

УДК 004.932.2

DOI: [10.26102/2310-6018/2025.48.1.016](https://doi.org/10.26102/2310-6018/2025.48.1.016)

Система анализа изображений ядросодержащих клеток костного мозга для формирования диагностического заключения в онкогематологии

Е.В. Поляков, В.В. Попов, В.В. Дмитриева

*Национальный исследовательский ядерный университет «МИФИ», Москва,
Российская Федерация*

Резюме. В работе представлена система анализа изображений ядросодержащих клеток костного мозга для формирования диагностического заключения в онкогематологии, направленной на решение проблемы построения конвейера обработки данных в автоматических анализаторах биомедицинских изображений. Актуальность исследования обусловлена необходимостью повышения надежности автоматического микроскопического анализа биомедицинских образцов, что является сложной задачей из-за высокой изменчивости и морфологической сложности исследуемых объектов. Одним из решений указанной проблемы является разработка веб-сервиса посредством загрузки, обработки и описание изображений с последующей классификацией на категории подтвержденных и неподтвержденных случаев. Данный веб-сервис обеспечивает кроссплатформенность и доступность, формирует открытую базу данных верифицированных изображений и предоставляет инструменты для обработки и анализа изображений, а также инструменты для корректировки врачом результатов обработки. Система не назначает лечение и не ставит диагнозы самостоятельно, а служит интеллектуальным инструментом для обработки, анализа и передачи результатов исследований в режиме реального времени. Результаты тестирования показали высокую точность работы системы: 91 % для нейросетевых методов и до 97 % для классических алгоритмов. Разработанная система позволяет проводить анализ модулей обработки данных для систем компьютерной микроскопии.

Ключевые слова: анализ биомедицинских изображений, выделение объектов, классификация ядросодержащих клеток, распознавание образов, онкогематология.

Для цитирования: Поляков Е.В., Попов В.В., Дмитриева В.В. Система анализа изображений ядросодержащих клеток костного мозга для формирования диагностического заключения в онкогематологии. *Моделирование, оптимизация и информационные технологии.* 2025;13(1). URL: <https://moitvvt.ru/ru/journal/pdf?id=1796> DOI: 10.26102/2310-6018/2025.48.1.016

A system for analyzing images of nucleated bone marrow cells for the formation of a diagnostic conclusion in oncohematology

E.V. Polyakov, V.V. Popov, V.V. Dmitrieva

*National Research Nuclear University MEPhI (Moscow Engineering Physics Institute),
Moscow, the Russian Federation*

Abstract. The paper presents a system for analyzing images of nucleated bone marrow cells to form a diagnostic conclusion in oncohematology, aimed at solving the problem of constructing a data processing pipeline in automatic analyzers of biomedical images. The relevance of the study is due to the need to improve the reliability of theof automatic microscopic analysis of biomedical samples, which is aa difficult task due to high variability and morphological complexity of the investigated objects. One solution to this problem is to develop a web service that uploads, processes and describes images, then classifies them into categories of confirmed and unconfirmed cases. This web service provides cross-platform and accessibility, builds an open database of verified images and providestools for processing

and analyzing images, as well as tools for correcting by the physician of the processing results. The system does not prescribe treatment and does not make diagnoses independently, but serves as an intelligent tool for processing, analyzing and transmission of research results in real time. The testing results showed high accuracy of the system: 91% for neural network methods and up to 97% for classical algorithms. The developed system allows for the analysis of data processing modules for computer microscopy systems.

Keywords: analysis of biomedical images, selection of objects, classification of nucleated cells, pattern recognition, oncohematology.

