Development of a MRI radiomic-based ML model to predict aggressiveness of prostate cancer

Odette Rios-Ibacache [1,3] Paola Caprile [1,3] José Dominguez [2,3] Cecilia Besa [2,3]

[1]Department of Physics, Pontifical Catholic University of Chile [2]Department of Radiology, School of Medicine, Pontifical Catholic University of Chile [3]Millennium Institute for Intelligent Healthcare Engineering - iHEALTH, Santiago, Chile

Introduction

The gold standard for the evaluation of prostate cancer (PCa) aggressiveness is the Gleason score (GS), which requires a histopathological analysis to discriminate between clinically significant (CS, $GS \leq 7$) and non-significant (non-CS, $GS = 6$) cases. The aim of this study was to develop a non-invasive tool able to predict the GS classification of PCa, based on the information extracted from multiparametric magnetic resonance imaging (mpMRI) and clinical data, by using machine learning (ML) tools.

Materials and Methods

This retrospective cohort included 86 adult male patients with positive biopsy for PCa, made by fusion technique (mpMRI-ultrasound) at Clinical Hospital UC-Christus between 2017 and 2021, with lesions greater than 5 mm. 2D segmentations of the target prostate lesions were made by experienced radiologists in T2 weighted (T2w)/Apparent Diffusion Coefficient (ADC) map images at a 3T scanner. Clinical indicators of the study cohort are summarized in Table 1.

x	$\bar{x} \pm \sigma$, [*range*], *median*
Age (*years*)	60.0 ± 8.2, [45.0,88.0],67.0
PSA (*ng/mL*)	8.6 ± 6.1, [0.6,34.5],7.2
PSA Density (ng/mL^2)	0.2 ± 0.2, [0.01, 0.9],0.2
PV (*mL*)	48.0 ± 25.2, [15.0, 175.0], 43.5
Pirads v2	***N %***
3	2.3
4	69.7
5	27.9
Gleason Score	***N %***
6	23.3
3 + 4	32.5
4 + 3	24.4
8	5.8
9	13.9

Table 1. Relevant clinical features of the study cohort.

A radiomic analysis was performed considering first order, textural and shape features, besides clinical/qualitative image information, such as PIRADS-v2. Splitting the dataset on train/test (80%) and validation sets (20%), univariate and multivariate models were built using manual and automatic feature selection algorithms. In order to evaluate the performance of the models, twofold cross-validation (CV) was employed. In particular, we used the Repeated Stratified KFold CV technique with 1000 repetitions, AUC as the performance metric. Different classifiers such as Logistic Regression (LR), K-Nearest Neighbors (KNN), Naive Bayes (NB), Support Vector Machine (SVM), and CART were evaluated.

The manual selection method was based on individual feature performance and correlation, using parametric and non-parametric statistical hypothesis tests, Pearson correlation, and predictive power with bootstrap AUC analysis. A comparison between models was performed using Frequentist and Bayesian correlated t-tests. This process is showed in Fig. 2.

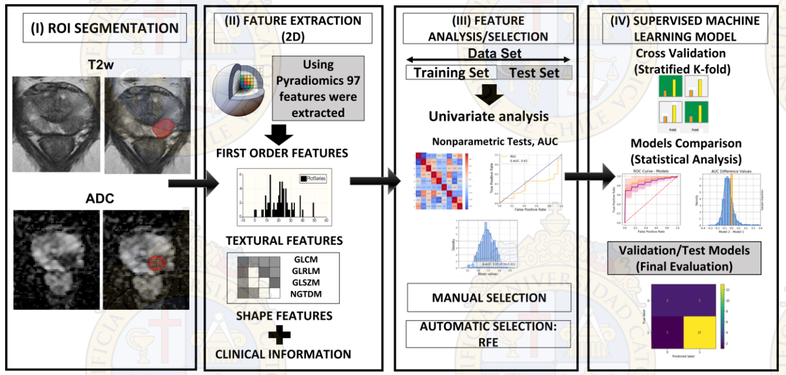

Fig. 1. Scheme of the MRI texture analysis and radiomics workflow, for patients with PCa. 2D features were extracted from ROIs from T2w and ADC images.

