

DOI: <https://doi.org/10.17816/DD321423>

Двойной просмотр результатов маммографии с применением технологий искусственного интеллекта: новая модель организации массовых профилактических исследований

Ю.А. Васильев¹, И.А. Тыров², А.В. Владзимирский¹, К.М. Арзамасов¹,
И.М. Шулькин¹, Д.Д. Кожихина¹, Л.Д. Пестренин¹

¹ Научно-практический клинический центр диагностики и телемедицинских технологий, Москва, Российская Федерация

² Департамент здравоохранения города Москвы, Москва, Российская Федерация

АННОТАЦИЯ

Обоснование. Доступность наборов медицинских данных и технологий разработки программного обеспечения на основе искусственного интеллекта в последние годы привела к увеличению количества решений для медицинской диагностики и маммографии в частности. Это программное обеспечение, зарегистрированное как медицинское изделие, может быть использовано для описания цифровых маммографий, что позволит в значительной мере сэкономить временные, материальные и кадровые ресурсы в здравоохранении при гарантированном сохранении качества профилактических исследований молочных желёз.

Цель — обосновать возможность и эффективность применения программного обеспечения на основе технологий искусственного интеллекта для первой интерпретации цифровых маммограмм при сохранении практики второго описания рентгеновских изображений врачом-рентгенологом.

Материалы и методы. Набор данных из 100 цифровых маммографических исследований, из них 50 — «Отсутствие целевой патологии», 50 — «Присутствие целевой патологии» (с признаками злокачественных новообразований), был обработан программным обеспечением на основе технологий искусственного интеллекта, зарегистрированным в Российской Федерации как медицинское изделие. Выполнен ROC-анализ. Ограничения исследования: значения метрик диагностической точности получены для версий программного обеспечения на основе технологий искусственного интеллекта, актуальных на конец 2022 года.

Результаты. При настройке на 80,0% чувствительность специфичность искусственного интеллекта составила 90,0% (95% ДИ 81,7–98,3), точность — 85,0% (95% ДИ 78,0–92,0). При настройке на 100% специфичность искусственный интеллект показал чувствительность 56,0% (95% ДИ 42,2–69,8), точность — 78,0% (95% ДИ 69,9–86,1). При настройке на 100% чувствительность специфичность искусственного интеллекта составила 54,0% (95% ДИ 40,2–67,8), точность — 77,0% (95% ДИ 68,8–85,2). Предложены два подхода, предусматривающие автономную первую интерпретацию цифровой маммографии посредством искусственного интеллекта. Первый подход заключается в оценке рентгеновского изображения с помощью искусственного интеллекта с более высокой чувствительностью, чем у двойного описания маммографии врачами-рентгенологами, при сопоставимом уровне специфичности. Второй подход подразумевает, что программное обеспечение на основе технологий искусственного интеллекта будет определять категорию маммографии («Отсутствие целевой патологии» или «Присутствие целевой патологии») с указанием степени своей «уверенности» в полученном результате в зависимости от «коридора», в который попадает предсказанное значение.

Заключение. Оба предложенных сценария использования программного обеспечения на основе технологий искусственного интеллекта с целью автономного первого описания цифровых маммограмм способны обеспечить качество диагностики, не уступающее двойному описанию снимков врачами-рентгенологами и даже превышающее его. Экономическая выгода от практической реализации данного подхода в масштабах страны может составлять от 0,6 до 5,5 млрд рублей ежегодно.

Ключевые слова: профилактические исследования; маммография; искусственный интеллект; диагностическая точность.

Как цитировать

Васильев Ю.А., Тыров И.А., Владзимирский А.В., Арзамасов К.М., Шулькин И.М., Кожихина Д.Д., Пестренин Л.Д. Двойной просмотр результатов маммографии с применением технологий искусственного интеллекта: новая модель организации массовых профилактических исследований // *Digital Diagnostics*. 2023. Т. 4, № 2. С. 93–104. DOI: <https://doi.org/10.17816/DD321423>

Рукопись получена: 17.03.2023

Рукопись одобрена: 25.04.2023

Опубликована: 25.05.2023

DOI: <https://doi.org/10.17816/DD321423>

Double-reading mammograms using artificial intelligence technologies: A new model of mass preventive examination organization

Yuriy A. Vasilev¹, Ilya A. Tyrov², Anton V. Vladzimirskyy¹, Kirill M. Arzamasov¹, Igor M. Shulkin¹, Daria D. Kozhikhina¹, Lev D. Pestrenin¹

¹ Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

² Moscow Health Care Department, Moscow, Russian Federation

ABSTRACT

BACKGROUND: In recent years, the availability of medical datasets and technologies for software development based on artificial intelligence technology has resulted in a growth in the number of solutions for medical diagnostics, particularly mammography. Registered as a medical device, this program can interpret digital mammography, significantly saving time, material, and human resources in healthcare while ensuring the quality of mammary gland preventive studies.

AIM: This study aims to justify the possibility and effectiveness of artificial intelligence-based software for the first interpretation of digital mammograms while maintaining the practice of a radiologist's second description of X-ray images.

MATERIALS AND METHODS: A dataset of 100 digital mammography studies (50 — “absence of target pathology” and 50 — “presence of target pathology,” with signs of malignant neoplasms) was processed by software based on artificial intelligence technology that was registered as a medical device in the Russian Federation. Receiver operating characteristic analysis was performed. Limitations of the study include the values of diagnostic accuracy metrics obtained for software based on artificial intelligence technology versions, relevant at the end of 2022.

RESULTS: When set to 80.0% sensitivity, artificial intelligence specificity was 90.0% (95% CI, 81.7–98.3), and accuracy was 85.0% (95% CI, 78.0–92.0). When set to 100% specificity, artificial intelligence demonstrated 56.0% sensitivity (95% CI, 42.2–69.8) and 78.0% accuracy (95% CI, 69.9–86.1). When the sensitivity was set to 100%, the artificial intelligence specificity was 54.0% (95% CI, 40.2–67.8), and the accuracy was 77.0% (95% CI, 68.8–85.2). Two approaches have been proposed, providing an autonomous first interpretation of digital mammography using artificial intelligence. The first approach is to evaluate the X-ray image using artificial intelligence with a higher sensitivity than that of the double-reading mammogram by radiologists, with a comparable level of specificity. The second approach implies that artificial intelligence-based software will determine the mammogram category (“absence of target pathology” or “presence of target pathology”), indicating the degree of “confidence” in the obtained result, depending on the corridor into which the predicted value falls.

