Title

Integrative Machine Learning Prediction of Prostate Biopsy Results From Negative Multiparametric MRI

Journal
Journal of Magnetic Resonance Imaging, 55(1)

ISSN
1053-1807

Authors
Zheng, Haoxin
Miao, Qi
Liu, Yongkai
et al.

Publication Date
2022

DOI
10.1002/jmri.27793

Peer reviewed

HHS Public Access
Author manuscript
J Magn Reson Imaging. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:
J Magn Reson Imaging. 2022 January ; 55(1): 100–110. doi:10.1002/jmri.27793.

Integrative Machine Learning Prediction of Prostate Biopsy Results from Negative Multiparametric MRI

Haoxin Zheng, MS[1,2], **Qi Miao, MD**[1,3], **Yongkai Liu, MS**[1], **Steven S. Raman, MD**[1], **Fabien Scalzo, PhD**[2,4], **Kyunghyun Sung, PhD**[1]

[1]Radiological Sciences, University of California — Los Angeles, Los Angeles, CA 90095, USA

[2]Computer Science, University of California — Los Angeles, Los Angeles, CA 90095, USA

[3]Department of Radiology, The First Affiliated Hospital of China Medical University, Shenyang City 110001, Liaoning Province, China

[4]Neurology, University of California — Los Angeles, Los Angeles, CA 90095, USA

Abstract

Background: Multiparametric MRI (mpMRI) is commonly recommended as a triage test prior to any prostate biopsy. However, there exists limited consensus on which patients with a negative prostate mpMRI could avoid prostate biopsy.

Purpose: To identify which patient could safely avoid prostate biopsy when the prostate mpMRI is negative, via a radiomics-based machine learning approach.

Study Type: Retrospective.

Subjects: 330 patients with negative prostate 3T mpMRI between January 2016 and December 2018 were included.

Field Strength / Sequence: 3.0T / T_2-weighted Turbo Spin Echo (TSE) imaging (T_2WI) and diffusion-weighted imaging (DWI).

Assessment: The integrative machine learning (iML) model was trained to predict negative prostate biopsy results, utilizing both radiomics and clinical features. The final study cohort comprised 330 consecutive patients with negative mpMRI (PI-RADS<3) who underwent systematic transrectal ultrasound-guided (TRUS) or MR-ultrasound fusion (MRUS) biopsy within six months. A secondary analysis of biopsy naïve sub-cohort (n=227) was also conducted.

Statistical Tests: The Mann-Whitney U test and Chi-Squared test were utilized to evaluate the significance of difference of clinical features between prostate biopsy positive and negative groups. The model performance was validated using leave-one-out cross-validation (LOOCV) and measured by AUC, sensitivity, specificity, and negative predictive value (NPV).

Results: Overall, 306/330 (NPV 92.7%) of the final study cohort patients had negative biopsies, and 207/227 (NPV 91.2%) of the biopsy naïve sub-cohort patients had negative biopsies. Our

Corresponding author info: Name: Miao Qi, **Address**: 757 Westwood Plaza, Los Angeles, CA, US, 90095, qmiao@mednet.ucla.edu, **Tel**: 424-440-9637, **FAX**: 310-825-6201.

iML model achieved NPVs of 98.3% and 98.0% for the study cohort and sub-cohort respectively, superior to prostate-specific antigen density (PSAD)-based risk assessment with NPVs of 94.9% and 93.9%, respectively.

Data Conclusion: The proposed iML model achieved high performance in predicting negative prostate biopsy results for patients with negative mpMRI. With improved NPVs, the proposed model can be used to stratify patients who in whom we might obviate biopsies, thus reducing the number of unnecessary biopsies.

Keywords

Multi-parametric MRI; Prostate cancer; Radiomics; Machine Learning

INTRODUCTION

Multi-parametric MR Imaging (mpMRI) is now the preferred imaging technique for noninvasive diagnosis of prostate cancer (PCa). mpMRI is increasingly performed prior to prostate biopsy to maximize yield of clinically significant prostate cancer (csPCa) and minimize error (1). In the standardized 5 point Likert score based Prostate Imaging Reporting and Data System, version 2.1 (PI-RADS v2.1), intermediate and high suspicion MRI based lesions (PI-RADS≥3) typically undergo MRI-targeted biopsy with or without systematic biopsies with positive tissue diagnosis between 29.7-82.4% (2–4). However, when mpMRI findings are of low suspicion, mpMRI negative (PI-RADS 1 or 2), there is a lack of consensus of whether to proceed with a systematic biopsy which contributes to patient morbidity including pain, bleeding, urinary obstruction and erection dysfunction (5). Several strategies have been proposed in patients with negative mpMRI to predict low risk of csPCa including use of serum biomarkers such as prostate-specific antigen density (PSAD) levels less than either 0.10 ng/ml/ml or 0.15 ng/ml/ml (6–10). However, the current PSAD-based risk assessments are limited by negative predictive value (NPV) of 83.1% to 93.4% (6–10).

Radiomics is an emerging field in quantitative imaging that aims to associate radiomic features with specific clinical endpoints (11–13). The radiomics features extracted from medical images can provide large-scale imaging information, and many studies have shown promising results on the PCa detection and aggressiveness assessment using radiomics features (14–21). The aim of the study is to construct and validate a radiomics-based model for predicting biopsy results in patients with negative MRI. Specifically, an integrative machine learning (iML) model was proposed combining visually negative (PI-RADS 1 or 2) MRI-based radiomics features with routine clinical information to predict the prostate biopsy results. The efficacy of using the integrative multi-scale features was validated in comparisons with other machine learning approaches using either only clinical information or only radiomics features. In addition, the NPV and overall performance of the proposed iML approach was compared with pre-existing PSAD-based strategies to predict risks of csPCa in patients with negative mpMRI (6–9).

MATERIALS AND METHODS

Study Population And MRI Data

The single arm observational study was performed in compliance with the United States Health Insurance Portability and Accountability Act (HIPAA) of 1996 and was approved by the institutional review board (IRB) with a waiver of the requirement for informed consent. The initial study cohort included all identified negative prostate 3 Tesla mpMRI cases by reviewing all clinical prostate MRI scans performed by a standard protocol via one of several 3 Tesla scanners: Siemens Magnetom Trio, Skyra, and Verio scanner (Siemens Medical Systems, Malvern, Pennsylvania, USA) from January 2016 to December 2018 at a single academic institution. All prostate mpMRI scans were acquired using a standardized imaging protocol following European Society of Urogenital Radiology (ESUR) PI-RADS guidelines (22). The detailed sequence parameters are listed in Supplementary Materials. Three genitourinary radiologists interpreted the mpMRI scans, as part of the clinical diagnostic procedure, where each had read 1,000-3000 prostate mpMRI scans with 10+ years of experience.

