May radiomic data predict prostate cancer aggressiveness?

Danila Germanese[4], Sara Colantonio[4], Claudia Caudai[5], Maria Antonietta Pascali[4], Andrea Barucci[2], Nicola Zoppetti[2], Simone Agostini[1], Elena Bertelli[1] Laura Mercatelli[1], Vittorio Miele[1], and Roberto Carpi[3]

[1] Azienda Ospedaliero Universitaria Careggi, Largo G. Brambilla 3, 50134 Florence, Italy
[2] "Nello Carrara" Institute of Applied Physics, IFAC-CNR, via Madonna del Piano 10, 50019 Florence, Italy
[3] Azienda USL Toscana Centro, Piazza Santa Maria Nuova 1, 50019, Florence, Italy
[4] ISTI-CNR, Pisa, Italy
[5] ITB-CNR, Pisa, Italy
{name.surname}@isti.cnr.it

Abstract. Radiomics can quantify tumor phenotypic characteristics non-invasively by defining a signature correlated with biological information. Thanks to algorithms derived from computer vision to extract features from images, and machine learning methods to mine data, Radiomics is the perfect case study of application of Artificial Intelligence in the context of precision medicine. In this study we investigated the association between radiomic features extracted from multi-parametric magnetic resonance imaging (mp-MRI)of prostate cancer (PCa) and the tumor histologic subtypes (using Gleason Score) using machine learning algorithms, in order to identify which of the mp-MRI derived radiomic features can distinguish high and low risk PCa.

Keywords: Machine Learning · Artificial Intelligence · Radiomics · Image Processing · Computer Vision · Prostate Cancer.

1 Introduction

In the paradigm of precision medicine, Radiomics is an -omic science, aiming at the improvement of diagnostic, prognostic, and predictive accuracy [1, 2].

Mining quantitative images features from clinical imaging, Radiomics uses advanced quantitative features to objectively and quantitatively describe tumor phenotypes. These features can be extracted from medical images using advanced mathematical algorithms [3] to discover tumor characteristics that may not be appreciated by the naked eye. Radiomic features can provide richer information about intensity, shape, size or volume, and texture of tumor phenotype that is distinct or complementary to that provided by clinical reports, laboratory test results, and genomic or proteomic assays.

Radiomics may thus provide great potential to capture important phenotypic information, such as intratumoral heterogeneity, subsequently providing valuable information for personalised therapy [4–7].

In this work, we aimed at implementing a machine learning-based automatic classification of PCa aggressiveness (Low-grade PCa vs. High-grade PCa) by using mp-MRI-based radiomic features. In particular we will focus on two different MRI maps, T2-weighted (T2w) MR imaging, and the Apparent Diffusion Coefficient (ADC) from diffusion-weighted MR imaging (DWI), both being valuable and well established parameters for differentiating PCa aggressiveness [17, 26–28].

PCa is among the most common cancers and the second leading cause of cancer-specific mortality among Western males, imposing a huge economic and social burden [8]. In general, patients with PCa and a Gleason Score[6] (GS) $<= 3+4$ (Low-grade PCa) have better survival rates, lower biochemical recurrence rate and lower prostate cancer-specific mortality in comparison to the patients with GS $>= 4+3$ (High-grade PCa) [9]. As a consequence, the early grading and stratification of PCa aggressiveness play a key role in the therapy management and in the evaluation of patient long-term survival.

Nevertheless, PCa aggressiveness assessed by biopsy may result in an incorrect diagnosis, in addition to patient discomfort. Moreover, GS evaluated from biopsies may differ from that assessed following radical prostatectomy due, for example, to an incomplete sampling [10–12]. Therefore, non-invasive and robust radiological image-based techniques that can help the clinicians in the evaluation of PCa aggressiveness are needed to enhance the quality of both clinical outcomes and patient care.

The role of machine learning techniques in analysing radiomic features have been investigated in many studies, e.g. for the discrimination of PCa from non-cancer prostate tissue [13–17], or in the classification of PCa with different GS [18, 19], or in the assessment of PCa aggressiveness [20]. In particular, texture-based radiomic features showed effectiveness in discriminating between cancer and non-cancer prostate tissue [21, 22] and in the assessment of PCa aggressiveness [23, 24].

