

УДК 004.8

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В РЕШЕНИИ ПРОБЛЕМЫ ОНКОПРОФИЛАКТИКИ: РЕТРОСПЕКТИВНОЕ ИССЛЕДОВАНИЕ

П. А. Филоненко¹ [0000-0002-6295-4470], В. Н. Кох² [0000-0002-9257-0259],

П. Д. Блинов³ [0009-0000-7583-7655]

^{1, 3}Sber AI Lab, г. Москва, Россия

²Sber AI, г. Москва, Россия

¹petr-filonenko@mail.ru, ²kokh.v.n@sber.ru, ³blinov.p.d@sber.ru

Аннотация

Исследована возможность эффективного решения задачи популяционной онкопрофилактики с помощью методов искусственного интеллекта (ИИ), прогнозирующих риск злокачественных новообразований (ЗНО) на основе минимального набора данных из электронной медицинской карты (ЭМК) – кодов медицинских диагнозов и услуг. Для решения поставленной задачи рассмотрен широкий спектр современных подходов, включающих методы классического машинного обучения, анализа выживаемости, глубокого обучения и больших языковых моделей (LLM). Численные эксперименты показали, что наилучшей способностью ранжирования пациентов по уровню риска ЗНО обладает градиентный бустинг, использующий модели анализа выживаемости в качестве дополнительных предикторов, что позволяет учитывать как популяционные, так и индивидуальные факторы риска ЗНО. Из данных ЭМК были сконструированы предикторы, включающие демографические характеристики, паттерны обращений за медицинской помощью и клинические маркеры. Это решение было протестировано в ретроспективных экспериментах под контролем профильных врачей-онкологов. В ретроспективном эксперименте с участием более 1.9 млн пациентов установлено, что в группу риска попадает до 5.4 раза больше пациентов с ЗНО при том же уровне медицинских обследований. Предложенный метод представляет собой масштабируемое решение, использующее исключительно коды диагнозов и услуг, не требующее специализированной инфраструктуры и интегрируемое в процесс онконастороженности, что делает его применимым для решения задач популяционной онкопрофилактики.

Ключевые слова: ИИ в медицине, популяционная онкопрофилактика, ретроспективные эксперименты.

ВВЕДЕНИЕ

Злокачественные новообразования остаются одной из ведущих причин смертности в мире, при этом эффективность их раннего выявления напрямую связана с прогнозом заболевания. В Российской Федерации в 2023 г. заболеваемость ЗНО составила около 461 новых случая на каждые 100 тыс. населения [1], что подчеркивает критическую важность развития эффективных методов популяционной онкопрофилактики.

Действующие программы профилактики демонстрируют ограниченную эффективность [2], а традиционные методы остаются дорогостоящими, трудозатратными и практически неприменимыми для масштабного популяционного применения [3], что создает разрыв между потребностью в раннем выявлении ЗНО и возможностями системы здравоохранения. Внедрение электронных медицинских карт (ЭМК) в сочетании с развитием методов искусственного интеллекта (ИИ) открывает новые возможности для автоматизированного анализа медицинских данных. Однако такие ИИ-решения для прогнозирования риска ЗНО требуют либо специализированных данных (например, биомаркеров [4], семейного анамнеза [5], генетических данных [6] и др.), либо специализированной инфраструктуры для развертывания вычислительно сложных решений, что существенно ограничивает их практическое применение в задачах массовой профилактики.

Исследование направлено на поиск, разработку и валидацию метода, способного эффективно решать задачу популяционной онкопрофилактики, основанной исключительно на кодах медицинских диагнозов и услуг, доступных в любой медицинской организации. Для этого: 1) проведен сравнительный анализ широкого спектра различных ИИ-решений и проверена их эффективность в ранжировании пациентов по уровню риска ЗНО; 2) под контролем профильных врачей-онкологов лучшее ИИ-решение прошло ретроспективную валидацию на предмет эффективного применения в задаче популяционной онкопрофилактики.

