REVIEW PAPER

JMS Journal of Medical Science

Ethical issues on artificial intelligence in radiology: how is it reported in research articles? The current state and future directions

Tomasz Piotrowski
Department of Electroradiology, Poznan University of Medical Sciences, Poland
https://orcid.org/0000-0002-8894-7412
Corresponding author: tomaszpiotrowski@ump.edu.pl

Joanna Kazmierska
Department of Electroradiology, Poznan University of Medical Sciences, Poland; Radiotherapy Department II, Greater Poland Cancer Centre, Poznan, Poland
https://orcid.org/0000-0001-7160-8150

Mirosława Mocydlarz-Adamcewicz
Department of Electroradiology, Poznan University of Medical Sciences, Poland; IT Department, Greater Poland Cancer Centre, Poznań, Poland
https://orcid.org/0000-0001-9201-9099

Adam Ryczkowski
Department of Electroradiology, Poznan University of Medical Sciences, Poland; Department of Medical Physics, Greater Poland Cancer Centre, Poznań, Poland
https://orcid.org/0000-0001-5106-5152

DOI: https://doi.org/10.20883/medical.e513

Keywords: ethics, radiology, artificial intelligence, diagnostics, medical imaging

Published: 2021-06-29

How to Cite: Piotrowski T, Kazmierska J, Mocydlarz-Adamcewicz M, Ryczkowski A. Ethical issues on artificial intelligence in radiology: how is it reported in research articles? The current state and future directions. JMS. 2021 Jun. 29;90(2):e513. doi:10.20883/medical.e513

ABSTRACT

Background. This paper evaluates the status of reporting information related to the usage and ethical issues of artificial intelligence (AI) procedures in clinical trial (CT) papers focussed on radiology issues as well as other (non-trial) original radiology articles (OA).

Material and Methods. The evaluation was performed by three independent observers who were, respectively physicist, physician and computer scientist. The analysis was performed for two groups of publications, i.e., for CT and OA. Each group included 30 papers published from 2018 to 2020, published before guidelines proposed by Liu et al. (Nat Med. 2020; 26:1364-1374). The set of items used to catalogue and to verify the ethical status of the AI reporting was developed using the above-mentioned guidelines.

Results. Most of the reviewed studies, clearly stated their use of AI methods and more importantly, almost all tried to address relevant clinical questions. Although in most of the studies, patient inclusion and exclusion criteria were presented, the widespread lack of rigorous descriptions of the study design apart from a detailed explanation of the AI approach itself is noticeable. Few of the chosen studies provided information about anonymization of data and the process of secure data sharing. Only a few studies explore the patterns of incorrect predictions by the proposed AI tools and their possible reasons.

Conclusion. Results of review support idea of implementation of uniform guidelines for designing and reporting studies with use of AI tools. Such guidelines help to design robust, transparent and reproducible tools for use in real life.

Introduction

Advances in radiology directly correlate with developments in imaging technology [1]. Developing new imaging modalities or increasing the efficacy of already implemented solutions improves the decision-making process in routine work of radiologists making their analysis more accurate. To implement new hardware solutions with dedicated software from factory to clinic, an appropriate certificate and regulation (e.g., Conformité Européenne, Food and Drug Administration, EU Medical Device Regulation) needs to be obtained and, then, the usefulness of solutions needs to be carefully evaluated in specific areas of usage. After this process is over, the use of these tools is clearly defined and established in routine work. The challenge starts when the machine ceases to be a tool in the hands of radiologists and becomes their advisor, e.g., decision support systems based on artificial intelligence (AI).

AI describes a range of techniques that allow computers to perform tasks that require human reasoning and problem-solving skills [2]. AI is encapsulated in software for which advanced mathematical algorithms (e.g., machine learning) are implemented to automate work or support human decisions [3,4]. AI is not a new concept in radiology. Over the last 10 years (from 2010 to 2020), over 6,000 original papers describing the implementation and use of AI methods in radiology have been published (source: authors' search with the PubMed engine). However, for radiologists, AI is a new tool that not only gives the radiologist the content for interpretation but also tries to interpret this content for them. This fact revolutionized common thinking about radiology tools and forced the radiologist community to redefine tools used in routine work, especially in legal and ethical terms. Indeed, in the last three years (from 2017 to 2020) over 100 statements, editorials, review and commentary articles have been published about ethical aspects of AI usage in radiology (source: authors' search with the PubMed engine). These papers focused on fundamental ethical aspects of diagnostics, as well as ethical issues connected to every step of the diagnostic process supported by AI. Neri *et al*. [5] emphasize that it is the radiologist who is responsible for diagnosis, not the AI tool, designed to support it. Patients should always sign informed consent for their data to be used in this non-conventional way. The radiologist should know how to use AI tools. AI operating patterns should be transparent and as clear as possible and, finally, when using AI tools radiologists need to take responsibility for the accuracy of the AI suggestion as it may bias their final diagnosis. The European and North American multi-society statement [6] is one of the essential papers that describe in detail every step of the diagnostic process supported by AI. This multi-institutional report identified three main areas of the process that require new regulations. These are: data processing, transparency of algorithms and trained models and the relationship between patients and radiologists. While the statement answers "*why*" it is needed, Brady and Neri [7] tried to answer "*how*" to do it. Showing the examples of how to resolve new challenges, they pointed out and highlighted that the main challenge was to anticipate how rapidly evolving systems might go wrong or could be abused and to prevent these possible outcomes before they occur [8]. While establishing correct rules of practice for AI is a key to its proper implementation in hospitals, correct reporting in scientific reports should not be forgotten either. At the end of 2020, a consensus statement was published on reporting trials involving AI procedures [9].

This study is a retrospective review of original articles published in the last three years in the field of radiology assessing the methods of reporting information related to the usage and ethical issues of AI procedures.