Despite a huge amount of works it is important to highlight that there is not a unanimous consent about the specific radiomic signature that is most effective in distinguishing PCa aggressiveness. In our opinion the origins of this failure can be sought in the lack of standardised and robust data, in the use of small dataset which are usually unable to explain all the variability of the real samples. A solution to this phenomenon could be obtained using shared imaging

[6] The Gleason grading system is used to help evaluate the prognosis of men with prostate cancer using samples from a prostate biopsy. The pathologist looks at how the cancer cells are arranged in the prostate and assigns a score on a scale of 3 to 5 from 2 different locations. Please note the notation: the first number is the most common grade in all the samples, while the second number is the highest grade of whats left. Gleason Score = the most common grade + the highest other grade in the samples

biobanks. Datasets originating from a single institution can be very useful to test algorithms and to begin to understand which radiomic features can be the most representatives for PCa, but the definition of a radiomic signature with a strong clinical impact requires a different kind of dataset. The dimension of the imaging dataset is obviously directly related to the clinical problem of interest and at to kind of algorithm implemented.

In the presented work, we aimed at implementing a machine learning-based system to automatically classify PCa aggressiveness (Low-grade PCa vs. High-grade PCa). We compared the results obtained using *(i)* the whole set of 851 radiomic features (*first-order statistics*, *shape-based 3D features*, *shape-based 2D features*, *Gray level Cooccurence Matrix features*, *Gray level Run Length Matrix features*, *Gray level Size Zone Matrix features*, *Neighbouring Gray Tone Difference Matrix features*, *Gray level Dependence Matrix features* and their wavelet transform which yields 8 decompositions per level - all possible combinations of applying either a High or a Low pass filter in each of the three dimensions) and *(ii)* only those calculated on the original image (107, without wavelet filtering); and considering three dataset *(i)* T2w, *(ii)* ADC, *(iii)* T2w + ADC.

The paper is organised as follows: Section 2 describes the whole radiomic process (image acquisition, image segmentation, feature extraction and selection, analysis and model building); Section 3 reports the achieved results; Section 4 concludes the paper.

2 METHODS AND MATERIALS

2.1 Patient Cohort

This retrospective study involved 125 patients who underwent a 1.5 T mp-MRI and free hand transperineal MRI/US fusion-guided targeted biopsy (MyLab-TM Twice Esaote).

From such cohort of patients, we selected 50 peripheral zone PCa patients for our pilot study, with a PI-RADS score[7] 3-5, corresponding to an intermediate-to-very high probability of malignancy. 57 lesions were biopsed and the histopathological result was as follow: 37 with GS $<= 3+4$, consistent with a less aggressive behaviour of the prostate cancer, and 20 with GS $>= 4+3$.

2.2 Image Acquisition

In this study all exams were performed using a 1.5 T MR scanner (Magnetom Aera, Siemens Healthcare, Erlangen, Germany) equipped with a pelvic phased-array 32-channels coils (Fig. 1).

[7] The PI-RADS v2 [25] (Prostate Imaging Reporting & Data System) assessment categories are based on the findings of mp-MRI, combining T2-weighted (T2W), diffusion weighted imaging (DWI) and dynamic contrast-enhanced (DCE) imaging. The PI-RADS assessment category determines the likelihood of clinically significant prostate cancer. A score, ranging from 1 to 5, is given accordingly to each imaging technique, with 1 being most probably benign (clinically significant cancer is highly unlikely to be present) and 5 being high suspicious for malignancy

Our acquisition protocol included:

- High-resolution T2w sequences in the axial (voxel size 0.6 x 0.6 x 3.0 mm), sagittal and coronal planes (voxel size 0.7 x 0.7 x 3.0 mm);
- T1w pre-contrast sequence in the axial plane (voxel size 0.8 x 0.8 x 5 mm);
- a multi-b DWI (range $0 - 2000$ s/mm^2, step of 500 s/mm^2, voxel size 0.8 x 0.8 x 3mm) EPI sequence from which corresponding ADC maps were automatically calculated using software on board the Siemens MRI console;
- Dynamic Contrast Enhancement (DCE) assessment with time intensity curves evaluation.