МАТЕРИАЛЫ И МЕТОДЫ

1. Постановка задачи

Задача прогнозирования риска ЗНО формулируется как бинарная классификация (рис. 1): в момент времени t_{pred} требуется оценить вероятность ЗНО (C00-C97 по МКБ-10) в следующие 12 месяцев, используя данные ЭМК за предшествующие N месяцев. Значение N выбрано равным 24, чтобы большинство ЭМК не были пустыми, а решение было доступно для массового применения. Такая постановка задачи является консistentной целям федерального проекта «Борьба с онкологическими заболеваниями» на 2025-2030 гг. Целевая переменная определяется как:

- **target = 1**, если выявлено ЗНО в период $[t_{\text{pred}}, t_{\text{pred}} + 12M]$;
- **target = 0** в противном случае.

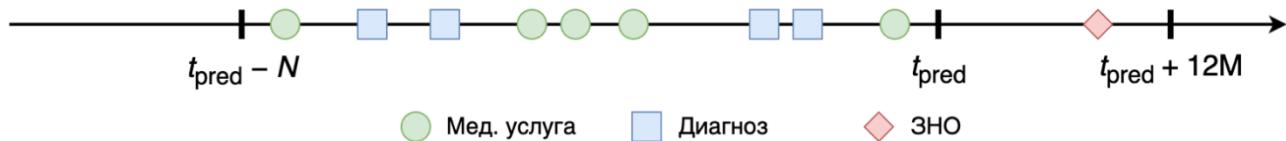


Рис. 1. Постановка задачи

Такая постановка позволяет решать задачу ранжирования пациентов, формируя группы риска для приоритетного прохождения обследований. Для этого используем метрику Average Precision (AP) [7], которая максимизирует долю верных ответов в верхней части списка ($\text{Precision@TOP} \rightarrow \text{MAX}$) и стабильна даже при экстремальном дисбалансе классов, что критически важно в задаче прогнозирования риска ЗНО. Для сравнения с другими решениями, известными в литературе, приведем значения ROC AUC.

2. Методы решения

Для решения поставленной задачи применимы самые разные методы, каждый из которых имеет схожий пайплайн: извлечение признаков из ЭМК \rightarrow ИИ-модель \rightarrow оценка вероятности $P(\text{ЗНО} | \text{ЭМК})$ или функции выживаемости $S(t | \text{ЭМК})$.

В данной работе мы рассматриваем следующие виды решений:

- 1) **Методы машинного обучения:** логистическая регрессия, случайный лес, градиентный бустинг (GBM);
- 2) **Модели выживаемости:** AFT-модель, случайные леса выживаемости, DeepHit [8], Deep Survival Machines [9];
- 3) **Глубокое обучение:** CoLES (*дообучение*) [10], BERT (*претрейн на медицинских текстах*) [11], Longformer (*претрейн на медицинских текстах*) [12];
- 4) **LLM-энкодеры:** DeepSeek-R1-Distill-Qwen-1.5B (*эмбеддинг последнего скрытого слоя*), Qwen3-Embedding-0.6B, GigaChat-Embeddings;
- 5) **LLM-конвейер:** LLM-суммаризация (*DeepSeek-R1*) → LLM-энкодер (*Qwen3-Emb*) → ИИ-модель (*GBM, DeepHit, LoRA-адаптер*);
- 6) **Ансамбль GBM и моделей выживаемости**¹ (*популяционные риски – оценки Каплана-Майера для каждого пола, индивидуальные риски – AFT-модель*): $P(\text{ЗНО} | \text{ЭМК}) = \text{GBM}(\text{ML-предикторы} \oplus \text{Предикторы-выживаемости})$, где \oplus – это конкатенация.

ЭКСПЕРИМЕНТЫ

1. Сравнение методов

Методы сравнивались на амбулаторных данных 175411 пациентов (18+) за период 2017-2021 гг. Для корректности выводов была проведена стратификация пациентов по полу и возрасту с проверкой однородности выборок [13]. Проверялись многомерная гипотеза однородности демографических характеристик $H_0: F_{\text{Train}}(x) = F_{\text{Validate}}(x) = F_{\text{Test}}(x)$ и одномерная гипотеза однородности времени до ЗНО $H_0: S_i(t) = S_j(t) \forall i \neq j, i, j \in \{\text{Train}, \text{Validate}, \text{Test}\}$. Минимальное p -value > 0.05, что подтверждает отсутствие систематических различий между выборками.