The study cohort patients met the following inclusion criteria: 1) clinical suspicion of PCa, (elevated PSA level with respect to the current age and/or abnormal digital rectum exam results); 2) 3T-mpMRI with all lesions scored as PIRADS 1 or 2 (MR negative); 3) standardized 12-14 core systematic transrectal ultrasound-guided (TRUS) biopsy with or without magnetic resonance ultrasound fusion (MRUS) within six months after 3T-mpMRI study (23); 4) serum prostate specific antigen (PSA) measured within six months prior to biopsy. All eligible cases were re-reviewed by another independent abdominal radiologist (X. X., 5 years of experience in clinical prostate MRI interpretation), and no discordant was observed. MRUS was used for a partial cohort to record and track biopsy site locations, and there was no difference between TRUS biopsy with and without MRUS. Patients with a known diagnosis of PCa, undergoing active surveillance, or PCa treatment (including surgery, focal therapy, radiation, or hormonal therapy), were excluded.

For patients with multiple mpMRI scans, we selected the mpMRI scan immediately preceding the first negative TRUS/MRUS biopsy. The detailed patient inclusion workflow is shown in Figure 1.

In all, 330 men, median age 63 years (IQR: 58-67), with either systematic TRUS (n=87) or MRUS (n=243) biopsy were included in the final study cohort, for the primary analysis. A secondary analysis on a biopsy naïve cohort (n=227) was conducted to further evaluate the performance in a less cancer enriched population (6–9).

Negative biopsy was defined as excluding csPCa (lack of primary or secondary Gleason Score (GS) ≥ 7) findings in each biopsy session (24). The following clinical information was evaluated: patient age, family history of PCa, prostate biopsy history, prostate volume, PSA, and PSAD. Other clinical information was incompletely available and thus not included in the study to avoid potential selection bias (25). All TRUS and MRUS biopsy cores were fixed in formalin, stained with hematoxylin and eosin (H&E) for histological evaluation performed by dedicated genitourinary pathologists as part of the routine clinical histopathological evaluation.

Integrative Machine Learning Model

The workflow for building our proposed iML model is shown in Figure 2. For a patient-basis prediction of positive or negative biopsy results, we used both apparent diffusion coefficient (ADC) maps and T_2-weighted images (T_2WI) from 3T mpMRI (4). The ADC maps were registered to T_2WI through rigid spatial transformation using voxel size and real-world coordinates information for each patient (14,26–28). After checking the quality of the registration, we found no observable discrepancies between T_2WI and ADC. The whole prostate gland was manually segmented on T_2WI slice-by-slice by the abdominal radiologist (X.X.; 5 years of experience in clinical prostate MRI interpretation) under the supervision of a senior genitourinary radiologist (Y.Y.; 20+ years of experience in clinical prostate MRI interpretation) using OsiriX MD (ver. 11.0.3). We then applied N4 bias field correction to T_2WI to compensate for the low-frequency intensity non-uniformities and applied z-score normalization to T_2WI and ADC images (29,30).

Radiomic features were extracted from T_2WI and ADC images after cropping the whole prostate, as shown in Figure 2. All the slices containing region of interest (ROI) of the whole prostate were used for feature extraction, and the mid-prostate slice was separately used to extract additional radiomics features. Among texture features, Gray-Level Cooccurrence Matrix (GLCM) and Gray-Level Run Length Matrix (GLRLM) were included using Pyradiomics package based on Python (31). A total of 300 radiomics features were extracted for each patient, including 32 shape-based, 38 first-order, and 80 texture features from each of the T_2WI and ADC images.

In order to pre-select important clinical features, significance levels, defined as $p < 0.05$, were calculated for all routine clinical information between prostate biopsy positive group and negative group. Specifically, given the six initial clinical characteristics, Mann-Whitney U test was applied for continuous-valued features (i.e., age, PSA, PSAD, prostate volume) after checking the data normality using Kolmogorov-Smirnov test. Chi-Square test was applied for categorical features (i.e., family PCa history, prostate biopsy history). The detailed patients' clinical information can be seen in Table 1. We selected the clinical features that have a significant difference ($p<0.05$) between the biopsy positive and negative groups. Finally, we combined the pre-selected clinical features and all radiomics features and applied the Sequential Floating Forwarding Selection (SFFS) algorithm for integrative feature selection (Figure 2) (32).

Model Comparison And Statistical Analysis

We used a quadratic-kernelized support vector machine (SVM) classifier with a class-balanced weight to train our proposed iML model. The model was validated by leave-one-out cross-validation (LOOCV) to reduce potential overfitting issues and also measure the evaluation results' variance (33–36). We first investigated the value of the iML approach by comparing the performance of iML with the models using only radiomics features or clinical features by DeLong test (37). All models were using the same classifier, the quadratic-kernelized SVM with the class-balanced weight. We then compared the prediction performance of the proposed iML with two conventional PSAD-based strategies: 1) PSAD<0.10 ng/ml/ml as with low risks of having csPCa (6) and 2) PSAD<0.15 ng/ml/ml as with low risks of having csPCa (7–9).

For each model, we identified the optimal cutoff point for the prediction of negative biopsy results by maximizing the Youden's index value (sensitivity+specificity-1) on ROC curves (38). NPV was calculated to measure the detection rate of true negative cases among all negative predictions, consistent with other studies (6–9). We further included sensitivity, specificity, and AUC in order to perform a more comprehensive evaluation to minimize the potential influence caused by data imbalance during model evaluation. Finally, all model comparisons were evaluated based on AUC with a 95% confidence interval (CI), and NPV, sensitivity, and specificity were calculated from ROC at the optimal cutoff point.

RESULTS

The patient clinical characteristics in the final study cohort and the biopsy naïve sub-cohort are summarized in Table 1. Clinical information including age, prostate volume, and PSAD were selected during the procedure of clinical feature selection because of the significant difference ($p<0.05$) between biopsy positive and negative groups. Based on our inclusion criteria, 306 patients had negative biopsies and 24 patients had positive biopsies among the final study cohort (n=330). 207 patients had negative biopsies and 20 patients had positive biopsies among the biopsy naïve sub-cohort (n=227).

