frontiers in Oncology

MINI REVIEW
published: 10 June 2021
doi: 10.3389/fonc.2021.631813

Articles That Use Artificial Intelligence for Ultrasound: A Reader's Guide

Ming Kuang[1,2*†], *Hang-Tong Hu*[1†], *Wei Li*[1], *Shu-Ling Chen*[1] *and Xiao-Zhou Lu*[3]

[1] Department of Medical Ultrasonics, Ultrasomics Artificial Intelligence X-Lab, Institute of Diagnostic and Interventional Ultrasound, The First Affiliated Hospital of Sun Yat-Sen University, Guangzhou, China, [2] Department of Hepatobiliary Surgery, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, [3] Department of Traditional Chinese Medicine, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

Artificial intelligence (AI) transforms medical images into high-throughput mineable data. Machine learning algorithms, which can be designed for modeling for lesion detection, target segmentation, disease diagnosis, and prognosis prediction, have markedly promoted precision medicine for clinical decision support. There has been a dramatic increase in the number of articles, including articles on ultrasound with AI, published in only a few years. Given the unique properties of ultrasound that differentiate it from other imaging modalities, including real-time scanning, operator-dependence, and multi-modality, readers should pay additional attention to assessing studies that rely on ultrasound AI. This review offers the readers a targeted guide covering critical points that can be used to identify strong and underpowered ultrasound AI studies.

Keywords: ultrasound, artificial intelligence, machine learning, deep learning, radiomics

OPEN ACCESS

Edited by:
Hui-Xiong Xu,
Tongji University, China

Reviewed by:
Jun Shi,
Shanghai University, China
Xin-Wu Cui,
Huazhong University of Science and Technology, China

***Correspondence:**
Ming Kuang
kuangm@mail.sysu.edu.cn

†These authors share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *21 November 2020*
Accepted: *12 May 2021*
Published: *10 June 2021*

Citation:
Kuang M, Hu H-T, Li W, Chen S-L and Lu X-Z (2021) Articles That Use Artificial Intelligence for Ultrasound: A Reader's Guide. Front. Oncol. 11:631813. doi: 10.3389/fonc.2021.631813

INTRODUCTION

By looking into pixels not readily visible to the human naked eyes, artificial intelligence (AI) has led medical imaging into the era of big data (1). Articles using conventional machine learning (ML) algorithms and deep learning, especially convolutional neural networks (CNN), have also become more numerous over the past several years. Studies have reported the use of AI in X-rays, computerized tomography (CT), magnetic resonance imaging (MRI), ultrasound, and other types of scans, and they have reported superior performance of AI to that of conventional methods in disease detection, characterization, and patient prognosis prediction (2–4).

Working groups of the Consolidated Standards of Reporting Trials-Artificial Intelligence (CONSORT-AI) and the Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence (SPIRIT-AI) have developed an extension to the core CONSORT 2010 items and 2013 SPIRIT statement that serves as a guidance for medical AI studies (5, 6). Given the rapid expansion of the literature published, *JAMA* has provided a reader's guide to assessing clinical AI articles (7), which reviewed the basics of machine learning and aspects of the clinical implementation of AI. The editorial board of *Radiology* also highlighted several crucial considerations meant to formalize AI methodology in medical imaging studies (8). However, when AI is used with ultrasound, issues become complicated for the current existing guides.

Ultrasound uses the reflection of the ultrasonic beam to reveal tissue structure. It is one of the most widely used methods of imaging in clinical practice. It serves as a mainstay in obstetricians, cardiology, interventional therapy guidance and post-treatment surveillance (9). Ultrasound-based radiomics studies, called ultrasomics (10), follow the standard three-step AI process for medical imaging: data preparation, model development and testing, and evaluation of clinical effectiveness (11). However, given ultrasound's unique properties of real-time scanning, operator-dependence, and multi-modality, some specific issues may influence the performance of AI models and the generalizability of a study's results. For example, operator dependence may influence the use of expert-dataset-based model training to the resident-dataset-based model testing and use in primary hospitals. In this minireview, we aim to provide the readers with an overview of how to assess medical imaging AI articles, including some specific points regarding ultrasound AI studies.

OBJECTIVE: IS THE CLINICAL SCENARIO CLEARLY DEFINED?

