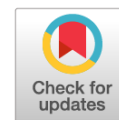


DOI: <https://doi.org/10.17816/DD501759>

Технологические дефекты программного обеспечения с искусственным интеллектом

В.В. Зинченко, К.М. Арзамасов, Е.И. Кремнева, А.В. Владзимирский, Ю.А. Васильев

Научно-практический клинический центр диагностики и телемедицины, Москва, Российская Федерация

АННОТАЦИЯ

Обоснование. Технологические дефекты в работе программного обеспечения с искусственным интеллектом являются критически важными при принятии решения о практической применимости и клинической ценности программного обеспечения с искусственным интеллектом.

Цель — анализ и систематизация технологических дефектов, возникающих при работе программного обеспечения с искусственным интеллектом для анализа медицинских изображений.

Материалы и методы. В рамках эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы проводится мониторинг технологических параметров для всех участвующих решений как на этапе апробации, так и на этапе опытной эксплуатации. В статье представлена графическая информация о среднем числе технологических дефектов для профилактического направления, модальность «Маммография», за 2021 год. Этот период выбран как наиболее показательный, характеризующийся активным развитием программного обеспечения с искусственным интеллектом с позиции увеличения технической стабильности их работы. С целью оценки применимости подхода по выявлению технологических дефектов аналогичный анализ проводился для направления обнаружения внутривенных кровоизлияний на компьютерных томограммах головного мозга за 2022–2023 годы.

Результаты. В ходе исследования было проанализировано программное обеспечение с искусственным интеллектом по модальностям «Маммография» (2 алгоритма) и «Компьютерная томография головного мозга» (1). Всего для модальности «Маммография» собрано 14 выборок по 20 исследований; для модальности «Компьютерная томография» — 12 выборок по 80 исследований. Для каждого типа дефекта были построены графики, а для каждой из модальностей были построены линии тренда. Коэффициенты уравнений линий трендов указывают на тенденцию к снижению числа технологических дефектов.

Заключение. Проведённый анализ позволяет проследить тенденцию к снижению числа технологических дефектов, что может свидетельствовать о доработке программного обеспечения с искусственным интеллектом и повышении его качества благодаря периодическому мониторингу. Кроме того, такой результат показывает универсальность использования как для профилактических методов, так и для экстренных.

Ключевые слова: искусственный интеллект; программное обеспечение с искусственным интеллектом; технологический мониторинг; технологические дефекты; Московский эксперимент.

Как цитировать:

Зинченко В.В., Арзамасов К.М., Кремнева Е.И., Владзимирский А.В., Васильев Ю.А. Технологические дефекты программного обеспечения с искусственным интеллектом // Digital Diagnostics. 2023. Т. 4, № 4 С. 593–604. DOI: <https://doi.org/10.17816/DD501759>

DOI: <https://doi.org/10.17816/DD501759>

Technological defects in software based on artificial intelligence

Viktoriya V. Zinchenko, Kirill M. Arzamasov, Elena I. Kremneva, Anton V. Vladzemyrskyy, Yuriy A. Vasilev

Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Moscow, Russian Federation

ABSTRACT

BACKGROUND: Technological defects in the use of artificial intelligence software are critical when deciding on the practical applicability and clinical value of artificial intelligence software.

AIM: To conduct an analysis and systematization of technological defects occurring when artificial intelligence software analyzes medical images.

MATERIALS AND METHODS: As part of the experiment on the use of innovative computer vision technologies for the analysis of medical images and further application in the Moscow healthcare system, technological parameters of all artificial intelligence software are monitored at the testing and operation stages of the trial. This article presents graphical information on the average number of technological defects in mass mammography screening in 2021. This period was chosen as the most indicative and characterized by the active development of artificial intelligence software and increased technical stability of its performance. To assess the applicability of the analysis for technological defects, a similar analysis was conducted for the direction of detection of intracranial hemorrhage on computed tomography scans of the brain for 2022–2023.

RESULTS: During the study, artificial intelligence software used for mammography (two algorithms) and brain computed tomography (one algorithm) were analyzed. Fourteen mammography samples were collected for technological monitoring during the identified period, each from 20 studies, and 12 brain computed tomography samples were obtained, each from 80 studies. Graphs were constructed for each type of defect, and trend lines were plotted for each modality. The coefficients of the trend line equations indicate a downward tendency in the number of technological defects.

CONCLUSION: This analysis allows tracing a downward trend in the number of technological defects, which may indicate a refinement of artificial intelligence software and an increase in its quality because of periodic monitoring. It also shows the versatility of use for both preventive and emergency methods.

Keywords: artificial intelligence; artificial intelligence software; technological monitoring; technological defects; Moscow experiment.

To cite this article:

Zinchenko VV, Arzamasov KM, Kremneva EI, Vladzemyrskyy AV, Vasilev YuA. Technological defects in software based on artificial intelligence. *Digital Diagnostics*. 2023;4(4):593–604. DOI: <https://doi.org/10.17816/DD501759>

Received: 20.06.2023

Accepted: 21.11.2023

Published online: 06.12.2023

DOI: <https://doi.org/10.17816/DD501759>

人工智能软件的技术缺陷

Viktoria V. Zinchenko, Kirill M. Arzamasov, Elena I. Kremneva, Anton V. Vladzmyrskyy, Yuriy A. Vasilev

Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies, Moscow, Russian Federation

简评

论证。人工智能软件性能方面的技术缺陷是确定人工智能软件实用性和临床价值的关键。

该研究的目的是对医学影像分析人工智能软件运行中的技术缺陷进行分析并使之系统化。

材料和方法。在莫斯科市进行了一项《使用创新计算机视觉技术进行医学图像分析并进一步应用于莫斯科市医疗系统的实验》。在实验框架内，对所有参与解决方案的技术参数进行监测。监测是在批准阶段和试运行阶段进行的。本文以图表形式介绍2021年“乳房摄影术”预防方向的平均技术缺陷数量。这一时期被选为最有意义的时期。这一时期的特点是从提高操作技术稳定性的角度出发，积极开发人工智能软件。为了评估该方法在发现技术缺陷方面的适用性，我们对2022–2023年脑部CT扫描颅内出血的检测方向进行了类似的分析。

