

ORIGINAL ARTICLE

Machine Learning Techniques in Prostate Cancer Diagnosis According to Prostate-Specific Antigen Levels and Prostate Cancer Gene 3 Score

Roberto Passera¹, Stefano De Luca², Cristian Fiori², Enrico Bollito³, Francesco Porpiglia²

¹Division of Nuclear Medicine, Department of Medical Science, University of Torino, San Giovanni Battista Hospital, Torino, Italy

²Division of Urology, Department of Oncology, University of Torino, San Luigi Gonzaga Hospital, Orbassano, Italy

³Division of Pathology, Department of Oncology, University of Torino, San Luigi Gonzaga Hospital, Orbassano, Italy

Purpose: To explore the role of artificial intelligence and machine learning (ML) techniques in oncological urology. In recent years, our group investigated the prostate cancer gene 3 (PCA3) score, prostate-specific antigen (PSA), and free-PSA predictive role for prostate cancer (PCa), using the classical binary logistic regression (LR) modeling. In this research, we approached the same clinical problem by several different ML algorithms, to evaluate their performances and feasibility in a real-world evidence PCa detection trial.

Materials and Methods: The occurrence of a positive biopsy has been studied in a large cohort of 1,246 Italian men undergoing first or repeat biopsy. Seven supervised ML algorithms were selected to build biomarkers-based predictive models: generalized linear model, gradient boosting machine, eXtreme gradient boosting machine (XGBoost), distributed random forest/extremely randomized forest, multilayer artificial Deep Neural Network, naïve Bayes classifier, and an automatic ML ensemble function.

Results: All the ML models showed better performances in terms of area under curve (AUC) and accuracy, when compared to LR model. Among them, an XGBoost model tuned by the autoML function reached the best metrics (AUC, 0.830), well overtaking LR results (AUC, 0.738). In the variable importance ranking coming from this XGBoost model (accuracy, 0.824), the PCA3 score importance was 3-fold and 4-fold larger, when compared to that of free-PSA and PSA, respectively.

Conclusions: The ML approach proved to be feasible and able to achieve good predictive performances with reproducible results: it may thus be recommended, when applied to PCa prediction based on biomarkers fluctuations.

Key Words: Prostatic neoplasms, Prostate biopsy, PCA3 score, Prostate-specific antigen, Machine learning

Received May 17, 2021

Revised June 2, 2021

Accepted June 30, 2021

Corresponding author:

Roberto Passera

Email: passera.roberto@gmail.com

https://orcid.org/0000-0003-3372-2248

INTRODUCTION

Screening using prostate-specific antigen (PSA) is characterized by low specificity for prostate cancer (PCa), since elevated PSA may be due to benign conditions, especially within a PSA range of 4–10 ng/mL. Only one fourth

of men with PCa suspicion go on to have a positive biopsy.¹ Prostate cancer gene 3 (PCA3), first described by Bussemakers et al.² in 1999, is a noncoding, prostate-specific mRNA highly overexpressed in 95% of PCa cells, with a median 66-fold upregulation compared with adjacent nonneoplastic prostatic cells. As the name implies,

Copyright © The Korean Urological Oncology Society and The Korean Prostate Society. All rights reserved.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

PCA3 is specific for PCa and is expressed only in this disease. Since 2012, PCA3 was approved as an auxiliary biomarker in the molecular diagnosis of PCa in the European Union, Canada, and the United States. Many studies investigated the diagnostic value of urine PCA3 in PCa, but results regarding its applicability in a clinical setting (first and/or repeat biopsy) have been inconclusive.³

Recently, in the era of big data, artificial intelligence technology, and machine learning (ML) techniques have been applied to analyze large amounts of data in medical field, and their adequacy and usefulness in diagnosis are increasing.⁴⁻⁶ This approach is playing an emerging role even in urology: pattern recognition and classification, confounder discrimination, cancer new markers identification and computer-assisted diagnosis, image processing and radiomics, computational biology, new surgical techniques validation, bridging clinical data with histopathological and genetics/genomics ones to build up a data warehouse.^{7,8} When dealing with PCa, ML can be applied to assist several procedures, like capsule segmentation, fusion-targeted biopsy, robotic-assisted surgical systems, just to digital pathology and automatic diagnostics.⁹⁻¹¹

In recent years, one of our main interests focused on the predictive role of PCA3 score for PCa, when combined with classical risk factors like PSA and free-PSA (%fPSA). Our first experience investigated this biomarker on a large real-world cohort of Italian men undergoing first or repeat biopsy for PCa. At that time, we used the logistic regression (LR) modeling to predict PCa detection rate at different PCA3 score values.¹² Actually, no studies evaluated the PCa prediction role of age, PSA, and PCA3 score using ML methods. The aim of the current research is to improve our past results by a modern approach, now proposing the use of several supervised ML algorithms to build biomarkers-based predictive models for PCa

diagnosis.

