



Artificial Intelligence-Powered Clinical Decision Support and Simulation Platform for Radiology Trainee Education

Chintan Shah¹ · Karapet Davtyan² · Ilya Nasrallah³ · R Nick Bryan³ · Suyash Mohan³

Received: 21 April 2022 / Revised: 1 September 2022 / Accepted: 5 October 2022 / Published online: 24 October 2022
© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2022

Abstract

Technological tools can redesign traditional approaches to radiology education, for example, with simulation cases and via computer-generated feedback. In this study, we investigated the use of an AI-powered, Bayesian inference-based clinical decision support (CDS) software to provide automated “real-time” feedback to trainees during interpretation of clinical and simulation brain MRI examinations. Radiology trainees participated in sessions in which they interpreted 3 brain MRIs: two cases from a routine clinical worklist (one without and one with CDS) and a teaching file-based simulation case with CDS. The CDS software required trainees to input imaging features and differential diagnoses, after which inferred diagnoses were displayed, and the case was reviewed with an attending neuroradiologist. An observer timed each case, including time spent on education, and trainees completed a survey rating their confidence in their findings and the educational value of the case. Ten trainees reviewed 75 brain MRI examinations during 25 reading sessions. Trainees had slightly lower confidence in their findings and diagnosis and rated the educational value slightly higher for simulation cases with CDS compared to clinical cases without CDS ($p < 0.05$). There were no significant differences in ratings of clinical cases with or without CDS. No differences in overall timing were found among the reading scenarios. Simulation cases with “CDS-provided feedback” may improve the educational value of interpreting imaging studies at a workstation without adding additional time. Further investigation will help drive innovation in trainee education, which may be particularly relevant in this era of increasing remote work and asynchronous attending review.

Keywords Artificial intelligence · Bayesian networks · Brain MRI · Education · Simulation

Introduction

Traditional radiology trainee education relies heavily on interpretation of live clinical cases in an apprenticeship-type model [1, 2]. Residents and fellows typically review and generate reports for imaging exams obtained on current outpatients, inpatients, or emergency room patients that have recently been scanned, and each exam is subsequently reviewed by an attending radiologist. Depending on the setting, the attending may review each case with the trainee

at the PACS workstation, independently review cases and provide in-person feedback, or, as is increasingly the case, remotely review images synchronously or asynchronously, and intermittently provide feedback via remote systems such as chat, screen sharing, screen captures, or simply report edits/addenda. This is supplemented by didactic or case-based conferences.

Although innovative approaches are improving the conference-based aspect of learning [3, 4], the larger practical component of this training has not changed considerably for decades. While the apprenticeship model offers many distinct advantages, such as real time-decision-making, reporting, and clinical correlation, there are limitations as well. Depending on the practice setting (i.e., community hospital, quaternary care specialty hospital, trauma center, cancer center, etc.), some pathologies may be over or under-represented for a well-rounded training. Based on a trainee’s individual random experience, some categories of disease may also be disproportionately represented. Further, a large

✉ Chintan Shah
shahc2@ccf.org

¹ Department of Radiology, Imaging Institute, Cleveland Clinic, 9500 Euclid Ave, Mail code S3, Cleveland, OH, USA

² University of Texas Medical Branch, Galveston, TX, USA

³ Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA

share of a typical clinical workload involves clinical follow-up for known lesions and relatively fewer opportunities for characterization of new lesions or disease. This challenge is often addressed via case conferences and teaching files [2], but such a setting loses the advantages of the clinical environment, and cannot be easily tailored for each trainee. Second, there is a high degree of variability in feedback and teaching from attendings, which may be due to the clinical case itself, the demands of the clinical service, and/or the aptitude and teaching skills of the attending radiologist.

Technological tools have considerable potential to redesign this approach [5], including artificial intelligence (AI)-based systems. For example, simulation cases can be assigned to individual trainees in order to address gaps in case mix [1, 2, 5]. This would allow a trainee to experience cases that they may have less experience in, but to do so in a “clinical-type environment,” allowing full interrogation of the images and generation of a typical report rather than viewing selected images in a conference. Additionally, AI applications are being rapidly developed for detection of findings, diagnosis of disease [6], and triage of examinations. These can theoretically provide automated feedback to a trainee, augmenting traditional feedback mechanisms and reducing variability.

