Artificial intelligence in medical ultrasonography: driving on an unpaved road

Young H. Kim

Department of Radiology, University of Massachusetts Medical School, Worcester, MA, USA

ULTRA SONO GRAPHY

Correspondence to:
Young H. Kim, MD, Department of Radiology, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, USA

Tel. +1-508-334-0792
Fax. +1-508-856-4910
E-mail: young.kim@umassmemorial.org

Introduction

Recent advances in deep-learning technology have brought revolutionary changes to artificial intelligence (AI) research and application across industries, yielding major innovations such as facial recognition and self-driving cars. Medicine is no exception, and radiology, which is based on the interpretation of image data obtained through various methods-and has often been compared with computer vision using pattern analysis-is anticipated to experience a major revolution. Despite expectations for increasing research and development of AI-empowered ultrasonography, the clinical implementation of AI in medical ultrasonography faces unique obstacles. It will be necessary to standardize image acquisition, regulate operator and interpreter qualification and performance, integrate clinical information, and provide performance feedback to maximize benefits for patient care.

Fundamental Limitations of Ultrasonographic Image Data

Computed tomography and magnetic resonance imaging generate predictable, reproducible imaging without much dependence on operator skill. However, the quality of ultrasound imaging is significantly dependent on the ability and experience of the person performing ultrasonography. Even the most competent ultrasound examiner may not produce high-quality diagnostic results depending on the patient's body habitus and compliance. Imaging is further dependent upon the angle of the probe meeting the skin, shadowing from superficial structures, and the depth of the target lesion. These limitations, in addition to common variables such as ultrasound machine operating settings and artifacts, make it difficult to apply AI in ultrasound examinations. As a result of these complex interactions, ultrasound image information as input data for AI applications is subject to various random noise from unidentifiable sources.

Standardizing Image Acquisition

In order to overcome these fundamental limitations, standardized protocols for the acquisition of static images and additional cine images are required. Cine images are routinely obtained to complement anatomic variance among patients and reduce the perception gap between the person performing ultrasonography and the person interpreting the images. Unfortunately, cine images often fail to provide a clear depiction of the intended image information since the ideal physical conditions for obtaining optimal images change with dynamic anatomy. Long-range cine swiping results in visually inconclusive features with minimal benefit. The best results of cine swiping can be produced

when applied to a very small or fixed anatomic structure. For example, in echocardiography and obstetric pelvic ultrasonography, where measurement and visual analysis are required, video clips could provide a full set of relevant structured data, allowing spatiotemporal analysis and maximizing the advantages of deep learning. To this end, in February 2020, a cardiac ultrasound software utilizing AI called Caption Guidance received the first Food and Drug Administration (FDA) approval for its algorithm, which can calculate the ejection fraction from the best auto-captured three-dimensional (3D) video clips [1]. The software allows clinicians to easily utilize 3D anatomical information diagnostically by using the AI algorithm to obtain optimal images. In addition, the AI algorithm is capable of recording loops of echo data that allow it to calculate the left ventricular ejection fraction that is in agreement with human experts [1]. Similarly, there have been advances in obstetric and gynecologic research pertaining to ultrasound, including the automatic detection of endometrial thickness and automatic classification of ovarian cysts [2]. The concept of 3D ultrasound acquisition using cine clips can be applied to many complex anatomic areas, including various joint diseases, fetal evaluation, neonatal disease, and binary classification of benign or malignant tumors. From these examples, it seems that software capable of overcoming the fundamental limitations of ultrasound acquisition will have the ability to produce standardized image data using 3D acquisition of focused anatomical areas of interest by limiting noise.