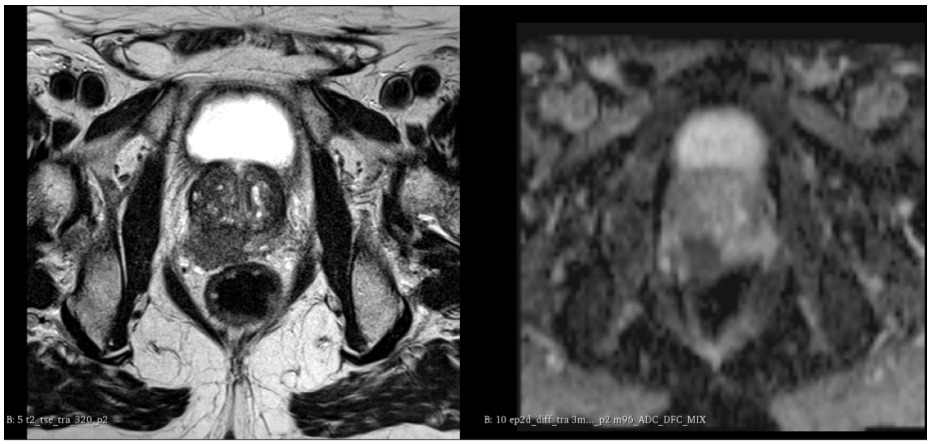

Fig. 1. Example of prostate mp-MRI images. Left: T2w; Right: ADC. Please note that slices are different in the 2 maps

2.3 Image Segmentation

Segmentation was performed on the two most representative sequences for PI-RADS assessment in clinical practice, T2w images and the ADC maps derived from the Diffusion Weighted Imaging (DWI).

Tumor regions were defined by manually drawing ROIs using the 3D Slicer software [29]. For consistency between ROIs, all depicted lesions were strictly segmented with the same criteria and visually validated by three radiologists (with different experience in reporting prostate mp-MRI (15, 5 and 1 year respectively) in consensus, both on T2w images and ADC maps (Fig. 2, Fig. 3).

2.4 Feature Extraction

Quantitative features were extracted both from original images and after applying wavelet transform for T2w dataset and for ADC dataset. All the feature

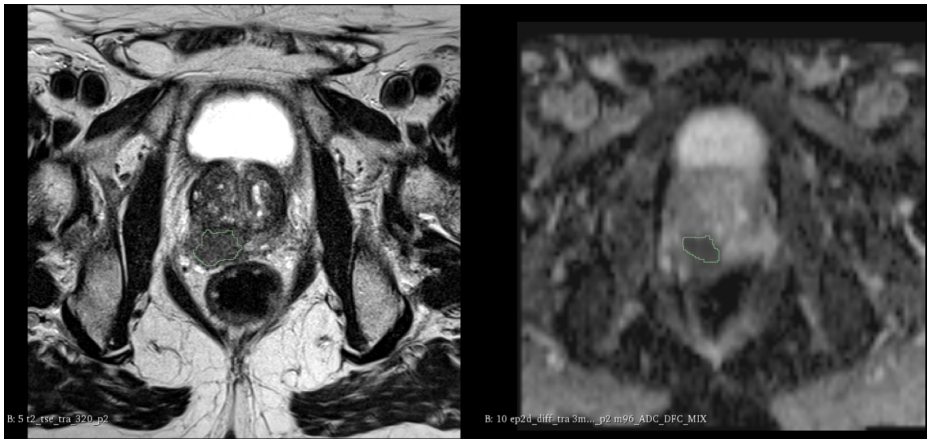

Fig. 2. Example of prostate mp-MRI images ROIs segmentation. Left: T2w; Right: ADC. The green line defines the border of the tumor.

