

ISSN 2782-3806
ISSN 2782-3814 (Online)
УДК 616-006.484:004.8

РАЗРАБОТКА МОДЕЛИ ПРОГНОЗИРОВАНИЯ ВТЭО НА ОСНОВЕ АВТОМАТИЧЕСКИ ОТОБРАННЫХ ПРИЗНАКОВ У ПАЦИЕНТОВ С ГЛИОМОЙ

Леонтьев С. С.¹, Симакова М. А.², Лукинов В. Л.³, Пищулов К. А.²,
Абрамян Л. К.², Угольников Е. А.³

¹ Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО», Санкт-Петербург, Россия

² Федеральное государственное бюджетное учреждение «Национальный медицинский исследовательский центр имени В. А. Алмазова» Министерства здравоохранения Российской Федерации, Научный центр мирового уровня «Центр персонализированной медицины», Санкт-Петербург, Россия

³ Федеральное государственное бюджетное учреждение науки «Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук», Новосибирск, Россия

Контактная информация:

Угольникова Екатерина Алексеевна,
ИВМиМГ СО РАН,
пр. Академика Лаврентьева, 6,
Новосибирск, Россия, 630090.
E-mail: giekooolis@gmail.com

Статья поступила в редакцию 02.10.2024
и принята к печати 31.10.2024

РЕЗЮМЕ

Венозные тромбоэмболические осложнения (ВТЭО) представляют серьезную угрозу для пациентов, проходящих лечение от рака, особенно на фоне прогрессирующих и метастатических заболеваний. В нейроонкологии частота возникновения ВТЭО зависит от расположения и стадии опухоли. Некоторые первичные и метастатические опухоли головного мозга имеют повышенную склонность к тромботическим событиям. В этом исследовании мы применили современные методы машинного обучения, в частности XGBoost, чтобы создать модели поиска предикторов, связанных с риском ВТЭО у пациентов с глиомой. Сравнивая точность диагностики наших моделей XGBoost с традиционными подходами логистической регрессии, мы стремимся углубить понимание прогнозирования ВТЭО в этой группе пациентов. Наши результаты дополняют растущий объем исследований по оценке риска тромбоза у онкологических больных и могут помочь в разработке персонализированных стратегий профилактики и лечения для снижения риска ВТЭО у пациентов с глиомой в стационаре.

Ключевые слова: ВТЭО, клиническая диагностика, машинное обучение, предсказательное моделирование

Для цитирования: Леонтьев С.С., Симакова М.А., Лукинов В.Л. и др. Разработка модели прогнозирования ВТЭО на основе автоматически отобранных признаков у пациентов с глиомой. Российский журнал персонализированной медицины. 2024;4(6):517-530. DOI: 10.18705/2782-3806-2024-4-6-517-530. EDN: SSZMPC

DEVELOPMENT OF A VTE PREDICTION MODEL BASED ON AUTOMATICALLY SELECTED FEATURES IN GLIOMA PATIENTS

Leontev S. S.¹, Simakova M. A.², Lukinov V. L.³, Pishchulov K. A.²,
Abramyan L. K.², Ugolnikova E. A.³

¹ ITMO University, Saint Petersburg, Russia

² Almazov National Medical Research Centre, World-Class Research Centre for Personalized Medicine, Saint Petersburg, Russia

³ Institute of Computational Mathematics and Mathematical Geophysics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

Corresponding author:

Ugolnikova Ekaterina A.,
ICMMG SB RAS,
Ac. Lavrentieva ave., 6, Novosibirsk, Russia,
630090.
E-mail: giekoolis@gmail.com

Received 02 October 2024; accepted 31
October 2024

ABSTRACT

Venous thromboembolism (VTE) is a serious threat to patients undergoing cancer treatment, especially in advanced and metastatic diseases. In neuro-oncology, the incidence of VTE depends on the location and stage of the tumor. Some primary and secondary brain tumors have an increased propensity for thrombotic events. In this study, we applied state-of-the-art machine learning methods, particularly XGBoost, to create models to search for predictors associated with the risk of VTE in glioma patients. By comparing the diagnostic accuracy of our XGBoost models with traditional logistic regression approaches, we aim to advance the understanding of VTE prediction in this patient population. Our results add to the growing body of research on thrombosis risk assessment in cancer patients and may help in the development of personalized prevention and treatment strategies to reduce the risk of VTE in hospitalized glioma patients.

Key words: clinical diagnosis, machine learning, predictive modeling, VTE

For citation: Leontev SS, Simakova MA, Lukinov VL, et al. Development of a VTE prediction model based on automatically selected features in glioma patients. Russian Journal for Personalized Medicine. 2024;4(6):517-530. (In Russ.) DOI: 10.18705/2782-3806-2024-4-6-517-530. EDN: SSZMPC

ВВЕДЕНИЕ

Венозные тромбоэмболические осложнения (ВТЭО) — это группа заболеваний, включающих в себя тромбоз поверхностных вен, тромбоз глубоких вен (ТГВ), венозную гангрену и тромбоэмболию легочной артерии (ТЭЛА). В области кардиоонкологии рак-ассоциированные тромбозы представляют собой серьезную проблему и тесно связаны с повышением ранней смертности от всех причин во время химиотерапии и хирургического вмешательства при злокачественных новообразованиях [1–3].

Построение моделей прогнозирования риска ВТЭО остается актуальной проблемой и по сей день [4–8].

Частота венозных тромбозов зависит от локализации опухоли и значительно увеличивается на поздних и метастатических стадиях.

