

DOI: <https://doi.org/10.17816/DD60635>



标准医疗日期(MosMedData)独立外部评价的算法在诊断的人工智能基础上

© Nikolay A. Pavlov, Anna E. Andreychenko, Anton V. Vladzimirskyy, Anush A. Revazyan, Yury S. Kirpichev, Sergey P. Morozov

Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

简评:

这篇文章介绍了一个独特的方法来创建附加说明的医疗日期，以测试基于人工智能技术的诊断解决方案。描述了数据集形成的四个阶段-计划，初始数据选择，标记和验证，文档。所举的例子是根据上述日期方法建立的。该方法是广泛而普遍的，因此可以应用于医学和卫生的其他领域，它是由人工智能技术和高数据技术的自动化和发展。

关键字: 关键词: 人工智能 医疗数据; 数据集 标记; 机器学习; 大数据; 确认。

引用本文:

Pavlov NA, Andreychenko AE, Vladzimirskyy AV, Revazyan AA, Kirpichev YS, Morozov SP. 标准医疗日期 (MosMedData) 独立外部评价的算法在诊断的人工智能基础上. *Digital Diagnostics*. 2021;2(1):49-65. DOI: <https://doi.org/10.17816/DD60635>

收到: 11.02.2021

接受: 23.03.2021

发布时间: 30.03.2021



DOI: <https://doi.org/10.17816/DD60635>

Эталонные медицинские датасеты (MosMedData) для независимой внешней оценки алгоритмов на основе искусственного интеллекта в диагностике

© Н.А. Павлов, А.Е. Андрейченко, А.В. Владзимирский, А.А. Ревазян, Ю.С. Кирпичев, С.П. Морозов

Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения г. Москвы, Москва, Российская Федерация

АННОТАЦИЯ

В статье описывается оригинальный подход к формированию аннотированных медицинских датасетов для проверки диагностических решений, основанных на технологиях искусственного интеллекта. Описаны 4 этапа формирования датасета — планирование, отбор исходных данных, разметка и верификация, документирование. Приведены примеры созданных по описанной методике датасетов. Методика является масштабируемой и универсальной, а значит, может быть использована в других областях медицины и здравоохранения, которые подлежат автоматизации и развитию с помощью технологий искусственного интеллекта и технологий больших данных.

Ключевые слова: искусственный интеллект; медицинские данные; датасет; разметка; машинное обучение; большие данные; верификация.

Как цитировать

Павлов Н.А., Андрейченко А.Е., Владзимирский А.В., Ревазян А.А., Кирпичев Ю.С., Морозов С.П. Эталонные медицинские датасеты (MosMedData) для независимой внешней оценки алгоритмов на основе искусственного интеллекта в диагностике // *Digital Diagnostics*. 2021. Т. 2, №1. С. 49–65. DOI: <https://doi.org/10.17816/DD60635>

DOI: <https://doi.org/10.17816/DD60635>

Reference medical datasets (MosMedData) for independent external evaluation of algorithms based on artificial intelligence in diagnostics

© Nikolay A. Pavlov, Anna E. Andreychenko, Anton V. Vladzemyrskyy, Anush A. Revazyan, Yury S. Kirpichev, Sergey P. Morozov

Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

ABSTRACT

The article describes a novel approach to creating annotated medical datasets for testing artificial intelligence-based diagnostic solutions. Moreover, there are four stages of dataset formation described: planning, selection of initial data, marking and verification, and documentation. There are also examples of datasets created using the described methods. The technique is scalable and versatile, and it can be applied to other areas of medicine and healthcare that are being automated and developed using artificial intelligence and big data technologies.

Keywords: artificial intelligence; medical data; dataset; marking; computer-assisted learning; big data; verification.

To cite this article

Pavlov NA, Andreychenko AE, Vladzemyrskyy AV, Revazyan AA, Kirpichev YS, Morozov SP. Reference medical datasets (MosMedData) for independent external evaluation of algorithms based on artificial intelligence in diagnostics. *Digital Diagnostics*. 2021;2(1):49–65. DOI: <https://doi.org/10.17816/DD60635>

Received: 11.02.2021

Accepted: 23.03.2021

Published: 30.03.2021

略语表

GB, MB, TB-数字存储量: 千兆字节, 兆字节, 太字节

数据集-Dataset (Data set) - 一个结构化的信息集, 结合一定的逻辑原理, 适合计算机处理的数据分析方法。数据集-具有四个主要阶段的复杂概念: 内容的存在 (观察, 值, 记录, 文件等); 目标的存在 (例如, 知识库, 用于特定任务的用途); 分组的存在 (将内容聚合和组织为集合, 集合等); 凝聚力的存在 (与主题的关系, 集成, 内容的逻辑集合等)

Unified medical information analysis system (UMIAS) - 莫斯科统一医疗信息和分析系统

URIS - 莫斯科统一放射信息服务

人工智能-创建能够执行任务的智能计算机程序的科学技术, 通常需要人类智能来解决这些问题

CT - CT扫描

K T 0-4 - 2020年莫斯科卫生部的医疗和远程医疗技术和科学实用临床诊断中心开发-COVID-19的CT征象分类。K T 0-标准和没有病毒性肺炎的CT征象。K T 1-磨砂玻璃类型的压实区域; 肺实质受累 $\leq 25\%$ 。K T 2-磨砂玻璃类型的压实区域; 肺实质受累 $25 - 50\%$ 。K T 3-磨砂玻璃类型的压实区域; 肺实质受累 $50 - 75\%$ 。K T 4 - 磨砂玻璃型肺

组织弥漫性压实, 固结区伴铺路石征。实质受累 $>75\%$

HIS - 医疗信息系统

Mammography - 乳腺钼靶

NDKT - 低剂量CT扫描

Thoracic organs - 胸部器官

X-ray - 射线照相

FLG - 胸透X光

COVID-19 - 2020年, 这种疾病的传播被世界卫生组织 (WHO) 视为大流行病-SARS-CoV-2病毒引起的传染病。根据第十版的《国际疾病分类》, 其编码为U07.1或U07.2分别取决于是否存在实验室鉴定的病毒

DICOM (Digital Imaging and Communications in Medicine) - 建立、储存、传送和显形数字化医疗图片及文件行业医疗标准

MeSH (Medical Subject Headings)

-包含关键医学术语的词库, 用于在医学和生物出版物的英文文本数据库中对文章进行索引, 分类和搜索, 就创建美国国家生物技术信息中心 (PubMed)。

README - **英语** “读我”, 带有可执行代码, 数据库或其他软件产品的文档的公认名称, 通常包含有关位于同一目录中的文件的基本信息。

SARS-CoV-2 - 单链壳层 (+) RNA病毒 **Beta-coronavirus**

论证:

人工智能 (AI) 技术的进步及其在各个领域 (特别是在医学领域) 的实际应用证明了技术在解决方案中的潜力, 这些解决方案包括自动诊断系统, 用于识别非结构化病历和理解自然语言的系统, 用于分析和预测的系统事件, 系统信息的自动分类和验证; 以及支持患者的自动聊天机器人 [1]。由于深机培训及相关的计算机图像识别和模式识别的迅速发展, 在所有领域都有相当多的注意力。自动诊断系统的应用侧重于对医疗图像的分析, 特别是对放射研究的分析 [2]。

对于实际的医疗保健而言, 自动化诊断过程是人口老龄化的首要任务之一, 这是提高可用性

的方法, 因此, 无法通过增加合格人员的数量来弥补诊断过程的数量, 就必需确保正确解释结果。因此, 及时提供医疗服务。这个问题在放射诊断中很严重 [3], 基于医生对图像的视觉分析。对于放射诊断中最现代的方法, 每个要解释的患者的二维图像数量可以达到1000个或更多。因此, 目前正是辐射诊断领域技术的积极发展; 而包括人工智能的概念。建立计算机视觉系统, 确保医疗图像的解释自动化。通过其他机器学习方法进行深度学习的一个显著特征是, 所创建模型的准确性, 可靠性和实用价值直接取决于在学习, 验证 (微调) 和测试过程中使用的数据的数量和质量 [4]。

这就是为什么在医学诊断中开发基于AI的解决方案的主要障碍之一是缺少经过验证的数据集（没有不完整和错误的信息）和高质量的数据集（统一的，为自动机器处理准备的数据集）[5]。

带注解数据集 [6]不仅是对AI的“训练”所必需的，特别是用于计算机神经网络的机器学习，还用于测试在其他数据上训练过的网络。

数据集要求不允许使用从医疗信息系统中进行的简单卸载，但是必须对数据进行大量操作，然后再将这些数据变成适合AI模型有效使用的带注释的数据集。医学数据与积极使用机器学习的其他领域（例如银行，服务）的数据之间的差异在于历史上建立的医学记录文化，结构的缺乏或结构的简化以及同一患者不同研究的比较有限彼此。目前关于医学数据集准备的文献仅以少量出版物为代表[7-9]。通过该出版物，作者旨在扩大与开发或测试AI的过程相关或涉及的医学专业人员之间基于医学数据的问题和基于医学数据准备数据集的功能的理解，以及程序员和数据处理专家，以改进基于I的算法的独立评价过程。

本文为放射诊断领域中使用AI技术的解决方案的客观（尽可能在每种特定情况下）测试提供了一种统一的方法（方法），以进行客观（尽可能在每种情况下）测试。描述我们提出的方法的各个阶段的过程中，我们给出了一些实际的数据集示例。我们在2019年9月至2020年12月期间使用存储在统一放射信息服务（ERIS EMIAS）中的莫斯科医疗门诊和住院机构的放射诊断部门的数据开发而成[10]。该条所述的基本原则可用于在其他医疗领域建立医疗日期。

方法与结果

数据集不同于医学数据的简单集合，因为它具有特殊的属性-数据统一和结构化；缺乏重大错误，错误的研究；是否存在其他信息（数据项的属性或特征类别和值）；俄罗斯联邦数据集等同于数据库，并由于智力活动而接受自愿国家注册。国外实践中数据集通常不仅以可下载的数据集形式发布，而且还以期刊中的科学出版物形式发布。每个数据集不仅在研究的组成方面而且在分类方式和标记方法上都是唯一的，并且创建数据集的过程具有探索性。即使在存在形成数据集的结构化方法的情况下，根据生成的数据集的目的，也可能会发生偏离，异常和某些阶段的变化。

整个过程可以分为四个主要阶段-规划，初始数据选择，标记和验证，文档编制（图1）。

规划阶段

像科学研究一样，数据集的准备工作始于计划阶段：

- 制定医学领域的临床和/或实践问题，并通过智能系统实现自动化；
- 汇编初始数据的特征和/或特征列表，计划在解决任务的过程中从智能系统接收有关该信息的信息，并有可能评估信息的正确性系统；
- 确定所生成数据集的选定特征和/或元素的特征值的验证方法；
- 资料来源的定义；
- 说明计划采取哪些步骤确保数据的安全；
- 确定列入标准，未将研究结果从日期中删除；

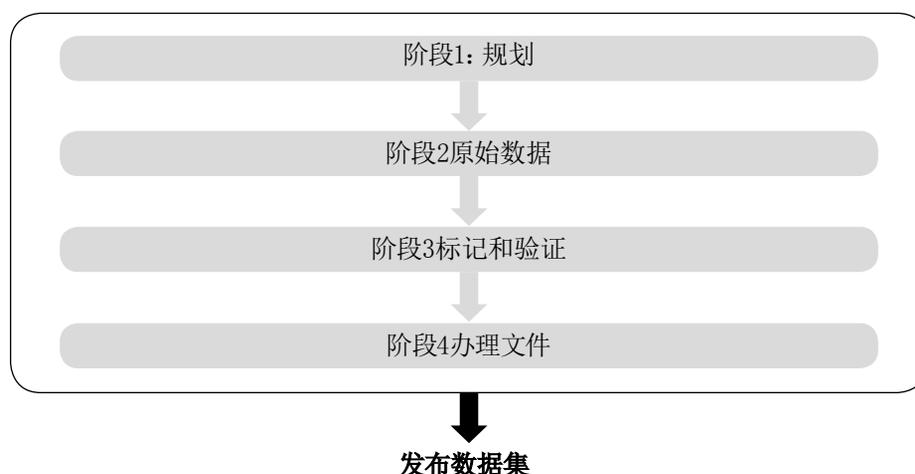


图 1形成医疗数据集的阶段。

- 确定所需数据的重要特征，不仅要评估准确性，而且要评估知识系统的可靠性和范围。

设置临床任务是数据集创建者面临的最重要任务之一。对它的关注不足导致在准备数据集期间以及在将基于AI的诊断算法引入临床实践期间已经突然出现的问题(图2)。

为了使任务与AI已确立为一项有前途技术的任务类别相对应，同时从临床专家的角度来看，AI具有重要的社会经济组成部分，这是一个由不同背景的专业人员组成的工作组，即临床医生，专家应参与定义任务，医学数据处理，研究工程师(机器学习或AI验证解决方案)以及访问和上传原始数据的管理员。

临床任务应允许数据集的创建者回答以下问题：

- 1) 应采用什么方式，什么程序，什么临床，人口统计学和类似信息作为算法求解的输入，以及什么应作为一个数据单元；
- 2) 使用AI技术应确定哪些功能；
- 3) 什么符号或符号组是理想的标志；
- 4) 解决问题的方法如何帮助临床专家；
- 5) 为了使用创建的数据集，需要多少数据单位(AI验证，机器学习等)。