Модели были обучены на выборке Train (54%), гиперпараметры оптимизированы с помощью Optuna на выборке Validate (23%). В табл. 1 показаны результаты сравнения методов на выборке Test (23%) с указанием 95%-х доверительных интервалов.

¹ <https://github.com/sb-ai-lab/Can-SAVE>

Из полученных результатов видно, что ансамбль градиентного бустинга и моделей выживаемости превосходит другие решения с AP 22.8%. Даже несмотря на то что у других решений значения ROC AUC выше, ансамбль обладает лучшей способностью ранжировать пациентов по уровню риска ЗНО, чем остальные методы. Покажем далее, на какие факторы опирается найденное ИИ-решение.

Табл. 1. Сравнительный анализ методов на тестовой выборке (95% ДИ)

Метод	Average Precision, %	ROC AUC, %
Логистическая регрессия	10.4 ± 1.3	83.4 ± 0.7
Случайный лес	10.2 ± 0.5	83.3 ± 0.6
Градиентный бустинг (GBM)	16.0 ± 1.8	78.6 ± 1.3
Ансамбль GBM и моделей выживаемости	22.8 ± 2.7	83.7 ± 1.7
AFT-модель	11.7 ± 1.7	84.8 ± 2.2
Случайные леса выживаемости	7.4 ± 0.3	78.6 ± 0.5
DeepHit	10.2 ± 2.5	86.4 ± 1.6
Deep Survival Machines	10.1 ± 0.5	82.3 ± 0.6
CoLES (fine-tuned)	10.3 ± 0.2	81.3 ± 0.2
BERT → GRU	15.1 ± 2.6	84.9 ± 0.8
Longformer → GBM	9.3 ± 0.2	77.7 ± 0.5
Qwen3-Emb → GBM	15.1 ± 0.9	86.9 ± 0.3
Qwen3-Emb → DeepHit	18.6 ± 0.7	88.5 ± 0.3
DeepSeek-R1 → GBM	16.4 ± 1.0	87.3 ± 0.5
GigaChat → GBM	18.5 ± 0.2	89.6 ± 0.1
DeepSeek-R1 → Qwen3-Emb → GBM	17.6 ± 1.0	88.1 ± 0.5
DeepSeek-R1 → Qwen3-Emb → DeepHit	17.4 ± 0.4	89.5 ± 0.2
DeepSeek-R1 → Qwen3-Emb → LoRA	19.3 ± 0.4	90.1 ± 0.2

2. Важность признаков

Для найденного ансамбля градиентного бустинга и моделей выживаемости проанализируем важность входящих предикторов. Для этого вычислим Feature Importance (*как часто предикторы используются при построении деревьев решений*) и Permutation Importance (*как сильно случайные перестановки значений предикторов влияют на целевую метрику*). В табл. 2 суммарные значения показателей важности предикторов по каждой группе признаков. Из полученных результатов очевидно, что модели выживаемости вносят ключевой вклад в прогностическую силу найденного ИИ-решения, поскольку их суммарные значения Feature Importance (39.692) и Permutation Importance (6.594) максимальны.

Табл. 2. Важность групп предикторов найденного ИИ-решения

Группа предикторов	Feature Importance	Permutation Importance
Социально-демографические признаки (<i>пол, возраст</i>)	$\Sigma = 21.792$	$\Sigma = 2.290$
Паттерны визитов (<i>время с первого визита; месяц визита; доля диагнозов от числа визитов</i>)	$\Sigma = 21.562$	$\Sigma = 6.322$
Клинические маркеры (<i>частотность диагнозов D37-D48, O20-O29; время с первого диагноза D00-D48, I00-I99, Q00-Q99; частотность медицинских услуг по иммунной системе</i>)	$\Sigma = 16.954$	$\Sigma = 1.542$
Модели выживаемости (<i>оценки Каплана-Мейера для мужчин (M), женщин (Ж) и M+Ж; AFT-модель; приращение риска оценок Каплана-Мейера через 12 месяцев; приращение риска AFT-модели через 12 месяцев</i>)	$\Sigma = 39.692$	$\Sigma = 6.594$