Было замечено, что некоторые виды рака, такие как рак поджелудочной железы, почек, яичников, легких, желудочно-кишечного тракта и опухоли головного мозга, имеют повышенную склонность к образованию тромбов. Эти опухоли делятся на первичные и вторичные, связанные с метастазированием. Двумя наиболее распространенными первичными опухолями головного мозга являются менингиома и глиальные опухоли, на долю которых приходится 35,6 % и 35,5 % случаев соответственно [9]. Глиобластома — наиболее распространенная злокачественная

опухоль головного мозга, составляющая 15,6 % случаев среди всех первичных опухолей головного мозга.

Целью данного исследования было сравнение методов XGBoost и логистической регрессии как инструментов для создания модели стратификации риска венозных тромбоэмболических событий у пациентов с глиомой.

МАТЕРИАЛЫ И МЕТОДЫ

С января 2021-го по май 2023 года в Национальном медицинском исследовательском центре имени В. А. Алмазова было проведено исследование, в которое вошли 286 последовательных пациентов с гистологически верифицированной глиомой, перенесших операцию. В группу вошли 133 (51,2 %) мужчины и 132 (49,8 %) женщины, средний возраст которых составил 54 [41; 63] года. Перед операцией всем пациентам было проведено УЗИ вен нижних конечностей для выявления тромбоза глубоких вен (ТГВ). Диагноз тромбоэмболии легочной артерии и тромбоза глубоких вен устанавливался в соответствии с современными клиническими рекомендациями [10].

Непрерывные переменные были проверены на нормальность с помощью теста Шапиро-Уилка. В связи с ненормальностью распределений данных для их описания использовали медиану с первым и третьим квартилями (Q1; Q3). Распределения сравнивали с помощью U-критерия Манна-Уитни.

Таблица 1. Клинические характеристики ретроспективной группы

Table 1. The clinical characteristics of the retrospective group

Показатель	Значения
	Медиана [Q1; Q3]
Индекс массы тела, кг/м ²	26,9 [23,57; 30,1]
	Частота возникновения [95 % ДИ]
Головная боль	68,3 % [59 %; 78 %]
Очаговый дефицит	32,8 % [29 %; 49 %]
Предыдущая химиотерапия	17,7 % [11 %; 28 %]
Предыдущая лучевая терапия	19,9 % [14 %; 29 %]
Размер опухоли > 5 см	55,7 % [42 %; 64 %]
Биопсия опухоли	7 % [5 %; 13 %]
Частичная резекция опухоли	49,7 % [42 %; 61 %]

Для категориальной переменной GRADE вычислялась частота (количество пациентов) в каждой категории и сравнивалась с помощью точного теста Фишера.

В исследовании определяли частоту бинарных переменных, указывающих на возникновение ВТЭО, клинические проявления новообразований и сопутствующую сердечно-сосудистую патологию; 95 % доверительные интервалы (95 % ДИ) оценивались по формуле Уилсона и сравнивались с помощью точного теста Фишера.

Модели сравнивались по площади под ROC-кривыми (AUC) с помощью теста Делонга.

Также сравнивались прогностические характеристики, связанные с развитием ВТЭ. Чувствительность и специфичность оценивались с помощью критерия Мак-Немара, а положительная и отрицательная прогностические ценности (PPV и NPV) сравнивались с помощью взвешенного обобщенного теста (WSG-тест).

Для решения проблемы множественных сравнений р-значения корректировались с помощью метода Бенджамини-Хохберга. Статистические гипотезы проверялись при уровне значимости

$p = 0,05$. Различия считались статистически значимыми, если $p < 0,05$.

ГРУППЫ ПАЦИЕНТОВ

Ретроспективная группа

В ретроспективную группу были включены 186 пациентов с подтвержденной глиомой (хирургическое лечение с 2019 по 2021 гг.). Частота возникновения ВТЭО составила 8,1 % ($n = 15$).

Средний возраст на момент включения в исследование составил 54 [42; 63] года, 54,3 % пациентов — женщины.

Клинические данные ретроспективной группы представлены в таблице 1.

Пациенты были разделены на 2 группы в зависимости от наличия ВТЭО. Спектр сопутствующих заболеваний в зависимости от группы представлен в таблице 2.

Факторы риска ВТЭО, такие как очаговый неврологический дефицит (ОР 8; 95 % ДИ [1,4; 18,7]), варикозное расширение вен нижних конечностей, размер опухоли более 5 см в одном из измерений, чаще встречались у пациентов с ВТЭО (табл. 3).

Таблица 2. Спектр сопутствующих заболеваний в зависимости от наличия венозных тромбозмболических осложнений для ретроспективной группы

Table 2. The spectrum of comorbid pathology depending on the presence of venous thromboembolic complications for retrospective group

Сопутствующее заболевание	Группа с ВТЭО ($n = 15$)	Группа без ВТЭО ($n = 171$)	Критерий Фишера, p
Гипертония	73 % [48 %; 89 %]	58 % [50 %; 65 %]	0,284
Дислипидемия	20 % [7 %; 45 %]	24 % [18 %; 31 %]	> 0,999
Ишемическая болезнь сердца	27 % [11 %; 52 %]	19 % [14 %; 26 %]	0,505
Нарушения ритма и проводимости	0 % [0 %; 20 %]	4 % [2 %; 7 %]	> 0,999
Хроническая сердечная недостаточность	30,5 % ($n = 7$)	6,5 % ($n = 5$)	> 0,999
Заболевания дыхательной системы	7 % [1 %; 30 %]	9 % [6 %; 14 %]	> 0,999
Аутоиммунные заболевания щитовидной железы	0 % [0 %; 22 %]	9 % [5 %; 14 %]	0,607
Заболевания пищеварительной системы	50 % [27 %; 73 %]	36 % [29 %; 43 %]	0,387
Заболевания мочеполовой системы	14 % [4 %; 40 %]	11 % [7 %; 17 %]	0,687