选择数据单元数量和研究特征的重要标准是将数据集应用于AI的目的。您可以按照用途对数据集进行以下分类：

- 1) 常用集：
 - 自检以检查AI是否符合技术要求；
 - 临床测试以评估AI准确性和生产率

标；

- “附加培训”，用于调整已经受过培训的AI模型；
- 机器学习可训练成为AI基础的新模型并解决新的临床问题；

2) 专用套件：

- 用于评估随时间变化的动态集(将多个数据项绑定到一个主题)；
- 技术缺陷，以在尝试分析缺陷研究时评估基于AI的诊断解决方案的稳定性和可靠性。

对于每种类型或型号的诊断设备，自测所需的研究单位数量通常作为单个示例进行计算；用于临床测试的动态数据集和数据集中的研究单位数量通常为几十个；用于训练和“再训练”的数据集可以包含数百到数万的研究。这些数量是粗略的估计，可能在广泛的范围内有所不同，取决于数据来源的研究、临床任务的复杂性，数据说明和其他因素的详细性和工作量。定义临床任务后，智能系统将根据该标准从逻辑上决定是将特定研究还是将图像中找到的区域分配给感兴趣的组(AI工作的基本诊断要求)。诊断要求包括在研究中对所需特征的正式描述，以及一系列特征和/或特征，作为今后的基础在数据点进行数据标记。该信息允许开发商更精确地配置解决方案来确定所需的特征，日期准备专家就数据的标记和验证问题起草指示。

对于获得的使用数据集的基于人工智能技术的

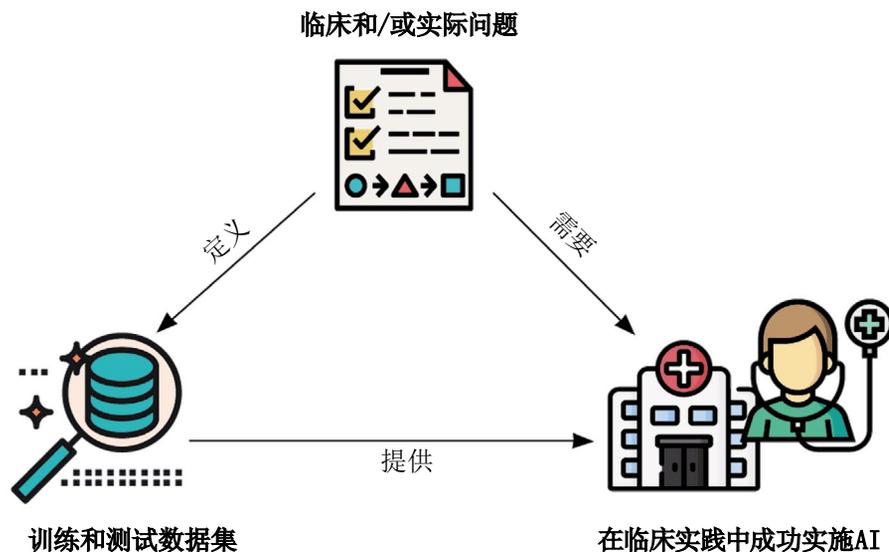


图 2 临床任务，数据集与常规临床实践中实施人工智能(AI)解决方案的成功之间的关系。



图 3莫斯科的数据集实验了根据此方法在计算机视觉领域中使用创新技术来分析医学图像并进一步在莫斯科的医疗保健系统中使用。

系统分析的价值和意义而言，至关重要是类别的平衡，即与各种特征和/或特征相关的数据集的研究分布的比例。最简单的情况下，要评估提供二分反应的智能诊断系统的性能，请在两个类别之间进行均等的划分（例如，根据AI工作的基本诊断要求，有50%的具有病理体征的研究和50%的无病理迹象的研究）。更复杂的情况下几个类之间的划分可能不均匀，并且取决于您计划以后使用的比较方法。

根据显着性状分为几类的研究可能还有其他差异-在临床方面（例如，由于发病年龄和性别模式，具有病理迹象的女性患者的患病率）以及技术方面（例如，是由于倾向于在具有较高分辨率的设备上进行的研究中倾向于直接指导已经确定病理的患者，因此存在人工抽样偏倚。为了避免有系统的错误，必须查明哪些迹象虽然不为临床任务所知，但却影响到诊断知识系统的工作，而在选择DATASET的研究时，力求在每一类中提供不同的例子。对于广泛的临床任务，将这些体征和特征系统化的问题（即，数据集中的类平衡问题）仍然是一个悬而未决的问题，目前正在积极研究中[9]。

计划阶段的最后，**确定源数据的来源，以及确定从数据集中包括，排除和排除研究的标准。**

为了创建最具代表性的数据集，数据源应尽可能与计划在未来引入AI解决方案的信息系统

相同或相关。对于莫斯科医疗保健来说，ERIS EMIAS是这种来源的一个例子，该系统将存储系统与莫斯科数十家门诊和住院医疗机构的放射诊断部门结合在一起。

列入和不列入标准往往是由临床和/或实际目标确定的，而排除标准通常是在处理原始数据时加以补充的，由于有些标准对数据集的结构和统一产生了不利影响。这些标准可以是医学上的（例如，年龄从18到99岁；是否存在目标器官的完整结构等）也可以是技术上的（例如，CT过滤器-软组织，卷积核芯-FC51等）数据统一必要用于AI解决方案性能评估工具的可靠性能（请看3点“标记和验证”）。

例如，给出了数据集“ MosMedData: 具有肺部聚焦的超低剂量计算机断层扫描的结果”¹。创建数据库的目的是确保检查自动系统（包括使用AI的系统）是否可以在ERIS UMIAS中工作的可能性。临床任务是寻找和确定肺癌筛查中的肺病灶。对于数据集，选择了DICOM格式的匿名计算机断层扫描（CT）研究，以特殊的超低剂量断层扫描模式进行（在增加的电压-135 kV下，辐射暴露的有效剂量小于1 mSv）。符合以下

1 Morozov S.P., Gonchar A.P., Nikolaev A.E. 等等。MosMedData: 超低剂量计算机断层扫描对肺部病变的研究结果（数据库）。2020年6月21日第2020622727号数据库的国家注册证书。

标准的一项胸部CT (CT) 扫描是每单位一个单位。

A. 选标准:

1. 患者的年龄在55岁以上且在75岁以下。
2. 吸烟经验超过30包/年 (30年每天至少1包, 或15年每天每天2包, 等等)。
3. 目前吸烟或不超过15年前戒烟。
4. 该研究是在第一轮肺癌筛查中以超低剂量CT扫描模式进行的。

B. 排除标准:

1. 第一轮超低剂量CT肺癌筛查后两年内发现肺癌。
2. 肺癌和/或肺部手术史 (不包括经皮肺活检)。
3. 癌症病史在不到5年前被确定, 不包括皮肤癌和宫颈癌 *in situ*。
4. 存在明显的心血管、免疫、呼吸系统和内分泌系统病理。此外, 预期寿命不到5年。
5. 呼吸系统急性疾病。
6. 过去的12周内进行了抗生素治疗。
7. 过去的一年中咯血或体重减轻超过10公斤。

