

ТЕХНИЧЕСКИЕ НАУКИ

TECHNICAL SCIENCES

Научная статья

УДК 004.932.2

ББК 32.818.1

А 45

DOI: 10.53598/2410-3225-2024-2-341-27-34

Алгоритмы распознавания объектов в видеопотоке и определение свойств их взаимного расположения

(Рецензирована)

Марат Вячеславович Алиев¹, Дмитрий Александрович Бербенцев²,
Валерий Олегович Немыкин³, София Маратовна Алиева⁴

^{1, 2, 4} Адыгейский государственный университет, Майкоп, Россия

¹ alievmarat@mail.ru

² dmitriy.berbencev@mail.ru

³ Майкопский государственный технологический университет, Майкоп, Россия,
West55555@list.ru

⁴ alievasofia172@gmail.com

Аннотация. Представлен алгоритм обнаружения и идентификации объектов с возможностью определения характеристик их взаимного расположения. Алгоритм основан на использовании нейросетевой модели YOLO. Продемонстрировано его практическое применение в медицине и дана оценка эффективности.

Ключевые слова: обнаружение объектов, нейронная сеть YOLO, режим реального времени, координаты ключевых точек, инъекция, угол ввода, видеокамера

Для цитирования: Алгоритмы распознавания объектов в видеопотоке и определение свойств их взаимного расположения / М. В. Алиев, Д. А. Бербенцев, В. О. Немыкин, С. М. Алиева // Вестник Адыгейского государственного университета. Сер.: Естественно-математические и технические науки. 2024. Вып. 2 (341). С. 27–34. DOI: 10.53598/2410-3225-2024-2-341-27-34

Original Research Paper

Algorithms for recognizing objects in a video stream and determining the properties of their relative position

Marat V. Aliev¹, Dmitry A. Berbentsev², Valery O. Nemykin³, Sofiya M. Alieva⁴

^{1, 2, 4} Adyghe State University, Maykop, Russia

¹ alievmarat@mail.ru

² dmitriy.berbencev@mail.ru

³ Maykop State Technological University, Maykop, Russia, West55555@list.ru

⁴ alievasofia172@gmail.com

Abstract. An algorithm for detecting and identifying objects with possibility of determining characteristics of their mutual arrangement is presented. The algorithm is based on the use of the YOLO neural network model. Its practical application in copper-zinc is demonstrated and an efficiency assessment is given.

Keywords: object detection, YOLO neural network, real-time mode, keypoints coordinates, injection, insertion angle, video camera

For citation: Algorithms for recognizing objects in a video stream and determining the properties of their relative position / M. V. Aliev, D. A. Berbentsev, V. O. Nemykin, S. M. Alieva // The Bulletin of the Adyghe State University. Ser. : Natural-Mathematical and Technical Sciences. 2024. Iss. 2 (341). P. 27–34. DOI: 10.53598/2410-3225-2024-2-341-27-34

Введение

Обнаружение объектов с возможностью определения характеристик их взаимного расположения – актуальное направление в развитии таких областей, как робототехника, производство, видеонаблюдение и медицина. Данный подход применим к решению многих задач. В исследовании [1] такого рода алгоритм позволяет определить положение зрелого плода в трехмерном пространстве, которое впоследствии используется роботом-сборщиком для позиционирования его руки при сборе урожая, а в работе [2] описывается его использование для определения угла ослабления болта. Однако существующих решений, осуществляющих контроль проведения различных видов инъекций, авторам неизвестны.

Ключевую роль в эффективности и безопасности данной процедуры играет правильный угол ввода, позволяющий минимизировать болевые ощущения и дискомфорт у пациента, а также обеспечивающий корректное попадание лекарственного средства [3].

Современные возможности машинного зрения и нейросетевых технологий позволяют реализовать тренажер, который будет идентифицировать конечность и медицинский прибор, а также определит их положение в пространстве, на основании чего будет произведен расчет угла ввода инъекции. Данный тренажер предназначен для повышения эффективности обучения и тренировок.

Обзор существующих решений в области детекции объектов

Среди алгоритмов по детекции и классификации можно выделить (Faster R-CNN) [4] и You Only Look Once (YOLO) [5].