Material and Methods

Literature search

An initial list of 4,301 items was generated by PubMed engine through the review of literature published over the last three years (from 2018 to 2020). When analysing the number of publications during the previous ten years (from 2010 to 2020), we noticed that more than 65% of articles were published in the last three years; hence we decided to limit our analysis to this period. The search was performed on 20 October 2020 using the terms 'artificial intelligence', 'machine learning' and 'deep learning' to identify published original articles for AI interventions in radiology. The search excluded review studies and statement or editorial articles.

The query box used during the search was: *((Radiology) AND ((Artificial intelligence) OR (Machine Learning) OR (Deep Learning))) NOT ((Review) OR (Statement) OR (Editorial))*.

In the next step, from the initial cohort, the PubMed engine built-in filter was used to extract 51 clinical trials (CT) written in English. After reading the abstracts, we narrowed the list of publications to 30 CTs that focused directly on diagnostic or interventional radiology (we excluded articles where radiology was just a tool) [10-39]. These articles constituted the first arm of the study. The second arm included 30 non-trial (original) articles (OA) randomly sampled from the initial cohort of articles [40-69]. Figure 1 shows the flow diagram that include each step of including/excluding process of the papers to this review [70].

Scope and method of analysis

The guideline published by Liu et al. [9] was used to develop the check list that was used to

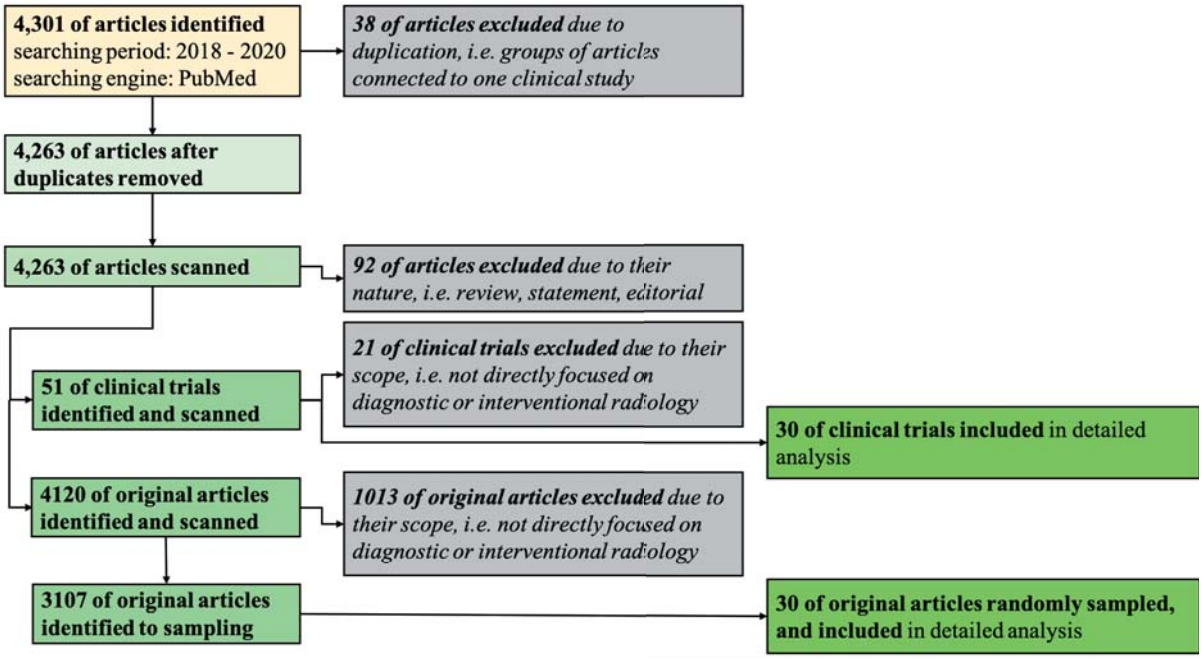

Figure 1. Flow of information through the different phases of review

Table 1. The set of items used to check the status of AI reporting in the analysed groups of articles

Item	The scope of the assessment
Q1	The title includes information on AI or, in the abstract, the use of AI intervention within the study was clearly stated.
Q2	The AI intervention was adequately justified in the context of the clinical pathway.
Q3	The inclusion and exclusion criteria at the level of input data as well as participants were stated.
Q4	Clear description of how the AI intervention was integrated into the study setting, including any onsite or offsite requirements.
Q5	Was the version of the AI algorithm stated?
Q6	Were patients informed and did they sign the consent?
Q7	Was data anonymization used and was the method described?
Q8	Were the data shared?
Q9	Description of how low quality or unavailable input data were assessed and handled.
Q10	Checking whether there was human-AI interaction in the handling of the input data, and what level of expertise was required of users.
Q11	Checking the explanations of how AI intervention outcomes contributed to decision-making or other elements of clinical practice.
Q12	How were potential harms described, i.e., description of any analysis of performance errors and how errors were identified, where applicable.
Q13	Checking whether information was provided on how AI intervention and/or its code can be accessed, including any restrictions to access or re-use.

assess accuracy, transparency and ethical issues of the AI reporting in specific parts of the article (Table 1).

Both groups (CT and OA) were independently scored for the items included in Table 1 by three observers: physicist (OBS_1), physician (OBS_2), and computer scientist (OBS_3). The assessment was made for each item on a two-stage scale (meet/fail). The Cohen's Kappa coefficient was used to measure inter-observer reliability. Separately for CT and OA, the maximum difference between observers' scores (MDO) was counted. The average percentage (AP) of the positive scores was calculated for CT and OA, and every item of Table 1. The results obtained for CT and OA were compared using Fisher's exact test. Moreover, the relative difference between AP for CT and OA was calculated.