Таблица 3. Факторы риска развития венозных тромбоэмболических событий (ВТЭ), разделенные на категории для ретроспективной группы

Table 3. Risk factors for the development of Venous Thromboembolic Events (VTE), divided into categories for retrospective group

Фактор риска	Группа с ВТЭО	Группа без ВТЭО	Критерий Фишера / U-критерий Манна-Уитни, p
Рецидив/прогрессирование опухоли	33 % [15 %; 58 %]	25 % [19 %; 32 %]	0,537
Очаговый неврологический дефицит	67 % [42 %; 85 %]	29 % [23 %; 37 %]	0,007*
Длительный постельный режим	33 % [15 %; 58 %]	4 % [2 %; 8 %]	< 0,001
История венозных тромбоэмболических событий	7 % [1 %; 30 %]	1 % [0 %; 4 %]	0,225
Варикозное расширение вен нижних конечностей	33 % [15 %; 58 %]	11 % [7 %; 16 %]	0,025*
Химиотерапия в прошлом	27 % [11 %; 52 %]	16 % [12 %; 23 %]	0,080
Лучевая терапия в прошлом	40 % [20 %; 64 %]	18 % [13 %; 24 %]	0,170
Индекс массы тела	26,99 [24,72; 30,86]	26,845 [23,66; 30,06]	0,742
Доза глюкокортикостероидов перед операцией	14 [11; 17]	8 [4; 16]	0,120
Наличие стереотаксической биопсии	7 % [1 %; 30 %]	7 % [4 %; 12 %]	> 0,999
Частичная резекция опухоли	36 % [16 %; 61 %]	52 % [44 %; 59 %]	0,278
Размер опухоли > 5 см в одном из измерений	67 % [42 %; 85 %]	42 % [35 %; 50 %]	0,101
Стадия в соответствии с классификацией GRADE	3–2 (14,3 %) 4–11 (78,6 %)	3–2 (14,3 %) 4–11 (78,6 %)	0,012*

* — статистически значимые различия.

** — псевдомедиана парных различий (пМЕД) с 95 % ДИ.

Проспективная группа

В проспективную группу были включены 100 пациентов, средний возраст которых составил 42,5 [40; 62,3] года, 60 % (n = 60) — мужчины. Частота ВТЭ составила 23 % (n = 23), среди которых тромбоз глубоких вен нижних конечностей имел место в 78 % (n = 18) случаев, тромбоз легочной артерии — в 13 % (n = 3), а сочетание ТГВ и ТЭЛА выявлено в 8,7 % (n = 2) случаев. Среднее время развития ВТЭО от момента операции составило 10 [5; 16] дней. Клинические данные пациентов проспективной группы представлены в таблице 4.

Спектр сопутствующих заболеваний в зависимости от наличия венозных тромбоэмболических осложнений для проспективной группы представлен в таблице 5.

Следует отметить, что в группе с ВТЭ значительно чаще встречались нарушения ритма и проводимости (26,1 % (n = 6) против 3,9 % (n = 3), p = 0,005), а также застойная сердечная недостаточность (ЗСН) (30,5 % (n = 7) против 6,5 % (n = 5), p = 0,002). Кроме того, наличие сопутствующих нарушений ритма повышало риск развития ВТЭО в 2,15 раза (0,67; 3,6), а наличие ХСН — в 1,84 раза (0,57; 3,1).

Таблица 4. Клинические характеристики проспективной группы

Table 4. The clinical characteristics of the prospective group

Показатель	Значения
	Медиана [Q1; Q3]
Индекс массы тела, кг/м ²	26,18 [23,55; 30,05]
Объем опухоли, см ³	60,98 [28,59; 99,7]
	Частота возникновения [95 % ДИ]
Головная боль	70 % [59 %; 79 %]
Очаговый дефицит	44 % [34 %; 55 %]
Предыдущая химиотерапия	18 % [11 %; 28 %]
Предыдущая лучевая терапия	19 % [12 %; 29 %]
Биопсия опухоли	12 % [6 %; 20 %]
Частичная резекция опухоли	43 % [32 %; 54 %]

Таблица 5. Спектр сопутствующих заболеваний в зависимости от наличия венозных тромбоэмболических осложнений для проспективной группы

Table 5. The spectrum of comorbid pathology depending on the presence of venous thromboembolic complications for prospective group

Сопутствующее заболевание	Группа с ВТЭО (n = 23)	Группа без ВТЭО (n = 77)	Критерий Фишера, p
Гипертония	52,1 % (n = 12)	41,5 % (n = 32)	0,474
Дислипидемия	56,5 % (n = 13)	31,2 % (n = 24)	0,047
Ишемическая болезнь сердца	8,7 % (n = 2)	3,8 % (n = 3)	0,324
Нарушения ритма и проводимости	26,1 % (n = 6)	3,9 % (n = 3)	0,005
Хроническая сердечная недостаточность	30,5 % (n = 7)	6,5 % (n = 5)	0,002
Заболевания дыхательной системы	8,7 % (n = 2)	6,4 % (n = 5)	0,659
Аутоиммунные заболевания щитовидной железы	8,7 % (n = 2)	12,3 % (n = 10)	0,812
Заболевания пищеварительной системы	30,4 % (n = 7)	19,4 % (n = 15)	0,267
Заболевания мочеполовой системы	8,7 % (n = 2)	11,6 % (n = 9)	0,687

Факторы риска развития ВТЭ, разделенные на категории, представлены в таблице 6.