Рассматривая количественные характеристики данных подходов, приведенные в статье [6], можно сделать вывод, что оба метода имеют достаточно высокие показатели (табл. 1).

Таблица 1

Результаты YOLOv8 и Faster R-CNN [6]
Table 1. YOLOv8 and Faster R-CNN results [6]

	YOLOv8	Faster R-CNN
Precision	0,898	0,894
Recall	0,785	0,915
F1	0,838	0,904
mAP50	0,863	0,908
mAP50–95	0,68	0,676

Метод Faster R-CNN превосходит YOLO по точности в области обнаружения и классификации. Тем не менее стоит отметить, что данная модель требует значительно больше времени как для обучения, так и для выполнения задачи. В связи с этим можно предположить, что при использовании детектора в режиме реального времени модель YOLO может оказаться предпочтительнее. Несмотря на небольшое снижение точности по сравнению с Faster R-CNN, она компенсирует это своей высокой скоростью работы, что делает ее эффективной в данном контексте.

Сбор данных

Анализ рассмотренных решений приводит к выводу, что наиболее подходящей моделью является YOLO, так как содержит в себе инструменты по обнаружению и

идентификации объекта, а также позволяет выявить его ключевые точки, при этом предоставляя высокую скорость обработки изображений, что необходимо при работе в режиме реального времени.

Тестирование алгоритма проводится на задаче определения угла ввода инъекции в руку, поэтому в качестве целей обнаружения выступают: шприц, инъекционная игла и рука.

Датасет формируется из изображений с динамическим фоном, а также применением аугментации. Этот метод преобразует исходные данные с помощью таких операций, как сдвиг, поворот, отражение, добавление шума и т. д., с целью повышения разнообразия набора изображений и его устойчивости к искажениям.

Набор состоит из 2 классов и 2137 изображений с динамически меняющимся фоном. Формат данных – JPG, используемая цветовая модель – RGB, размер преобразован в 640x640 пикселей. Далее производится разделение на обучающую и тестовую выборку в следующих соотношениях. Разделение датасета показано на рисунке 1.

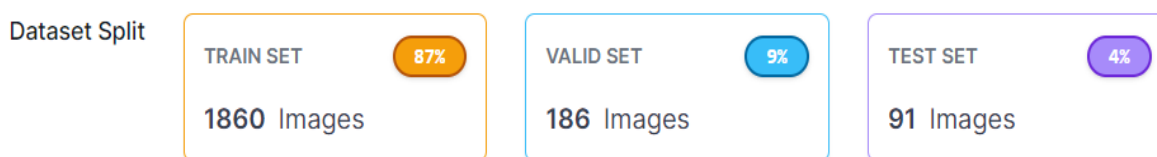


Рис. 1. Разделение датасета

Fig. 1. Dataset partitioning

Обучение

В исследовании используется модель YOLOv8-Pose, целью которой является выделение ключевых точек для идентификации различных частей тела человека. Расположение представляется в виде набора координат в формате $[x, y]$. Pose позволит выделить ключевые точки руки и шприца, которые впоследствии будут использованы для расчета угла при вводе инъекции. Данную модель можно обучить на любом наборе данных, в котором каждое изображение маркируется с учетом выделения точек, характеризующих тот или иной класс. Этот метод имеет преимущество перед обычной ограничивающей рамкой, так как она не является мобильной и не позволяет эффективно использовать асимметричные характеристики объектов, что понижает точность их распознавания.

Остается подготовить модель на ранее собранном датасете. Для этого используем уже предобученную нейронную сеть, дополнив ее своими данными.

Чтобы получить более наглядное представление о ходе и качестве обучения модели, построим график зависимости функции потерь и оценок (ассигасу – доля правильных ответов, precision – точность, recall – полнота) от номера эпохи обучения. Результат представлен на рисунке 2.

Функция потерь сравнивает прогнозируемый результат модели с истинным выходом и выдает оценку, показывающую вероятность принадлежности изображения x_i к классу y_i :

$$L = \frac{1}{N} \sum_i L_i(f(x_i, W), y_i), \quad (1)$$

где L – общее число потерь, N – количество данных для обучения, x – исходное изображение, y – метки классов, f – классификатор, W – весовая матрица.

