



Evaluation of prostate MRI: can machine learning provide support where radiologists need it?

Alexander D. J. Baur¹ · Tobias Penzkofer^{1,2}

Received: 7 April 2019 / Accepted: 16 April 2019 / Published online: 9 May 2019
© European Society of Radiology 2019

Magnetic resonance imaging of the prostate: where do we stand?

Prostate cancer (PCa) is a heterogeneous disease and tumor grading is a major predictor of prognosis [1]. Multiparametric magnetic resonance imaging (mpMRI) is a valuable tool to non-invasively detect PCa and has even proven superior compared to systematic biopsy [2]. In order to improve and standardize this technique and its application, guidelines for acquisition, interpretation, and reporting of mpMRI of the prostate have been established [3]. Using a current version of these guidelines, the Prostate Imaging–Reporting and Data System (PI-RADS) version 2, and its decision rules, high sensitivities and specificities for the detection of clinically significant PCa can be reached [4]. A major limitation is an only moderate to good interreader agreement [5]. In order to address this issue, an updated version of PI-RADS, version 2.1, has been published just recently [6].

What is machine learning and what are the challenges?

Over the past years the promises of the application of machine learning (ML) in radiology have created a hype with increasing amounts of papers being published and increasing

numbers of companies and researchers from different fields being involved. ML describes a broad class of analysis algorithms that are not programmed with explicit decision rules but build models for semi-autonomous or autonomous predictions inferred from ideally large and high-quality training datasets [7]. ML can incorporate vast amounts of parameters derived from imaging as well as clinical parameters and combine them in ways not easily comprehensible by humans to calculate the probability of a diagnosis. In supervised ML, sufficiently large training datasets have to be labeled based on a reliable reference standard, as otherwise the algorithms' decisions will be flawed. Several specific types of ML algorithms exist, and their description as well as the discussion of associated challenges would go far beyond the scope of this comment.

Due to its ability to incorporate large amounts of parameters ML might help to identify associations between features and selected diseases or disease states yet unknown or too subtle for human recognition and subsequently improve diagnostics. In addition, ML could help to overcome limitations of human image interpretation such as a suboptimal interreader agreement. However, its application can be unsatisfying for radiologists, as we have been trained using an approach to image interpretation, where findings on different MRI sequences are supposed to represent changes in tissue composition more or less specific for a selected disease. These associations often have been evaluated extensively in studies before being used in clinical routine or even being integrated into decision rules for assigning a diagnosis.

Ignoring the hype around ML, technical challenges, and the potential dissatisfaction one might experience, another aspect is very important when discussing potential applications: Experienced clinicians and radiologists must define tasks and applications, where ML can have a positive impact on patient care. An example for such an application of ML in patients undergoing mpMRI of the prostate has been given by Wang et al. they were able to show that ML algorithms incorporating PI-RADS version 2 assessment categories and mpMRI

This comment refers to the article available at <https://doi.org/10.1007/s00330-019-06244-2>.

✉ Alexander D. J. Baur
alexander.baur@charite.de

¹ Klinik für Radiologie, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Augustenburger Platz 1, 13353 Berlin, Germany

² Berlin Institute of Health (BIH), Anna-Louisa-Karsch-Str. 2, 10178 Berlin, Germany

radiomics achieve higher predictive values for PCa compared to PI-RADS alone and therefore might help to further improve the diagnostic performance of mpMRI [8].

What can we learn from the current study and where do we go from here?

In this issue of European Radiology, Antonelli et al. present the results of a study on selected supervised ML algorithms for the detection of Gleason 4 pattern PCa based on parameters derived from mpMRI acquired using a standardized imaging protocol as well as selected clinical parameters. The authors used the results of template as well as MRI-targeted biopsies as a high-quality reference standard and a relatively large number of lesions as a training dataset. They were able to show that selected supervised ML algorithms can predict the presence of Gleason 4 pattern in equivocal or suspicious lesions previously contoured by an experienced radiologist in the peripheral and transition zone. Furthermore, they were able to show that selected ML algorithms outperformed experienced radiologists. As said before, PCa is a heterogeneous disease and tumor grading is a major predictor of prognosis, especially the presence and percentage of Gleason 4 pattern [1]. Therefore, the task selected for ML in this study is of high clinical relevance, as a potential tool for the detection of clinically significant PCa in patients before biopsy as well as for patients with biopsy-proven low-grade PCa that consider undergoing or already undergo active surveillance [9]. It would be interesting to evaluate, what effect the contouring of lesions by different readers has and if the tested ML algorithms can help to overcome the limitation of an only suboptimal interreader agreement in interpretation of mpMRI by humans. The evaluation of ML as a tool for clinical decision making or patient stratification in studies assessing patient outcome as an endpoint in the previously described scenarios is eagerly awaited.