The objective of a medical imaging AI study should comply with two principles: first, it must be derived from clinical practical needs, and second, it must be applicable to AI technique. For example, un-enhanced ultrasound is recommended for monitoring populations at high risk of liver cancer (12), so it would be a risk stratification tool. An unenhanced ultrasound AI tool would ideally increase the detection rate of liver lesions and assist in risk assessment. When transformed into AI tasks, target recognition and classification are both technically feasible.

MATERIALS AND METHODS: IS THERE AN INDEPENDENT TESTING DATASET BESIDES THE TRAINING AND VALIDATION SETS?

AI models are prone to overfitting. Both conventional ML and CNN algorithms can vary greatly in performance across different data sources (13). After a model is trained using the training set, its hyperparameters must be tuned in the validation set (also called the tuning set) for better generalizability. If multiple models had been trained, the validation set could also be used to select models. Once a model is finalized, its performance must be evaluated in a testing set, which has no overlap with the training or validation sets. Ideally, the testing set comes from other centers, which involves data from different ultrasound devices and vendors, and patients with different demographic characteristics. A study that reports generalizable results in an independent testing dataset would be much more valuable than a study that relies on internal validation or single-dataset-based cross-validation.

MATERIALS AND METHODS: IS THE IMAGE PROCESSING PROCEDURE CLEARLY DESCRIBED?

A clear description of the image processing procedure is vital for the assessment of study repeatability and reproducibility. Readers should pay attention to the ultrasound data acquisition process and the validity of the data range. Questions below should be raised when acquiring such information. Is the data collected retrospectively or prospectively? Which modality does the study apply? Is it radio frequency signal, grayscale, elastography, doppler imaging, contrast-enhanced ultrasound (CEUS), or transferring between modalities (14, 15)? Also, the number of pictures per patient enrolled for the training or testing and whether the patients' clinical data are involved in the AI development should be inspected.

In terms of ultrasound data preprocessing, each step should be presented clearly. Ultrasound images are derived from various devices produced by different radiologists. Ultrasound is highly operator-dependent (16, 17), which causes variations in image quality, target lesion identification, and selection of representative sections. Cropping is widely adopted in image processing in medical AI studies, and it filters out most irrelevant, non-lesion information, and for the ultrasound, reduces image heterogeneity by adjusting size and depth. Augmentation can enrich data diversity, and it can simulate the common causes of image heterogeneity as observed under real-world conditions in ultrasound examinations (18, 19). For example, resizing reduces resolution variation of different devices, rotation simulates scanning from different angles and sections, and contrast adjustment simulates variation in gain and dynamic range.

MATERIALS AND METHODS: IS THE ALGORITHM FOR MODELING SUITABLE?

Conventional ML algorithms such as logistic regression, support vector machine (SVM), random forest, and Naïve Bayes have much fewer parameters than deep learning algorithms. For example, SVM has only 13 parameters to be adjusted, while the ResNet-50 has an amount of 2.3×10^7 parameters. Thus, conventional ML algorithms require far less training than deep learning algorithms do (20). With a limited sample size, such as a set of only hundreds of images (not videos), conventional ML algorithms are preferred (21). However, with thousands or millions of images, deep learning algorithms, principally CNN in imaging analysis, are recommended. The minimum number of training images needed varies across different tasks and algorithms and may only be determined by evaluating the relationship between its increase and changes in model performance.

Algorithms' clinical intelligibility, which means the level of understandability of an algorithm in a clinical way, should also be considered. There has not been any ultrasound-specific

imaging analysis algorithms reported. Instead, model algorithm selection is primarily based on the type of task. Ultrasound has multiple modalities. CEUS videos record a lesion's hemodynamic information revealed by the dynamic perfusion of microbubble contrast agents. Multi-phase image features can be extracted by simply analyzing frames from each phase but the time sequencing features were missing. Recurrent neural networks (RNNs) such as long short-term memory (LSTM) or gated recurrent units can be incorporated to these time-dimension-related tasks (18). Previous studies using LSTM in CEUS reported excellent performance (22, 23). The application of clinically explicable AI algorithms to modeling renders the study findings more clinically acceptable.

MATERIALS AND METHODS: IS THE AI ALGORITHM PUBLICLY AVAILABLE?