Сформируем матрицу, показывающую точность прогнозирования классификатора в отношении двух и более классов. Результат представлен на рисунке 3. В результате обучения будет получена модель с наилучшими показателями.

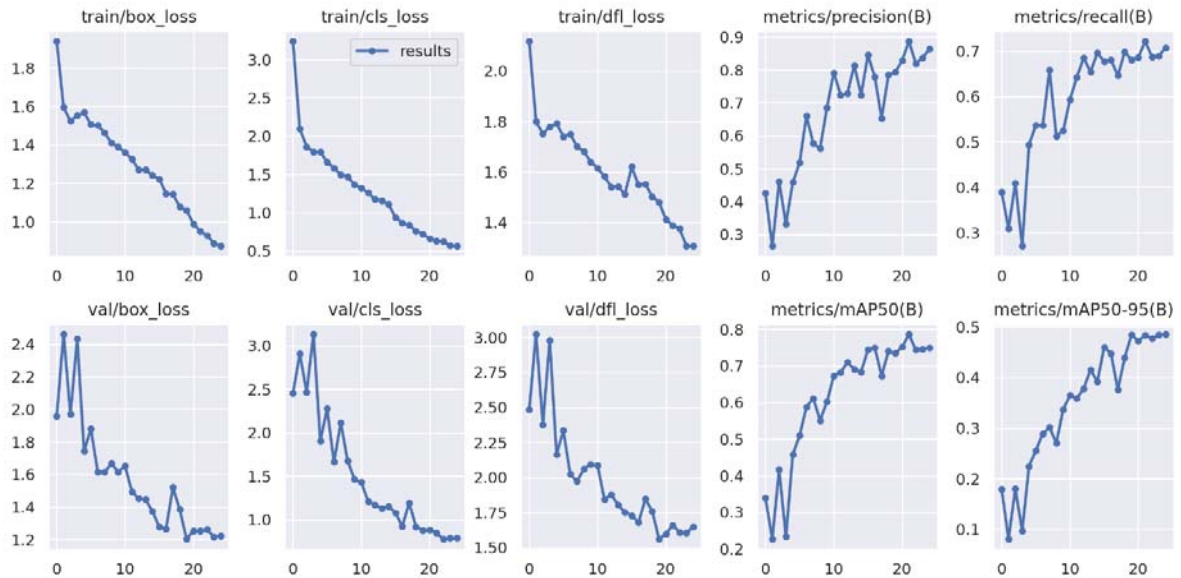


Рис. 2. Результаты оценок качества и функции потерь от периода обучения
 Fig. 2. Results of quality assessments and loss function from the study period