There were nine total integrative features comprised of six radiomics and three clinical features and are summarized in Table 2. Figure 3 shows representative examples of 3T mpMRI-based radiomics features, stratified as negative (top) and positive (bottom) biopsies. The six radiomics features consisted of three shape and three texture features. The shape features (Minor Axis_Length, Major Axis_Length and Least Axis_Length) descripted the shape and size information of the ROI region of prostate, and the texture features (Sum Squares, Gray Level Non-Uniformity and Run Length Non-Uniformity) descripted the texture information of the ROI region of prostate, on T_2WI and ADC images. With negative MRI, the selected radiomics features show different visual patterns between two groups (A and B vs. C and D), as shown in the spider plots.

Figure 4A and C show the ROC comparisons between the proposed iML model and machine learning models with an individual feature group in two patient cohorts. The proposed iML approach achieved the highest AUC ($p<0.05$), compared with the models using an individual group of radiomics or clinical features, in both cohorts (Table 3). The AUC, and sensitivity, specificity and NPV that based on the optimum cutoff points of the iML approach were 0.798 (95% CI, 0.711-0.885), 83.3%, 75.2%, and 98.3% respectively in the final study cohort, which improved the AUC of [13.2%, 17.5%] compared with clinical-only and radiomics-only models ($p<0.05$), respectively. For the biopsy naïve cohort, the iML approach reached AUC, sensitivity, specificity and NPV of 0.749 (95% CI, 0.645-0.854), 85.0%, 72.0%, and 98.0% respectively. It thus improved the AUC of [10.3%, 29.4%], compared with clinical-only and radiomics-only models ($p<0.05$), respectively.

The comparison results between the PSAD-based risk prediction methods and the iML model conducted on the same study population are shown in Table 4 and Figure 4B and 4D. The iML approach achieved higher specificity/sensitivity ($p<0.05$) while keeping the similar sensitivity/specificity to the results conducted by using thresholds of PSAD=0.10 ng/ml/ml and of PSAD=0.15 ng/ml/ml (Fig. 4B and D) for both cohorts (6–9). Moreover, for both final study cohort and biopsy naïve cohort, our proposed iML approach achieved NPVs of [98.3%, 98.0%], showed improvement compared with PSAD-based risk prediction methods, resulted in NPVs of [94.6%, 93.9%] for PSAD<0.10 ng/ml/ml and NPVs of [94.9%, 93.9%] for PSAD<0.15 ng/ml/ml (6–9).

Comparisons of prediction performances using different approaches are shown in Table 4. Specifically, for the final study cohort, the iML approach improved the results conducted by using a threshold of PSAD=0.10 ng/ml/ml and a threshold of PSAD=0.15 ng/ml/ml on sensitivity of [17.7%, 66.6%], specificity of [88.5%, 2.4%] and NPV of [3.9%, 3.6%] (6–9). For the biopsy naïve cohort, the iML approach improved the results on sensitivity of [21.4%, 70.0%], specificity of [+60.4%, −3.2%] and NPV of [4.4%, 4.4%], respectively. Figure 5 visualized the prediction results using PSAD-based approaches and iML on both final study cohort and biopsy naïve cohort. The histograms show iML had the highest true positive ratio and the smallest true negative ratio among all methods (see Table 4 for statistical comparison between different methods).

DISCUSSION

We proposed an integrative machine learning (iML) model as a potential triage test to obviate biopsy when 3T mpMRI was negative. Our findings showed that integrating both MRI and clinical information helped improve the prediction of the biopsy results ($p<0.05$), compared with the machine learning approaches conducted by individually using either MRI-based radiomics features or clinical features.

Recent review studies reported that common strategies of using PIRADS<3 as a triage test to obviate biopsy resulted in NPVs with a range of 80.5% to 92.3%, and the PSAD-based assessment improved the NPVs to be in the range of 83.1% to 93.4% for predicting negative biopsy results among patients with negative MRI (10,39). In this study, the final study cohort had NPV of 92.7% and improved to 98.3% using the iML approach. This performance of the negative biopsy results was higher than other studies' NPVs, ranged from 89.0% to 89.9%, and the PSAD-based assessment with NPVs, ranged from 83.1 to 93.4% (7,8,10). In the biopsy naïve cohort, iML improved NPV from 91.2 to 98.0%, also higher than reported by other existing studies and the PSAD-based assessment strategies (6,9,10). Furthermore, our results on both patient cohorts also achieved improvements in sensitivity/specificity with a small cost of specificity/sensitivity in comparison with reported specificities and sensitivities (6,7,9).

Prior studies had shown the MRI-based radiomics features had excellent performance for the prediction and aggressiveness assessment of PCa (14–21). With the similar settings as the previous studies, our study also took the first-order, shape and texture features into consideration as the MRI-based radiomics features in order to comprehensively extracted the PCa-related information from T_2WI and ADC images. The study showed the improved performances of using the integration of radiomics features and clinical information via machine learning when predicting patient-basis negative biopsy results, compared with the situation using individual feature groups only, or using the pre-existing PSAD-based methods.

Limitations

Our study includes a few limitations. The study included 330 patients with negative MRI who underwent systematic biopsies within six months. The study cohort was identified after investigating all in-house prostate 3T mpMRI scans for three years at a single academic institution (n=2,679). Although the size of the dataset was relatively small and contained imbalanced distribution between positive and negative biopsies, the data characteristic was similar to the previously investigated studies due to the study objectives (6–9). Moreover, the study didn’t conduct a separate study on the cohort of patients that have a prior negative biopsy (n=103) was conducted due to the limited number of positive biopsy cases (n=4). Additionally, we used LOOCV for the model evaluation due to the limited number of data with class imbalance (33), consistent with other studies when only limited data was available (34–36). Our future works would include continuous collection of available data to evaluate our model with an external testing set. We believe this will further solidify our findings.

Conclusion

In conclusion, the negative biopsy results were highly predictable among patients with negative prostate MRI using the integrative machine learning (iML) model. The integration of MRI-based radiomics and clinical features improved the performance in predicting negative biopsy results. The proposed iML model outperformed the existing PSAD-based strategies with NPV of 98.3%, in the final study cohort, and NPV of 98.0%, in the biopsy naïve sub-cohort, respectively. It can thus be used to stratify patients who should obviate biopsies, potentially reducing the number of unnecessary biopsies for patients with negative prostate MRI.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Grant support:

This work was supported by the National Institutes of Health (NIH) R01-CA248506 and funds from the Integrated Diagnostics Program, Department of Radiological Sciences & Pathology, David Geffen School of Medicine at UCLA.