C. 排除标准:

1. 第一轮莫斯科肺癌筛查中没有肺病灶。最终数据集中的研究数量目标值-300-足以测试基于AI的自动诊断系统 (总数为312个单位)。

2. 初始数据的选择阶段

访问初始数据的源之后选择初始 (“原始”) 数据的阶段开始。获取 (卸载) 它们的方法取决于存储数据的源和方法。

既可以在医疗机构的常规诊断过程中收集医疗数据 (MeSH: Routinely Collected Health Data), 也可以直接由患者和/或其亲戚和社工收集数据 (MeSH: Patient Generated Health Data), 以及有针对性的数据收集的结果 (例如在临床试验期间)。常规收集的数据通常具有广泛的参数可变性, 并允许您创建最具代表性的数据集。使用临床试验期间收集的数据时值得注意的 (1) 纳入标准, 根据研究的设计规定, 以及在准备数据集时限制研究的可能性, 不得纳入或排除研究对象, (2) 受研究能力限制的数据量。

数字化非电子文档几乎没有意义; 存储在外部介质上的文档通常结构不佳, 并且数字化和/或从其他介质传输数据的成本可能很高 (例如, 传输存储在CD-ROM上的放射线图像数据库)。医

疗信息系统 (MIS; MeSH: Health Information Systems) 的存在简化了卸载过程, 因为它使您可以应用过滤器并通过诸如某些特定研究或诊断的迹象来选择必要的研究。但是, 并非针对所有临床任务, 必要的信息都包含在电子病历中: 可以将符合临床任务标准的患者列表与MIS分开生成。这些列表中选择患者的研究需要花费大量时间。

选择 “原始” 数据的一般原则是:

- 1) 选择尽可能多的研究模式和程序;
- 2) 保存必要的数量的临床任务所需的信息 (包括描述研究结果的文本文件, 临床诊断的病人, 结束了医疗事故, 等等);
- 3) 可能的情况下将 “现场” 的研究排除在外, 而不离开数据采集机构的信息轮廓。

在选择阶段也应适用将研究纳入未来的东盟海关数据自动化系统的标准和将研究纳入该系统的标准。在选择国际投资协定的研究时, 可直接进行这项工作, 也可在卸载后立即进行这项工作 (医疗组织的信息轮廓已经超出了范围)。必须注意的是, 这一步骤可能会导致将日期减少几十次。

选择研究时, 还应铭记在第一阶段确定的职等平衡。

例如, 对于数据集 “MosMedData: 阶段1的描述中提到具有肺部病变的超低剂量计算机断层扫描结果”, 第二阶段可以包括以下步骤:

- 1) 选择进行低剂量CT扫描研究以筛查恶性肿瘤的MIS患者;
- 2) 根据入选和排除标准分析选定患者的电子病历 (生活史, 以前的疾病史, 以前的研究数据) 以选择患者, 就第1阶段成立;
- 3) 决定根据所需的班级平衡将研究纳入数据集中。

3. 标记和验证阶段

标记是指确定日期数据项的特征值或特性的过程。基于标记可以对元素进行分类, 然后将它们分配给特定的组。对于标记, 选择初始数据时已经可用的信息 (追溯标记) 和选择阶段之后具有医学教育和/或工作经验的专家进行标记 (预期标记) [9]。

对于追溯标记, 可以使用随附文件中的数据 (例如, 仪器研究结果的结论文本), MIS, 电子病历等。这方面的一个例子是由研究仪器自动生成

并储存在原始数据中的元数据。回顾性标记的明显优势在于,由于大多数准备工作是由数据科学家完成的,因此医疗保健专业人员所需的时间大大减少。

预期标记涉及医疗专业人员在数据集“饱和”过程中的积极参与以及其他信息,例如,使您可以有效地将数据集的元素划分为类和类别。放射诊断学中,标记最常被理解为按类别对研究进行分类(所选疾病是否存在放射学体征)以及与所需体征相对应的感兴趣区域的图形名称(用于例如,多发性硬化症的脱髓鞘病灶在大脑的MR图像上)。参与程度可以分为或多或少的昂贵:在第一种情况下,专家被要求勾勒出感兴趣区域的轮廓,在第二种情况下-用简单的几何图形指定其坐标。

专家意见是确定数据特征或特性的值的最重要因素的情况下,由两名独立专家同时阅读研究是合理的。如果两位专家不一致,则将有争议的研究发送给第三位更有资格的专家(基于实际经验,学位或其他标准)。由三位专家阅读后仍然有争议的研究可能会被认为有争议,并从数据集中排除。根据我们准备的数据集(包括对OGK进行100次CT检查以及各种呼吸系统病理征象)的实践,经过两次独立读数,多达1/4项研究可能引起争议;由第三位更有资格的专家阅读后,多达4%的研究仍可能会引起争议(其医学经验超过5年)。

进行前瞻性标记之前,有必要确定每位专家的研究范围,标记符号的标准,允许以文本,图形或其他方式指定所需标记的软件,以准备标记医生的说明。如果可能的话,准备此类说明的过程应包括在计划阶段定义临床任务的同一工作组。

标记验证为开发人员提供了一定程度的“信任”标记或评估智能系统工作的专家。标记验证可以分为:

- 低至低(存在发现的事实)-根据文档;
- 平均(发现的分类)-基于专家的意见;
- 高(确定的诊断) - 基于更敏感的研究方法或动态观察的结果(在一定时间间隔后重复相同的方法)。

在第4图标记类型的分类数据集的一部分可以具有一个类别,而另一数据集可以具有不同的类别;在同一数据集中允许使用追溯标记和预期标记的组合。标记过程的重要组成部分是随

附文档中的正确文档(请看第4点“文档编制阶段”)。

对于追溯标记和预期标记,可以使用各种技术和编程语言使用各种数据自动化工具(例如,查看医学成像结果并创建二进制蒙版,分析数据库)(C / C ++, Python, Kotlin, Java等)[11]。

文档阶段

数据集通过所有之前的阶段并准备好转移给第三方后,便被视为“准备发布”。数据集的发布伴随着第一个主要版本(1.0.0)的发布,以及随附文档(README文件)的准备和发布。

准备数据集的过程中,不可避免地会忽略某些标准,当最终用户直接使用数据集时(专家们会根据AI验证解决方案的专家或使用机器学习的研究人员)会弹出这些标准。对数据集进行调整应对所有流程参与者和用户透明。数据集版本控制可跟踪此类更改。

我们提出了以下原始方法来解决所描述的问题,这是语义版本控制的一种变体[12]:

1. Major程序版本:当数据集的重要参数发生变化时(与临床任务,目的,数据标记和验证原理相关),该值会增加。
2. Minor程序版本:在替换,添加或删除数据集中的数据单元时增加,而无需更改数据集的其他重要参数;同时,学习或验证算法可以使用新的次要版本,而无需更改代码。发行新的Major程序版本时,Minor程序版本设置为0。
3. Patch程序版本:在调整随附文档,更正拼写错误和标记文件中的其他错误时增加,而数据集中数据单元的数量和质量不变。发行新的Major程序版本和/或Minor程序版本时,Patch程序版本设置为0。

为了便于使用数据集,将Markdown格式的名为README.md的文件和Adobe PDF格式的生成的README.pdf放置在根目录中。README文件结构的统一方法将允许将来为所有已发布的数据集组织方便的搜索和过滤。README文件的基本结构在5图提出了,但是如果需要可以将其他部分添加到文件中。

为了方便保存报告,一个单一的数据集登记册具有实用价值,例如在13表[13]。

建议登记的最低限度字段:

意义	A		B		C	
	预期		预期		回顾性	
	像素遮罩		区域坐标		元数据	
1 确诊		乳腺癌 (组织学数据)		乳腺癌 (组织学数据)		乳腺癌 (组织学数据)
2 发现的分类		BI-RADS 2		BI-RADS 2		BI-RADS 2
3 发现的可用性		“没有焦点” “有焦点”		“没有焦点” “有焦点”		“没有焦点” “有焦点”

图 4按劳动成本和核查程度对标记进行分类。

1. 注册表项的顺序编号。
2. 内部代码对于当前注册表和/或机构中的数据集是唯一的。
3. 数据集的目的, 范围。
4. 方式/程序 (适合于在MIC中进行搜索和选择的研究特征)。
5. 所需的体征和/或目标病理 (如果可能, 请注明国际疾病分类的代码)
6. 数据单位的定义。
7. 数据单位数 (如果可能, 表示以MB, GB或TB为单位的数据输出量)。
8. 标记类, 指示每个类中的记录数。

讨论

本文提出了一种实验方法, 用于形成医学数据集 (数据集), 以用于开发和评估医学诊断智能系统, 使用AI技术。

使用大型MIS (ERIS UMIAS) 作为数据集的数据源是其代表性的一定保证。诊断过程中实施后的AI算法的性能参数很可能与在此类数据集验证期间获得的参数相对应。同时, 有必要考虑诊断设备组的可变性以及正在执行的研究的物理参数的变化, 以尝试在数据集中呈现最广泛的研究。数据集中显示的来自不同制造商的设备可变性值对于微调AI系统的阈值以确保其可靠运行可能具有实际重要性[14]。

<p>数据集名称</p> <hr/> <p>组织徽标</p> <hr/> <p>摘要 (1-2句)</p> <hr/> <p>Disclaimer</p> <ul style="list-style-type: none"> • 数据集分配 • 许可条款 (基本) • 禁止使用的特殊条件 <hr/> <p> 执照指定</p> <hr/> <p>分页符</p>	<p>一般信息</p> <ul style="list-style-type: none"> • 名称 • 内部代码 • 标记类 • 关键词: • 语言 • 资金 • 版本 • 固定链接 • 出版日期 <p>融合和作者</p> <ul style="list-style-type: none"> • 作者 • 融合 <p>数据集结构</p> <ul style="list-style-type: none"> • 目录结构图和文件 • 根目录文件描述 • 目录命名原则说明 • 文件命名原则说明 	<p>数据概述</p> <table border="1"> <thead> <tr> <th>变数</th> <th>数值</th> </tr> </thead> <tbody> <tr> <td> </td> <td> </td> </tr> <tr> <td> </td> <td> </td> </tr> <tr> <td> </td> <td> </td> </tr> </tbody> </table> <ul style="list-style-type: none"> • 研究准备的特点 • 数据标记原则 • 数据标记验证日期 <p>使用和传播规则</p> <ul style="list-style-type: none"> • 许可证 • 复制品 • 推荐的引用形式 • 传播规则 	变数	数值						
变数	数值									

图 5README文件的基本结构。

使用ERIS UMIAS的另一个优势是几乎可以无限地访问成千上万种不同模态的射线研究，从而可以创建具有极为不同的技术，人口统计和临床特征集的数据集。这样的变化保证了生成的数据集的价值不仅用于评估准确性，而且还保证了正在开发和测试的AI系统的可伸缩性和可靠性。

在7个方向的辐射诊断中创建25个数据集时，总共开发了1400多个数据单元（研究），包括在莫斯科实施计算机领域使用创新技术的实验期间，开发并测试了该提议的方法。用于医学图像分析的视觉系统以及在系统中的进一步应用莫斯科的医疗保健[15]（看图3）。表格中提供了完整的数据集列表。本条所述规定应符合基准日期标准，包括在辐射诊断中的智能技术临床试验方法建议[16]。

莫斯科实验期间保证对算法进行两个阶段的独立外部评价（函数和校准测试，分别如下）：首先，使用相对较小的数据集（最多5个数据单元）来检查阅读和处理研究的技术可行性；在第二阶段，使用中等大小的数据集（平均100到200个数据单元）将处理AI研究的结果与经过验证的标记进行比较。如果作为初始测试的结果，基于AI的解决方案的开发人员收到了有关改进其解决方案的建议，则可以在其他数据集上进行重新测试。

作为数据集发布后的数据集生命周期的重要部分是在相关出版物和手稿中科学地展示工作结果。支持在公共数据集上免费发布信息的门户之一是medRxiv²，这是一种生物医学预印本服务。该服务的优势是无需对出版物进行外部同行评审，这可以使社区尽快了解其工作结果。MedRxiv门户上提出 [17]数据集的出版物示例。

应当指出，这种方法生成的数据集已被国内外研究团队成功使用，如最近的出版物所证明的 [18, 19]。实践中使用工作结果可以确认所制定方法和方法的及时性和充分性。

进行必要的更改后，该技术不仅可以完全或部分地用于放射诊断的其他领域，而且可以在实际医学的其他领域（在医疗活动过程中积累主要电子信息）的其他领域完全使用（脑电图，心电图和其他记录生理信号，床边复苏监测仪的

记录，现代实验室设备（例如化学分析仪）的日志记录等）。特别在形成具有心血管疾病迹象的心电图数据集的过程中，成功地测试了制定临床和/或实际问题的原则与MIS一起使用以卸载初始数据，在实验模式下进行标记和记录的一般原则。该技术可以包含在国家标准中，从而确保在国家一级用于教学和测试AI技术的医学数据集的连续性和统一性。

一个热门讨论的问题是医学数据，特别是放射线研究结果的非人格化问题。当前没有使医学图像匿名化的普遍接受的标准。处理此类数据的专业人员必须遵循合理的逻辑，以防止泄露患者的医疗秘密和个人数据。应该记住的是，放射线研究的结果本身可以用作个人数据的来源：例如，可以从头部切口重建面部颅骨软组织的三维图像，这反过来又可以充分识别一个人。尽管在这种情况下缺乏明确的人格化立法规范或标准，但数据集的作者也应做出决定，从临床和/或实际任务开始，从研究中删除头部软组织的记录根据数据集的目的。