Таблица 6. Факторы риска развития венозных тромбоэмболических событий (ВТЭ), разделенные на категории для проспективной группы

Table 6. Risk factors for the development of Venous Thromboembolic Events (VTE), divided into categories for prospective group

Фактор риска	Группа с ВТЭО (n = 23)	Группа без ВТЭО (n = 77)	Критерий Фишера / U-критерий Манна-Уитни, p
Рецидив/прогрессирование опухоли	26,1 % (n = 6)	20,7 % (n = 16)	0,570
Очаговый неврологический дефицит	56,5 % (n = 13)	37,6 % (n = 29)	0,149
Длительный постельный режим	47,8 % (n = 11)	5,1 % (n = 4)	< 0,001
История венозных тромбоэмболических событий	30,4 (n = 7)	0 %	< 0,001
Варикозное расширение вен нижних конечностей	34,7 % (n = 8)	3,8 % (n = 3)	< 0,001
Химиотерапия в прошлом	22,7 % (n = 5)	14,2 % (n = 11)	0,340
Лучевая терапия в прошлом	31,8 % (n = 7)	18,8 % (n = 14)	0,170
Индекс массы тела	28 [25,3; 29,9]	26 [22; 29]	0,238
Доза глюкокортикостероидов перед операцией	12 [10; 12]	4 [0; 8]	< 0,001
Наличие стереотаксической биопсии	0 %	8 % (n = 6)	0,162
Частичная резекция опухоли	45,5 % (n = 10)	45,8 % (n = 33)	0,975
Размер опухоли > 5 см в одном из измерений	52,1 % (n = 12)	48 % (n = 36)	0,726
Объем опухоли	62 [32; 84]	62 [22; 110]	0,890
Стадия в соответствии с классификацией GRADE	3–3 (13 %) 4–20 (87 %)	1–5 (6,8 %) 2–16 (21,2 %) 3–15 (20,2 %) 4–50 (37 %)	0,012*

ПОСТРОЕНИЕ МОДЕЛИ

Метрики

При построении модели для максимизации были выбраны такие параметры, как специфичность (Specificity) и прецизионность (чувствитель-

ность) (Precision). Это означает, что модели с более высокой суммой специфичности и точности будут считаться более качественными.

Формулы для расчета метрик выглядят следующим образом (1) и (2):

$$Specificity = True\ Positive\ (TP) / (True\ Negative\ (TN) + False\ Positive\ (FP)) \tag{1}$$

$$Precision = True\ Positive\ (TP) / (True\ Positive\ (TP) + False\ Positive\ (FP)) \tag{2}$$

Первичная обработка данных

В первую очередь было проведено преобразование категориальных признаков (таким признаком оказалась только «Группа крови»). Для преобразования был выбран метод One-Hot-Encoding с удалением первого столбца.

Характеристики пола и роста также были удалены из исходного набора данных, поскольку они явно не влияют на результаты медицинских исследований и не отражают реальных различий при постановке диагноза, а их включение в итоговую модель может привести к статистическим ошибкам.

Кроме того, были исключены параметры, по которым имелось менее 80 % данных.

На заключительном этапе обработки первичных данных были исключены записи, для которых не был указан параметр наличия ВТЭО.

Отбор признаков для финальной модели

Модели XGBoost были выбраны как в качестве конечной модели, так и в качестве модели отбора признаков. Эти модели были разделены на тестовые и обучающие наборы в соответствии с правилами, описанными в предыдущих главах.

Затем оставшиеся параметры последовательно включались в модель. Выбирались те параметры, которые приводили к наибольшему увеличению специфичности. Процесс начинался с одного параметра, затем добавлялся другой, и так продолжалось до тех пор, пока не были изучены наборы параметров от 5 до 10 штук. Окончательная модель показала наиболее благоприятные результаты с 7 параметрами.

Выбранные параметры затем автоматически передавались в основную модель для ее обучения.

Таблица 7. Метрики полученной модели

Table 7. Model result metrics

Метрика	Значения [95 % ДИ]
Специфичность (specificity)	93 % [87 %; 99 %]
Чувствительность (precision)	77 % [58 %; 94 %]
Точность (accuracy)	84 % [76 %; 90 %]
Полнота (recall)	61 % [42 %; 77 %]
F1-мера (F1-score)	68 % [51 %; 81 %]

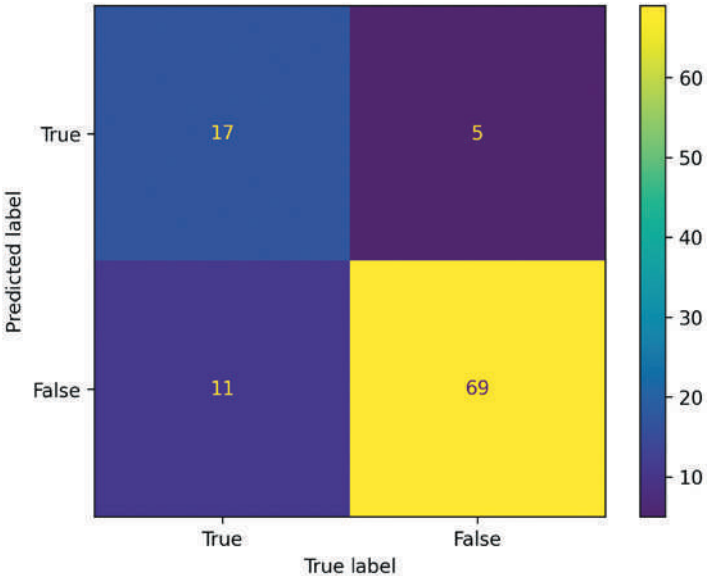


Рис. 1. Таблица точности диагностического тестирования

Figure 1. Diagnostic testing accuracy table

Построение модели

Далее была построена финальная модель, которая динамически получает на вход параметры, полученные на предыдущем этапе. Сама модель также представляет собой XGBoost, но с параметрами, отличными от тех, которые использовались при выборе параметров.