REFERENCES

1. Brizmohun Appayya M, Adshead J, Ahmed HU, et al. National implementation of multi-parametric magnetic resonance imaging for prostate cancer detection — recommendations from a UK consensus meeting. BJU International 2018;122:13–25. [PubMed: 29699001]

2. Schoots IG, Roobol MJ, Nieboer D, Bangma CH, Steyerberg EW, Hunink MG. Magnetic resonance imaging-targeted biopsy may enhance the diagnostic accuracy of significant prostate

cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis. Eur Urol 2015;68:438–450. [PubMed: 25480312]

3. Sathianathen NJ, Konety BR, Soubra A, et al. Which scores need a core? An evaluation of MR-targeted biopsy yield by PIRADS score across different biopsy indications. Prostate Cancer and Prostatic Diseases 2018;21:573–578. [PubMed: 30038389]
4. Turkbey B, Rosenkrantz AB, Haider MA, et al. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. Eur Urol 2019;76:340–351. [PubMed: 30898406]
5. Liss MA, Ehdaie B, Loeb S, et al. An Update of the American Urological Association White Paper on the Prevention and Treatment of the More Common Complications Related to Prostate Biopsy. J Urol 2017;198:329–334. [PubMed: 28363690]
6. Hansen NL, Barrett T, Kesch C, et al. Multicentre evaluation of magnetic resonance imaging supported transperineal prostate biopsy in biopsy-naive men with suspicion of prostate cancer. BJU Int 2018;122:40–49. [PubMed: 29024425]
7. Oishi M, Shin T, Ohe C, et al. Which Patients with Negative Magnetic Resonance Imaging Can Safely Avoid Biopsy for Prostate Cancer? J Urol 2019;201:268–276. [PubMed: 30189186]
8. Distler FA, Radtke JP, Bonekamp D, et al. The Value of PSA Density in Combination with PI-RADS for the Accuracy of Prostate Cancer Prediction. J Urol 2017;198:575–582. [PubMed: 28373135]
9. Otti VC, Miller C, Powell RJ, Thomas RM, McGrath JS. The diagnostic accuracy of multiparametric magnetic resonance imaging before biopsy in the detection of prostate cancer. BJU Int 2019;123:82–90. [PubMed: 29804315]
10. Pagniez MA, Kasivisvanathan V, Puech P, Drumez E, Villers A, Olivier J. Predictive Factors of Missed Clinically Significant Prostate Cancers in Men with Negative Magnetic Resonance Imaging: A Systematic Review and Meta-Analysis. The Journal of urology 2020;204:24–32. [PubMed: 31967522]
11. Gillies RJ, Kinahan PE, Hricak H. Radiomcis: Images Are More than Pictures, They Are Data. Radiology 2016;278:563–577. [PubMed: 26579733]
12. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. CancerData 2016;41:366–373.
13. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–762. [PubMed: 28975929]
14. Yan C, Peng Y, Li X. Radiomics analysis for prostate cancer classification in multiparametric magnetic resonance images. International Conference on Biological Information and Biomedical Engineering. Hangzhou, China: IEEE; 2019. p. 247–250.
15. Vignati A, Mazzetti S, Giannini V, et al. Texture features on T2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness. Phys Med Biol 2015;60:2685–2701. [PubMed: 25768265]
16. Gnep K, Fargeas A, Gutierrez-Carvajal RE, et al. Haralick textural features on T2 - weighted MRI are associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer. J Magn Reson Imaging 2017;45:103–117. [PubMed: 27345946]
17. Nketiah G, Elschot M, Kim E, et al. T2-weighted MRI-derived textural features reflect prostate cancer aggressiveness: preliminary results. Eur Radiol 2017;27:3050–3059. [PubMed: 27975146]
18. Algohary A, Viswanath S, Shiradkar R, et al. Radiomic features on MRI enable risk categorization of prostate cancer patients on active surveillance: Preliminary findings. Journal of Magnetic Resonance Imaging 2018;48:818–828.
19. Castillo TJM, Starmans MPA, Niessen WJ, Schoots I, Klein S, Veenland JF. Classification of prostate cancer: High grade versus low grade using a radiomics approach. International Symposium on Biomedical Imaging. Venice, Italy: IEEE; 2019. p. 1319–1322.
20. Varghese B, Chen F, Hwang D, et al. Objective risk stratification of prostate cancer using machine learning and radiomics applied to multiparametric magnetic resonance images. Scientific Reports 2019;9:1–10. [PubMed: 30626917]

21. Zhang GM, Han YQ, Wei JW, et al. Radiomics Based on MRI as a Biomarker to Guide Therapy by Predicting Upgrading of Prostate Cancer From Biopsy to Radical Prostatectomy. J Magn Reson Imaging 2020;52:1239–1248. [PubMed: 32181985]
22. Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. European Urology 2016;69:16–40. [PubMed: 26427566]
23. Sonn GA, Natarajan S, Margolis DJ, et al. Targeted biopsy in the detection of prostate cancer using an office based magnetic resonance ultrasound fusion device. J Urol 2013;189:86–91. [PubMed: 23158413]
24. Matoso A, Epstein JI. Defining clinically significant prostate cancer on the basis of pathological findings. Histopathology 2019;74:135–145. [PubMed: 30565298]
25. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. Nephron Clin Pract 2010;115:94–99.
26. Cao R, Mohammadian Bajgiran A, Afshari Mirak S, et al. Joint Prostate Cancer Detection and Gleason Score Prediction in mp-MRI via FocalNet. IEEE Trans Med Imaging 2019;38:2496–2506. [PubMed: 30835218]
27. Cao R, Zhong X, Shakeri S, et al. Prostate Cancer Detection and Segmentation in Multi-parametric MRI via CNN and Conditional Random Field. International Symposium on Biomedical Imaging. Venice, Italy: IEEE; 2019. p. 1900–1904.
28. Novak CL, Liu P, Aylward S, et al. A prostate cancer computer-aided diagnosis system using multimodal magnetic resonance imaging and targeted biopsy labels. Medical Imaging 2013: Computer-Aided Diagnosis; 8670.
29. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 2010;29:1310–1320. [PubMed: 20378467]
30. Hectors SJ, Cherny M, Yadav KK, et al. Radiomics Features Measured with Multiparametric Magnetic Resonance Imaging Predict Prostate Cancer Aggressiveness. J Urol 2019;202:498–505. [PubMed: 30958743]
31. Van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 2017;77:104–107.
32. Zongker D, Jain A. Algorithms for Features Selection: An Evaluation. International Conference on Pattern Recognition. Vienna, Austria, Austria: IEEE; 1996.
33. Hawkins DM, Basak SC, Mills D. Assessing Model Fit by Cross-Validation. Journal of Chemical Information and Computer Science 2003;43:579–586.
34. Peng Y, Jiang Y, Yang C, et al. Quantative Analysis of Multiparametric Prostate MR Images: Differentiation between Prostate Cancer and Normal Tissue and Correlation with Gleason Score - A Computer-aided Diagnosis Development Study. Radiology 2013;267:787–796. [PubMed: 23392430]
35. Sutton EJ, Dashevsky BZ, Oh JH, et al. Breast cancer molecular subtype classifier that incorporates MRI features. J Magn Reson Imaging 2016;44:122–129. [PubMed: 26756416]
36. Pei C, Sun Y, Zhu J, et al. Ensemble Learning for Early-Response Prediction of Antidepressant Treatment in Major Depressive Disorder. J Magn Reson Imaging 2020;52:161–171. [PubMed: 31859419]
37. DeLong ER, Delong DM, Clarke-Pearon DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics 1988;44:837–845. [PubMed: 3203132]
38. Marcus D Ruopp NJP, Brain W. Whitcomb, and Enrique F. Schisterman. Youden Index and Optimal Cut-Point Estimated from Observations Affected by a Lower Limit of Detection. Biometrical Journal 2008;50:419–430. [PubMed: 18435502]
39. Moldovan PC, Van den Broeck T, Sylvester R, et al. What Is the Negative Predictive Value of Multiparametric Magnetic Resonance Imaging in Excluding Prostate Cancer at Biopsy? A Systematic Review and Meta-analysis from the European Association of Urology Prostate Cancer Guidelines Panel. European Urology 2017;72:250–266. [PubMed: 28336078]

