Can texture features computed from the joint intensity distribution of different MRI sequences accurately predict prostate cancer grade?

Vasilis Stavrinides, Lina Carmona Echeverria, Hayley C. Whitaker

Division of Surgery and Interventional Science, University College London, London, UK

Correspondence to: Hayley C. Whitaker. Division of Surgery and Interventional Science, University College London, Charles Bell House, 43-45 Foley Street, London W1T 7TS, UK. Email: Hayley.whitaker@ucl.ac.uk.

The diagnostic landscape of prostate cancer has evolved rapidly, from prostate-specific antigen (PSA) testing to exciting new technologies that allow visualization of the disease, moving away from random sampling to targeted biopsies. Multiparametric magnetic resonance imaging (mpMRI) is a new modality that combines T2-weighted (T2W), diffusion-weighted (DW) and dynamic contrast-enhanced (DCE) sequences, each designed to reveal specific microstructural features typically associated with malignancy such as increased vascularity and cellularity. Recent studies such as PROMIS and PRECISION have shown that mpMRI has a high negative predictive value, optimises cancer detection rates and reduces the number of biopsies needed in the active surveillance setting, where patients with lower risk disease are regularly monitored after diagnosis rather than immediately treated (1-3). Despite these developments, a significant challenge in the post-mpMRI era is the sudden emergence of new, previously uncharacterised MRI phenotypes, the clinical importance of which is not yet fully clear. The radiological grading of prostate mpMRI is now routine in clinical practice and relies on assigning a degree of suspicion to a particular MRI lesion or area, thus helping clinicians decide whether a biopsy or intervention is warranted. These systems work fairly well and provide a common language for radiologists, urologists and pathologists. Nonetheless, they rely on a significant degree of subjectivity, have moderate reproducibility among radiologists and ultimately discretize multiple continuous variables with the potential of losing valuable predictive and prognostic information (4).

Texture analysis refers to a collection of techniques that quantify the grey-level patterns and pixel interrelationships in an image in order to recognise patterns of variation often imperceptible to the human eye (5). In many studies, this involves selecting regions of interest (ROIs) in an MR image and extracting a set of statistical features using a (usually grey-level) co-occurrence matrix (GLCM), a mathematical construct that expresses the frequency of grey-level combinations within the ROI in a tabular form. Texture analysis has been applied in the field of prostate cancer with significant success. In a well-known study, Wibmer and colleagues used T2WI and DWI from 147 patients who underwent a radical prostatectomy (RP) and calculated the ability of Haralick features to differentiate cancer from non-cancer using whole-mount pathology as a reference standard (6). The authors found that this distinction is possible and that features such as energy and entropy on ADC maps correlate with cancer grade expressed by the Gleason score. Other authors have used template mapping biopsies as the reference standard (7). It is important to note that GLCM-based feature extraction, although very common in the literature, is not the only possibility. For example, spectral features calculated by wavelet functions are sometimes used for analysis, but such methods are more

computationally intensive and less popular (8).

Generally, co-occurrence matrix-based analyses are applied in one modality at a time and in the prostate mpMRI domain features are extracted separately for T2WI, DWI and DCE, while their joint distribution tends to be disregarded. This is somewhat counterintuitive, considering it is the simultaneous, combined review of all mpMRI sequences by the expert radiologist that confers mpMRI its full potential. In their paper, Chaddad and colleagues propose a new approach for extracting radiomic features from prostate MRIs by using a joint intensity matrix (JIM) that calculates the joint intensity distribution between 3D images of two different modalities (T2WI and DWI) (9). This is a significant departure from the usual GLCM-based approaches, which generally tend to ignore encoding relationships between different mpMRI sequences. The JIM computation described could be a significant step towards fully exploiting intensity values from all three mpMRI modalities for the purpose of deriving clinically useful radiomic signatures.

In more detail, the authors analysed data obtained from 99 patients with biopsy-confirmed, MRI-localised prostate cancer from the SPIE-AAPM-NCI and the Cancer Imaging Archive (TCIA). The patients were divided into three Gleason grade groups (3+3, 3+4 and ≥4+3) and GLCM/JIM were computed for all ROIs in order to extract quantitative features using 19 functions originally proposed by Haralick. To select the features with the greatest ability to discriminate between the three different Grade groups, a combination of non-parametric analysis of variance and correlation analyses was used while accounting for multiple comparisons through a Holm-Bonferroni correction. Finally, GLCM and JIM-derived features were used to train a random forest algorithm to classify tumours to a specific Gleason group in a binary fashion (i.e., one grade category against the rest).

Interestingly, five JIM-derived features (contrast, homogeneity, difference variance, dissimilarity and inverse difference) were shown to be significantly different across the three Gleason groups. Area under the curve (AUC) evaluation of the random forest classifier demonstrated that JIM-based features performed better than GLCM-based or standard characteristics alone, although no formal statistical comparison was made. In addition, combining JIM and GLCM features increased AUC values even further (reaching 78.4%, 82.35% and 64.76% for Gleason 3+3, 3+4 and ≥4+3, respectively).

From a clinical perspective this development could have interesting applications, such as the detection of Gleason upgrading in patients managed by mpMRI-based active surveillance. In this scenario, where the main objective is not cancer detection (as the diagnosis has already been made) but the accurate and well-timed recognition of pathological progression in an otherwise fit and well patient, it very well might be that a JIM-based approach could yield previously unobtainable, clinically useful features that are superior to conventionally computed ones.