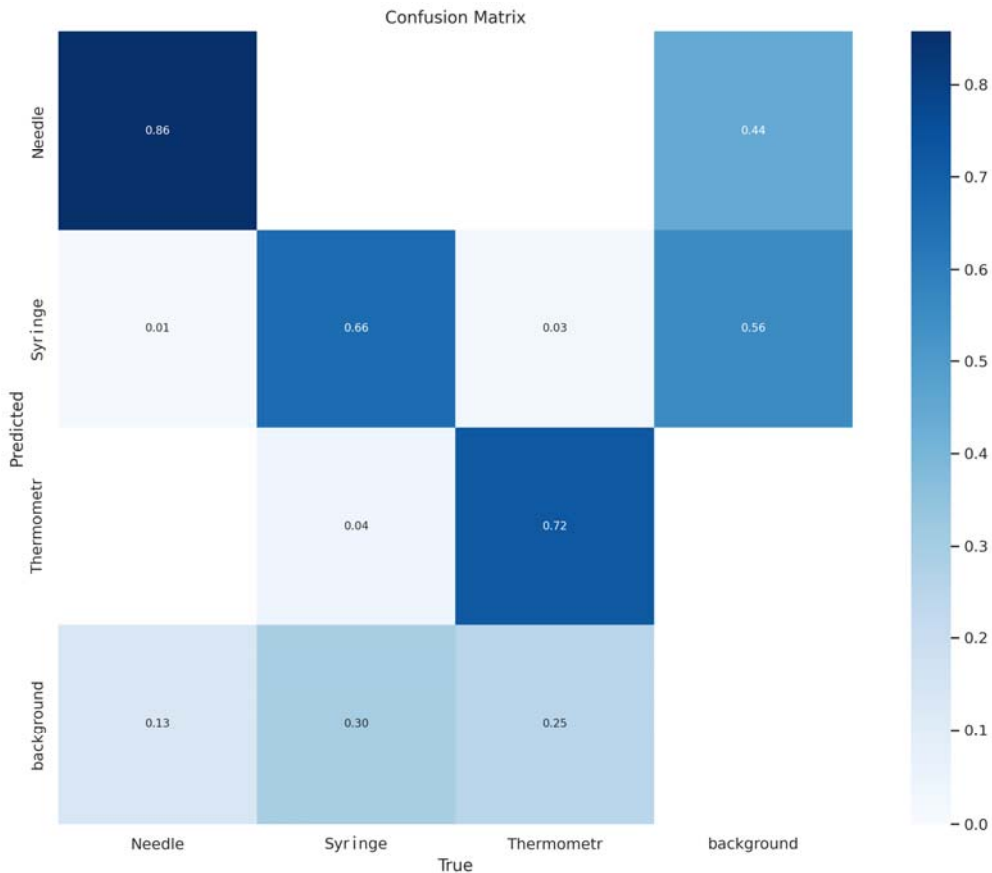


Рис. 3. Матрица оценки успешности классификации и ошибок
 Fig. 3. Evaluation matrix for classification success and errors

Определение взаимного расположения объектов

Существует несколько видов инъекций [3], иллюстрация которых представлена на рисунке 4.

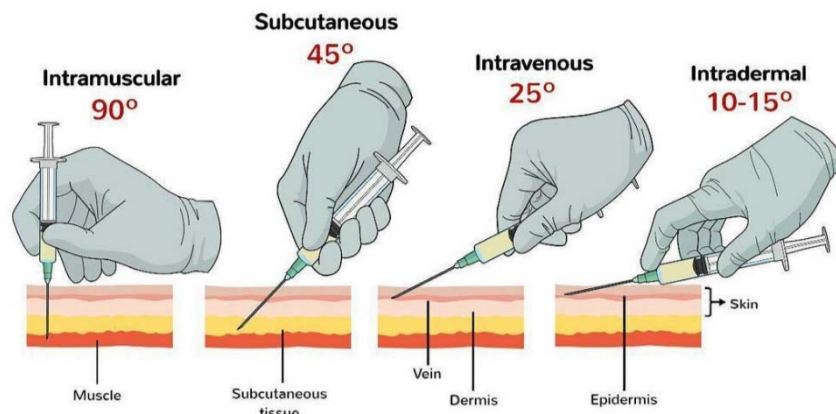


Рис. 4. Виды инъекций [3]

Fig. 4. Types of injections

Как можно заметить, угол измеряется относительно той части тела, куда производится ввод шприца, в нашем случае это рука.

Используя модель POSE, можно выделить ключевые точки тела человека. Фильтруя результаты по индексам, получаем координаты локтя и запястья. Это и будет основанием нашего угла (табл. 2).

Таблица 2

Индексы ключевых точек руки

Table 2. Indexes of key points of the hand

Индекс	Ключевая точка
7	Левый локоть
8	Правый локоть
9	Левое запястье
10	Правое запястье

Для детекции шприца используется ранее полученная модель, которая в случае обнаружения вернет координаты центра данного объекта. Однако одной ключевой точки для измерения угла недостаточно, отсюда и возникает необходимость обнаружения центра инъекционной иглы. Пример неправильного положения шприца представлен на рисунке 5.

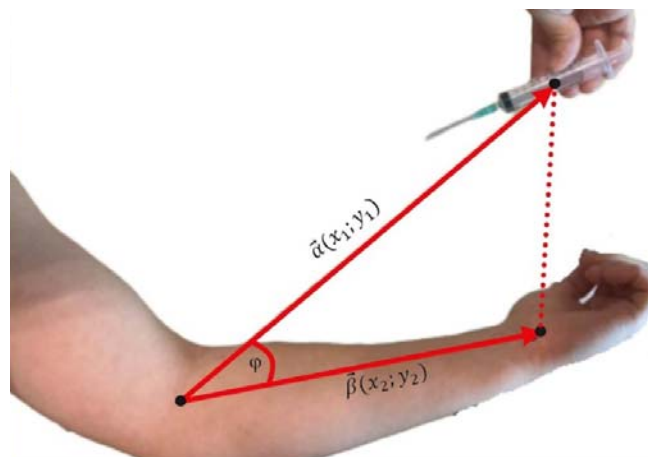


Рис. 5. Пример неправильного положения шприца

Fig. 5. Example of incorrect syringe position

В дальнейшем можно классифицировать инъекционные иглы и шприцы, так как существует большое число разновидностей, предназначенных для конкретных задач. Это поможет определить область применения данных инструментов и гарантировать их правильный выбор.