CONCLUSIONS: Both proposed approaches for using artificial intelligence-based software for the autonomous first interpretation of digital mammograms can provide diagnostic quality comparable to, if not superior to, double-image reading by radiologists. The economic benefit from the practical implementation of this approach nationwide can range from 0.6 to 5.5 billion rubles annually.

Keywords: artificial intelligence; diagnostic accuracy; mammography; preventive medicine.

To cite this article

Vasilev YuA, Tyrov IA, Vladzimirskyy AV, Arzamasov KM, Shulkin IM, Kozhikhina DD, Pestrenin LD. Double-reading mammograms using artificial intelligence technologies: A new model of mass preventive examination organization. *Digital Diagnostics*. 2023;4(2):93–104. DOI: <https://doi.org/10.17816/DD321423>

Received: 17.03.2023

Accepted: 25.04.2023

Published: 25.05.2023

DOI: <https://doi.org/10.17816/DD321423>

使用人工智能对乳腺X线摄影结果进行两次查看：一种组织大规模预防性研究的新模式

Yuriy A. Vasilev¹, Ilya A. Tyrov², Anton V. Vladzimirskyy¹, Kirill M. Arzamasov¹, Igor M. Shulkin¹, Daria D. Kozhikhina¹, Lev D. Pestrenin¹

¹ Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

² Moscow Health Care Department, Moscow, Russian Federation

简评

论证。近年来，医疗数据集和人工智能软件技术开发的可达性，使得医疗诊断，特别是乳腺X线摄影的解决方案激增。这种登记为医疗设备的软件可被用于描述数乳腺X线摄影，这允许提供医疗服务时在很大程度上节省时间、物质和人力的资源，同时确保乳房预防性检查的质量。

该研究的目的是证明使用基于人工智能技术的软件对数字乳房X光照片进行第一次解读的可用性和有效性，同时保持放射科医生对X射线图像进行第二次描述的做法。

材料和方法。我们用在俄罗斯联邦登记为医疗设备的基于人工智能技术的软件处理了100张数字乳房X光照片的数据集，其中50张为“无目标病变”，50张为“存在目标病变”（有恶性肿瘤症状）。进行了ROC分析。研究局限性：诊断准确性度量值是基于人工智能技术的软件版本获得的，是在2022年底有效的。

结果。当设置为80.0%的灵敏度时，人工智能显示出90.0%的特异度（95% CI 81.7–98.3）和85.0%的准确性（95% CI 78.0–92.0）。当设置为100%的特异度时，人工智能显示出56.0%的灵敏度（95% CI 42.2–69.8）和78.0%的准确性（95% CI 69.9–86.1）。当设置为100%灵敏度时，人工智能的特异度为54.0%（95% CI 40.2–67.8），准确性为77.0%（95% CI 68.8–85.2）。提出了两种方法，涉及通过人工智能对数字乳腺X线摄影进行的第一次自主解读。第一种方法是利用人工智能评估X射线图像，其灵敏度高于放射科医生进行的双重乳腺X线摄影描述，特异度水平相当。第二种方法是基于人工智能技术的软件将对乳腺X线摄影进行分类（“无目标病变”或“存在目标病变”），表明其对结果的“信心”程度，取决于预测值所处的“走廊”。

结论。使用基于人工智能技术的软件对数字乳房X光照片进行第一次自主描述的两方案都能提供与放射科医生对图像进行的双重描述相同甚至更高的诊断质量。在全国范围内在实践中实现这种方法的经济效益可能是每年6亿至55亿卢布。

关键词：预防性研究；乳腺X线摄影；人工智能；诊断准确性。

To cite this article

Vasilev YuA, Tyrov IA, Vladzimirskyy AV, Arzamasov KM, Shulkin IM, Kozhikhina DD, Pestrenin LD. 使用人工智能对乳腺X线摄影结果进行两次查看：一种组织大规模预防性研究的新模式. *Digital Diagnostics*. 2023;4(2):93–104. DOI: <https://doi.org/10.17816/DD321423>

收到: 17.03.2023

接受: 25.04.2023

发布日期: 25.05.2023

BACKGROUND

Breast malignancies are a significant problem from medical, socioeconomic, and demographic points of view, heading the list of cancers and leading the mortality causes of the female population. A steady increase in prevalence from 45.24 to 53.43 per 100,000 population was observed from 2011 to 2019; later, it sharply decrease to 47.39 per 100,000 population in 2020, with a renewed rise in 2021 [1]. Such a pattern is accounted for by the suspension of mass screening during the COVID-19 pandemic. Conversely, it was a completely substantiated decision, and the situation demonstrates how vulnerable the healthcare system is. The emergency-related resource re-allocation took its toll on the socially significant diseases. However, even outside the pandemic, breast malignancies are still underdiagnosed, and the rate of newly diagnosed advanced diseases is high: as much as 27.0% of new cases are classified as stage III–IV cancer. A remarkable positive trend is noteworthy: the prevalence-to-incidence ratio over the reporting period is increasing steadily. In 2011, it was 9.5, while in 2021, it increased to 11.9. It is indicative of gradual improvements in the quality and efficacy of breast malignancy treatment [2].

Thus, the optimization of mass screening is warranted to expand coverage and population compliance; increase capacity, quality, and cost-efficiency; and enable sustainability and continuous accessibility. Given the evident progress in anti-cancer therapies, accomplishing these goals will take breast malignancy treatment to the next level.