For citation: Polyakov E.V., Popov V.V., Dmitrieva V.V. A system for analyzing images of nucleated bone marrow cells for the formation of a diagnostic conclusion in oncohematology. *Modeling, Optimization and Information Technology*. 2025;13(1). (In Russ.). URL: <https://moitvvt.ru/journal/pdf?id=1796> DOI: 10.26102/2310-6018/2025.48.1.016

Введение

Исследования, направленные на повышение надежности автоматического микроскопического анализа биомедицинских образцов, проводятся уже более 50 лет. Научные и инженерные исследования были направлены как на автоматизацию рутинных операций получения изображений образцов, так и на создание методов и алгоритмов анализа и интерпретации изображений [1–2]. Применяющиеся при диагностике многочисленной группы патологий анализа микроскопических объектов биоматериалов относятся к группе наиболее трудных как для визуального анализа врачом, например, при анализе резидуальных лимфо- и миелобластов, так и для автоматического анализатора [3–4]. Как следствие, существует острая необходимость разработки интеллектуальных систем анализа микроскопических изображений. Это обусловлено, прежде всего, сложностью морфологии исследуемых объектов, их высокой изменчивостью, многочисленностью, сходством признаков, наличием редких форм, не полностью контролируемым влиянием на морфологию технологий подготовки препаратов и другими факторами [3, 5, 6].

Последние достижения машинного обучения в области компьютерного зрения позволили исследователям применять конвейеры машинного обучения для автоматизации анализа медицинских изображений. Тем не менее, существуют серьезные проблемы, связанные с применением этих конвейеров в области анализа медицинских изображений [7].

Цель работы – разработка системы анализа изображений ядросодержащих клеток костного мозга для формирования диагностического заключения в онкогематологии и анализа конвейеров машинного обучения при автоматизации обработки медицинских изображений.

Постановка проблемы

1. Опыт диагностики онкогематологических заболеваний формируется не только при непосредственном участии врача в процедуре обследования, но и при самостоятельном изучении подобных случаев в медицинской практике [8–10].

2. Отсутствует единая открытая база данных верифицированных изображений. Подобные базы данных существуют в рамках клиники или локальных программно-аппаратных комплексов, что требует покупки не только лицензии на использование программного обеспечения, но и самого аппаратного комплекса, совместимого с программным обеспечением.

3. Сложность обработки, анализа и верификации изображений связана, в первую очередь, с отсутствием единых стандартов проведения последовательного цифрового

анализа биомедицинских изображений, а также единой базы знаний закономерных связей физических параметров идентифицируемых объектов с фактическими данными подтвержденных случаев [7, 11].

4. Отсутствует возможность оперативного поиска изображений по содержанию описания.

Для решения вышеперечисленных проблем предлагается разработка веб-сервиса для загрузки, обработки и описания изображений с последующей классификацией на категории подтвержденных и неподтвержденных случаев.

Данный веб-сервис позволит [6]:

- 1) обеспечить кроссплатформенность и доступность решения;
- 2) сформировать открытую базу данных изображений и их описаний;
- 3) обеспечить фоновую обработку изображений с целью сегментации, поиска, анализа параметров основных объектов;
- 4) предоставить интуитивно понятные инструменты для корректировки врачом результатов обработки.

Задача распознавания атипичных клеток

Рассматриваемая задача на основе анализа изображений, полученных с препаратов крови и костного мозга, включает в себя следующие подзадачи:

- Предварительная обработка изображения – улучшение изображения.
- Обнаружение объекта – локализация объекта, осуществляемая посредством сегментации изображения и определения положения интересующего объекта.
- Извлечение признаков и обучение – использование статистических методов или методов глубокого обучения для выявления наиболее значимых шаблонов изображений и уникальных особенностей, которые помогут классификатору различать разные классы (типы объектов).
- Классификация изображения по предопределенным классам, используя подходящий метод, который сравнивает шаблоны изображений с целевыми шаблонами.

Важной характеристикой системы является способность анализировать данные диагностических исследований с различных типов медицинского оборудования, независимо от его производителя, в рамках одной системы и в едином интерфейсе.

При этом система не назначает лечение и не ставит диагнозы самостоятельно – это всего лишь интеллектуальный инструмент, который обрабатывает, анализирует и автоматически передает результаты исследований в режиме реального времени.

