

Научная статья

УДК 616-091

doi: <https://doi.org/10.19163/1994-9480-2024-21-1-3-8>

Искусственный интеллект, машинное обучение и нейронные сети в морфологии

Н.А. Дорофеев^{1,2}✉, Г.Л. Снигур^{1,2}, М.Ю. Фролов^{1,2}, А.В. Смирнов^{1,2}, Д.А. Сасов³, А.В. Зубков^{1,3}¹ Волгоградский государственный медицинский университет, Волгоград, Россия² Волгоградский медицинский научный центр, Волгоград, Россия³ Волгоградский государственный технический университет, Волгоград, Россия

Аннотация. Наука о данных – область информатики, занимающаяся прикладным анализом данных – зарекомендовала себя как ценный инструмент во многих областях жизни человека, в том числе медицине в целом и морфологии в частности. Чтобы в полной мере понять и раскрыть потенциал входящих в него методов, необходимо иметь представление об их сущности, возможностях и ограничениях. В статье приводится краткая характеристика ключевых понятий и актуальных достижений в области искусственного интеллекта, машинного обучения и нейронных сетей в морфологии.

Ключевые слова: Data Science, наука о данных, морфология, искусственный интеллект, машинное обучение, нейронные сети

Original article

doi: <https://doi.org/10.19163/1994-9480-2024-21-1-3-8>

Artificial intelligence, machine learning and neural networks in morphology

N.A. Dorofeev^{1,2}✉, G.L. Snigur^{1,2}, M.Yu. Frolov^{1,2}, A.V. Smirnov^{1,2}, D.A. Sasov³, A.V. Zubkov^{1,3}¹ Volgograd State Medical University, Volgograd, Russia² Volgograd Medical Research Center, Volgograd, Russia³ Volgograd State Technical University, Volgograd, Russia

Abstract. Data science, the field of computer science dealing with applied data analysis, has established itself as a valuable tool in many areas of human life, including medicine in general and morphology in particular. In order to fully understand and unlock the potential of the methods included in it, it is necessary to have an idea of their essence, capabilities and limitations. The article provides a brief description of the key concepts and current achievements in the field of artificial intelligence, machine learning and neural networks in morphology.

Keywords: Data Science, data science, morphology, artificial intelligence, machine learning, neural networks

ВВЕДЕНИЕ

Цифровизация медицины в целом и морфологии в частности в последние десятилетия развивается быстрыми темпами. Благодаря совершенствованию вычислительной техники на бюджетном ноутбуке можно производить операции, доступные ранее лишь крупным компаниям и исследовательским центрам. Возможность бесплатного использования целого ряда передовых технологий, адаптации их под собственные нужды позволила заниматься разработкой и внедрением инновационных продуктов небольшим командам. *Data science (DS)*, русск. Наука о данных) – область информатики, занимающаяся изучением и прикладным применением методов поиска, обработки и визуализации данных – заняла прочное место в повседневной жизни современного человека, помогая автоматизировать не только рутинные, но и более сложные манипуляции, ранее недоступные для выполнения машиной.

С *DS* связаны такие понятия как искусственный интеллект (ИИ), машинное обучение (МО) и нейронные сети (НС). ИИ в прикладном смысле – раздел информатики, направленный на создание вычислительных систем, программ, имитирующих разумные действия и рассуждения [1, 2, 3, 4, 5]. Данный термин не является новым, так как активно используется уже на протяжении нескольких десятилетий, а технологии, которые за ним стоят, нашли практическое применение в большинстве сфер жизни человека. Однако ИИ – очень широкое понятие, включающее в себя все – от простых математических расчетов и алгоритмов, наиболее часто встречающихся в реальности, до искусственного разума, описанного фантастами.