All tests were performed at the significance level $\alpha = 0.05$, using XLSTAT software (Addinsoft SARL, New York, USA) in an MS Excel environment (Microsoft Corp., Redmond, WA, USA).

Results

The analysis includes articles prepared 'on the eve' of the publication of Liu *et al*. guidelines [9]. Therefore, the criteria used to assess the AI reporting in the studied articles were not available for the authors of the cited studies at the time of publications.

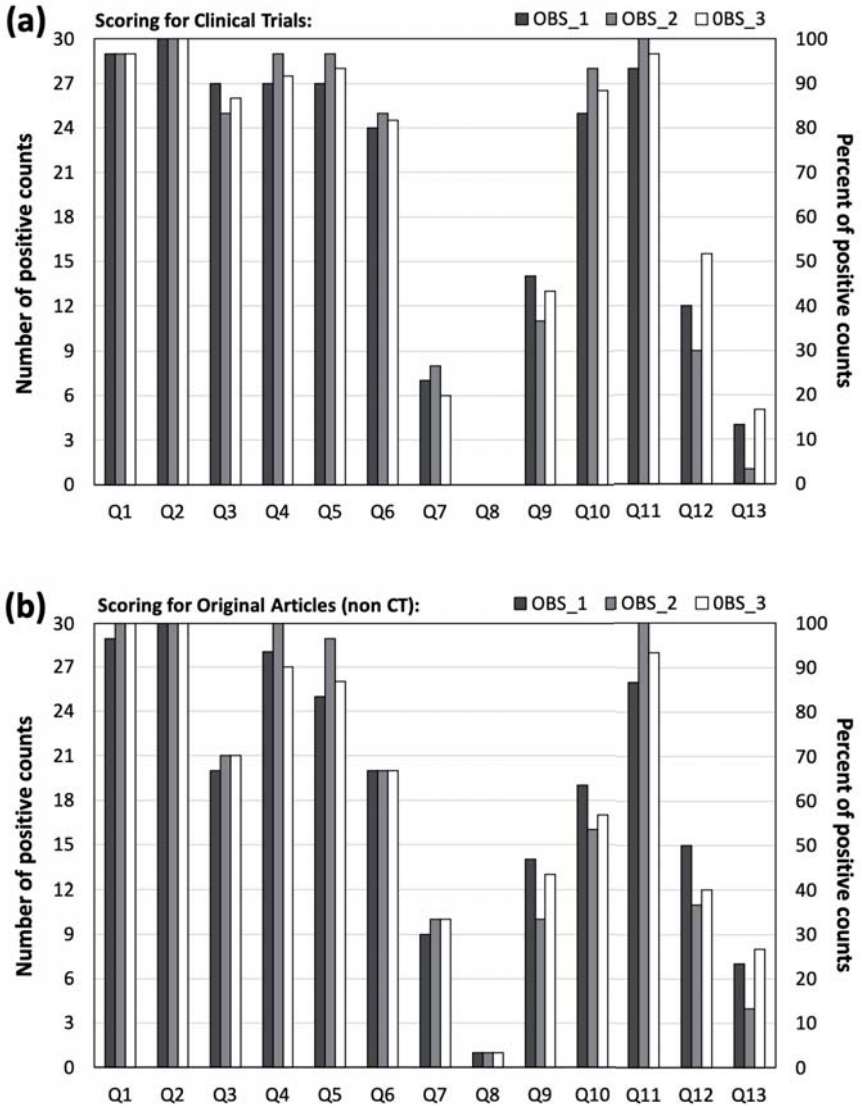

Figure 2. Positive counts from three observers related to (a) clinical trial and (b) original articles. The value of Q8 for clinical trials (a) was zero for each observer. Abbreviations: Q1-Q13 – the items described in Table 1; OBS_1 – first observer (physicist); OBS_2 – second observer (physician); OBS_3 – third observer (computer scientist)

Figure 2 shows the scores granted by every observer for CT (Figure 2a) and OA (Figure 2b) groups. The highest MDOs was 4 (13%) for the CT as well as for the OA group. While, in the CT group, these MDOs were connected to Q12 and Q13 items, in the OA group they were also linked to Q5, Q9, and Q11. The analysis of Cohen's Kappa coefficients (Figure 3) confirmed the lowest agreement between observers' scoring for Q5, Q11, Q13. Small Cohen's Kappa value was also observed for Q4 where MDO's were relatively high (i.e., 7% for CT and 10% for OA). All obtained

Item	OBS_1 *vs* OBS_2	OBS_1 *vs* OBS_3	OBS_2 *vs* OBS_3
Q1	0.66	0.66	1.00
Q2	1.00	1.00	1.00
Q3	0.90	1.00	0.95
Q4	0.31	1.00	0.31
Q5	0.37	0.84	0.47
Q6	0.96	1.00	0.96
Q7	0.92	1.00	0.92
Q8	1.00	1.00	1.00
Q9	0.76	0.93	0.83
Q10	1.00	0.96	0.96
Q11	0.27	0.64	0.49
Q12	0.76	0.97	0.73
Q13	0.58	0.90	0.50

The scale of agreement quality according to Kappa values:

0.8 - 1.0	almost perfect
0.6 - 0.8	substantial
0.4 - 0.6	moderate
0.2 - 0.4	fair agreement
0.0 - 0.2	small

Figure 3. Cohen's Kappa coefficients for agreement of the judges' answers. Abbreviations: Q1-Q13 – the items described in Table 1; OBS_1 – first observer (physicist); OBS_2 – second observer (physician); OBS_3 – third observer (computer scientist)

Table 2. The maximum differences between observers' scores and the relative difference between the average percentage of the positive scores for clinical trial and original articles groups. Statistical comparison performed by Fisher's exact test on the significance level equal to 0.05

