

DOI: <https://doi.org/10.17816/DD321971>

# Методология тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики

Ю.А. Васильев, А.В. Владзимирский, О.В. Омелянская, К.М. Арзамасов,  
С.Ф. Четвериков, Д.А. Румянцев, М.А. Зеленова

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

## АННОТАЦИЯ

**Обоснование.** Мировая сумма инвестиций в компании по разработке программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики составила 80 млн долларов в 2016 году, 152 млн долларов — в 2017 и, ожидаемо, продолжает расти. Активная деятельность компаний-производителей программного обеспечения должна соответствовать существующим клиническим, биоэтическим, правовым и методологическим основам и стандартам. Как на национальном, так и на международном уровне не существует единых стандартов и протоколов проведения испытаний и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики.

**Цель** — разработать универсальную методологию тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики, направленную на повышение его качества и внедрение в практическое здравоохранение.

**Материалы и методы.** В ходе аналитического этапа был проведён обзор литературы по базам данных PubMed и eLIBRARY. Практический этап включал апробацию разработанной методологии в рамках Эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы.

**Результаты.** Разработана методология тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики, направленная на повышение качества данного программного обеспечения и его внедрение в практическое здравоохранение. Методология состоит из 7 этапов: самотестирование, функциональное тестирование, калибровочное тестирование, технологический мониторинг, клинический мониторинг, обратная связь и доработка.

**Заключение.** Отличительными особенностями методологии являются цикличность этапов тестирования, мониторинга и доработки программного обеспечения, приводящие к постоянному повышению его качества, наличие подробных требований к результатам его работы, участие врачей в его оценке. Методология позволит разработчикам программного обеспечения достичь высоких результатов и продемонстрировать достижения в различных направлениях, а пользователям — сделать осознанный и уверенный выбор среди программ, прошедших независимую и всестороннюю проверку качества.

**Ключевые слова:** программное обеспечение; искусственный интеллект; рентгенология; диагностическая визуализация; методология; контроль качества.

## Как цитировать:

Васильев Ю.А., Владзимирский А.В., Омелянская О.В., Арзамасов К.М., Четвериков С.Ф., Румянцев Д.А., Зеленова М.А. Методология тестирования и мониторинга программного обеспечения на основе технологий искусственного интеллекта для медицинской диагностики // *Digital Diagnostics*. 2023. Т. 4, № 3. С. 252–267. DOI: <https://doi.org/10.17816/DD321971>

DOI: <https://doi.org/10.17816/DD321971>

# Methodology for testing and monitoring artificial intelligence-based software for medical diagnostics

Yuriy A. Vasilev, Anton V. Vladzimirskyy, Olga V. Omelyanskaya, Kirill M. Arzamasov, Sergey F. Chetverikov, Denis A. Rumyantsev, Maria A. Zelenova

Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

## ABSTRACT

**BACKGROUND:** The global amount of investment in companies developing artificial intelligence (AI)-based software technologies for medical diagnostics reached \$80 million in 2016, rose to \$152 million in 2017, and is expected to continue growing. While software manufacturing companies should comply with existing clinical, bioethical, legal, and methodological frameworks and standards, there is a lack of uniform national and international standards and protocols for testing and monitoring AI-based software.

**AIM:** This objective of this study is to develop a universal methodology for testing and monitoring AI-based software for medical diagnostics, with the aim of improving its quality and implementing its integration into practical healthcare.

**MATERIALS AND METHODS:** The research process involved an analytical phase in which a literature review was conducted on the PubMed and eLibrary databases. The practical stage included the approbation of the developed methodology within the framework of an experiment focused 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.

**RESULTS:** A methodology for testing and monitoring AI-based software for medical diagnostics has been developed, aimed at improving its quality and introducing it into practical healthcare. The methodology consists of seven stages: self-testing, functional testing, calibration testing, technological monitoring, clinical monitoring, feedback, and refinement.

**CONCLUSION:** Distinctive features of the methodology include its cyclical stages of monitoring and software development, leading to continuous improvement of its quality, the presence of detailed requirements for the results of the software work, and the participation of doctors in software evaluation. The methodology will allow software developers to achieve significant outcomes and demonstrate achievements across various areas. It also empowers users to make informed and confident choices among software options that have passed an independent and comprehensive quality check.

**Keywords:** software; artificial intelligence; radiology; diagnostic imaging; methodology; quality control.

## To cite this article:

Vasilev YuA, Vladzimirskyy AV, Omelyanskaya OV, Arzamasov KM, Chetverikov SF, Rumyantsev DA, Zelenova MA. Methodology for testing and monitoring artificial intelligence-based software for medical diagnostics. *Digital Diagnostics*. 2023;4(3):252–267. DOI: <https://doi.org/10.17816/DD321971>

DOI: <https://doi.org/10.17816/DD321971>

# 基于人工智能技术的医疗诊断软件测试和监测方法学

Yuriy A. Vasilev, Anton V. Vladzimirskyy, Olga V. Omelyanskaya, Kirill M. Arzamasov, Sergey F. Chetverikov, Denis A. Rumyantsev, Maria A. Zelenova

Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

## 简评

**论证。**2016年，全球对基于人工智能技术开发医疗诊断软件的公司投资额为8000万美元，2017年为1.52亿美元，并预料还将继续增长。软件公司的积极活动必须符合现有的临床、生物伦理、法律和方法学原理和标准。在国家和国际范围，基于人工智能技术的软件还没有统一的测试和监测标准和协议。

**该研究的目的是**开发一种通用方法，用于测试和监测基于人工智能技术的医疗诊断软件，以提高其质量和在实际医疗中的应用。

**材料和方法。**在分析阶段，对PubMed和eLIBRARY数据库进行了文献综述。实用阶段包括在《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》框架内对所开发的方法学进行了测试，并将其进一步应用于莫斯科的医疗保健系统。

**结果。**我们开发了一套基于人工智能技术的医疗诊断软件测试和监测方法，旨在提高该软件的质量，并将其应用于实用保健中。该方法包括7个阶段：自测试、功能测试、校准测试、技术监测、临床监测、反馈和改进。

**结论。**该方法的显著特点是对软件进行周期性的监测和改进，从而不断提高其质量；对软件性能结果并医生参与软件评估提出详细要求。该方法可使软件开发人员在各个领域取得优异成绩并展示成就，也可使用户在通过独立、全面质量控制的程序中做出明智、自信的选择。

**关键词：**软件；人工智能；放射学；诊断成像；方法学；质量控制。

## 引用本文：

Vasilev YuA, Vladzimirskyy AV, Omelyanskaya OV, Arzamasov KM, Chetverikov SF, Rumyantsev DA, Zelenova MA. 基于人工智能技术的医疗诊断软件测试和监测方法学. *Digital Diagnostics*. 2023;4(3):252–267. DOI: <https://doi.org/10.17816/DD321971>

收到: 06.04.2023

接受: 15.06.2023

发布日期: 30.08.2023

## 论证

2016年，全球对基于人工智能（AI）技术开发医疗诊断软件（software）的公司的投资额为8000万美元，2017年为1.52亿美元，并预料还将继续增长[1]。2019年在俄罗斯，基于人工智能技术的医疗诊断软件得到广泛推广，当时莫斯科市政府决定开展一项仍在进行（2023年）的大规模科学研究，即《关于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》（以下简称实验）<sup>1</sup>。