Также в этот перечень входит МО – набор методов решения задач путем обучения алгоритма на схожих задачах [1, 3, 4, 5]. Результатом так называемого обучения – предоставления набора данных с необходимыми примерами и применения к ним определенной

методики расчетов – является математическая модель, способная с известной эффективностью решать задачу какого-либо класса. Среди этих задач наиболее часто встречаются классификация, когда входные данные делятся на известные классы; кластеризация, отличающаяся от предыдущей заранее неизвестным перечнем категорий, рассчитываемым непосредственно в процессе обучения; регрессия, применяемая для прогнозирования будущих значений; снижение размерности данных и поиск ассоциаций.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

Для решения этих типов задач разработано огромное множество математических алгоритмов, которые подразделяются на две большие категории – обучение «с учителем» (классификация, регрессия) и «без учителя» (кластеризация, снижение размерности и поиск ассоциаций) [2, 4]. В первом случае набор данных для построения модели МО, содержащий примеры решаемой задачи, включает в себя как признаки, по которым производится расчёт, так и «правильный ответ», результат, который должен быть получен в итоге. Во втором случае имеются только входные данные, зачастую потому, что необходимый результат неизвестен в принципе, и с помощью алгоритма вычисляется возможный вариант решения задачи, производится поиск неясных признаков.

Методы МО включают в себя как простые, понятные, но слабо применимые на практике, такие как линейная регрессия, так и более сложные, показывающие хорошие и отличные результаты, но также имеющие свои ограничения – алгоритмы на основе «деревьев решений», НС (подсемейство моделей МО) и другие.

Для оценки эффективности и отбора моделей используются несколько основных общепринятых метрик: *accuracy* (русс. точность), *precision* (русс. точность), *recall* (англ. отзыв) и *AUC* (англ. *the Area Under a Curve*, русс. площадь под кривой). Они по-разному связаны с положительными (англ. *Positive, P*) и отрицательными (англ. *Negative, N*) истинными (англ. *True, T*) и ложными (англ. *False, F*) результатами расчётов.

Accuracy представляет из себя долю правильных ответов (как *TP*, так и *TN*), что является интуитивно понятной оценкой, но она становится мало полезной в задачах, когда данные разделены на группы (классы) неравномерно (один или несколько из них встречаются значительно реже либо чаще других).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision рассчитывается как доля *TP* среди всех ответов, названных моделью положительными (как правильно, так и ложно).

$$Precision = \frac{TP}{TP + FP}$$

Третья мера – *recall* – является долей *TP* ответов, найденных моделью среди всех релевантных элементов, то есть насколько хорошо *TP* отличаются от *FN*. Во всех предыдущих случаях чем выше точность модели, тем ближе значение метрики стремится к единице (ста процентам).

$$Recall = \frac{TP}{TP + FN}$$

Последняя указанная мера – *AUC* – представлена так называемой площадью под кривой ошибок (англ. *ROC-curve, Receiver Operating Characteristic curve*), которая строится в координатах доли *TP*, являющихся значением *recall*, и доли *FP* (рис. 1). Ее оценка похожа на предыдущие тем, что чем эффективнее модель, тем ближе показатель *AUC* к единице (ста процентам), но при этом показатель метрики «0,5» (50 %) означает полное отсутствие предсказательной силы, случайное определение *TP* и *FP* с равной вероятностью. Также есть и другие способы оценки эффективности, также используемые в ряде случаев – *F*-мера (агрегированные вместе *precision* и *recall*, лучшее значение также стремится к единице), *AUC* под кривой *precision-recall*, логистическая функция потерь и другие.

Программы и приложения, основанные на технологиях ИИ и МО, незаметно для многих стали подручными инструментами, используемыми каждый день. Коснулись они и такой консервативной отрасли, как медицина. Хотя в ней и имеется целый ряд понятных и во многом оправданных барьеров для внедрения современных методов обработки всевозможных медицинских данных, в мировом научном сообществе уже имеются примеры использования ИИ в качестве помощника врача. Современный ИИ в большинстве своем относится к так называемому «слабому», предназначенному для решению какой-либо конкретной задачи, но его результативность в конкретном приложении может равняться показываемой профильным специалистом или превышать ее по разным параметрам.