结果。本研究分析了“乳房摄影术”（2种算法）和“脑计算机断层扫描”（1种）模式的人工智能软件。在“乳房X射线照相术”模式中，共收集了14个样本，共有20项研究。在“脑计算机断层扫描”模式中，共收集了12个样本，共有80项研究。我们对每种缺陷类型都绘制了图表，对每种模式绘制了趋势线。趋势线公式的系数表明了，技术缺陷的数量呈下降趋势。

结论。通过分析，我们发现了减少技术缺陷数量的趋势。这可能表明人工智能软件的完善，以及通过定期监测，软件质量的提升。此外，这一结果还显示使用预防和应急方法的通用性。

关键词：人工智能；人工智能软件；技术监测；技术缺陷；莫斯科实验。

引用本文：

Zinchenko VV, Arzamasov KM, Kremneva EI, Vladzmyrskyy AV, Vasilev YuA. 人工智能软件的技术缺陷. *Digital Diagnostics*. 2023;4(4):593–604. DOI: <https://doi.org/10.17816/DD501759>

收到: 20.06.2023

接受: 21.11.2023

发布日期: 06.12.2023

BACKGROUND

Artificial intelligence (AI)-based software can assist healthcare professionals (HCPs) with routine and complex tasks and improve the quality, accessibility, and speed of patient care [1–3]. This was largely possible because of the continuity of foreign and domestic experience with AI in healthcare [4–7] and the experiment on the use of innovative computer vision technologies for analysis of medical images and further application in the Moscow healthcare system (hereinafter referred to as “the Experiment”). The Experiment aimed to conduct a scientific study of the possibility of using medical decision support methods in the Moscow healthcare system based on data analysis using advanced innovative technologies. Requirements were developed for AI-based software results in 21 areas of diagnostic radiology. Currently, results obtained using >50 AI-based solutions are available for HCPs. More than 10 million studies have been processed at the end of September 2023.

The use of new technologies in healthcare requires mandatory compliance with safety regulations. Therefore, the development, deployment, and use of AI-based software should be monitored [8]. Furthermore, AI-based software requires special control during operation because it can produce biased results when used on a population other than the one used to train it [9,10].

Some tests are used as part of the Experiment to control the quality of processing AI-based software research. Self-testing is the first step, which is designed to understand the technological AI compatibility of the software and research (input) data submitted for processing. The next step is functional testing. It determines the presence of the declared AI-based software functions, including its performance. The AI-based software is evaluated from both technical and clinical perspectives by technical and medical experts. Calibration testing is the step of determining the performance metrics of AI-based software with the area under the ROC curve as the main indicator.

If all tests are successfully passed, the software is allowed to work with AI, and based on the work results, the technological and clinical monitoring of algorithms is performed. According to international studies, the technological tests (monitoring of technological parameters) are an integral part of product tests and are performed as part of comprehensive tests for the possible use in real clinical practice [11]. Therefore, this study focused on monitoring technological defects.

STUDY AIMS

This study aimed to evaluate the technological defects of AI-based software based on the results of the Experiment, analyze and statistically process them, and assess the impact on the safety and quality of AI-based software in clinical practice.

MATERIALS AND METHODS

Conditions of the study

For all studies analyzed by the AI-based software during the reporting period for the “mammography” modality in 2021, monitoring was performed in accordance with the categories of errors shown in Table 1 (left column, according to order no. 51 of the Moscow Healthcare Department; January 26, 2021) [12]:

- **Group A:** The time to analyze a study exceeds 6.5 min. The time limit was derived as the average time required to describe an AI-based software study to obtain results suitable for use by a radiologist.
- **Group B:** No results from the evaluated studies.
- **Group C:** The images included in the AI-based software results do not match those of the native (source) study (they are damaged). In rare cases, changing the metadata can change the settings when viewing studies, making it more difficult to visualize the original image.
- **Group D:** Incorrect operation of the declared AI-based software functions that complicates the HCP’s work or makes it impossible to perform with adequate quality, including cropping of images, changes in brightness/contrast, missing description of results, and missing markers of abnormalities.
- **Group E:** Other violations of the integrity and content of the study file, limiting its diagnostic interpretation, including off-target markings and AI-based software analysis based on incorrect anatomy.
- **Group F:** A modification of the original series of studies. In 2022, the errors were restructured. This was considered when processing the monitoring data for the CT modality (Table 1, right column).

Study Duration

Monitoring was performed monthly until the end of the use of the AI-based software in the Experiment. The reporting monitoring period is one calendar month. Based on the data from days 10 and 20 of each month, an interim report for monitoring of Group A and B errors was prepared and sent to the AI-based software manufacturer.

For samples of mammograms and brain CTs, the article provides information on defects from March to December 2021 and from May 2022 to May 2023, respectively. For different AI-based software, the frequency of monitoring is different owing to variations in the time of entry into the Experiment and time of improvement after receiving feedback.

