

DOI: <https://doi.org/10.17816/DD430355>

# Подготовка набора данных компьютерной томографии органов брюшной полости пациентов с аневризмой абдоминального отдела аорты

М.Р. Коденко<sup>1, 2</sup>, Т.А. Макарова<sup>3</sup><sup>1</sup> Научно-практический клинический центр диагностики и телемедицинских технологий, Москва, Российская Федерация<sup>2</sup> Московский государственный технический университет имени Н.Э. Баумана, Москва, Российская Федерация<sup>3</sup> Российская медицинская академия непрерывного профессионального образования, Москва, Российская Федерация

## АННОТАЦИЯ

**Обоснование:** технологии искусственного интеллекта (ИИ) активно внедряют в обработку и анализ диагностических медицинских изображений. Точность и надёжность алгоритмов ИИ определяются объёмом и качеством обучающих наборов данных. В настоящее время существует потребность в увеличении наборов данных открытого доступа, в частности КТ-ангиографических исследований брюшной аорты (КТА). Ограничения существующих наборов данных КТА брюшной аорты: бинарная разметка (классификация исследования в целом), малое количество исследований. Кроме того, большинство исследований не содержит признаков патологии аорты, что, с учётом вариабельности данной патологии, значительно ограничивает их использование для обучения ИИ, так как целевой задачей таких алгоритмов является обнаружение патологии.

**Цель:** подготовка набора данных КТА пациентов с аневризмой брюшного отдела аорты.

**Методы:** на примере набора данных КТА-исследований с аневризмой брюшного отдела аорты рассмотрены этапы и особенности создания набора данных для обучения ИИ в соответствии с методическими рекомендациями. На основе базовых диагностических требований для выбранной клинической задачи было сформировано техническое задание на подготовку набора данных, рассчитан необходимый объём выборки и определён оптимальный сценарий разметки. На следующем этапе был выполнен отбор исходных данных КТ органов брюшной полости в Единой радиологической информационной системе, проведены анонимизация данных и полуавтоматическая разметка области интереса (стенка и русло аорты) с применением инструмента 3D Slicer и её верификация экспертом-рентгенологом, документирование промежуточных результатов.

**Результаты:** рассчитанный объём выборки составил 100 исследований, содержащих артериальную фазу исследования, толщина среза не более 1,2 мм, баланс классов «норма»: «патология» был выбран 1:4. Проведена частичная разметка (50% исследований) данных.

**Заключение:** разработана методология для подготовки наборов данных КТА. Сформированный набор данных, при соблюдении необходимых процедур, будет выложен в открытый доступ и может быть использован для обучения и тестирования алгоритмов ИИ, проведения научных исследований.

**Ключевые слова:** искусственный интеллект; набор данных; КТА; аневризма брюшной аорты.

## КАК ЦИТИРОВАТЬ

Коденко М.Р., Макарова Т.А. Подготовка набора данных компьютерной томографии органов брюшной полости пациентов с аневризмой абдоминального отдела аорты // *Digital Diagnostics*. 2023. Т. 4, № 1 Supplement. С. 90–92. DOI: <https://doi.org/10.17816/DD430355>

## СПИСОК ЛИТЕРАТУРЫ

1. Павлов Н.А., Андрейченко А.Е., Владимировский А.В., и др. Эталонные медицинские датасеты (MosMedData) для независимой внешней оценки алгоритмов на основе искусственного интеллекта в диагностике // *Digital Diagnostics*. 2021. Т. 2, № 1. С. 49–65. doi: <https://doi.org/10.17816/DD60635>
2. Центр диагностики и телемедицины [интернет]. Набор данных для селф-теста диагностического MosMedData КТ с признаками аневризмы аорты тип III. Дата обращения: 05.06.2023. Доступ по ссылке: <https://mosmed.ai/datasets/mosmeddata-kt-s-priznakami-anevrizmi-aorti-tip-iii>.

Рукопись получена: 15.05.2023

Рукопись одобрена: 05.06.2023

Опубликована Online: 10.07.2023

3. Aorta segmentation [Internet]. Дата обращения: 05.06.2023. Доступ по ссылке: <https://www.kaggle.com/datasets/licethyaneth/aorta-segmentation>.
4. Регламент подготовки наборов данных с описанием подходов к формированию репрезентативной выборки данных. Ч. 1 : методические рекомендации : [препринт] / сост. С.П. Морозов, А.В. Владимирский, А.Е. Андрейченко, и др. Москва : ГБУЗ «НПКЦ ДиТ ДЗМ», 2021. 40 с. (Лучшие практики лучевой и инструментальной диагностики ; вып. 103).
5. Базовые диагностические требования к результатам работы ИИ [интернет]. Дата обращения: 05.06.2023. Доступ по ссылке: <https://mosmed.ai/ai/docs/>.
6. Homeyer A., Geißler C., Schwen L.O., et al. Recommendations on compiling test datasets for evaluating artificial intelligence solutions in pathology // *Mod Pathol*. 2022. Vol. 35, N 12. P. 1759–1769. doi: 10.1038/s41379-022-01147-y

DOI: <https://doi.org/10.17816/DD430355>

## Preparation of abdominal computed tomography data set for patients with abdominal aortic aneurysm

Maria R. Kodenko<sup>1,2</sup>, Tatiana A. Makarova<sup>3</sup>

<sup>1</sup> Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

<sup>2</sup> Bauman Moscow State Technical University, Moscow, Russian Federation