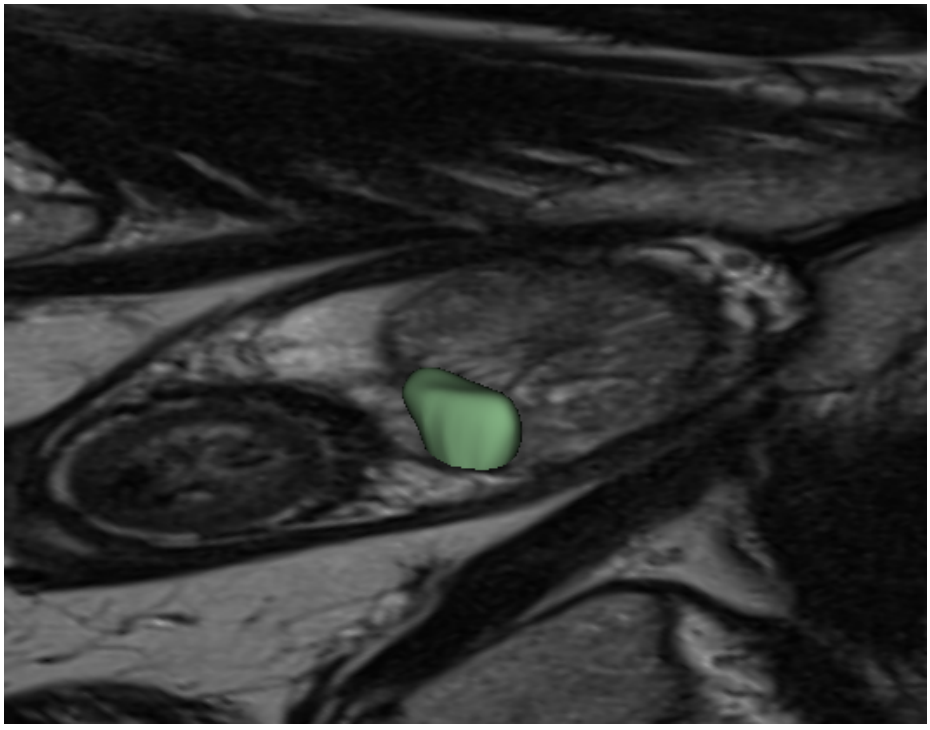

Fig. 3. Example of prostate mp-MRI image 3D segmentation showing the entire tumour volume used for radiomic analysis.

classes were computed: *shapes features*, *first- order statistics features*, *second-order statistics features* (that included the so called *texture features*) and *higher-order statistics features*, for a total of 851 features.

The features were evaluated using a home-made software based on the open-source python package pyradiomics [30].

2.5 Feature Selection and Classification

The analyses, implemented on MATLAB® R2018 platform, were carried out by considering the whole set of features (851) and only those calculated on the original image (107, without wavelet transform), on three dataset: T2w, ADC, T2w+ADC. Radiomics raw data were firstly normalized across all patients by using quantile normalization.

Then, for each dataset, a correlation analysis was run to detect redundancy. Pearson's correlation coefficient was calculated, and one feature was dropped from those pairs of features showing high correlation (> 0.95, p-value < 0.05) and, hence, more linear dependence.

A feed-forward feature selection method was applied to select the most discriminative radiomic features. A predictive model was devised to distinguish low-grade (GS $<= 3 + 4$) from intermediate/high-grade (GS $>= 4 + 3$) PCa. A non-linear Support Vector Machine (SVM) was used as the classifier. Starting from an empty feature set, the implemented selection method created candidate feature subsets by sequentially adding each of the features not yet selected. At each step, 10-fold cross-validation was applied to get the prediction accuracy for each candidate feature subset. The process was repeated until the criterion value (that is, the mis-classification error) reached the global minimum.

3 RESULTS

3.1 Radiomic signatures building

In Table 1, the built radiomic signatures are shown for each dataset (T2w, ADC, T2w+ADC) and according to the set of features that was considered for the analysis, that means, the whole set of 851 features (F_{851}) and the features computed only on the original image (F_{107}).