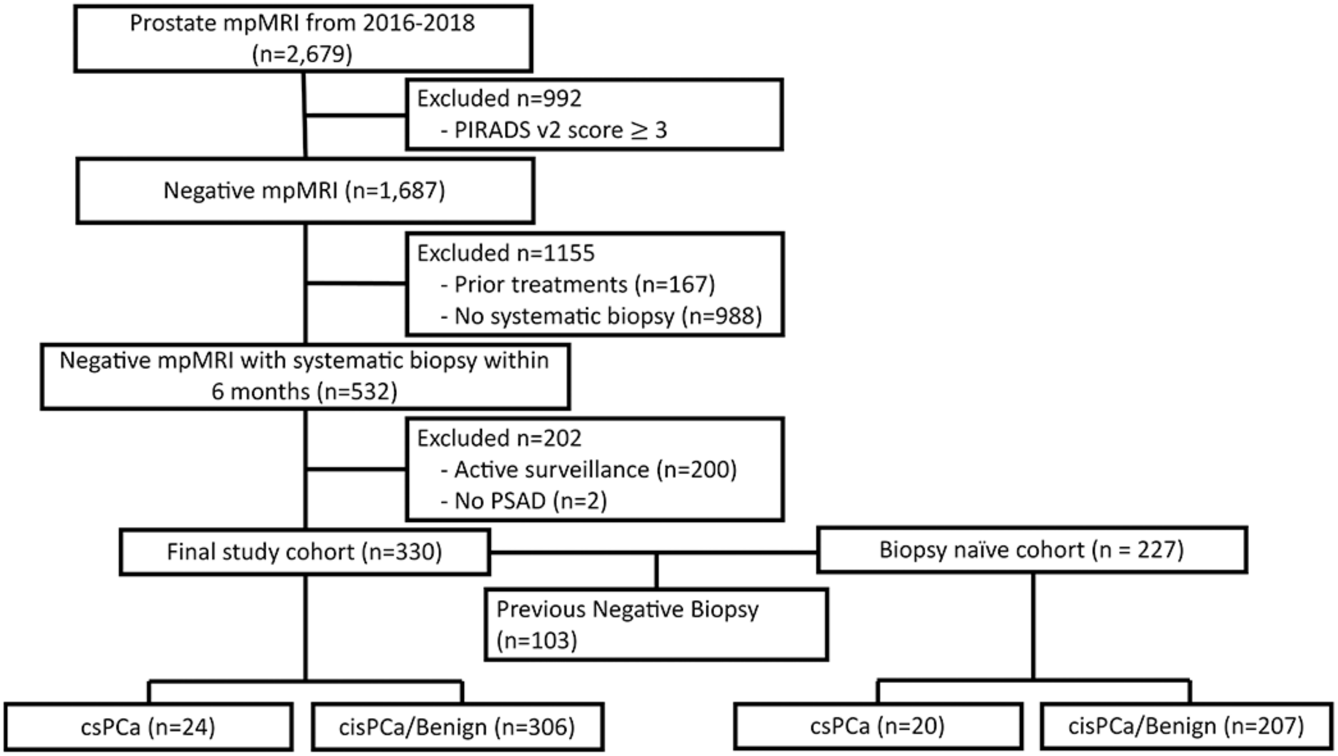

Figure 1:

Patients cohorts selection pipeline. Finally, we generated two cohorts: a final study cohort (n=330) and a sub-cohort of biopsy naïve sub-cohort (n=227), which were used for further model construction, validation, and evaluation.

