Artificial intelligence for ultrasonography: unique opportunities and challenges

SEONG HO PARK

Guest Editor
ULTRASONOGRAPHY

The application of artificial intelligence (AI) technology to medical imaging has recently brought about tremendous excitement, and AI is making its way into clinical practice, thanks to the technical prowess of current deep learning technology compared with the machine learning methods of the past, the wide availability of digital medical images, and the increased capabilities of computing hardware [1–4]. AI has been tried for ultrasonography in various organs and systems, such as the thyroid, musculoskeletal system, breast, and abdomen, as discussed in detail in the focused review articles of this special issue [5–8], albeit not as extensively as some other radiological imaging modalities such as chest X-rays [9]. The potential role of AI is anticipated to enhance the quality of ultrasonographic images, to provide various forms of diagnostic support (e.g., automated characterization of findings on ultrasonographic images; extraction of quantitative or predictive information from ultrasonographic images, which is difficult for a human examiner to do based on visual observations; and automated detection or segmentation of various structures on ultrasonographic images), and to improve workflow efficiency [10]. The list of specific examples of AI applications to ultrasonography is expected to grow in the future.

AI algorithms may augment the diagnostic accuracy and capability of ultrasonography examiners and are hoped to be particularly helpful for less-experienced examiners [11–15]. Ultrasonography is more widely used in clinical practice than computed tomography (CT) or magnetic resonance imaging (MRI), and it is performed by a more diverse range of medical professionals with varying levels of expertise, some of whom perform better than others. Typically, a single examiner interprets the findings and makes decisions on the fly while performing the examination. As a result, the greater operator-dependency and subjectivity of ultrasonography compared with CT or MRI are well-known issues. Therefore, one of the most eagerly anticipated benefits of applying AI in ultrasonography would be reduced variability between examiners. In this regard, AI may offer a unique opportunity to improve the performance of ultrasonography by removing variability between examiners. Nonetheless, it should be noted that the very nature of ultrasonography also poses challenges in the development and clinical implementation of AI for ultrasonography.

First, the operator-dependency and subjectivity of ultrasonography introduce additional variability

Seong Ho Park, MD, PhD

Department of Radiology and Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea   Tel. +82-2-3010-5984, Fax. +82-2-476-4719, E-mail: seongho@amc.seoul.kr

in the acquisition of imaging data. These factors could exacerbate the limited generalizability of current AI systems built with deep learning [16]. The finally obtained ultrasonographic images are determined by how the examiner captures them. Thus, the results of AI depend on how the target structure is represented and defined by the examiner in the captured image [17] and, furthermore, by whether the target object is correctly identified and captured at all, unless an entire 3-dimensional volume scan is used, such as those obtained using automated breast ultrasound systems. For the same reason, considerable discrepancies may exist between the dataset collected to train an AI algorithm and the imaging data generated in real-world practice to be fed into the AI system. Therefore, even for a highly sophisticated AI system to work correctly, some degree of competency of the human examiner, at least sufficing to scan the patient properly, still matters [17]. Moreover, standardization of scanning and image acquisition, depending on the diagnostic task, would be critical for the successful application of AI to ultrasonography, which requires human expertise. In some sense, the successful application of AI to ultrasonography creates an impetus for standardizing and ensuring the quality of examinations performed by humans.

Second, the more widespread use of ultrasonography in clinical practice and its relatively easy accessibility require extra caution when interpreting the results of AI used with ultrasonography. The results given by AI, which capitalizes on the associations between input features and outcome states, are probabilistic. Therefore, unlike the results provided by tests based on cause-effect relationships, the results of AI algorithms should generally not be regarded as fixed results. A positive result from a test that finds a clear causal determinant to make the diagnosis can be accepted as a fixed result regardless of other factors. An illustrative example is the reverse transcription polymerase chain reaction (RT-PCR) test for severe acute respiratory syndrome coronavirus 2. A positive RT-PCR test result is an immutable proof of the presence of the virus, as this test finds the RNA of the virus, as long as extraordinary cases of residual RNA being detected in convalescent patients are excluded. In contrast, the interpretation of AI results is affected substantially by the pretest probability and the relevant spectrum of disease manifestation [18]. An AI algorithm typically applies a threshold to a probability-like internal raw algorithm output to generate the final categorical result shown to the user (e.g., cancer vs. benign) or may present the raw output in the form of probability (e.g., a 65% probability of cancer). Both the accuracy of the probability scale and the optimal threshold are profoundly affected by the pretest probability and disease manifestation spectrum, which are, in turn, determined by the baseline characteristics of the patient and the clinical setting.