Item	MDO		$\lvert AP_{CT} - AP_{OA} \rvert$	Fisher's exact test
	CT	OA		
	Values in numbers and (%)			
Q1	0 (0%)	1 (3%)	2%	p = 0.621
Q2	0 (0%)	0 (0%)	0%	p = 1.000
Q3	2 (7%)	1 (3%)	18%	p = 0.007
Q4	2 (7%)	3 (10%)	2%	p = 0.767
Q5	2 (7%)	4 (13%)	4%	p = 0.433
Q6	1 (3%)	0 (0%)	15%	p = 0.028
Q7	2 (7%)	1 (3%)	9%	p = 0.244
Q8	0 (0%)	0 (0%)	3%	p = 0.246
Q9	3 (10%)	4 (13%)	1%	p = 1.000
Q10	3 (10%)	3 (10%)	31%	p < 0.001
Q11	2 (7%)	4 (13%)	3%	p = 0.497
Q12	4 (13%)	4 (13%)	2%	p = 0.881
Q13	4 (13%)	4 (13%)	10%	p = 0.104

MDO - the maximum difference between observers' scores; CT - the group including Clinical Trials; OA - the group including Original Articles that are not CT; AP - the average percentage of the positive scores

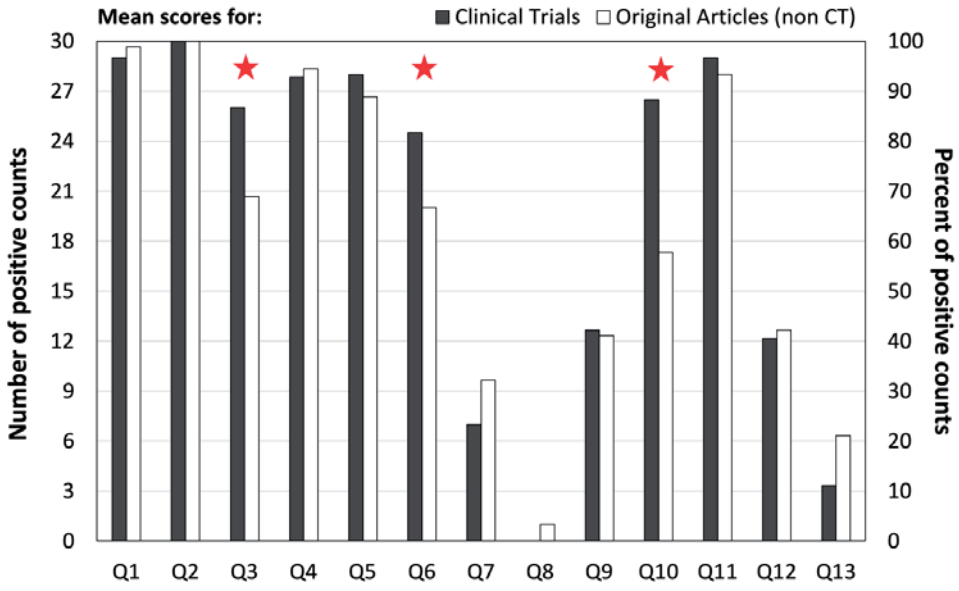

Figure 4. Averaged positive counts for the groups of clinical trial and original articles. Abbreviations: Q1-Q13 – the items described in Table 1. Red asterisk: statistically significant difference between scores granted to clinical trials and original articles

Cohen's Kappa values ranged on the scale proposed by Landis and Koch [71] from "*fair agreement*" to "*almost perfect*" level.

Table 2 shows detailed information of the MDOs in the CT and the OA groups, the relative differences between the average percentage of the positive scores counted in CT and OA, and the statistic results of CT *vs* OA comparison for every item from Table 1.

Figure 4 shows the AP of positive counts for the CT and the OA. The results lower than 50% of the passing checks in both groups were noted for five items - Q7, Q8, Q9, Q12, and Q13. While for Q3, Q6, and Q10 the passing checks were above 50% (ranged from 58% to 88%), the AP counted for CT and OA differed significantly among themselves (Fisher's exact test performed at α=0.05) (Table 2). The highest passing checks (> 90%) with smallest differences between CT and OA were observed for Q1, Q2 and Q11.

By analysing the items, we noted that only for Q7 and Q8, the passing checks were higher for OA than CT. It should be noted that the Q8 item was scored as incorrect with only one article assessed as meeting those criteria.

Discussion

The checklist proposed by Liu *et al*. [9] puts forward important criteria for safe and effective integration of AI into clinical practice, defining clear criteria of study design, data management and patients' rights to privacy. The criteria list used to score the articles, presented in Table 1, was based directly on the checklist of Liu et al. and contains all the key criteria presented by them. Although not validated yet, such proposal is a good starting point for future guidelines for authors and editors, regarding minimal standard criteria for publication of studies integrating AI tools. The assessment criteria used were not published at the time of publication of reviewed articles. Therefore, the results of our study should be interpreted as an indication of areas where the authors of future AI articles should put higher attention, based on published Liu recommendation.

Most of the reviewed studies, as expected, clearly stated their use of AI methods and, more importantly, almost all tried to address relevant clinical questions (Q1, Q2). Although in most of the studies patient inclusion and exclusion criteria were presented, they lacked widespread rigorous descriptions of the study design (Q3, Q4, Q10, Q11) apart from a detailed explanation of the AI approach itself (Q5). These concerns fit the broader discussion about transparency and reproducibility in AI research which includes reporting of data selection and flow. Additionally, using data collected in routine clinical practice - as opposed to highly curated datasets, e.g. from clinical trials - can produce the 'garbage-in, garbage out' phenomenon due to low quality or missing data points, creating the risk of wrong clinical decision. Unfortunately, this problem is rarely addressed in the reviewed publications (Q9).