软件公司的积极活动必须符合现有的临床、生物伦理、法律和方法学原理和标准[1]。根据俄罗斯法律，为了将基于人工智能技术的软件用于实用保健，而不是用于实验情境，必须按照法律规定的程序将其作为医疗器械进行国家注册，从而使软件获得Roszdravnadzor<sup>2</sup>颁发的注册证书（RC）。

在将软件送交注册之前，一个强制性的准备阶段是进行技术和临床试验，以验证声明的软件功能<sup>3</sup>。同时，由于基于人工智能技术的软件的特殊性，在国家和国际范围内还没有统一的试验标准和协议，也没有关于工作程序和决策原则的方便用户的信息[2]。美国食品药品监督管理局（Food and Drug Administration, FDA）也正在制定评估和管理基于人工智能技术软件的精准建议[1]。由于无法可靠地确认软件是否符合对其提出的要求，可能存在极为不利的后果：用户对软件不信任，阻碍软件在实用保健中的应用，软件的使用失去积极的社会经济效益，阻碍医疗保健的整体发展[3]。

软件收到RC后，将进行注册后临床监测。监测的目的是确保其在实用保健中使用的安全<sup>4</sup>。然而，现有的监测要求对不同的医疗器械都是通用的，也不考虑到基于人工智能技术的医疗诊断软件的特殊性[4]。根据欧亚经济委员会理事会的决议，在收到RC后的三年内每年应进行一次对第三类风险医疗器械（包括基于人工智能技术的软件）的监测<sup>5</sup>。然而，由于医疗数据的可变性较高，且难以预测周围条件（如流行病情况）的变化，因此有必要进行更频繁的监测[5]。在监测过程中，有可能会发现软件性能的批评意见。在这种情况下，有必要对软件进

行改进，一旦软件得到改进，就需要重新进行测试和监测。

医疗诊断中可通过临床试验对基于人工智能技术的软件进行研究，而回顾性队列研究是最合适的临床试验类型[1]。这种评估软件的方法也有很多缺点。主要缺点是在实践中获得的软件性能结果与研究结果之间存在差异[1]。在实用保健中，采用基于人工智能技术的医疗诊断软件的方法不完善的一个普遍例子是第一个用于乳腺X线筛查的计算机辅助诊断（computer-aided diagnostic, CAD）系统的负面经验。大型多中心研究结果表明了，由于使用该软件，乳腺癌的检出率提高了2-10%[6]。1998年，该软件获得了FDA的批准，以医生开始在实用保健中使用该软件。然而，该软件在实用保健中并没有取得积极的效果，甚至导致了乳房摄影检查的检出率下降，假阳性率上升[6]。文献中提出的一种解释是，不同经验水平的放射科医生使用新技术的方式不同。经验丰富的医生没有注意到这一点。而经验较少的放射科医生则因新技术带来的虚假安全感而犯错。第二种解释是，该软件未能有效检测出某些癌症，早期研究未发现这一点[1]。

因此，虽然伦理和法律问题是基于人工智能技术软件的最大特点，但还有一个重要的方法问题，可以表述为缺乏一种通用的、全面的方法来测试和监测基于人工智能技术的医疗诊断软件；这种方法应以提高软件质量和进一步在实用保健中采用为目标[7]。综上所述，制定这样一种方法似乎迫在眉睫。值得注意的是，该方法不会取代现有的、经法律批准的软件安全和效率评估方法，而是会独立存在，并作为一种补充，从而增加软件成功获得俄罗斯联邦卫生监督局注册证书的机会，并在获得批准后进一步评估和改进软件，以便在实用保健中有效采用。

**该研究的目的是**开发一种通用方法，用于测试和监测基于人工智能技术的医疗诊断软件，以提高其质量并在实际医疗保健中的应用。

## 材料和方法

### 研究设计

本文是一项分析性研究。所介绍的方法是在分析文献和作者自己经验的基础上制定的。

<sup>1</sup> 莫斯科市政府于2019年11月21日第1543-PP号决议《关于使用创新计算机视觉技术进行医学图像分析和进一步应用于莫斯科市医疗系统的实验》。访问方式：<https://docs.cntd.ru/document/563879961>。

<sup>2</sup> 俄罗斯联邦政府于2020年11月24日第1906号决议《关于对《医疗器械国家注册规则》进行修改》。访问方式：<http://publication.pravo.gov.ru/Document/View/0001202011270010>。

<sup>3</sup> 联邦法2011年11月21日第323-FZ号《关于俄罗斯联邦公民健康保护的基本原则》。第38条《医疗器械》。访问方式：[https://www.consultant.ru/document/cons\\_doc\\_LAW\\_121895/ddcfddbdbb49e64f085b65473218611b4bb6cd65/](https://www.consultant.ru/document/cons_doc_LAW_121895/ddcfddbdbb49e64f085b65473218611b4bb6cd65/)。

<sup>4</sup> 俄罗斯联邦卫生部于2020年9月15日第980n号命令《关于医疗器械安全监测程序的批准》。访问方式：<https://docs.cntd.ru/document/566006416>。

<sup>5</sup> 欧亚经济委员会理事会于2015年12月22日第174号决议《关于医疗器械安全、质量和有效性监测规则的批准》。访问方式：<https://www.alt.ru/tamdoc/15kr0174/>。

## 制定方法

制定该方法的过程分为两个阶段，即分析和实践。

在分析阶段，对现有方法进行了研究。在电子科学图书馆（PubMed和eLIBRARY）中对2018年至2023年（最近5年）的文献进行了综述。搜索词包括“methodology for evaluation artificial intelligence in radiology”，“методология оценки искусственного интеллекта в рентгенологии”（放射学人工智能评估方法）。我们通过阅读文章标题和简评评估其相关性后，将其纳入分析范围。共分析了22篇文章[1-22]和5项规范性法律文件<sup>6</sup>。

然后，在《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》框架内对该方法进行了测试。该实验是在莫斯科统一医疗信息和分析系统（Unified Medical Information and Analytical System, EMIAS）的统一放射科信息系统（Unified Radiological Information System, URIS）基础上进行的。本文给出该方法测试结果的一部分作为说明。

## 样本量的统计论证

在评估的不同阶段，我们确定了以下样本研究的数量。

1. 在自测试阶段，数据集的大小没有规定，取决于软件所解决的临床问题<sup>7</sup>。自测试、功能测试和校准测试阶段中使用的数据集是根据专家共识数据准备的。在某些情况下，数据集是根据组织学检查结果准备的（例如，用于评估恶性肿瘤）。《数据集准备规定》给出准备数据集过程的详情[19]。
2. 在功能测试阶段中，我们准备好了5次检查的数据集（根据GOST R 8.736-2011，多次测量被理解为至少四次测量）<sup>8</sup>。真实值被认为是专家医生的总结报告。专家医生是在该专业领域拥有5年以上经验的医生，在数据流上描述特定领域（特定模式和目标病理）的研究。专

家医生已接受过关于使用基于人工智能技术软件的指导。这一阶段至少有一名技术专家和一名专家医生参与。

3. 在校准测试阶段中，我们准备好了100次检查的数据集，具有50/50类平衡（50%的检查有目标病理，50%的检查无目标病理）<sup>9</sup>[20, 21]。这一阶段至少有一名技术专家和一名专家医生参与。
4. 在技术监测阶段中，我们考虑了报告期内软件对缺陷“a、b”（基于缺陷检测自动化）进行分析的所有检查，以及对缺陷“c-d”进行分析的80次检查样本<sup>10</sup>[20, 21]。这一阶段至少有一名技术人员参与。
5. 在临床监测阶段中，对上述80次检查样本进行了审查。真值被认为是专家医生的总结报告<sup>11</sup>[20, 21]。这一阶段有一名专家医生参与。