In this study, we investigated the use of AI-based clinical decision support (CDS) software to provide automated

“real-time” feedback to trainees during interpretation of clinical and teaching file (TF)-based simulation brain MRI examinations. We hypothesized that trainees would rate simulation and CDS cases as having greater educational value, and that utilizing CDS would not require more time.

Materials and Methods

Neuroradiology fellows and diagnostic radiology residents were enrolled to participate in this IRB-approved study. Informed consent was obtained from all individual participants included in the study.

Trainees were asked to interpret brain MRI examinations during reading sessions. Each session consisted of 3 brain MRIs: two live clinical cases (one without and one with CDS) and a TF simulation case with CDS. We used a Bayesian inference-based CDS system for probabilistic diagnosis of lesions at brain MRI (Galileo CDS, Austin, TX) [6]. The education-focused interface for this application requires trainees to input imaging features and their top three differential diagnostic considerations for each case (Fig. 1). The imaging features which trainees were asked to input included signal information (appearance on T1, T2, FLAIR, and diffusion-weighted imaging and presence of contrast enhancement or susceptibility) and spatial information

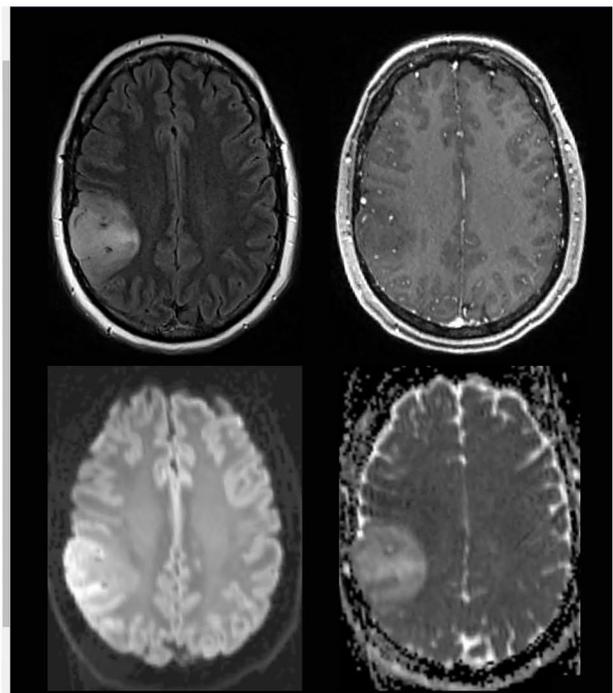
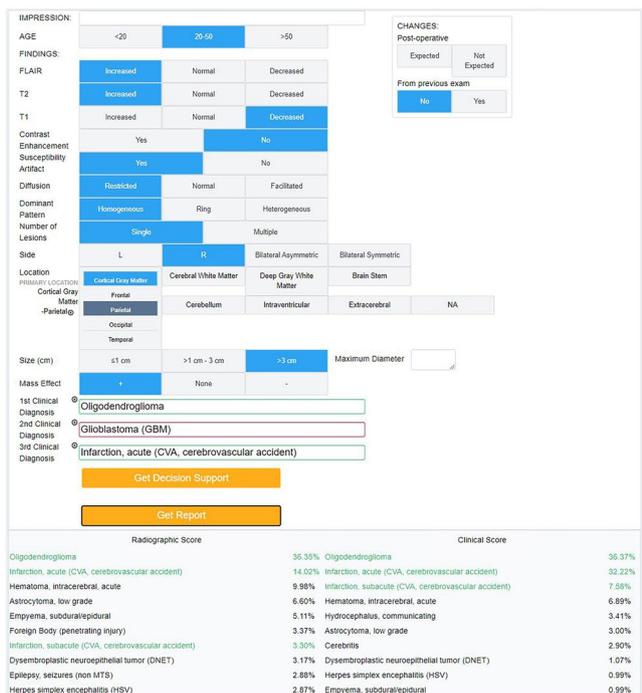


Fig. 1 Clinical decision support (CDS) system with simulation case. **A** CDS Interface for education, requiring selection of imaging features and entry of differential diagnoses. Computer-generated probabilistic differential diagnoses are displayed at the bottom. **B** Example

teaching file-based simulation case demonstrating an infiltrative non-enhancing FLAIR-hyperintense lesion in the right parietal lobe with areas of mild restricted diffusion

(configuration (homogeneous/heterogeneous/ring), number of lesions (single/multiple), laterality, general location (cortex, white matter, deep gray, brainstem, cerebellar, intraventricular, extracerebral), specific location (i.e., lobes), size, and mass effect), as depicted in Fig. 1. The software then displays inferred diagnoses, with probabilities based on imaging features alone (referred to as “radiographic score” by the application), or also incorporating pre-test probability based on prevalence information and age (referred to as “clinical score” by the application). Each disease entity is listed with hyperlinks to educational resources. The differential possibilities can then be adjusted as deemed appropriate. The software then generates a report including a description based on the imaging features and the top diagnoses, which is exported to the reporting software.