3. Ретроспективные эксперименты

Предлагаемое ИИ-решение было протестировано в ретроспективных экспериментах под контролем профильных врачей-онкологов в условиях, приближенных к реальным. Для этого под руководством профильных врачей онкологов в 5 регионах Российской Федерации был проведен эксперимент, демонстрирующий способность формировать группы риска пациентов с ЗНО в сравнении с контрольной группой на основе диспансеризации взрослого населения РФ.

Эксперимент состоял из следующих действий: 1) оценить риск ЗНО каждого пациента в выборке с помощью ИИ-решения; 2) сформировать группы риска 1%, 3%, 5% от численности выборки (такой дополнительный поток пациентов не перегрузит систему здравоохранения); 3) передать списки групп риска контролирующем врачам-онкологам для верификации числа верных ЗНО в каждой группе; 4) сравнить с аналогичными результатами контрольной группы. В ретро-эксперименте анализировались ЭМК из 5 регионов РФ численностью более 1,9 млн пациентов (мужчин 43%, женщин 57%), охватывающих периоды прогнозирования 2018-2024 гг. Пациенты были включены в исследование, если были не младше 18 лет на момент t_{pred} и у них отсутствовало ЗНО в анамнезе жизни. Результаты данного эксперимента представлены в табл. 3.

Табл. 3. Результаты сравнения случаев ЗНО в группах риска 1%-5%

Группа риска, количество пациентов		Контрольная группа, ЗНО	ИИ-решение, ЗНО	Прирост
%	чел.			
Регион 1 (численность: 92 985)				
1%	930	9	41	4,4x
3%	2 790	28	94	3,4x
5%	4 649	46	133	2,9x
Регион 2 (численность: 112 620)				
1%	1 126	11	60	5,3x
3%	3 378	34	117	3,5x
5%	5 631	56	178	3,2x

Регион 3 (численность: 165 355)				
1%	1 653	15	74	4,9x
3%	4 960	35	97	2,8x
5%	8 267	58	117	2,0x
Регион 4 (численность: 651 697)				
1%	6 516	85	315	3,7x
3%	19 550	254	658	2,6x
5%	32 584	424	933	2,2x
Регион 5 (численность: 889 293)				
1%	8 893	80	434	5,4x
3%	26 679	240	781	3,3x
5%	44 465	400	1103	2,8x

Как видно из представленных результатов, ИИ-решение способно эффективно формировать группы риска ЗНО, превышая от 2.0 до 5.4 раза пациентов с ЗНО в каждой группе риска (1%-5%) в сравнении с текущим состоянием процесса выявления ЗНО. Это значит, что если выполнить ИИ-сканирование пациентов целого региона, то работа с группой риска размером от 1% до 5% от численности региона способна повысить показатели выявляемости ЗНО, не перегрузив систему здравоохранения. Таким образом, полученные результаты ретроспективного эксперимента подтверждают, что методы ИИ даже на минимально доступных данных способны повысить качество результатов популяционной онкопрофилактики.

ЗАКЛЮЧЕНИЕ

Найденное решение на основе ИИ доказывает возможность эффективного решения задачи популяционной онкопрофилактики с помощью ИИ-методов с использованием исключительно кодов медицинских диагнозов и услуг. Ансамбль моделей выживаемости и градиентного бустинга превосходит другие рассмотренные подходы при минимальных требованиях к данным и вычислительным ресурсам.

Результаты ретроспективной валидации на более чем 1.9 млн пациентов под контролем профильных онкологов подтвердили клиническую значимость

найденного решения и его способность существенно повысить эффективность популяционной онкопрофилактики. Представленное решение естественным образом встраивается в существующий медицинский процесс онконастороженности, направляя пациентов из группы риска к специалистам первичного звена для принятия решения о дополнительном обследовании и направлении к врачу-онкологу.

