLM_SRHGE ($p=0,041$), GLRLM_RLNU ($p=0,039$). GLRLM — матрица длин серий уровня серого. Однофакторная логистическая регрессия показала, что статистически значимо влияли на исход: SHAPE_Compacity[onlyFor3DROI] ($R^2=15\%$) и PARAMS_ZSpatialResampling ($R^2=18\%$). Методом многофакторной логистической регрессии была построена прогностическая модель, которая учитывает линейные зависимости. Модель включает 3 показателя, которые совместно статистически значимо влияют на исход ($R^2=23\%$): SHAPE_Sphericity[onlyFor3DROI], PARAMS_ZSpatialResampling и GLRLM_RLNU.

Для описания нелинейных связей, была построена другая модель на основе алгоритма «Дерево решений». В неё вошли 4 показателя ($R^2=58\%$): DISCRETIZED_HISTO_Entropy_log10 (случайность распределения), SHAPE_Sphericity[onlyFor3DROI], PARAMS_ZSpatialResampling и GLRLM_SRE.

Стекинг алгоритмов, состоящий в расчёте среднеарифметического между предсказаниями алгоритмов многофакторной логистической регрессии и «Дерева решений», позволил построить модель, которая учитывает линейные и нелинейные зависимости. Модель включает 5 параметров ($R^2=77\%$). Построенная модель легла в основу программы калькулятора [7], внедрённой в данный момент в практику врача-рентгенолога.

Заключение. Новая модель, построенная на основе карт измеряемого коэффициента диффузии, работает лучше (площадь под ROC-кривой 99,0% [97,7; 100,0]), чем существующие модели с площадью под ROC-кривой 83,6% [78,3; 88,9], которые показывают значительную вариативность ($I^2=71\%$). Повышение точности работы модели произошло за счёт применения стекинга технологий машинного обучения, который позволил учесть как линейное, так и нелинейное влияние переменных на исход.

Received: 28.01.2024

Accepted: 13.02.2024

Published online: 30.06.2024

Ключевые слова: бипараметрическая магнитно-резонансная томография; рак предстательной железы; радиомика; текстурный анализ; машинное обучение.

Как цитировать:

Кузнецов А.И. Разработка прогностической модели для диагностики рака предстательной железы на основе радиомики бипараметрической магнитно-резонансной томографии карт измеряемого коэффициента диффузии и стекинга алгоритмов машинного обучения // Digital Diagnostics. Т. 5, № S1. С. 80–82. DOI: <https://doi.org/10.17816/DD626145>

СПИСОК ЛИТЕРАТУРЫ

1. Mottet N., van den Bergh R.C.N., Briers E., et al. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer—2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent // Eur Urol. 2021. Vol. 79, N 2. P. 243–262. doi: 10.1016/j.eururo.2020.09.042
2. Здоровоохранение в России. 2021: статистический сборник / под ред. П.А. Смелова, С.Ю. Никитиной. Москва : Федеральная служба государственной статистики, 2021.
3. Shah V., Turkbey B., Mani H., et al. Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging // Med Phys. 2012. Vol. 39, N 7. P. 4093–4103. doi: 10.1118/1.4722753
4. He D., Wang X., Fu C., et al. MRI-based radiomics models to assess prostate cancer, extracapsular extension and positive surgical margins // Cancer Imaging. 2021. Vol. 21, N 1. P. 46. doi: 10.1186/s40644-021-00414-6
5. Lu Y., Li B., Huang H., et al. Biparametric MRI-based radiomics classifiers for the detection of prostate cancer in patients with PSA serum levels of 4–10 ng/mL // Front Oncol. 2022. Vol. 12. doi: 10.3389/fonc.2022.1020317
6. Chen T., Li M., Gu Y., et al. Prostate Cancer Differentiation and Aggressiveness: Assessment With a Radiomic-Based Model vs. PI-RADS v2 // Journal of Magnetic Resonance Imaging. 2019. Vol. 49, N 3. P. 875–884. doi: 10.1002/jmri.26243
7. Свидетельство о регистрации программы для ЭВМ 2023669718/ 19.09.2023. Щепкина Е.В., Крючкова О.В., Кузнецов А.И. Программное обеспечение для прогнозирования вероятности выявления злокачественных новообразований (рака) предстательной железы у мужчин на основе Мрт-радиомики. Режим доступа: <https://www.elibrary.ru/item.asp?id=54657407> Дата обращения: 28.01.2024. EDN: FNYFEB

DOI: <https://doi.org/10.17816/DD626145>

Development of a prognostic model for diagnosis of prostate cancer based on radiomics of biparametric magnetic resonance imaging apparent diffusion coefficient maps and stacking of machine learning algorithms

Anton I. Kuznetsov

Moscow Aviation Institute (National Research University), Moscow, Russia

ABSTRACT

BACKGROUND: Prostate cancer is one of the most common cancers among men [1, 2]. In recent years, a number of prognostic models based on texture analysis of biparametric magnetic resonance images have been created. The research has shown that radiomics features extracted from apparent diffusion coefficient maps are the most reproducible [3]. However, the models were limited in accuracy, since they are built using a single machine learning algorithm, which takes into account only linear dependences [4–6].

AIM: Increasing the accuracy of a prognostic model diagnosing prostate cancer through the use of stacking machine learning algorithms that takes into account not only linear, but also nonlinear dependencies based on radiomics of biparametric magnetic resonance imaging apparent diffusion coefficient maps.