Even being generalizable among different datasets in a given study, especially for studies carried out in a single center, AI performance still needs a broad verification. The existing public medical imaging data sets are minimal (24), and no public ultrasound dataset exists. Authors are encouraged to make their AI models publicly available *via* such websites as GitHub (https://github.com/) to allow independent validation, fine-tuning, and updating. A study reporting publicly available AI algorithms may improve its results' reliability in this way.

RESULTS: HOW DO THE RESULTS PRODUCED BY THE AI MODEL COMPARE TO THOSE PRODUCED BY EXPERT RADIOLOGISTS?

Medical AI must be evaluated against the performance of radiology experts (8). The value of a prospectively designed AI performance testing procedure can be determined by comparing its performance to that of human experts under real-world conditions. In retrospectively designed studies, missing data, and data mismatch regarding the target lesion are unavoidable in datasets collected from clinical practice, considering which is beyond AI's ability (25). Radiologists make ultrasound diagnosis in real time during face-to-face examinations, where they receive far more information than retrospective image review does. The common study design usually underestimates radiologists' performance and renders meaningful evaluation of medical AI difficult.

Combing clinician experience and AI's advantages can render imaging more efficient and accurate than either alone (26). Because ultrasound offers diagnosis in real time and is heavily dependent on the operator, ultrasound AI's performance should be compared to that of radiologists with varied experiences to develop a viable human-AI interaction strategy (27). Ideally, this strategy would involve dynamic assessment during an ultrasound examination. A specific application scenario based AI developing and testing study would have considerable practical value.

RESULTS: ARE THE EVALUATION INDEXES SUITABLE?

For detection and classification purposes, an AI model is first evaluated by the receiver operating characteristic curve (ROC) or precision-recall curve (PRC), and further by its accuracy, error rate or F1 value. However, in medical imaging analysis programs, performance is assessed based on indicators of clinical significance, such as sensitivity and specificity for diagnosis and prediction programs (28, 29), detection rate for disease screening and lesion detection (30, 31), κ and dice coefficient for inter-annotator agreement and overlapping in radiotherapy planning (32, 33). For example, for a screening task model, detection rate and sensitivity would be the primary indexes for model evaluation, while for diagnostic tasks, high specificity or positive predictive value would be the top priority. A specifically preferred high evaluation index can be achieved using an appropriate cutoff value for AI outputs but not necessarily by the default of 0.5 or the Youden index.

DISCUSSION: ARE THE RESULTS COMPARED TO STATE-OF-ART REPORTS?

AI results should be compared to state-of-art reports, both the previous studies of the same design and these using other imaging modalities, traditional methods, or guideline recommendations. Readers should keep in mind that results without independent tests or internally validated results are not comparable to studies reporting independently tested results, no matter how good the statistics are relative to state-of-art results. A well-designed study with practical results is much more valuable than studies with flawed design but with good statistical results.

DISCUSSION: WHAT IS THE UNSOLVED PROBLEM OF THE PRESENT WORK?

Limitations of medical AI studies are often the challenge of future work. For example, what situation wouldn't the AI system be implemented when considering that AI performance errors and failure cases could influence clinical practice decision-making? What are the latent factors keeping AI systems from generalizing to other centers and populations, given the hardware requirements, algorithm versions, data quality, and processing procedures? How can these be solved in further study? Is the sample size large enough to build a robust model? The relationship between the training dataset size and model performance should be evaluated, as Dunnmon et al. (34) in the research reporting that the AI performance benefited little after a certain number of images were used for training.

CONCLUSION

Given ultrasound's unique properties, readers should pay additional attention when assessing an AI study that relies on ultrasound than those that rely on other imaging modalities. Here, we list several crucial points to help readers distinguish strong ultrasound AI articles from underpowered articles. With more formalized standards for medical AI studies published in the future, ultrasound AI studies may better benefit the clinical practice.

AUTHOR CONTRIBUTIONS

Conception, design, and final approval of the manuscript, MK. Preparing the main manuscript, H-TH. Editing and review of the manuscript, WL, S-LC. Revision of the manuscript, X-ZL. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by the National Natural Science Foundation of China (NO. 81971630).