MATERIALS AND METHODS

The original study took place in 3 Italian institutions (San Luigi Gonzaga Hospital Orbassano, Gradenigo Hospital Torino, and San Raffaele Hospital Milano) and recruited 3,571 men, who consecutively underwent PCA3 testing between October 2008 and December 2010. A total of 3,446 urine samples (96.5%) had adequate levels of PCA3 and PSA mRNAs to calculate the PCA3 score. All patients (n=1,246, 36.1%) who underwent ≥ 1 biopsy after PCA3 assessment as of December 31, 2010, were enrolled. Seven hundred and thirty-one subjects had their first biopsy due to a serum PSA ≥ 2.5 ng/mL after ruling out the presence of urinary tract infections and/or inflammation with clinical history, urine cultures and digital rectal examination (DRE); the remaining 515 ones had 1 or 2 previous negative biopsies and underwent repeat biopsy due to PSA elevation persistency. The current study is a reanalysis of the original cohort dataset by ML techniques, without any impact either on patient's clinical history or future treatment decision. Due to the retrospective observational nature of this research and according to Italian law (Agenzia Italiana del Farmaco-AIFA, Guidelines for observational studies, March 20, 2008), no formal approval from the local Institutional Review Board/Independent Ethics Committee was needed.

1. Statistical Analysis

At first, the determinants for a positive biopsy (dependent variable, target) have been estimated by the multivariate binary LR model. Eight predictors (independent variables, features) were tested as PCa risk factors: 4 continuous (age, PSA, %fPSA, and PCA3 score) and 4 categorical

(family history for PCa, DRE, high-grade prostatic intraepithelial neoplasia [HG-PIN]). The continuous variables were reported as median-interquartile range (IQR) while the categorical ones as absolute/relative frequencies. Two different inferential tests were applied, the Mann-Whitney and the Fisher exact test, for continuous and categorical covariates respectively. All reported p-values were obtained by the 2-sided exact method at the conventional 5% significance level. Data were analyzed as of February 2021 using R 4.0.5 package H2O version 3.32.1.1 (R Foundation for Statistical Computing, Vienna, Austria).¹³

2. Development and Validation of ML Models

At a second step, 6 different supervised ML algorithms for binomial classification were trained and cross-validated for target prediction (biopsy result for PCa) using the same 8 features; these estimation processes have been performed by H2O for R, an open-source distributed ML platform.¹⁴ The ML algorithms were generalized linear model (GLM), gradient boosting machine (GBM), eXtreme Gradient Boosting machine (XGBoost), distributed random forest (DRF)/eXtremely randomized forest (XRT), multilayer artificial Deep Neural Network (DNN), and naïve Bayes classifier (NB).¹⁵⁻²⁰ Moreover, the modeling process has been performed by H2O AutoML too, an automatic supervised ML ensemble function that sequentially trains, cross-validates and tunes an ordered series of ML models, ranking them by performance metrics: 3 XGBoosts, a fixed grid of GLMs, 1 DRF, 5 GBMs, 1 DNN, 1 XRT, a random grid of XGBoost, and 2 Stacked Ensemble models too, the former containing all the models, the latter only the best ones from each algorithm class.²¹ Automatic ML algorithm searches for the optimal combination of a collection of prediction

algorithms stacking together various classifiers and it is considered among the newest frontiers for ML, often challenging the predictions deriving from a manual ML hyperparameters tuning.

For all the models, the target was balanced in the training and test data via resampling (either oversampling the minority class or undersampling the majority one) and the missing values were replaced by the Multiple Imputation by Chained Equations procedure.²² Sample size is quite critical in ML modeling: not to lose statistical power and conversely from our 2012 study, we investigated the whole 1,246 patients (instances) cohort, disregarding if they underwent either first or repeat biopsy. Therefore, the original dataset was randomly splitted for train into 80% training frame and 20% test one. After the training phase and to decrease the risk of model overfitting, a 5-fold cross-validation was used to compare the classifiers and produce a single estimation: the training frame was split into 5 folds, using 4 of them for training and 1 for cross-validation, replying 5 times with each fold used once as a test frame. Model performances have been investigated on the test set and the whole training/cross-validation/test procedure has been replied 20 times for estimation stability, each time using a different training/test split partitioning. The best prediction performance was identified by the area under curve (AUC) of the receiver operating characteristic (ROC) curve. The ROC curve shows the trade-off between false positive rate and true positive rate and its AUC allows to compare learning algorithms for binary classification better than accuracy. The AUC represents the likelihood that a positive case (patient with positive biopsy) is ranked higher than a negative one (patient with negative biopsy), considering all possible thresholds: the higher AUC, the better PCa detection performance.²³ Conversely, the accuracy measures, for a given threshold (0.5 by default),

the percentage of correctly classified cases, regardless of which class (negative or positive biopsy) they belong to. Dealing with binary classification, AUC is used to evaluate how well a model is able to distinguish between true positives and false positives, while accuracy to estimate the number of correct predictions made as a ratio of all predictions.