Результаты

На этапе выбора параметров для передаваемых данных были получены следующие 7 параметров:

[«D-димер», «ИМТ», «парез нижних конечностей», «легочная недостаточность», «варикозное расширение вен», «гипертония», «дислипидемия»].

Модель смогла достичь следующих показателей по основным метрикам (табл. 7).

Таблица точности диагностического тестирования представлена на рисунке 1.

Область под ROC-кривой составляет 0,77 (рис. 2).

Также были построены Shap-значения для финальной модели и дерева результирующей модели (см. рис. 3, 4).

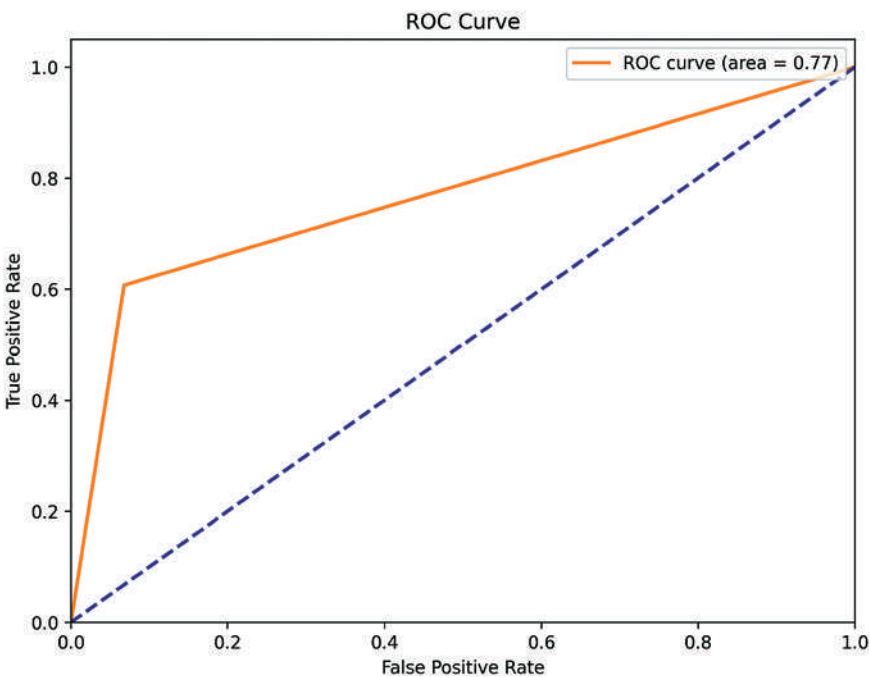


Рис. 2. ROC-кривая для XGBoost модели

Figure 2. ROC curve of XGBoost model

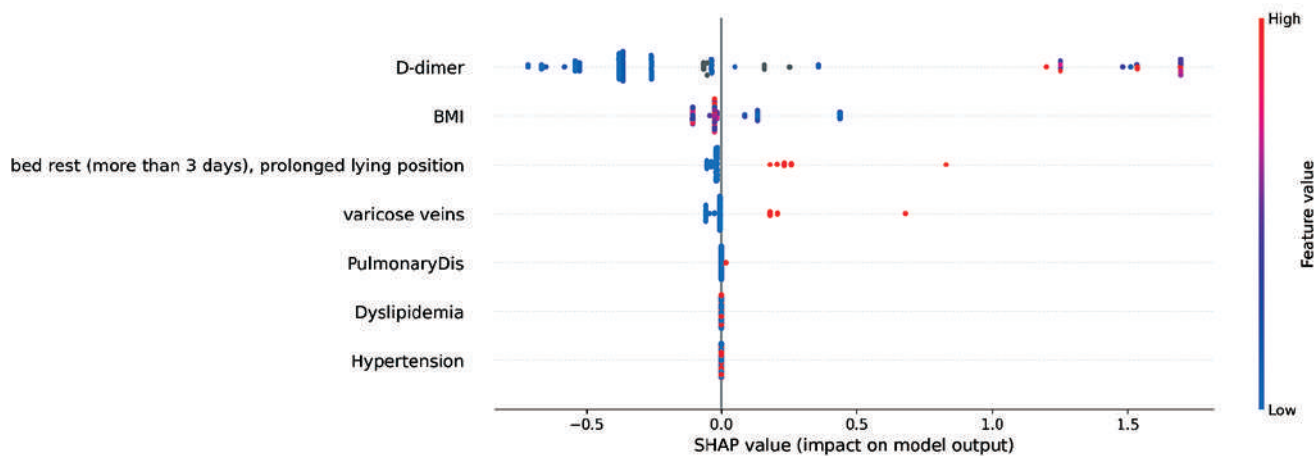


Рис. 3. Shap-значения

Figure 3. Shap values

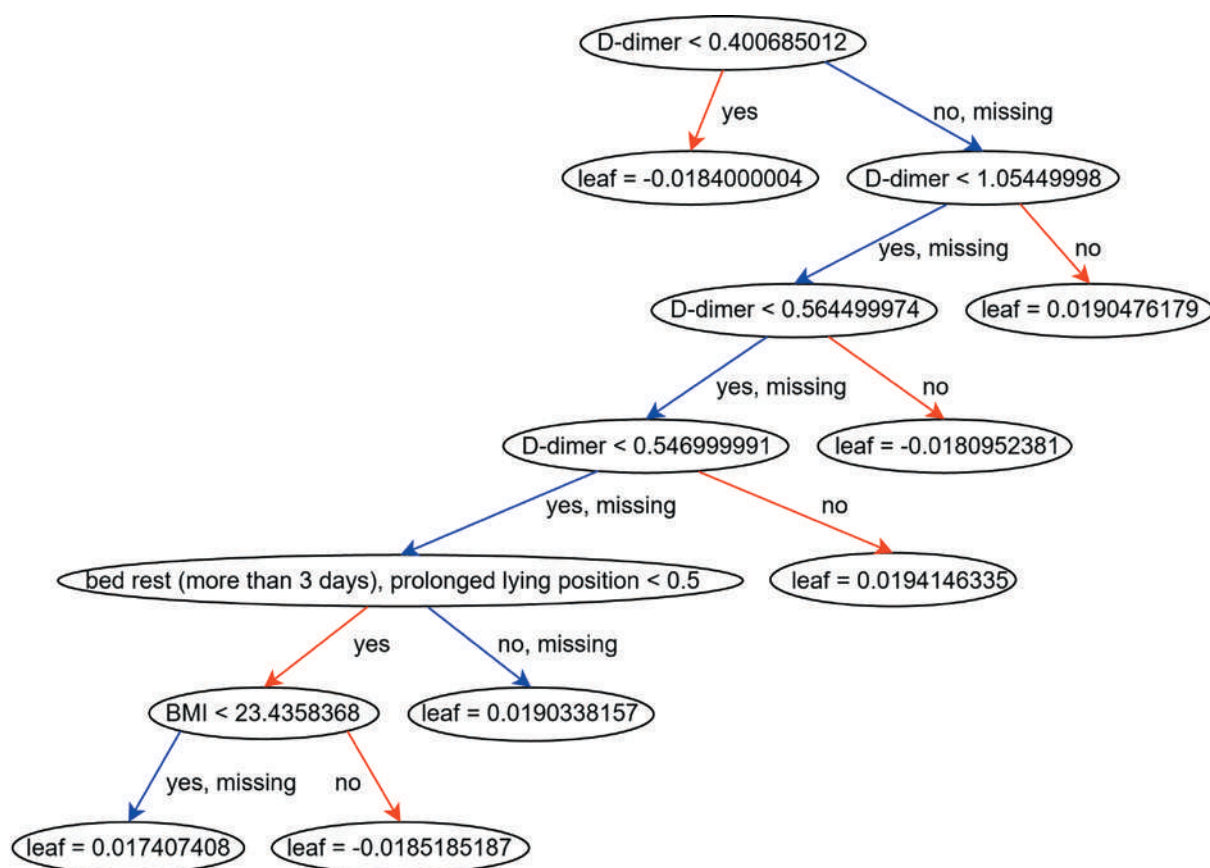


Рис. 4. Дерево результирующей модели

Figure 4. Result model tree

Сравнение с логистической регрессией

Значимые предикторы ВТЭО были определены путем построения однофакторных логистических регрессионных моделей. Независимые предикторы

развития ВТЭО были определены с помощью построения многофакторной логистической регрессионной модели. Данные однофакторного и многофакторного регрессионных анализов представлены в таблице 8.

Таблица 8. Данные многофакторного и однофакторного регрессионных анализов