Подход к формированию диагностического заключения в онкогематологии и анализа конвейеров машинного обучения для автоматизации обработки медицинских изображений

Веб-сервис загрузки, обработки и описания биомедицинских изображений с последующим формированием диагностического заключения позволяет:

- 1) обеспечить кроссплатформенность и доступность решения;
- 2) сформировать открытую базу данных изображений и их описаний;
- 3) провести фоновую обработку изображений с целью сегментации, поиска и оценки параметров основных биомедицинских объектов;
- 4) предоставить интуитивно понятные инструменты для корректировки врачом результатов обработки.

Реализованные элементы конвейера анализа изображений и формирования диагностического заключения в онкогематологии:

- Авторизация, регистрация и личный кабинет пользователя.

– Загрузка и предварительный просмотр изображения.
– Фоновая обработка изображений для обнаружения и классификации ядродержащих клеток крови и костного мозга и просмотра результатов по этапам анализа. Данный элемент основан на многошаговой процедуре, включающей в себя ряд алгоритмов предобработки, выделения и описания обнаруженных объектов на изображениях. Для работы с нейронной сетью применялись библиотеки Ultralytics и PyTorch, а в качестве архитектуры глубокого обучения была выбрана современная модель YOLOv8.

– Вычисление основных параметров обнаруженных объектов: морфологических, текстурных, характеристик.

Для тестирования реализованы и встроены модули в веб-интерфейс с использованием языка Python. Процесс тестирования состоит из следующих этапов:

– создание в базе данных записи о пользователе с правами администратора;
– авторизация пользователя с правами администратора на сайте;
– загрузка изображений через панель администратора;
– получение подробного отчета о результатах выполнения обработки на каждом этапе.

Таким образом, разработанные элементы позволяют загрузить и обработать изображения с любого устройства, имеющего доступ к интернету, что, в свою очередь, позволяет обеспечить:

- 1) кроссплатформенность и доступность решения;
- 2) возможность формирования открытой базы данных изображений и их описаний;
- 3) фоновую обработку изображений с целью сегментации, поиска и оценки параметров основных объектов биомедицинских изображений;
- 4) разработку интуитивно понятных инструментов для корректировки врачом результатов обработки;
- 5) автоматическое создание обучающих наборов данных для модуля поиска изображений по описаниям, а также для автоматической классификации найденных объектов биомедицинских изображений и выявление закономерностей между параметрами для последующего формирования диагностического заключения.

При этом решается проблема диагностики онкогематологических заболеваний, связанных с отсутствием единой открытой базы данных верифицированных изображений, сложностью обработки, анализа и верификации изображений, отсутствием возможности оперативного поиска изображений по содержанию описания, что, в свою очередь, позволит улучшить статистику выявления онкогематологических заболеваний на ранних стадиях и, как следствие, снизить уровень смертности.

Анализ результатов и их обсуждение

Результат работы по анализу конвейеров обработки данных показан на Рисунке 1, где представлен фрагмент интерфейса, демонстрирующего анализ изображений ядродержащих клеток крови и костного мозга на базе построенного конвейера. Изображения масштабированы для визуального восприятия, при этом вся информация сохранена, как в реальном приложении. На Рисунках 1а–г представлена работа классических методов выделения на изображениях ядер клеток костного мозга, на Рисунке 1д приведено входное изображение нейронной сети, на Рисунке 1е – результат работы нейронной сети с архитектурой YOLOv8.