После того, как нами были получены все необходимые данные, производим расчет двух углов, чтобы убедиться, что шприц и игла находятся в правильном положении и направлены в одну точку. В случае, если углы не совпадают, мы можем предпринять дополнительные меры, такие как корректировка положения прибора относительно нашей руки. Визуализация правильного определения угла продемонстрирована на рисунке 6.

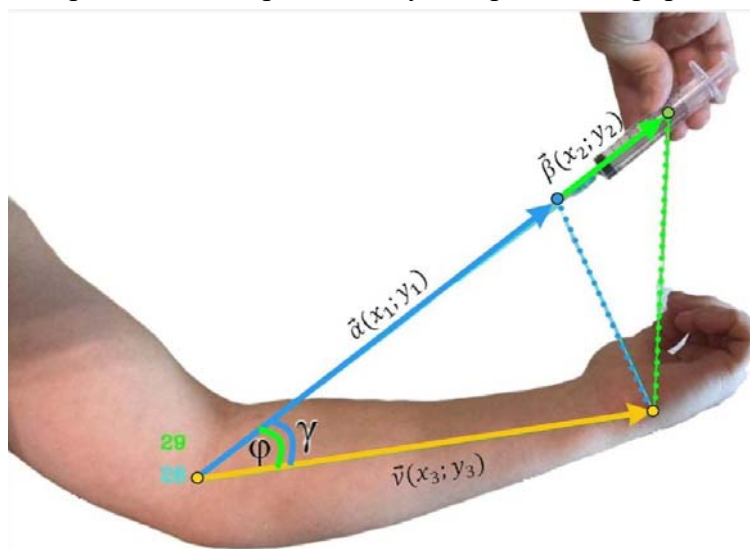


Рис. 6. Визуализация правильного определения угла

Fig. 6. Visualization of correct angle determination

Расчет углов производится на основании формулы скалярного произведения векторов:

$$\vec{a} \cdot \vec{v} = |\vec{a}| \cdot |\vec{v}| \cdot \cos \gamma . \quad (2)$$

Откуда получаем формулу для нахождения угла между векторами:

$$\gamma = \arccos \left(\frac{\vec{a} \cdot \vec{v}}{|\vec{a}| \cdot |\vec{v}|} \right) . \quad (3)$$

Предлагаемый алгоритм безусловно позволит правильно вычислить угол между объектами, однако для максимально точной работы требуется подтверждение того, что угол между иглой и прямой, соединяющей точку ввода и точку на запястье, должен составлять 0° , как показано на рисунке 7. Эту потребность можно удовлетворить путем анализа данных, полученных с дополнительной камерой, которая находится над исследуемыми объектами.

Также необходимо дообучить модель на более объемном наборе данных, учитывая результаты прогнозирования предыдущего обучения, для повышения точности детекции.

Достоинством использования YOLO является возможность экспорта моделей в различные форматы. Так, OpenVINO (Open Visual Inference & Neural Network Optimization) обеспечивает ускорение вычислений на процессоре, а использование технологии CUDA (Compute Unified Device Architecture) позволяет увеличить вычислительную производительность благодаря использованию графических процессоров Nvidia, что играет важную роль при детекции в режиме реального времени.