Technological monitoring was performed by a group of experts, including technical specialists and radiologists, who received additional training in monitoring and instruction in working with specific AI-based software. Moreover, to report the performed monitoring, a unified internal reporting form

Table 1. Criteria correlation for technological defects in orders from the Moscow Department of Health

Technological defects according to order no. 51 of the Moscow Department of Health dated January 26, 2021 (for mammography data presented in the article)	Technological defects according to Order No. 160 of the Moscow Department of Health dated November 3, 2022 (restructured) (for the brain CT data presented in the article)
Group A: study analysis time >6.5 min	Group A: analysis time of one study >6.5 min
Group B: missing results of the studies evaluated	Group B: missing results of the studies evaluated
–	Group C: incorrect operation of the declared functions of AI-based software, which complicates the work of radiologist or makes it impossible to perform it with proper quality
D2: no additional series	C1: no additional series
D3: no DICOM SR	C2: no DICOM SR
D4: presence of two or more DICOM SRs	C3: presence of two or more DICOM SRs
D5: no name for AI-based software	C4: no name for AI-based software
D6: missing information about the AI-based software version	C5: missing information about the AI-based software version
–	Group D: defects related to the display of the image area
C1: images are cropped	D1: images of additional series are cropped
C2: brightness/contrast changed	D2: brightness/contrast of the additional series does not match the original image
C3: not all necessary images were evaluated	D3: not all necessary images were evaluated
D1: complete absence of AI-based software results	Excluded
D7: no warning label “For research/scientific use only”	D4: no warning label “For research/scientific use only”
D8: missing markings of abnormalities	F: defects related to clinical work
E1: inconsistent DICOM SR information and additional series	Excluded
F: change to the original study series	D5: change to the original study series
–	Group E: other violations of the integrity and content of the study file that limit its diagnostic interpretation, including
E2: off-target markings	E1: off-target markings
E3: incorrect anatomy, projection, or series were analyzed	E2: incorrect anatomy, projection, or series were analyzed

SR: structure report

and technological monitoring instructions were developed and used.

Statistical analysis

A pseudorandomly selected dataset (study sample) was used for testing during technology monitoring, with the following proportions considered: 25% of studies with no abnormalities detected by the AI-based software (no-abnormality group) and 75% of studies with abnormalities detected (abnormality group). Selected studies with AI-based software results were assessed for technological errors. The study was assigned to the abnormality group if it exceeded the optimal threshold set during testing; otherwise, it was classified as a no-abnormality study [13,14].

In 2021, the pseudorandom sample size of the Experiment was 20 studies per month. This was still the pilot phase of the project. The nomogram power level was 42.5%, with a statistical significance level of 0.05. The standard difference between the sample elements was 0.79 [15]. In a full-scale project after 2021, with the use of risk analysis, the sample

size was 80 studies (see the justification in the article by Chetverikov et al. [13]). These 80 exams formed the monthly sample for brain CT scans.

RESULTS

In total, 14 samples from 20 studies were used in technology monitoring for the mammography modality. From March to December 2021, the generated pseudorandom samples were sent monthly for testing to all working (not under development) AI-based software.

To evaluate the applicability of the method for identifying technological defects, a similar analysis of generated pseudorandom samples was performed for the brain CT modality for the detection of intracranial hemorrhage. From May 2022 to May 2023, 80 studies per month were submitted to test the AI-based software (12 samples of 80 studies in total).

To represent changes in technological defects over time, general statistics were used for all technological monitoring

results of the AI-based software from March to December 2021 (for the mammography modality) or from May 2022 to May 2023 (for the CT modality). The number of technological defects was calculated as a percentage of the total number of studies in the dataset.

Figure 1 shows the changes in the average number of technological defects for the mammography modality from March 2021 to December 2021. The y-axis represents the presence of defects (expressed as a percentage of the total number of studies in the sample), and the x-axis represents the reporting period in months. Figure 2 provides similar information for the brain CT modality (from May 2022 to May 2023).

The left column of Table 1 shows the defects of the mammography modality. As shown in Figure 1, at the beginning

of the study period, most errors belonged to groups C, D, and B. At the end of the study period, only Group C errors remained, although their percentage decreased significantly.

Moreover, the right column of Table 1 presents the errors of the CT modality. As shown in Figure 2, the percentage of errors in relation to the sample was lower for all the errors, except for Group B, for the brain CT modality for the detection of intracranial hemorrhage than for the other modalities analyzed. The percentage of Group D and E errors decreased, whereas Group B errors showed a wide range every month.

To quantify this trend, the corresponding trend lines were added. These were linear functions $k \times x + b$, where k indicates the slope of the approximation curve, i.e., it indicates a tendency to increase or decrease the number of defects,

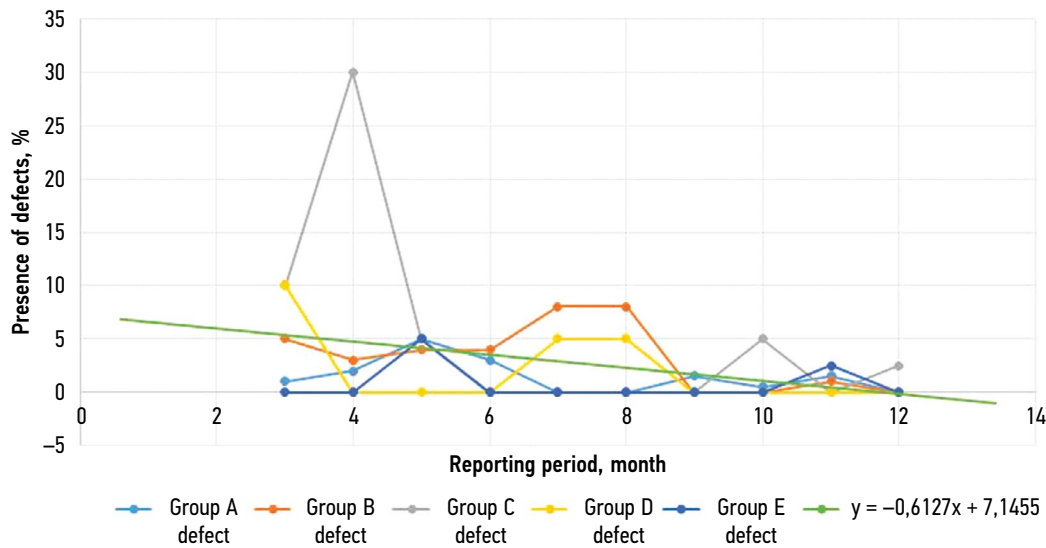


Fig. 1. Changes in detection of the average number of each technological defect for software based on artificial intelligence for mammography. Defects are divided into groups in accordance with order no. 51 of the Moscow Department of Health dated January 26, 2021.