Для исследований использовались данные базы данных цветных изображений с разметкой [12]. Набор содержит 14 классов объектов: Лимфоцит (657), Метамиелоцит

(191), Моноцит (201), Миелоцит (249), Нормобласты (500), Палочкоядерный нейтрофил (164), Сегментоядерный нейтрофил (476), Бласт (10163), Промиелоцит (4), Эозинофил (78), Базофил (3), Магакариоцит (1), Артефакты (5), плазматическая клетка (13). Балансировка классов, помимо добавления вырезанных и обработанных изображений ядер клеток, включала пространственные преобразования. Выборка разделялась на обучающую, валидационную и тестовую в соотношении 80 / 10 / 10. При исследовании классических методов машинного обучения, выборка состояла только из выделенных ядер и разделялась в соотношении 80 / 20. Анализ результатов обнаружения и классификации проводился экспертом.

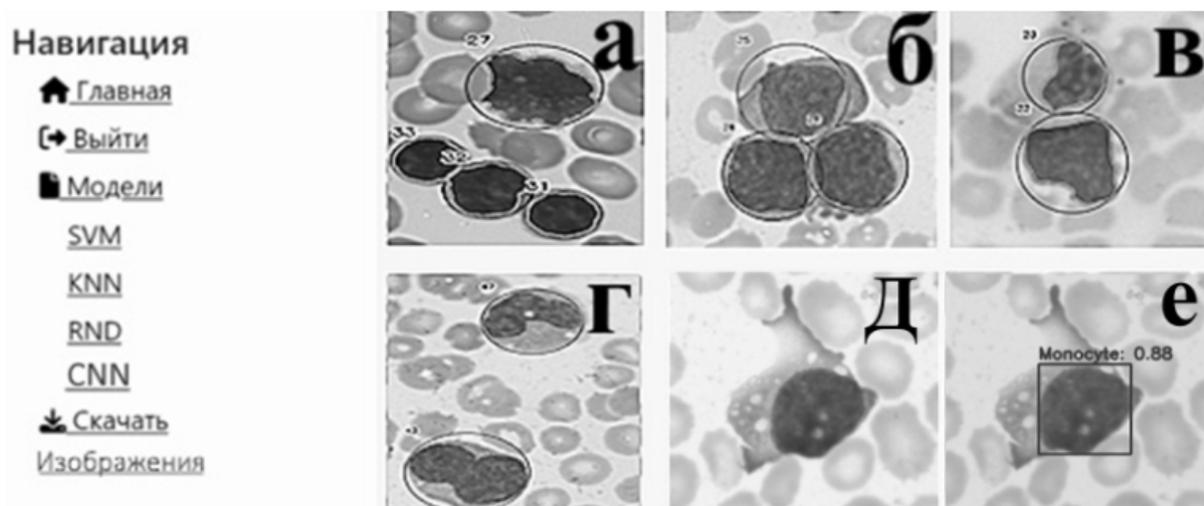


Рисунок 1 – Фрагмент интерфейса, демонстрирующего анализ конвейеров обработки изображений ядродержащих клеток крови и костного мозга

Figure 1 – A fragment of the interface demonstrating the analysis of image processing pipelines of blood and bone marrow nucleus-containing cells

Экспериментальные исследования позволили оценить точность работы системы. Исследование классификации с применением разработанной системы позволило определить оптимальные гиперпараметры и улучшило точность классификаторов. Гиперпараметры, используемые при обучении модели YOLOv8, включают: размер входного изображения (`img_size=640`); скорость обучения (`learning_rate=0.001`); количество образцов, обрабатываемых одновременно (`batch_size=16`); количество полных проходов по обучающему набору данных (`epochs=80`); алгоритм оптимизации, используемый для обновления весов (`optimizer='Adam'`); функция, используемая для оценки качества предсказаний модели (`loss_function='CrossEntropy'`). Результат обучения нейронной сети представлен на Рисунке 2. На рисунке 2а представлен график изменения точности (`precision`) модели в процессе обучения, на Рисунке 2б – график метрики `mAP@0.5` (mean Average Precision при пороге 0.5), показывающий среднюю точность по всем классам при заданном уровне перекрытия (`IoU = 0.5`) между предсказанными и истинными ограничивающими рамками.

