SHORT COMMUNICATION

Changes in software as a medical device based on artificial intelligence technologies

Victoria Zinchenko[1] · Sergey Chetverikov[3] · Ekaterina Akhmad[1] · Kirill Arzamasov[3] ·
Anton Vladzymyrskyy[4] · Anna Andreychenko[2,3] · Sergey Morozov[5]

Abstract

Purpose   to develop a procedure for registering changes, notifying users about changes made, unifying software as a medical device based on artificial intelligence technologies (SaMD-AI) changes, as well as requirements for testing and inspections—quality control before and after making changes.

Methods   The main types of changes, divided into two groups—major and minor. Major changes imply a subsequent change of a SaMD-AI version to improve efficiency and safety, to change the functionality, and to ensure the processing of new data types. Minor changes imply those that SaMD-AI developers can make due to errors in the program code. Three types of SaMD-AI testings are proposed to use: functional testing, calibration testing or control, and technical testing.

Results   The presented approaches for validation SaMD-AI changes were introduced. The unified requirements for the request for changes and forms of their submission made this procedure understandable for SaMD-AI developers, and also adjusted the workload for the Experiment experts who checked all the changes made to SaMD-AI.

Conclusion   This article discusses the need to control changes in the module of SaMD-AI, as innovative products influencing medical decision making. It justifies the need to control a module operation of SaMD-AI after making changes. To streamline and optimize the necessary and sufficient control procedures, a systematization of possible changes in SaMD-AI and testing methods was carried out.

Keywords   Artificial intelligence · Medical software based on artificial intelligence technologies · Software as a medical device · Modifications · Changes · Validation

✉ Ekaterina Akhmad
   e.ahmad@npcmr.ru

1   The Department of Innovative Technologies, State Budget-Funded Health Care Institution of the City of Moscow "Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department", 24 Petrovka Str., Bldg. 1, 127051 Moscow, Russia

2   Department of Physics and Engineering, ITMO University, Kronverkskiy Prospekt, 49, 197101 Saint Petersburg, Russia

3   The Department of Medical Informatics, Radiomics and Radiogenomics, State Budget-Funded Health Care Institution of the City of Moscow "Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department", 24 Petrovka Str., Bldg. 1, 127051 Moscow, Russia

4   Deputy Director for Science, State Budget-Funded Health Care Institution of the City of Moscow "Research and Practical Clinical Center for Diagnostics and Telemedicine

Technologies of the Moscow Health Care Department", 24 Petrovka Str., Bldg. 1, 127051 Moscow, Russia

5   Radiology, State Budget-Funded Health Care Institution of the City of Moscow "Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department", 24 Petrovka Str., Bldg. 1, 127051 Moscow, Russia

Introduction

The intensive development of artificial intelligence technologies in medicine is a new and promising direction. Software as a medical device based on artificial intelligence technologies (SaMD-AI) can help users in routine and complex tasks, improve a quality, availability and rate of medical care provided for patients [1–3]. At the same time, the stages of development, commissioning and application of SaMD-AI should be monitored by both developers according the qual-

ity management system and regulatory authorities allowing medical devices on the market.

The World Health Organisation (WHO) identified several ethical principles that should be followed during the process of SaMD-AI development and use [4,5]. Among others aspects, the WHO noted the rule for transparency, explainability of AI technologies. SaMD-AI should be transparent, which means it should give access to information about the algorithm and data used by this system. SaMD-AI should also be explainable, i.e. provide description of how it came up with a given result [6]. Towards the users, causability is one of the most important concepts, which indicates the measurable level of causal understanding that human expert achieves [7]. According to these rules and principles, SaMD-AI users should always be provided with clear information about the current SaMD-AI performance and description of how it works.

However, SaMD-AI has a non-physical nature and like other software as a medical device can be changed during its clinical implementation [8]. Besides, SaMD-AI should be regularly updated in case of better training data availability, clarity in tasks definition or other technological improvement [9]. This software can also be modified in case of encouraging technical errors [8,9]. Besides, this updating can be on the different levels of the SaMD-AI architecture [10]. All these causes can lead to potential discrepancy between the current state of SaMD-AI and the one described previously. Thus, software developers and other stakeholders should review documentation and descriptions of SaMD-AI performance according to the changes made, as it is required by the ethical principles.

This issue was first identified by us during the Experiment– a scientific and practical study to investigate the possibility of using computer vision technologies that include SaMDs-AI in diagnostic radiology (further, the Experiment) [11]. The objective of this project is to assess SaMD-AI safety and efficiency in the Moscow radiology departments [12]. To implement this project, SaMD-AI is being integrated into the Unified Radiological Information Service (URIS) of the Unified Medical Information Analysis System, routinely operating in Moscow [13]. At the time of this writing (September 23, 2021), 21 SaMDs-AI participate in the Experiment, which were connected to 299 Moscow medical facilities—908 diagnostic devices and 538 doctors were involved. A total of 1,662,208 studies in 13 different directions were analyzed (chest computed tomography (CT) or low-dose computed tomography (LDCT) to detect cancer, COVID-19; chest X-ray to detect different pathologies; mammography to detect breast cancer; head CT; and etc). An important value of the Experiment lies in its scale and possibility to study different scenarios of the SaMD-AI application, which was possible due to the use of such a large-scale information system of Moscow Healthcare as URIS.