伦理审查本研究是在当地伦理委员会以前批准的研究（NCT04489992号）《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》（莫斯科实验）框架内进行的。

## 结果

根据文献综述的结果，我们找到了关于基于人工智能技术的医疗诊断软件评估的不同阶段的研究：软件的验证[1, 5, 8, 9]、监测[10]以及采用[7, 11-13]和监管[14, 15]。与此同时，我们没有发现统一的方法来测试和监测基于人工智能技术的医疗诊断软件。有一些关于基于人工智能技术的软件生命周期的研究[16]。不过，它们主要针对的不是医疗软件，而是其他活动领域的软件。这些文章没有考虑到基于人工智能技术的医疗诊断软件的特殊性。此外，也有关于基于人工智能技术软件的研究和科学出版物编写的指南，但不可能借助这些指南对软件进行测试和监测[17, 18]。需要强调的是，我们也没有发现关

<sup>6</sup> 俄罗斯联邦莫斯科政府于2019年11月21日第1543-PP号决议（<https://docs.cntd.ru/document/563879961>）；俄罗斯联邦政府于2020年11月24日第1906号决议（<http://publication.pravo.gov.ru/Document/View/0001202011270010>）；2011年11月21日第323-FZ号联邦法第38条（[https://www.consultant.ru/document/cons\\_doc\\_LAW\\_121895/](https://www.consultant.ru/document/cons_doc_LAW_121895/)）；俄罗斯联邦卫生部于2020年9月15日第980n号命令（<https://docs.cntd.ru/document/566006416>）；莫斯科卫生局于2023年2月16日第134号命令（[https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)）。

<sup>7</sup> 诊断和远程医疗中心。官方网站。数据集。访问方式：<https://mosmed.ai/datasets/>。

<sup>8</sup> GOST R 8.736-2011。俄罗斯联邦国家标准。国家保证测量统一系统。直接多次测量。测量结果处理方法。基本规定。访问方式：<https://docs.cntd.ru/document/1200089016>。

<sup>9</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式：[https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)。

<sup>10</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式：[https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)。

<sup>11</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式：[https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)。

于测试和监测后软件修订的出版物。同时，软件修订对于提高软件质量和在实用保健中成功采用来说至关重要。

为此，作者开发了一套基于人工智能技术的医疗诊断软件测试和监测，旨在提高该软件的质量，并将其应用于在实用保健中。该方法包括7个阶段。如图1所示。每个阶段的目的、本质和结果如下。

## 自测试

自测试阶段旨在评估软件与输入数据的技术兼容性。软件开发商（或供应商）可访问开放数据集。该数据集包含Dicom (Digital Imaging and Communications in Medicine) 文件。这些文件是诊断检查的匿名示例<sup>12</sup>。数据集提供以下参数结构：模式、诊断程序类型、诊断设备的制造商和型号[19]。

软件与数据的兼容性有助于在技术上将软件集成到医疗机构的放射信息网络中，并从功能测试阶段开始进行进一步评估<sup>13</sup>。

## 功能测试

在功能测试阶段中，要对供应商申报的软件功能的可用性和可操作性进行验证。测试从技术和临床角度进行。从技术角度来看，软件的评估标准如下：检查的优先级排序（分诊）；软件是否提供额外系列的图像；是否提供额外系列的名称；是否在额外系列的图像上提供软件的图示

符号；是否在图像上和DICOM SR中标注“仅供研究使用”的警告指示；是否存在系列同步的可能性；是否显示病理存在的可能性；是否指出病理类别；是否提供DICOM SR协议的完整结构（图2、3）。

这一部分功能测试由受过技术教育的专家根据基本功能要求进行。在俄罗斯联邦，这些要求由莫斯科市国家预算卫生机构“莫斯科卫生局诊断和远程医疗技术科学实践临床中心”制定<sup>14</sup>。从医学角度来看，对软件功能的评估由专家医生根据莫斯科卫生局诊断和远程医疗技术科学实践临床中心制定的基本诊断要求进行<sup>15</sup>。基本诊断要求包括软件答复的必选和可选内容、提交答复的格式和形式等标准。基本功能和基本诊断要求既有对所有软件的共同要求，也有具体要求。具体要求由软件所解决的临床问题决定。

如果发现严重不符合项，软件测试就会停止，直到供应商消除其原因为止。不符合基本功能要求的不符合项被认为是严重的，因为它们会对医生的工作过程产生负面影响，并直接或间接地影响病人的生命和健康（图4、5）。

在供应商公司消除了严重不符合项的原因后，应重新进行功能测试。申请方应有机会重新进行该阶段的测试，但不得超过2次。在收到测试结果不合格的协议后，申请方可在不受限制的时间内进行第一次重测。第二次重测不得早于收到上次测试结果不合格的协议后3个月。如果第二次重测不成功，可向申请方提供另一个科学实践合

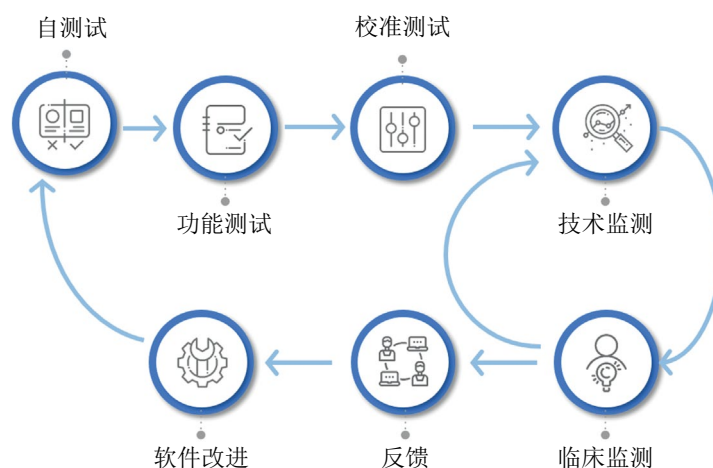


图1. 关于测试和监测基于人工智能技术的医疗诊断软件的方法。

<sup>12</sup> 诊断和远程医疗中心。官方网站。数据集。访问方式：<https://mosmed.ai/datasets/>。

<sup>13</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式：[https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)。

<sup>14</sup> 人工智能服务结果的基本功能要求。访问方式：[https://mosmed.ai/documents/218/Базовые\\_функциональные\\_требования\\_29.11.2022.pdf](https://mosmed.ai/documents/218/Базовые_функциональные_требования_29.11.2022.pdf)。

<sup>15</sup> 人工智能服务结果的基本诊断要求。访问方式：[https://mosmed.ai/documents/226/Базовые\\_диагностические\\_требования\\_22\\_02\\_2023.pdf](https://mosmed.ai/documents/226/Базовые_диагностические_требования_22_02_2023.pdf)。

