OPEN ACCESS

Digital Medicine ARTICLE

The Innovation Oncology

Multimodal deep learning model for predicting homologous recombination deficiency in prostate cancer: A multi-cohort study

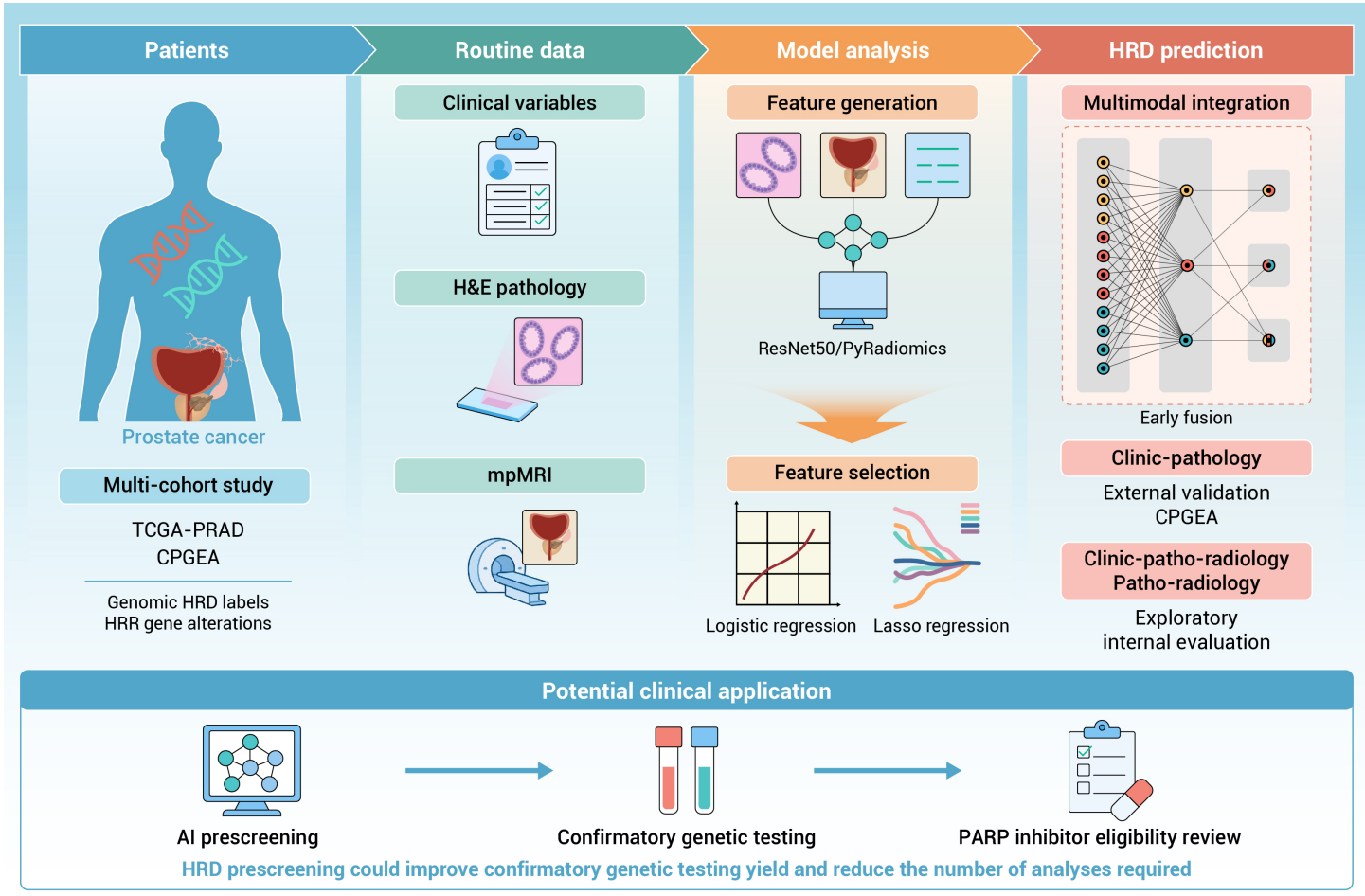
Longxin Deng,[1,2,8] Zijian Song,[2,8] Qianwen Zhang,[3,8] Wenhui Zhang,[1,8] Na Ta,[4,8] Yan Zhu,[4,8] Han Wu,[1] Lingxuan Zhu,[1] Yanan Gao,[1] Yaoming Li,[5] Xiao Wang,[6] Yang Yang,[7] Wei Zhang,[1] Yancheng Lai,[1] Yan Wang,[1] Min Qu,[1] Zhenyang Dong,[1] Huamao Ye,[1] Linhui Wang,[1,*] Gao Xu,[1,*] and Rui Chen[2,*]

*Correspondence: drchenrui@foxmail.com (R.C.); gaoxu.changhai@foxmail.com (G.X.); wanglinhuicz@163.com (L.W.)

Received: May 17, 2026; Accepted: October 20, 2026; Published Online: October 22, 2026; https://doi.org/10.59717/j.xinn-oncol.2026.100043

© 2026 The Author(s). This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- A model combined clinical data, tissue images, and magnetic resonance imaging in prostate cancer.
- Image-based models showed predictive signal for homologous recombination deficiency.
- Combined imaging showed a favorable trend in a small internal validation set.
- Larger prospective cohorts are needed before clinical use.