3.2 Diagnostic performance of radiomic signatures

All the built radiomic signatures were used to train a non-linear Support Vector Machine (SVM) classifier. It was trained on 40 cases (26 GS $<= 3+4$ and 14 GS $>= 4+3$) and tested on the remaining 17 (11 GS $<= 3+4$ and 6 GS $>= 4+3$).

As shown in Figure 4, the best performance was obtained with the T2w+ADC radiomic signature built including the wavelet parameters, with an overall accuracy of 94.12%, 100% sensitivity, 90,9% specificity (just one case misclassified).

We obtained in the other cases: 88.23% accuracy (60% sens, 100% spec) for T2w+ADC (9 features without wavelet); 88.23% accuracy (71,42% sens, 83,33%

Table 1. The computed radiomic signatures and criterion values (CV) for each dataset.

	$\mathbf{F}_{851}$	$\mathbf{F}_{107}$
T2w	wavelet-LHL glszm Zone Entropy wavelet-LHH glcm Joint Entr. wavelet-HLL glszm Size Zone Non Unif. original glcm Idmn wavelet-LHL first ord. Root Mean Sq. orig. glcm Sum Entropy wavelet-LLH first ord. Entr. CV: 0.086	orig. first ord. Total Energy orig. glszm Size Zone Non Unif. Normal. orig. shape Max. 2D Diameter Row original glcm Idmn orig. ngtdm Strength orig. gldm Large Dep. High Gray Lev. Emph. orig. glrlm Long Run High Gray Lev. Emph. orig. glszm Size Zone Non Unif. orig. gldm Dependence Non Unif. Normal. orig. first ord. Energy orig. glcm Correlation orig. glcm Idm orig. glcm Sum Entropy CV: 0.069
ADC	wavelet-LHH glrlm Run Len. Non Unif. wavelet-LHH first ord. Tot. Energy wavelet-LLL glcm Correlation wavelet-LHH first ord. Root Mean Sq. wavelet-HHL glszm Gray Lev. Non Unif. wavelet-LLL ngtdm Contrast wavelet-HLL first ord. Root Mean Sq. wavelet-LLH glcm Id wavelet-HLH first ord. 10 Percent. CV: 0.024	orig. first ord. Entr. orig. ngtdm Contrast orig. shape Minor Axis Length orig. first order Uniformity orig. ngtdm Complexity orig. glcm Contrast orig. glcm Diff. Average orig. shape Least Axis Length orig. glrlm Long Run High Gray Lev. Emph. orig. glcm Joint Entropy orig. glrlm High Gray Lev. Run Emph. orig. glcm Difference Variance orig. gldm Small Dep. Low Gray Lev. Emph. orig. glszm Size Zone Non Unif. CV: 0.155
T2w + ADC	wavelet-HLL glcm Joint Energy orig. glszm Size Zone Non Unif. Nor. wavelet-HLL gldm Dep. Non Unif. wavelet-LHH first ord. Tot. En. wavelet-HLL glcm Imc1 wavelet-LLH glcm Correlation wavelet-HLH first ord. Range wavelet-LHL glszm Gray Leve. Non Unif. wavelet-LLH glcm Correlation wavelet-HLL gldm Low Gray Lev. Emph. CV: 0.000	orig. glcm Idmn orig. shape Minor Axis Length orig. ngtdm Complexity orig. glszm Size Zone Non Unif. orig. glcm Diff. Average orig. glcm MCC orig. glszm Gray Lev. Non Unif. Normal. orig. first ord. Maximum orig. glrlm Gray Level Non Unif. CV: 0.053