RESULTS

The main patients' characteristics are reported in Table 1. Seven hundred fifty cases had complete data, while for the other 496 a missing replacement was needed, mostly for %fPSA; the size for training frame was 996, while 250 for the test one.

Among the 1,246 participants, whose median (IQR) age was 67 years (61–72 years), a positive biopsy was found in 325 of them (26.1%). When comparing the 2 subcohorts (negative vs. positive

biopsy), PSA as well %fPSA and PCA3 score were statistically significant different, being their median values 6.5 ng/mL versus 7.4 ng/mL, 16 versus 13, and 35 versus 63, respectively ($p < 0.001$ for every comparison). Likewise, age and DRE had a different distribution between the 2 subcohorts, while a family history for PCa and the occurrence of HG-PIN was not associated to a major risk of positive biopsy.

In the multivariate binary LR model with all the 8 features, the main risk factors for PCa occurrence were PSA (odds ratio [OR], 1.07), %fPSA (OR, 0.94), and PCA3 score (OR, 1.01) ($p < 0.001$ for every biomarker, Table 2). Using AUC as a measure of model performance for PCa detection rate, that from the logistic model was 0.738: this is our reference for ML models.

Table 3 shows the median (IQR) and the best AUC as well the accuracy obtained by all the ML classifiers on the test set (best AUC stays for

Table 1. Main patients' characteristics

Characteristic	Whole cohort (n=1,246)	Negative biopsy (n=921)	Positive biopsy (n=325)	Missing values	p-value
Age (yr)	67 (61–72)	67 (61–71)	68 (63–72)	5 (0.4)	0.001
PSA (ng/mL)	6.7 (5.1–9.0)	6.5 (5.0–8.4)	7.4 (5.5–10.0)	-	<0.001
%fPSA	15 (11–20)	16 (12–21)	13 (9–18)	488 (39.2)	<0.001
PCA3 score	39 (16–70)	35 (14–56)	63 (31–115)	-	<0.001
Family history for PCa	92 (7.4)	64 (7.0)	28 (8.6)	1 (0.08)	0.325
HG-PIN	78 (6.3)	65 (7.1)	13 (4.0)	4 (0.3)	0.061

Values are presented as median (interquartile range) or number (%).

PSA: prostate-specific antigen, fPSA: free-PSA, PCA3: Prostate cancer gene 3, PCa: prostate cancer, HG-PIN: high-grade prostate intraepithelial neoplasia.

Table 2. Uni- and multivariate binary logistic regression models

Variable	Univariate models			Multivariate model		
	OR	95% CI	p-value	OR	95% CI	p-value
Age	1.03	1.01–1.05	0.009	1.09	0.99–1.03	0.430
PSA	1.07	1.03–1.12	0.002	1.07	1.03–1.12	0.002
%fPSA	0.94	0.91–0.96	<0.001	0.94	0.92–0.96	<0.001
PCA3 score	1.01	1.01–1.01	<0.001	1.01	1.01–1.01	<0.001
Family history for PCa	1.09	0.62–1.91	0.762	1.34	0.73–2.46	0.358
HG-PIN	0.41	0.19–0.90	0.026	0.58	0.26–1.31	0.172

OR: odds ratio, CI: confidence interval, PSA: prostate-specific antigen, fPSA: free-PSA, PCA3: Prostate cancer gene 3, PCa: prostate cancer, HG-PIN: high-grade prostate intraepithelial neoplasia.

the highest among all the 20 modeling runs). Notably, all the 6 algorithms were able to overtake multivariate logistic model performance, ranging their best AUC from 0.772 to 0.808 and their accuracy from 0.769 to 0.824, always based on the same 8 features set.

The AutoML function had better performances: the top model was an XGBoost one, with AUC 0.830 and accuracy 0.824: 197 of 250 biopsies were correctly classified with an global error rate equal to 21.2%, while the marginal error was 17.8% for the 180 of 250 negative biopsies, and 30.0% for the 70 of 250 positive ones.

The graphics helps to perform an explanatory model analysis for AutoML models. Fig. 1 reports the model ranking for the best AutoML run:

Table 3. ML models classification metrics (AUC and accuracy) estimated on test dataset

Algorithm	Median (IQR) AUC	Best model AUC/accuracy
AutoML (XGBoost)	0.778 (0.748–0.790)	0.830/0.824
GLM	0.758 (0.743–0.767)	0.793/0.780
GBM	0.742 (0.726–0.762)	0.787/0.769
XGBoost	0.731 (0.705–0.746)	0.772/0.769
DRF	0.762 (0.740–0.779)	0.808/0.776
DNN	0.711 (0.687–0.740)	0.776/0.784
NB	0.731 (0.723–0.748)	0.774/0.780

ML: machine learning, AUC: area under curve, IQR: interquartile range, GLM: generalized linear model, GBM: gradient boosting machine, XGBoost: eXtreme gradient boosting machine, DRF: distributed random forest, XRT: eXtremely Randomized Forest, DNN: multilayer artificial Deep Neural Network, NB: naïve Bayes classifier.