Одним из наиболее популярных и наглядных разделов *DS* в настоящее время стало «компьютерное зрение» – направление анализа изображений для получения из них нужной информации, основанное в большинстве случаев на различных вариациях НС. Вполне ожидаемо оно начало находить свое применение в диагностике заболеваний, особенно в рентгенологии и морфологии. Активное развитие анализа рентгенограмм и томограмм обусловлено сложившимся процессом получения рентгеновских снимков, уже включающем в себя хранение оцифрованных изображений, вывод их на монитор

для оценки специалистом. При этом возможность практического применения цифровых технологий в патологической анатомии появилась сравнитель-

но недавно – с вводом сканирующих микроскопов, позволяющих получить полное изображение микропрепарата (англ. *whole slide image, WSI*).

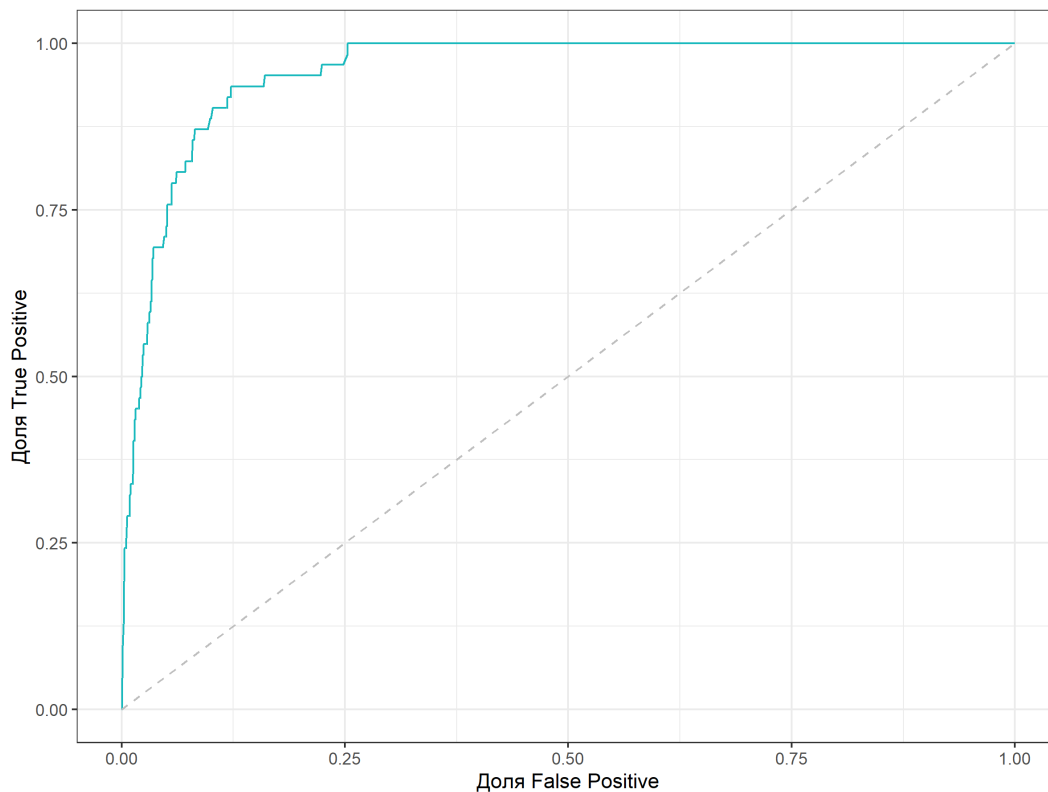


Рис. 1. Пример построения ROC-кривой. В данном случае показатель AUC равен 0,969

Интерес исследователей в данной области значительно вырос, что отражают публикации по запросам *artificial intelligence in pathology* (русс. ИИ в патологии), *machine learning in pathology* (русс. МО в патологии), *neural network in pathology* (русс. НС в патологии). Их количество в библиотеке *PubMed.gov* выросло с десятков в конце двадцатого века, до сотен и тысяч к настоящему времени. При этом чаще употребляется понятие ИИ, а НС уступили в последние годы второе место МО (рис. 2). Русскоязычные работы по этой тематике также имеются, но встречаются значительно реже и представлены в основном обзорными исследованиями, либо изучением алгоритмического ИИ в рамках цифровой патологии [6, 7, 8]. Однако есть и оригинальные публикации, затрагивающие МО и НС [9].