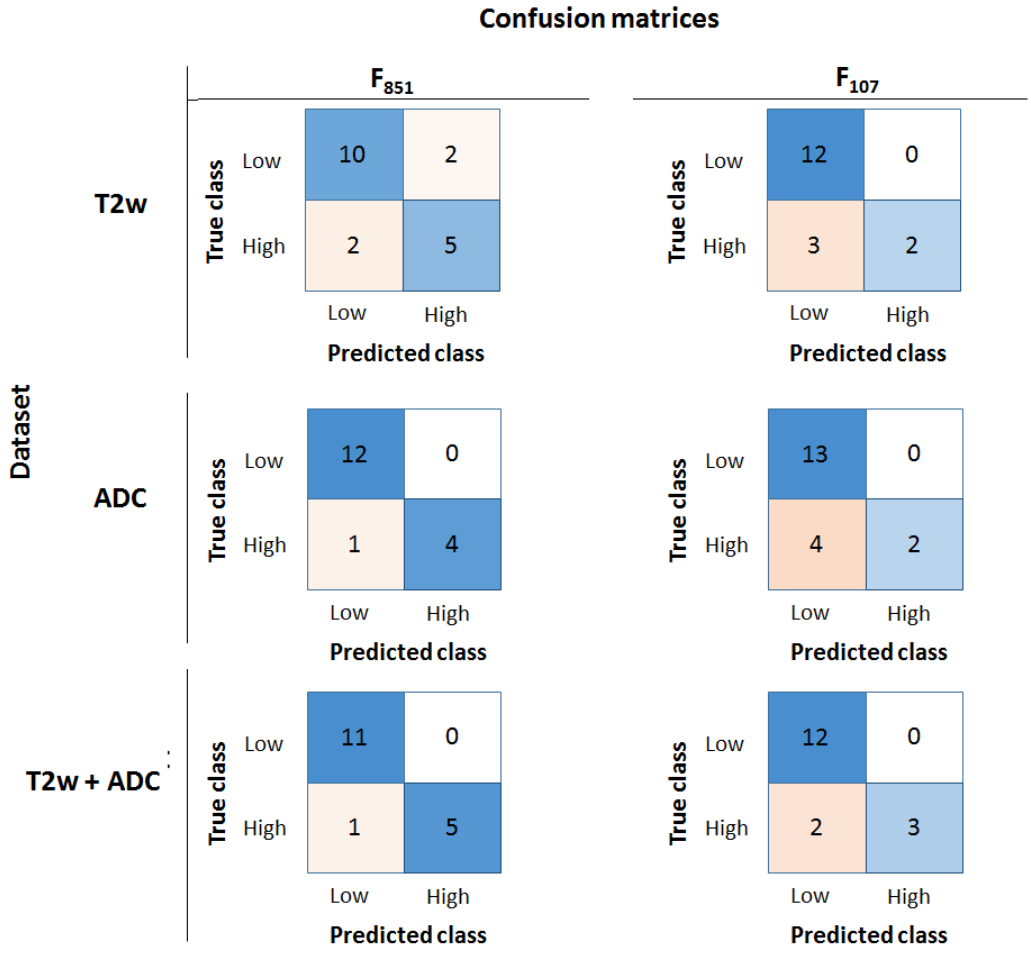

Fig. 4. Confusion matrices for all the investigated cases.

spec) for T2w (7 features with wavelet); 82.35% accuracy (60% sens, 100% spec) for T2w (13 features without wavelet); 94.11% accuracy (80% sens, 100% spec) for ADC (9 features with wavelet); 78.94% accuracy (50% sens, 100% spec) for ADC (14 features without wavelet).

4 CONCLUSION

In this study, we evaluated the potential role of radiomic features in predicting the aggressiveness of prostate cancer compared with bioptic Gleason score. We compared the prediction power of six radiomic signatures, selected from three dataset (T2w MRI-based radiomic features dataset, ADC MRI-based radiomic features dataset, and the combination of both) and using both the whole set of computed features, that integrated also the ones computed on the wavelet transformed images, and the set of features that included the features calculated on the original images only.

The ADC dataset with the whole set of features gave good accuracy in discriminating between high vs low risk PCa. Also, the combination of ADC and T2w radiomic features, along with the inclusion of wavelet filtering, seemed to add discriminative information to the lesions classification.