A small discrepancy in the evaluation between a clinical observer (OBS_2) and a technical-scientific observer (OBS_1, OBS_3) for the items (Q9, Q10, Q11, Q12, Q13), do not change the final assessment of the quality of the articles analyzed. Rather, a possible hypothesis for future research studies is that different level of perception of ethics among different observers could have an impact on patient data and study management when applied to clinical use.

Another important part of every study protocol is informed consent of the participants (Q6). Many studies reported that local Institutional Review Board (IRB) had waived patients' consent. It can be understood when the study uses retrospective data but for prospective trials, even when images are the only subject of research, such consent should be mandatory. Very few of the chosen studies provided information about anonymization of data and process for safe sharing (Q7, Q8). The discussion regarding data ownership is still ongoing, and it is not clear who should be responsible for the evaluation of trade-off between the potential benefit for future patients and privacy concerns when patient data is released. Is patients' consent needed for sharing their data publicly or with non-medical companies or is IRB judgement sufficient? With the increasing involvement of non-medical technology companies like Google or Facebook in healthcare and the associated quest for sensitive medical data, these questions will become even more relevant in the near future [72,73]

Only a few studies explore the patterns of incorrect predictions by the proposed AI tools and possible reasons. Such analysis is impor-

tant for the evaluation of model performance as every error carries a potential cost and risk for the patient and the clinician who is fully responsible for the decision made or augmented by an AI model (Q12).

To gain the trust of clinicians, AI tools designed for use in clinical routine should be robust and transparent. Studies proposing these tools must be transparent and reproducible. According to 2020 State of AI report [74], only 15% of AI studies made the code used to train and validate the proposed models publicly available. This is clearly seen in the set of studies included in this review as only five (5%) of them are accompanied by an open-source code and metadata (Q13). The international research community have recently raised a concern about the replicability of AI research regarding a publication on an AI tool for breast cancer diagnosis built by Google researchers with no open access to the code [72]. In their comments, the authors argue that sharing key materials, like code and metadata, would allow verification of results by other scientists. Without this, published results are rather like a "promotion of closed technology" [75].

Conclusion

Recommendation on how to report results of studies with use or development of AI tools are important and should be implemented by authors and editors to increase robustness and replicability of their work. The review shows that authors of studies using AI tools should put more emphasis on the accurate description of the study design to increase transparency and reproducible of their works.

Acknowledgements

Conflict of interest statement
The authors declare no conflict of interest.

Funding sources
There are no sources of funding to declare.

References

1. Kruskal JB, Eisenberg R, Sosna J, Yam CS, Kruskal JD, Boiselle PM. Quality improvement in radiology: basic principles and tools required to achieve success. RadioGraphics 2011; 31(6):1499-1509. DOI: 10.1148/rg.316115501
2. Academy of Royal Medical Colleges: Artificial Intelligence in Healthcare. The report published on 28.01.2019 at the Academy website: http://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare
3. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. Nat Rev Cancer 2018; 18:500–10. DOI: 10.1038/s41568-018-0016-5
4. Kazmierska J, Hope A, Spezi E, Beddar S, Nailon WH, Osong B, et al. From multisource data to clinical decision aids in radiation oncology: the need for a clinical data science community. Radiotherapy and Oncology 2020; (Online ahead of print). DOI: 10.1016/j.radonc.2020.09.054
5. Neri E, Coppola F, Miele V, Bibbolino C, Grassi R. Artificial intelligence: Who is responsible for the diagnosis? La Radiol Med 2020; 125:517–521. DOI: 10.1007/s11547-020-01135-9
6. Geis JR, Brady A, Wu CC, Spencer J, Ranschaert E, Jaremko JL, et al. Ethics of artificial intelligence in radiology: summary of the joint European and North American multisociety statement. Radiology 2019; 293:436-440. DOI: 10.1148/radiol.2019191586
7. Brady AP, Neri E. Artificial intelligence in radiology – ethical considerations. Diagnostics (Basel) 2020; 10:231. DOI: 10.3390/diagnostics10040231
8. Kohli M, Geis R. Artificial Intelligence and Radiology. J Am Coll Radiol. 2018; 15:1317–1319. DOI: 10.1016/j.jacr.2018.05.020
9. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020; 26:1364-1374. DOI: 10.1038/s41591-020-1034-x.
10. Baskaran L, Ying X, Xu Z, Al'Aref SJ, Lee BC, Lee SE, et al. Machine learning insight into the role of imaging and clinical variables for the prediction of obstructive coronary artery disease and revascularization: An exploratory analysis of the CONSERVE study. PLoS One. 2020; 15(6):e0233791. DOI: 10.1371/journal.pone.0233791.
11. Leung DG, Bocchieri AE, Ahlawat S, Jacobs MA, Parekh VS, Braverman V, et al. Longitudinal functional and imaging outcome measures in FKRP limb-girdle muscular dystrophy. BMC Neurol. 2020; 20(1):196. DOI: 10.1186/s12883-020-01774-5.
12. Eresen A, Li Y, Yang J, Shangguan J, Velichko Y, Yaghmai V, et al. Preoperative assessment of lymph node metastasis in Colon Cancer patients using machine learning: a pilot study. Cancer Imaging. 2020; 20(1):30. DOI: 10.1186/s40644-020-00308-z.
13. Huang Z, Liu D, Chen X, Yu P, Wu J, Song B, et al. Retrospective imaging studies of gastric cancer: Study protocol clinical trial (SPIRIT Compliant). Medicine (Baltimore). 2020; 99(8):e19157. DOI: 10.1097/MD.0000000000019157.
14. Eisenberg E, McElhinney PA, Commandeur F, Chen X, Cadet S, Goeller M, et al. Deep learning-based quantification of epicardial adipose tissue volume and attenuation predicts major adverse cardiovas-