To participate in the Experiment, SaMD-AI developers submit an application that includes a set of documents to be reviewed by experts of the Experiment. Given the satisfactory results of the review, the developer receives a positive decision. After that, SaMD-AI is technically integrated with the URIS by providing technical access to the URIS system and verifying that data exchange between SaMD-AI and URIS is correct. Next, SaMD-AI receives studies that correspond to its area of application from all diagnostic devices located in medical facilities of the Moscow Healthcare Department. All medical facilities that took part in the Experiment are subordinate agencies of the Moscow public healthcare system [14,15]. Heads of these facilities were instructed to provide assistance in the implementation of the Experiment. Medical facilities received the SaMD-AI operation results, but the doctors used them and gave their feedback on a voluntary basis. Following a successful integration of SaMD-AI with the URIS, the participants in the Experiment (SaMD-AI developers) receive monetary grants, which are paid once a month. Further benefits of participation in the Experiment for the developers include the possibility of obtaining SaMD-AI testing protocols and receiving an expert opinion on SaMD-AI operation and further product development. As an additional incentive for radiologists connected to SaMD-AI, Most Active User contests were occasionally organized in medical facilities.

As part of the Experiment, SaMD-AI developers can test their product in radiology departments of Moscow medical facilities, even though not all of SaMDs-AI have a registration certificate as a medical device at the time of participation. Before connecting SaMD-AI to the automated radiologist's workstation, experts (the data scientist and radiologists) perform a number of testings and allow operating only those SaMDs-AI which provide a sufficient accuracy and efficiency [14,16]. Besides, the quality control and monitoring tests are performed already at the stage when SaMDs-AI are embedded into actual radiology workflow in the Experiment [14]. While performing these procedures, experts discovered a few cases of significant degradation of several SaMD-AI performance characteristics, which was caused by some modifications made by SaMD-AI developers. However, as it is known, changes in a single SaMD-AI module can lead to a change in the functional characteristics of other SaMD-AI modules, related not only to the AI-based software [17]. As for the need for confirmation of the quality level of SaMD-AI after changes, notification from the SaMD-AI developers and additional testing from experts should be performed taking into account that the level of testing should be also determined by the causability of the SaMD-AI.

Thus, in order to provide the level of explainability, transparency and causability, as well as to maintain SaMD-AI quality, safety and efficiency, it is necessary to develop a procedure for registering changes, notifying users about changes
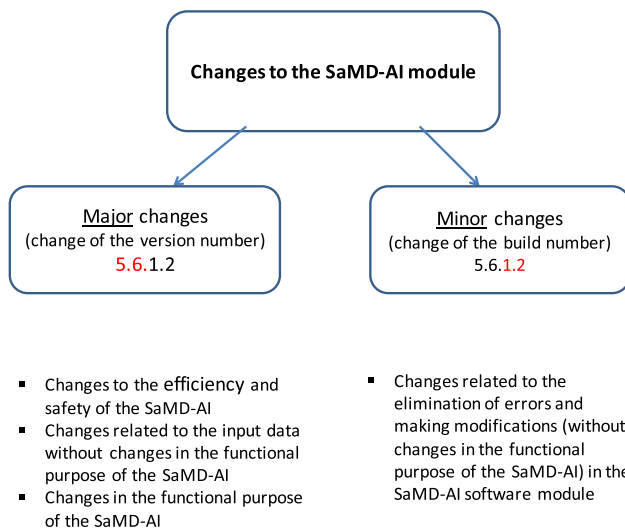
Fig. 1 Variants of SaMD-AI changes

made, unifying SaMD-AI changes, as well as requirements for testing and inspections—quality control before and after making changes.

Methods

SaMD-AI developers can make changes to the software module of their products according to their own improvements, feedback from users or recommendations from the Experiment experts, or to correct errors. In this article, the following practical approach to stratification of changes in software is proposed, which can be also applied in other areas outside of diagnostic radiology. Figure 1 shows the main types of changes, divided into two groups—major and minor [10,18]. Any changes to SaMD-AI are aimed at the specific result. Appendix B presents the types of SaMD-AI changes in details with some examples. Major changes are divided into three types; they imply a subsequent change of a SaMD-AI version [19]:

1. Changes made to improve the efficiency and safety of SaMD-AI. This type of changes includes improvements in SaMD-AI performance indicators, which may be a result of its changes during an additional training on new datasets within the declared functionality of SaMD-AI and the same type of input data (Appendix B, Fig. 5).
2. Changes made to ensure a processing of new data types – a developer modifies the input data used by SaMD-AI. However, the functional purpose of SaMD-AI is not changed (Appendix B, Fig. 6).
3. Changes in the functional purpose – these changes lead to a change in the influence degree of the SaMD-AI oper-

ation results and conditions for its use (Appendix B, Fig. 7).

Minor changes implies to those that SaMD-AI developers can make due to errors in the program code, a need for modifications due to incorrect data display in the user interface, etc. (Appendix B, Fig. 8). In this case, a SaMD-AI version number is unchanged, and a build number is changed.