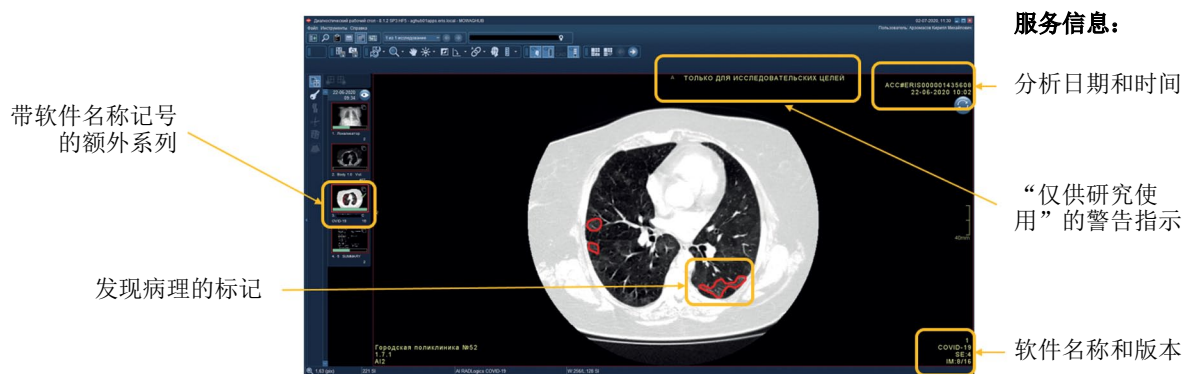


图2. 有图像的基于人工智能技术软件成果的主要组成部分：基准作品示例。

<p>DICOM SR1——关于检查和服务的服务信息</p>		资料编号 ERIS 00001990912 检查日期 12.08.2020 0:47 报告日期 12.08.2020 1:08 状态 服务名称 COVID-19 警告 仅供研究使用 软件版本 1.8.0 分析日期和时间 2020-09-12 1:08	
<p>DICOM SR2——关于服务及其目的的信息</p>		服务目的 模式: CT 解剖区域: 胸部 患者: 成人 目的: 寻找肺部COVID-19肺炎迹象并评估其病变的严重程度 简明用户手册: 病变用红色轮廓标出	
<p>DICOM SR4——关于检查结果的人工智能服务报告</p>		结论 人工智能处理结果 COVID-19: 在左上叶, 0%的肺实质受累 (0分) 在左下叶, 16.1%的肺实质受累 (2分) 在右上叶, 0%的肺实质受累 (0分) 在右中叶, 0%的肺实质受累 (0分) 在右下叶, 0%的肺实质受累 (0分) 双肺总计: 8%。严重程度——CT1	
<p>DICOM SR5——发现病理的详细说明</p>		COVID-19迹象 发现到病变的切片数: 114 肺切片总数: 269 阳性切片数占总切片数的百分比: 42%	
		<p>DICOM SR3——关于如何使用服务的信息（用户手册）</p>	

图3. 有DICOM SR的基于人工智能技术软件成果的主要组成部分：基准作品示例。

作项目<sup>16</sup>。如果没有发现严重不符合项，软件就进入校准测试阶段<sup>17</sup>。

### 校准测试

在校准测试阶段中，要确定关于软件诊断准确度的指标。主要指标是ROC曲线下面积（Area under the ROC Curve, AUC）。当通过使用尤登指数和最大化阴性和阳性预测值对ROC曲线进行分析时，可以确定最佳激活阈值。其他确定

的指数包括灵敏度、特异度、准确度、阳性预测值、阴性预测值。还确定每个检查的最短分析时间、平均分析时间和最长分析时间。真阳性、假阴性、假阳性、真阴性的数量以四栏表的形式呈现。某些指标的阈值如下：AUC——不低于0.81或0.91（取决于临床任务）；接受、处理检查和传输分析结果的所需时间——不超过6.5分钟；成功处理检查的比重——不低于90%<sup>18</sup>[21]。

<sup>16</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式：[https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)。

<sup>17</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式：[https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)。

<sup>18</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式：[https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)。

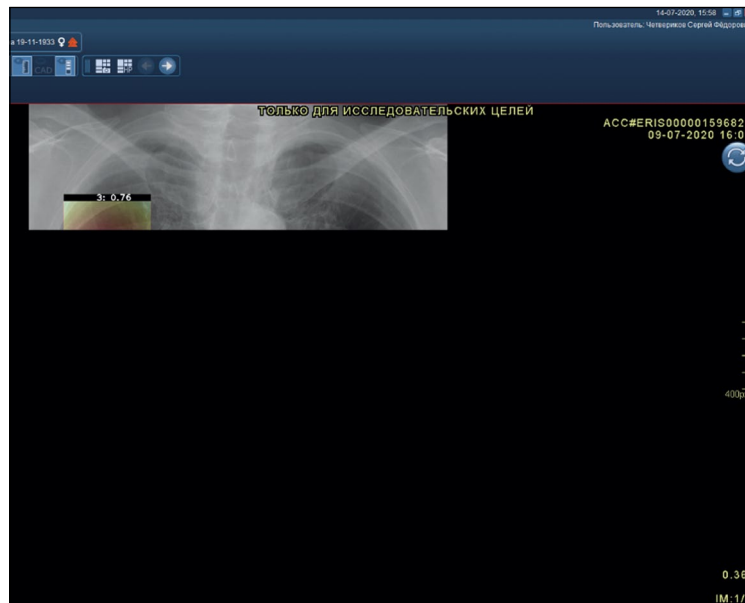


图4。基于人工智能技术的软件另一系列的截图：不符合基本功能要求的严重不符合项。

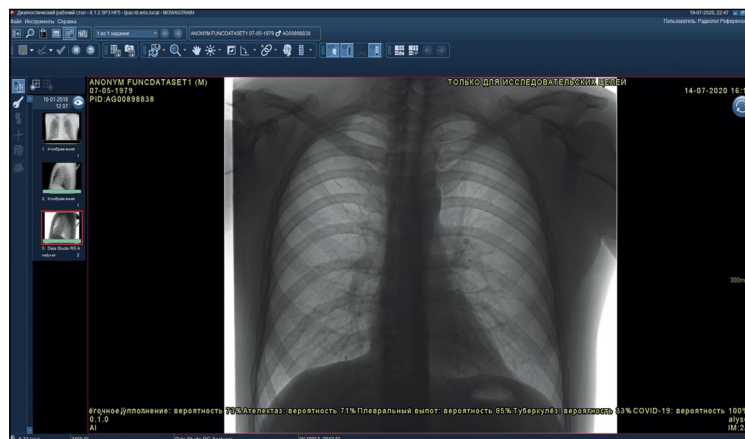


图5。图片说明叠置：不符合基本功能要求的严重不符合项。

校准测试的结果是其协议书（图6）。协议书可能包含严重和非严重不符合项。不符合上述阈值和方法建议的重大偏差的不符合项被认为是严重的[21]。在发现这些不符合项的情况下，软件测试就会停止，直到消除为止。如果没有不符合项，软件就有机会在定期监测阶段中开始对检查的前瞻性分析。定期监测包括技术和临床监测<sup>19</sup>。

## 技术监测

技术监测是从技术角度定期检查软件性能结果的阶段。这一阶段对于及时缺陷检测、及时质量控制和防止软件在放射科医生的实践中不正常运行是必要的。在这一阶段可以检测到的缺陷分为以下几类：

- 处理一份检查报告的时间超过6.5分钟；
- 缺乏分析检查的结果；
- 软件声明功能的不正确操作，妨碍放射科医生的工作或使其无法高质量地完成工作；
- 与图像区域显示有关的缺陷；
- 其他检查结果档案完整性和内容的失调，从而限制对检查结果的诊断解释。