cular events in asymptomatic subjects. Circ Cardiovasc Imaging. 2020; 13(2):e009829. DOI: 10.1161/CIRCIMAGING.119.009829.
15. Wang D, Xu J, Zhang Z, Li S, Zhang X, Zhou Y, et al. Evaluation of rectal cancer circumferential resection margin using faster region-based convolutional neural network in high-resolution magnetic resonance images. Dis Colon Rectum. 2020; 63(2):143-151. DOI: 10.1097/DCR.0000000000001519.
16. Hilbert A, Ramos LA, van Os HJA, Olabarriaga SD, Tolhuisen ML, Wermer MJH, et al. Data-efficient deep learning of radiological image data for outcome prediction after endovascular treatment of patients with acute ischemic stroke. Comput Biol Med. 2019; 115:103516. DOI: 10.1016/j.compbiomed.2019.103516.
17. Sheth SA, Lopez-Rivera V, Barman A, Grotta JC, Yoo AJ, Lee S, et al. Machine learning-enabled automated determination of acute ischemic core from computed tomography angiography. Stroke. 2019; 50(11):3093-3100. DOI: 10.1161/STROKEAHA.119.026189.
18. Bielak L, Wiedenmann N, Nicolay NH, Lottner T, Fischer J, Bunea H, et al. Automatic tumor segmentation with a convolutional neural network in multiparametric MRI: influence of distortion correction. Tomography. 2019; 5(3):292-299. DOI: 10.18383/j.tom.2019.00010.
19. Bhattarai S, Klimov S, Aleskandarany MA, Burrell H, Wormall A, Green AR, et al. Machine learning-based prediction of breast cancer growth rate in vivo. Br J Cancer. 2019; 121(6):497-504. DOI: 10.1038/s41416-019-0539-x.
20. Toivonen J, Montoya Perez I, Movahedi P, Merisaari H, Pesola M, Taimen P, et al. Radiomics and machine learning of multisequence multiparametric prostate MRI: Towards improved non-invasive prostate cancer characterization. PLoS One. 2019; 14(7):e0217702. DOI: 10.1371/journal.pone.0217702.
21. van Gastel MDA, Edwards ME, Torres VE, Erickson BJ, Gansevoort RT, Kline TL. Automatic measurement of kidney and liver volumes from MR images of patients affected by autosomal dominant polycystic kidney disease. J Am Soc Nephrol. 2019; 30(8):1514-1522. DOI: 10.1681/ASN.2018090902.
22. Kurata A, Fukuyama N, Hirai K, Kawaguchi N, Tanabe Y, Okayama H, et al. On-site computed tomography-derived fractional flow reserve using a machine-learning algorithm - clinical effectiveness in a retrospective multicenter cohort. Circ J. 2019; 83(7):1563-1571. DOI: 10.1253/circj.CJ-19-0163.
23. Park A, Chute C, Rajpurkar P, Lou J, Ball RL, Shpanskaya K, et al. Deep learning-assisted diagnosis of cerebral aneurysms using the HeadXNet model. JAMA Netw Open. 2019; 2(6):e195600. DOI: 10.1001/jamanetworkopen.2019.5600.
24. Shayesteh SP, Alikhassi A, Fard Esfahani A, Miraie M, Geramifar P, Bitarafan-Rajabi A, et al. Neo-adjuvant chemoradiotherapy response prediction using MRI based ensemble learning method in rectal cancer patients. Phys Med. 2019; 62:111-119. DOI: 10.1016/j.ejmp.2019.03.013.
25. Fiehler J, Thomalla G, Bernhardt M, Kniep H, Berlis A, Dorn F, et al. ERASER. Stroke. 2019; 50(5):1275-1278. DOI: 10.1161/STROKEAHA.119.024858.
26. Gates EDH, Lin JS, Weinberg JS, Hamilton J, Prabhu SS, Hazle JD, et al. Guiding the first biopsy in glioma patients using estimated Ki-67 maps derived from MRI: conventional versus advanced imaging. Neuro Oncol. 2019; 21(4):527-536. DOI: 10.1093/neuonc/noz004.
27. Shan QY, Hu HT, Feng ST, Peng ZP, Chen SL, Zhou Q, et al. CT-based peritumoral radiomics signatures to predict early recurrence in hepatocellular carcinoma after curative tumor resection or ablation. Cancer Imaging. 2019; 19(1):11. DOI: 10.1186/s40644-019-0197-5.
28. Cho H, Lee JG, Kang SJ, Kim WJ, Choi SY, Ko J, et al. Angiography-based machine learning for predicting fractional flow reserve in intermediate coronary artery lesions. J Am Heart Assoc. 2019; 8(4):e011685. DOI: 10.1161/JAHA.118.011685.
29. Tu Y, Ortiz A, Gollub RL, Cao J, Gerber J, Lang C, et al. Multivariate resting-state functional connectivity predicts responses to real and sham acupuncture treatment in chronic low back pain. Neuroimage Clin. 2019; 23:101885. DOI: 10.1016/j.nicl.2019.101885.
30. Togo R, Hirata K, Manabe O, Ohira H, Tsujino I, Magota K, et al. Cardiac sarcoidosis classification with deep convolutional neural network-based features using polar maps. Comput Biol Med. 2019; 104:81-86. DOI: 10.1016/j.compbiomed.2018.11.008.
31. Stuckey TD, Gammon RS, Goswami R, Depta JP, Steuter JA, Meine FJ 3rd, et al. Cardiac phase space tomography: a novel method of assessing coronary artery disease utilizing machine learning. PLoS One. 2018; 13(8):e0198603. DOI: 10.1371/journal.pone.0198603.
32. Wang XH, Jiao Y, Li L. Identifying individuals with attention deficit hyperactivity disorder based on temporal variability of dynamic functional connectivity. Sci Rep. 2018; 8(1):11789. DOI: 10.1038/s41598-018-30308-w.
33. de Jong EEC, van Elmpt W, Rizzo S, Colarieti A, Spitaleri G, Leijenaar RTH, et al. Applicability of a prognostic CT-based radiomic signature model trained on stage I-III non-small cell lung cancer in stage IV non-small cell lung cancer. Lung Cancer. 2018; 124:6-11. DOI: 10.1016/j.lungcan.2018.07.023.
34. Jun Y, Eo T, Kim T, Shin H, Hwang D, Bae SH, et al. Deep-learned 3D black-blood imaging using automatic labelling technique and 3D convolutional neural networks for detecting metastatic brain tumors. Sci Rep. 2018; 8(1):9450. DOI: 10.1038/s41598-018-27742-1.
35. Citak-Er F, Firat Z, Kovanlikaya I, Ture U, Ozturk-Isik E. Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3T. Comput Biol Med. 2018; 99:154-160. DOI: 10.1016/j.compbiomed.2018.06.009.
36. Mokhtari F, Rejeski WJ, Zhu Y, Wu G, Simpson SL, Burdette JH, et al. Dynamic fMRI networks predict success in a behavioral weight loss program among