ARTICLE Digital Medicine

OPEN ACCESS

Multimodal deep learning model for predicting homologous recombination deficiency in prostate cancer: A multi-cohort study

Longxin Deng,[1,2,8] Zijian Song,[2,8] Qianwen Zhang,[3,8] Wenhui Zhang,[1,8] Na Ta,[4,8] Yan Zhu,[4,8] Han Wu,[1] Lingxuan Zhu,[1] Yanan Gao,[1] Yaoming Li,[5] Xiao Wang,[6] Yang Yang,[7] Wei Zhang,[1] Yancheng Lai,[1] Yan Wang,[1] Min Qu,[1] Zhenyang Dong,[1] Huamao Ye,[1] Linhui Wang,[1,*] Gao Xu,[1,*] and Rui Chen[2,*]

[1]Department of Urology, Shanghai Changhai Hospital, Second Military Medical University, Shanghai 200433, China
[2]Department of Urology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China
[3]Department of Radiology, Shanghai Changhai Hospital, Second Military Medical University, Shanghai 200433, China
[4]Department of Pathology, Shanghai Changhai Hospital, Second Military Medical University, Shanghai 200433, China
[5]Department of Urology, Daping Hospital, Army Medical University, Chongqing 400042, China
[6]Department of Urology, First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310003, China
[7]Department of Clinical Laboratory, Nanjing Jinling Hospital, Nanjing University School of Medicine, Nanjing 210002, China
[8]These authors contributed equally
*Correspondence: drchenrui@foxmail.com (R.C.); gaoxu.changhai@foxmail.com (G.X.); wanglinhuicz@163.com (L.W.)

Received: May 17, 2026; Accepted: October 20, 2026; Published Online: October 22, 2026; https://doi.org/10.59717/j.xinn-oncol.2026.100043

© 2026 The Author(s). This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Citation: Deng L., Song Z., Zhang Q., et al. (2026). Multimodal deep learning model for predicting homologous recombination deficiency in prostate cancer: A multi-cohort study. *The Innovation Oncology* 1:100043.

Accurate prediction of homologous recombination deficiency (HRD) status is crucial for effective management of prostate cancer (PCa). However, genetic testing is expensive and inaccessible. We propose a multimodal deep learning approach that integrates clinical information, Hematoxylin & Eosin (H&E)-stained whole-slide images (WSIs), and multiparametric magnetic resonance imaging (MRI) to predict the HRD status of patients with PCa. Patients from The Cancer Genome Atlas (TCGA; n=486) and three Chinese hospitals (n=210) were used to develop and evaluate the prediction models. The Pathology Signature, Radiology Signature, and Patho-Radiology Signature showed predictive performance for HRD status in external validation or internal held-out evaluation (area under the receiver operating characteristic curve [AUC]=0.815, 0.833, and 0.900, respectively). Notably, four interpretable pathological features were identified in the Pathology Signature. These signatures may serve as a prescreening tool to select patients for confirmatory genetic testing. This multimodal framework may be adaptable to the prediction of molecular alterations in other malignancies, subject to independent validation. Because radiology and multimodal models were evaluated in small modality-complete subsets, these estimates should be interpreted as exploratory trends.

INTRODUCTION

Prostate cancer (PCa) is the most commonly diagnosed malignancy and the fifth leading cause of cancer-related death in men worldwide.[1] PCa is a malignancy that is characterized by diverse molecular alterations.[2] Homologous recombination deficiency (HRD) is a frequently observed molecular alteration in PCa.[3] HRD leads to defective DNA break repair, increased somatic copy number alterations, genomic instability, and oncogenesis.[4,5] HRD has emerged as a potent biomarker for selecting patients for Poly(ADP-Ribose)-polymerase (PARP) inhibitor (PARPi) treatment of various cancers, including advanced-stage PCa, as demonstrated in the PROfound study.[6] However, HRD detection typically relies on genetic sequencing, which is associated with limited accessibility, high cost, and a long detection period.[7] Considering that the prevalence of HRD is approximately 5-10% in primary PCa and approximately 13% in advanced-stage PCa,[8] ten genetic tests are required to identify one patient with HRD.[3] Therefore, there is an urgent need for a rapid, low-cost, and accessible method to detect HRD status or, at the very least, prescreening of patients with a higher risk of HRD status. In addition, patients who consider HRD detection usually have advanced-stage PCa. They typically have Hematoxylin & Eosin (H&E)-stained tissue sections and multiparameter Magnetic Resonance Imaging (MRI) scans for disease staging and therapeutic evaluation. Consequently, utilizing these existing datasets to develop a predictive model for HRD status could optimize the diagnostic efficiency and could advance the clinical management of PCa.