<sup>3</sup> Russian Medical Academy of Continuous Professional Education, Moscow, Russian Federation

### ABSTRACT

**BACKGROUND:** Artificial intelligence (AI) technologies are actively implemented in the processing and analysis of diagnostic medical images. The accuracy and reliability of AI algorithms are determined by the amount and quality of training data sets. Currently, a need exists for increased open access data sets, particularly abdominal aortic CT angiographic studies (CTA). Limitations of existing abdominal aortic CTA data sets are binary labeling (classification of the entire study) and small number of examinations. In addition, most examinations do not contain signs of aortic pathology, which, given its variability, significantly limits their use for AI training, since the target of such algorithms is the detection of pathology.

**AIM:** To prepare a CTA data set for patients with abdominal aortic aneurysm.

**METHODS:** Using the CTA data set with signs of abdominal aortic aneurysm, the stages and features of data set creation for AI training in accordance with the methodological recommendations were considered. Given the basic diagnostic requirements for the selected clinical task, the terms of reference for the preparation of the data set were formed, the required sample size was calculated, and the optimal annotation scenario was determined. The next stage included the selection of initial CT data of abdominal organs in the Unified Radiology Information System, anonymization of data, semi-automatic labeling and of the area of interest (aortic wall and aortic bed) using the 3D Slicer tool and its verification by an examining radiologist, and documentation of intermediate results.

**RESULTS:** The calculated sample volume included 100 scans, containing the arterial phase, with a slice thickness of up to 1.2 mm. The balance of “normal vs. pathology” classes was chosen to be 1:4. Partial annotation of the data (50%) was performed.

**CONCLUSIONS:** A methodology for preparing CTA data sets was developed. The generated dataset, if the necessary procedures are followed, will be placed in the public domain and may be used for training and testing AI algorithms and conducting scientific research.

**Keywords:** artificial intelligence; dataset; CTA; abdominal aortic aneurysm.

### FOR CITATION

Kodenko MR, Makarova TA. Preparation of abdominal computed tomography data set for patients with abdominal aortic aneurysm. *Digital Diagnostics*. 2023;4(1S):90–92. DOI: <https://doi.org/10.17816/DD430355>

Received: 15.05.2023

Accepted: 05.06.2023

Published Online: 10.07.2023

## REFERENCES

1. Pavlov NA, Andreychenko AE, Vladzimirskiy AV, et al. Reference medical datasets (MosMedData) for independent external evaluation of algorithms based on artificial intelligence in diagnostics. *Digital Diagnostics*. 2021;2(1):49–65. (In Russ). doi: <https://doi.org/10.17816/DD60635>
2. *Tsentr diagnostiki i telemeditsiny [Internet]. Nabor dannykh dlya self-testa diagnosticheskogo MosMedData KT s priznakami anevrizmy aorty tip III* [cited 2023 Jun 05]. Available from: <https://mosmed.ai/datasets/mosmeddata-kt-s-priznakami-anevrizmi-aorti-tip-iii>. (In Russ).
3. *Aorta segmentation* [Internet] [cited 2023 Jun 05]. Available from: <https://www.kaggle.com/datasets/licethyaneth/aorta-segmentation>.
4. *Reglament podgotovki naborov dannykh s opisaniem podkhodov k formirovaniyu reprezentativnoi vyborki dannykh*. P. 1: metodicheskie rekomendatsii: [preprint]. Morozov SP, Vladzimirskii AV, Andreichenko AE, et al. Moscow: GBUZ “NPKTs DiT DZM”, 2021. 40 p. (Luchshie praktiki luchevoi i instrumental'noi diagnostiki; iss. 103). (In Russ).
5. *Bazovye diagnosticheskie trebovaniya k rezul'tatam raboty II* [Internet] [cited 2023 Jun 05]. Available from: <https://mosmed.ai/docs/>. (In Russ).
6. Homeyer A, Geißler C, Schwen LO, et al. Recommendations on compiling test datasets for evaluating artificial intelligence solutions in pathology. *Mod Pathol*. 2022;35(12):1759–1769. doi: 10.1038/s41379-022-01147-y

## ИНФОРМАЦИЯ ОБ АВТОРАХ

**Коденко Мария Романовна;**

ORCID: <https://orcid.org/0000-0002-0166-3768>;  
eLibrary SPIN: 5789-0319; e-mail: [m.r.kodenko@yandex.ru](mailto:m.r.kodenko@yandex.ru)

**Макарова Татьяна Андреевна;**

ORCID: <https://orcid.org/0000-0002-6274-5636>;  
e-mail: [makarova97.mail@gmail.com](mailto:makarova97.mail@gmail.com)

## AUTHORS' INFO

**Maria R. Kodenko;**

ORCID: <https://orcid.org/0000-0002-0166-3768>;  
eLibrary SPIN: 5789-0319; e-mail: [m.r.kodenko@yandex.ru](mailto:m.r.kodenko@yandex.ru)

**Tatiana A. Makarova;**

ORCID: <https://orcid.org/0000-0002-6274-5636>;  
e-mail: [makarova97.mail@gmail.com](mailto:makarova97.mail@gmail.com)

\* Автор, ответственный за переписку / Corresponding author