“a、b”缺陷的自动监测是对报告期内软件分析的所有检查进行的。“c-d”缺陷的半自动监测是对80个检查样本进行的。为了对缺陷进行正确评估，我们开发了软件运行监测内部报告表格，其中包含技术缺陷监测指示（图7）。图8给出“胸部X射线照相术”技术缺陷平均数量的图表信息：缺陷数量呈下降趋势。

技术监测的结果就是技术监测报告（图9）。

<sup>19</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式：[https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)。



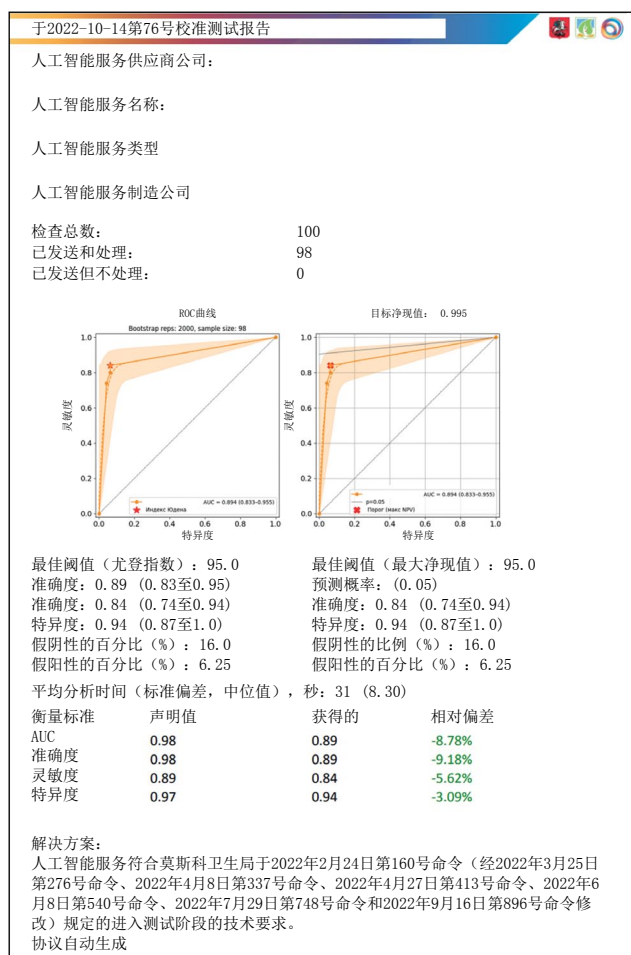


图6. 校准测试协议书的示例。

如果检测到的缺陷比重超过10%，则停止该软件的检查路由，直到消除缺陷原因为止。如果检测到的缺陷比重不超过10%，则软件运行和其定期监测继续进行<sup>20</sup>。

### 临床监测

在定期监测期间，放射科医生也会对软件结果进行临床评估。两个主要评估标准是病理的解释（结论）和定位（标记）。在评估过程中，医生可选择以下响应选项：完全一致、不正确评估、假阳性和假阴性。例如，如果医生完全同意软件的报告，则选择“解释：完全一致”。如果医生部分同意软件的报告，则选择“解释：不正确评估”。例如，医生同意存在病理，但不同意其细节。反之，医生同意细节，但不同意关于病理机率或其严重程度的总体结论。当医生完全不同意软件的报告时，就会选择“解释：假阳性”和“解释：假阴性”（图10）。

临床评估结果将输入上述的监测内部报告表格。然后我们将其上传到监测软件模块，以软件模块自动生成监测报告。

根据定期监测的结果，做出一种结论，其中“软件继续参与实验”、“实验参与者需要对软件运行进行修改”、“停止软件参与实验，直到对软件运行进行修改为止”<sup>21</sup>。

### 反馈

为了评估软件的实际意义，放射科医生的反馈阶段是必要的。反馈表位于程序窗口的放射科医



图7. 基于人工智能技术的软件运行监测内部报告表格。

<sup>20</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式：[https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)

<sup>21</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式：[https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)

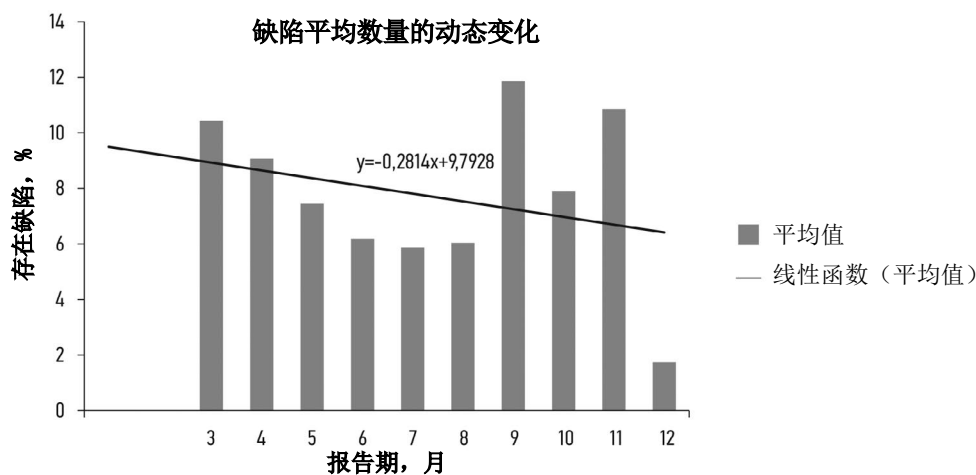


图8. “胸部X射线照相术”模式软件的技术缺陷动态变化。

服务运行技术参数监测报告	
1. 人工智能服务供应方公司: 人工智能服务名称: 人工智能服务制造公司: 实验中的人工智能服务标识符:	
2. 报告期:	
3. 检查类型:	
4. 临床目标:	
5. 检查总数:	
5.1. 根据URIS EMIAS的送交服务分析的报告期检查	16108个
5.2. 其中根据URIS EMIAS的独一无二检查	16108个*
6. 通过监测的检查次数	
6.1. 通过人工控制的检查	20个
7. 有缺陷的检查数量:	
7.1. 含有技术缺陷“a”, 附录1	32个
7.2. 含有技术缺陷“b”, 附录2	808个
7.3. 含有技术缺陷“c”-“f”, 附录3	1个
8. 检查比重:	
8.1. 相对于15300次检查, 含有技术缺陷“a”	0%
8.2. 相对于16108次检查, 含有技术缺陷“b”	5%
8.3. 相对于20次检查, 含有技术缺陷“c”-“f”	5%
9. 没有缺陷的检查数量	
10. 解决方案:	15267个
<b>人工智能服务继续参与实验</b>	
11. 备注:	
报告发布日期:	报告自动生成
负责人姓名:	报告自动生成
*根据URIS EMIAS, 有0个非唯一检查	

图9. 技术监测报告的示例。

生自动工作站中(图11)。放射科医生可以同意或不同意软件的报告。如果不同意, 医生要选择不同意的原因。主要原因是技术缺陷和诊断不准确。医生对软件分析的所有检查中的5%提供反馈就足够了。此外, 我们还通过对医生的问卷调查收集反馈。这样就可以评估他们对软运行作的满意度<sup>22</sup>。

## 改进

如果在功能测试、校准测试和定期监测阶段发现对软件运行的重要批评, 则停止软件测试, 直到消除备注的原因为止。改进由供应商方负责, 对医疗机构来说是一个“黑盒子”。如果有必要进行改进, 如果改进不涉及最初声明的功能、技术架构的变更, 也不涉及软件诊断准确度指标的变更, 则申请方在改进后可立即进入该方法的下一阶段。