A SaMD-AI developer should describe the changes being made and send a notification about them to users, register them in its documentation and inform the expert group (as part of the Experiment). Then, if necessary, the new version of SaMD-AI is tested using the dataset [20]. A type of testing is determined by the change made to SaMD-AI, their correlation is provided in Appendix A. According to the Experiment experience, 3 types of SaMD-AI testings are proposed to use: functional testing, calibration testing or control and technical testing [16].

Functional testing

During functional testing, the availability and operability of the SaMD-AI functions declared by the developer are checked, as well as their compliance with the requirements, an example of which is described and presented on the Experiment website (https://mosmed.ai/en/).

Functional testing is performed on a small dataset that includes images with no target pathology, images with target pathology, as well as images with defects. Defects mean nonstandard for the software content of diagnostic study results: incorrect or cropped anatomical area, noisy images, etc. This is necessary to check the reliability and stability of the software operation. Testing includes the assessment of SaMD-AI from both technical and medical points of view. A technical part is responsible for the correct display of all required areas and labeling in the radiologist's automated workstation. A medical part is performed by expert doctors in terms of assessing the correctness of the solution to the clinical task of SaMD-AI, as well as identifying serious violations in the SaMD-AI operation (for example, missing obvious pathological findings).

The result of functional testing should be a protocol for assessing the completeness and correctness of the implemented SaMD-AI functionality. The protocol may contain significant comments and recommendations.

All comments can be conditionally divided into three groups:

1. Affect a patient safety and a doctor's performance—the absence of implementation of the declared functionality by the developer; comments affecting a radiologist's performance; deviations in the SaMD-AI functionality from

the basic functional requirements; irreversible distortion of the initial study data.

2. Do not affect a patient safety, but affect a doctor's performance—functional deficiencies that do not meet generally accepted standards for presenting results of the radiological study interpretation.

3. Do not affect a patient safety, do not affect a doctor's performance—minor deficiencies that need to be eliminated for a more convenient, intuitive and efficient performance of the radiologist.

Calibration testing

As part of the calibration testing, the optimal cut-off threshold is calculated, i.e. at which level and above it, a probability of the pathology according to SaMD-AI is present in the analyzed study, as well as the assessment of SaMD-AI diagnostic accuracy metrics [20]. In the verified dataset for a calibration testing, a part of the studies corresponds to the category 'no target pathology' and the other part—to the category 'with target pathology' in equal proportions (balanced data with a 50/50 ratio of studies with changes and without changes). A study assignment to one of the categories occurs in accordance with the clinical task of SaMD-AI. Based on the obtained probability of pathology, diagnostic accuracy metrics and a study processing time are calculated. As a diagnostic accuracy metric, the area under the ROC curve is estimated for calibration testing using balanced dataset. In case of using imbalanced dataset, specific methods for efficiency evaluation are chosen, for example, a concordant partial and a new partial c statistic measures [21].

During a calibration testing, a protocol is prepared indicating SaMD-AI efficiency parameters, a study processing time, etc. For example, in the Experiment, threshold values for the diagnostic accuracy metrics are set up. Relative to these values, a decision on the further allowance of SaMD-AI (or an updated version/assembly) to operate is made. Performing calibration testing when changing SaMD-AI ensures a sufficient level of efficiency and safety during operation.

Control and technical testing

When making changes to SaMD-AI related to the expansion of the ability to process input data from a new diagnostic device, a control and technical testing (hereinafter referred to as CTT) is carried out in the number of two studies sent from each diagnostic device. CTT is performed by prospectively validating the studies analyzed by SaMD-AI. During CTT, technological testing of SaMD-AI takes place according to the parameters specified in the list of basic functional requirements (listed on the website mosmed.ai) and within the declared functionality. An average study processing time is also monitored.

In the protocol, the CTT results reflect information about SaMD-AI and performed testing, the parameters declared by a developer are checked and the corresponding results are entered. Upon a successful completion of testing, SaMD-AI is allowed to process data from a new device.

Results

The presented approaches for controlling SaMD-AI changes were introduced into the Experiment processes. Examples of test results after making changes to SaMD-AI are described below.

Example 1 SaMD-AI No.1 was connected to the Experiment for the processing of chest X-ray (fluorography) and performed a triage of examinations for norm/pathology without specifying a localization of pathology. In the process of testing SaMD-AI, the developers submitted an application for adding visualization of the pathology zone (red contour). After the functional testing, it was revealed that a detection of pathologies occurs outside the zone of interest—the lungs (Fig. 2). A decision was made to refuse SaMD-AI to change the version and to send it back for revision.

Example 2 SaMD-AI No.2 was connected to the Experiment for detecting COVID-19 on CT scan. Prior to making changes to SaMD-AI No.2, areas of contouring that were not pathological were identified (Fig. 3a). The developer made changes to the segmentation algorithm in such a way that contouring of pathological area became more correct, which was confirmed during a functional testing (Fig. 3 b).