为维持医学AI技术市场的增长速度，应尽可能考虑免费提供数据集，但要遵守上述所有匿名化条件。使用诸如arXiv (<https://arxiv.org>)，medRxiv (<https://medrxiv.org>)和Zenodo (<https://zenodo.org>)之类的门户发布描述数据集的文章。开放数据集有大量的公共存储库，并且可以通过它们进行整体搜索，例如Google's Dataset Search³。不仅确保合法访问数据集，而且吸引AI开发人员社区关注的方法之一是在Kaggle⁴等平台上进行AI开发人员之间的在线竞赛。

一个有希望的发展方向是“疾病的数字孪生”的使用-有关各种概况（社会，人口统计，行为等）的患者的广泛信息集，用于形成患有特定疾病的患者的特征性统计标志。使用此类信息可以创建更具代表性的医学数据集，包括对临床和/或实际任务具有重要意义的最广泛的疾病迹象和因素。创建“疾病的数字孪生”的基础首先是分析和处理从“患者的数字孪生”获得的非个人信息，其中包含关于患者的尽可能广泛的多样化信息。

² 访问模式：<https://medrxiv.org>。访问日期：15.01.2021

³ 访问模式：<https://datasetsearch.research.google.com>。访问日期：15.01.2021。

⁴ 访问模式：<https://kaggle.com>访问日期：15.01.2021。

表格使用上述实验方法开发的医学数据集列表

No	内部代码	约定	方式/程序	所需的体征和/或目标疾病	单位数据	几单数据点	标记类 (数据单元数)
1	DS_FT-I_CT_OGK_CANCER	FT-I	CT OGK	肺癌	CT检查	5	没有病理迹象 (2), 有病理迹象 (2), 有技术缺陷 (1)
2	DS_FT-I_CT_OGK_COVID	FT-I	CT OGK	病毒性肺炎 (COVID-19)	CT检查	5	无病理迹象 (2), 有病理迹象 (2)
3	DS_FT-II_CT_OGK_COVID	FT-II	CT OGK	病毒性肺炎 (COVID-19)	CT检查	4	无病理迹象 (2), 有病理迹象 (2)
4	DS_FT-I_LDCT_OGK_CANCER	FT-I	小剂量CT扫描	肺癌	CT检查	5	没有病理迹象 (2), 有病理迹象 (2), 有技术缺陷 (1)
5	DS_FT-I_MMG_CANCER	FT-I	Mammography	乳癌	MMG研究	5	无病理迹象 (2), 有病理迹象 (2)
6	DS_FT-II_MMG_CANCER	FT-II	Mammography	乳癌	MMG研究	4	无病理迹象 (2), 有病理迹象 (2)
7	DS_FT-I_DX_OGK_PAT	FT-I	RG OGK	呼吸病理学	研究射线照相	5	没有病理迹象 (2), 有病理迹象 (2), 有技术缺陷 (1)
8	DS_FT-II_DX_OGK_PAT	FT-II	RG OGK	呼吸病理学	研究射线照相	5	无病理迹象 (2), 有病理迹象 (2)
9	DS_FT-I_DX_OGK_COVID	FT-I	RG OGK	病毒性肺炎 (COVID-19)	研究射线照相	4	无病理迹象 (2), 有病理迹象 (2)
10	DS_FT-II_DX_OGK_COVID	FT-II	RG OGK	病毒性肺炎 (COVID-19)	研究射线照相	4	无病理迹象 (2), 有病理迹象 (2)
11	DS_FT-I_FLG_OGK_PAT	FT-I	FLG OGK	呼吸病理学	FLH研究	4	无病理迹象 (2), 有病理迹象 (2)
12	DS_FT-II_FLG2_OGK_PAT	FT-II	FLG OGK	呼吸病理学	FLH研究	4	无病理迹象 (2), 有病理迹象 (2)
13	DS_CT-I_CT_OGK_CANCER	KIT-I	CT OGK	肺癌	CT检查	100	无病理迹象 (50), 有病理迹象 (50)
14	DS_CT-I_CT_OGK_COVID	KIT-I	CT OGK	病毒性肺炎 (COVID-19)	CT检查	100	无病理迹象 (50), 有病理迹象 (50)

No	内部代码	约定	方式/程序	所需的特征和/或目标疾病	单位数据	几单数据点	标记类 (数据单元数)
15	DS_CT-II_CT_OGK_COVID	K1T-II	CT OGK	病毒性肺炎 (COVID-19)	CT检查	100	无病理迹象 (50), 有病理迹象 (50)
16	DS_CT-II_CT_OGK_COVID_2	K1T-I	CT OGK	病毒性肺炎 (COVID-19)	CT检查	125	CT0 (25), CT1 (25), CT2 (25), CT3 (25), CT4 (25)
17	DS_CT-II_CT_OGK_COVID_3	K1T-I	CT OGK	病毒性肺炎 (COVID-19)	CT检查	200	CT0 (100), CT1 (25), CT2 (25), CT3 (25), CT4 (25)
18	DS_CT-I_LDCT_OGK_CANCER	K1T-I	小剂量CT扫描	肺癌	CT检查	100	
19	DS_CT-I_MMG_CANCER	K1T-I	Mammography	乳癌	MMG研究	100	
20	DS_CT-II_MMG_CANCER	K1T-II	Mammography	乳癌	MMG研究	100	
21	DS_CT-I_DX_OGK_CANCER	K1T-I	RG OGK	呼吸病理学	研究射线照相	100	无病理迹象 (50), 有病理迹象 (50)
22	DS_CT-II_DX_OGK_CANCER	K1T-II	RG OGK	呼吸病理学	研究射线照相	100	
23	DS_CT-I_DX_OGK_COVID	K1T-I	RG OGK	病毒性肺炎 (COVID-19)	研究射线照相	100	
24	DS_CT-I_FLG_OGK_CANCER	K1T-I	FLG OGK	呼吸病理学	FLH研究	100	
25	DS_CT-II_FLG_OGK_CANCER	K1T-II	FLG OGK	呼吸病理学	FLH研究	100	

注意: FT-I-主要功能测试; FT-II-重新功能测试; K1T-I-初步校准测试; K1T-II-重新校准测试; CT-计算机断层扫描; MMG-乳腺X线摄影; RG-射线照相; FLG-荧光照相; OGK-胸部器官; K1T-I-毛玻璃类型的压实区域。肺实质受累≤25%; K1T-II-毛玻璃类型的压实区域。累及肺实质25-50%; K1T-III-毛玻璃类型的压实区域。累及肺实质50-75%; K1T-IV-肺组织的弥漫性压实, 固结区伴铺路石征肺实质受累>75% [13]。