Artificial intelligence (AI), particularly deep learning (DL), has shown great promise in identifying pathological subtypes, molecular alterations, and prognostic outcomes in cancer patients.[9] Prediction of Microsatellite Instability (MSI) in gastrointestinal cancer is one of the most investigated applications. Kather et al. proposed the application of deep learning in the analysis of conventional Hematoxylin & Eosin (H&E) histology to predict microsatellite instability based on multiple international cohorts.[10] Yamashita et al. proposed DL models that surpassed the performance of experienced gastrointestinal pathologists in predicting MSI on H&E-stained whole-slide images (WSIs) in colorectal cancer.[11]

However, despite the progress made in utilizing deep learning methods for predicting molecular alterations in various cancer types, a comprehensive review highlighted the limited exploration in this field in the context of PCa.[12] Chinnaiyan et al. reported the application of DL in the analysis of conventional pathological slides to predict the presence of E26 transformation-specific (ETS)-related gene (ERG) fusions in a cohort of 392 cases.[13] Mark et al. reported the application of DL-based methods for speckle-type POZ protein (SPOP) mutation prediction in 177 PCa patients, but the study was not officially published.[14] The main limitation of these studies may be the small sample size due to poor data accessibility, as there were a limited number of PCa cases with available pathology images and molecular alteration data. Furthermore, studies aimed at predicting molecular alterations in PCa using Magnetic Resonance Imaging (MRI) data are still in the preliminary stages and involve a limited number of patients.[15]

Recently, multimodal data were analyzed using DL algorithms in the risk stratification of high-grade serous ovarian cancer,[16] in the prediction of response to programmed cell death protein 1/programmed death-ligand 1 [PD-(L)1] blockade in patients with non-small cell lung cancer,[17] and in the prediction of molecular classification of endometrial cancer.[18] In a previous study, we presented the Chinese Prostate Cancer Genetic and Epigenetic Atlas (CPGEA) cohort, which is one of the largest multi-omics datasets of PCa, encompassing genomic, transcriptome, and epigenetic data.[19] In this study, for the first time, we collected and analyzed the pathological and imaging data of patients in the CPGEA cohort. Using external validation for pathology and fixed internal held-out evaluation, with training-only cross-validation, for radiology and multimodal analyses, we aimed to develop and evaluate multimodal models for predicting the HRD status in patients with PCa.

MATERIALS AND METHODS

This study complied with relevant ethical regulations. The study protocols were approved by the Institutional Review Board of Shanghai Changhai Hospital (approval number: CHEC2022-151). This study was registered with the Chinese Clinical Trial Registry (ChiCTR2200064329). Given the retrospective nature of this study, the requirement for individual informed consent was waived with explicit approval granted by the Institutional Review Board of Shanghai Changhai Hospital. Participants were not compensated. Two cohorts of hormone-sensitive PCa patients were included in this study. The study design is illustrated in Figure 1.