older adults. Neuroimage. 2018; 173:421-433. DOI: 10.1016/j.neuroimage.2018.02.025.
37. Dey D, Gaur S, Ovrehus KA, Slomka PJ, Betancur J, Goeller M, et al. Integrated prediction of lesion-specific ischaemia from quantitative coronary CT angiography using machine learning: a multicentre study. Eur Radiol. 2018; 28(6):2655-2664. DOI: 10.1007/s00330-017-5223-z.
38. Johnson C, Price G, Khalifa J, Faivre-Finn C, Dekker A, Moore C, et al. A method to combine target volume data from 3D and 4D planned thoracic radiotherapy patient cohorts for machine learning applications. Radiother Oncol. 2018; 126(2):355-361. DOI: 10.1016/j.radonc.2017.11.015.
39. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. Circ Res. 2017; 121(9):1092-1101. DOI: 10.1161/CIRCRESAHA.117.311312.
40. van der Velden BHM, Janse MHA, Ragusi MAA, Loo CE, Gilhuijs KGA. Volumetric breast density estimation on MRI using explainable deep learning regression. Sci Rep. 2020; 10(1):18095. DOI: 10.1038/s41598-020-75167-6.
41. Hsu YC, Tsai YH, Weng HH, Hsu LS, Tsai YH, Lin YC, et al. Artificial neural networks improve LDCT lung cancer screening: a comparative validation study. BMC Cancer. 2020; 20(1):1023. DOI: 10.1186/s12885-020-07465-1.
42. Kwon G, Ryu J, Oh J, Lim J, Kang BK, Ahn C, et al. Deep learning algorithms for detecting and visualising intussusception on plain abdominal radiography in children: a retrospective multicenter study. Sci Rep. 2020; 10(1):17582. DOI: 10.1038/s41598-020-74653-1.
43. Fu F, Wei J, Zhang M, Yu F, Xiao Y, Rong D, et al. Rapid vessel segmentation and reconstruction of head and neck angiograms using 3D convolutional neural network. Nat Commun. 2020; 11(1):4829. DOI: 10.1038/s41467-020-18606-2.
44. Winther H, Hundt C, Ringe KI, Wacker FK, Schmidt B, Jürgens J, et al. A 3D deep neural network for liver volumetry in 3T contrast-enhanced MRI. Rofo. 2020; (Online ahead of print). DOI: 10.1055/a-1238-2887.
45. Shao L, Yan Y, Liu Z, Ye X, Xia H, Zhu X, et al. Radiologist-like artificial intelligence for grade group prediction of radical prostatectomy for reducing upgrading and downgrading from biopsy. Theranostics. 2020; 10(22):10200-10212. DOI: 10.7150/thno.48706.
46. Lång K, Dustler M, Dahlblom V, Åkesson A, Andersson I, Zackrisson S. Identifying normal mammograms in a large screening population using artificial intelligence. Eur Radiol. 2020; (Online ahead of print). DOI: 10.1007/s00330-020-07165-1.
47. Olive-Gadea M, Crespo C, Granes C, Hernandez-Perez M, Pérez de la Ossa N, Laredo C, et al. Deep learning based software to identify large vessel occlusion on noncontrast computed tomography. Stroke. 2020; 51(10):3133-3137. DOI: 10.1161/STROKEAHA.120.030326.
48. Zhang C, Zhao J, Niu J, Li D. New convolutional neural network model for screening and diagnosis of mammograms. PLoS One. 2020; 15(8):e0237674. DOI: 10.1371/journal.pone.0237674.
49. Kim JH, Kim JY, Kim GH, Kang D, Kim IJ, Seo J, et al. Clinical validation of a deep learning algorithm for detection of pneumonia on chest radiographs in emergency department patients with acute febrile respiratory illness. J Clin Med. 2020; 9(6):1981. DOI: 10.3390/jcm9061981.
50. Bhat CS, Chopra M, Andronikou S, Paul S, Wener-Fligner Z, Merkoulovitch A, et al. Artificial intelligence for interpretation of segments of whole body MRI in CNO: pilot study comparing radiologists versus machine learning algorithm. Pediatr Rheumatol Online J. 2020; 18(1):47. DOI: 10.1186/s12969-020-00442-9.
51. Kim MS, Park HY, Kho BG, Park CK, Oh IJ, Kim YC, et al. Artificial intelligence and lung cancer treatment decision: agreement with recommendation of multidisciplinary tumor board. Transl Lung Cancer Res. 2020; 9(3):507-514. DOI: 10.21037/tlcr.2020.04.11.
52. Drozdov I, Forbes D, Szubert B, Hall M, Carlin C, Lowe DJ. Supervised and unsupervised language modelling in Chest X-Ray radiological reports. PLoS One. 2020; 15(3):e0229963. DOI: 10.1371/journal.pone.0229963.
53. Kann BH, Hicks DF, Payabvash S, Mahajan A, Du J, Gupta V, et al. Multi-institutional validation of deep learning for pretreatment identification of extranodal extension in head and neck squamous cell carcinoma. J Clin Oncol. 2020; 38(12):1304-1311. DOI: 10.1200/JCO.19.02031.
54. Mall S, Brennan PC, Mello-Thoms C. Can a machine learn from radiologists' visual search behaviour and their interpretation of mammograms - a deep-learning study. J Digit Imaging. 2019; 32(5):746-760. DOI: 10.1007/s10278-018-00174-z.
55. Seidler M, Forghani B, Reinhold C, Pérez-Lara A, Romero-Sanchez G, Muthukrishnan N, et al. Dual-energy CT texture analysis with machine learning for the evaluation and characterization of cervical lymphadenopathy. Comput Struct Biotechnol J. 2019; 17:1009-1015. DOI: 10.1016/j.csbj.2019.07.004.
56. Mohammadi S, Mohammadi M, Dehlaghi V, Ahmadi A. Automatic segmentation, detection, and diagnosis of abdominal aortic aneurysm (AAA) using convolutional neural networks and hough circles algorithm. Cardiovasc Eng Technol. 2019; 10(3):490-499. DOI: 10.1007/s13239-019-00421-6.
57. Akselrod-Ballin A, Chorev M, Shoshan Y, Spiro A, Hazan A, Melamed R, et al. Predicting breast cancer by applying deep learning to linked health records and mammograms. Radiology. 2019; 292(2):331-342. DOI: 10.1148/radiol.2019182622.
58. Yanagawa M, Niioka H, Hata A, Kikuchi N, Honda O, Kurakami H, et al. Application of deep learning (3-dimensional convolutional neural network) for the prediction of pathological invasiveness in lung adenocarcinoma: A preliminary study. Medicine (Baltimore). 2019; 98(25):e16119. DOI: 10.1097/MD.0000000000016119.
59. Gan K, Xu D, Lin Y, Shen Y, Zhang T, Hu K, et al. Artificial intelligence detection of distal radi-