Currently, the most common type of screening for breast malignancies is mammography. Following effective regulations, screening mammograms are subject to double reading, i.e., the images obtained with each patient should be viewed and interpreted by two independent radiologists. Such practice has been proved expedient by domestic and foreign authors. The cumulative rate of pathologic change detection is higher with double reading. Single reading lowers the sensitivity for all categories of breast imaging reporting and data system (BI-RADS) compared with double reading. Moreover, single reading is associated with various negative consequences for the examined patients [3, 4], although double reading also has its downsides, such as resource-intensiveness, quality issues, and funding difficulties.

Resource-intensiveness. In primary healthcare, two radiologists are required to interpret every screening image, with the vast majority of them being “target changes not found.” There is a risk for these positions to be filled in fictitiously to cover staff shortage, which would affect women’s health badly. Meanwhile, given the actual need for screening mammograms and the rate of the equipment fleet growth, the shortage of staff for these purposes is expected.

Quality issues. Interpreting mammograms requires specific skills in a narrow subarea of modern radiology. This worsens the staff shortage: formally employing more

radiologists will not contribute to timely precise detection of breast malignancies.

Funding difficulties. Double payment only applies if there is a separate service of mammography interpretation, and one payment is allocated for mammography scanning and its interpretation, whereas the other covers interpretation only. If this is not the case, funding difficulties arise: the payment is allocated only for combined mammography scanning and mammogram interpretation. Therefore, underfunding is common when only a single payment available, not covering the second mammogram reading.

The potential problems outlined can be solved with the use of artificial intelligence (AI) in interpreting mammograms.

High-quality readings obtained with such technology have been reported in the literature. Indeed, certain AI-based solutions have diagnostic precision similar to an average radiologist. The cumulative sensitivity, specificity, and area under the receiver operating curve (ROC) were 75.4, 90.6, and 0.89% for AI and 73.0, 88.6, and 0.85% for a radiologist, respectively, and no significant differences were found [5]. A meta-analysis [6] showed that the cumulative sensitivity, specificity, and area under the ROC for AI-based mammography interpretation were 91.4%, 91.6%, and 94.5%, respectively. The similarity coefficients for the intersection-over-union accuracy for abnormal change localizations segmented by AI and a radiologist were 0.86 and 0.96, respectively [7]. In our opinion, AI-based solutions should be implemented not just as discrete systems backing up the decisions made by doctors but as independent computerized processes.

This study aimed to justify a model for mass mammography screening using AI technology.

MATERIALS AND METHODS

The study was part of the Experiment for Computer Vision Innovations Used in Medical Image Analysis and its further use in the healthcare system of Moscow (hereinafter, the Moscow experiment), which took place in 2020 and was funded by the Government of Moscow (mosmed.ai). This study was feasible because of the accuracy of AI-based software demonstrated with 61,497 mammograms over the first year of the Moscow experiment [8].

Study design

This study has a mixed design, i.e., using a retrospective diagnostic study for the quantitative component and an analytical study for the qualitative component.

AI technology

AI-based software should be registered as a medical device to be used in routine clinical practice. AI-based software products by OOO “Medical Screening Systems” (Reg. No. RZN 2021/14449) and OOO “Third Opinion Platform” (Reg. No. RZN 2022/16534) intended for computerized

mammogram interpretation are registered as medical devices in the Russian Federation.

Such a product by one of the above companies was used for the study. As part of the Moscow experiment, the AI-based software was integrated into the Unified Radiology Information Service (ERIS) of the Unified Medical Information and Analytical System (EMIAS) of Moscow. Digital mammography images in the DICOM format were the input for the AI-based software. After the analysis, a text interpretation (DICOM SR) was generated, along with abnormal areas mapping (DICOM SC) and an overall disease probability. AI-generated results appeared on ERIS EMIAS along with the raw data. The overall probability values of cancer were used in the study. The correctness of BI-RADS rating and the precision of abnormal finding localization were not evaluated in this study.

Dataset

This study included 100 digital mammograms obtained as part of breast cancer screening, of which 50 demonstrated abnormalities. The mean age of the examined patients was 63 ± 6 years.

The images were categorized into “target changes not found” or “target changed detected” based on the consensus of two assessors (radiologists with >5 years of experience in mammography). The primary inclusion criterion for “target changed detected” was histological verification. Different opinions of assessors were an exclusion criterion for “target changes not found.” The exclusion criteria for both categories were age <18 years and low-quality images (PGMI score 1) identified by the assessor at the mapping stage for dataset preparation.

Mammography abnormalities consistent with BI-RADS categories 3–5 were classified as “target changed detected.” Mammography results consistent with BI-RADS categories 1 or 2, i.e., without any suspicion of breast malignancy, were classified as “target changes not found.”

The distribution based on the American College of Radiology types was as follows: A, $n = 26$; B, $n = 16$; C, $n = 5$; and D, $n = 3$ in “target changes not found,” and A, $n = 15$; B, $n = 24$; C, $n = 11$; and D, $n = 0$ in “target changed detected.”

Images included in the dataset were obtained with FUJIFILM Corporation (Japan) mammography machines. The following healthcare providers contributed to the dataset: City Polyclinic (CP) No. 22 Branch No. 1, Diagnostic Clinical Center No. 1, CP No. 8, CP No. 36, CP No. 22, CP No. 209, Diagnostic Center No. 2 Branch No. 4, Consultation and Diagnostic Polyclinic No. 121, Clinical and Diagnostic Center No. 4, and M.P. Konchalovsky City Clinical Hospital with the Moscow Healthcare Department, Outpatient Department No. 3.

Ethics review

The study was based on the results of the Moscow experiment and was approved by the Ethics Committee (Protocol abstract No. 2 NEK MRO RORR dated February 20, 2020; ClinicalTrials ID: NCT04489992).