ACKNOWLEDGMENTS

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found onlineat: https://www.frontiersin.org/articles/10.3389/fonc.2021.631813/full#supplementary-material

REFERENCES

1. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial Intelligence in Radiology. *Nat Rev Cancer* (2018) 18(8):500–10. doi: 10.1038/s41568-018-0016-5
2. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer (Oxford Engl 1990)* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
3. Carin L, Pencina MJ. On Deep Learning for Medical Image Analysis. *Jama* (2018) 320(11):1192–3. doi: 10.1001/jama.2018.13316
4. Dong Y, Wang QM, Li Q, Li LY, Zhang Q, Yao Z, et al. Preoperative Prediction of Microvascular Invasion of Hepatocellular Carcinoma: Radiomics Algorithm Based on Ultrasound Original Radio Frequency Signals. *Front Oncol* (2019) 9:1203. doi: 10.3389/fonc.2019.01203
5. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension. *Nat Med* (2020) 26(9):1364–74. doi: 10.1038/s41591-020-1034-x
6. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for Clinical Trial Protocols for Interventions Involving Artificial Intelligence: The SPIRIT-AI Extension. *Nat Med* (2020) 26(9):1351–63. doi: 10.1038/s41591-020-1037-7
7. Liu Y, Chen PC, Krause J, Peng L. How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. *Jama* (2019) 322(18):1806–16. doi: 10.1001/jama.2019.16489
8. 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* (2019) 294(3):487–9. doi: 10.1148/radiol.2019192515
9. Muse ED, Topol EJ. Guiding Ultrasound Image Capture With Artificial Intelligence. *Lancet (London England)* (2020) 396(10253):749. doi: 10.1016/S0140-6736(20)31875-4
10. Li W, Huang Y, Zhuang BW, Liu GJ, Hu HT, Li X, et al. Multiparametric Ultrasomics of Significant Liver Fibrosis: A Machine Learning-Based Analysis. *Eur Radiol* (2019) 29(3):1496–506. doi: 10.1007/s00330-018-5680-z
11. Yin R, Jiang M, Lv WZ, Jiang F, Li J, Hu B, et al. Study Processes and Applications of Ultrasomics in Precision Medicine. *Front Oncol* (2020) 10:1736. doi: 10.3389/fonc.2020.01736
12. Morgan TA, Maturen KE, Dahiya N, Sun MRM, Kamaya AAmerican College of Radiology Ultrasound Liver I, et al. Us LI-RADS: Ultrasound Liver Imaging Reporting and Data System for Screening and Surveillance of Hepatocellular Carcinoma. *Abdominal Radiol* (2018) 43(1):41–55. doi: 10.1007/s00261-017-1317-y
13. De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease. *Nat Med* (2018) 24(9):1342–50. doi: 10.1038/s41591-018-0107-6
14. Han X, Wang J, Zhou W, Chang C, Ying S, Shi J. Deep Doubly Supervised Transfer Network for Diagnosis of Breast Cancer with Imbalanced Ultrasound Imaging Modalities. In: A.L Martel, et al. (eds) *Medical Image Computing and Computer Assisted Intervention - MICCAI 2020. MICCAI 2020. Lecture Notes in Computer Science*. 12266 doi: 10.1007/978-3-030-59725-2_14
15. Fei X, Shen L, Ying S, Cai Y, Zhang Q, Kong W, et al. Parameter Transfer Deep Neural Network for Single-Modal B-Mode Ultrasound-Based Computer-Aided Diagnosis. *Cogn Comput* (2020) 12(6):1252–64. doi: 10.1007/s12559-020-09761-1
16. Todsen T, Tolsgaard MG, Olsen BH, Henriksen BM, Hillingsø JG, Konge L, et al. Reliable and Valid Assessment of Point-of-Care Ultrasonography. *Ann Surg* (2015) 261(2):309–15. doi: 10.1097/SLA.0000000000000552
17. Chou R, Cuevas C, Fu R, Devine B, Wasson N, Ginsburg A, et al. Imaging Techniques for the Diagnosis of Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis. *Ann Internal Med* (2015) 162(10):697–711. doi: 10.7326/M14-2509
18. Ramachandram D, Taylor GW. Deep Multimodal Learning: A Survey on Recent Advances and Trends. *IEEE Signal Process Magazine* (2017) 34(6):96–108. doi: 10.1109/MSP.2017.2738401
19. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A Guide to Deep Learning in Healthcare. *Nat Med* (2019) 25(1):24–9. doi: 10.1038/s41591-018-0316-z
20. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications. *CA: Cancer J Clin* (2019) 69(2):127–57. doi: 10.3322/caac.21552
21. Zhang H, Guo L, Wang D, Wang J, Bao L, Ying S, et al. Multi-Source Transfer Learning Via Multi-Kernel Support Vector Machine Plus for B-Mode Ultrasound-Based Computer-Aided Diagnosis of Liver Cancers. *IEEE J Biomed Health Inf* (2021). PP(99):1–1. doi: 10.1109/JBHI.2021.3073812
22. Azizi S, Bayat S, Yan P, Tahmasebi A, Kwak JT, Xu S, et al. Deep Recurrent Neural Networks for Prostate Cancer Detection: Analysis of Temporal Enhanced Ultrasound. *IEEE Trans Med Imaging* (2018) 37(12):2695–703. doi: 10.1109/TMI.2018.2849959
23. Sharma H, Droste R, Chatelain P, Drukker L, Papageorghiou AT, Noble JA. Spatio-Temporal Partitioning and Description of Full-Length Routine Fetal Anomaly Ultrasound Scans. *Proc IEEE Int Symposium Biomed Imaging* (2019) 16:987–90. doi: 10.1109/ISBI.2019.8759149
24. 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(3):590–606. doi: 10.1148/radiol.2018180547