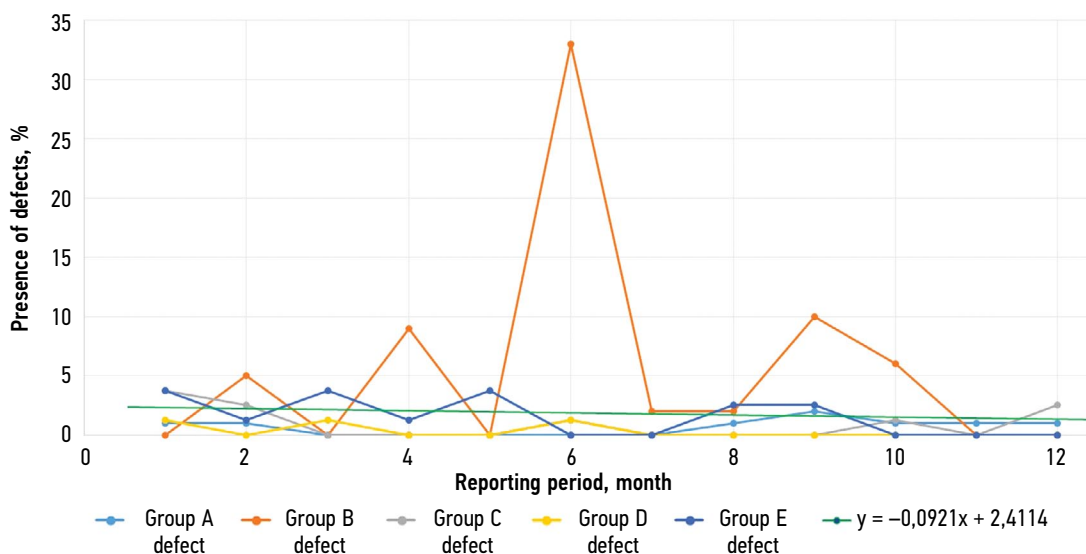


Fig. 2. Changes in the detection of the average number of each technological defect for software based on artificial intelligence for the brain computed tomography modality (presence or absence of intracranial hemorrhage). Defects are divided into groups in accordance with order no. 160 of the Moscow Department of Health dated November 3, 2022.

and b corresponds to the number of defects at the beginning of the monitoring. The approximation was performed for all AI-based software for individual modalities, and the entire data set was approximated at once (Figures 1 and 2). It is possible to predict changes in technological defect removal for each AI-based software product separately or for the entire range if k is known.

Figures 3, 4, 5, and 6 show examples of technological errors in AI-based software.

DISCUSSION

According to the results obtained and evaluated, the mammography modality shows an excellent trend in reducing

the number of technological defects (Figure 1, trend line). The AI-based software with the brain CT modality has a more uniform trend of decreasing the number of technological defects (Figure 2, trend line), despite the values of Group B defects. The reason for this is that the fluctuation of some technological defects is related to the automatic detection, fast feedback, and prompt improvement of the AI-based software by the manufacturer (version change or bug fix).

The AI-based software versions for the mammography modality were changed in September–October 2021, and the average number of Group B and D defects decreased (Figure 1). This may indicate the successful maintenance of AI-based software, which in turn may indicate the effective use of the presented methodology for monitoring technology.

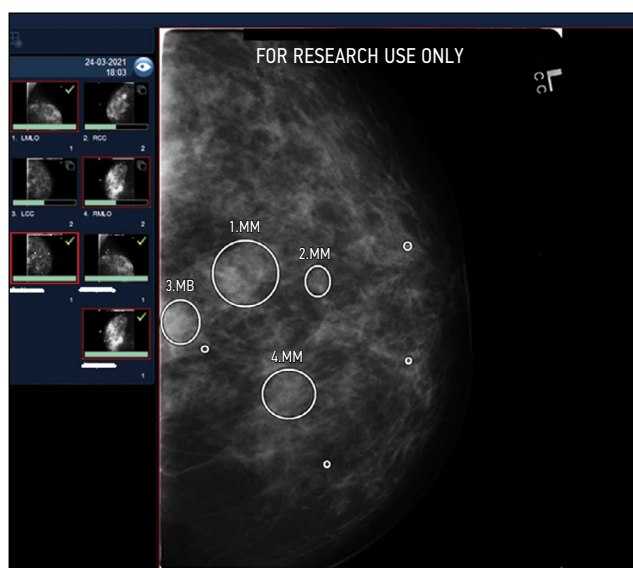


Fig. 3. Defect: not all necessary images have been evaluated. Modality: mammography.



Fig. 4. Defect: off-target markings; Modality: mammography.

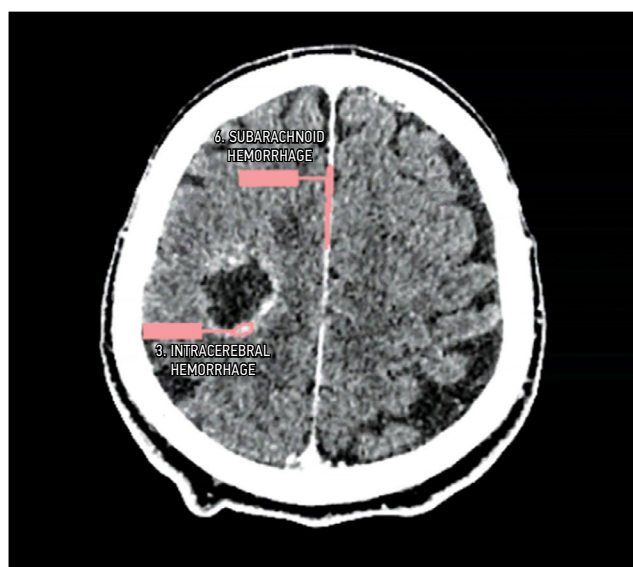


Fig. 5. Defect: an incorrect series was evaluated (contrast-enhanced computed tomography instead of native on). Modality: computed tomography.