Statistical analysis

Receiver operating curve (ROC) analysis available from a web tool was used for statistical data processing (<https://roc-analysis.mosmed.ai/>) [9, 10]. The true values were binary (0 for “target changes not found”; 1 for “target changes detected”). The result was the probability of cancer determined by AI-based software. Data obtained in the CSV format were uploaded to an electronic form, after which the web tool plotted a ROC. Cutoff values corresponding to the leftmost point with 100% sensitivity and the rightmost point with 100% specificity were determined in the interactive mode. Later, other diagnostic accuracy parameters for the established cutoff value were analyzed. A classic 2×2 contingency table was used for the analysis. Correct classification of a mammogram as “target changed detected” was considered true positive, and correct classification of a mammogram as “target changes not found” was considered true negative. Incorrect classification of a normal mammogram as “target changed detected” was considered false positive, and incorrect classification of an abnormal mammogram as “target changes not found” was considered false negative.

All statistical parameters presented in the results were calculated using 95% confidence interval (CI) by bootstrapping with 1,000 iterations.

RESULTS

The following model was suggested: the attending physician refers the patient to mammography screening in accordance with the current regulations and clinical guidelines. The X-ray technologist performs the examination. The resulting digital mammograms are sent to the archive of medical images as part of the medical information system of a medical organization and/or a healthcare information system of a constituent entity of the Russian Federation. The first reading is performed by the software (an approved AI-based medical device). Following the first reading, an automatic electronic medical record¹ is formed in the information system, containing (a) the series of images with graphic marks and/or a temperature map of abnormal areas, if any; (b) a structured report, brief user guide, conclusion, details, and cancer probability. The second reading is performed by a radiologist. Based on the second reading, an electronic interpretation protocol and conclusion are provided in the information system.

¹ Medical records generated automatically by the approved medical devices that do not require the electronic signature of a healthcare professional (in accordance with Order of the Ministry of Health of the Russian Federation No. 947n dated September 7, 2020, On approval of the procedure for electronic documents turnover in healthcare).

A study was performed to establish whether AI-based software can provide the required level of diagnostic accuracy. The ROC for the studied AI-based software is shown in Fig. 1. The distribution of mammograms by groups (true positive, true negative, false negative, and false positive) depending on the selected cutoff value is presented in Table 1.

When setting a cutoff value of 0.93 corresponding to 100.0% specificity, AI correctly identified mammography from the “target changes detected” group, i.e., no false positives were recorded. Of the 50 mammograms in the “target changes detected” group, AI correctly identified 28 (56.0%) images at the specified threshold setting. With these settings of the AI-based software, sensitivity, specificity, and diagnostic

accuracy were 56.0% (95% CI 42.2–69.8), 100.0% (95% CI 100.0–100.0), and 78.0% (95% CI 69.9–86.1), respectively.

When setting a cutoff value of 0.25 corresponding to 100.0% sensitivity, no false negatives were observed, and 27 true negatives were identified (54.0% of all images in the “target changes not found” group). With these settings of the AI-based software, the sensitivity, specificity, and diagnostic accuracy were 100.0% (95% CI 100.0–100.0), 54.0% (95% CI 40.2–67.8), and 77.0% (95% CI 68.8–85.2), respectively.

When setting a cutoff value of 0.82 to maximize the Youden index, the sensitivity was 80.0%, 45 were true-negative results (90.0% of all images in the “Target changes not found” group), and 40 were true-positive results (80.0%