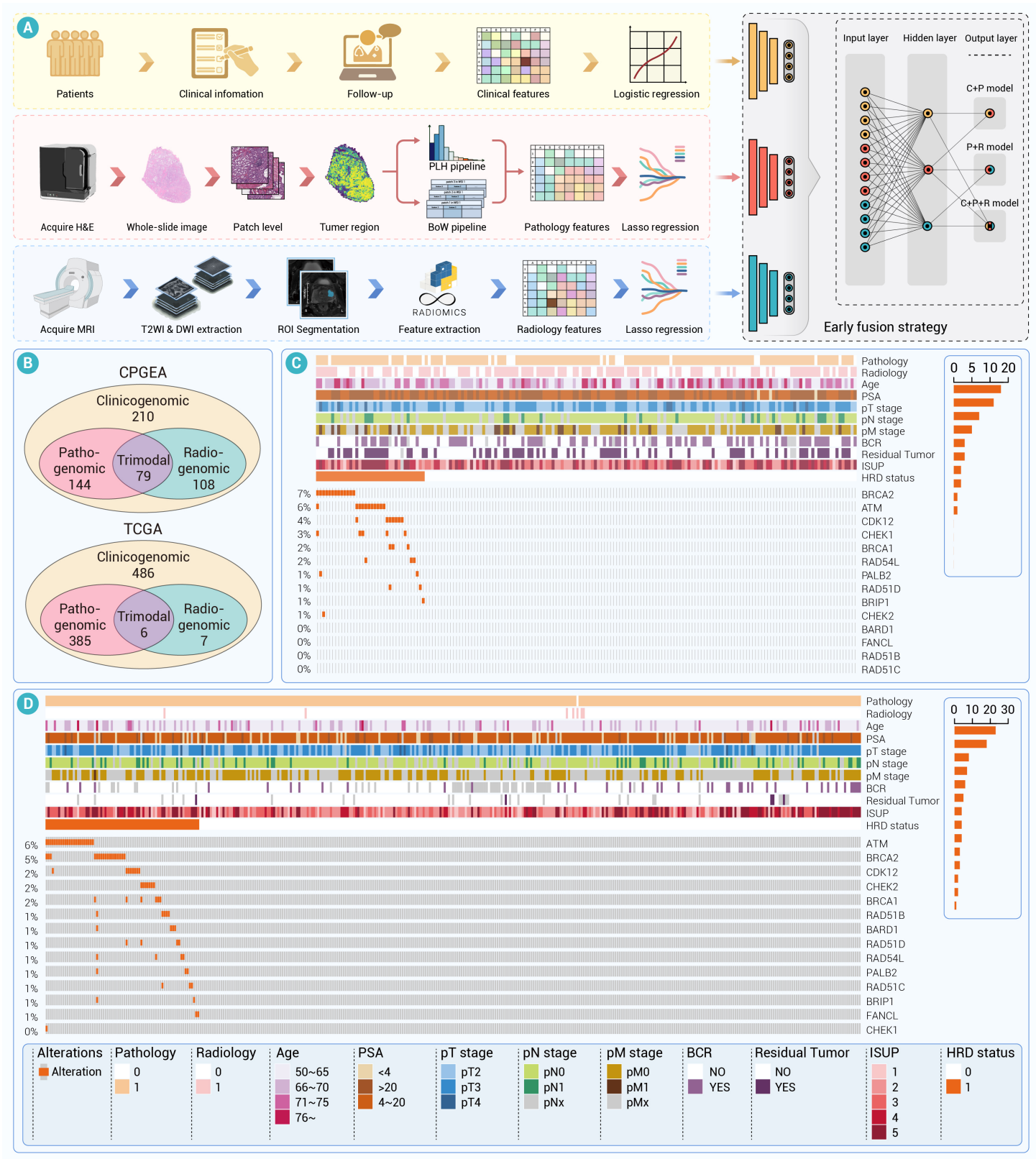


Figure 1. Schematic outline of the study and multimodal cohort characteristics (A) Workflow of this study. The multimodal models incorporate clinical, genomic, pathological, and radiological features for predicting homologous recombination deficiency (HRD) status in prostate cancer patients. (B) The Venn diagram depicts the distribution of patients across clinical, genomic, pathological, and radiomic categories. A multimodal cohort heatmap displays clinical, genomic, pathological, and radiomic features for the patients in (C) the TCGA-PRAD cohort and (D) the CPGEA cohort. The stacked bar chart reveals the type of genetic mutation and the presence of corresponding clinical, genomic, pathological, and radiomic features in each patient. HRD=homologous recombination deficiency. TCGA-PRAD=The Cancer Genome Atlas Prostate Adenocarcinoma. CPGEA= Chinese Prostate Cancer Genetic and Epigenetic Atlas. BCR=Biochemical Recurrence. ISUP=International Society of Urological Pathology.

Cohort description

The clinicogenomic analytic cohort comprised 486 patients from The Cancer Genome Atlas Prostate Adenocarcinoma (TCGA-PRAD) cohort and 210 pathologically confirmed patients from CPGEA, all with available clinical variables and genomic HRD labels.[19,20] Diagnostic H&E-stained WSIs were initially available for 386 TCGA-PRAD and 157 CPGEA patients; after slide

quality control and confirmation of valid pathology features, 385 and 144 patients, respectively, were retained for pathology modeling. Eligible MRI data after quality control were available for 7 TCGA-PRAD and 108 CPGEA patients. The CPGEA cohort was sequenced and reported previously,[19] and its WSIs and MRI data were curated for the present study. Clinicopathological information was obtained from a prospectively maintained database. Model-specific eligibility required the corresponding clinical, genomic, pathology, and/or radiology data, as detailed in the study design and Figure 1. For clarity, MRI data were initially available for 118 patients (8 TCGA-PRAD and 110 CPGEA); after excluding three inadequate scans (1 TCGA-PRAD and 2 CPGEA), 115 eligible patients remained (7 TCGA-PRAD and 108 CPGEA).

Study design

For the analysis of pathology and clinical data, we selected the TCGA-PRAD cohort as the training cohort and the CPGEA cohort as the test cohort. Because few TCGA-PRAD patients had MRI, patients with eligible MRI were pooled and randomly assigned to training and test sets at an 80:20 ratio (Figure S1). All data-dependent preprocessing and feature-selection steps, including missing-value imputation, categorical encoding, scaling, correlation filtering, and least absolute shrinkage and selection operator (LASSO) feature selection, were fitted using only the training data within each outer split. The LASSO penalty and model hyperparameters were selected by inner cross-validation, and the fitted transformations were applied unchanged to the held-out data. For same-patient comparisons, the Patho-Clinical and Radio-Clinical models used the same 385/144 and 92/23 training/test splits as their corresponding Pathology and Radiology models, respectively; Pathology, Radiology, and Patho-Radiology were compared in a fixed 85-patient cohort (68 training and 17 test patients). The distributions of annotator assignments and equipment types did not differ significantly between the training and test sets (Figure S2).

Inferring genetic alteration status

The HRD status was ascertained based on alterations in 14 homologous recombination repair (HRR)-related genes, as previously demonstrated (including ATM, BRCA1, BRCA2, BARD1, BRIP1, CDK12, CHEK1, CHEK2, FANCL, PALB2, RAD51B, RAD51C, RAD51D, and RAD54L).[21] Patients were assigned to the HRD subtype if they had at least one nonsynonymous mutation or deep deletion in the HRR-related genes.

Radiology pipeline

A flowchart of the radiology pipeline is shown in Figure 2. Prostate MRI was performed on a 3.0-T MRI scanner with an abdominal phase array coil, following a 4h fasting period and enema treatment with glycerin (20 mL). Routine sequences included sagittal T2-weighted imaging (T2WI), axial high-resolution T2WI, and axial diffusion-weighted imaging (DWI). Table S1 shows the axial T2WI and DWI parameters used for model development. Each radiologist traced the outer contour of the prostate lesions on the tumor-containing axial section using Insight Segmentation and Registration Toolkit-SNAP v3.8.0 software. Handcrafted features were extracted using PyRadiomics (https://pyradiomics.readthedocs.io/en/latest/).