Live clinical cases were chosen as the first available outpatient brain MRI from specified worklists in the radiology information system. These could be either unenhanced or gadolinium-enhanced examinations and were required to be single-study brain MRI examinations (i.e., exams with an associated additional study such as an MRI of the spine or MR angiography were excluded, as were focused examinations such as MRI pituitary). TF cases used for simulation cases were chosen at random from an anonymized TF with established diagnoses based on clinical or pathological follow-up; these were complete exams including all images from all acquired sequences and comprised a subset of the cases originally used for training of the Bayesian network for the CDS. Cases were considered to be of low complexity if they were normal or negative or only had white matter disease (i.e., small vessel ischemic disease, migraine, demyelinating disease, etc.) and to be of higher complexity for any other pathology (i.e., sub-acute infarct, glioblastoma, pleomorphic xanthoastrocytoma, etc.).

For all types of cases, trainees (1) reviewed images in PACS, (2) used CDS (if applicable), (3) performed clinical correlation as needed with the electronic medical record or online resources, (4) dictated and drafted a report, and (5) reviewed the exam, including images and draft report, with an attending neuroradiologist. An observer timed all parts of this process using a multi-channel stopwatch and observed interaction with the software, including whether trainees altered their own top 3 diagnoses after seeing the CDS results and whether trainees utilized the educational links. For simulation cases, the observer also provided the established diagnosis after attending review. Trainees completed a survey after each case, rating their confidence in their findings and the educational value of the case on a 5-point scale, with 1 as high and 5 as low, and also an overall survey at the completion of the reading session. Trainee ratings for each type of case were compared using pairwise Mann–Whitney *U* tests. Timings for each type of case were compared using pairwise 2 sample *t*-tests.

Finally, we performed a post hoc analysis of the trainee performance on identifying key features. For the TF and clinical cases with CDS in which the copies of the original reports from the CDS system were available, we reviewed the reports to determine the features chosen by the trainees for the six “signal information” selections described above. The images were reviewed in PACS by an attending neuroradiologist to determine the true key features. The trainee performance in identifying these was then compared between the Clinical and TF cases with CDS, using a Mann–Whitney *U* test.

Results

75 MRI examinations were reviewed during 25 reading sessions by 9 neuroradiology fellows (postgraduate year (PGY 6) and 1 radiology resident (PGY 3)). Sessions were completed in the second quarter or early third quarter of the academic year. Of the live clinical cases interpreted without CDS, clinical cases interpreted with CDS, and the TF cases interpreted with CDS, 56%, 56%, and 84% were considered of high complexity, respectively.

Confidence and Educational Value Ratings

Trainees had slightly lower confidence in their findings for TF cases with CDS compared to clinical cases (with or without CDS) and had slightly lower confidence in their diagnosis for TF cases with CDS compared to clinical cases without CDS ($p < 0.05$, Table 1). TF cases with CDS were also rated as having greater value for identification of topics for further reading and greater overall educational value ($p < 0.05$, Table 1), compared to clinical cases without CDS. There were no significant differences in ratings of clinical cases with or without CDS.

Timing Results

No differences in overall timing were found between case types. The overall mean (\pm SD) time spent on a case by a trainee was 18.2 (\pm 10.0) min. TF cases had significantly lower time required for clinical correlation ($p < 0.05$, Table 2).

Observer Assessment and Overall Session Survey

The observer noted that trainees adjusted their top 3 considerations in the CDS interface in 44% of cases for both clinical cases and TF cases read with CDS and utilized the educational links in 4% of cases for each type of case.