XGBoost is the top classifier and 4 different XGBoost models appear among the top 10. Fig. 2 reports an heatmap with the frequencies of identical predictions; those of most AutoML models are quite correlated (especially XGBoosts and GBMs), while XRT/DRFs whose frequencies are not. Fig. 3 represents the variable importance across all AutoML models, after it has been scaled between 0 and 1: the contribute of PCA3 score is clearly prevalent, especially in XGBoosts and GBMs. In Fig. 4, the scaled variable importance for the top XGBoost model is plotted: the PCA3 score sharply leads this ranking, being the most critical feature for a positive biopsy. Finally, Figs. 5–7 show 3 partial dependence plots for PCA3 score, %fPSA, and PSA, respectively: while the x-axis reports the biomarker values, the y-axis represents the likelihood of a positive biopsy. The marginal effect that these features exert on the target follows a different pattern: as for %fPSA and PSA, the probability of PCa changes progressively without any cutoff. For PCA3 score instead, this risk follows a bimodal distribution: it jumps up from around 25% to 60%, when the PCA3 score increases from around 80 to 120.

Of note, it's possible to make individual interpretations for any single patient too, e.g., for the 2 patients with the lowest and highest probability of positive biopsy, estimating the Break Down profile that shows the contribution of every feature to

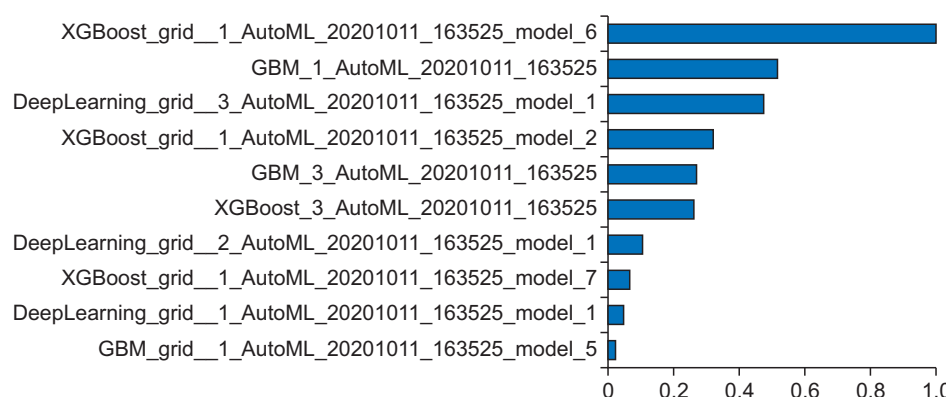


Fig. 1. AutoML (all models), model ranking. XGBoost: eXtreme gradient boosting machine, GBM: gradient boosting machine.

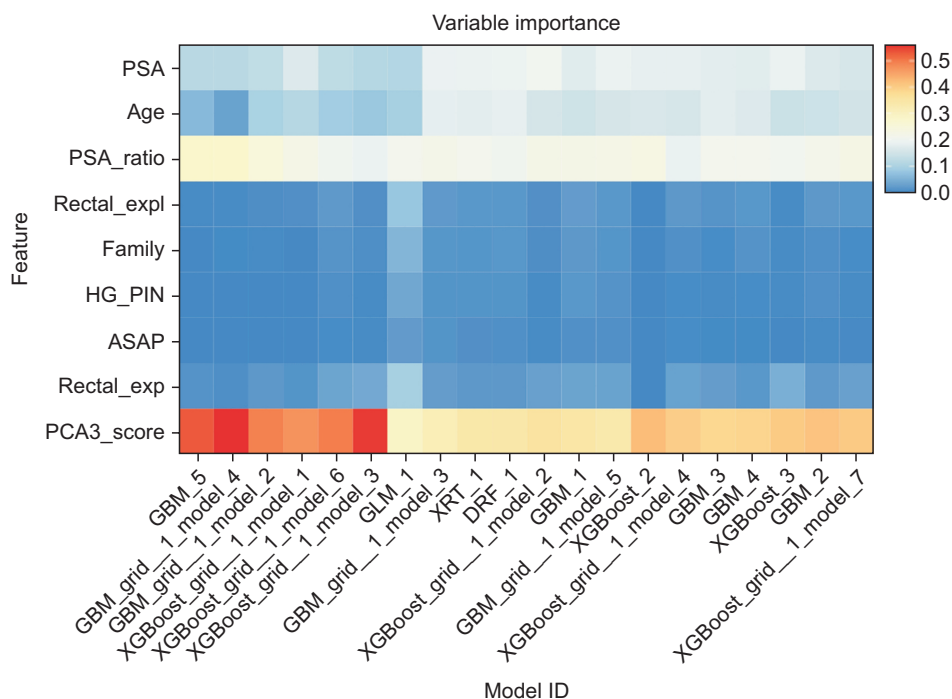


Fig. 2. AutoML (all models), variable importance. PSA: prostate-specific antigen, HG-PIN: high-grade prostate intraepithelial neoplasia, ASAP: Atypical Small Acinar Proliferation, PCA3: Prostate cancer gene 3, GBM: gradient boosting machine, XRT: eXtremely Randomized Forest, DRF: distributed random forest, GLM: generalized linear model, XGBoost: eXtreme gradient boosting machine.