Работы исследовательского уровня были доступны и ранее, с накоплением базы микрофотографий с помощью цифровых камер – дополнительного модуля, редко используемого в повседневной работе врача-патологоанатома. Однако полученные по такому сценарию модели МО могут в дальнейшем использоваться в том числе для анализа *WSI*.

Опыт, накопленный по теме компьютерного зрения в морфологии, представлен в публикациях авто-

ров по всему миру. Исследователи в большей степени занимаются вопросами онкологии, решая целый ряд задач. В их число входят определение первичного опухолевого очага, дифференциация инвазивных форм рака *in situ* и различных доброкачественных изменений, поиск метастазов, анализ экспрессии биомаркеров по результатам иммуногистохимического окрашивания, подсчет прогностически значимых показателей и индексов, сегментация изображений – выделение тех или иных структур, а также такие в большей степени исследовательское направление, как прогнозирование экспрессии конкретных маркеров по микропрепаратам, окрашенным гематоксилином и эозином.

Среди органов, систем и патологических состояний внимание обычно уделяется наиболее встречающимся. Одним из них является рак молочной железы, наиболее изученный в указанном контексте. Целью одних работ была дифференциальная диагностика злокачественного поражения от незлокачественного с/без выделением(-ия) карцином *in situ* и гиперплазии. Эффективность разных моделей достигала 83,3 % (*accuracy*), 81 % (*precision*), 0,962 (*AUC*), а чувствительность и специфичность 89 и 80 % соответственно, что было сопоставимо с результатами

патологов либо превосходило последние [1, 3]. На замороженных срезах показатель *accuracy* достигал 94,96 % [2]. При этом модели, обученные на гистологическую классификацию опухолей, пока-

зывали *accuracy* в 90,66–93,81 % [5]. Менее впечатляющие результаты были получены при оценке степени дифференцировки злокачественных новообразований – 69 % *accuracy* [5].