When asked whether CDS improved the educational experience of reading a case, trainees indicated an average

Table 1 Trainee ratings of confidence and educational value by case type

	Confidence in lesion identification 1 = high; 5 = low	Confidence in differential diagnosis 1 = high; 5 = low	Usefulness to identify questions for attending 1 = high; 5 = low	Usefulness to identify topics for further reading 1 = high; 5 = low	Overall educational value 1 = high; 5 = low
Clinical only	1.48 ± 0.59	1.44 ± 0.65	2.28 ± 0.89	2.4 ± 0.91	2.2 ± 0.65
Clinical + CDS	1.64 ± 0.91	1.88 ± 1.01	2.08 ± 0.64	2.16 ± 0.62	2.08 ± 0.57
TF + CDS	2.2 ± 1.04^{*,**}	2.16 ± 1.07[*]	1.84 ± 0.69	1.8 ± 0.82[*]	1.8 ± 0.71[*]

^{*} $p < 0.05$ (TF + CDS vs. clinical only), Mann Whitney U test

^{**} $p < 0.05$ (TF + CDS vs. clinical + CDS), Mann Whitney U test

rating of 2.2 on a 5-point scale, with 1 being significantly improved and 5 being significantly worsened. On a scale of 1 (frequently) to 5 (never), on average trainees reported that they uncommonly to rarely used the educational links (average rating 3.6/5). Trainees were neutral on the usefulness of the report generation (average rating 2.8/5) and felt that the impact of CDS on their workflow was small or neutral (average rating 2.6/5).

Key Feature Identification

Key feature selection information was available for 19 TF cases and 19 clinical cases with CDS. For the six “signal information” key features, trainees identified most features correctly for most cases, with an average of 5.37 out of 6 in the TF cases and 5.74 in the clinical cases with CDS (median of 6 in both groups). There was no significant difference between the two case types ($p = 0.25$).

Discussion

In this study, we found that TF simulation cases interpreted with CDS were rated as having a higher educational value by trainees, compared to live clinical cases interpreted without CDS. Trainees also found greater difficulty with these cases, noting lower confidence in their findings and diagnosis. This suggests simulation cases with automated feedback can play an important role in improving the educational experience and can target areas in which trainees are weaker or have

less experience. The differences may have been in part due to 84% of TF cases being in the higher complexity category, compared to 56% of the clinical cases. The simulation cases offered more opportunity for assessment of novel and less-common lesions, which is primarily what CDS can be helpful for, both in terms of educational feedback and clinical utility. There was no observed difference in educational value when interpreting clinical cases with or without CDS, although this may be due to the relatively small sample size, in particular with a smaller proportion of novel and less-common lesions. In the cohort of outpatient examinations chosen for the live clinical cases, even in the higher complexity category, many were for follow up of known pathology, a setting in which CDS would have lower utility compared to examinations with newly diagnosed pathology.

Simulation is not new in radiology training and education [1, 2]. However, most commonly simulation focuses on clinical scenarios such as contrast reaction management [7] and training of procedural skills such as those for image-guided procedures [8–10]. In terms of diagnostic imaging, some simulation has focused on improving acquisition in operator-dependent settings such as fluoroscopic upper gastrointestinal exams [11]. Simulation for education in the interpretive side of radiology is less common but an emerging field. Interactive simulation software has been used to improve the educational experience of medical students during a radiology rotation by allowing them to interact with cases rather than simply observe [12]. Additional examples of applications include a simulation software integrated within PACS developed to help trainees prepare for independent call [13],

Table 2 Average time for image interpretation activities by case type

Mean ± SD	Observation	Clinical correlation	Education	Transcription	Other	Total
Clinical only	7.63 ± 3.57	1.47 ± 2.13	4.25 ± 3.71	5.14 ± 5.92	0.67 ± 1.42	18.73 ± 10.8
Clinical + CDS	7.52 ± 5.06	0.93 ± 1.36	5.06 ± 3.09	5.36 ± 3.76	1.62 ± 3.87	20.31 ± 11.19
TF + CDS	6.44 ± 4.66	0.26 ± 0.76[*]	4.53 ± 3.47	4.06 ± 2.52	0.44 ± 0.83	15.46 ± 7.16
Overall	7.20 ± 4.45	0.88 ± 1.59	4.61 ± 3.40	4.85 ± 4.28	0.91 ± 2.45	18.17 ± 9.96