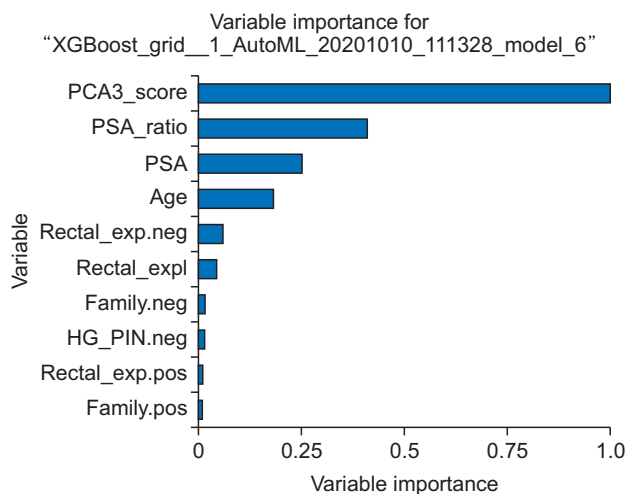


Fig. 3. AutoML (XGBoost), heat map for variable importance. XGBoost: eXtreme gradient boosting machine, PCA3: Prostate cancer gene 3, PSA: prostate-specific antigen, HG-PIN: high-grade prostate intraepithelial neoplasia.

target prediction (always working with AutoML XGBoost top model). Even if 130 patients had a negative biopsy, his probability of PCa was very high (91.0%), mostly due to his extreme PCA3 score: the ML model thus underlines the potential risk that this man has had a false negative result at

biopsy (Table 4).

DISCUSSION

The management of PCa poses difficult challenges, mainly due to the lackness of ideal tools to predict its occurrence. At present, up to two-thirds of patients undergoing a systematic prostate biopsy have a negative histological finding, depending on the decision to perform biopsy on PSA, a sensitive but highly unspecific biomarker. To avoid unnecessary prostate biopsies, multiparametric magnetic resonance imaging (mpMRI) and serum/urine biomarkers such as free-PSA, total/free-PSA ratio and 4Kscore, prostate health index (PHI), and PCA3 score may be used to diagnose PCa.^{1-3,24}

In a recent systematic review and meta-analysis, Muñoz Rodriguez and Perdomo²⁵ documented an overall sensitivity of 69% and specificity of 65%, at a PCA3 score cutoff of 35. Additionally, the PCa occurrence OR was 4.24 (95% confidence interval, 3.49–5.17), and the area under the curve 0.734.