25. Doshi-Velez F, Perlis RH. Evaluating Machine Learning Articles. *Jama* (2019) 322(18):1777–9. doi: 10.1001/jama.2019.17304
26. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. *Am J Surg Pathol* (2018) 42(12):1636–46. doi: 10.1097/PAS.0000000000001151
27. Moga TV, Popescu A, Sporea I, Danila M, David C, Gui V, et al. Is Contrast Enhanced Ultrasonography a Useful Tool in a Beginner's Hand? How Much Can a Computer Assisted Diagnosis Prototype Help in Characterizing the Malignancy of Focal Liver Lesions? *Med Ultrasonography* (2017) 19(3):252–8. doi: 10.11152/mu-936
28. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep Learning Radiomics of Shear Wave Elastography Significantly Improved Diagnostic Performance for Assessing Liver Fibrosis in Chronic Hepatitis B: A Prospective Multicentre Study. *Gut* (2019) 68(4):729–41. doi: 10.1136/gutjnl-2018-316204
29. Eun NL, Kang D, Son EJ, Park JS, Youk JH, Kim JA, et al. Texture Analysis With 3.0-T MRI for Association of Response to Neoadjuvant Chemotherapy in Breast Cancer. *Radiology* (2020) 294(1):31–41. doi: 10.1148/radiol.2019182718
30. Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-Time Automatic Detection System Increases Colonoscopic Polyp and Adenoma Detection Rates: A Prospective Randomised Controlled Study. *Gut* (2019) 68(10):1813–9. doi: 10.1136/gutjnl-2018-317500
31. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *Jama* (2017) 318(22):2199–210. doi: 10.1001/jama.2017.14580
32. Oktay O, Nanavati J, Schwaighofer A, Carter D, Bristow M, Tanno R, et al. Evaluation of Deep Learning to Augment Image-Guided Radiotherapy for Head and Neck and Prostate Cancers. *JAMA Network Open* (2020) 3(11): e2027426. doi: 10.1001/jamanetworkopen.2020.27426
33. Liang S, Tang F, Huang X, Yang K, Zhong T, Hu R, et al. Deep-Learning-Based Detection and Segmentation of Organs at Risk in Nasopharyngeal Carcinoma Computed Tomographic Images for Radiotherapy Planning. *Eur Radiol* (2019) 29(4):1961–7. doi: 10.1007/s00330-018-5748-9
34. Dunnmon JA, Yi D, Langlotz CP, Ré C, Rubin DL, Lungren MP. Assessment of Convolutional Neural Networks for Automated Classification of Chest Radiographs. *Radiology* (2019) 290(2):537–44. doi: 10.1148/radiol.2018181422

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.