^{*} $p < 0.05$ (TF + CDS vs. clinical only), 2 sample t -test

simulation software designed to assess trainee performance and readiness for call [14, 15], and simulation cases to improve detection of pulmonary embolism (PE) by trainees [16]. In a similar vein, our results also indicate that educational experience can be enhanced with simulation software. However, as opposed to simulating only the image viewing experience and subsequently asking targeted quiz questions or about the presence or absence of specific pathology such as PE, our efforts were aimed at “high-fidelity” simulation including generation of a “clinical-type radiology report” including “findings” and “impression.” This approach encourages not only identification and characterization of a lesion, but also other clinically relevant practices such as identification of complications, mass effect, and important regions of involvement. Furthermore, our implementation is unique in that it applies automated computer-generated feedback to simulation, whereby a software algorithm provides feedback as to what differential diagnoses should be considered based on image findings and pre-test probability alone, regardless of the established diagnosis. This is an important real-life consideration, which is sometimes overlooked upon presentation of TF examples of rare pathologies—a radiologist must often consider multiple potential diagnoses. However, this should be considered an augmentation of traditional learning methods and not a replacement for case review with an attending radiologist, given that the AI tool focuses on diagnostic considerations, whereas the attending radiologist can provide richer clinical context and teaching regarding important interpretative considerations beyond lesion characterization.

This work has several limitations. The participants could not be blinded to the reporting mechanism and to whether a case was a simulation vs. clinical case, which could have introduced bias into the results. The sample size of the study is also relatively small, which may have reduced power. We did not validate or provide automated feedback regarding the key features input by trainees, which could result in poor differential considerations presented to trainees. The automated feedback was focused on the differential diagnoses, whereas feedback regarding imaging features was provided during attending review, which remains an integral component of radiology education. However, in a post hoc analysis, we found that trainees identified the key features correctly most of the time for most cases, suggesting that these are lower-level observational skills more quickly learned by trainees and that incorrect key features did not impair the CDS-provided diagnostic considerations in the majority of cases. Although we might expect that the trainees would have more difficulty with the key features in the TF cases based on the survey responses, the study was not powered to detect such a small difference, particularly in this post hoc analysis. A larger difference might also be expected for a higher-level task such as differential diagnosis. Future

advances in this software for automatic key feature extraction from the images are under development, which would allow automated feedback on the inputs as well.

Participants’ behavior may have been changed by the presence of an observer timing them, and their actions such as the time spent on image review, and interacting with the software and educational links may have been different if there was no perceived time pressure. Finally, the live clinical setting was limited to outpatient exams for logistical purposes, but CDS may have different utility in inpatient/emergency setting, where types and presentations of pathology may be different.

To this end, future work would benefit from inclusion of inpatient/emergency cases to allow greater frequency of new pathology and inclusion of simulation TF cases that require assessment of new lesions without CDS. In the future of radiology education, incorporation of simulation training and automated feedback alongside traditional teaching methods would allow targeted cases to be presented to trainees based on their own needs and allow a consistent mechanism of immediate feedback to augment traditional feedback.

Conclusion

Teaching file-based simulation cases with “CDS-provided feedback” may improve the educational value of interpreting imaging studies at a workstation without adding additional time. Further investigation will help drive innovation in trainee education, which may be particularly relevant in this COVID-19 era with physical distancing.

Author Contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by CS, KD, and SM. Supervision was provided by IMN, RNB, and SM. The first draft of the manuscript was written by CS, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported in part by research funding from Galileo CDS, Inc.

Declarations

Ethics Approval The study was approved by the Institutional Review Board at the University of Pennsylvania.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Competing Interests Author RNB serves as founder, Chairman, and Chief Scientific Officer for Galileo CDS. Author SM has grant funding from Galileo CDS, Inc., Novocure, Inc., and NIH/NCI and serves as consultant for Northwest Biotherapeutics, AI Integrated Radiological Solutions Medical, and Qynapse SAS. The remaining authors have no relevant financial or non-financial interests to disclose.