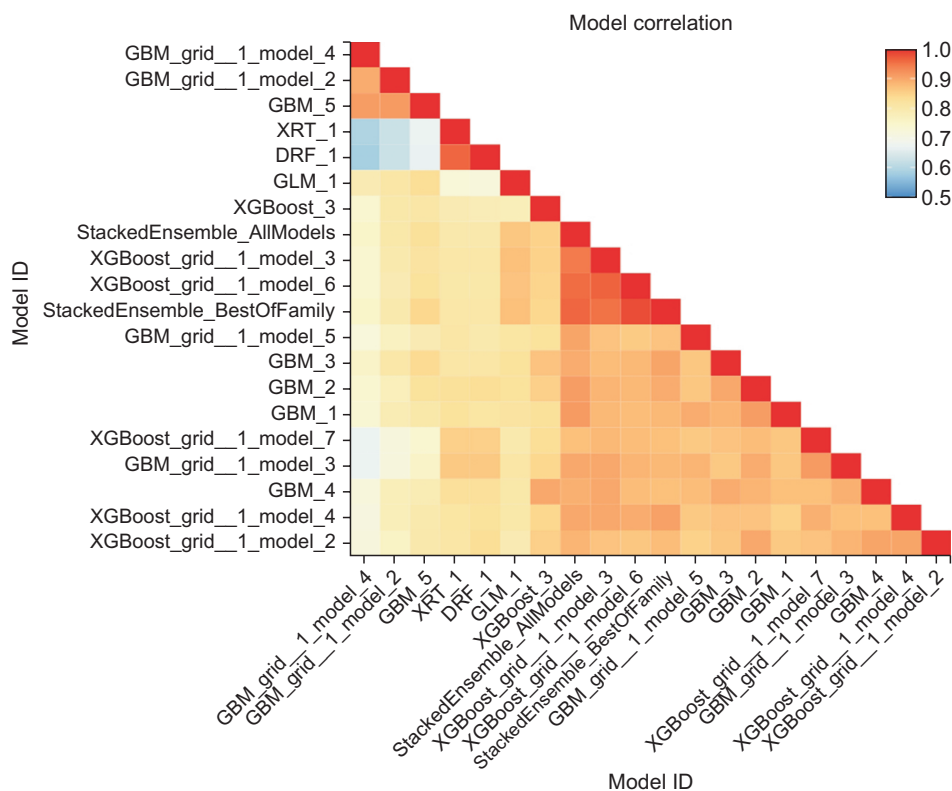


Fig. 4. AutoML (all models), model correlation. GBM: gradient boosting machine, XGBoost: eXtreme gradient boosting machine, GLM: generalized linear model, XRT: eXtremely Randomized Forest, DRF: distributed random forest.

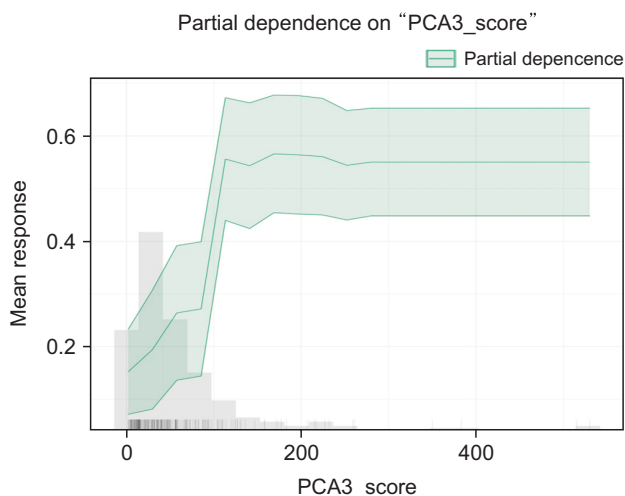


Fig. 5. AutoML (XGBoost), partial dependence plot for prostate cancer gene 3 (PCA3) score. XGBoost: eXtreme gradient boosting machine.

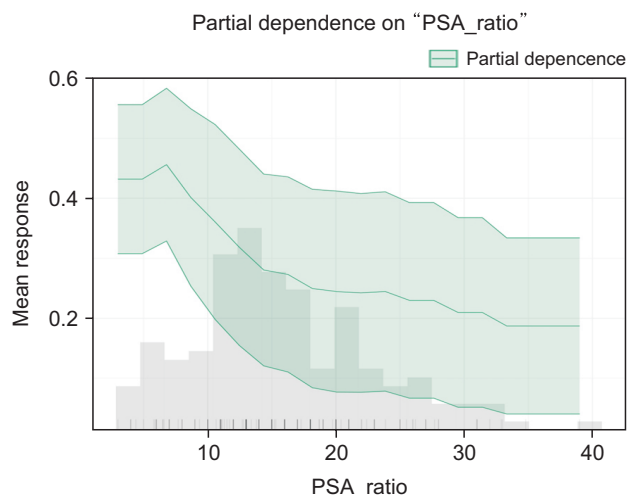


Fig. 6. AutoML (XGBoost), partial dependence plot for %fPSA. XGBoost: eXtreme gradient boosting machine, PSA: prostate-specific antigen.

In our previous study, we investigated the role of PCA3 score on a large real-world cohort of Italian men undergoing first or repeat biopsy for PCa.¹² At that time, we used the LR model to predict PCa detection rate at different PCA3 score values.

We confirmed the usefulness of the PCA3 score determination among men who had a previous negative biopsy and an elevated PSA level. In this subgroup, a sensitivity of 73.2% and a specificity of 75.5% were documented using a cutoff of 39, its

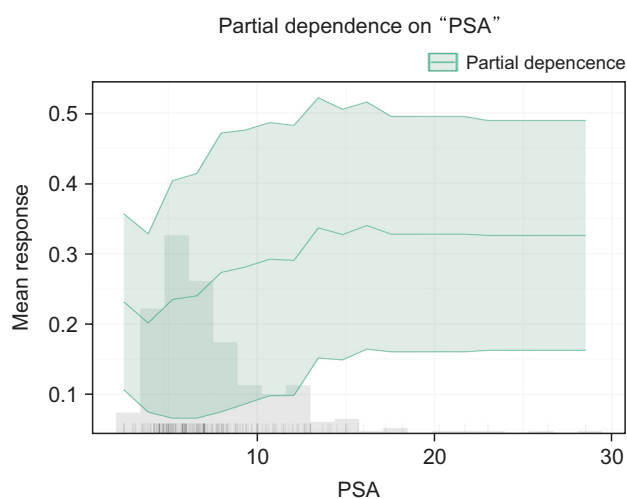


Fig. 7. AutoML (XGBoost), partial dependence plot for PSA. XGBoost: eXtreme gradient boosting machine, PSA: prostate-specific antigen.

median value.