The idea would be to ground on the latter result and build a radiomic signature which include both ADC and T2w radiomic features, in accordance with the fact that also PI-RADS assessment uses a combination of mp-MRI T2W and DWI findings. However, a deeper investigation will be carried on a larger, multicentre dataset with a more balanced distribution to confirm such results.

The identification of a robust and validated radiomic signature would be fundamental to move precision medicine forward. Indeed, in combination with other omics data, radiomic signatures can then be used for the development of diagnostic and prognostic models, describing phenotypic patterns connected to biological or clinical end points, aiming at tailoring of the therapies based on patients needs and at the monitoring of the response to care.

References

1. Lambin, P., and Leijenaar, R. T. H. and Deist, T. M. and Peerlings, J. and de Jong, E. E. C. and van Timmeren, J. and Sanduleanu, S. and Larue, R T. H. M. and Even, A. J. G. and Jochems, A. and van Wijk, Y. and Woodruff, H and van Soest, J. and Lustberg, T. and Roelofs, E. and van Elmpt, W. and Dekker, A. and Mottaghy, F. M. and Wildberger, J. E. and Walsh, S., Radiomics: the bridge between medical imaging and personalized medicine, Nature Reviews Clinical Oncology,**14**: 749 EP – (2017), https://doi.org/10.1038/nrclinonc.2017.141
2. Robert J. Gillies, Paul E. Kinahan, and Hedvig Hricak, Radiomics: Images Are More than Pictures, They Are Data, Radiology, **278**2: 563 – 577 (2016), https://doi.org/10.1148/radiol.2015151169
3. Larue, R.T.H.M. et al., Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thiknesses: a comprehensive phantom study. In: Acta Oncol., vol. 56, pp. 1544-1553 (2017).

4. Barucci, A., Farnesi, D., Ratto, F., Pini, R., Carpi, R., Olmastroni, M., Materassi, M., Romei, C., Taliani, A., Exposing Cancer's Complexity Using Radiomics in Clinical Imaging. An Investigation on the Role of Histogram Analysis as Imaging Biomarker to Unravel Intra-Tumour Heterogeneity. In: 2018 IEEE Workshop on Complexity in Engineering (COMPENG), 2018, pp. 1-5. https://doi.org/10.1109/CompEng.2018.8536244
5. Stoyanova R, Takhar M, Tschudi Y, Ford JC, Solrzano G, Erho N, Balagurunathan Y, Punnen S, Davicioni E, Gillies RJ, Pollack A. Prostate cancer radiomics and the promise of radiogenomics. Transl Cancer Res 2016;5(4):432-447. https://doi.org/10.21037/tcr.2016.06.20
6. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014; 5: 4006. https://doi.org/10.1038/ncomms5006
7. Avanzo, M., Stancanello, J., El Naga, I.: Beyond imaging: The promise of radiomics. Physica Medica **38**, 122–139 (2017), https://doi.org/10.1016/j.ejmp.2017.05.071
8. Bray F. et al., Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. In: CA: A Cancer Journal for Clinicians, vol. 68, pp. :394424 (2018).
9. Ahmed H.U. et al., Transatlantic Consensus Group on Active Surveillance and Focal Therapy for Prostate Cancer. In: BJU Int., vol. 109, pp.: 1636-1647 (2012).
10. King, C.R. and Long, J.P., Prostate biopsy grading errors: a sampling problem? In: Int J. Cancer, vol. 90, pp.: 326-330 (2000).
11. Epstein J.I., Feng, Z., Trock, B.J. and Pierorazio, P.M., Upgrading and downgrading of prostate cancer from biopsy to radical prostatectomy: incidence and predictive factors using the modified Gleason grading system and factoring in tertiary grades. In: Eur Urol, vol. 61, pp.: 1019-1024 (2012).
12. Berglung, R.K. et al., Pathological upgrading and up staging with immediate repeat biopsy in patients elegible for active surveillance. In: J. Urol, vol. 180, pp.: 1964-1967 (2008).
13. Peng, Y. et al, Quantitative analysis of multiparametric prostate MR images: Differentiation between prostate cancer and normal tissue and correlation with Gleason scorea computer-aided diagnosis development study. In: Radiology, vol. 267, pp.: 787-796 (2013).
14. Tiwari, P., Viswanath, S., Kurhanewicz, J., Sridhar, A. and Madabhushi, A., Multimodal wavelet embedding representation for data combination (MaWERiC): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. In: NMR Biomed., vol. 25, pp.: 607-619 (2012).
15. Moradi, M. et al., Multiparametric MRI maps for detection and grading of dominant prostate tumors. In: J. Magn. Reson. Imaging, vol. 35, pp.: 1403-1413 (2012).
16. Barucci, A., Bastiani, P., Carpi, R., Fondelli, S., Giannetti, A., Olmastroni, M., Pini, R., Ratto, F., Rucco, M., Zatelli, G., Esposito, M. 301. Prostate cancer Radiomics using multiparametric MR imaging: an exploratory study. In: Proc. of 10th Congress of the Associazione Italiana di Fisica Medica - AIFM, Physica Medica: European Journal of Medical Physics, vol. 56, pp. 246 ,Elsevier, (2018). https://doi.org/10.1016/j.ejmp.2018.04.310
17. Mazaheri, Y. et al., Prostate cancer: Identification with combined diffusion weighted MR imaging and 3D 1H MR spectroscopic imaging-correlation with pathologic findings. In: Radiology, vol. 246, pp.: 480-488 (2008).
18. Wibmer, A. 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. In: Eur. Radiol., vol. 25, pp.: 2840-2850 (2015).