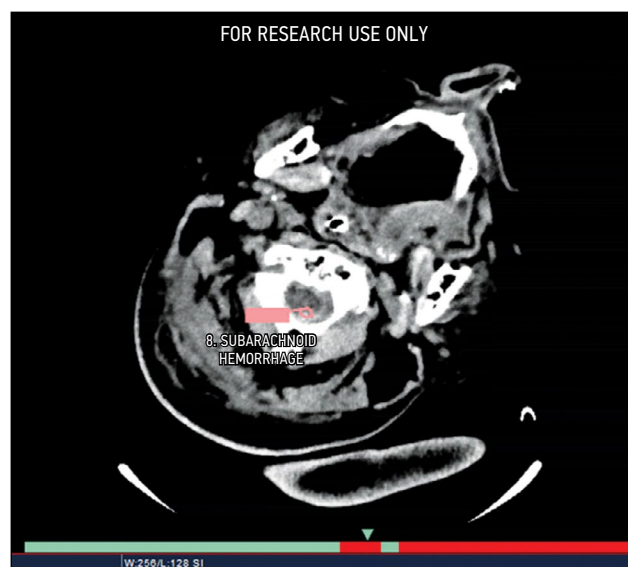


Fig. 6. Defect: off-target markings, contrast-enhanced computed tomography instead of native computed tomography. Modality: computed tomography.

Note that identifying technological defects within technological monitoring may be crucial in comprehensive testing aimed at safer, higher quality, and more efficient software operation using AI not only in diagnostic radiology but also in general healthcare. The analysis showed that the quality of AI-based software increases as the number of defects decreases. Therefore, AI-based software gains more trust from users, minimizes defects, and helps HCPs [16,17].

Restructuring technological defects

In 2022, the groups of technological defects were restructured based on the results of the monitoring of technological defects and their analysis presented in this article. According to the updated group classification (Table 1, right column), AI-based software defects for the brain CT modality were monitored for the presence or absence of intracranial hemorrhage. Group A and B defects were reviewed automatically, whereas Group C, D, and E defects required manual review by experts. The updated list of technological defects is presented in order no. 160 of the Moscow Healthcare Department dated November 3, 2022, which is still valid [18]. The reasons for increasing the number of studies in the sample have been discussed by Chetverikov et al. [13]. Such restructuring of technological defects optimized the work of experts analyzing the AI-based software monitoring results.

In addition, based on the results of technological monitoring of AI-based software under experimental

conditions, technological defects in accordance with the order dated 2021 for the mammography modality can be divided into three groups regarding the safety of AI-based software as a medical product:

- **Defects that affect the safety of patients and work of HCPs:** failure to implement functions declared by the manufacturer; comments that influence a radiologist or complicate their work; and irreversible damage to original research data. This group includes, for example, Group D (D2, D3, D4) and F defects. Separately, a D7 defect (absence of a warning label "For research/scientific use only") should be considered. This defect can only occur in the research setting and can never occur when using AI-based software as a medical device.
- **Defects that do not affect the safety of patients but affect the work of HCPs:** functional defects that do not conform to generally accepted standards for the presentation of research interpretation results. This group includes Group E and C (C1, C2, C3) defects.
- **Defects that do not affect the safety of patients or the work of HCPs:** minor defects that need to be removed to make the work of HCPs more convenient, intuitive, and efficient. This group includes D5, D6, and D8 defects.

For the CT modality, because of the restructuring of the defects (Table 1), three safety subgroups were presented from November 2023 until present, according to the 2021 order:

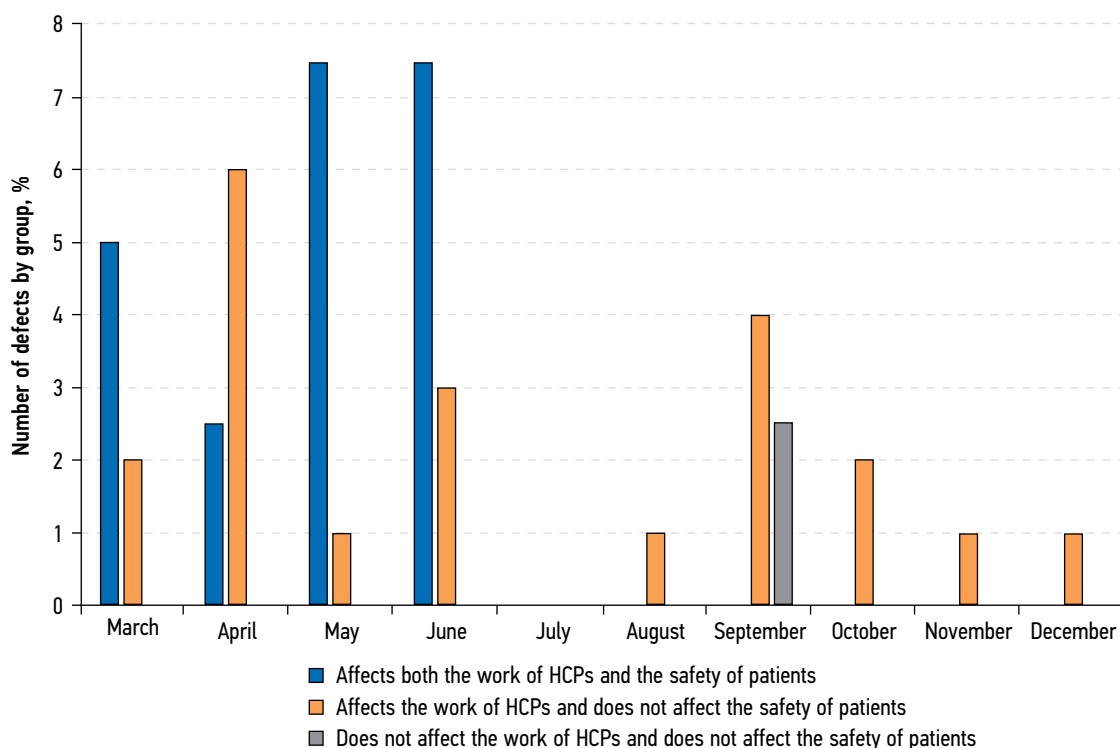


Fig. 7. Number of defects in each group over time; modality: mammography.