Takeuchi reported that the prediction rate for PCa improved by about 5%–10% when using a multilayer artificial neural network compared to classical LR in 334 patients who underwent 3.0T mpMRI before trans rectal ultrasound-guided prostate biopsy.²⁶ The authors applied various ML algorithms to calculate PCa prediction rate comparing LR to ML approach. In patients aged >75 years and with a PSA level of 2.5–10 ng/mL, LR analysis showed the best prediction rate (74.6%); however, ML methods performed better than LR in other patient groups. In particular, the Random Forest model (prediction rate, 70.5%–72.4%) overtook LR results (prediction rate, 65.6%–70.0%) in patients aged 65–74 years.

It is worth noting that our retrospective case series included patients from 2008 to 2010. At that time, the use of mpMRI as a triage test was not already diffused in Europe, especially in Italy. The first experiences with the use of mpMRI before surgery were published in 2011,²⁷ whilst the use of mpMRI as a guidance for target biopsies were developed in Italy from 2017.²⁸

Nevertheless, in the current PCa diagnostic

Table 4. The 2 patients with lowest and highest probability of positive biopsy, as identified by the AutoML(XGBoost) best model

Variable	Best patient (#110)	Worst patient (#130)
Age (yr)	75	65
PSA (ng/mL)	5.21	8.16
%fPSA	47	15
PCA3 score	5	331
Family history for PCa	Negative	Positive
DRE	Negative	Negative
HG-PIN	Negative	Negative
Biopsy result	Negative	Negative
Probability of positive biopsy	12.4%	91.0%

XGBoost: eXtreme gradient boosting machine, PSA: prostate-specific antigen, fPSA: free-PSA, PCA3: Prostate cancer gene 3, PCa: prostate cancer, DRE: digital rectal examination, HG-PIN: high-grade prostate intraepithelial neoplasia.

scenario, MRI cannot be used for screening purposes yet, being time-consuming and not homogeneously available everywhere, as well requiring experienced radiologists. A possible solution could be the development of a ML-based clinical decision support system based on biomarkers, as in this case PCA3 score and PSA, to stratify patients according to their risk of PCa progression, properly selecting candidates for mpMRI, thus reducing unnecessary prostate biopsies.

The present study has been planned to investigate ML modeling performances in the prediction of PCa, to confirm the role of PCA3 score as a key determinant for positive biopsy, to test ML techniques feasibility and reproducibility in a real-world urological context.

Each of the ML models, GLM (AUC, 0.793), GBM (AUC, 0.787), XGBoost (AUC, 0.772), DRF (AUC, 0.808), DNN (AUC, 0.776) and NB (AUC, 0.774), outperformed the classical LR model (AUC, 0.738). Notably, the top model has been set up with the AutoML function (AutoML XGBoost; AUC, 0.830): this represents an interesting improvement in PCa detection, when compared with our previous results: a biomarker-driven PCa prediction appears to be feasible by a ML approach.

As for PCA3 score, all ML algorithms recognized it as a main predictor of positive biopsy; it would be of value to test by ML how this biomarker performs when associated either to other ones (4Kscore, PHI, PSA density) or to mpMRI, histopathological and genetics data, and for specific subcohorts like e.g., subjects with “gray zone” PSA.

As for ML techniques feasibility, the H2O ML platform is freely available both for R and Python, among the most diffused open-source programming languages. Furthermore, a brand new ML addendum like the explanatory model analysis, here presented for the best model AutoML (XGBoost), forms a graphics set that can remarkably help the reader in understanding what happened inside the “black box” model.

Among the methodological limitations of this research, the biases depending on the retrospective observational design could be overtaken by a randomized controlled trial. Second, the absence of external validation; since the development cohort was retrieved from only 3 institutions, the risk of selection bias regarding patient population or biopsy indications should be overcome by external validation. Third, all ML models are affected by the sample size: in artificial intelligence investigations, instances are never excessive and no sample size could be really defined as adequate. Finally, when analyzing today a 10-year-old cohort, it’s important to remember that diagnosis and therapy for PCa have been changed and updated so much. About that, 10 years ago mpMRI was not widespread in Italy and its use was reserved to a minority of cases and only in a repeat biopsy setting. At that time, in all studies on urine or serum biomarkers, the gold standard for PCa detection was the pathological examination of multiple nontargeted systematic TRUS-guided prostate biopsies. Intrinsically, this approach implies that no cancer predicted by the biomarker could count for a biopsy missed cancer.

CONCLUSIONS

In our experience, the ML approach may be recommended, when applied to PCa prediction based on biomarkers fluctuations. It proved to be feasible and had better performances with reproducible results in terms of AUC and accuracy, when compared to the LR model.

CONFLICT OF INTEREST

The authors claim no conflicts of interest.