本文介绍的方法可以对数据集及其生命周期进行系统化和标准化,以供以后在测试智能系统(包括基于AI的系统)中使用,并注册经过测试的系统以在医疗保健领域中进一步使用。这样逐步详细的方法来形成数据集将使开发人员能够客观地评估其产品和监管者-使用在建议的方法基础上创建的数据集来确保评估过程的客观性和透明性。

结论

建立基于人工智能技术的医疗数据集,以培训和验证诊断系统,这一任务特别重要,因为这将有助于建立一个有效的医疗信息系统积极发展这一领域。本文所述的原始方法可作为创建各种方法和数据类型的医学数据集的准备和标准

化的完整方法的起点,并确定成功实际应用的必要条件和因素这种方法论。

附加信息

资金来源。作者声称这项研究没有资金支持。

利益冲突。作者声明,没有明显的和潜在的利益冲突相关的发表这篇文章。

作者贡献所有作者都确认其作者符合国际ICMJE标准(所有作者为文章的概念,研究和准备工作做出了重大贡献,并在发表前阅读并批准了最终版本)。最大的贡献分配如下: N.A. Pavlov-出版物结构的形成,出版物文本的编写,开发数据集的方法,数据集示例的开发; S.P. Morozov-研究的一般概念; A.E. Andreichenko-研究设计开发,出版文字撰写的监督,文字编辑; A.V. Vladzimirsky-数据集形成方法的科学证实; A.A. Revazyan-编写文献综述,参与数据集示例的编写; Yu.S. Kirpichev -参与数据集示例的编写。

REFERENCES

- Gusev AV. Prospects for neural networks and deep machine learning in creating health solutions (Complex medical information system, Russian). *Vrach i Informatsionnye Tekhnologii*. 2017;(3):92–105. (In Russ).
- Ranschaert ER, Morozov S, Algra PR, eds. Artificial intelligence in medical imaging. Cham: Springer International Publishing; 2019. doi: 10.1007/978-3-319-94878-2
- Griffith B, Kadom N, Straus CM. Radiology Education in the 21st Century: Threats and Opportunities. *J Am Coll Radiol*. 2019;16(10):1482–1487. doi: 10.1016/j.jacr.2019.04.003
- Savadjiev P, Chong J, Dohan A, et al. Demystification of AI-driven medical image interpretation: past, present and future. *Eur Radiol*. 2019;29(3):1616–1624. doi: 10.1007/s00330-018-5674-x
- Ng A. What artificial intelligence can and can't do right now. *Harvard Business Review*; 2016. Available from: <https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now>
- Renear H, Sacchi S, Wickett KM. Definitions of dataset in the scientific and technical literature. *Proceedings of the American Society for Information Science and Technology*. 2010;47(1):1-4. doi: 10.1002/meet.14504701240
- Tan SL, Gao G, Koch S. Big data and analytics in healthcare. *Methods Inf Med*. 2015;54(6):546–547. doi: 10.3414/ME15-06-1001
- Kohli MD, Summers RM, Geis JR. Medical image data and datasets in the era of machine learning—whitepaper from the 2016 C-MIMI meeting dataset session. *J Digit Imaging*. 2017;30(4):392–399. doi: 10.1007/s10278-017-9976-3
- Willeminck MJ, Koszek WA, Hardell C, et al. Preparing medical imaging data for machine learning. *Radiology*. 2020;295(1):4–15. doi: 10.1148/radiol.2020192224
- Morozov SP, Shelekhov PV, Vladzimirsky AV. Modern approaches to the radiology service improvement. *Health Care Standardization Problems*. 2019;(5-6):30–34. (In Russ). doi: 10.26347/1607-2502201905-06030-034
- Kulberg NS, Gusev MA, Reshetnikov RV, et al. Methodology and tools for creating training samples for artificial intelligence systems for recognizing lung cancer on CT images. *Health Care Russian Federation*. 2021;64(6):343–350. doi: 10.46563/0044-197x-2020-64-6-343-350
- Preston-Werner T. Semantic Versioning 2.0.0 [Internet]. Available from: <https://semver.org>
- Morozov SP, Protsenko DN, Smetanina SV, et al. Radiation diagnostics of coronavirus disease (COVID-19): organization, methodology, interpretation of results: Preprint No.CDT — 2020 — II. Version 2 from 17.04.2020. The series “Best practices of radiation and instrumental diagnostics”. Issue 65. Moscow : Scientific and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Department of Health; 2020. 80 p. (In Russ). Available from: <https://tele-med.ai/biblioteka-dokumentov/luhevaya-diagnostika-koronavirusnoj-bolezni-covid-19-organizaciya-metodologiya-interpretaciya-rezultatov>
- Pavlov N. ECR 2021: Value of technical stratification of medical datasets for AI services. Moscow, 2021. [Internet]. Available from: <https://connect.myesr.org/course/ai-in-breast-imaging/>
- Morozov SP, Vladzimirsky AV, Andreychenko A, et al. Moscow experiment on computer vision in radiology: involvement and participation of radiologists. *Vrach i informacionnye tehnologii*. 2020;(4):14–23. doi: 10.37690/1811-0193-2020-4-14-23
- Morozov SP, Vladzimirsky AV, Klyashtorny VG, et al. Clinical acceptance of software based on artificial intelligence technologies (radiology). Series “Best practices in medical imaging”. Issue 57. Moscow; 2019. 45 p.
- Morozov SP, Andreychenko AE, Pavlov NA, et al. MosMedData: Chest CT scans with COVID-19 related findings dataset. *medRxiv*. 2020. doi: 10.1101/2020.05.20.20100362
- Sushentsev N, Bura V, Kotniket M, et al. A head-to-head comparison of the intra- and interobserver agreement of COVID-RADS

and CO-RADS grading systems in a population with high estimated prevalence of COVID-19. *BJR Open*. 2020;2(1):20200053. doi: 10.1259/bjro.20200053

19. Jin C, Chen W, Caoet Y, et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. *Nat Commun*. 2020;11(1):5088. doi: 10.1038/s41467-020-18685-1