References

1. S. H. Sabir, S. Aran, and H. Abujudeh, "Simulation-based training in radiology.," *J. Am. Coll. Radiol.*, vol. 11, no. 5, pp. 512–7, May 2014, <https://doi.org/10.1016/j.jacr.2013.02.008>.
2. A. L. Chetlen *et al.*, "Conventional Medical Education and the History of Simulation in Radiology," *Acad. Radiol.*, vol. 22, no. 10, pp. 1252–1267, Oct. 2015, <https://doi.org/10.1016/J.ACRA.2015.07.003>.
3. O. A. Awan, F. Shaikh, B. Kalbfleisch, E. L. Siegel, and P. Chang, "RSNA Diagnosis Live: A Novel Web-based Audience Response Tool to Promote Evidence-based Learning.," *Radiographics*, vol. 37, no. 4, pp. 1111–1118, Jul. 2017, <https://doi.org/10.1148/rg.2017160107>.
4. E. I. Rubio, M. J. Bassignani, M. A. White, and W. E. Brant, "Effect of an audience response system on resident learning and retention of lecture material," *Am. J. Roentgenol.*, vol. 190, no. 6, Jun. 2008, <https://doi.org/10.2214/AJR.07.3038>.
5. M. T. Duong *et al.*, "Artificial intelligence for precision education in radiology.," *Br. J. Radiol.*, vol. 92, no. 1103, p. 20190389, Nov. 2019, <https://doi.org/10.1259/bjr.20190389>.
6. A. M. Rauschecker *et al.*, "Artificial intelligence system approaching neuroradiologist-level differential diagnosis accuracy at brain MRI," *Radiology*, vol. 295, no. 3, pp. 626–637, Jun. 2020, <https://doi.org/10.1148/radiol.2020190283>.
7. K. Pippin *et al.*, "Implementing Contrast Reaction Management Training for Residents Through High-Fidelity Simulation," *Acad. Radiol.*, vol. 26, no. 1, pp. 118–129, Jan. 2019, <https://doi.org/10.1016/J.ACRA.2018.06.006>.
8. A. R. Faulkner, A. C. Bourgeois, Y. C. Bradley, K. B. Hudson, R. E. Heidel, and A. S. Pasciak, "Simulation-based educational curriculum for fluoroscopically guided lumbar puncture improves operator confidence and reduces patient dose.," *Acad. Radiol.*, vol. 22, no. 5, pp. 668–73, May 2015, <https://doi.org/10.1016/j.acra.2014.12.024>.
9. N. Fulton, J. Buehe, J. Gollamudi, and M. Robbin, "Simulation-Based Training May Improve Resident Skill in Ultrasound-Guided Biopsy," vol. 207, no. 6, pp. 1329–1333, Sep. 2016, <https://doi.org/10.2214/AJR.16.16161>.
10. M. Mendiratta-Lala, T. R. Williams, V. Mendiratta, H. Ahmed, and J. W. Bonnett, "Simulation Center Training as a Means to Improve Resident Performance in Percutaneous Noncontinuous CT-Guided Fluoroscopic Procedures With Dose Reduction," , vol. 204, no. 4, pp. W376–W383, Mar. 2015, <https://doi.org/10.2214/AJR.14.13420>.
11. E. C. Benya, M. R. Wyers, E. K. O'Brien, V. Nandhan, and M. D. Adler, "Upper gastrointestinal fluoroscopic simulator for neonates with bilious emesis.," *Pediatr. Radiol.*, vol. 45, no. 9, pp. 1413–6, Aug. 2015, <https://doi.org/10.1007/s00247-015-3330-z>.
12. M. V. Friedman, J. L. Demertzis, T. J. Hillen, J. R. Long, and D. A. Rubin, "Impact of an Interactive Diagnostic Case Simulator on a Medical Student Radiology Rotation," vol. 208, no. 6, pp. 1256–1261, Mar. 2017, <https://doi.org/10.2214/AJR.16.17537>.
13. A. J. Towbin, B. E. Paterson, and P. J. Chang, "Computer-based Simulator for Radiology: An Educational Tool," vol. 28, no. 1, pp. 309–316, Jan. 2008, doi: <https://doi.org/10.1148/RG.281075051>.
14. P. A. Gondim Teixeira *et al.*, "Radiology resident MR and CT image analysis skill assessment using an interactive volumetric simulation tool - the RadioLOG project.," *Eur. Radiol.*, vol. 27, no. 2, pp. 878–887, Feb. 2017, <https://doi.org/10.1007/s00330-016-4384-5>.
15. S. Ganguli, M. Camacho, C. S. Yam, and I. Pedrosa, "Preparing first-year radiology residents and assessing their readiness for on-call responsibilities: Results over 5 years," *Am. J. Roentgenol.*, vol. 192, no. 2, pp. 539–544, Feb. 2009, <https://doi.org/10.2214/AJR.08.1631>.
16. J. Williams, T. S. Sato, and B. Policeni, "Pulmonary Embolism Teaching File: A Simple Pilot Study for Rapidly Increasing Pulmonary Embolism Recognition among New Residents Using Interactive Cross-sectional Imaging," *Acad. Radiol.*, vol. 20, no. 8, pp. 1048–1051, Aug. 2013, <https://doi.org/10.1016/J.ACRA.2012.12.020>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.