Understanding Deep Learning

Computer-aided diagnosis (CAD) has been utilized in medical imaging, including ultrasonography, for decades. Image information obtained by ultrasonography provides the foundation of input data for the development of a machine learning algorithm, either using traditional handcrafted feature extraction or automated feature extraction via deep learning [3]. As the input data are the results of numerous complex factors, it is difficult to have a well-classified, structured ultrasound image data set. Variables in the ultrasound image acquisition process produce enormous amounts of unintentional image data, making it difficult for either the conventional method or deep learning to create a usable algorithm in practice. The traditional handcrafted feature extraction method is heavily dependent on the domain expert's knowledge, which has a high possibility of bias due to dependence on a single expert and a potentially skewed selection of the best-quality images for the training set. The conventional method is also time-intensive because an expert must annotate images manually. Deep-learning technology with a convolutional neural network is known to have the best performance in image pattern recognition. This method could bypass

the image selection process and extract relevant information from image data through multiple processing layers that human cognitive function cannot recognize [3]. It is uncertain whether large amounts of input data could guarantee good outcomes in the computer analysis of ultrasound imaging due to the fundamental limitations of high variance and high bias in image acquisition. Well-curated input data that have meaningful display objects in the research environment may not be applicable in real-life situations. In addition, the algorithm may not be generalizable for use in different clinical settings and patients with different demographic characteristics from those present at the start of the data gathering process.

Opportunities for Clinical AI Applications

There is substantial research in the literature regarding ultrasound AI applications in areas such as the thyroid, breast, and prostate, which are located on the superficial aspect of the body and are less affected by overlying structures, for the binary task of tumor characterization. Many CAD systems have been developed to improve the diagnostic accuracy of thyroid and breast nodules with a certain degree of success. However, despite the high sensitivity of CAD systems, they have failed due to variable diagnostic performance and low specificity [4]. Recent research has demonstrated improved diagnostic performance for breast cancer using quantitative color Doppler radiomics features [5]. Research has been further expanded to demonstrate the feasibility of neural networks for activity scoring of arthritis with impressive results compared to a human expert [6]. Further research on musculoskeletal disease, including cartilage segmentation, is expected to progress in the future [7]. Better results could be obtained by adding color Doppler or elastography to high-quality grayscale images in the analysis.

Regulation of Operator and Interpreter Qualification and Performance

Despite the necessity for expert image acquisition to aid diagnostic accuracy, operator capabilities have never been examined thoroughly in the literature. Institutional quality assurance is not intended to provide individual performance feedback. Anecdotal malpractice claims are rare and cannot be used as feedback for real improvement of operator skills. Ultrasound images are a calibrated image representation of the operator's knowledge-based perception, which requires a significant learning curve for ultrasound use and interpretation. Performing ultrasonography requires a high level of training that may take years to develop. A well-trained radiologist or sonographer can distinguish between normal and abnormal structures by subjective judgement. Although there may

be some minor variation, the imaging results are generally agreeable among experts. However, even an expert radiologist's interpretation of ultrasound imaging is subject to error owing to an inadequate examination by the sonographer. If the operator misses the lesion when scanning the patient, there is no way that either a human or an AI system can detect the lesion based on the provided images.

Medical ultrasound is increasingly used by non-radiology-trained medical personnel due to the widespread dissemination of affordable portable ultrasound equipment and the need for urgent clinical decisions at the bedside. This leads to ultrasound examinations being performed by non-radiology-trained physicians in the emergency department or intensive care unit without extensive quality assurance or a standardized training system compared to traditional radiology [8]. In most developed countries, a governing body stipulates the training requirements and licensure of sonographers and physicians (radiologists). While there is a regulatory system for continuing education requirements and quality assessment for recertification in the United States, it is not intended for the evaluation of individual performance. An ideal system would have institutional practice guidelines to set the scope of individuals who perform or interpret ultrasound examinations.

Integration of Clinical Information

Deep-learning ultrasound has been studied for pattern recognition and classification based on hepatic echogenicity to make a diagnosis of fatty liver disease [9]. Subsequently, hepatic fibrosis grading using a deep-learning technique with ultrasound shear wave elastography has been developed [10,11]. While these studies provide important insights, further investigations in patients with many etiologies are necessary, as different morphologic changes over time may be expected in relation to different etiologies.

A CAD algorithm, based on a morphologic feature analysis, has been developed for the differential diagnosis of ovarian lesions. AI has tremendous capabilities to integrate clinical information such as family history, genetic information, menstruation status, and hormonal treatment with image analysis, which could improve the diagnostic ability to differentiate benign and malignant ovarian lesions and could add predictive information for management. Clinical information should be integrated into image data analysis tools both to improve performance and to maximize clinical usefulness.