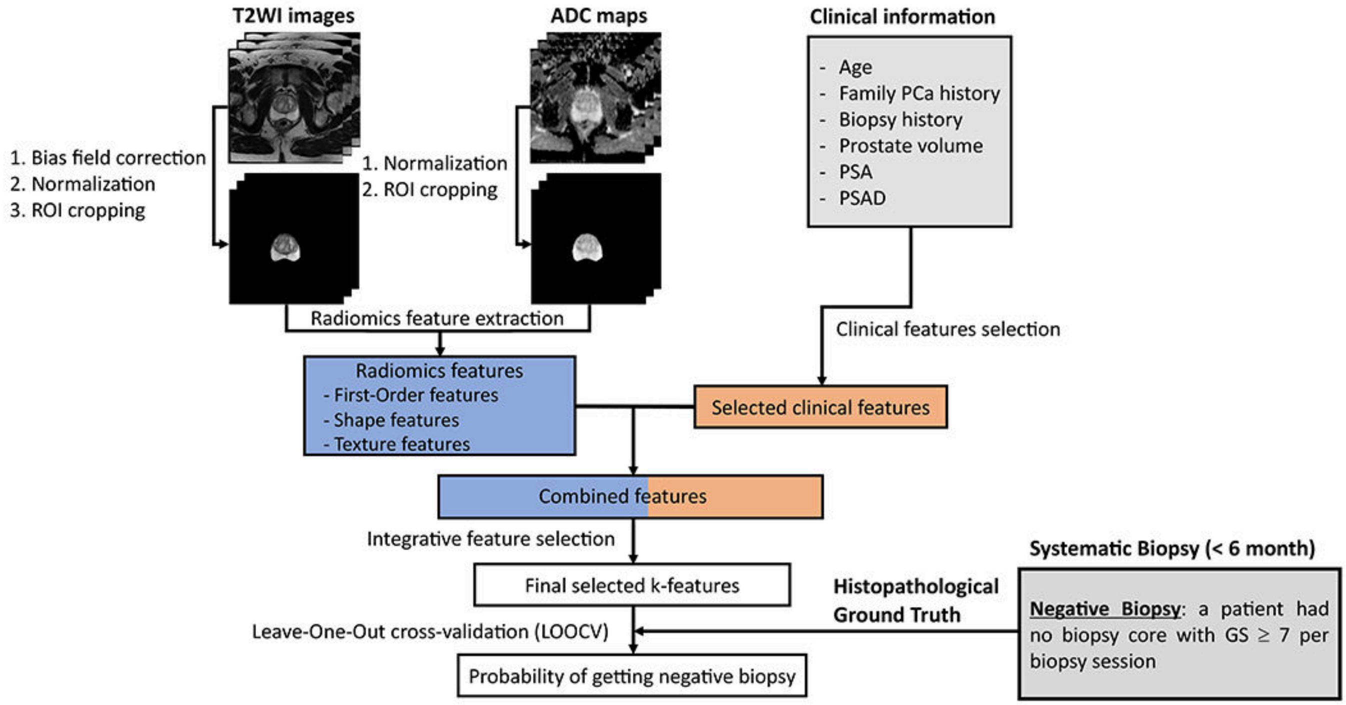

Figure 2:
Workflow of building the integrative machine learning (iML) model for predicting negative biopsy results. The three inputs of the model are the patient's clinical information, T_2WI, and ADC images. First, clinical features were selected from all clinical information, and radiomics features were extracted from the T_2WI and ADC images that have been pre-processed and cropped based on ROI. Then, integrative feature selection was made based on the combination of the two categories of features. Finally, with the selected features, Leave-one-out cross-validation (LOOCV) was performed to evaluate the model's predictability.

Author Manuscript

Author Manuscript

Author Manuscript

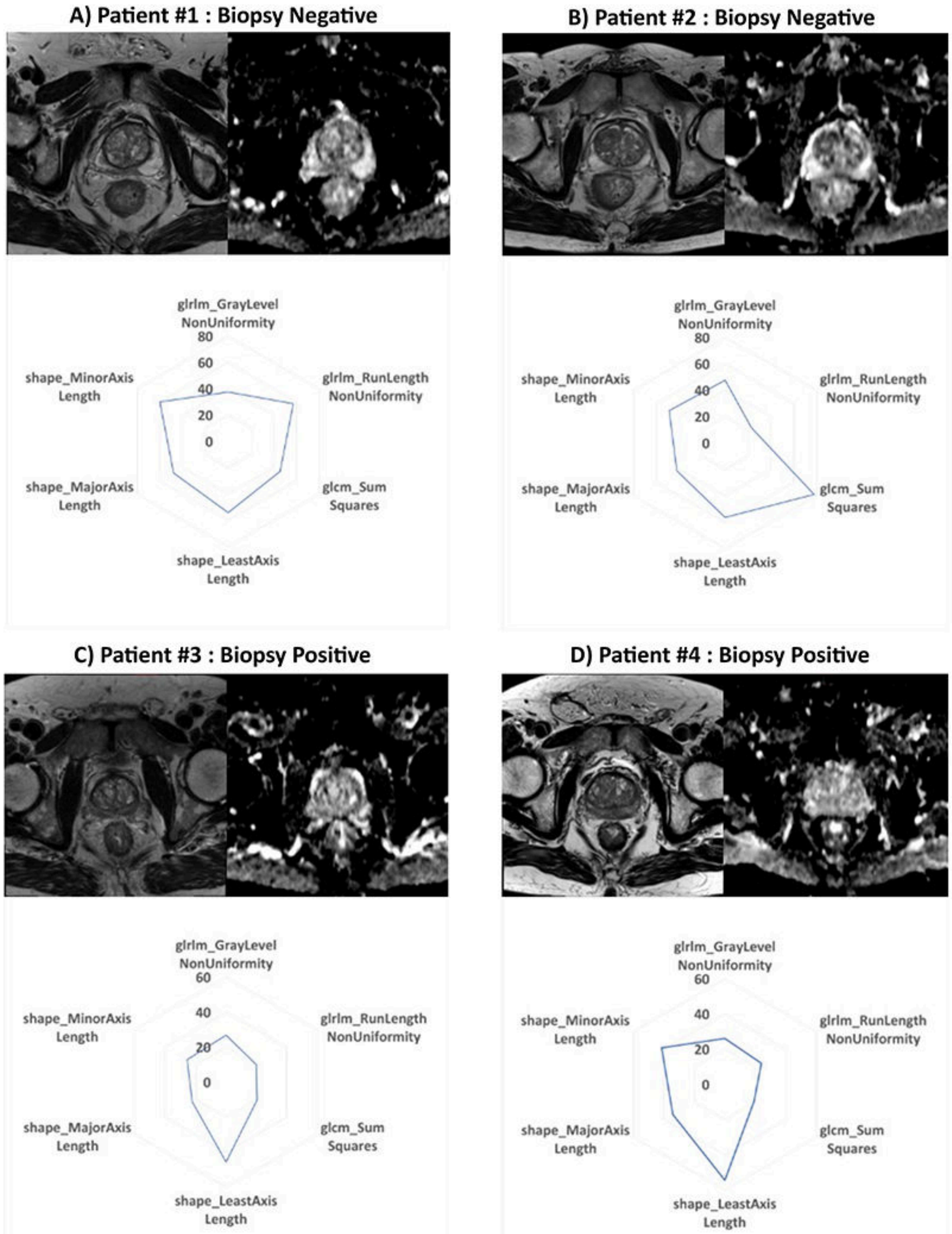

Figure 3:

Visualizations of mpMRI (T_2WI and ADC) images and the values of their corresponding radiomics features for patients with negative mpMRI; A) and B): patients with negative biopsies, and C) and D): patients with positive biopsies. Visualizations of the radiomics feature values are shown in spider plots, where the length of a feature's spoke is proportional to the value of that feature relative to that feature's maximum values across all patients. The numbers adjacent to each level of polygon represents the proportion value of the spoke at that level.