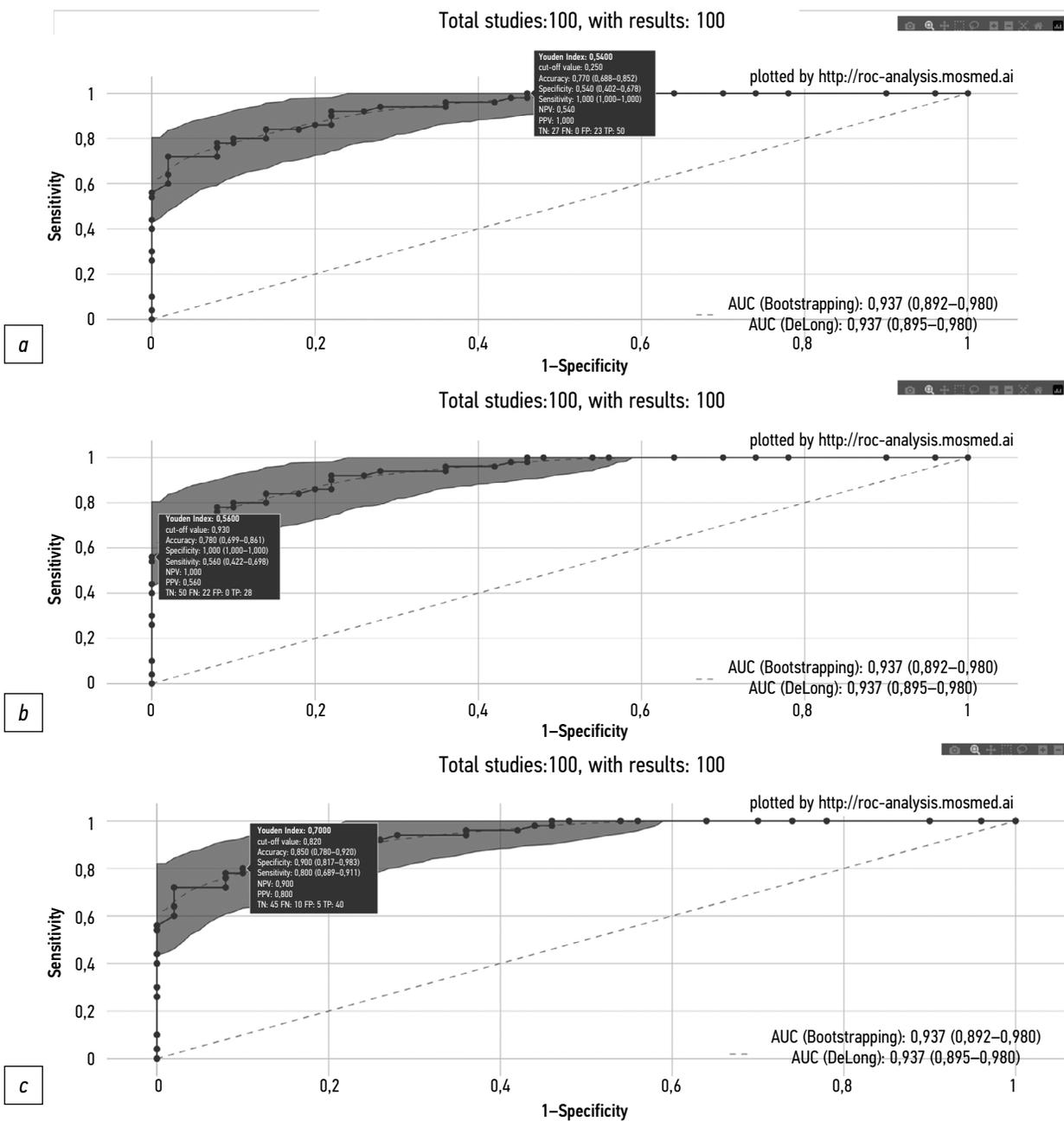


Fig. 1. ROC for AI-based software. The highlight shows the 95% confidence interval. Experimental values corresponding to 100.0% sensitivity (a), 100.0% specificity (b), and 80.0% sensitivity (c) are highlighted individually. For each experimental point, the rectangle shows the diagnostic accuracy metrics at the corresponding cutoff value.

Table 1. 2 × 2 contingency table for different threshold values

Group	No. of scans	Result	100% sensitivity	100% specificity	Balanced sensitivity and specificity
Target changes detected	50	True positive	50	28 56.0%**	40 80.0%**
		False positive	23	0	5
Target changes not found	50	True negative	27 54.0%*	50	45 90.0%*
		False negative	0	22	10

Note: * The percentage of true-negative results is calculated from the total “target changes not found” images. ** The percentage of true-positive results is calculated from the total “target changes detected” images.

of all images in the “target changes detected” group). With these settings of the AI-based software, the sensitivity, specificity, and diagnostic accuracy were 80.0% (95% CI 68.9–91.1), 90.0% (95% CI 81.7–98.3), and 85.0% (95% CI 78.0–92.0), respectively.

DISCUSSION

Summary of the primary outcome

Setting different cutoff values for AI-based software used for description and interpretation of mammography data allows achieving the sensitivity, specificity, and accuracy metrics that correspond to or exceed those for double reading of mammogram by radiologists.

Discussion of the primary outcome

In the Russian Federation, approximately 8.2 million mammographies² are performed annually as part of screening, and their double reading requires a significant amount of time, staff, and financial resources. The use of AI-based software for the first review of mammograms will reduce the above costs while maintaining or even improving diagnostic quality. Two approaches were proposed to setting up AI-based software for the first reading of mammograms.

The first approach involves the use of AI-based software with balanced values of sensitivity and specificity. In our case, the sensitivity was 80.0%, which exceeds the sensitivity of double reading of mammograms by radiologists determined in reviews (72.0%–73.0%) [5, 6]. In this study, the specificity of AI (90.0%) is not inferior to that of two radiologists (88.0%–98.0%) [5, 6]. AI-based software in combination with the assessment by one radiologist will have a higher overall accuracy of mammography interpretation than interpretation by only one radiologist, which is confirmed by a number of scientific publications [11–13]. An electronic medical record containing a conclusion on the category of the image (“target changes not found” or “Target changes detected”) will be generated by AI-based software with this approach.

The second approach implies that the AI-based software will determine the category of the image (“target changes not found” or “target changes detected”), indicating the degree of its “confidence” in the result. The general concept of the method is shown in Fig. 2.

As mentioned in the Results, cutoff values were determined for the predicted values at 100% sensitivity and 100% specificity (0.25 and 0.93, respectively) when plotting the ROC for the AI-based software. Based on these data, it is proposed to contribute to the predicted value to one of three “corridors,” which correspond to different classification results and different degrees of AI “confidence”:

- 1) Green corridor: the predicted values are within the range of 0–0.25 and correspond to the category “target changes not found” with 100% confidence,
- 2) Red corridor: the predicted values are within the range of 0.93–1.0 and correspond to the category “target changes detected” with 100% confidence,
- 3) Yellow corridor: the predicted values are within the range of 0.25–0.93 inclusive and correspond to the “target changes not found” or “target changes detected” category; however, the probability of correct classification is <100%.

The predicted value and color of the corridor in which it falls are proposed to be added to the description of mammography by the AI-based software. With this information, the radiologist making the second reading after the AI step will know how much they can rely on the results. This will help the doctor stay alert when examining a mammogram from the yellow corridor. In the long term, this approach can increase the confidence of radiologists in the results of AI-based software because, despite the current high level of accuracy, AI is not yet able to correctly classify 100% of the analyzed images with a high degree of confidence.

The advantage of the second approach is not only the ability of AI to categorize some mammograms as “target changes not found” or “target changes detected” with a 100% degree of confidence (green and red corridors for the

² I.E. Tyurin. 2020 Report by the Chief Independent Expert of the Ministry of Health of Russia on radiation and instrumental diagnostics [electronic resource]. Accessed at: https://static-0.minzdrav.gov.ru/system/attachments/attaches/000/056/620/original/Отчет_за_2020_год_Тюрин.pdf?1624967722.