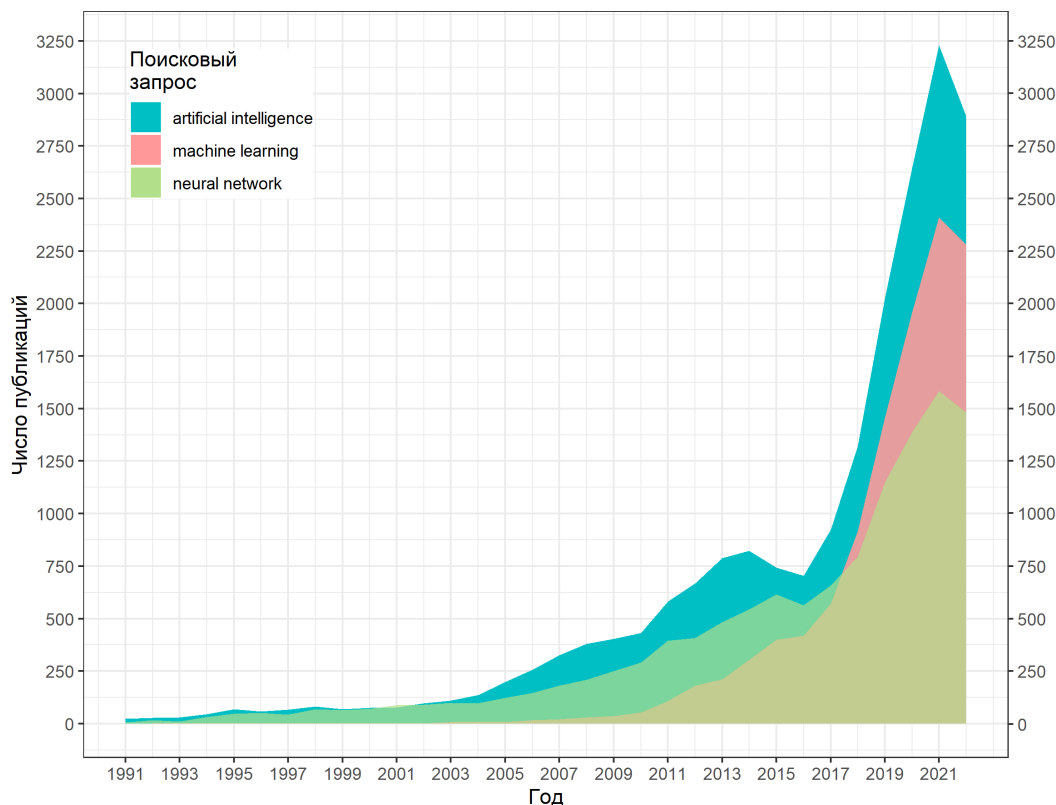


Рис. 2. Ежегодная динамика публикации работ по запросам ИИ, МО и НС в патологии, по данным PubMed.gov

Другие модели, обученные в рамках соревнования *CAMELYON16*, целью которого было распознавание метастазов рака молочной железы в лимфоузлах, показали лучшее значение *AUC*, равное 0,925, что было несколько меньше, чем результаты патологов (0,9664) [3]. Однако при использовании результатов работы моделей специалисты показали результат в 0,9948, что обеспечивало снижение процента ошибок человека примерно на 85 % [3]. По результатам других авторов, лучшая модель показывала 0,994 *AUC*, тогда как патологи лишь 0,810 [3]. Согласно ещё одному исследованию, презентованная авторами модель, показывающая 100-процентную чувствительность, может снизить нагрузку на врача благодаря скринингу микропрепаратов на 58,2 % [1].

Также МО использовалось для подсчёта митозов, индекса *Ki-67*, индекса пролиферации, иммунных клеток, *HER2*, *ER*, *PR* статуса, показывая во многих случаях убедительные результаты [5].

Кроме рака молочной железы активно разрабатываются вопросы детекции и классификации злокачественных опухолей легкого, простаты, мозга, яичников, шейки матки, желудка, толстой кишки,

щитовидной железы, нейроэндокринных опухолей поджелудочной железы, меланомы, остеосаркомы [1, 2, 3, 5, 10]. Модели обучают классифицировать рак простаты по Глиссону, что является одной из проблемных тем современной онкоморфологии ввиду значительных расхождений оценок между разными специалистами и центрами [1, 2]. Менее охвачены, но тоже исследуются в контексте применения МО темы сердечной недостаточности, глютеновой болезни [1].

Отдельно стоит отметить, что с помощью анализа гистологических микропрепаратов можно не только отвечать на вопросы морфологии, но и на прямую получать прогностическую информацию, полезную лечащему врачу. Так, дополняя набор микрофотографий для обучения моделей данными об исходах, можно оценивать прогноз для конкретного пациента уже только по входному микропрепарату [2, 5, 10].

ЗАКЛЮЧЕНИЕ

Информатизация и цифровизация здравоохранения и морфологии предоставляют возможность автоматизации процессов различного уровня сложности. С помощью методов *DS*, в частности ИИ, МО и НС,

исследователями из разных стран разрабатываются способы анализа изображений микропрепаратов, имеющие ценность как для морфологов, так и для клиницистов. Несмотря на наличие трудностей и ограничений, которые необходимо преодолеть, прежде чем современные технологии обработки медицинских данных войдут в повседневную практику врача-патологоанатома, опубликованные результаты уже показывают их потенциал в улучшении диагностики и, как следствие, лечения широкого спектра заболеваний.