Example 3 SaMD-AI No.3 was intended for processing chest X-ray. Analysing operation results of the first version of SaMD-AI No.3, it was identified more than 40 of defects associated with the complete or partial absence of results, as well as the presence of findings outside the target organ (Fig. 4a), which means a high number of false responses by SaMD-AI. The Experiment experts recommended making changes to SaMD-AI to eliminate errors. After the first change to SaMD-AI No.3, the following results were obtained: a satisfactory visualization based on the functional testing results, to be exact, a localization type was changed from a heatmap to a green contouring (Fig. 4 b); however, unsatisfactory quality of work was identified. The experts sent back SaMD-AI No.3 again to fix the defects, however, the subsequent changes resulted in negative outcomes compared to the operation quality of the previous version (AUC decreased by 16%), and therefore a transition to the latest updated version was not agreed upon.

Based on the experience obtained in the Experiment, we can propose to unify the principles of SaMD-AI version/build numbering, for example, in the form of Arabic numerals (1.0.0).

Fig. 2 SaMD-AI No.1 operation results after making a change. Errors – localization of pathology outside the organ of interest – the lungs (**a**, **b**)
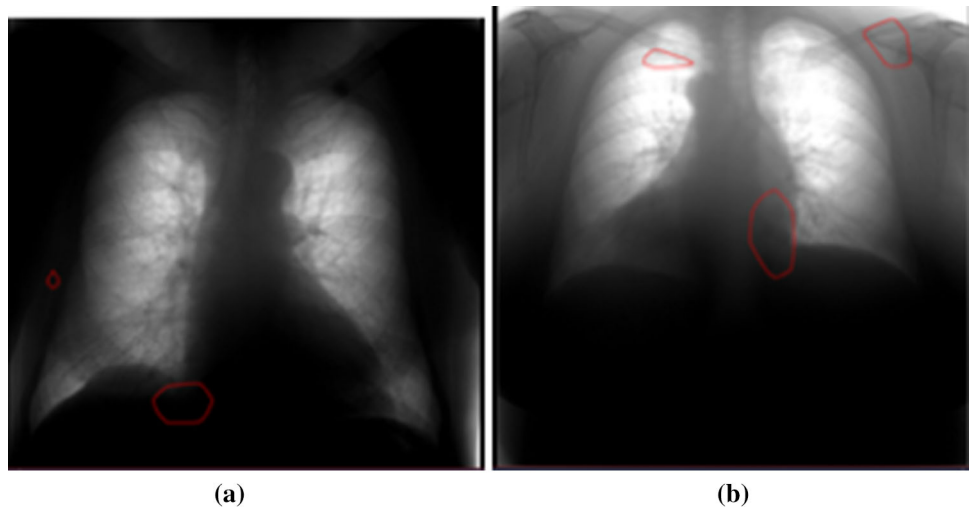

(a)   (b)

Fig. 3 SaMD-AI No.2 operation results before making changes (**a**) and after (**b**)
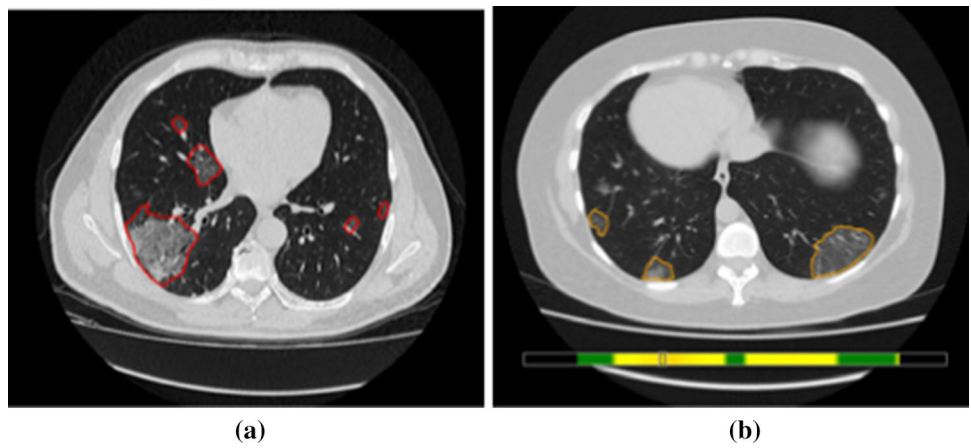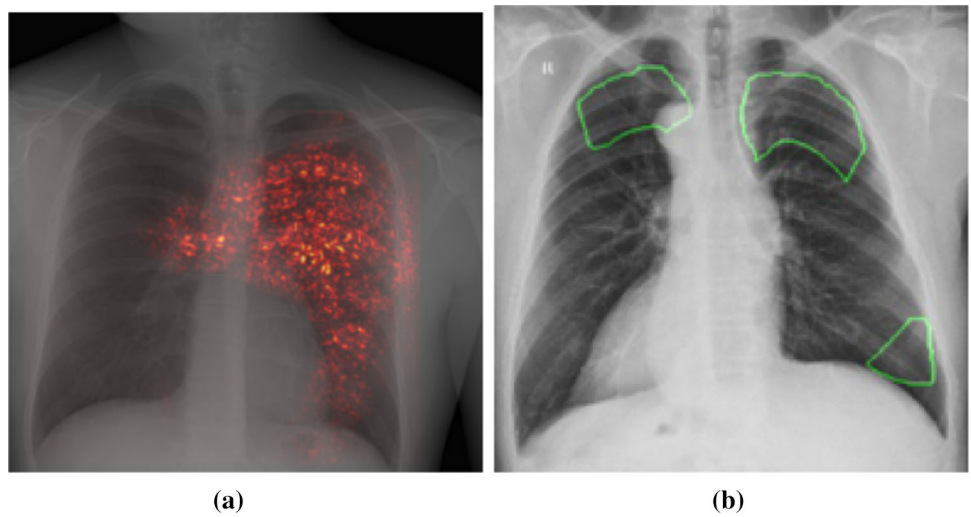