19. Fehr, D. et al., Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. In: Proc. Natl. Acad. Sci U S A., vol. 112, pp.: 6265-6273 (2015).
20. Chen, T. , Li, M. , Gu, Y. , Zhang, Y. , Yang, S. , Wei, C. , Wu, J. , Li, X. , Zhao, W. and Shen, J., Prostate Cancer Differentiation and Aggressiveness: Assessment With a RadiomicBased Model vs. PIRADS v2. J. Magn. Reson. Imaging, **49**: 875–884 (2019), https://doi.org/10.1002/jmri.26243
21. Sidhu, H.S. et al., Textural analysis of multiparametric MRI detects transition zone prostate cancer. In: Eur Radiol., vol. 27, pp.: 1-11 (2017).
22. Khalvati, F., Wong, A., Haider, M.A., Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature model. In: BMC Med. Imaging, vol. 15, pp.:27 (2015).
23. Vignati, A. et al., Texture features on T2-weighted magnetic resonance imaging: New potential biomarkers for prostate cancer aggressiveness. In: Phys Med Biol, vol. 60, pp.: 2685-2701 (2015).
24. Nketiah, G. et al., T2-weighted MRI-derived textural features reflect prostate cancer aggressiveness: Preliminary results. In: Eur Radiol, vol. 27, pp.: 3050-3059 (2016).
25. Weinreb, J.C., et al. PI-RADS prostate imaging - reporting and data systems: 2015, version 2. In: Eur. Urol.; 69:16-40 (2016).
26. Langer, D.L. et al., Prostate tissue composition and MR measurements: Investigating the relationships between ADC, T2, K(trans), v(e), and corresponding histologic features. In: Radiology, vol. 255, pp.: 485-494 (2010).
27. Oto, A. et al., Diffusion-weighted and dynamic contrast-enhanced MRI of prostate cancer: Correlation of quantitative MR parameters with Gleason score and tumor angiogenesis. In: AJR Am J. Roentgenol, vol. 197, pp.: 13821390 (2011).
28. Nagarajan, M.B. et al., Classification of small lesions in breast MRI: Evaluating the role of dynamically extracted texture features through feature selection. In: J. Med. Biol. Eng., vol. 33, pp.:33 (2013).
29. Fedorov, A., et al. 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. In: Magnetic Resonance Imaging, 30(9):1323-41 (2012).
30. van Griethuysen, J. J. M., et al. Computational Radiomics System to Decode the Radiographic Phenotype. In: Cancer Research, 77(21), e104e107 (2017).