Institutional and Individual Feedback

The American College of Radiology has developed the Ultrasound Liver Imaging Reporting and Data System (US LI-RADS) algorithm to standardize interpretation utilizing two components: detection scores and visualization scores [12]. Son et al. [13] evaluated the US LI-RADS algorithm and reported that 86% of hepatocellular carcinomas were missed among patients with an elevated body mass index and moderate/severe fatty infiltration. While AI could provide further insight by integrating demographic information on the incidence of hepatocellular carcinoma, it is uncertain how the individual performance of the operator and interpreter affects tumor detection. AI has the ability to provide both institutional and individual performance feedback. Current AI has the potential to eventually provide performance feedback for operators with integrated clinical information and demographics. Ultimately, AI ultrasonography will not simply assist novice operators, but will help to improve outcome-based performance via continuous feedback with system improvement.

Future Development Directions

The application of AI in ultrasonography will likely have the greatest impact in fields where there is no significant difference between beginners and experts in the initial image acquisition process. The more complex the image acquisition process is, the more difficult it may be to apply AI tools. Further research will likely focus on an easily approachable organ with the application of elastography, color Doppler ultrasonography, and contrast-enhanced Doppler ultrasonography [14]. Three-dimensional acquisition of anatomy has the potential to expand the efficacy of AI applications. Recent advances have been made in research and clinical application of point-of-care ultrasound with the intent to problem-solve single tasks in specific anatomic areas of interest, including the detection of ascites, pleural effusion, pericardial effusion, pneumothorax, and more recently-the complicated diagnosis of coronavirus disease 2019 pneumonia [15,16]. It is hoped that the addition of AI-empowered tools could eventually help non-radiology-trained clinicians in certain clinical scenarios. In a recent article, point-of-care chest ultrasonography demonstrated a sensitivity of 75%, a specificity of 100%, a positive predictive value of 100% and a negative predictive value of 94.9% for the diagnosis of pneumothorax, though there are various situations that decrease the performance of ultrasonography such as subcutaneous emphysema [15]. An AI-enabled ultrasound diagnostic algorithm for pneumothorax has been investigated and proven clinically effective in a small number of patients [17].

Regulatory Aspects of AI in Ultrasound

AI technology has continued to evolve with research advances following FDA approval. Previously, FDA-approved AI applications

had locked-in models that were an obstacle to the adaptive evolution of AI algorithms. To overcome this regulatory problem, in January 2021, the FDA announced an action plan including a Predetermined Change Control Plan, which allows future modification to their original devices within a predefined range of "what to change" and "how to learn" while maintaining safety [18].

Conclusion

AI-empowered ultrasonography has the potential to further accelerate the use of medical ultrasound in various clinical settings with broad usage by medical personnel. The application of AI in ultrasonography could help to assist physicians in the diagnosis and triage of patients. The standardization of ultrasound examinations and qualifications for operators and interpreters should be discussed in medical disciplines, institutional leadership, and governing bodies [8]. These discussions are essential in the looming era of AI. Before using any AI tools, each institution should conduct an internal validation process to verify whether it is suitable for their patients and practitioners, as there is a lack of evidence-based non-randomized prospective studies to validate the efficacy of AI tools [19]. Otherwise, the increasing use of ultrasonography coupled with AI assistant tools could result in wasted resources, malpractice caused by misdiagnoses, and eventually a great burden on medical institutions and their patients.

ORCID: Young H. Kim: https://orcid.org/0000-0002-0206-5663

Conflict of Interest

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

References

1. Schneider M, Bartko P, Geller W, Dannenberg V, Konig A, Binder C, et al. A machine learning algorithm supports ultrasound-naive novices in the acquisition of diagnostic echocardiography loops and provides accurate estimation of LVEF. Int J Cardiovasc Imaging 2021;37:577-586.

2. Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. Ultrasound Obstet Gynecol 2020;56:498-505.

3. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-444.

4. Zhao WJ, Fu LR, Huang ZM, Zhu JQ, Ma BY. Effectiveness evaluation of computer-aided diagnosis system for the diagnosis of thyroid nodules on ultrasound: a systematic review and meta-analysis. Medicine (Baltimore) 2019;98:e16379.

5. Moustafa AF, Cary TW, Sultan LR, Schultz SM, Conant EF, Venkatesh SS, et al. Color Doppler ultrasound improves machine learning diagnosis of breast cancer. Diagnostics (Basel) 2020;10:631.

6. Andersen JKH, Pedersen JS, Laursen MS, Holtz K, Grauslund J, Savarimuthu TR, et al. Neural networks for automatic scoring of arthritis disease activity on ultrasound images. RMD Open 2019;5:e000891.

7. Antico M, Sasazawa F, Dunnhofer M, Camps SM, Jaiprakash AT, Pandey AK, et al. Deep learning-based femoral cartilage automatic segmentation in ultrasound imaging for guidance in robotic knee arthroscopy. Ultrasound Med Biol 2020;46:422-435.

8. Stasi G, Routi EM. A critical evaluation in the delivery of the ultrasound practice: the point of view of the radiologist. Ital J Med 2015;9:5-10.

9. Biswas M, Kuppili V, Edla DR, Suri HS, Saba L, Marinhoe RT, et al. Symtosis: a liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm. Comput Methods Programs Biomed 2018;155:165-177.

10. Brattain LJ, Ozturk A, Telfer BA, Dhyani M, Grajo JR, Samir AE. Image processing pipeline for liver fibrosis classification using ultrasound shear wave elastography. Ultrasound Med Biol 2020;46:2667-2676.

11. Gatos I, Tsantis S, Spiliopoulos S, Karnabatidis D, Theotokas I, Zoumpoulis P, et al. A machine-learning algorithm toward color analysis for chronic liver disease classification, employing ultrasound shear wave elastography. Ultrasound Med Biol 2017;43:1797-1810.

12. Morgan TA, Maturen KE, Dahiya N, Sun MR, Kamaya A; American College of Radiology Ultrasound Liver Imaging and Reporting Data System (US LI-RADS), et al. US LI-RADS: ultrasound liver imaging reporting and data system for screening and surveillance of hepatocellular carcinoma. Abdom Radiol (NY) 2018;43:41-55.

13. Son JH, Choi SH, Kim SY, Jang HY, Byun JH, Won HJ, et al. Validation of US liver imaging reporting and data system version 2017 in patients at high risk for hepatocellular carcinoma. Radiology 2019;292:390-397.

14. Turco S, Frinking P, Wildeboer R, Arditi M, Wijkstra H, Lindner JR, et al. Contrast-enhanced ultrasound quantification: from kinetic modeling to machine learning. Ultrasound Med Biol 2020;46:518-543.

15. Jahanshir A, Moghari SM, Ahmadi A, Moghadam PZ, Bahreini M. Value of point-of-care ultrasonography compared with computed tomography scan in detecting potential life-threatening conditions in blunt chest trauma patients. Ultrasound J 2020;12:36.

16. Yassa M, Mutlu MA, Birol P, Kuzan TY, Kalafat E, Usta C, et al. Lung ultrasonography in pregnant women during the COVID-19 pandemic: an interobserver agreement study among obstetricians. Ultrasonography 2020;39:340-349.

17. Mehanian C, Kulhare S, Millin R, Zheng X, Gregory C, Zhu H, et al.

Deep learning-based pneumothorax detection in ultrasound videos. In: Wang Q, Gomez A, Hutter J, McLeod K, Zimmer V, Zettinig O, editors. Smart ultrasound imaging and perinatal, preterm and paediatric image analysis. Cham: Springer, 2019;74-82.

18. FDA Releases Artificial Intelligence/Machine Learning Action Plan [Internet]. Silver Spring, MD: U.S. Food and Drug Administration, 2021 [cited 2021 Jan 31]. Available from: https://www.fda. gov/news-events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan.

19. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ 2020;368:m689.