(a)   (b)

Fig. 4 SaMD-AI operation results No.3: **a** defects in the first version of SaMD-AI No.3 – localization of pathology on the heart, spine, abdominal organs; **b** after making the second changes – areas without pathological changes are localized, labeling outside the target organ

(a)   (b)

Discussion

The work performed on structuring the types of changes in SaMD-AI, which was based on the FDA's proposals [10], and defining certain checking and testing for each type of changes. The developed method allowed us to maintain a quality level, safety and efficiency of SaMD-AI when dynamic changes were made at the approbation stage during the Experiment. The unified requirements for the request for changes and forms of their submission made this procedure understandable for SaMD-AI developers, and also adjusted the workload for the Experiment experts who checked all the changes made to SaMD-AI. Besides, radiologists working with SaMD-AI are promptly informed about changes in the software, which enables them to accurately identify the product being used and thereby to build confidence in it, since all versions of SaMD-AI go through the appropriate control stages before users start applying SaMD-AI in their practice.

The Experiment is a unique research study with a large-scale infrastructure implementation of SaMDs-AI and a remarkable work aimed at proving the ability of the software to process data correctly and providing objective evidence of clinical effectiveness of SaMD-AI. This evaluation includes both clinical metrics of performance and, for example, time characteristics [22]. The design of this project requires performing SaMD-AI testing at different stages: at the preparatory stage, during the deployment and at the final evaluation stage. Possible challenges associated with the degrading SaMD-AI performance during the deployment stage result in the need for a procedure for SaMD-AI changes notification and quality control. This procedure was developed and optimized on various use cases with an active involvement of the stakeholders (software developers, end-user physicians, data scientist, healthcare managers, etc.).

The necessity to provide clear information about SaMD-AI changes is determined by ethical principles that are intended for all stakeholders [4]. One of these recommendations includes the requirements for general AI trustworthiness that can be achieved by following the provisions of regulations and standards. IEC 62304 requires to develop a software configuration management plan; however, the rules are described in general and do not take into account specific characteristics of SaMD-AI [23]. ISO 14971 provides the requirements for creating risk management files that should also describe general assessment of possible medical device changes and their risk evaluation [24]. According to DIN SPEC 92001-1, SaMD-AI can be changed at the operation stage of the product life cycle due to the continuous learning [25]. ISO/IEC TR 29119-11 introduces the term 'evolution' as a SaMD-AI characteristic [26]. Evolution is determined by the changes in SaMD-AI behaviour as a consequence of self-learning or updates of usage profile. Another recommendation for developers after making changes to SaMD-AI is

to ensure internal quality control of released changes and provide the results of their internal tests to users of the new version/build of SaMD-AI [8]. However, the weak point of the above standards is their generalized approach, which means that the responsibility for determining and classifying SaMD-AI changes lies with the developers and quality management system.

To address the range of SaMD-AI changes, the Regulatory Guidelines of the Singapore Health Sciences Authority employ a risk-based approach [27]. This approach relates to the changes presented in our study; however, in our work, we also propose testing methods for different types of changes.

During clinical evaluation of SaMD-AI at the development stage, initial data about the SaMD-AI effectiveness is received [22,28]. Depending on the SaMD-AI type, several standards for this study are known [29–31]. However, as it was mentioned above, SaMD-AI can change its performance characteristics due to self-learning or other modifications. That is why it is significantly important for users to have information about recent updates and current SaMD-AI performance metrics. These requirements are also mentioned in various guidelines for evaluation of commercial SaMDs-AI [32] and can be part of self-assessment of SaMD-AI provided by the developers [33]. Our methodology presented in this article would be beneficial for helping users, such as physicians and healthcare managers, to achieve their goals by deploying a high-quality SaMD-AI application.

The approaches to control SaMD-AI at the testing stage, presented in this paper, can be used in the future to monitor these software products at the operational stage [34,35]. This is relevant at the moment, since requirements for this procedure, and safety and efficiency assurance of SaMD-AI at the entire stage of operation are just being formed.

Testing SaMD-AI performance is a necessary and important task, primarily to control a sufficient level of efficiency and minimize false results that may affect medical decisions. One of the goals of the strategies for the artificial intelligence development in different countries is to increase confidence in working with AI technologies, which is based on reliable and reproducible performance results.

The proposed approach is applicable for future implementation as a base for SaMD-AI changes quality control, especially in the field of diagnostic radiology. Our proposal was developed and applied for a number of the most prominent use cases for SaMD-AI application in radiology on different cases from the Experiment, which makes it more valid and beneficial.