СПИСОК ЛИТЕРАТУРЫ

1. Гусев А.В. Перспективы нейронных сетей и глубокого машинного обучения в создании решений для здравоохранения // *Врач и информационные технологии*. 2017. № 3. С. 92–105.
2. Ranschaert E.R., Morozov S., Algra P.R., eds. *Artificial intelligence in medical imaging*. Cham: Springer International Publishing; 2019. doi: 10.1007/978-3-319-94878-2
3. Griffith B., Kadom N., Straus C.M. Radiology Education in the 21st Century: Threats and Opportunities // *J Am Coll Radiol*. 2019. Vol. 16, N 10. P. 1482–1487. doi: 10.1016/j.jacr.2019.04.003
4. Savadjiev P., Chong J., Dohan A., et al. Demystification of AI-driven medical image interpretation: past, present and future // *European Radiology*. 2019. Vol. 29, N 3, P. 1616–1624. doi: 10.1007/s00330-018-5674-x
5. Ng A. What artificial intelligence can and can't do right now. *Harvard Business Review*; 2016. Available from: <https://hbr.org/2016/11/what-artificial-intelligence-can-and-cant-do-right-now>
6. Renear H., Sacchi S., Wickett K.M. Definitions of dataset in the scientific and technical literature // *Proceedings of the American Society for Information Science and Technology*. 2010. Vol. 47, N 1. P. 1–4. doi: 10.1002/meet.14504701240
7. Tan S.L., Gao G., Koch S. Big data and analytics in health-care // *Methods Inf Med*. 2015. Vol. 54, N 6. P. 546–547. doi: 10.3414/ME15-06-1001
8. Kohli M.D., Summers R.M., Geis J.R. Medical image data and datasets in the era of machine learning—whitepaper from the 2016 C-MIMI meeting dataset session // *J Digit Imaging*. 2017. Vol. 30, N 4. P. 392–399. doi: 10.1007/s10278-017-9976-3
9. Willeminck M.J., Koszek W.A., Hardell C., et al. Preparing medical imaging data for machine learning // *Radiology*. 2020. Vol. 295, N 1. P. 4–15. doi: 10.1148/radiol.2020192224
10. Морозов С.П., Шелехов П.В., Владимирский А.В. Современные стандартизованные подходы к совершенствованию службы лучевой диагностики // *Проблемы стандартизации в здравоохранении*. 2019. № 5-6. С. 30–34. doi: 10.26347/1607-2502201905-06030-034
11. Kulberg N.S., Gusev M.A., Reshetnikov R.V., et al. Methodology and tools for creating training samples for artificial intelligence systems for recognizing lung cancer on CT images // *Health Care Russian Federation*. 2021. Vol. 64, N 6. P. 343–350. doi: 10.46563/0044-197x-2020-64-6-343-350

12. Preston-Werner T. Semantic Versioning 2.0.0 [Internet]. Available from: <https://semver.org>

13. Морозов С.П., Проценко Д.Н., Сметанина С.В. и др. Лучевая диагностика коронавирусной болезни (COVID-19): организация, методология, интерпретация результатов : препринт № ЦДТ — 2020 — II. Версия 2 от 17.04.2020. Серия «Лучшие практики лучевой и инструментальной диагностики». Вып. 65. Москва : ГБУЗ НПКЦ ДиТ ДЗМ, 2020. 80 с. Режим доступа: <https://tele-med.ai/biblioteka-dokumentov/luchevaya-diagnostika-koronavirusnoj-bolezni-covid-19-organizaciya-metodologiya-interpretaciya-rezultatov>. Дата обращения: 15.01.2021.

14. Pavlov N. ECR 2021: Value of technical stratification of medical datasets for AI services. Moscow, 2021. [Internet]. Available from: <https://connect.mysr.org/course/ai-in-breast-imaging/>

15. Морозов С.П., Владимирский А.В., Ледихова Н.В. и др. Московский эксперимент по применению компьютерного зрения в лучевой диагностике: вовлеченность врачей-рентгенологов // *Врач и информационные технологии*. 2020. № 4. С. 14–23. doi: 10.37690/1811-0193-2020-4-14-23

16. Morozov S.P., Vladzimirskyy A.V., Klyashtornyy V.G., et al. Clinical acceptance of software based on artificial intelligence technologies (radiology). Series «Best practices in medical imaging». Issue 57. Moscow; 2019. 45 p.

17. Morozov S.P., Andreychenko A.E., Pavlov N.A., et al. MosMed-Data: Chest CT scans with COVID-19 related findings dataset // *medRxiv*. 2020. doi: 10.1101/2020.05.20.20100362

18. Sushentsev N., Bura V., Kotniket M., et al. A head-to-head comparison of the intra- and interobserver agreement of COVID-RADS and CO-RADS grading systems in a population with high estimated prevalence of COVID-19 // *BJR Open*. 2020. Vol. 2, N 1. P. 20200053. doi: 10.1259/bjro.20200053

19. Jin C., Chen W., Caoet Y., et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis // *Nat Commun*. 2020. Vol. 11, N 1. P. 5088. doi: 10.1038/s41467-020-18685-1

AUTHORS' INFO

*Nikolay A. Pavlov, MD;

address: 28-1, Srednyaya Kalitnikovskaya street, 109029, Moscow, Russia; tel.: +7 (495) 276-04-36;
ORCID: <https://orcid.org/0000-0002-4309-1868>;
eLibrary SPIN: 9960-4160; e-mail: n.pavlov@npcmr.ru

Anna E. Andreychenko, PhD;

ORCID: <https://orcid.org/0000-0001-6359-0763>;
eLibrary SPIN: 6625-4186;
e-mail: a.andreychenko@npcmr.ru

ОБ АВТОРАХ

*Павлов Николай Александрович;

адрес: Россия, 109029, Москва, ул. Средняя Калитниковская, д. 28, стр. 1; тел.: +7 (495) 276-04-36;
ORCID: <https://orcid.org/0000-0002-4309-1868>;
eLibrary SPIN: 9960-4160; e-mail: n.pavlov@npcmr.ru

Андрейченко Анна Евгеньевна, к.ф.-м.н.;

ORCID: <https://orcid.org/0000-0001-6359-0763>;
eLibrary SPIN: 6625-4186;
e-mail: a.andreychenko@npcmr.ru

Anton V. Vladzimirskyy, MD, Dr. Sci. (Med.);
ORCID: <https://orcid.org/0000-0002-2990-7736>;
eLibrary SPIN: 3602-7120; e-mail: a.vladzimirsky@npcmr.ru

Anush A. Revazyan;
ORCID: <http://orcid.org/0000-0003-1589-2382>;
e-mail: anushrevazyan@gmail.com

Yury S. Kirpichev;
ORCID: <http://orcid.org/0000-0002-9583-5187>;
eLibrary SPIN: 3362-3428; e-mail: y.kirpichev@npcmr.ru

Sergey P. Morozov, MD, Dr. Sci. (Med.), Professor;
ORCID: <http://orcid.org/0000-0001-6545-6170>;
eLibrary SPIN: 8542-1720; e-mail: morozov@npcmr.ru

Владзимирский Антон Вячеславович, д.м.н.;
ORCID: <https://orcid.org/0000-0002-2990-7736>;
eLibrary SPIN: 3602-7120; e-mail: a.vladzimirsky@npcmr.ru

Ревазян Ануш Артуровна;
ORCID: <http://orcid.org/0000-0003-1589-2382>;
e-mail: anushrevazyan@gmail.com

Кирпичев Юрий Сергеевич;
ORCID: <http://orcid.org/0000-0002-9583-5187>;
eLibrary SPIN: 3362-3428; e-mail: y.kirpichev@npcmr.ru

Морозов Сергей Павлович, д.м.н., профессор;
ORCID: <http://orcid.org/0000-0001-6545-6170>;
eLibrary SPIN: 8542-1720; e-mail: morozov@npcmr.ru