Минимальные требования к инфраструктуре делают решение доступным для широкого внедрения в различных системах здравоохранения, что открывает новые возможности для раннего выявления онкологических заболеваний и снижения смертности от ЗНО.

СПИСОК ЛИТЕРАТУРЫ

1. Каприн А. Д., Старинский В. В., Шахзадова А. О. Злокачественные новообразования в России в 2023 году (заболеваемость и смертность) / Под ред. А. Д. Каприна, В. В. Старинского, А. О. Шахзадовой. М.: МНИОИ им. П. А. Герцена — филиал ФГБУ «НМИЦ радиологии» Минздрава России, 2024. 276 с. ISBN 978-5-85502-298-8.
2. Cenin D. R., Tinmouth J., Naber S. K., Khalaf N., Rabeneck L., Tinmouth J. M., Earle C. C., Hilsden R. J., Leddin D., Rostom A., Issaka R. B., Heitman S. J., Lansdorp-Vogelaar I. Calculation of stop ages for colorectal cancer screening based on comorbidities and screening history // Clinical Gastroenterology and Hepatology. 2021. Vol. 19, No. 3. P. 547–555. <https://doi.org/10.1016/j.cgh.2020.05.038>
3. Ratushnyak S., Hoogendoorn M., van Baal P. H. M. Cost-effectiveness of cancer screening: health and costs in life years gained // American Journal of Preventive Medicine. 2019. Vol. 57, No. 6. P. 792–799. <https://doi.org/10.1016/j.amepre.2019.07.027>
4. Alexander M., Burbury K. A systematic review of biomarkers for the prediction of thromboembolism in lung cancer — Results, practical issues and proposed strategies for future risk prediction models // Thrombosis Research. 2016. Vol. 148. P. 63–69. <https://doi.org/10.1016/j.thromres.2016.10.020>
5. Jacobs M. F. Predicting cancer risk based on family history // eLife. 2021. Vol. 10. e73380. <https://doi.org/10.7554/eLife.73380>

6. *Wang X., Oldani M. J., Zhao X., Huang X., Qian D.* A review of cancer risk prediction models with genetic variants // *Cancer Informatics*. 2014. Vol. 13, Suppl. 2. P. 19–28. <https://doi.org/10.4137/CIN.S13788>
7. *Zhu M.* Recall, precision and average precision // *Technical Report*. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, 2004. 6 p.
8. *Lee C., Zame W. R., Yoon J., van der Schaar M.* DeepHit: A deep learning approach to survival analysis with competing risks // *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018. Vol. 32, No. 1. P. 2314–2321. <https://doi.org/10.1609/aaai.v32i1.11842>
9. *Nagpal C., Li X., Dubrawski A.* Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks // *IEEE Journal of Biomedical and Health Informatics*. 2021. Vol. 25, No. 8. P. 3163–3175. <https://doi.org/10.1109/JBHI.2021.3052441>
10. *Babaev D., Ovsov N., Kireev I., Ivanova M., Gusev G., Nazarov I., Tuzhilin A.* CoLES: Contrastive learning for event sequences with self-supervision // *Proceedings of the 2022 International Conference on Management of Data (SIGMOD '22)*. New York, NY, USA: ACM, 2022. P. 1190–1199. <https://doi.org/10.1145/3514221.3526129>
11. *Blinov P., Kokh V.* Medical profile model: scientific and practical applications in healthcare // *IEEE Journal of Biomedical and Health Informatics*. 2023. Vol. 28, No. 1. P. 450–458. <https://doi.org/10.1109/JBHI.2023.3295631>
12. *Yalunin A., Nesterov A., Umerenkov D.* RuBioRoBERTa: a pre-trained biomedical language model for Russian language biomedical text mining // *arXiv preprint*. 2022. arXiv:2204.03951. <https://doi.org/10.48550/arXiv.2204.03951>
13. *Philonenko P., Postovalov S.* The new robust two-sample test for randomly right-censored data // *Journal of Statistical Computation and Simulation*. 2019. Vol. 89, No. 8. P. 1357–1375. <https://doi.org/10.1080/00949655.2019.1577858>