Conclusion

The approaches to control SaMD-AI changes presented in this paper allow making a verification procedure clearly

for SaMD-AI developers, experts, as well as for users. Compliance with these principles allows to maintain the level of safety and efficiency of SaMD-AI operation when implementing in medical facilities as well as in medical networks. The proposed approaches allow unifying requirements, streamline procedures for making changes while optimizing the cost for users, but maintaining the quality level.

Declarations

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent For this type of study formal consent is not required.

Appendix A The list of abbreviations

CT - Computed tomography

CTT - Control and technical testing

LDCT - Low-dose computed tomography

ROC - Receiver operating characteristic

SaMD-AI - a Software as a medical device based on artificial intelligence technologies

URIS - Unified radiological information service

Appendix B Types of changes to AI services

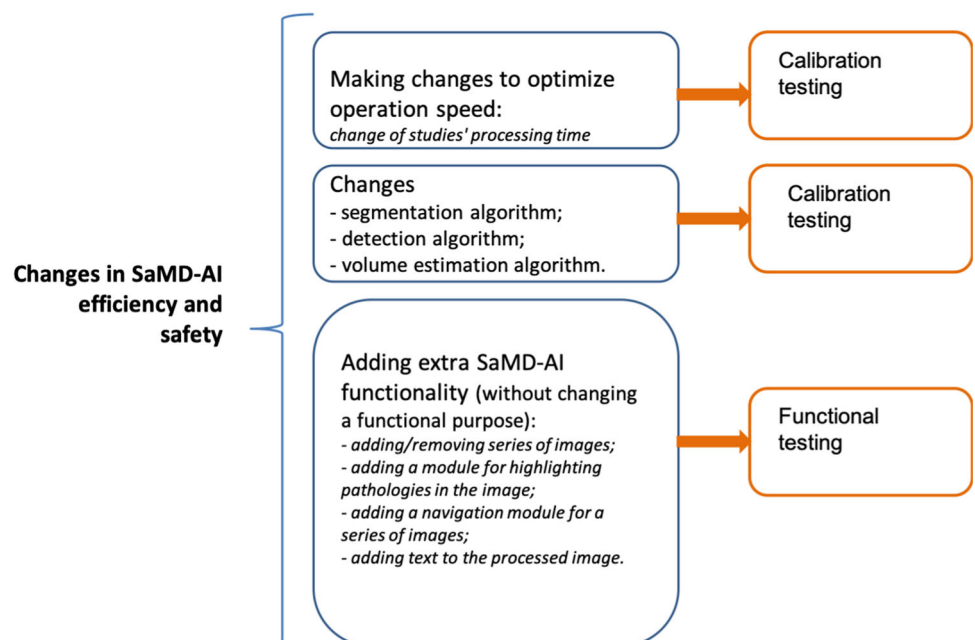Fig. 5 "Changes in SaMD-AI efficiency and safety"

Fig. 6 "Changes related to input data without changes in the functional purpose of the SaMD-AI"
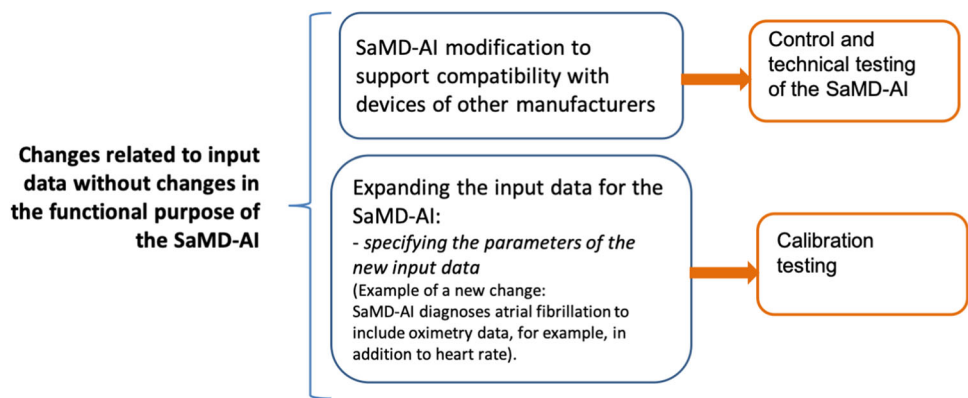