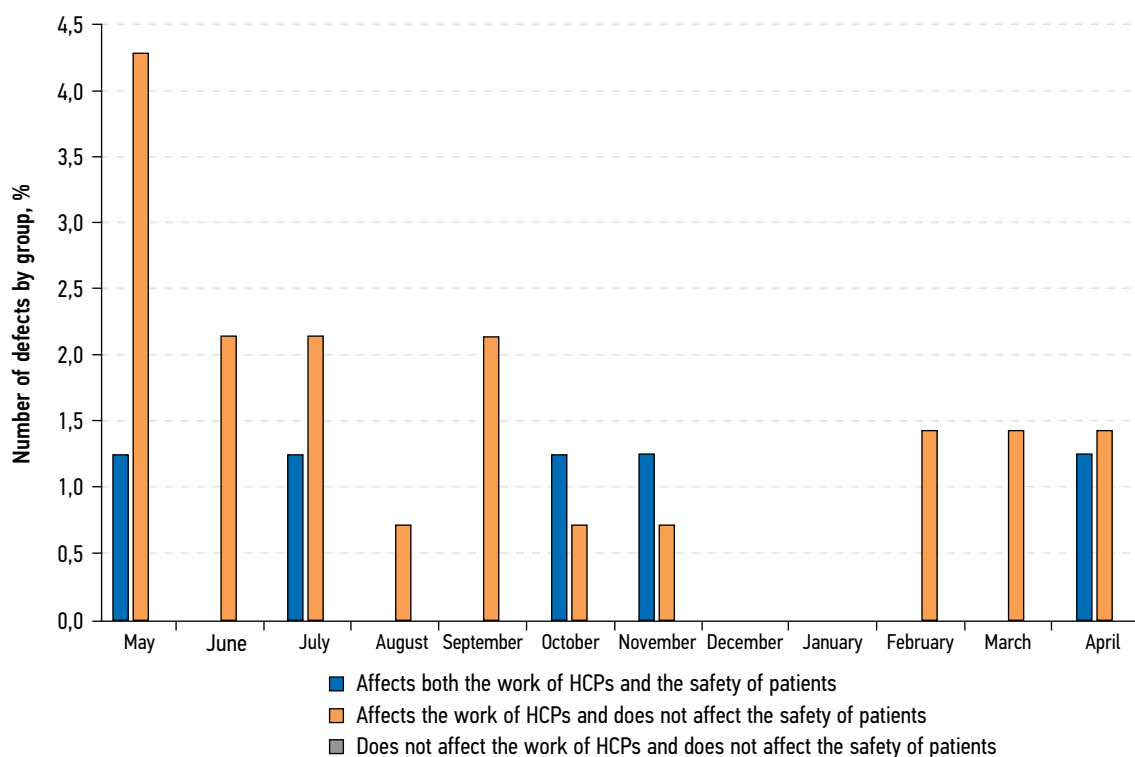


Fig. 8. Number of defects in each group over time; modality: computed tomography.

- **Defects that affect the safety of patients and work of HCPs:** Group C defects (C1, C2, C3) and D4 and D5 defects.
- **Defects that do not affect the safety of patients but affect the work of HCPs:** Group E and D defects (D1, D2, D3).
- **Defects that do not affect the safety of patients and work of HCPs:** C4 and C5 defects.

Figures 7 and 8 show graphical information on the changes in the number of errors by group and month for both modalities.

For the mammography modality (Figure 7), defects that affect patient safety and HCP work were no longer detected after June because of the AI-based software update. Furthermore, defects that affect HCP work but do not affect patient safety tend to decrease by the end of the study period.

For the brain CT modality, the most common defects (those that affect the HCP work but do not affect the patient safety) do not show a clear downward trend.

The methodology presented in the present study allows the monitoring of the technical stability of algorithms. This is of great practical importance when evaluating AI-based software and ensuring its safety. The methodology used to monitor the operation of AI-based software on the stream allowed identifying technological defects and improving solutions, which ultimately led to increasing the technological stability of AI-based software, as shown in the example of brain CT analysis data. Therefore, the

developed methodology proved to be an effective and universal tool for increasing the technical stability of AI-based software.

CONCLUSION

This study presents a list of the main technological defects that occur when implementing AI-based software, as well as a methodology for monitoring technological defects based on regular random control testing, which increases the technical stability of AI-based software. The developed AI-based software testing methodology for identifying technological defects is presented as part of monitoring the safety, quality, and efficiency of AI-based software testing in real world clinical practice.

ADDITIONAL INFORMATION

Funding source. Analysis of technological defects in computed tomography dataset with or without intracranial bleeding was funded by Russian Science Foundation Grant № 22-25-20231, <https://rscf.ru/project/22-25-20231/>.

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

Authors' contribution. All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work. The contribution is distributed as follows:

V.V. Zinchenko — structuring and analysis of the results obtained (mammography modality), writing the manuscript of the article; K.M. Arzamasov — obtaining technological monitoring data, analyzing the results obtained, correcting the manuscript of the article; E.I. Kremneva — structuring and analysis of the results

obtained (computed tomography modality), writing the manuscript of the article; A.V. Vladzimirskiy — review of the manuscript of the article, formation of the research hypothesis; Yu.A. Vasilev — formation of the research hypothesis, general management of the research.