AI IN CANCER PREVENTION: A RETROSPECTIVE STUDY

P. A. Philonenko¹ [0000-0002-6295-4470], V. N. Kokh² [0000-0002-9257-0259],
P. D. Blinov³ [0009-0000-7583-7655]

^{1, 3}Sber AI Lab, Moscow, Russia

²Sber AI, Moscow, Russia

¹petr-filonenko@mail.ru, ²kokh.v.n@sber.ru, ³blinov.p.d@sber.ru

Abstract

This study investigates the feasibility of effectively solving population-scale cancer screening problems using artificial intelligence (AI) methods that predict malignant neoplasm risk based on minimal electronic health record (EHR) data – medical diagnosis and service codes. To address the formulated problem, we considered a broad spectrum of modern approaches, including classical machine learning methods, survival analysis, deep learning, and large language models (LLMs). Numerical experiments demonstrated that gradient boosting using survival analysis models as additional predictors possesses the best ability to rank patients by cancer risk level, enabling consideration of both population-level and individual risk factors for malignant neoplasms. Predictors constructed from EHR data include demographic characteristics, healthcare utilization patterns, and clinical markers. This solution was tested in retrospective experiments under the supervision of specialized oncologists. In the retrospective experiment involving more than 1.9 million patients, we established that the risk group captures up to 5.4 times more patients with cancer at the same level of medical examinations. The investigated method represents a scalable solution using exclusively diagnosis and service codes, requiring no specialized infrastructure and integrable into oncological vigilance processes, making it applicable for population-scale cancer screening.

Keywords: AI in medicine, cancer prevention, retrospective experiments.

REFERENCES

1. Kaprin A. D., Starinskiy V. V., Shakhzadova A. O. Malignant neoplasms in Russia in 2023 (incidence and mortality) / Ed. by A. D. Kaprin, V. V. Starinskiy, A. O. Shakhzadova. Moscow: P. A. Herzen Moscow Oncology Research Institute — Branch of

the National Medical Research Radiological Centre of the Ministry of Health of Russia, 2024. 276 p. ISBN 978-5-85502-298-8. (In Russian).

2. *Cenin D. R., Tinmouth J., Naber S. K., Khalaf N., Rabeneck L., Tinmouth J. M., Earle C. C., Hilsden R. J., Leddin D., Rostom A., Issaka R. B., Heitman S. J., Lansdorp-Vogelaar I.* Calculation of stop ages for colorectal cancer screening based on comorbidities and screening history. *Clinical Gastroenterology and Hepatology*, 2021, vol. 19, no. 3, pp. 547–555. <https://doi.org/10.1016/j.cgh.2020.05.038>
3. *Ratushnyak S., Hoogendoorn M., van Baal P. H. M.* Cost-effectiveness of cancer screening: health and costs in life years gained. *American Journal of Preventive Medicine*, 2019, vol. 57, no. 6, pp. 792–799. <https://doi.org/10.1016/j.amepre.2019.07.027>
4. *Alexander M., Burbury K.* A systematic review of biomarkers for the prediction of thromboembolism in lung cancer — Results, practical issues and proposed strategies for future risk prediction models. *Thrombosis Research*, 2016, vol. 148, pp. 63–69. <https://doi.org/10.1016/j.thromres.2016.10.020>
5. *Jacobs M. F.* Predicting cancer risk based on family history. *eLife*, 2021, vol. 10, e73380. <https://doi.org/10.7554/eLife.73380>
6. *Wang X., Oldani M. J., Zhao X., Huang X., Qian D.* A review of cancer risk prediction models with genetic variants. *Cancer Informatics*, 2014, vol. 13, suppl. 2, pp. 19–28. <https://doi.org/10.4137/CIN.S13788>
7. *Zhu M.* Recall, precision and average precision. Technical Report, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, 2004, 6 p.
8. *Lee C., Zame W. R., Yoon J., van der Schaar M.* DeepHit: A deep learning approach to survival analysis with competing risks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, vol. 32, no. 1, pp. 2314–2321. <https://doi.org/10.1609/aaai.v32i1.11842>
9. *Nagpal C., Li X., Dubrawski A.* Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks. *IEEE Journal of Biomedical and Health Informatics*, 2021, vol. 25, no. 8, pp. 3163–3175. <https://doi.org/10.1109/JBHI.2021.3052441>
10. *Babaev D., Ovsov N., Kireev I., Ivanova M., Gusev G., Nazarov I., Tuzhilin A.* CoLES: Contrastive learning for event sequences with self-supervision. *Proceedings of*