Pathology pipeline

A flowchart of the pathology pipeline is shown in Figure 3. All WSIs were digitally captured using a 20× objective lens. The task of annotating the H&E WSIs to identify the tumor region was performed by two fellowship-trained uropathologists using the QuPath software (https://github.com/qupath). The WSIs were partitioned into 512×512-pixel patches. The 50-layer residual network (ResNet50) algorithm was then used to predict the tumor region based on annotations by pathologists. The patch label and corresponding probability were extracted and applied to generate WSI-level features via Patch Likelihood Histogram (PLH) and bag-of-words (BoW) pipelines. WSI-level features were applied to construct pathological models for HRD status prediction.

Multimodal integration

Before early fusion, binary clinical variables were represented numerically, while ordinal variables, including pathological tumor (T), node (N), and metastasis (M) stages and International Society of Urological Pathology (ISUP) grade, were retained as ordered numerical features. No one-hot or latent encoding was applied. Numerical features from each modality were standardized feature-wise using the mean and standard deviation estimated from the training data, and the same parameters were applied unchanged to held-out data. After modality-specific feature selection, the standardized clinical, pathology, and radiology features were concatenated without additional modality-specific weighting and entered into ten machine-learning algorithms, each of which generated an HRD probability for each case. Model discrimination was summarized using the area under the curve (AUC) for receiver operating characteristic (ROC-AUC) and precision-recall (PR-AUC) curves, calibration curves, and decision-curve analysis (DCA). Results were displayed for the fixed Patho-Radiology analytic cohort (n=85), including the training set (n=68), held-out test set (n=17), and a descriptive all-cohort view. Pairwise ROC-AUC differences among the Pathology, Radiology, and Patho-Radiology scores were evaluated on the same patients using two-sided paired DeLong tests.

Visualization of Pathology Signature prediction with Gradient-weighted Class Activation Mapping (Grad-CAM)

Grad-CAM class localization maps are created by visualizing the gradients flowing into the final convolutional layer of the network immediately before the fully connected layers. Because the convolutional layers contain class-specific spatial information from the input image that is lost in the fully connected layers, this is the optimal point for map generation. Grad-CAM requires neither any modifications to the existing model architecture nor any retraining of the model.

Survival analysis

Associations of actual or signature-predicted HRD status with biochemical recurrence (BCR) were assessed using Cox proportional hazards models, which provided hazard ratios (HRs), 95% confidence intervals (CIs), and corresponding P values. These survival analyses were conducted independently of HRD classifier development. For the survival analyses, HRD status was coded as 1 and homologous recombination proficient (HRP) status as 0, with BCR specified as the event of interest. Kaplan–Meier curves and two-sided log-rank tests were used to visualize and compare BCR-free survival. Complete-case denominators and BCR event counts were reported for each signature and cohort analysis.

RESULTS

Cohorts and clinical characteristics

The clinicogenomic analytic cohort was defined by availability of both clinical variables and genomic HRD status and was verified using non-missing Clinical k-nearest-neighbor (KNN) scores. It comprised 486 patients from TCGA-PRAD and 210 from CPGEA (N=696). Patient characteristics for these patients are summarized in Table 1. The model-specific analytic sets were: Clinical, 486 TCGA-PRAD and 210 CPGEA; Pathology, 385 and 144; Radiology, 7 and 108; and Patho-Radiology, 6 and 79, respectively. Actual HRD was present in 90/486 (18.5%) TCGA-PRAD patients and 39/210 (18.6%) CPGEA patients. All score-file IDs were matched to the master clinical workbook by UNIQUEID (Figure 1).

Data are shown as mean (SD), median [IQR], or n (%). Table 1 includes the clinicogenomic analytic cohort of 696 patients with both clinical variables and genomic HRD labels, verified by available Clinical KNN scores; model-specific analytic sample sizes were defined separately by score availability. HRD was coded 1 and HRR-proficient/HRP was coded 0. Age was compared using Welch's t test, PSA using the Wilcoxon rank-sum test, and categorical variables using Pearson's chi-square test or Fisher's exact test as appropriate; missing observations were reported but excluded from hypothesis tests. Three-year BCR-free survival was estimated by Kaplan-Meier and compared using the log-rank test. Abbreviations: SD, standard deviation; IQR, interquartile range; HRP, homologous recombination proficient; HRD, homologous recombination deficient; ISUP, International Society of Urological Pathology; BCR, biochemical recurrence.