如果申请方的改进涉及最初声明的功能、技术架构的变更, 也涉及软件诊断准确度指标的变更, 则无论之前软件方法处于哪个阶段, 都要重复进行功能测试和校准测试<sup>23</sup>。

## 讨论

本文介绍一种基于人工智能技术的医疗诊断软件测试和监测方法, 旨在提高其质量, 并将其应

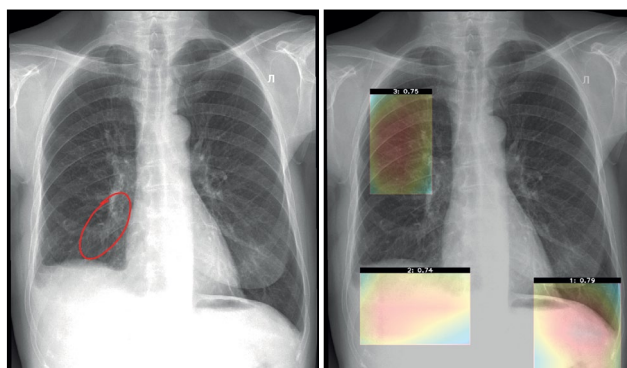


图10. 假阴性(没有右肺下叶亚段膨胀不全的定位): 不符合基本诊断要求的非严重不符合项。

<sup>22</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式: [https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)。

<sup>23</sup> 莫斯科卫生局于2023年2月16日第134号命令《关于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程序和条件批准》。访问方式: [https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf)。

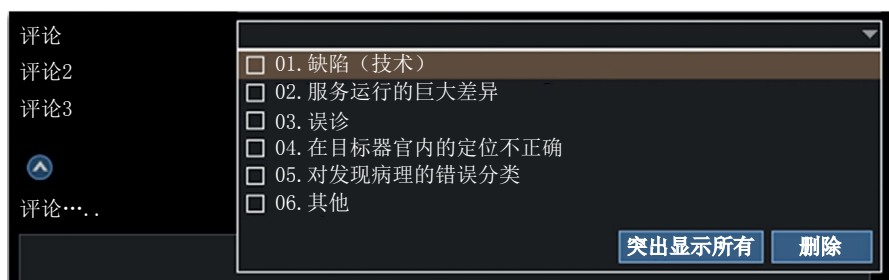


图11. 用户界面反馈窗口的内容。

用于实用保健中。开发该方法的原因包括：1) 现有的规范性文件中没有对基于人工智能技术的医疗诊断软件的测试和监测提出具体要求，2) 在市场上现有的各种程序中，没有关于医疗机构选择软件的规范性原则。本方法既不违反法定要求，又考虑到基于人工智能技术的医疗诊断软件的特殊性。该方法包括7个独特而清晰的阶段。这些阶段以科学研究成果为依据[1-4, 19-21]，并得到法律文件<sup>24</sup>的支持。

该方法的一个重要特点是制定基本功能和基本诊断要求。这些要求用于功能测试阶段<sup>25</sup>。缺陷和要求的系统化是独一无二的（审查过的研究没有提供详细描述）。将不符合项分为严重和非严重两种似乎尤为重要。这种划分对软件开发人员和用户都很方便。在世界范围内，美国放射医师学会数据科学研究所的文件是众所周知的。这些文件描述了通过使用软件解决的临床任务、预期的输入和输出数据<sup>26</sup>。

该方法的另一个重要特点是必须根据本地数据对软件进行校准（校准测试阶段），随后根据真实数据流进行验证（定期监测阶段）。根据国外的系统综述[22]，只有6%基于人工智能技术的软件经过了外部验证阶段。验证可分为“广义”验证和“狭义”验证[8]。“狭义”验证的目的是评估产品的“正确性”，即产品的使用结果是否符合使用目的。它可以包括临床验证和可用性评估。“广义”验证包括“狭义”验证，也是质量控制的同义词，即确保软件是采用最佳实践和方法开发的。它可以包括算法分析、软件测试和以文件研究。在这种情况下，对软件的内部结构进行评估，软件成为一个“白盒子”[8]。

值得单独指出的是，该方法包括在发现严重不符合项后的软件改进阶段。正是软件改进有助于逐步减少技术缺陷数量和提高软件诊断准确度指

标。因此，该方法将使基于人工智能技术的医疗诊断软件开发人员在各个方面取得重大成果。该方法将使用户能够在具有独立质量检查的软件中做出明智而坚定的选择。最终这将促使软件在实用保健中得到采用，有助于降低放射科医生的劳动成本，并提高诊断检查的解释效率，即实现利用人工智能技术实现流程自动化的最初目标。

这种方法不能替代在既定程序框架内的医疗器械注册过程。不过，监管机构可以采用整个方法或其中的某些阶段和方法来评估基于人工智能技术软件的安全性和有效性。该方法也可作为制造商质量管理体系的一部分。该方法既可能被软件开发商使用，例如在起草注册后临床监测计划的过程中，该计划应作为医疗器械注册文件集的一部分；也可被医疗机构使用，为特定条件和目的选择最合适的软件[4]。该方法可无限期应用。它既符合欧亚经济委员会关于3年监测的要求，也符合FDA关于整个产品生命周期监测的建议。

具有基于人工智能技术的软件注册证书，并不能证明就没有必要完成所介绍方法的全部或部分阶段。至少有两个原因需要完成所有阶段。首先，注册证书可能是在某些诊断设备上测试该软件时获得的，而该软件在其他设备上运行时，结果可能会发生变化。其次，注册证书可能是为解决某个临床问题而获得的，今后开发人员可能会增加软件的功能。

本文所述的例子来自放射学领域。不过，这种方法也适合临床医学其他领域使用的基于人工智能技术的软件。在这种情况下，需要对个别表格进行调整：技术缺陷清单、临床评估等。

## 该研究的局限性

该方法的局限性在于将制造公司与执行评估的机构分开。在许多方法中，软件开发和评估由一

<sup>24</sup> 莫斯科政府于2019年11月21日第1543-PP号决议 (<https://docs.cntd.ru/document/563879961>)；莫斯科卫生局于2023年2月16日第134号命令 ([https://mosmed.ai/documents/227/приказ\\_ДЗМ\\_134\\_от\\_16.02.2023.pdf](https://mosmed.ai/documents/227/приказ_ДЗМ_134_от_16.02.2023.pdf))。

<sup>25</sup> 人工智能服务结果的基本功能要求 ([https://mosmed.ai/documents/218/Базовые\\_функциональные\\_требования\\_29.11.2022.pdf](https://mosmed.ai/documents/218/Базовые_функциональные_требования_29.11.2022.pdf))；人工智能服务结果的基本诊断要求 ([https://mosmed.ai/documents/226/Базовые\\_диагностические\\_требования\\_22\\_02\\_2023.pdf](https://mosmed.ai/documents/226/Базовые_диагностические_требования_22_02_2023.pdf))。

<sup>26</sup> ACR Data Science Institute Releases Landmark Artificial Intelligence Use Cases. 2018. 访问方式：<https://www.acr.org/Media-Center/ACR-News-Releases/2018/ACR-Data-Science-Institute-Releases-Landmark-Artificial-Intelligence-Use-Cases>。

家公司完成（“从概念到实施”的方法）[16]。而在本方法中，评估是接近采用阶段时由第三方机构进行的。尽管如此，开发人员在开发早期阶段中的错误仍有可能被发现，但开发人员可能更难纠正这些错误。