the 2022 International Conference on Management of Data (SIGMOD '22), New York, NY, USA, ACM, 2022, pp. 1190–1199. <https://doi.org/10.1145/3514221.3526129>

11. *Blinov P., Kokh V.* Medical profile model: scientific and practical applications in healthcare. *IEEE Journal of Biomedical and Health Informatics*, 2023, vol. 28, no. 1, pp. 450–458. <https://doi.org/10.1109/JBHI.2023.3295631>

12. *Yalunin A., Nesterov A., Umerenkov D.* RuBioRoBERTa: a pre-trained biomedical language model for Russian language biomedical text mining. *arXiv preprint*, 2022, arXiv:2204.03951. <https://doi.org/10.48550/arXiv.2204.03951>

13. *Philonenko P., Postovalov S.* The new robust two-sample test for randomly right-censored data. *Journal of Statistical Computation and Simulation*, 2019, vol. 89, no. 8, pp. 1357–1375. <https://doi.org/10.1080/00949655.2019.1577858>

СВЕДЕНИЯ ОБ АВТОРАХ



ФИЛОНЕНКО Петр Александрович – окончил Новосибирский государственный технический университет (ныне – НГТУ-НЭТИ). В 2018 году защитил диссертацию на соискание кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики. В настоящее время работает руководителем направления по исследованию данных в Sber AI Lab. Область научных интересов: прогнозирование риска заболеваний, теория выживаемости, проверка статистических гипотез, машинное обучение.

Petr Aleksandrovich PHILONENKO – graduated from Novosibirsk State Technical University (now NSTU-NETI). In 2018, he defended his dissertation for the degree of Candidate of Technical Sciences in specialty 05.13.17 Theoretical foundations of computer science. Currently, he works as a Senior Data Scientist at the Sber AI Lab. Research interests: risk prediction, survival analysis, hypothesis testing, machine learning.

email: petr-filonenko@mail.ru

ORCID: 0000-0002-6295-4470



КОХ Владимир Николаевич – окончил Сибирский государственный медицинский университет. В настоящее время работает исполнительным директором по анализу данных в Sber AI. Область научных интересов: ИИ в медицине, R&D на основе искусственного интеллекта в здравоохранении.

Vladimir Nikolaevich KOKH – graduated from Siberian State Medical University. Currently, he works as an Executive Director of Data Analysis at Sber AI. Research interests: AI in medicine, artificial intelligence-based R&D in healthcare.

email: kokh.v.n@sber.ru

ORCID: 0000-0002-9257-0259



БЛИНОВ Павел Дмитриевич – окончил Вятский государственный университет по специальности прикладная математика и информатика, закончил в 2011 году. В 2016 году защитил кандидатскую диссертацию на соискание учёной степени кандидата технических наук по специальности 05.13.17 "Теоретические основы информатики". В настоящее время работает исполнительным директором по исследованию данных в Sber AI Lab. Область научных интересов: обработка естественного языка, методы машинного обучения, интеллектуальный анализ медицинских данных.

Pavel Dmitrievich BLINOV – graduated with a degree in Applied Mathematics and Computer Science from Vyatka State University in 2011. In 2016, he defended his dissertation and was awarded the degree of Candidate of Sciences (Ph.D. equivalent) in "Theoretical Foundations of Computer Science" (specialty code 05.13.17). He is currently the Executive Director of Data Science at Sber AI Lab. His research interests include natural language processing, machine learning methods, and medical knowledge mining.

email: blinov.p.d@sber.ru

ORCID: 0009-0000-7583-7655

Материал поступил в редакцию 10 октября 2025 года