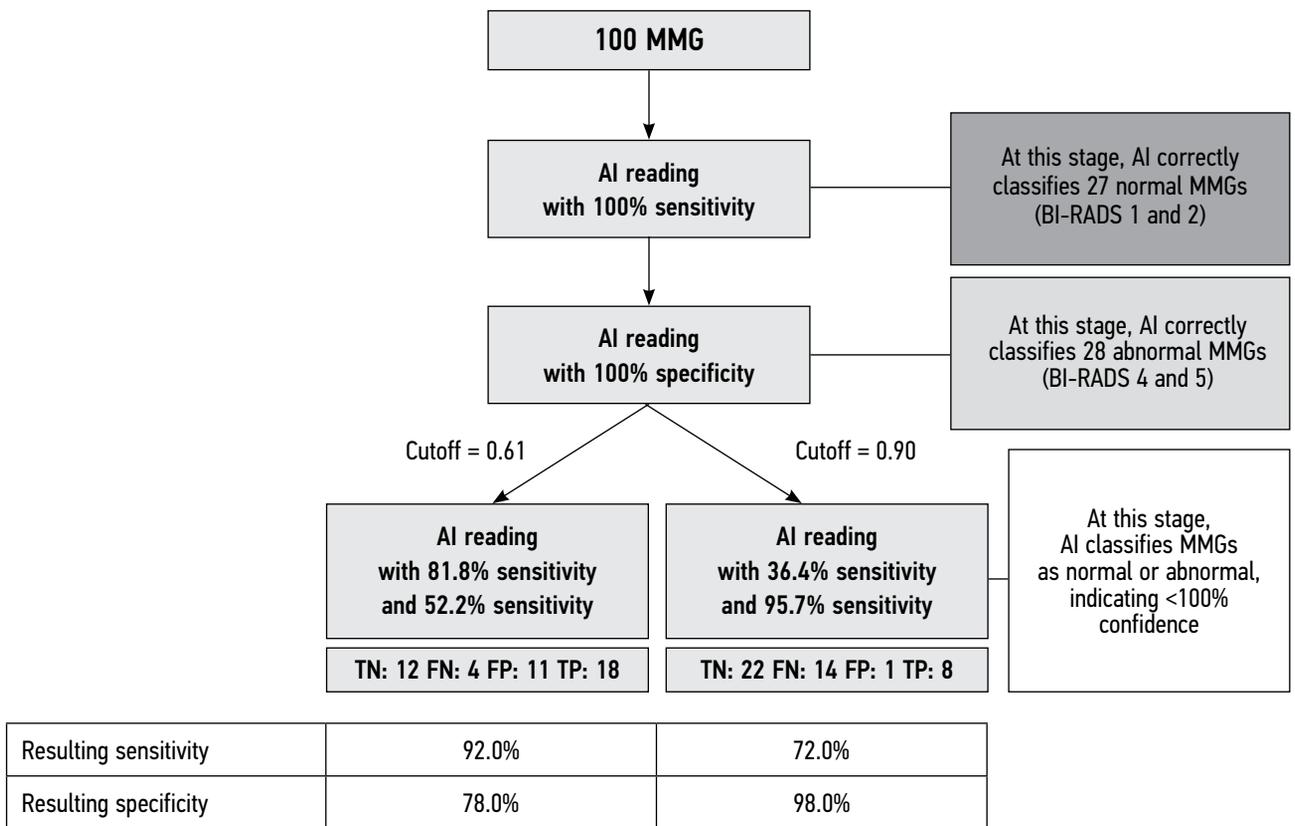


Fig. 2. Concept of an approach to the first mammogram reading using artificial intelligence involving binary image classification with an indication of the degree of confidence of the AI-based software in the results obtained.

Note: AI, artificial intelligence; FN, false negative; FP, false positive; MMG, mammogram; TN, true negative; TP, true positive.

predicted values) but also the ability to change the cutoff value to balance the sensitivity and specificity of AI-based software for the analysis of mammograms that fall into the yellow corridor. Depending on the clinical task, a higher sensitivity can be set, which will provide better detection of pathology with the lowest number of false-negative results or higher specificity to reduce the number of false-positive results (Table 2) [5, 6].

The results of this study demonstrate the possibility of using AI-based software for the first reading of mammograms; however, in the future, the software must be optimized to more effectively distinguish between “target changes not found” and “target changes detected.” Only high-quality mammograms were initially selected for this study.

However, modern AI-based software has the function of offline mammography quality control. When introduced into routine practice, AI-based software can perform technical assessment of image quality and clinical assessment [8].

Economic justification of AI-based double reading of mammograms

As part of the study, payment rates for medical care provided under the territorial program of compulsory medical insurance adopted in the constituent entities of the Russian Federation for 2023 were analyzed.

In 19 constituent entities (22.4% of all constituent entities of the Russian Federation), a separate payment for medical service A06.20.004 mammography (provided as

Table 2. Sensitivity and specificity with different approaches to mammogram reading

Screening mammography results	Sensitivity, %	Specificity, %
Double reading by two radiologists *	72.0–73.0	88.0–98.0
First approach to using AI for the first mammogram reading (binary classification)	80.0	90.0
Second approach to using AI for the first mammogram reading (binary classification with a degree of confidence) with a cutoff value of 0.61	92.0	78.0
Second approach to using AI for the first mammogram reading (binary classification with a degree of confidence) with a cutoff value of 0.90	72.0	98.0

Note: * Based on literature data [5, 6].

part of screening) is available. In 4 out of 19 constituent entities, a separate payment for medical service A06.30.002 description and interpretation of radiographic images (second reading of mammograms) is also available. In all other constituent entities of the Russian Federation, the payment for a comprehensive service (without specifying the mammography-related services included in it) is charged at the first stage of screening of the adult population.

The cost of mammogram description varies from 114.97 to 1034.93 rubles. As of March 1, 2023, the description and interpretation of mammography data using AI is available only as part of the Moscow experiment [8]. According to the rate agrees for medical care provided under the territorial program of compulsory medical insurance in Moscow, this medical service costs 239.00 rubles³.