在定期监测阶段中，软件会分析大量检查（超过1000个）。由于资源、专家医生人数和工作时间有限，无法控制所有检查的质量。虽然软件自动生成有代表性的伪随机检查样本，但在定期监测阶段中，仍有可能因系统性抽样误差而出现遗漏。

## 研究展望

1. 公布关于使用介绍的方法进行软件评估的结果（假设：根据介绍的方法进行软件评估，可提高基于人工智能技术的医疗诊断软件的诊断准确度和实用性指标）。
2. 使用介绍的方法对获得和未获得俄罗斯联邦卫生监督局注册证书的软件进行比较。
3. 在方法中创造另一个测试阶段，该阶段的目的是评估关于软件处理“不满意”检查的结果（不适合该软件的解剖区域、模式、伪影、不正确的病人定位、植入物和其他医疗设备）。

## 结论

我们开发了一套基于人工智能技术的医疗诊断软件测试和监测方法，旨在提高该软件的质量，并将其应用于实用保健中。该方法包括7个阶段：自测试、功能测试、校准测试、技术监测、临床监测、反馈和改进。该方法学的显著特点是对软件进行周期性的监测和改进，从而不断提高其质量；对软件性能结果并医生参与软件评估提

出详细要求。该方法可使软件开发人员在各个领域取得优异成绩并展示成就，也可使用户在通过独立、全面质量控制的程序中做出明智、自信的选择。

## ADDITIONAL INFORMATION

**Funding source.** This article was prepared by a group of authors as a part of the research and development effort titled “Development of a platform for improving the quality of AI services for clinical diagnostics”, No. 123031400006-0 in accordance with the Order No. 1196 dated December 21, 2022 “On approval of state assignments funded by means of allocations from the budget of the city of Moscow to the state budgetary (autonomous) institutions subordinate to the Moscow Health Care Department, for 2023 and the planned period of 2024 and 2025” issued by the Moscow Health Care Department.

**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. Vasiliev — development of the concept, approval of the final version of the manuscript; A.V. Vladzimirsky — development of the concept, approval of the final version of the manuscript; O.V. Omelyanskaya — development of methodology, approval of the final version of the manuscript; K.M. Arzamasov — concept development, research, editing and approval of the final version of the manuscript; S.F. Chetverikov — development of methodology, research; D.A. Rummyantsev — literature review, writing and editing the text of the article; M.A. Zelenova — editing the text of the article.

## СПИСОК ЛИТЕРАТУРЫ

1. Oakden-Rayner L., Palme L.J. Artificial intelligence in medicine: Validation and study design. In: Ranschart E., Morozov S., Algra P., eds. Artificial intelligence in medical imaging. Cham: Springer, 2019. P. 83–104.
2. Морозов С.П., Зинченко В.В., Хоружая А.Н., и др. Стандартизация искусственного интеллекта в здравоохранении: Россия выходит в лидеры // Врач и информационные технологии. 2021. № 2. С. 12–19. doi: 10.25881/18110193\_2021\_2\_12
3. 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. № 1. С. 9–17. doi: 10.22328/2079-5343-2020-11-1-9-17
4. 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
5. Tanguay W., Acar P., Fine B., et al. Assessment of radiology artificial intelligence software: A validation and evaluation framework // Can Assoc Radiol J. 2023. Vol. 74, N 2. P. 326–333. doi: 10.1177/08465371221135760
6. Kohli A., Jha S. Why CAD failed in mammography // J Am Coll Radiol. 2018. Vol. 15, N 3, Pt. B. P. 535–537. doi: 10.1016/j.jacr.2017.12.029
7. Recht M.P., Dewey M., Dreyer K., et al. Integrating artificial intelligence into the clinical practice of radiology: Challenges and recommendations // Eur Radiol. 2020. Vol. 30, N 6. P. 3576–3584. doi: 10.1007/s00330-020-06672-5
8. Higgins D.C., Johner C. Validation of artificial intelligence containing products across the regulated healthcare industries // Ther Innov Regul Sci. 2023. Vol. 57, N 4. P. 797–809. doi: 10.1007/s43441-023-00530-4
9. Rudolph J., Schachtner B., Fink N., et al. Clinically focused multi-cohort benchmarking as a tool for external validation of artificial intelligence algorithm performance in basic chest radiography analysis // Sci Rep. 2022. Vol. 12, N 1. P. 12764. doi: 10.1038/s41598-022-16514-7
10. Allen B., Dreyer K., Stibolt R., et al. Evaluation and real-world performance monitoring of artificial intelligence models in clinical

practice: Try it, buy it, check it // *J Am Coll Radiol*. 2021. Vol. 18, N 11. P. 1489–1496. doi: 10.1016/j.jacr.2021.08.022

11. Strohm L., Hehakaya C., Ranschaert E.R., et al. Implementation of artificial intelligence (AI) applications in radiology: Hindering and facilitating factors // *Eur Radiol*. 2020. Vol. 30, N 10. P. 5525–5532. doi: 10.1007/s00330-020-06946-y

12. Sohn J.H., Chillakuru Y.R., Lee S., et al. An open-source, vendor agnostic hardware and software pipeline for integration of artificial intelligence in radiology workflow // *J Digit Imaging*. 2020. Vol. 33, N 4. P. 1041–1046. doi: 10.1007/s10278-020-00348-8

13. Wichmann J.L., Willeminck M.J., De Cecco C.N. Artificial intelligence and machine learning in radiology: Current state and considerations for routine clinical implementation // *Invest Radiol*. 2020. Vol. 55, N 9. P. 619–627. doi: 10.1097/RLI.0000000000000673

14. 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 // *J Am Coll Radiol*. 2021. Vol. 18, N 3, Pt. A. P. 413–424. doi: 10.1016/j.jacr.2020.09.060

15. Milam M.E., Koo C.W. The current status and future of FDA-approved artificial intelligence tools in chest radiology in the United States // *Clin Radiol*. 2023. Vol. 78, N 2. P. 115–122. doi: 10.1016/j.crad.2022.08.135

16. De Silva D., Alahakoon D. An artificial intelligence life cycle: From conception to production // *Patterns (NY)*. 2022. Vol. 3, N 6. P. 100489. doi: 10.1016/j.patter.2022.100489

17. Cerdá-Alberich L., Solana J., Mallol P., et al. MAIC-10 brief quality checklist for publications using artificial intelligence and

medical images // *Insights Imaging*. 2023. Vol. 14, N 1. P. 11. doi: 10.1186/s13244-022-01355-9

18. Vasey B., Novak A., Ather S., et al. DECIDE-AI: A new reporting guideline and its relevance to artificial intelligence studies in radiology // *Clin Radiol*. 2023. Vol. 78, N 2. P. 130–136. doi: 10.1016/j.crad.2022.09.131

19. Регламент подготовки наборов данных с описанием подходов к формированию репрезентативной выборки данных. Москва: Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, 2022. 40 с. (Лучшие практики лучевой и инструментальной диагностики; Часть 1).

20. 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. 19–27. doi: 10.17691/stm2023.15.2.02

21. Морозов С.П., Владимировский А.В., Кляшторный В.Г., и др. Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика). Москва: Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, 2019. 33 с.

22. Kim D.W., Jang H.Y., Kim K.W., et al. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers // *Korean J Radiol*. 2019. Vol. 20, N 3. P. 405–410. doi: 10.3348/kjr.2019.0025

## REFERENCES

1. Oakden-Rayner L, Palme LJ. Artificial intelligence in medicine: Validation and study design. In: Ranschaert E, Morozov S, Algra P, eds. *Artificial intelligence in medical imaging*. Cham: Springer; 2019. P. 83–104.