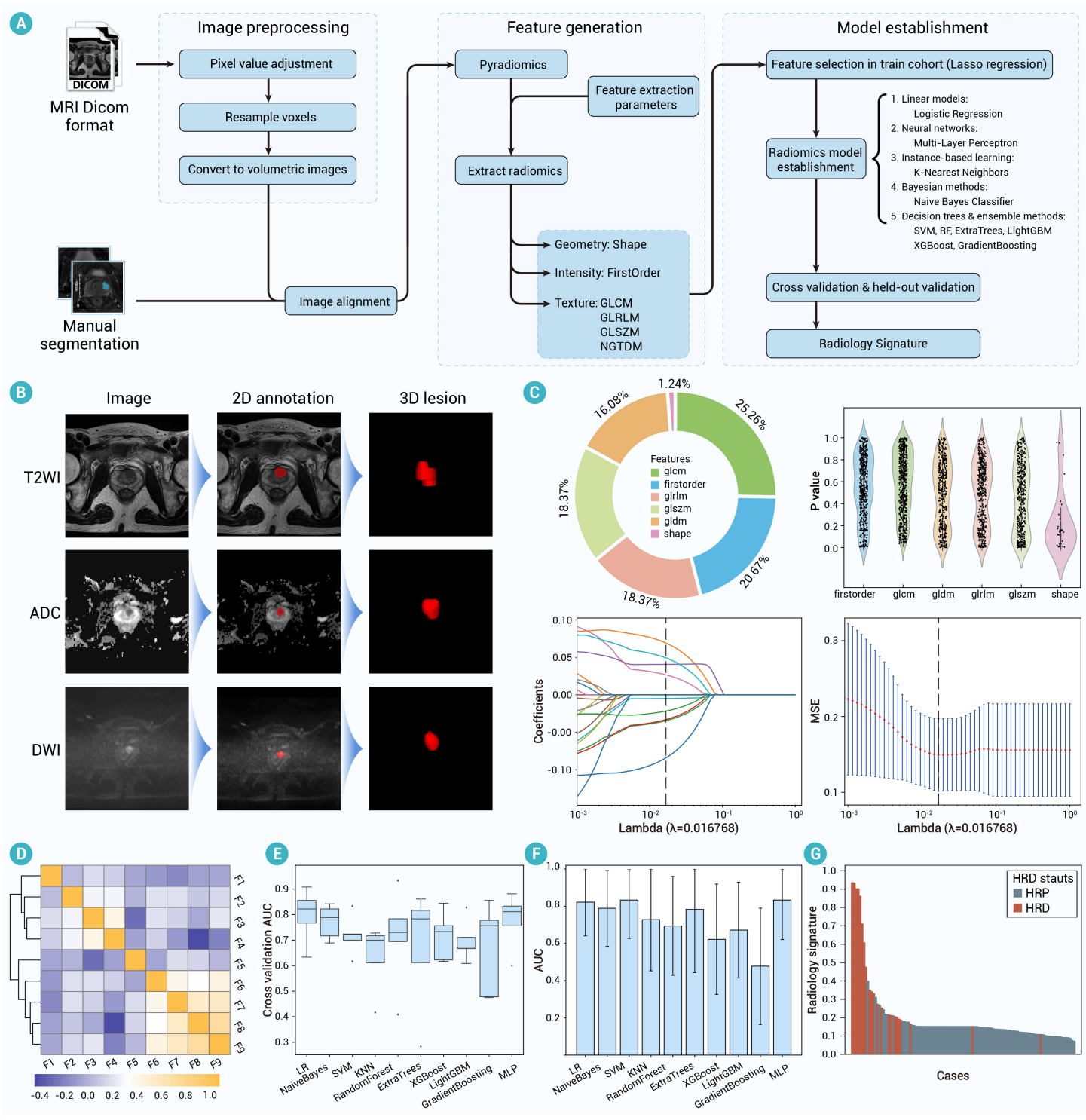


Figure 2. Establishment and validation of Radiology Signature (A) Radiology pipeline across image pre-processing, feature generation, and model establishment. (B) Pre-processing of the MRI images. (C) Feature generation and selection. The number, proportion, and *P* values of handcrafted features generated with PyRadiomics are summarized. Nine features of nonzero coefficients were selected by the least absolute shrinkage and selection operator (LASSO) logistic regression model to establish the radiology models. (D) Heatmap indicating the pairwise Spearman correlations among the nine selected radiology features, denoted as F1 to F9. (E) The distributions of AUCs obtained for the different radiology models during five-fold cross-validation. Boxes indicate the median and interquartile range, whiskers extend to 1.5 times the interquartile range, and points beyond the whiskers represent outliers. (F) Test-set AUCs of the radiology models. Error bars represent DeLong 95% confidence intervals calculated from patient-level predictions. (G) The waterfall plot of the Radiology Signature based on SVM indicating predicted Radiology Signature score and HRD status. AUC=area under the curve. SVM=Support Vector Machine. LASSO=least absolute shrinkage and selection operator. HRD=Homologous Recombination Deficiency. Feature identifiers in panel D: F1, wavelet_LHH_glcm_ClusterShade; F2, wavelet_LLH_glszm_ZoneVariance; F3, wavelet_LHL_glszm_LargeAreaHighGrayLevelEmphasis; F4, original_shape_LeastAxisLength; F5, wavelet_HHL_glszm_GrayLevelNonUniformityNormalized; F6, wavelet_LHL_firstorder_Skewness; F7, log_sigma_5_0_mm_3D_glszm_SmallAreaLowGrayLevelEmphasis; F8, log_sigma_4_0_mm_3D_glcm_MaximumProbability; F9, wavelet_HLL_firstorder_Median.