It is critical for AI users to understand that the same AI result could be correct for one patient but not for another, right in one hospital but not in another hospital, and so on, depending on patients' baseline characteristics and the clinical setting. The limited generalizability of AI algorithms for medical diagnosis and prediction (i.e., the substantial variability in AI accuracy across patients and hospitals) is a well-known phenomenon, often described as "overfitting" in a broad sense [2,18–23]. This problem is primarily due to epidemiological factors, as mentioned above (pretest probability and disease manifestation spectrum), or, more simply, a disparity between training data and real-world data, rather than technical/mathematical overfitting [2,18–20]. This pitfall may be especially pronounced for AI algorithms for ultrasonography, as ultrasonography examinations are often used in a wide range of clinical settings and patients, and are performed by a diverse range of medical professionals with varying expertise. Ultrasonography systems are also more diverse, with more vendors and versions, than CT or MRI. While one might expect AI to be more helpful for less-experienced examiners, ironically, less-experienced examiners may be more likely to have difficulties in appraising AI results and more vulnerable to developing a complacent attitude of merely accepting the AI results without the necessary appraisal. Such complacency would ultimately compromise the accuracy of ultrasonography examinations. The fact that ultrasonography is typically performed and interpreted on the fly by a single practitioner may further increase the risk. Consequently, the human expertise of the examiner, including adequate knowledge and experience in ultrasonography examinations, sound clinical and epidemiological knowledge, and ideally some knowledge about AI as well, would be crucial for maximizing the benefits that AI may provide.

The issue of overfitting underscores the importance of an adequate external validation of an AI algorithm in various real-world clinical settings where it is intended to be used [16,18,24–34]. For all the reasons explained above, perhaps, the importance of sufficient external validation should be even more strongly emphasized for AI applications to ultrasonography. A recent systematic review of studies that evaluated AI algorithms for the diagnostic analysis of medical imaging found that only 6% of such studies published in peer-review journals performed some form of external validation (whether they were otherwise methodologically adequate or not) [35]. Future research on AI for ultrasonography should emphasize the external validation of developed algorithms, in addition to the development of novel algorithms. Rigorous external validation helps to clarify the boundaries of when an AI algorithm maintains its anticipated accuracy and when it fails, and can thus help assure the users of conditions where the AI system can be used safely and effectively. Furthermore, establishing a mechanism to

deliver such information to the end-users of AI more effectively and explicitly would also be an important next step [36].

Third, the operator-dependency of ultrasonography makes prospective research studies to validate AI even more essential. The effect of a computerized decision support system such as AI depends on not only its technical analytic capability, but also on how the computerized results are presented to and acted upon by human practitioners. Considering the expected operator-dependency and variability in generating the ultrasonography image data and in acting upon AI results in on-the-fly decision-making during real-time examinations, there could be meaningful differences between an analysis of retrospectively collected images and natural clinical practice. Studies on AI for ultrasonography have so far mostly been retrospective. More prospective studies that involve actual interactions between human examiners and AI systems should be performed.

AI research in healthcare is accelerating rapidly, with numerous potential applications being demonstrated. However, there are currently limited examples of such techniques being successfully deployed in clinical practice [1,16]. The introduction of AI into medicine is just beginning, and there remain multitudes of challenges to overcome, including difficulties in obtaining sufficiently large, curated, high-quality, representative datasets, deficiencies in robust clinical validation, and technical limitations such as the "black box" nature of AI algorithms, to name just a few [1,16,37]. These challenges are all relevant to AI for ultrasonography. This article highlighted a few additional points that are unique to AI as applied to ultrasonography and need to be addressed for the successful development and clinical implementation of AI for ultrasonography. In summary, the nature of how ultrasonography examinations are performed and utilized demands extra attention to the following issues regarding AI for ultrasonography. It is crucial to maintain the human expertise of examiners, in terms of both ultrasonography itself and the related clinical and epidemiological knowledge. Standardization of scanning and image acquisition, depending on the diagnostic tasks that AI is used to perform, is also critical. The importance of sufficient external validation of AI algorithms is especially significant for AI used with ultrasonography. Prospective research studies that involve actual interactions between human examiners and AI systems, rather than analyses of retrospectively collected images, should also be conducted.