In this study, two approaches to determining the required amount of funding for screening mammograms in the Russian Federation were analyzed. The first approach was to perform calculations based on the cost of medical services in Moscow for 2023. A description of mammography by a radiologist costs 178.00 rubles. Therefore, a double reading of each mammogram by radiologists will cost 356.00 rubles. In turn, mammography description by AI and a radiologist, as mentioned above, costs 239.00 rubles. Thus, with an average number of annual mammograms of 8.2 million in Russia, double reading by two radiologists will cost 2.9 billion rubles, and double reading by AI and a radiologist will cost 1.9 billion rubles. Potential savings through the use of AI-based software may account to 1.0 billion rubles annually. The second approach was to perform calculations based on the cost of medical services in the constituent entities of the Russian Federation for 2023. The percentage of money saved was also considered, thanks to the interpretation of mammography by AI and a radiologist compared with double reading by two radiologists in Moscow, which amounted to the following:

$$\frac{178,00 \times 2 - 239,00}{178,00 \times 2} \times 100\% = 32,8\%.$$

Mammogram reading by a radiologist in the constituent entities of the Russian Federation costs 114.97–1034.93 rubles, which means that double reading by radiologists costs 229.94–2069.86 rubles. Assuming that the interpretation of mammography by AI and a radiologist in the constituent entities of the Russian Federation is cheaper than double reading by radiologists by 32.8% (as is the case in Moscow), the resulting cost of double mammogram reading by AI

and a radiologist will range from 154.51 to 1390.94 rubles. Thus, with an average number of annual mammograms of 8.2 million in Russia, double reading by two radiologists will cost 1.8–16.9 billion rubles, whereas double reading by AI and a radiologist will cost 1.2–11.4 billion rubles. Potential savings due to the use of AI-based software at the national level may amount to 0.6–5.5 billion rubles annually.

Study limitations

The determined values of diagnostic accuracy metrics are valid for AI-based software versions as of the end of 2022. For patients in the “target changes not found” group, changes over time on the BI-RADS scale were not evaluated, which can be regarded as a study limitation.

CONCLUSION

The results of this study show the feasibility and prospects of using AI for the first reading of mammograms. AI-based software (registered as a medical device) has sensitivity and specificity non-inferior or superior to those of two radiologists. The model for using AI-based software for the first reading combined with the second reading by a radiologist allows for nationwide economic benefits amounting to 0.6–5.5 billion rubles annually.

ADDITIONAL INFORMATION

Funding source. This article was prepared by a group of authors as a part of the research and development effort titled “Evidence-based methodologies for sustainable development of artificial intelligence in medical imaging”, registration number to EGISU: 123031500004-5.

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. Yu.A. Vasilev, I.A. Tyrov — research concept, final proofreading of the text; A.V. Vladzmyrsky — research concept and design, writing and editing of the text; K.M. Arzamasov — research concept and design, analysis of the data obtained, writing and editing of the text; I.M. Shulkin — research design, collection and processing of materials, final proofreading the text; D.D. Kozhikhina — research design, collection and processing of materials; L.D. Pestrenin — analysis of the data obtained, writing the text.

REFERENCES

1. Malignant neoplasms in Russia in 2021 (morbidity and mortality). Ed by A.D. Kaprin, V.V. Starinsky, A.O. Shahzadova. Moscow; 2022. 252 p. (In Russ).
2. The state of oncological assistance to the population of Russia in 2021. Ed by A.D. Kaprin, V.V. Starinsky, A.O. Shahzadova. Moscow; 2022. 239 p. (In Russ).

³ Rate agreement for medical care provided under the Territorial Program of Compulsory Medical Insurance for Moscow in 2023 (Concluded in Moscow on December 30, 2022). Accessed at: <https://www.consultant.ru/law/review/208575247.html>.

3. Chen Y, James JJ, Michalopoulou E, et al. Performance of radiologists and radiographers in double reading mammograms: The UK national health service breast screening program. *Radiology*. 2023;306(1):102–109. doi: 10.1148/radiol.212951
4. Euler-Chelpin MV, Lillholm M, Napolitano G, et al. Screening mammography: Benefit of double reading by breast density. *Breast Cancer Res Treat*. 2018;171(3):767–776. doi: 10.1007/s10549-018-4864-1
5. Hickman SE, Woitek R, Le EP, et al. Machine learning for workflow applications in screening mammography: Systematic review and meta-analysis. *Radiology*. 2022;302(1):88–104. doi: 10.1148/radiol.2021210391
6. Liu J, Lei J, Ou Y, et al. Mammography diagnosis of breast cancer screening through machine learning: A systematic review and meta-analysis. *Clin Exp Med*. 2022. doi: 10.1007/s10238-022-00895-0
7. Rozhkova NI, Rojtberg PG, Varfolomeeva AA, et al. Neural network-based segmentation model for breast cancer X-ray screening. *Sechenov medical journal*. 2020;11(3):4–14 (In Russ). doi: 10.47093/2218-7332.2020.11.3.4-14
8. Vasilev JA, Vladzimirskyy AV. Computer vision in radiology: The first stage of the Moscow experiment: Monograph. Moscow: Izdatel'skie resheniya; 2022. 388 p. (In Russ).
9. Patent RUS № 2022617324/05.04.2022. Byul. № 4. Morozov SP, Andreichenko AE, Chetverikov SF, et al. A web-based tool for performing ROC analysis of diagnostic test results. Available from: <https://www.elibrary.ru/item.asp?id=48373757>. Accessed: 10.03.2023. (In Russ).
10. Morozov SP, Vladzimirskyy AV, Klyashtornyy VG, et al. Clinical acceptance of software based on artificial intelligence technologies (radiology). Moscow; 2019. 45 p. (Ser. Best practices in medical imaging).
11. Schaffter T, Buist DS, Lee CI, et al. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. *JAMA Netw Open*. 2020;3(3):e200265. doi: 10.1001/jamanetworkopen.2020.0265
12. Wan Y, Tong Y, Liu Y, et al. Evaluation of the combination of artificial intelligence and radiologist assessments to interpret malignant architectural distortion on mammography. *Front Oncol*. 2022;(12):880150. doi: 10.3389/fonc.2022.880150
13. Leibig C, Brehmer M, Bunk S, et al. Combining the strengths of radiologists and AI for breast cancer screening: A retrospective analysis. *Lancet Digit Health*. 2022;4(7):e507–e519. doi: 10.1016/S2589-7500(22)00070-X