Antonelli et al. also evaluated the best combination of parameters used as an input for the selected ML algorithms. Remarkably, the input data is comprised of relatively few and established parameters such as prostate-specific antigen density, apparent diffusion coefficient values, and maximum enhancement resulted in an optimal Gleason 4 pattern prediction. This can be seen as reassuring regarding the way radiologists interpret imaging and also indicates how in the future ML might not just potentially replace humans, but also help to identify useful parameters readily available in clinical practice and subsequently improve our way of interpreting imaging and defining rules to do so.

As Antonelli et al. already state in their paper, external validation of the selected algorithms as well as further training with much larger datasets should be performed. Only if ML can be used successfully using imaging from different scanners and acquired with potentially even slightly different imaging protocols as input data, it will be applicable in a broader context. If ML stands this test and a benefit for patients can be proven, it will prevail and be a welcomed addition to our current way of performing radiology.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Alexander D. J. Baur.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Outside of the current work, Alexander D. J. Baur has received payments as a speaker for Bayer Pharma AG and Uropharm AG and research support from Bayer Pharma AG (no personal payments).

Tobias Penzkofer received grant support from the Berlin Institute of Health within the Clinician Scientist Programme and received research support from Siemens Healthcare and Philips Healthcare (no personal payments). Outside of the current work, Tobias Penzkofer is involved in clinical trials with AGO, Aprea AB, Astellas Pharma Global Inc., AstraZeneca, Celgene, Genmab A/S, Incyte Corporation, Lion Biotechnologies, Inc., Millennium Pharmaceuticals, Inc., Morphotec Inc., MSD, Tesaro Inc., and Roche.

Statistics and biometry A statement regarding statistics and biometry is not applicable.

Informed consent A statement regarding informed consent is not applicable.

Ethical approval A statement regarding ethical approval is not applicable.

References

1. Epstein JI, Zelefsky MJ, Sjoberg DD et al (2016) A contemporary prostate cancer grading system: a validated alternative to the Gleason score. *Eur Urol* 69:428–435
2. Ahmed HU, El-Shater Bosaily A, Brown LC et al (2017) Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *Lancet* 389:815–822
3. Padhani AR, Weinreb J, Rosenkrantz AB, Villeirs G, Turkbey B, Barentsz J (2019) Prostate imaging-reporting and data system steering committee: PI-RADS v2 status update and future directions. *Eur Urol* 75:385–396
4. Woo S, Suh CH, Kim SY, Cho JY, Kim SH (2017) Diagnostic performance of prostate imaging reporting and data system version 2 for detection of prostate cancer: a systematic review and diagnostic meta-analysis. *Eur Urol* 72:177–188

5. Rosenkrantz AB, Ginocchio LA, Cornfeld D et al (2016) Interobserver reproducibility of the PI-RADS version 2 lexicon: a multicenter study of six experienced prostate radiologists. *Radiology* 280:793–804
6. Turkbey B, Rosenkrantz AB, Haider MA et al (2019) Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. *Eur Urol*. <https://doi.org/10.1016/j.eururo.2019.02.033>
7. Kohli M, Prevedello LM, Filice RW, Geis JR (2017) Implementing machine learning in radiology practice and research. *AJR Am J Roentgenol* 208:754–760
8. Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang YD (2017) Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol* 27:4082–4090
9. Barrett T, Haider MA (2017) The emerging role of MRI in prostate cancer active surveillance and ongoing challenges. *AJR Am J Roentgenol* 208:131–139

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