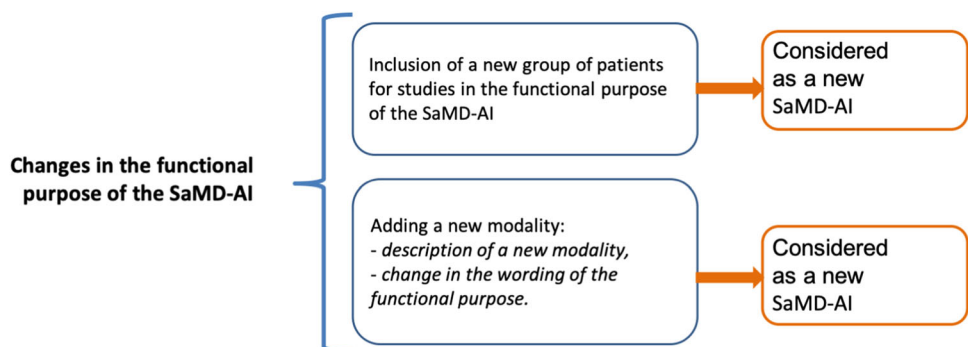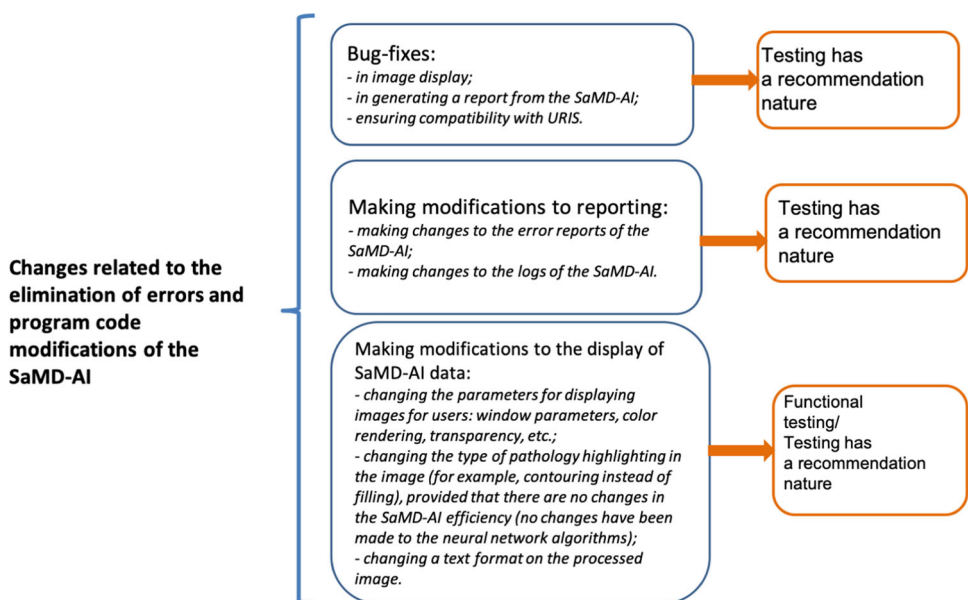**Fig. 7** "Changes in the functional purpose of the SaMD-AI"

Fig. 8 "Changes related to the elimination of errors and program code modifications of the SaMD-AI"

References

1. He J, Baxter S, Xu J, Zhou X, Zhang K (2019) The practical implementation of artificial intelligence technologies in medicine. Nat Med 25:30–36. https://doi.org/10.1038/s41591-018-0307-0
2. Ranschaert E, Morozov S, Algra P (eds) (2019) Artificial intelligence in medical imaging. Springer, Cham
3. Gusev A, Dobridnyuk S (2017) Artificial intelligence in medicine and healthcare. Inform Soc 4–5:78–93
4. The World Health Organization guidance. Ethics and governance of artificial intelligence for health. https://www.who.int/publications/i/item/9789240029200
5. Sharova D, Zinchenko V, Akhmad E, Mokienko O, Vladzymyrskyy A, Morozov S (2021) On the issue of ethical aspects of the artificial intelligence systems implementation in healthcare. Dig Diagnost 2(3):356–368
6. Muller H, Mayrhofer M, Van Veen E, Holzinger A (2021) The ten commandments of ethical medical ai. Computer 54(7):119–123