2. Morozov SP, Zinchenko VV, Khoruzhaya AN, et al. Standardization of artificial intelligence in healthcare: Russia is becoming a leader. *Doctor Inform Technol*. 2021;(2):12–19. (In Russ). doi: 10.25881/18110193\_2021\_2\_12

3. Mello AA, Utkin LV, Trofimova TN. Artificial intelligence in medicine: The current state and main directions of development of intellectual diagnostics. *Radiation Diagnost Therap*. 2020;(1):9–17. (In Russ). doi: 10.22328/2079-5343-2020-11-1-9-17

4. Zinchenko VV, Arзамасов KM, Четвериков SF, et al. Methodology of post-registration clinical monitoring for software using artificial intelligence technologies. *Modern Technol Med*. 2022;14(5):15–25. (In Russ). doi: 10.17691/stm2022.14.5.02

5. Tanguay W, Acar P, Fine B, et al. Assessment of radiology artificial intelligence software: A validation and evaluation framework. *Can Assoc Radiol J*. 2023;74(2):326–333. doi: 10.1177/08465371221135760

6. Kohli A, Jha S. Why CAD failed in mammography. *J Am Coll Radiol*. 2018;15(3 Pt B):535–537. doi: 10.1016/j.jacr.2017.12.029

7. Recht MP, Dewey M, Dreyer K, et al. Integrating artificial intelligence into the clinical practice of radiology: Challenges and recommendations. *Eur Radiol*. 2020;30(6):3576–3584. doi: 10.1007/s00330-020-06672-5

8. Higgins DC, Johner C. Validation of artificial intelligence containing products across the regulated healthcare industries. *Ther Innov Regul Sci*. 2023;57(4):797–809. doi: 10.1007/s43441-023-00530-4

9. Rudolph J, Schachtner B, Fink N, et al. Clinically focused multi-cohort benchmarking as a tool for external validation of artificial intelligence algorithm performance in basic chest radiography analysis. *Sci Rep*. 2022;12(1):12764. doi: 10.1038/s41598-022-16514-7

10. Allen B, Dreyer K, Stibolt R, et al. Evaluation and real-world performance monitoring of artificial intelligence models in clinical practice: Try it, buy it, check it. *J Am Coll Radiol*. 2021;18(11):1489–1496. doi: 10.1016/j.jacr.2021.08.022

11. Strohm L, Hehakaya C, Ranschaert ER, et al. Implementation of artificial intelligence (AI) applications in radiology: Hindering and facilitating factors. *Eur Radiol*. 2020;30(10):5525–5532. doi: 10.1007/s00330-020-06946-y

12. Sohn JH, Chillakuru YR, Lee S, et al. An open-source, vendor agnostic hardware and software pipeline for integration of artificial intelligence in radiology workflow. *J Digit Imaging*. 2020;33(4):1041–1046. doi: 10.1007/s10278-020-00348-8

13. Wichmann JL, Willeminck MJ, De Cecco CN. Artificial intelligence and machine learning in radiology: Current state and considerations for routine clinical implementation. *Invest Radiol*. 2020;55(9):619–627. doi: 10.1097/RLI.0000000000000673

14. Larson DB, Harvey H, Rubin DL, et al. Regulatory frameworks for development and evaluation of artificial

intelligence-based diagnostic imaging algorithms: Summary and recommendations. *J Am Coll Radiol.* 2021;18(3 Pt A):413–424. doi: 10.1016/j.jacr.2020.09.060

15. Milam ME, Koo CW. The current status and future of FDA-approved artificial intelligence tools in chest radiology in the United States. *Clin Radiol.* 2023;78(2):115–122. doi: 10.1016/j.crad.2022.08.135

16. De Silva D, Alahakoon D. An artificial intelligence life cycle: From conception to production. *Patterns (NY).* 2022;3(6):100489. doi: 10.1016/j.patter.2022.100489

17. Cerdá-Alberich L, Solana J, Mallol P, et al. MAIC-10 brief quality checklist for publications using artificial intelligence and medical images. *Insights Imaging.* 2023;14(1):11. doi: 10.1186/s13244-022-01355-9

18. Vasey B, Novak A, Ather S, et al. DECIDE-AI: A new reporting guideline and its relevance to artificial intelligence studies in radiology. *Clin Radiol.* 2023;78(2):130–136. doi: 10.1016/j.crad.2022.09.131

19. Regulations for the preparation of data sets with a description of approaches to the formation of a representative sample of data.

Moscow: Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Department of Health of the City of Moscow; 2022. 40 p. (Best practices in radiological and instrumental diagnostics; Part 1). (In Russ).

20. Chetverikov S, Arzamasov KM, Andreichenko AE, et al. Approaches to sampling for quality control of artificial intelligence systems in biomedical research. *Modern Technol Med.* 2023;15(2):19–27. (In Russ). doi: 10.17691/stm2023.15.2.02

21. Morozov SP, Vladzimirsky AV, Klyashtorny VG, et al. Clinical trials of software based on intelligent technologies (radiation diagnostics). Moscow: Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Department of Health of the City of Moscow; 2019. 33 p. (In Russ).

22. Kim DW, Jang HY, Kim KW, et al. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers. *Korean J Radiol.* 2019;20(3):405–410. doi: 10.3348/kjr.2019.0025

## AUTHORS' INFO

### \* Denis A. Rumyantsev;

address: 24/1 Petrovka street, 127051 Moscow, Russia;  
ORCID: 0000-0001-7670-7385;  
eLibrary SPIN: 8734-2085;  
e-mail: RumyantsevDA3@zdrav.mos.ru

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

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

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

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

### Olga V. Omelyanskaya;

ORCID: 0000-0002-0245-4431;  
eLibrary SPIN: 8948-6152;  
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

### Sergei F. Chetverikov, Cand. Sci. (Engin.);

ORCID: 0000-0002-3097-8881;  
eLibrary SPIN: 3815-8870;  
e-mail: ChetverikovSF@zdrav.mos.ru

### Maria A. Zelenova;

ORCID: 0000-0001-7458-5396;  
eLibrary SPIN: 3823-6872;  
e-mail: ZelenovaMA@zdrav.mos.ru

## ОБ АВТОРАХ

### \* Румянцев Денис Андреевич;

адрес: Россия, 127051, Москва, ул. Петровка, д. 24, стр. 1;  
ORCID: 0000-0001-7670-7385;  
eLibrary SPIN: 8734-2085;  
e-mail: RumyantsevDA3@zdrav.mos.ru

### Васильев Юрий Александрович, канд. мед. наук;

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

### Владзими́рский Анто́н Вячеславо́вич, д-р мед. наук;

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

### Омелянская Ольга Васильевна;

ORCID: 0000-0002-0245-4431;  
eLibrary SPIN: 8948-6152;  
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

### Четвериков Сергей Федорович, канд. тех. наук;

ORCID: 0000-0002-3097-8881;  
eLibrary SPIN: 3815-8870;  
e-mail: ChetverikovSF@zdrav.mos.ru

### Зеленова Мария Александровна;

ORCID: 0000-0001-7458-5396;  
eLibrary SPIN: 3823-6872;  
e-mail: ZelenovaMA@zdrav.mos.ru

\* Corresponding author / Автор, ответственный за переписку