Для архитектуры нейронной сети YOLOv8 получены следующие характеристики: точность (`precision`) – 91 %, F1-мера – 80 %, `mAP@0.5` – 81 %.

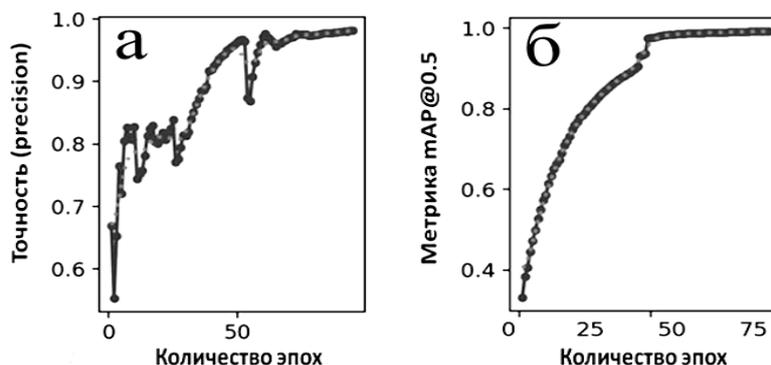


Рисунок 2 – Результат обучения нейронной сети: а – график точности (precision); б – график mAP@0.5 метрики

Figure 2 – Neural network training result: a – precision graph; b – graph of mAP@0.5 metric

Один из результатов работы конвейера с применением нейронной сети глубокого обучения для распознавания цветных изображений ядросодержащих клеток по типам, представлен на Рисунке 3.

```
80 epochs completed in 49.741 hours.
Optimizer stripped from runs/detect/train/weights/last.pt, 52.1MB
Optimizer stripped from runs/detect/train/weights/best.pt, 52.1MB

Validating runs/detect/train/weights/best.pt...
Ultralytics YOLOv8.2.38 Python-3.9.2 torch-2.3.1+cud118 CUDA:0 (NVIDIA GeForce RTX 4060 Ti, 15985MiB)
Model summary (fused): 218 layers, 25852498 parameters, 0 gradients, 78.8 GFLOPs
```

Class	Images	Instances	Box(P)	R	mAP50	mAP50-95	100%
all	14115	60603	0.943	0.93	0.935	0.889	
Lymphocyte	2131	2748	0.966	0.984	0.992	0.95	
Metamyelocyte	579	668	0.969	0.981	0.993	0.969	
Monocyte	410	492	0.948	0.982	0.992	0.959	
Myelocyte	795	936	0.97	0.987	0.992	0.975	
Normoblasts	42	46	0.968	0.826	0.882	0.826	
Rod neutrophil	687	807	0.97	0.993	0.993	0.965	
Segmentonuclear neutrophil	1135	1372	0.984	0.992	0.995	0.96	
blast	10899	42831	0.989	0.998	0.994	0.979	
Promyelocyte	13	13	0.89	1	0.995	0.988	
Eosinophil	160	172	0.977	1	0.995	0.94	
Basophil	10	10	0.87	1	0.977	0.895	
Magakariocyte	2	2	1	0	0.0259	0.0234	
Bad cells	1913	4367	0.965	0.986	0.992	0.957	
Plasma cell	11	13	1	0.904	0.937	0.865	
artifacts	39	42	0.93	0.953	0.971	0.825	
gumpricht shadows	2701	4438	0.957	0.991	0.99	0.928	
methoses	103	108	0.968	1	0.995	0.962	
normoblasts: oxyfilny	1038	1191	0.968	0.977	0.987	0.942	
macrophages	20	20	0.786	1	0.99	0.979	
normoblasts: erythroblasts	19	20	0.821	0.92	0.895	0.787	
normoblasts: basophilic	64	64	0.889	1	0.991	0.92	
normoblasts: polychromatophilic	233	243	0.959	0.992	0.994	0.957	

```
Speed: 0.3ms preprocess, 6.0ms inference, 0.0ms loss, 1.0ms postprocess per image
Results saved to runs/detect/train
```