MATERIALS AND METHODS: A single-center cohort retrospective study of patients with suspected prostate cancer was conducted in the X-ray Diagnostics and Tomography Department of the United Hospital and Polyclinic (Moscow, Russia) from 2017 to 2023. The presence of prostate cancer was confirmed by biopsy or radical prostatectomy. Statistical analyses was performed using Python 3.11.

RESULTS: The study involved 67 men aged 60 [54; 66] years, of which 57 were diagnosed with prostate cancer, and 10 — with benign prostate formation. The LIFEx software identified 96 radiomic features.

Statistically significant differences were found for: PARAMS_ZSpatialResampling (the voxel size of the image: Z dimension) ($p=0.001$), SHAPE_Sphericity[onlyFor3DROI] (how spherical a Volume of Interest is) ($p=0.006$), SHAPE_Compacity[onlyFor3DROI]

Рукопись получена: 28.01.2024

Рукопись одобрена: 13.02.2024

Опубликована online: 30.06.2024

(how compact the Volume of Interest is) ($p=0.004$), GLRLM_HGRE ($p=0.039$), GLRLM_SRHGE ($p=0.041$), GLRLM_RLNU ($p=0.039$), where GLRLM — Grey-Level Run Length Matrix. Univariate logistic regression showed that SHAPE_Compacity[onlyFor3DROI] ($R^2=15\%$) and PARAMS_ZSpatialResampling ($R^2=18\%$) had a statistically significant effect on the outcome. First, using the multivariate logistic regression method, a prognostic model was built that takes into account only linear dependencies. The model includes 3 features that together have a statistically significant effect on the outcome ($R^2=23\%$): SHAPE_Sphericity[onlyFor3DROI], PARAMS_ZSpatialResampling and GLRLM_RLNU.

To describe nonlinear relationships, another model was built based on the “Decision Tree” algorithm. It included 4 indicators ($R^2=58\%$): DISCRETIZED_HISTO_Entropy_log10 (the randomness of the distribution), SHAPE_Sphericity[onlyFor3DROI], PARAMS_ZSpatialResampling and GLRLM_SRE.

Stacking of algorithms, which consists of calculating the arithmetic mean between the predictions of the multivariate logistic regression and “Decision Tree” algorithms, made it possible to construct a model that takes into account both linear and nonlinear dependencies. The model includes 5 features ($R^2=77\%$). The constructed model formed the basis of the developed calculator program [7], currently introduced into a radiology practice.

CONCLUSION: The new model built on the basis of apparent diffusion coefficient maps performs better (area under ROC-curve 99.0% [97.7; 100.0]) than the existing models with area under ROC-curve 83.6% [78.3; 88.9], which also show high heterogeneity ($I^2=71\%$). The accuracy of the new model was increased due to the use of stacking machine learning technologies, which made it possible to take into account both linear and nonlinear effects from features on the outcome.

Keywords: biparametric magnetic resonance imaging; prostate cancer; radiomics; texture analysis; machine learning.

To cite this article:

Kuznetsov AI. Development of a prognostic model for diagnosis of prostate cancer based on radiomics of biparametric magnetic resonance imaging apparent diffusion coefficient maps and stacking of machine learning algorithms. *Digital Diagnostics*. 2024;5(S1):80–82. DOI: <https://doi.org/10.17816/DD626145>

REFERENCES

- Mottet N, van den Bergh RCN, Briers E, et al. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer—2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol*. 2021;79(2):243–262. doi: 10.1016/j.eururo.2020.09.042
- Smelov PA, Nikitina SYu, editors. Health care in Russia. 2021: statistical compendium. Moscow: Federal'naya sluzhba gosudarstvennoy statistiki; 2021. (In Russ).
- Shah V, Turkbey B, Mani H, et al. Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging. *Med Phys*. 2012;39(7):4093–4103. doi: 10.1118/1.4722753
- He D, Wang X, Fu C, et al. MRI-based radiomics models to assess prostate cancer, extracapsular extension and positive surgical margins. *Cancer Imaging*. 2021;21(1):46. doi: 10.1186/s40644-021-00414-6
- Lu Y, Li B, Huang H, et al. Biparametric MRI-based radiomics classifiers for the detection of prostate cancer in patients with PSA serum levels of 4–10 ng/mL. *Front Oncol*. 2022;12. doi: 10.3389/fonc.2022.1020317
- Chen T, Li M, Gu Y, et al. Prostate Cancer Differentiation and Aggressiveness: Assessment With a Radiomic-Based Model vs. PI-RADS v2. *Journal of Magnetic Resonance Imaging*. 2019;49(3):875–884. doi: 10.1002/jmri.26243
- Certificate for the computer software 2023669718/ 19.09.2023. Schepkina EV, Kruchkova OV, Kuznetsov AI. Computer Software for the Prognosis of Probability of Prostate Cancer in Patients Based on MRI. Available from: <https://www.elibrary.ru/item.asp?id=54657407> [cited 2024 Jan 28]. (In Russ). EDN: FNYFEB

ОБ АВТОРЕ

Кузнецов Антон Игоревич;
ORCID: 0000-0003-2182-5792;
eLibrary SPIN: 8824-9080;
e-mail: drednout5786@yandex.ru

AUTHOR'S INFO

Anton I.Kuznetsov;
ORCID: 0000-0003-2182-5792;
eLibrary SPIN: 8824-9080;
e-mail: drednout5786@yandex.ru