REFERENCES

1. Vladzimirskii AV, Vasil'ev YuA, Arzamasov KM, et al. *Computer Vision in Radiologic Diagnostics: the First Stage of Moscow experiment*. Vasil'ev YuA, Vladzimirskii AV, editors. Publishing solutions; 2022. (In Russ).
2. Ranschaert ER, Morozov S, Algra PR, editors. *Artificial Intelligence in Medical Imaging*. Berlin: Springer; 2019. doi: 10.1007/978-3-319-94878-2
3. Gusev AV, Dobridnyuk SL. Artificial intelligence in medicine and healthcare. *Information Society Journal*. 2017;(4-5):78–93. (In Russ).
4. Shutov DV, Sharova DE, Abuladze LR, Drozdov DV. Artificial intelligence in clinical physiology: How to improve learning agility. *Digital Diagnostics*. 2023;4(1):81–88. doi: 10.17816/DD123559
5. Meldo AA, Utkin LV, Trofimova TN. Artificial intelligence in medicine: current state and main directions of development of the intellectual diagnostics. *Diagnostic radiology and radiotherapy*. 2020;11(1):9–17. doi: 10.22328/2079-5343-2020-11-1-9-17
6. Recht MP, Dewey M, Dreyer K, et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. *European radiology*. 2020;30(6):3576–3584. doi: 10.1007/s00330-020-06672-5
7. Larson DB, Harvey H, Rubin DL, et al. Regulatory Frameworks for Development and Evaluation of Artificial Intelligence-Based Diagnostic Imaging Algorithms: Summary and Recommendations. *Journal of the American College of Radiology*. 2021;18(3 Pt A):413–424. doi: 10.1016/j.jacr.2020.09.060
8. Zinchenko V, Chetverikov S, Ahmad E, et al. Changes in software as a medical device based on artificial intelligence technologies. *International Journal of Computer Assisted Radiology and Surgery*. 2022;17:1969–1977. doi: 10.1007/s11548-022-02669-1
9. Nomura Y, Miki S, Hayashi N, et al. Novel platform for development, training, and validation of computer-assisted detection/diagnosis software. *International Journal of Computer Assisted Radiology and Surgery*. 2020;15(4):661–672. doi: 10.1007/s11548-020-02132-z
10. Methodological recommendations on the procedure for expert examination of quality, efficiency and safety of medical devices (in terms of software) for state registration under the national system FGBU «VNIIMT» Roszdravnadzor. Moscow; 2021. (In Russ).
11. Pemberton HG, Zaki LAM, Goodkin O, et al. Technical and clinical validation of commercial automated volumetric MRI tools for dementia diagnosis — a systematic review. *Neuroradiology*. 2021;63:1773–1789. doi: 10.1007/s00234-021-02746-3
12. Order of the Moscow City Health Department No. 51 dated 26.01.2021 «On approval of the procedure and conditions for conducting an experiment on the use of innovative technologies in the field of computer vision to analyze medical images and further application in the health care system of the city of Moscow in 2021». (In Russ).
13. Chetverikov SF, Arzamasov KM, Andreichenko AE, et al. Approaches to Sampling for Quality Control of Artificial Intelligence in Biomedical Research. *Modern Technologies in Medicine*. 2023;15(2):19. doi: 10.17691/stm2023.15.2.02
14. Zinchenko VV, Arzamasov KM, Chetverikov SF, et al. Methodology for Conducting Post-Marketing Surveillance of Software as a Medical Device Based on Artificial Intelligence Technologies. *Modern Technologies in Medicine*. 2022;14(5):15–25. doi: 10.17691/stm2022.14.5.02
15. Altman DG. Statistics and ethics in medical research: III How large a sample? *British medical journal*. 1980;281(6251):1336. doi: 10.1136/bmj.281.6251.1336
16. Tyrov IA, Vasilyev YuA, Arzamasov KM, et al. Assessment of the maturity of artificial intelligence technologies for healthcare: methodology and its application based on the use of innovative computer vision technologies for medical image analysis and subsequent applicability in the healthcare system of Moscow. *Medical Doctor and IT*. 2022;4(4):76–92. doi: 10.25881/18110193_2022_4_76
17. Vladzimirskiy AV, Gusev AV, Sharova DE, et al. Health Information System Maturity Assessment Methodology. *Medical Doctor and IT*. 2022;3(3):68–84. doi: 10.25881/18110193_2022_3_68
18. Order of the Moscow City Health Department No. 160 dated 03.11.2022 «On Approval of the Procedure and Conditions for Conducting an Experiment on the Use of Innovative Technologies in Computer Vision for Analyzing Medical Images and Further Application in the Moscow City Health Care System in 2022». (In Russ).

СПИСОК ЛИТЕРАТУРЫ

1. Владзими́рский А.В., Васи́льев Ю.А., Арза́масов К.М. и др. Компьютерное зрение в лучевой диагностике: первый этап Московского эксперимента / под ред. Ю.А. Васи́льева, А.В. Влади́мирского. Издательские решения, 2022.
2. Ranschaert E.R., Morozov S., Algra P.R., editors. *Artificial Intelligence in Medical Imaging*. Berlin : Springer, 2019. doi: 10.1007/978-3-319-94878-2
3. Гусев А.В., Добри́днюк С.Л. Искусственный интеллект в медицине и здравоохранении // Информационное общество. 2017. № 4-5. С. 78–93.
4. Шутов Д.В., Шарова Д.Е., Абуладзе Л.П., Дроздов Д.В. Системы искусственного интеллекта в клинической физиологии: как сделать их обучение эффективным? // Digital Diagnostics. 2023. Т. 4, № 1. С. 81–88. doi: 10.17816/DD123559
5. Мелдо А.А., Уткин Л.В., Трофимова Т.Н. Искусственный интеллект в медицине: современное состояние и основные направления развития интеллектуальной диагностики // Лучевая диагностика и терапия. 2020. Т. 11, № 1. С. 9–17. doi: 10.22328/2079-5343-2020-11-1-9-17
6. Recht M.P., Dewey M., Dreyer K., et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and