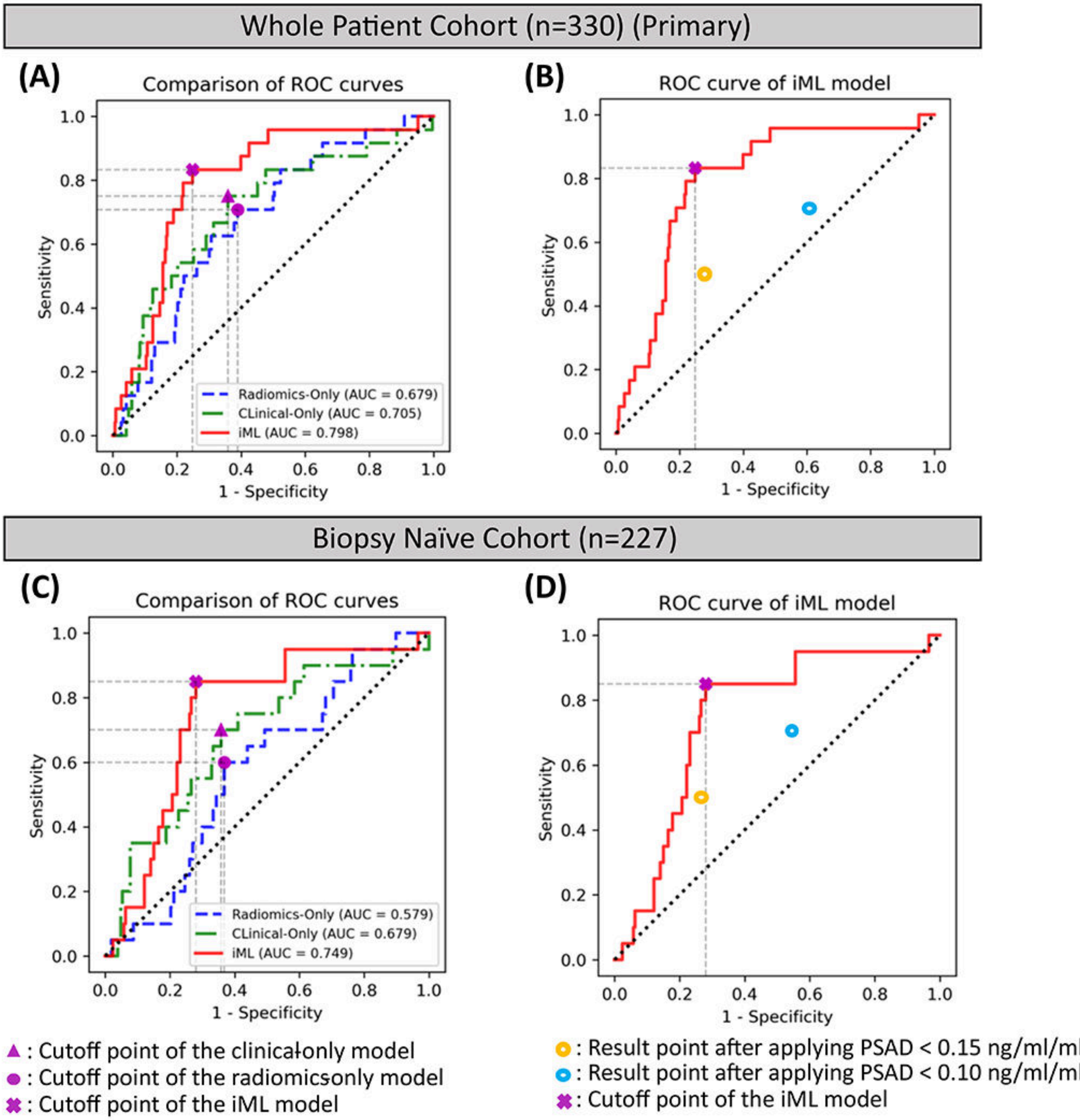

Figure 4:

Comparisons between iML and machine learning approaches using individual feature groups for both patient cohorts. Red solid, blue dash, and green dot-dash curves are the ROC curves of the radiomics-only, clinical-only, and iML models. Horizontal and vertical gray dash lines of each optimal cutoff point aimed to visualize sensitivity and “1-specificity” value on each ROC curve.