СПИСОК ЛИТЕРАТУРЫ

1. Злокачественные новообразования в России в 2021 году (заболеваемость и смертность) / под ред. А.Д. Каприна, В.В. Старинского, А.О. Шахзадовой. Москва, 2022. 252 с.
2. Состояние онкологической помощи населению России в 2021 году / под ред. А.Д. Каприна, В.В. Старинского, А.О. Шахзадовой. Москва, 2022. 239 с.
3. Chen Y., James J.J., Michalopoulou E., et al. Performance of radiologists and radiographers in double reading mammograms: The UK national health service breast screening program // *Radiology*. 2023. Vol. 306, N 1. P. 102–109. doi: 10.1148/radiol.212951
4. Euler-Chelpin M.V., Lillholm M., Napolitano G., et al. Screening mammography: Benefit of double reading by breast density // *Breast Cancer Res Treat*. 2018. Vol. 171, N 3. P. 767–776. doi: 10.1007/s10549-018-4864-1
5. Hickman S.E., Woitek R., Le E.P., et al. Machine learning for workflow applications in screening mammography: Systematic review and meta-analysis // *Radiology*. 2022. Vol. 302, N 1. P. 88–104. doi: 10.1148/radiol.2021210391
6. Liu J., Lei J., Ou Y., et al. Mammography diagnosis of breast cancer screening through machine learning: A systematic review and meta-analysis // *Clin Exp Med*. 2022. doi: 10.1007/s10238-022-00895-0
7. Рожкова Н.И., Ройтберг П.Г., Варфоломеева А.А., и др. Сегментационная модель скрининга рака молочной железы на основе нейросетевого анализа рентгеновских изображений // *Сеченовский вестник*. 2020. Т. 11, № 3. С. 4–14. doi: 10.47093/2218-7332.2020.11.3.4-14
8. Компьютерное зрение в лучевой диагностике: первый этап Московского эксперимента. Монография / под ред. Ю.А. Васильева, А.В. Владзимиурского. Москва: Издательские решения, 2022. 388 с.
9. Патент РФ на изобретение № 2022617324/05.04.2022. Бюл. № 4. Морозов С.П., Андрейченко А.Е., Четвериков С.Ф., и др. Веб-инструмент для выполнения ROC-анализа результатов диагностических тестов. Режим доступа: <https://www.elibrary.ru/item.asp?id=48373757>. Дата обращения: 10.03.2023.
10. Morozov S.P., Vladzimirskyy A.V., Klyashtornyy V.G., et al. Clinical acceptance of software based on artificial intelligence technologies (radiology). Moscow, 2019. 45 p. (Сер. Best practices in medical imaging).
11. Schaffter T., Buist D.S., Lee C.I., et al. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms // *JAMA Netw Open*. 2020. Vol. 3, N 3. P. e200265. doi: 10.1001/jamanetworkopen.2020.0265
12. Wan Y., Tong Y., Liu Y., et al. Evaluation of the combination of artificial intelligence and radiologist assessments to interpret malignant architectural distortion on mammography // *Front Oncol*. 2022. N 12. P. 880150. doi: 10.3389/fonc.2022.880150
13. Leibig C., Brehmer M., Bunk S., et al. Combining the strengths of radiologists and AI for breast cancer screening: A retrospective analysis // *Lancet Digit Health*. 2022. Vol. 4, N 7. P. e507–e519. doi: 10.1016/S2589-7500(22)00070-X

AUTHORS' INFO

* **Lev D. Pestrenin**, Junior Research Associate;
address: 24/1 Petrovka street, 127051 Moscow, Russia;
ORCID: 0000-0002-1786-4329;
eLibrary SPIN: 7193-7706;
e-mail: PestreninLD@zdrav.mos.ru

ОБ АВТОРАХ

* **Пестренин Лев Дмитриевич**, мл. науч. сотр.;
адрес: Россия, 127051, Москва, ул. Петровка, д. 24, стр. 1;
ORCID: 0000-0002-1786-4329;
eLibrary SPIN: 7193-7706;
e-mail: PestreninLD@zdrav.mos.ru

* Corresponding author / Автор, ответственный за переписку

Yuriy A. Vasilev, MD, Cand. Sci. (Med);
ORCID: 0000-0002-0208-5218;
eLibrary SPIN: 4458-5608;
e-mail: npcmmr@zdrav.mos.ru

Ilya A. Tyrov;
ORCID: 0000-0001-9337-624X;
eLibrary SPIN: 8625-3458;
e-mail: npcmmr@zdrav.mos.ru

Anton V. Vladzimirskyy, MD, Dr. Sci. (Med), Professor;
ORCID: 0000-0002-2990-7736;
eLibrary SPIN: 3602-7120;
e-mail: npcmmr@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

Igor M. Shulkin;
ORCID: 0000-0002-7613-5273;
eLibrary SPIN: 5266-0618;
e-mail: npcmmr@zdrav.mos.ru

Daria D. Kozhikhina;
ORCID: 0000-0001-7690-8427;
eLibrary SPIN: 5869-3854;
e-mail: npcmmr@zdrav.mos.ru

Васильев Юрий Александрович, канд. мед. наук;
ORCID: 0000-0002-0208-5218;
eLibrary SPIN: 4458-5608;
e-mail: npcmmr@zdrav.mos.ru

Тыров Илья Александрович;
ORCID: 0000-0001-9337-624X;
eLibrary SPIN: 8625-3458;
e-mail: npcmmr@zdrav.mos.ru

Владзimirский Антон Вячеславович, д-р мед. наук, профессор;
ORCID: 0000-0002-2990-7736;
eLibrary SPIN: 3602-7120;
e-mail: npcmmr@zdrav.mos.ru

Арзамасов Кирилл Михайлович, канд. мед. наук;
ORCID: 0000-0001-7786-0349;
eLibrary SPIN: 3160-8062;
e-mail: ArzamasovKM@zdrav.mos.ru

Шулькин Игорь Михайлович;
ORCID: 0000-0002-7613-5273;
eLibrary SPIN: 5266-0618;
e-mail: npcmmr@zdrav.mos.ru

Кожихина Дарья Дмитриевна;
ORCID: 0000-0001-7690-8427;
eLibrary SPIN: 5869-3854;
e-mail: npcmmr@zdrav.mos.ru