Рисунок 3 – Результат тестирования на обученной модели для исследуемого набора данных
Figure 3 – Result of testing on the trained model for the investigated dataset

Среди классических методов стоит отметить ансамблевый метод на основе обучения множества деревьев решений (Extremely Randomized Trees), точность классификации которого составила 96 %, и градиентный бустинг (Gradient Boosting Classifier), который достиг точности 97 %. Классификаторы: метод линейных опорных векторов (Linear Support Vector Classifier), логистическая регрессия (Logistic Regression) и упаковочный классификатор (Bagging Classifier) показали неудовлетворительные точности классификации (около 50 %).

При применении машинного обучения в рутинной клинической практике исследователю необходимо предусмотреть набор программного обеспечения для обеспечения автоматизации процесса коррекции обученной модели и повторное дообучение, а также обратить внимание на тип маркера, назначаемый объектам, которые,

в ходе разметки данных, не были определены (артефакты). При анализе ядросодержащих клеток выбор стоит делать в направлении методов глубокого обучения при прочих равных условиях.

Для дальнейшего расширения наборов данных и проведения исследований на изменяющихся данных необходимо закрепить постоянный адрес в интернете за веб-сервисом, обеспечив его доступность и тестирование в открытом доступе.

Заключение

В результате работы разработана система анализа изображений ядросодержащих клеток костного мозга для формирования диагностического заключения в онкогематологии и исследования конвейеров машинного обучения для автоматизации анализа медицинских изображений, направленная на повышение точности и эффективности обработки большого объема данных, а также снижение нагрузки на специалистов. Система позволяет проводить анализ удаленно, без участия узкоспециализированных врачей, что актуально в труднодоступных регионах.

Эксперимент, проведенный в рамках данной работы, позволил оценить точность работы системы. Исследование классификации с применением разработанной системы позволило определить оптимальные гиперпараметры и улучшило точность классификаторов.

Дальнейшим развитием системы является ее масштабирование и проведение исследований с привлечением широкого круга специалистов и сбора статистики по работе с системой.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Kreiss L., Jiang S., Li X., Xu S., Zhou K.C., Lee K.C., Mühlberg A., Kim K., Chaware A., Ando M., Barisoni L., Seung A.L., Zheng G., Lafata K.J., Friedrich O., Horstmeyer R. Digital staining in optical microscopy using deep learning – a review. *Photonix*. 2023;4. <https://doi.org/10.1186/s43074-023-00113-4>
2. Yao J., Huang X., Wei M., Han W., Xu X., Wang R., Chen J., Sun L. High-Efficiency Classification of White Blood Cells Based on Object Detection. *Journal of Healthcare Engineering*. 2021;2021. <https://doi.org/10.1155/2021/1615192>
3. Medovy V.S., Volkov G.D., Strela N.M., Pervushkin I.V. An Adaptable Cloud-Based Multiple-Unit Laboratory Microscopy System. *Biomedical Engineering*. 2021;55:36–40. <https://doi.org/10.1007/s10527-021-10066-2>
4. Samorodov A.V. Biotechnological Systems for Automated Microscopy of Cytology Specimens. *Biomedical Engineering*. 2019;52(6):387–390. <https://doi.org/10.1007/s10527-019-09853-9>
5. Поляков Е.В. Исследование характеристик шума на изображениях в системах компьютерной микроскопии. *Вестник ТГТУ*. 2020;26(4):598–603.
Polyakov E.V. A study of noise characteristics on images in computer microscopy systems. *Transactions TSTU*. 2020;26(4):598–603. (In Russ.)
6. Дмитриева В.В., Тупицын Н.Н., Поляков Е.В., Носова Е.М., Палладина А.Д., Цыпьяк В.И., Либерис К.А. Медицинская информационная система с применением web-технологий для диагностики острых лимфобластных лейкозов и минимальной остаточной болезни. *Безопасность информационных технологий*. 2021;28(3):44–55. <https://doi.org/10.26583/bit.2021.3.03>
Dmitrieva V.V., Tupitsyn N.N., Polyakov E.V., Nosova E.M., Palladin A.D., Tsyplyak V.I., Liberis K.A. Medical information system based web technologies for the