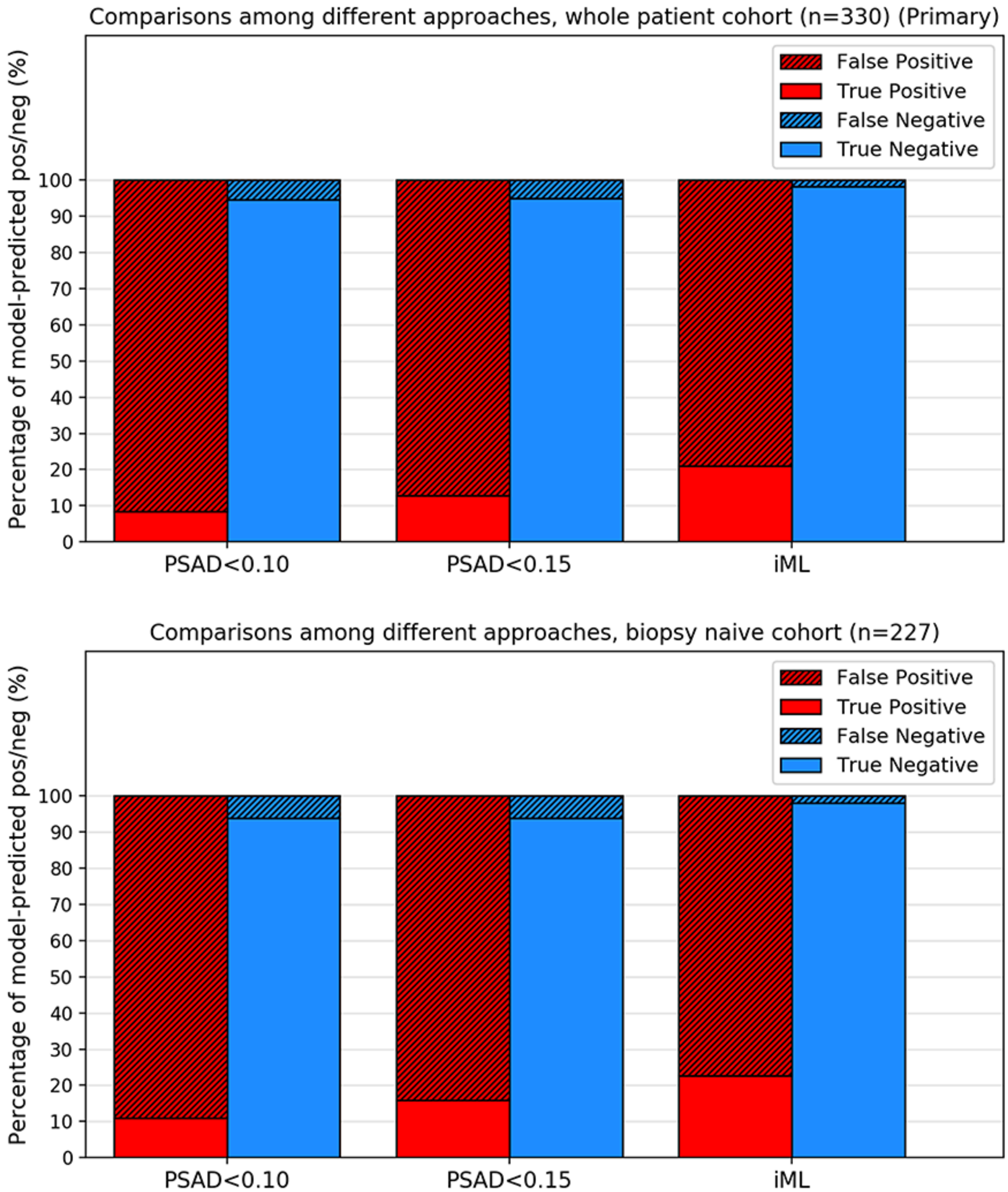

Figure 5:

Histogram Visualization of performance comparisons between our proposed iML approach and PSAD-based strategies from other studies, for both final study cohorts (left) and biopsy naïve cohort (right). Performances are measured by the percentage of true positive, false positive, true negative, and false negative in both cohorts. Red bars reveal prediction performance based on cases that are predicted as having biopsy positive, and blue bars reveal prediction performances based on cases that are predicted as having biopsy negative.

Table 1:

Clinical information for 1) final study cohort and 2) biopsy naïve sub-cohort and the p-values that reflect the significance of the difference between biopsy positive and negative group within each cohort, respectively.

Feature Name	Final Study Cohort (N=330)				Biopsy Naïve Sub-cohort (N=227)			
	Overall	Biopsy Positive	Biopsy Negative	P-value	Overall	Biopsy Positive	Biopsy Negative	P-value
No. of men (count {% of overall})	330 {100}	24 {7.3}	306 {92.7}	-	227 {100}	20 {8.8}	207 {91.2}	-
Prostate Volume (cc) (median {IQR})	55 {39-73}	49 {26-65}	55 {40-77}	0.02	53 {36-68}	42 {26-57}	54 {38-70}	0.01
Age (yr) (median {IQR})	63 {58-67}	65 {62-70}	62 {58-67}	0.02	62 {57-67}	65 {63-68}	62 {57-66}	0.02
PSA (ng/ml) (median {IQR})	6.3 {4.6-8.9}	6.7 {4.5-8.4}	6.3 {4.6-9.0}	0.46	5.7 {4.4-8.0}	6.3 {4.5-8.1}	5.7 {4.4-8.0}	0.32
PSAD(ng/ml/ml) (median {IQR})	0.11 {0.08-0.16}	0.15 {0.09-0.22}	0.11 {0.08-0.15}	0.02	0.11 {0.08-0.15}	0.15 {0.09-0.22}	0.11 {0.08-0.15}	0.02
PCa Family History (Yes/No: 1/0) (count {% of overall})	70 {100}	7 {10.0}	63 {90.0}	0.46	42 {100}	7{16.7}	35{83.3}	0.09
Prostate Biopsy History (Yes/No: 1/0) (count {% of overall})	103 {100}	4 {3.9}	99 {96.1}	0.17	-	-	-	-

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Nine selected features after integrative feature selection.

Selected Features	Type	Imaging Sequence
Gray Level Non-uniformity	GLRLM	ADC
Run Length Non-uniformity	GLRLM	ADC
Sum Squares	GLCM	T_2WI
Least Axis Length	Shape	ADC/ T_2WI
Major Axis Length	Shape	ADC/ T_2WI
Minor Axis Length	Shape	ADC/ T_2WI
Age	Clinical Information	--
PSAD	Clinical Information	--
Prostate Volume	Clinical Information	--

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

Comparisons of prediction performances of the proposed iML approach and the machine learning approaches that were clinical-only or radiomics-only for both final study cohort and biopsy naïve sub-cohort, respectively. P-values were calculated by DeLong test, for comparisons between AUCs of models using each individual feature group and the proposed iML model.

Method	Final Study Cohort (N=330)				
	AUC [%95 CI]	**Sensitivity (%)**	**Specificity (%)**	**NPV (%)**	**p-value**
Clinical-only	0.705 [0.589, 0.821]	75.0	64.1	97.0	0.011
Radiomics-only	0.679 [0.571, 0.787]	70.8	61.2	96.4	0.006
iML	**0.798 [0.711, 0.885]**	**83.3**	**75.2**	**98.3**	–
Method	**Biopsy Naïve Sub-cohort (N=227)**				
	AUC [%95 CI]	**Sensitivity (%)**	**Specificity (%)**	**NPV (%)**	**p-value**
Clinical-only	0.679 [0.553, 0.805]	70.0	64.3	95.7	<0.001
Radiomics-only	0.579 [0.464, 0.694]	60.0	63.3	94.2	0.046
iML	**0.749 [0.645, 0.854]**	**85.0**	**72.0**	**98.0**	–

Table 4:

Comparisons of prediction performances of the proposed iML approach and approaches that using PSAD-based risk assessments for both final study cohort and biopsy naïve sub-cohort, respectively. P-values were calculated by Chi-square test, for comparisons of each measurement between PSAD-based prediction models and the proposed iML model.

Method	Final Study Cohort (N=330)					
	Sensitivity (%)	p value	Specificity (%)	p value	NPV (%)	p value
PSAD < 0.10	70.8	0.303	39.9	<0.001	94.6	0.048
PSAD < 0.15	50.0	0.014	73.2	0.579	94.9	0.044
iML	**83.3**	–	**75.2**	–	**98.3**	–
Method	**Biopsy Naïve Sub-cohort (N=227)**					
	Sensitivity (%)	**p value**	**Specificity (%)**	**p value**	**NPV (%)**	**p value**
PSAD < 0.10	70.0	0.451	44.9	<0.001	93.9	0.171
PSAD < 0.15	50.0	0.018	**74.4**	0.579	93.9	0.010
iML	**85.0**	–	72.0	–	**98.0**	–