Results

The best model found was multivariate, obtained using the automatic feature selection algorithm RFE, with LR as estimator with 9 features, including image (T2w and ADC) and clinical info. The train/test mean AUC was 0.91 ± 0.06, with a validation AUC of 0.91 for a classification of high-lower aggressiveness (GS≥7 vs GS=6). The Bayesian tests confirmed that our best model performed better than the best univariate and multivariate models considering only image features or clinical info., with probability values of 0.95, 0.69 and 0.78, respectively.

Data set	N^o	$AUC_{\bar{x}}$ (σ) 95%C.I.	AUC_{val}	acc_{val}
T2w	5	0.85 (0.06) [0.72 - 0.94]	0.54	0.72
ADC	6	0.82 (0.09) [0.57 - 0.94]	0.71	0.89
CL	4	0.84 (0.07) [0.71 - 0.97]	0.84	0.72
T2w + CL	5	0.88 (0.06) [0.75 - 0.97]	0.88	0.83
ADC + CL	6	0.86 (0.07) [0.71 - 0.96]	0.55	0.83
T2w+ADC	6	0.87 (0.05) [0.77 - 0.95]	0.55	0.72
T2w+ADC+CL	9	0.91 (0.06) [0.75 - 0.99]	0.91	0.83

Table 2. Test and validation results of the best models selected for each dataset. LR as best classifier for all.

We also applied the Bayesian Test, between our best model with PIRADS-v2 and PSA-D based model, resulting in 0.85 and 0.90. As Figure 2 shows.

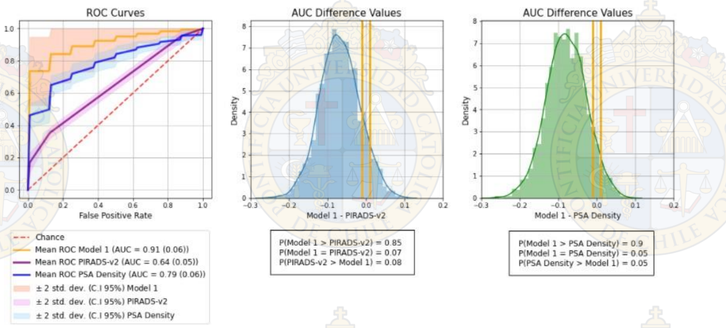

Fig. 2. Receive Operator Curves (left) and probabilities distribution (right) of 1000 repetitions of KFold CV classification. Orange vertical bars indicate the interval of 0.01 of difference. Where Model 1 represents the best multivariate model incorporating clinical and imaging features.

Conclusions

This study confirmed the existence of radiomic information with a high potential to predict the aggressiveness. Combining MRI-based radiomic and clinical/qualitative information can significantly improve the model performance to classify PCa aggressiveness. An additional cohort will be required to evaluate the applicability of this tool in a multi-center and multi-scanner setting.

Limitations and Future Work

Our model has limitations that affects the robustness of the carried work, one of them is the reduced amount (86) of patients. Another one is the lack of an external cohort to test them, to decide considerably from the expected values (C.I). Due to some testing scores showed in Table 2 are around 0.5, as future work we plan to reevaluate the data to know if the first split was not representative of the distribution on every set. Other important point to take into account for a future improvement and robustness acquisition of the dataset, and consequently, the construction of the model, is the variation of the features among consecutive MRI scans. Our next step is analyze the robustness using data augmentation and the incorporation of an automatic segmentation algorithm of lesions based on DL tools to know how both affect to the models.

References

1 Fehr, D.et al (2015). Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. Proceedings of the National Academy of Sciences of the USA.

2 Kuhn, M. Johnson, K. (2013). Applied Predictive Modeling. Springer.

3 Stanzione, A.et al (2020). Prostate MRI radiomics: A systematic review and radiomic quality score assessment. European journal of radiology.

4 Makowski Marcus R. (2021). De Novo Radiomics Approach Using Image Augmentation and Features From T1 Mapping to Predict Gleason Scores in Prostate Cancer. Investigative Radiology.

5 Sun, Y. et al (2019). Automatic stratification of prostate tumour aggressiveness using multiparametric MRI: a horizontal comparison of texture features. Acta oncologica (Stockholm, Sweden)

6 Vignati, A., et al. (2015). Texture features on T2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness. Physics in medicine and biology.