ORCID: Seong Ho Park: https://orcid.org/0000-0002-1257-8315

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

1. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44-56.
2. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional neural networks for radiologic images: a radiologist's guide. Radiology 2019;290:590-606.
3. Do S, Song KD, Chung JW. Basics of deep learning: a radiologist's guide to understanding published radiology articles on deep learning. Korean J Radiol 2020;21:33-41.
4. Lee JG, Jun S, Cho YW, Lee H, Kim GB, Seo JB, et al. Deep learning in medical imaging: general overview. Korean J Radiol 2017;18:570-584.
5. Ha EJ, Baek JH. Application of machine learning and deep learning to thyroid imaging: where do we stand? Ultrasonography 2021;40:23-29.
6. Shin Y, Yang J, Lee YH, Kim S. Artificial intelligence in musculoskeletal ultrasound imaging. Ultrasonography 2021;40:30-44.
7. Kim J, Kim HJ, Kim C, Kim WH. Artificial intelligence in breast ultrasound. Ultrasonography 2021 (in press).
8. Song KD. Current status of deep learning applications in abdominal ultrasonography. Ultrasonography 2020 Sep 2 [Epub]. https://doi.org/10.14366/usg.20085.
9. Hwang EJ, Park CM. Clinical implementation of deep learning in thoracic radiology: potential applications and challenges. Korean J Radiol 2020;21:511-525.
10. Yi J, Kang HK, Kwon JH, Kim KS, Park MH, Seong YK, et al. Technology trends and applications of deep learning in ultrasonography: image quality enhancement, diagnostic support, and improving workflow efficiency. Ultrasonography 2021;40:7-22.
11. Choi JS, Han BK, Ko ES, Bae JM, Ko EY, Song SH, et al. Effect of a deep learning framework-based computer-aided diagnosis system on the diagnostic performance of radiologists in differentiating between malignant and benign masses on breast ultrasonography. Korean J Radiol 2019;20:749-758.
12. Zhang J, Wang Y, Yu B, Shi X, Zhang Y. Application of computer-aided diagnosis to the sonographic evaluation of cervical lymph nodes. Ultrason Imaging 2016;38:159-171.
13. Yoo YJ, Ha EJ, Cho YJ, Kim HL, Han M, Kang SY. Computer-aided diagnosis of thyroid nodules via ultrasonography: initial clinical experience. Korean J Radiol 2018;19:665-672.
14. Jeong Y, Kim JH, Chae HD, Park SJ, Bae JS, Joo I, et al. Deep learning-based decision support system for the diagnosis of neoplastic gallbladder polyps on ultrasonography: preliminary results. Sci Rep 2020;10:7700.
15. Wang Y, Choi EJ, Choi Y, Zhang H, Jin GY, Ko SB. Breast cancer classification in automated breast ultrasound using multiview convolutional neural network with transfer learning. Ultrasound Med Biol 2020;46:1119-1132.

16. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019;17:195.

17. Jeong EY, Kim HL, Ha EJ, Park SY, Cho YJ, Han M. Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators. Eur Radiol 2019;29:1978-1985.

18. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology 2018;286:800-809.

19. Park SH, Kim YH, Lee JY, Yoo S, Kim CJ. Ethical challenges regarding artificial intelligence in medicine from the perspective of scientific editing and peer review. Sci Ed 2019;6:91-98.

20. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. PLoS Med 2018;15:e1002683.

21. Lee JH, Joo I, Kang TW, Paik YH, Sinn DH, Ha SY, et al. Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. Eur Radiol 2020;30:1264-1273.

22. Li X, Zhang S, Zhang Q, Wei X, Pan Y, Zhao J, et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. Lancet Oncol 2019;20:193-201.

23. Ting DS, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images From multiethnic populations with diabetes. JAMA 2017;318:2211-2223.

24. Hamon R, Junklewitz H, Sanchez Martin JI. Robustness and explainability of artificial intelligence: from technical to policy solutions. Luxembourg: Publications Office of the European Union, 2020.

25. Mehta MC, Katz IT, Jha AK. Transforming global health with AI. N Engl J Med 2020;382:791-793.

26. Mutasa S, Sun S, Ha R. Understanding artificial intelligence based radiology studies: what is overfitting? Clin Imaging 2020;65:96-99.

27. Bluemke DA, Moy L, Bredella MA, Ertl-Wagner BB, Fowler KJ, Goh VJ, et al. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers-from the Radiology Editorial Board. Radiology 2020;294:487-489.

28. Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. Science 2019;363:810-812.

29. Yu KH, Kohane IS. Framing the challenges of artificial intelligence in medicine. BMJ Qual Saf 2019;28:238-241.

30. Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in health care: how can we know it works? J Am Med Inform Assoc 2019;26:1651-1654.

31. Tang A, Tam R, Cadrin-Chenevert A, Guest W, Chong J, Barfett J, et al. Canadian Association of Radiologists white paper on artificial intelligence in radiology. Can Assoc Radiol J 2018;69:120-135.

32. Park SH, Do KH, Choi JI, Sim JS, Yang DM, Eo H, et al. Principles for evaluating the clinical implementation of novel digital healthcare devices. J Korean Med Assoc 2018;61:765-775.

33. Nevin L; PLOS Medicine Editors. Advancing the beneficial use of machine learning in health care and medicine: toward a community understanding. PLoS Med 2018;15:e1002708.

34. Nsoesie EO. Evaluating artificial intelligence applications in clinical settings. JAMA Netw Open 2018;1:e182658.

35. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers. Korean J Radiol 2019;20:405-410.

36. Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. NPJ Digit Med 2020;3:41.

37. Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing medical imaging data for machine learning. Radiology 2020;295:4-15.