LETTER TO THE EDITOR

Lack of evidence and criteria to evaluate artificial intelligence and radiomics tools to be implemented in clinical settings

Qian Zhou[1] · Yi-heng Cao[2] · Zhi-hang Chen[3]

Dear Sir,

We read with interest the recent review by Sollini et al. entitled "Toward clinical application of image mining: a systematic review on artificial intelligence and radiomics" [1]. The authors aimed to systematically review literature on artificial intelligence (AI) and radiomics to evaluate the quality and progress of the image mining research in the medical field. They developed a classification tool according to phases of clinical trials to classify the development process of AI and radiomics research. They concluded that adapting tools with knowledge from the drug development process could represent the most effective way for radiomics and AI algorithms to become the standard of care tools.

Well conducted phase III clinical trials of drug development may demonstrate that a new drug compared with standard care or placebo is effective and safe to be put on the market. This type of trial typically represents the highest quality and level of evidence in a single study [2]. However, there are some important limitations in Sollini et al.'s image mining tools development process. The authors classified phase III of the development process as being prospective, validated, and with at least 100 patients. Although they classified eight studies as phase III, but one of them did not provide any details about how they collected data prospectively [3]. Three other studies did not mention how their algorithms can be applied in clinical practices [4,5,6]. Our critical evaluation of these eight "Phase III"

studies raises questions about the appropriateness of the criteria the authors use to evaluate the application of these studies into clinical settings. Moreover, in their discussion, the authors pointed out that "the sample size should be big enough to minimise the effects of overfitting, be comprehensive of the "outliers", and, consequently, be reliable when used for the assessment of unseen patients" and, indeed, that is correct. However, later in the section they "arbitrarily chose 100 samples as the threshold for trial phases categorisation" which contradicts their previous statement. All of these issues mean it is not appropriate to include these studies as phase III studies in their proposed development process tool. Because their classification is inspired by clinical trials of drug development, the classification criteria for phase III trials should be stricter to avoid including poor quality studies. Therefore, there remains a lack of evidence justifying the criteria used to evaluate artificial intelligence and radiomics algorithms for implementation in clinical settings.

Further reading of the review identified additional issues that require clarification. In the "Search, eligibility criteria and study selection", the authors provided a query and limit date to help readers to find the references. However, this query does not correspond to the number of references reported in the paper. In the Supplementary Material "Table S2. Summary of the high-quality articles on image mining (n=171)", the numbers of "Type of clinical trial" and "Type of validation" are not correct. The sum of the frequencies gives a total of 169 and 170 for the "Type of clinical trial" and "Type of validation" respectively whereas 171 would be expected.

This article is part of the Topical Collection on Advanced Image Analyses (Radiomics and Artificial Intelligence)

✉ Qian Zhou
    zhouq49@mail.sysu.edu.cn

1   Department of Medical Statistics, Clinical Trials Unit, The First Affiliated Hospital of Sun Yat-sen University, 58 Zhong Shan Er Road, Guangzhou 510080, China

2   ESIEE Paris, Cité Descartes BP99, 93162 Noisy-le-Grand, France

3   Department of Liver Surgery, The First Affiliated Hospital of Sun Yat-sen University, 58 Zhong Shan Er Road, Guangzhou 510080, China

Authors' contributions Qian Zhou: study design, article writing, final approval of the manuscript. Yi-heng Cao: study design, article writing, final approval of the manuscript. Zhi-hang Chen: study design, article writing, final approval of the manuscript.

Data availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Compliance with ethical standards

Competing interests The authors declare that they have no conflict of interest.

Ethics approval and consent to participate Not applicable.

Abbreviations AI, artificial intelligence

References

1. Sollini M, Antunovic L, Chiti A, Kirienko M. Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. Eur J Nucl Med Mol Imaging. 2019. https://doi.org/10.1007/s00259-019-04372-x.

2. Piantadosi S. Clinical trials: a methodologic perspective. 2nd ed: Wiley; 2006.

3. Moon WK, Shen YW, Huang CS, Chiang LR, Chang RF. Computer-aided diagnosis for the classification of breast masses in automated whole breast ultrasound images. Ultrasound Med Biol. 2011;37: 539–48.

4. 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;29:1496–506.

5. Liu Y, Zhang Y, Cheng R, Liu S, Qu F, Yin X, et al. Radiomics analysis of apparent diffusion coefficient in cervical cancer: a preliminary study on histological grade evaluation. J Magn Reson Imaging. 2019;49:1–11.

6. Bodduluri S, Newell JD, Hoffman EA, Reinhardt JM. Registration-based lung mechanical analysis of chronic obstructive pulmonary disease (COPD) using a supervised machine learning framework. Acad Radiol. 2013;20:527–36.