- recommendations // *European radiology*. 2020. Vol. 30, N 6. P. 3576–3584. doi: 10.1007/s00330-020-06672-5
7. Larson D.B., Harvey H., Rubin D.L., et al. Regulatory Frameworks for Development and Evaluation of Artificial Intelligence-Based Diagnostic Imaging Algorithms: Summary and Recommendations // *Journal of the American College of Radiology*. 2021. Vol. 18, N 3 Pt A. P. 413–424. doi: 10.1016/j.jacr.2020.09.060
8. Zinchenko V., Chetverikov S., Ahmad E., et al. Changes in software as a medical device based on artificial intelligence technologies // *International Journal of Computer Assisted Radiology and Surgery*. 2022. Vol. 17. P. 1969–1977. doi: 10.1007/s11548-022-02669-1
9. Nomura Y., Miki S., Hayashi N., et al. Novel platform for development, training, and validation of computer-assisted detection/diagnosis software // *International Journal of Computer Assisted Radiology and Surgery*. 2020. Vol. 15, N 4. P. 661–672. doi: 10.1007/s11548-020-02132-z
10. Методические рекомендации по порядку проведения экспертизы качества, эффективности и безопасности медицинских изделий (в части программного обеспечения) для государственной регистрации в рамках национальной системы ФГБУ «ВНИИ-ИМТ» Росздравнадзора. Москва, 2021.
11. Pemberton H.G., Zaki L.A.M., Goodkin O., et al. Technical and clinical validation of commercial automated volumetric MRI tools for dementia diagnosis — a systematic review // *Neuroradiology*. 2021. Vol. 63. P. 1773–1789. doi: 10.1007/s00234-021-02746-3
12. Приказ Департамента здравоохранения города Москвы № 51 от 26.01.2021 «Об утверждении порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских

- изображений и дальнейшего применения в системе здравоохранения города Москвы в 2021 году»
13. Четвериков С.Ф., Арзамасов К.М., Андрейченко А.Е., и др. Подходы к формированию выборки для контроля качества работы систем искусственного интеллекта в медико-биологических исследованиях // *Современные технологии в медицине*. 2023. Т. 15, № 2. С. 19. doi: 10.17691/stm2023.15.2.02
14. Зинченко В.В., Арзамасов К.М., Четвериков С.Ф., и др. Методология проведения пострегистрационного клинического мониторинга для программного обеспечения с применением технологий искусственного интеллекта // *Современные технологии в медицине*. 2022. Т. 14, № 5. С. 15–25. doi: 10.17691/stm2022.14.5.02
15. Altman D.G. Statistics and ethics in medical research: III How large a sample? // *British medical journal*. 1980. Vol. 281, N 6251. P. 1336. doi: 10.1136/bmj.281.6251.1336
16. Тыров И.А., Васильев Ю.А., Арзамасов К.М., и др. Оценка зрелости технологий искусственного интеллекта для здравоохранения: методология и ее применение на материалах Московского Эксперимента по компьютерному зрению в лучевой диагностике // *Врач и информационные технологии*. 2022. № 4. С. 76–92. doi: 10.25881/18110193_2022_4_76
17. Владзимирский А.В., Гусев А.В., Шарова Д.Е., и др. Методика оценки уровня зрелости информационной системы для здравоохранения // *Врач и информационные технологии*. 2022. № 3. С. 68–84. doi: 10.25881/18110193_2022_3_68
18. Приказ Департамента здравоохранения города Москвы от 03.11.2022 № 160 «Об утверждении порядка и условий проведения эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы в 2022 году».

AUTHORS' INFO

* Viktoria V. Zinchenko;

address: 24-1 Petrovka str., 1127051, Moscow, Russian Federation;
ORCID: 0000-0002-2307-725X;
eLibrary SPIN: 4188-0635;
e-mail: ZinchenkoVV1@zdrav.mos.ru

Kirill M. Arzamasov, MD, Cand. Sci. (Med.);

ORCID: 0000-0001-7786-0349;
eLibrary SPIN: 3160-8062;
e-mail: ArzamasovKM@zdrav.mos.ru

Elena I. Kremneva, MD, Cand. Sci. (Med.);

ORCID: 0000-0001-9396-6063;
eLibrary SPIN: 8799-8092;
e-mail: KremnevaEI@zdrav.mos.ru

Anton V. Vladzimirskyy, MD, Dr. Sci. (Med.);

ORCID: 0000-0002-2990-7736;
eLibrary SPIN: 3602-7120;
e-mail: VladzimirskijAV@zdrav.mos.ru

Yuriy A. Vasilev, MD, Cand. Sci. (Med.);

ORCID: 0000-0002-0208-5218;
eLibrary SPIN: 4458-5608;
e-mail: VasilevYA1@zdrav.mos.ru

ОБ АВТОРАХ

* Зинченко Виктория Валерьевна;

адрес: Российская Федерация, 127051, Москва, ул. Петровка, д. 24, стр. 1;
ORCID: 0000-0002-2307-725X;
eLibrary SPIN: 4188-0635;
e-mail: ZinchenkoVV1@zdrav.mos.ru

Арзамасов Кирилл Михайлович, канд. мед. наук;

ORCID: 0000-0001-7786-0349;
eLibrary SPIN: 3160-8062;
e-mail: ArzamasovKM@zdrav.mos.ru

Кремнева Елена Игоревна, канд. мед. наук;

ORCID: 0000-0001-9396-6063;
eLibrary SPIN: 8799-8092;
e-mail: KremnevaEI@zdrav.mos.ru

Владзимирский Антон Вячеславович, д-р мед. наук;

ORCID: 0000-0002-2990-7736;
eLibrary SPIN: 3602-7120;
e-mail: VladzimirskijAV@zdrav.mos.ru

Васильев Юрий Александрович, канд. мед. наук;

ORCID: 0000-0002-0208-5218;
eLibrary SPIN: 4458-5608;
e-mail: VasilevYA1@zdrav.mos.ru

* Corresponding author / Автор, ответственный за переписку