ORCID

Roberto Passera, <https://orcid.org/0000-0003-3372-2248>

Stefano De Luca, <https://orcid.org/0000-0002-2596-9796>

Cristian Fiori, <https://orcid.org/0000-0002-6345-1322>

Enrico Bollito, <https://orcid.org/0000-0002-8483-893X>

Francesco Porpiglia, <https://orcid.org/0000-0002-0752-4857>

REFERENCES

1. Catalona WJ, Partin AW, Slawin KM. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial. *JAMA* 1998;279:1542-7.
2. Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, et al. DD3: a new prostate-specific gene, highly over expressed in prostate cancer. *Cancer Res* 1999;59:5975-9.
3. Lee D, Shim SR, Ahn, ST, Oh MM, Moon DG, Park HS, et al. Diagnostic performance of the prostate cancer antigen 3 test in prostate cancer: systematic review and meta-analysis. *Clin Genitourin Cancer* 2020;18:402-8.
4. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med* 2019;380:1347-58.
5. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC*

- Med Res Methodol 2019;19:64.
6. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol* 2019;6:391-403.
 7. Chen J, Remulla D, Nguyen JH, Liu AY, Dasgupta P, Hung AJ. Current status of artificial intelligence applications in urology and their potential to influence clinical practice. *BJU Int* 2019;124:567-77.
 8. Suarez-Ibarrola R, Hein S, Reis G, Gratzke C, Miernik A. Current and future applications of machine and deep learning in urology: a review of the literature on urolithiasis, renal cell carcinoma, and bladder and prostate cancer. *World J Urol* 2020;38:2329-47.
 9. Checcucci E, Autorino R, Cacciamani GE, Amparore D, De Cillis S, Piana A, et al. Artificial intelligence and neural networks in urology: current clinical applications. *Minerva Urol Nefrol* 2020;72:49-55.
 10. Lee J, Yang SW, Lee S, Hyon YK, Kim J, Jin L, et al. Machine learning approaches for the prediction of prostate cancer according to age and the prostate-specific antigen level. *Korean J Urol Oncol* 2019;17:110-7.
 11. Barlow H, Mao S, Khushi M. Predicting high-risk prostate cancer using machine learning methods. *Data* 2019;4:129.
 12. Bollito E, De Luca S, Cicilano M, Passera R, Grande S, Maccagnano C, et al. Prostate cancer gene 3 urine assay cutoff in diagnosis of prostate cancer. *Anal Quant Cytol Histol* 2012;34:96-104.
 13. The R Project for Statistical Computing [Internet]. Vienna (Austria): The R Project for Statistical Computing; [cited 2021 Jul 30]. Available from: <https://www.R-project.org/>.
 14. H2O, H2O version 3.33.0.5217 [Internet]. H2O.ai 2020; [cited 2021 Jul 30]. Available from: <https://github.com/h2oai/h2o-3>.
 15. Nelder J, Wedderburn RWM. Generalized linear models. *J R Statist Soc Series A General* 1972;135:370-84.
 16. Hastie T, Tibshirani R, Friedman JJH. The elements of statistical learning. New York: Springer; 2001.
 17. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: 22nd SIGKDD Conference on Knowledge Discovery and Data Mining, 2016.
 18. Breiman L. Random forests. *Mach Lear* 2001;45:5-32.
 19. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006;313:504-7.
 20. Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
 21. LeDell E, Poirier S. H2O AutoML: scalable automatic machine learning. In: 7th ICML Workshop on Automated Machine Learning, 2020.
 22. Van Buuren S. Flexible imputation of missing data. Boca Raton: Chapman & Hall/CRC; 2018.
 23. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Letters* 2006;27:861-74.
 24. Porpiglia F, Manfredi M, Mele F, Cossu M, Bollito E, Veltri A, et al. Diagnostic pathway with multiparametric magnetic resonance imaging versus standard pathway: results from a randomized prospective study in biopsy-naïve patients with suspected prostate cancer. *Eur Urol* 2017;72:282-8.
 25. Muñoz Rodríguez SV, Perdomo HAG. Diagnostic accuracy of prostate cancer antigen 3 (PCA3) prior to first prostate biopsy: a systematic review and meta-analysis. *Can Urol Assoc J* 2020;14:E214-9.
 26. Takeuchi T, Hattori-Kato M, Okuno Y, Iwai S, Mikami K. Prediction of prostate cancer by deep learning with multilayer artificial neural network. *Can Urol Assoc J* 2019;13:E145-50.
 27. Sciarra A, Barentsz J, Bjartell A, Eastham J, Hricak H, Panebianco V, et al. Advances in magnetic resonance imaging: how they are changing the management of prostate cancer. *Eur Urol* 2011;59:962-77.
 28. Manfredi M, De Luca S, Fiori C. Multiparametric prostate MRI for prostate cancer diagnosis: is this the beginning of a new era? *Minerva Urol Nefrol* 2017;69:628-9.