



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

Martina Sollini^{1,2} · Margarita Kirienko²

Published online: 30 August 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Keywords Radiomics · Artificial intelligence · Image mining · Phase classification · Trial phases

We appreciated the comments on our recent review entitled “Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics” [1] by Zhou et al [2]. This allows us to further explain the rationale for our proposal of the research path in image mining field.

Zhou et al. affirm that there is “lack of evidence and criteria to be used to evaluate artificial intelligence and radiomics algorithms to be implemented in clinical settings.” The classification we describe is a proposal for the research path, objectivized by the QUADAS-2 score in combination with the established drug development path. As we pointed out in the discussion section, the described phase distinction is our group proposal, arbitrary by definition, inspired by the revised literature, methodological and modeling considerations.

They commented on the proposed criteria for the phase III study classification. In particular, they disapprove our proposal of the sample size of 100 cases. Whatever the cut-off for the patient cohort size we could propose would have been prone to criticism. Nonetheless, in view of the reviewed literature, we chose 100 sample size as reasonable given the necessity of

the prospective study design and at least 7/8 QUADAS-2 score to define a study as belonging to the phase III class.

Zhou et al. affirm that we upgraded 4 papers as phase III studies. The colleagues pointed out that in one study, the authors did not report “how they collected the data prospectively.” Indeed, we assigned 7/8 QUADAS-2 score to the study because of insufficient clarity on the patient selection methods. We interpreted the study to be prospective since the authors declared to have obtained informed consent from all the patients that have been ratified by the local ethics committee; additionally, unique scanner and uniform acquisition settings were used. Moreover, Zhou et al. stated that additional three studies did not mention “how their algorithms can be applied into clinical practices.” However, this aspect did not constitute a criterion for phase classification; therefore, a misclassification did not occur. Irrespective of a possible misclassification, the proposed criteria for the research path towards clinical implementation of the image mining tools are sufficiently robust and easily applicable. Nonetheless, any discussion and improvement are welcome for maturation of radiomics and artificial intelligence-based algorithms.

Zhou et al. support more strict criteria for phase III studies, but they do not propose any alternative. As already pointed out, it is our opinion that a constructive discussion will only benefit the scientists involved in the field.

As for the search strategy and paper selection, we suppose that Zhou et al. did not get the same number of records, since we canceled those reports that our search algorithm retrieved multiple times at each key words' combination.

Finally, Zhou et al. correctly spotted one error in supplementary table 2 reporting the number of articles per type of validation and we are grateful for their comment.

We want to underline that the number of articles per QUADAS-2 score is correct. The revised portion of the table is shown in Table 1.

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

This reply refers to the article available at <https://doi.org/10.1007/s00259-019-04493-3>

✉ Margarita Kirienko
margarita.kirienko@icloud.com

Martina Sollini
martina.sollini@cancercenter.humanitas.it

¹ Nuclear Medicine, Humanitas Clinical and Research Center - IRCCS, via Manzoni, 56, 20089 Rozzano, Milan, Italy

² Department of Biomedical Sciences, Humanitas University, via Rita Levi Montalcini, 4, 20090 Pieve Emanuele, Milan, Italy

Table 1 Summary of the high-quality articles on image mining ($n = 171$)

		AI ($n = 27$)	Radiomics ($n = 141$)	Combined radiomics-AI ($n = 3$)
Type of validation	Independent	20	101	2
	Internal	7	40	1
QUADAS-2 score	7	13	93	1
	8	14	48	2

AI artificial intelligence

We understand that Zhou et al. consider the image mining tools as immature, and the studies of insufficient quality to be introduced into clinical practice. Overall, this is in line with our conclusion.

In our opinion, further methodological and reporting improvements, adherence to standardized processing and description, and interdisciplinary team engagement are necessary to allow timely introduction of the image mining tools into daily clinical practice.

Author contributions MS and MK conceptualized and approved the manuscript.

Funding information MK PhD scholarship was funded by the AIRC grant (IG-2016-18585).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical statement In view of the nature of the present article an ethical approval was considered unnecessary.

Reference

1. Sollini M, Antunovic L, Chiti A, Kirienco 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. Zhou Q, Cao Y, Chen Z. Lack of evidence and criteria to evaluate artificial intelligence and radiomics tools to be implemented in clinical settings. *Eur J Nucl Med Mol Imaging* (2019).

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