us fractures: a comparison between the convolutional neural network and professional assessments. Acta Orthop. 2019; 90(4):394-400. DOI: 10.1080/17453674.2019.1600125.
60. Lassau N, Estienne T, de Vomecourt P, Azoulay M, Cagnol J, Garcia G, et al. Five simultaneous artificial intelligence data challenges on ultrasound, CT, and MRI. Diagn Interv Imaging. 2019; 100(4):199-209. DOI: 10.1016/j.diii.2019.02.001.
61. Tran GS, Nghiem TP, Nguyen VT, Luong CM, Burie JC. Improving accuracy of lung nodule classification using deep learning with focal loss. J Healthc Eng. 2019; 2019:5156416. DOI: 10.1155/2019/5156416.
62. Ko SY, Lee JH, Yoon JH, Na H, Hong E, Han K, et al. Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. Head Neck. 2019; 41(4):885-891. DOI: 10.1002/hed.25415.
63. Li S, Xiao J, He L, Peng X, Yuan X. The tumor target segmentation of nasopharyngeal cancer in CT images based on deep learning methods. Technol Cancer Res Treat. 2019; 18:1533033819884561. DOI: 10.1177/1533033819884561.
64. Hae H, Kang SJ, Kim WJ, Choi SY, Lee JG, Bae Y, et al. Machine learning assessment of myocardial ischemia using angiography: Development and retrospective validation. PLoS Med. 2018; 15(11):e1002693. DOI: 10.1371/journal.pmed.1002693.
65. Adams M, Chen W, Holcdorf D, McCusker MW, Howe PD, Gaillard F. Computer vs human: Deep learning versus perceptual training for the detection of neck of femur fractures. J Med Imaging Radiat Oncol. 2019; 63(1):27-32. DOI: 10.1111/1754-9485.12828.
66. Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, et al. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet. 2018; 392(10162):2388-2396. DOI: 10.1016/S0140-6736(18)31645-3.
67. Sumathipala Y, Lay N, Turkbey B, Smith C, Choyke PL, Summers RM. Prostate cancer detection from multi-institution multiparametric MRIs using deep convolutional neural networks. J Med Imaging (Bellingham). 2018; 5(4):044507. DOI: 10.1117/1.JMI.5.4.044507.
68. Tong N, Gou S, Yang S, Ruan D, Sheng K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. Med Phys. 2018; 45(10):4558-4567. DOI: 10.1002/mp.13147.
69. Kim T, Heo J, Jang DK, Sunwoo L, Kim J, Lee KJ, et al. Machine learning for detecting moyamoya disease in plain skull radiography using a convolutional neural network. EBioMedicine. 2019; 40:636-642. DOI: 10.1016/j.ebiom.2018.12.043.
70. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009; 6(7): e1000097. DOI:10.1371/journal.pmed1000097
71. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33 (1):159–174. DOI:10.2307/2529310.
72. McKinney MS, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature 2020; 577:89-94. DOI: 10.1038/s41586-019-1799-6
73. Zbontar J, Knoll F, Sriram A, Muckley M, Bruno M, Defazio A, Parente M, et al. fastMRI: an open dataset and benchmarks for accelerated MRI. arXiv:1811.08839v2
74. Benaich N, Hogarth I. https://www.stateof.ai access: 19 Nov. 2020
75. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Massive Analysis Quality Control (MAQC) Society Board of Directors, Waldron L, et al. Transparency and reproducibility in artificial intelligence. Nature 2020; 586:14-16. DOI: 10.1038/s41586-020-2766-y