СПИСОК ИСТОЧНИКОВ

1. Acs B., Rantalainen M., Hartman J. Artificial intelligence as the next step towards precision pathology. *Journal of internal medicine*. 2020;1(288):62–81.
2. Chang H.Y., Jung C.K., Woo J.I. et al. Artificial intelligence in pathology. *Journal of pathology and translational medicine*. 2019;1(53):1–12.
3. Cui M., Zhang D. Y. Artificial intelligence and computational pathology. *Laboratory Investigation*. 2021;4(101):412–422.
4. Förtsch S., Klauschen F., Hufnagl P., Roth W. Artificial intelligence in pathology. *Deutsches Ärzteblatt International*. 2021;12(118):199.
5. Jiang Y., Yang M., Wang S. et al. Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer communications*. 2020;4(40):154–166.
6. Ганцев Ш.Х., Франц М.В. Искусственный интеллект как инструмент поддержки в принятии решений по диагностике онкологических заболеваний. *Медицинский вестник Башкортостана*. 2018;13;4(76):67–71.
7. Лебедев Г.С., Шадеркин И.А., Тертычный А.С., Шадеркина А.И. Цифровая патоморфология: создание системы автоматизированной микроскопии. *Журнал телемедицины и электронного здравоохранения*. 2021;7(4):27–47. doi: 10.29188/2712-9217-2021-7-4-27-47.
8. Никитаев В.Г., Проничев А.Н., Сельчук В.Ю. и др. Цифровая медицина: искусственный интеллект в телемедицинских технологиях диагностики онкологических заболеваний. *Управление развитием крупномасштабных систем (MLSD '2019)*. 2019:1125–1128. doi: 10.25728/mlsd.2019.1.1125.
9. Борбат А.М., Лишук С.В. Первый российский набор данных гистологических изображений патологических процессов молочной железы. *Врач и информационные технологии*. 2020;3:25–30. doi: 10.37690/1811-0193-2020-3-25-30.
10. Yoshida H., Kiyuna T. Requirements for implementation of artificial intelligence in the practice of gastrointestinal pathology. *World Journal of Gastroenterology*. 2021;21(27):2818.

REFERENCES

1. Acs B., Rantalainen M., Hartman J. Artificial intelligence as the next step towards precision pathology. *Journal of internal medicine*. 2020;1(288):62–81.
2. Chang H.Y., Jung C.K., Woo J.I. et al. Artificial intelligence in pathology. *Journal of pathology and translational medicine*. 2019;1(53):1–12.
3. Cui M., Zhang D. Y. Artificial intelligence and computational pathology. *Laboratory Investigation*. 2021;4(101):412–422.
4. Förtsch S., Klauschen F., Hufnagl P., Roth W. Artificial intelligence in pathology. *Deutsches Ärzteblatt International*. 2021;12(118):199.
5. Jiang Y., Yang M., Wang S. et al. Emerging role of deep learning-based artificial intelligence in tumor pathology. *Cancer communications*. 2020;4(40):154–166.
6. Gantsev Sh.Kh., Frants M.V. Artificial intelligence as a tool for decision-making support in the diagnostics of oncological diseases. *Meditinskii vestnik Bashkortostana = Bashkortostan Medical Journal*. 2018;13;4(76):67–71. (In Russ.).
7. Lebedev G.S., Shaderkin I.A., Tertychny A.S., Shaderkina A.I. Digital pathomorphology: creation of an automated microscopy system. *Zhurnal telemeditsiny i elektronnoogo zdavookhraneniya = The journal of telemedicine and e-Health*. 2021;7(4):27–47. (In Russ.) doi: 10.29188/2712-9217-2021-7-4-27-47.
8. Nikitaev V.G., Pronichev A.N., Selchuk V.Yu. et al. Digital medicine: artificial intelligence in telemedicine technologies for the diagnosis of cancer. *Upravlenie razvitiem krupnomasshtabnykh sistem (MLSD '2019) = Management of Large Scale Systems Development (MLSD '2019)*. 2019:1125–1128. (In Russ.) doi: 10.25728/mlsd.2019.1.1125.
9. Borbat A.M., Lishchuk S.V. The first Russian breast pathology histologic images data set. *Vrach i informatsionnye tekhnologii = Medical doctor and information technologies*. 2020;3:25–30. (In Russ.) doi: 10.37690/1811-0193-2020-3-25-30.
10. Yoshida H., Kiyuna T. Requirements for implementation of artificial intelligence in the practice of gastrointestinal pathology. *World Journal of Gastroenterology*. 2021;21(27):2818.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Информация об авторах