Table 8. The data from the multi-factor and single-factor regression analyses

Ковариации	Однофакторные модели	Многофакторные модели
	p	p
Постельный режим (более 3 дней), длительное положение лежа	< 0,001*	< 0,001*
Уровень Д-димера, FEU	< 0,001*	0,006*
Уровень тромбоцитов, *10 ⁹ /л	0,006*	0,099
Возраст на момент включения	0,010*	0,067
Лучевая терапия	0,044*	0,047*

Обобщенная характеристика модели по данным ROC-анализа представлена в таблице 9 и на рисунке 5.

Таблица 9. Сводная характеристика логистической регрессии

Table 9. The summary characteristic of the logistic regression

Параметр	Значения [95 % ДИ]
Специфичность	78,6 % [49,2 %; 95,3 %]
Чувствительность	93,5 % [88,4 %; 96,8 %]
Предсказательная ценность положительного результата	52,4 % [29,8 %; 74,3 %]
Предсказательная ценность отрицательного результата	98 % [94,2 %; 99,6 %]
Положительное отношение правдоподобия	12,1 [6,3; 23,4]
Отрицательное отношение правдоподобия	0,2 [0,1; 0,6]

На следующем этапе была проведена валидация модели на проспективной выборке из 100 пациентов

с глиомами ЦНС, проходивших лечение в Национальном медицинском исследовательском центре имени В. А. Алмазова в период с 2022 по 2023 годы. Данные валидации представлены в таблице 10.