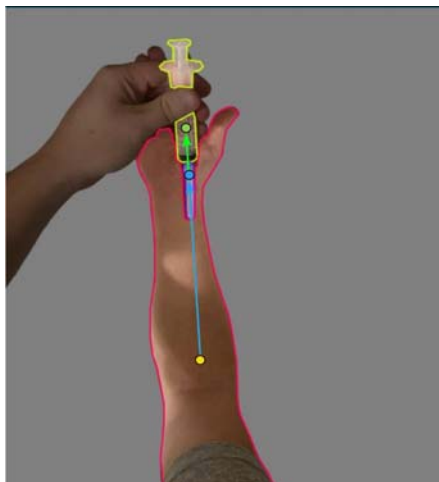


Рис. 7. Пример использования дополнительной камеры
Fig. 7. Example of using an additional camera

Заключение

В статье представлен разработанный авторами алгоритм распознавания объектов с возможностью определения характеристик их взаимного расположения, на базе модели YOLOv8-POSE. Продемонстрирована его эффективность при расчете угла ввода инъекции. Алгоритм ориентирован на улучшение процесса обучения не только медработников, но и военнослужащих, не имеющих должного медицинского опыта и практики.

Примечания

1. Tomato 3D pose detection algorithm based on keypoint detection and point cloud processing / Du Xiaoqiang, Meng Zhichao, Ma Zenghong, Lu Wenwu // Computers and Electronics in Agriculture. 2023. Vol. 212. P. 108056. DOI: <https://doi.org/10.1016/j.compag.2023.108056>
2. Qizhe Lu, Yicheng Jing, Xuefeng Zhao. Bolt Loosening Detection Using Key-Point Detection Enhanced by Synthetic Datasets // Applied Sciences. 2023. Vol. 13, No. 3. P. 2020. DOI: <https://doi.org/10.3390/app13032020>
3. Алгоритмы выполнения инъекций : метод. рекомендации / сост. А. А. Черемисина, Е. Н. Казакова, В. Г. Камалутдинова [и др.]; Фармацевтический колледж. Красноярск : КрасГМУ, 2013. 30 с.
4. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks / Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. Vol. 39, Iss. 6. P. 1137–1149. DOI: <https://doi.org/10.1109/TPAMI.2016.2577031>
5. You Only Look Once: Unified, Real-Time Object Detection // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) / J. Redmon, S. Divvala, R. Girshick, A. Farhadi. 2016. P. 779–788. DOI: <https://doi.org/10.1109/CVPR.2016.91>
6. Ступина А. А. Исследование возможности распознавания животных в искусственной среде. 2023. URL: https://www.researchgate.net/publication/376982240_Issledovanie_vozmoznosti_raspoznvania_zivotnyh_v_iskusstvennoj_srede

References

1. Tomato 3D pose detection algorithm based on keypoint detection and point cloud processing / Du Xiaoqiang, Meng Zhichao, Ma Zenghong, Lu Wenwu // Computers and Electronics in Agriculture. 2023. Vol. 212. P. 108056. DOI: <https://doi.org/10.1016/j.compag.2023.108056>
2. Qizhe Lu, Yicheng Jing, Xuefeng Zhao. Bolt Loosening Detection Using Key-Point Detection Enhanced by Synthetic Datasets // Applied Sciences. 2023. Vol. 13, No. 3. P. 2020. DOI: <https://doi.org/10.3390/app13032020>

3. Algorithms for performing injections : method. recommendations / comp. A. A. Chermisina, E. N. Kazakova, V. G. Kamalutdinova [et al.]; College of Pharmacy. Krasnoyarsk : KrasGMU, 2013. 30 p.

4. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks / Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017. Vol. 39, Iss. 6. P. 1137–1149.
DOI: <https://doi.org/10.1109/TPAMI.2016.2577031>

5. You Only Look Once: Unified, Real-Time Object Detection // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) / J. Redmon, S. Divvala, R. Girshick, A. Farhadi. 2016. P. 779–788. DOI: <https://doi.org/10.1109/CVPR.2016.91>

6. Stupina A. A. Exploring the possibility of recognizing animals in an artificial environment. 2023.

URL: https://www.researchgate.net/publication/376982240_Issledovanie_vozmozhnosti_raspoznavania_zivotnyh_v_iskusstvennoj_srede

Авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию 12.04.2024; одобрена после рецензирования 25.04.2024; принята к публикации 26.04.2024.

The authors declare no conflicts of interests.

The article was submitted 12.04.2024; approved after reviewing 25.04.2024; accepted for publication 26.04.2024.

© М. В. Алиев, Д. А. Бербенцев, В. О. Немькин, С. М. Алиева, 2024