Никита Алексеевич Дорофеев – младший научный сотрудник, лаборант-исследователь, Волгоградский государственный медицинский университет, Волгоград, Россия; ✉ dorofeev.na@mail.ru, <https://orcid.org/0000-0003-2679-7524>

Григорий Леонидович Снигур – доктор медицинских наук, доцент, заведующий кафедрой, Волгоградский государственный медицинский университет, Волгоград, Россия; glsnigur@volgmed.ru, <https://orcid.org/0000-0002-8612-6186>

Максим Юрьевич Фролов – кандидат медицинских наук, доцент, заведующий лабораторией, Волгоградский государственный медицинский университет, Волгоград, Россия; clinpharmrussia@mail.ru, <https://orcid.org/0000-0002-0389-560X>

Алексей Владимирович Смирнов – доктор медицинских наук, доцент, заведующий кафедрой патологической анатомии, Волгоградский государственный медицинский университет, Волгоград, Россия; alexeysmirnov.volgsmu@gmail.com, <https://orcid.org/0000-0001-5351-6105>

Дмитрий Александрович Сасов – магистрант по специальности «Программное обеспечение автоматизированных систем», Волгоградский государственный технический университет, Волгоград, Россия; s-dima29@yandex.ru

Александр Владимирович Зубков – ведущий инженер-программист Волгоградский государственный медицинский университет, Волгоградский государственный технический университет, Волгоград, Россия; zubkov.alexander.v@gmail.com

Статья поступила в редакцию 20.06.2023; одобрена после рецензирования 25.10.2023; принята к публикации 15.02.2024.

Competing interests. The authors declare that they have no competing interests.

Information about the authors

Nikita A. Dorofeev – Junior Researcher, Laboratory Assistant, Volgograd State Medical University, Volgograd, Russia; dorofeev.na@mail.ru, <https://orcid.org/0000-0003-2679-7524>

Grigory L. Snigur – Doctor of Medical Sciences, Associate Professor, Head of the Department, Volgograd State Medical University, Volgograd, Russia; glsnigur@volgmed.ru, <https://orcid.org/0000-0002-8612-6186>

Maxim Yu. Frolov – Candidate of Medical Sciences, Associate Professor, Head of the Laboratory, Volgograd State Medical University, Volgograd, Russia; clinpharmrussia@mail.ru, <https://orcid.org/0000-0002-0389-560X>

Alexey V. Smirnov – Doctor of Medical Sciences, Associate Professor, Head of the Department of Pathological Anatomy, Volgograd State Medical University, Volgograd, Russia; alexeysmirnov.volggmu@gmail.com, <https://orcid.org/0000-0001-5351-6105>

Dmitry A. Sasov – a master’s student in the specialty “Software for automated Systems”, Volgograd State Technical University, Volgograd, Russia; s-dima29@yandex.ru

Alexander V. Zubkov – Leading Software Engineer Volgograd State Medical University, Volgograd State Technical University, Volgograd, Russia; zubkov.alexander.v@gmail.com

The article was submitted 20.06.2023; approved after reviewing 25.10.2023; accepted for publication 15.02.2024.