Clinical Signature

Clinical variables were not significantly associated with the HRD status (Figure S3). Although age and International Society of Urological Pathology (ISUP) grade may be associated with HRD, the predictive value of clinical variables is limited. In the TCGA-PRAD cohort, clinical variables such as age, prostate-specific antigen (PSA) level, pathological tumor-node-metastasis (TNM) stage, residual tumor, and ISUP grade were not significantly associated with HRD status in either the univariate or multivariate logistic regres-

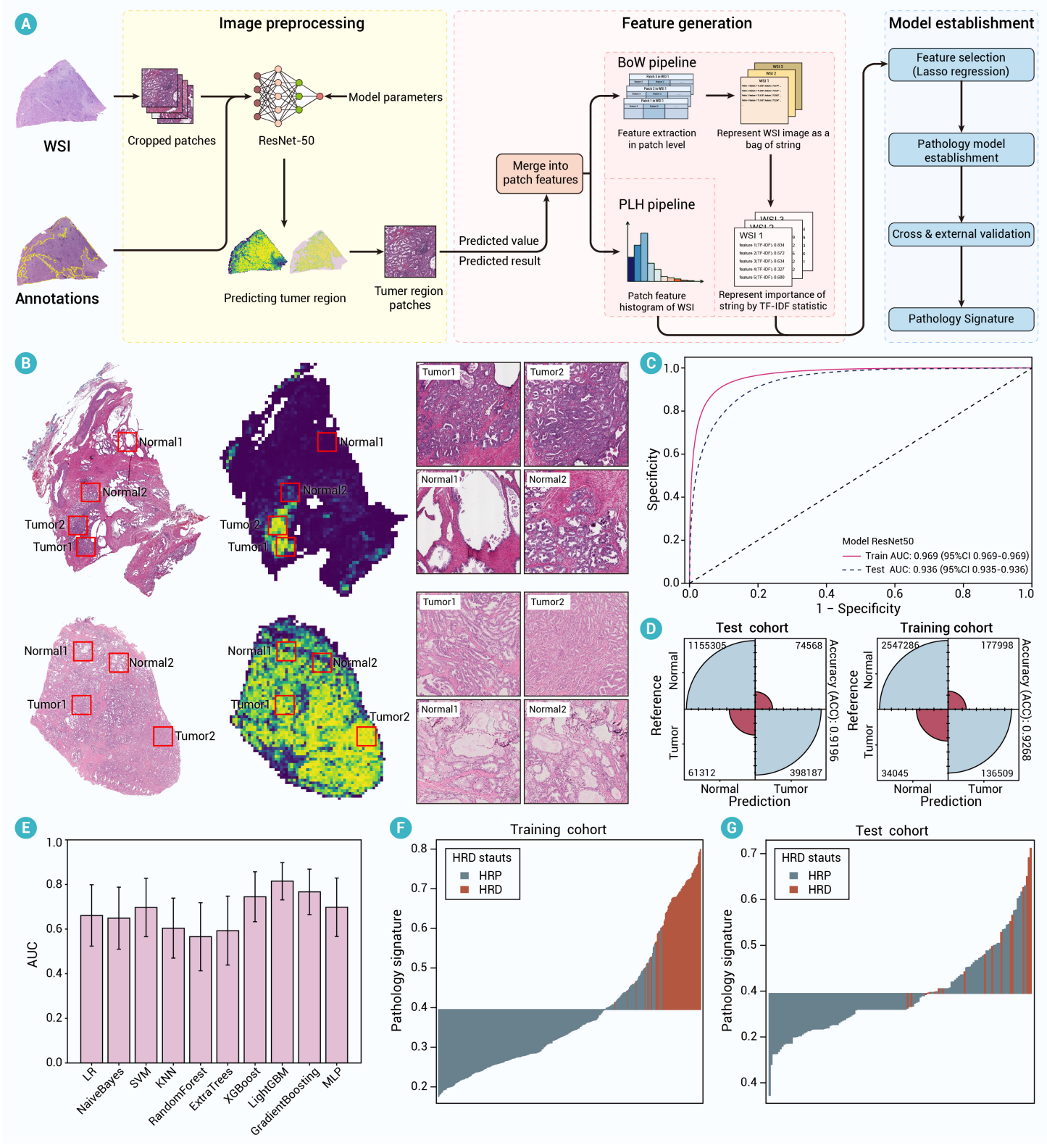


Figure 3. Establishment and validation of the Pathology Signature (A) Flow-chart of pathology pipeline, including image pre-processing, feature generation, and model establishment. (B) Patch-level tumor region prediction was performed using a ResNet50-based model. Examples from the TCGA-PRAD and the CPGEA cohort were illustrated, indicating high concordance of actual and predicted tumor regions. (C) The ROC plot of the ResNet50 model for distinguishing tumor and non-tumor patches in the training and test cohorts. (D) The confusion matrix plot of the ResNet50 model for distinguishing tumor and non-tumor patches in the training and test cohorts. (E) The performance of the pathology models in the test cohort. Bars indicate test-cohort AUCs, and error bars represent DeLong 95% confidence intervals calculated from patient-level predictions. The Pathology Signature was established based on LightGBM. (F) The waterfall plot depicted the association between the Pathology Signature score and the HRD status in the training and test cohorts. (G) Distribution of the Pathology Signature score and the HRD status in the training and test cohorts. HRD=Homologous Recombination Deficiency. TCGA=The Cancer Genome Atlas Prostate Cancer. CPGEA= Chinese Prostate Cancer Genetic and Epigenetic Atlas.

sion analyses. We constructed the Clinical Signature based on a k-nearest-neighbor (KNN) model with data from the TCGA-PRAD cohort and validated it in the CPGEA cohort, yielding an area under the curve (AUC) of 0.560. The Clinical Signature showed limited discrimination for HRD status.