Таблица 10. Данные внутренней валидации модели на проспективной выборке

Table 10. The internal validation data of the model on the prospective sample

Параметр	Значения [95 % ДИ]
Специфичность	95 % [87 %; 99 %]
Чувствительность	47 % [23 %; 72 %]
Предсказательная ценность положительного результата	73 % [39 %; 94 %]
Предсказательная ценность отрицательного результата	85 % [77 %; 94 %]
Положительное отношение правдоподобия	10,2 [3,03; 34,35]
Отрицательное отношение правдоподобия	0,56 [0,35; 0,87]

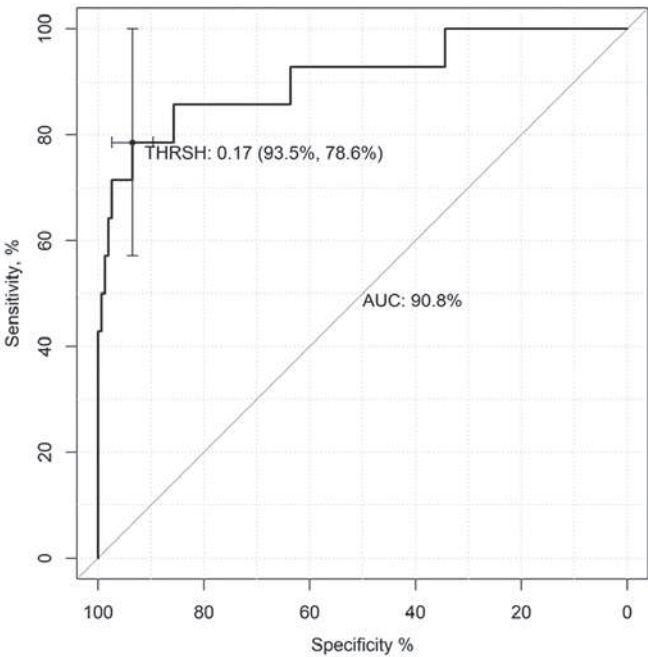


Рис. 5. ROC-кривая для модели многофакторной логистической регрессии

Figure 5. Roc curve of multifactor logistic regression model

СРАВНЕНИЕ МОДЕЛЕЙ

Сравнение моделей в проспективной выборке представлено в таблицах 11–14 и на рисунке 6. Разница в общем количестве пациентов объясняется неполнотой данных.

Таблица 11. Таблица точности диагностического тестирования модели XGBoost для всех данных

Table 11. Diagnostic testing accuracy table of XGBoost model for all data

	Результат +	Результат -	Всего
Тест +	11	5	16
Тест -	7	69	76
Всего	18	74	92

Таблица 12. Таблица точности диагностического тестирования модели XGBoost для скорректированных данных многофакторной логистической регрессии

Table 12. Diagnostic Testing Accuracy Table of XGBoost model for adjusted with multi-factor logistic regression data

	Результат +	Результат -	Всего
Тест +	10	5	15
Тест -	7	60	67
Всего	17	65	82

Таблица 13. Таблица точности диагностического тестирования модели многофакторной логистической регрессии

Table 13. Diagnostic Testing Accuracy Table of multifactor logistic regression model

	Результат +	Результат -	Всего
Тест +	8	3	11
Тест -	9	65	71
Всего	17	65	82

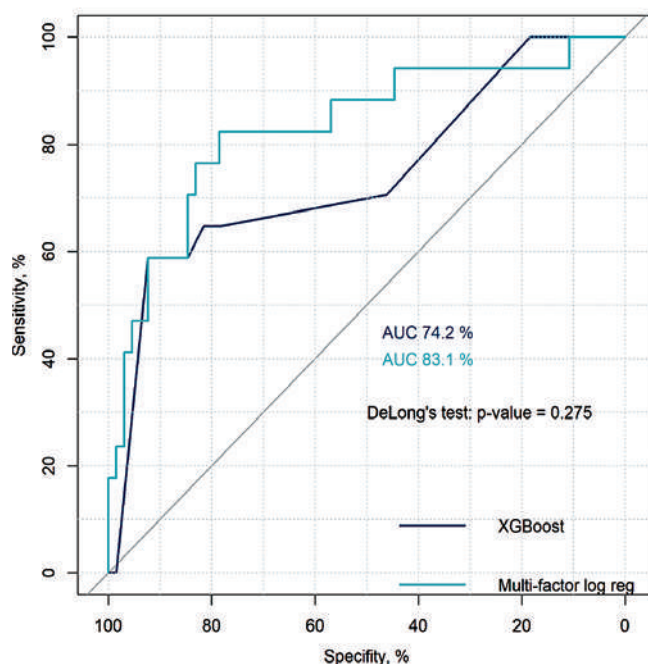


Рис. 6. Сравнение ROC-кривых моделей для проспективных данных

Figure 6. Comparing ROC curves of models on prospective data

Статистически значимых различий в характеристиках моделей не обнаружено.

ЗАКЛЮЧЕНИЕ

В данном исследовании была построена модель XGBoost для прогнозирования развития венозной тромбоземболии (ВТЭО) у пациентов с глиомой. Данная модель была основана на анализе автоматически выбранных параметров. Для построения модели был использован алгоритм XGBoost, оптимизированный на основе полученных параметров.

Модель XGBoost продемонстрировала хорошую производительность по ключевым метрикам, включая специфичность, точность, отзыв и F1-score. Для оценки качества модели использовались анализ ошибок и ROC-кривая, а для иллюстрации важности параметров были получены значения шейпов.

Кроме того, модель была сравнена с моделью логистической регрессии, выявившей значимые предикторы развития ВТЭО. Валидация модели на независимой выборке пациентов подтвердила ее способность к обобщению на внешние данные.

Конфликт интересов / Conflict of interest

Авторы заявили об отсутствии потенциального конфликта интересов. / The authors declare no conflict of interest.