- diagnosis of acute lymphoblastic leukemia and minimal residual disease. *IT Security (Russia)*. 2021;28(3):44–55. (In Russ.). <https://doi.org/10.26583/bit.2021.3.03>
7. Fazeli S., Samiei A., Lee T.D., Sarrafzadeh M. Beyond Labels: Visual Representations for Bone Marrow Cell Morphology Recognition. In: *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI), 26–29 June 2023, Houston, USA*. IEEE; 2023. pp. 111–117. <https://doi.org/10.1109/ICHI57859.2023.00025>
 8. Deshpande N.M., Gite S., Aluvalu R. A review of microscopic analysis of blood cells for disease detection with AI perspective. *PeerJ Computer Science*. 2021;7. <https://doi.org/10.7717/peerj-cs.460>
 9. Yildirim M., Çinar A. Classification of White Blood Cells by Deep Learning Methods for Diagnosing Disease. *Revue d'Intelligence Artificielle*. 2019;33(5):335–340.
 10. Dehkharghanian T., Mu Y., Ross C., Sur M., Tizhoosh H.R., Campbell C.J.V. Cell projection plots: A novel visualization of bone marrow aspirate cytology. *Journal of Pathology Informatics*. 2023;14. <https://doi.org/10.1016/j.jpi.2023.100334>
 11. Cheng Z., Li Y. Improved YOLOv7 Algorithm for Detecting Bone Marrow Cells. *Sensors*. 2023;23(17). <https://doi.org/10.3390/s23177640>
 12. Поляков Е.В., Тупицын Н.Н., Серебрякова И.Н., Палладина А.Д.; правообладатель федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский ядерный университет «МИФИ» (НИЯУ МИФИ). База данных клеток костного мозга больных острым лимфобластным лейкозом: заявка № 2023620945: заявл. 06.04.2023: опубл. 20.04.2023. Свидетельство о государственной регистрации базы данных № 2023621283 Российская Федерация.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Поляков Евгений Валерьевич, кандидат технических наук, доцент кафедры медицинской физики, Национальный исследовательский ядерный университет МИФИ, Москва, Российская Федерация.
e-mail: evpolyakov@mephi.ru
ORCID: [0000-0002-5346-6504](https://orcid.org/0000-0002-5346-6504)

Evgeniy V. Polyakov, Candidate of Engineering Sciences, Associate Professor of the Department of Medical Physics, National Research Nuclear University Mephi (Moscow Engineering Physics Institute), Moscow, the Russian Federation.

Попов Владимир Викторович, студент, Национальный исследовательский ядерный университет МИФИ, Москва, Российская Федерация.
e-mail: popovvladimir618618@gmail.com

Vladimir V. Popov, student, National Research Nuclear University Mephi (Moscow Engineering Physics Institute), Moscow, the Russian Federation.

Дмитриева Валентина Викторовна, кандидат технических наук, доцент кафедры электрофизических установок, Национальный исследовательский ядерный университет МИФИ, Москва, Российская Федерация.
e-mail: vydmitriyeva@mephi.ru
ORCID: [0000-0002-9202-6691](https://orcid.org/0000-0002-9202-6691)

Valentina V. Dmitrieva, Candidate of Engineering Sciences, Associate Professor of the Department of electrophysical systems, National Research Nuclear University Mephi (Moscow Engineering Physics Institute), Moscow, the Russian Federation.

Статья поступила в редакцию 10.01.2025; одобрена после рецензирования 03.02.2025; принята к публикации 05.02.2025.

The article was submitted 10.01.2025; approved after reviewing 03.02.2025; accepted for publication 05.02.2025.