Despite optimism regarding the clinical applicability of this and many similar quantitative imaging studies, there is a need for caution. The AUC for detecting cancers with Gleason ≥4+3 or greater (which are of particular clinical interest) was only moderately high. Such findings reinforce general concerns regarding the analytical validation (i.e., the measurement of accuracy, precision, repeatability, reproducibility and feasibility) and the qualification (i.e., demonstration of surrogacy and association with a clinical endpoint) of quantitative imaging metrics that have long been raised (10). These are not due to a lack of novel and interesting ideas, but more due the fact that most studies on prostate MRI to-date are single-centre, have a small sample size and focus heavily on very particular patient populations, which could significantly bias results and prevent generalizability and clinical application.

The performance of machine learning classifiers (random forests, support vector machines and neural networks are all being evaluated) and the results of highly dimensional multivariate analyses heavily depend on the structure of the underlying dataset. Therefore, it is crucial that analytical validation and qualification of an imaging biomarker is performed in a population appropriate for the clinical question at hand. For example, using RP specimens as a reference standard is common in texture analysis papers as it allows MR image registration to specific prostate areas. However, using primarily RP as a reference standard could result in an unacceptably high false positive rate of imaging features in the real diagnostic setting, where patients with low risk cancers or benign conditions are regularly seen. To overcome this, the ideal validation cohort for diagnostic biomarkers should include patients with a variety of underlying pathologies. Equally, features associated with pathological progression in active surveillance patients should be rigorously tested for reproducibility, repeatability and surrogacy in large, regularly imaged cohorts.

Adding one more layer of complexity to the overall problem, there is substantial heterogeneity in the way various authors address high data dimensionality. This is

a constant difficulty in imaging research as long lists of potential markers can be generated from a small number of patients. This means that false discovery rates often have to be tightly constrained through the selection of a subset of features and the use of statistical learning for parameter estimation rather than maximum likelihood (11). In their paper, Chaddad *et al.* extracted the best performing features from a highly dimensional dataset using non-parametric analysis of variance and correlation analysis, but extraction is sometimes done using methods such as sequential forward floating feature selection. This approach has been used by Litjens *et al.* to isolate computer-extracted features that distinguish cancer from benign confounding conditions (12). Alternatively, Wibmer and colleagues used generalized estimating equations for similar purposes, while other authors average features of the same type or resort to dimensionality reduction through analysis of principal components (6,13).

This variability contributes towards a general feeling of lack of standardization, which makes evidence synthesis extremely difficult and brings in mind the dictum: "If you torture the data long enough, it will confess". It would be a missed opportunity if quantitative imaging followed the footsteps of genetic biomarker research, where it has been repeatedly shown that in order to achieve even moderate overlap between two lists of predictive genes several thousand discovery samples are necessary, a standard that most published papers do not conform to (14). This is something that needs to be addressed and the imaging community is making considerable effort to devise road maps for imaging biomarker development in cancer research (15). Such roadmaps advocate parallel tracks for technical validation, increased standardization, continuous re-assessment of existing imaging biomarkers, the publication of all findings (including false-positive or false-negative), rigorous statistical methodology to avoid overfitting and the implementation of multicentre studies for biomarker qualification, however costly or complex.

In conclusion, computer-extracted texture features using a joint intensity rather than a simple grey-level co-occurrence matrix appear to be good at discriminating low from high-grade cancer on bi-parametric MRI. This could be a significant step towards calculating features in a way more consistent with the multiparametric approach currently used for prostate cancer risk stratification, especially if this calculation can be extended to incorporate all three mpMRI modalities. These results have to be corroborated by other authors and validated in large, multicentre cohorts with a wider spectrum of pathologies and clinical presentations, but they are encouraging and could also have clinical utility in the active surveillance setting.

Acknowledgements

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet 2017;389:815-22.

2. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. N Engl J Med 2018;378:1767-77.

3. Schoots IG, Petrides N, Giganti F, et al. Magnetic Resonance Imaging in Active Surveillance of Prostate Cancer: A Systematic Review. Eur Urol 2015;67:627-36.

4. Rosenkrantz AB, Ginocchio LA, Cornfeld D, et al. Interobserver Reproducibility of the PI-RADS Version 2 Lexicon: A Multicenter Study of Six Experienced Prostate Radiologists. Radiology 2016;280:793-804.

5. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern Syst 1973;SMC3:610-21.

6. Wibmer A, Hricak H, Gondo T, et al. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. Eur Radiol 2015;25:2840-50.

7. Sidhu HS, Benigno S, Ganeshan B, et al. Textural analysis of multiparametric MRI detects transition zone prostate cancer. Eur Radiol 2017;27:2348-58.

8. Kassner A, Thornhill RE. Texture Analysis: A Review of Neurologic MR Imaging Applications. Am J Neuroradiol 2010;31:809-16.

9. Chaddad A, Kucharczyk MJ, Niazi T. Multimodal Radiomic Features for the Predicting Gleason Score of

Prostate Cancer. Cancers (Basel) 2018;10. doi: 10.3390/cancers10080249.

10. Abramson RG, Burton KR, Yu JP, et al. Methods and challenges in quantitative imaging biomarker development. Acad Radiol 2015;22:25-32.

11. Pers TH, Albrechtsen A, Holst C, et al. The validation and assessment of machine learning: a game of prediction from high-dimensional data. PLoS One 2009;4:e6287.

12. Litjens GJ, Elliott R, Shih NN, et al. Computer-extracted features can distinguish noncancerous confounding disease from prostatic adenocarcinoma at multiparametric MR imaging. Radiology 2016;278:135-45.

13. Kuess P, Andrzejewski P, Nilsson D, et al. Association between pathology and texture features of multi parametric MRI of the prostate. Phys Med Biol 2017;62:7833-54.

14. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. Proc Natl Acad Sci 2006;103:5923-8.

15. O'Connor JP, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017;14:169-86.