Таблица 14. Сравнение точности диагностических моделей

Table 14. Comparing diagnostic models accuracy

Название	1. Модель XGBoost на первичных проспективных данных	2. Модель XGBoost на скорректированных данных	3. Многофакторная логистическая регрессия на скорректированных данных	Сравнение 1–3 2–3
	Значение [95 % ДИ]	Значение [95 % ДИ]	Значение [95 % ДИ]	P
Apparent prevalence	0,17[0,10; 0,27]	0,18[0,11; 0,28]	0,13[0,07; 0,23]	0,423
True prevalence	0,20[0,12; 0,29]	0,21[0,13; 0,31]	0,21[0,13; 0,31]	–
Sensitivity	0,61[0,36; 0,83]	0,59[0,33; 0,82]	0,47[0,23; 0,72]	0,414
Specificity	0,93[0,85; 0,98]	0,92[0,83; 0,97]	0,95[0,87; 0,99]	0,480
Positive predictive value	0,69[0,41; 0,89]	0,67[0,38; 0,88]	0,73[0,39; 0,94]	0,711
Negative predictive value	0,91[0,82; 0,96]	–	0,87[0,77; 0,94]	0,483

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Nicholson M, Chan N, Bhagirath V, et al. Prevention of venous thromboembolism in 2020 and beyond. J Clin Med. 2020;9(8):2467. Available from: <http://dx.doi.org/10.3390/jcm9082467>
- Kearon C, Akl EA, Ornelas J, et al. Antithrombotic therapy for VTE disease: CHEST guideline and expert panel report. Chest. 2016;149(2):315–52. Available from: <http://dx.doi.org/10.1016/j.chest.2015.11.026>
- Connors JM, Levy JH. COVID-19 and its implications for thrombosis and anticoagulation. Blood. 2020;135:2033–2040. DOI:10.1182/BLOOD.2020006000.
- Xu Q, Lei H, Li X, et al. Machine learning predicts cancer-associated venous thromboembolism using clinically available variables in gastric cancer patients. Heliyon. 2023;9(1):e12681. Available from: <http://dx.doi.org/10.1016/j.heliyon.2022.e12681>.
- He L, Luo L, Hou X, et al. Predicting venous thromboembolism in hospitalized trauma patients: a combination of the Caprini score and data-driven machine learning model. BMC Emerg Med. 2021;21(1):60. Available from: <http://dx.doi.org/10.1186/s12873-021-00447-x>.
- Lin C-C, Chen C-C, Li C-I, et al. Derivation and validation of a clinical prediction model for risks of venous thromboembolism in diabetic and general populations. Medicine(Baltimore)[Internet]. 2021;100(39):e27367. Available from: <http://dx.doi.org/10.1097/MD.00000000000027367>.
- Gerotziafas GT, Papageorgiou L, Salta S, et al. I. Updated clinical models for VTE prediction

in hospitalized medical patients. Thromb Res. 2018;164 Suppl 1:S62–9. Available from: <http://dx.doi.org/10.1016/j.thromres.2018.02.004>.

- Beal EW, Tumin D, Chakedis J, et al. Identification of patients at high risk for post-discharge venous thromboembolism after hepato-pancreato-biliary surgery: which patients benefit from extended thromboprophylaxis? HPB. 2018;20 (7):621–630. DOI:10.1016/j.hpb.2018.01.004.

- Lee E-J, Chang C-H, Wang L-C, et al. Two primary brain tumors, meningioma and glioblastoma multiforme, in opposite hemispheres of the same patient. J Clin Neurosci. 2002;9(5):589–91. Available from: <http://dx.doi.org/10.1054/jocn.2002.1086>.

- Farge D, Frere C, Connors JM, et al. 2022 international clinical practice guidelines for the treatment and prophylaxis of venous thromboembolism in patients with cancer, including patients with COVID-19. Lancet Oncol. 2022;23(7):e334–47. Available from: [http://dx.doi.org/10.1016/S1470-2045\(22\)00160-7](http://dx.doi.org/10.1016/S1470-2045(22)00160-7).

Информация об авторах:

Леонтьев Сергей Сергеевич, студент ФГАОУ ВО «Национальный исследовательский университет ИТМО»;

Симакова Мария Александровна, к.м.н., старший научный сотрудник — руководитель НИГ кардионкологии, НЦМУ «Центр персонализированной медицины», ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России;

Лукинов Виталий Леонидович, к.ф.-м.н., ведущий научный сотрудник лаборатории численного анализа стохастических дифференциальных уравнений, ИВМиМГ СО РАН;

Пищулов Константин Анатольевич, младший научный сотрудник НИГ кардиоонкологии, НЦМУ «Центр персонализированной медицины», ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России;

Абрамян Левон Кароевич, старший специалист по анализу данных, НЦМУ «Центр персонализированной медицины», ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России;

Угольникова Екатерина Алексеевна, инженер лаборатории численного анализа стохастических дифференциальных уравнений, ИВМиМГ СО РАН.

Authors information:

Leontev Sergei S., student, ITMO University;

Simakova Maria A., Candidate of Medical Sciences, Senior Researcher — Head of cardio-oncology Research Group, World-Class Research Centre for Personalized Medicine, Almazov National Medical Research Centre;

Lukinov Vitaliy L., Lead researcher (Laboratory of numerical analysis of SDE), ICMMG SB RAS;

Pishchulov Konstantin A., Junior Research, Cardio-Oncology Research Group, World-Class Research Centre for Personalized Medicine, Almazov National Medical Research Centre;

Abramyan Levon K., Senior Data Scientist, World-Class Research Centre for Personalized Medicine, Almazov National Medical Research Centre;

Ugolnikova Ekaterina A., Engineer (Laboratory of numerical analysis of SDE), ICMMG SB RAS.