7. Holzinger A, Malle B, Saranti A, Pfeifer B (2021) Towards multimodal causability with graph neural networks enabling information fusion for explainable ai. Inform Fusion 71:28–37
8. Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations: IMDRF-SaMD WGN12FINAL:2014. http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf
9. Larson D, Harvey H, Rubin D, Irani N, Tse J, Langlotz C (2020) Regulatory frameworks for development and evaluation of artificial intelligence-based diagnostic imaging algorithms: Summary and recommendations. J Am College of Radiol 18(3):413–424. https://doi.org/10.1016/j.jacr.2020.09.060
10. The United States Food and Drug Administration: Proposed Regulatory Framework for Modifications to Artificial Intelligence Machine Learning (AI ML)-Based Software as a Medical Device (SaMD) - Discussion Paper. https://www.fda.gov/media/122535/download (2019)
11. Morozov S, Vladzymyrskyy A, Ledikhova N, Andreychenko A, Arzamasov K, Balanjuk E, Gombolevskij V, Ermolaev S, Zhivodenko V, Idrisov I, Kirpichev J, Logunova T, Nuzhdina V, Omeljanskaja O, Rakovchen V, Slepushkina A (2020) Moscow experiment on computer vision in radiology: involvement and participation of radiologists. Phys Inform Technol 4:14–23
12. Experiment on the use of innovative computer vision technologies for analysis of medical images in the Moscow healthcare system. https://www.clinicaltrials.gov/ct2/show/NCT04489992 (2019)
13. Morozov S (2021) With the unified radiological information system (uris), the moscow health care department has laid the foundations for connected care. Health Manag org 21(6):348–351
14. Andreychenko A, Logunova T, Gombolevskiy V, Nikolaev A, Vladzymyrskyy V, AV abd Sinitsyn SP M (2022) A methodology for selection and quality control of the radiological computer vision deployment at the megalopolis scale. medRxiv 02(12):22270663 . https://doi.org/10.1101/2022.02.12.22270663
15. Order No. 160 of 2022 'On approving the Procedure and conditions for conducting the Experiment on the application of innovative computer vision technologies for the analysis of medical images and further use in the Moscow healthcare system in 2022', paragraph 3
16. Andreychenko A, Gombolevskiy V, Vladzymyrskyy A, Morosov S (2021) How to select and fine-tune an AI service for practical application of radiology. Book of Abstract Eur Cong Radiol RPS, pp. 3–7
17. Sculley D, Holt G, Golovin D, Davydov E, Phillips T (2015) Hidden technical debt in machine learning systems. NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems 2:2503–2511
18. Regulation (Eu) 2017745 The European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 200183EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90385EEC and 9342EEC. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02017R0745-20200424
19. Pavlov N, Andreychenko A, Vladzymyrskyy A, Revazyan A, Kirpichev Y, Morozov S (2020) Reference medical datasets mosmeddata for independent external evaluation of algorithms based on artificial intelligence in diagnostics. Dig Diagnost 2(1):49–65
20. Morozov S, Vladzymyrskyy A, Klyashtornyy V, Andreychenko A, Kulberg N, Gombolevsky V, Sergunova K (2019) Clinical acceptance of software based on artificial intelligence technologies (radiology) vol. 57, p. 45. Best practices in medical imaging, Moscow . https://mosmed.ai/documents/35/CLINICAL_ACCEPTANCE_19.02.2020_web__1_.pdf
21. Carrington A, Fieguth P, Qazi H, Holzinger A, Chen H, Mayr F, Manuel D (2020) A new concordant partial auc and partial c statistic for imbalanced data in the evaluation of machine learning algorithms. BMC Med Inform Decis Mak (20(4)), 1–12
22. Software as a Medical Device (SaMD): Clinical Evaluation: IMDRFSaMD WGN41FINAL:2017. https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-170921-samd-n41-clinical-evaluation_1.pdf
23. IEC 62304:2006 Medical device software – Software life cycle processes
24. ISO 14971:2019 Medical devices – Application of risk management to medical devices
25. DIN SPEC 92001-1 Artificial intelligence - Life cycle processes and quality requirements - Part 1: Quality Meta Model
26. ISO/IEC TR 29119-11:2020 Software and systems engineering – Software testing – Part 11: Guidelines on the testing of AI-based systems
27. Health Science Authority: Regulatory Guidelines for Software Medical Devices – A Lifecycle Approach. https://www.hsa.gov.sg/docs/default-source/announcements/regulatory-updates/regulatory-guidelines-for-software-medical-devices--a-lifecycle-approach.pdf
28. Higgins D (2021) Onramp for regulating artificial intelligence in medical products. Adv Intell Syst 3:2100042
29. Collins G, Dhiman P, Andaur Navarro C, Ma J, Hooft L, Reitsma J, Logullo P, Beam A, Peng L, Van Calster B, van Smeden M, Riley R, Moons K (2021) Protocol for development of a reporting guideline (tripod-ai) and risk of bias tool (probast-ai) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 11:048008
30. Sounderajah V, Ashrafian H, Golub R, Shetty S, De Fauw J, Hooft L, Moons K, Collins G, Moher D, Bossuyt P, Darzi A, Karthikesalingam A, Denniston A, Mateen B, Ting D, Treanor D, King D, Greaves F, Godwin J, Pearson-Stuttard J, Harling L, McInnes M, Rifai N, Tomasev N, Normahani P, Whiting P, Aggarwal R, Vollmer S, Markar S, Panch T, Liu X (2021) On behalf of the stard-ai steering committee. developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the stard-ai protocol. BMJ Open (11), 047709
31. Liu X, Rivera S, Moher D, Calvert M, Denniston A, Ashrafian H, Beam A, Chan A, Collins G, Darzi A, Deeks J, ElZarrad M, Espinoza C, Esteva A, Faes L, Ferrante di Ruffano L, Fletcher J, Golub R, Harvey H, Haug C, Holmes C, Jonas A, Keane P, Kelly C, Lee A, Lee C, Manna E, Matcham J, McCradden M, Monteiro J, Mulrow C, Oakden-Rayner L, Paltoo D, Panico M, Price G, Rowley S, Savage R, Sarkar R, Vollmer S, Yau C (2020) Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the consort-ai extension. Nat Med 26:1364–1374
32. Omoumi P, Ducarouge A, Tournier A, Harvey H, Kahn CJ, Louvet-de Verchère F, Pinto Dos Santos D, Kober T, Richhiardi J (2021) To buy or not to buy-evaluating commercial ai solutions in radiology (the eclair guidelines). Eur Radiol 31:3786–3796
33. Cabitza F, Campagner A (2021) The need to separate the wheat from the chaff in medical informatics: introducing a comprehensive checklist for the (self)-assessment of medical ai studies. Int J Med Inform 153:104510
34. Decision of the Board of the Eurasian Economic Commission No.174 dated December 22, 2015 On approval of the rules for monitoring the safety, quality and effectiveness of medical devices
35. European Union: Annual report on European SMEs. https://op.europa.eu/en/publication-detail/-/publication/cadb8188-35b4-11ea-ba6e-01aa75ed71a1/language-en (2018-2019)