Table 1. Clinical characteristics of the clinicogenomic analytic cohort in CPGEA and TCGA-PRAD according to actual HRD status.

Characteristic	CPGEA			TCGA-PRAD			Total		
	HRP (N=171)	HRD (N=39)	P value	HRP (N=396)	HRD (N=90)	P value	HRP (N=567)	HRD (N=129)	P value
AGE			0.8220			0.0123			0.0322
Mean (SD), years	68.4 (8.42)	68.6 (5.12)		60.6 (6.73)	62.7 (7.13)		63.0 (8.10)	64.5 (7.11)	
Missing	0 (0.0%)	0 (0.0%)		0 (0.0%)	0 (0.0%)		0 (0.0%)	0 (0.0%)	
PSA			0.6476			0.0942			0.0807
Median [IQR], ng/mL	18.9 [10.9, 38.4]	22.9 [12.3, 34.5]		7.3 [5.0, 10.7]	8.2 [5.3, 13.5]		8.6 [5.7, 17.0]	10.7 [6.1, 23.3]	
Missing	0 (0.0%)	0 (0.0%)		9 (2.3%)	3 (3.3%)		9 (1.6%)	3 (2.3%)	
Pathological T stage			0.1747			0.6284			0.2637
T2	86 (50.3%)	20 (51.3%)		154 (38.9%)	31 (34.4%)		240 (42.3%)	51 (39.5%)	
T3	79 (46.2%)	15 (38.5%)		229 (57.8%)	55 (61.1%)		308 (54.3%)	70 (54.3%)	
T4	6 (3.5%)	4 (10.3%)		7 (1.8%)	3 (3.3%)		13 (2.3%)	7 (5.4%)	
Tx	0 (0.0%)	0 (0.0%)		6 (1.5%)	1 (1.1%)		6 (1.1%)	1 (0.8%)	
Pathological N stage			0.6509			0.2148			0.2319
N0	127 (74.3%)	27 (69.2%)		275 (69.4%)	63 (70.0%)		402 (70.9%)	90 (69.8%)	
N1	19 (11.1%)	6 (15.4%)		58 (14.6%)	18 (20.0%)		77 (13.6%)	24 (18.6%)	
Nx	25 (14.6%)	6 (15.4%)		63 (15.9%)	9 (10.0%)		88 (15.5%)	15 (11.6%)	
Pathological M stage			0.0672			0.0824			0.3754
M0	105 (61.4%)	16 (41.0%)		183 (46.2%)	47 (52.2%)		288 (50.8%)	63 (48.8%)	
M1	20 (11.7%)	7 (17.9%)		0 (0.0%)	1 (1.1%)		20 (3.5%)	8 (6.2%)	
Mx	46 (26.9%)	16 (41.0%)		213 (53.8%)	42 (46.7%)		259 (45.7%)	58 (45.0%)	
Residual tumour			0.1976			1.0000			0.2874
YES	61 (35.7%)	18 (46.2%)		4 (1.0%)	1 (1.1%)		65 (11.5%)	19 (14.7%)	
NO	103 (60.2%)	19 (48.7%)		368 (92.9%)	83 (92.2%)		471 (83.1%)	102 (79.1%)	
Missing	7 (4.1%)	2 (5.1%)		24 (6.1%)	6 (6.7%)		31 (5.5%)	8 (6.2%)	
ISUP grade group			0.3395			0.0677			0.0186
1	15 (8.8%)	2 (5.1%)		40 (10.1%)	5 (5.6%)		55 (9.7%)	7 (5.4%)	
2	49 (28.7%)	7 (17.9%)		126 (31.8%)	18 (20.0%)		175 (30.9%)	25 (19.4%)	
3	34 (19.9%)	6 (15.4%)		74 (18.7%)	23 (25.6%)		108 (19.0%)	29 (22.5%)	
4	29 (17.0%)	7 (17.9%)		49 (12.4%)	12 (13.3%)		78 (13.8%)	19 (14.7%)	
5	44 (25.7%)	16 (41.0%)		107 (27.0%)	32 (35.6%)		151 (26.6%)	48 (37.2%)	
Missing	0 (0.0%)	1 (2.6%)		0 (0.0%)	0 (0.0%)		0 (0.0%)	1 (0.8%)	
3-year BCR-free survival (Kaplan-Meier)			0.7344			0.2984			0.3405
Estimate, % (95% CI)	61.9% (54.6-70.0)	62.7% (48.5-81.1)		83.7% (79.2-88.5)	88.3% (80.2-97.2)		76.0% (71.9-80.3)	79.5% (71.7-88.2)	
At risk at 36 months, n	82	20		132	25		214	45	
BCR events / evaluable, n/N	74/158	15/36		50/341	7/77		124/499	22/113	
Missing event/time	13 (7.6%)	3 (7.7%)		55 (13.9%)	13 (14.4%)		68 (12.0%)	16 (12.4%)	

Radiology Signature

Because eligible MRI data after quality control were available for only seven TCGA-PRAD cases, the 115 MRI cases (7 TCGA-PRAD and 108 CPGEA) were pooled and divided into a fixed 92-patient training set and a 23-patient held-out test set (Figure S1). Five-fold cross-validation for model selection and tuning was confined to the training set, and the untouched held-out set was used for final evaluation. We extracted 1,132 shape, first-order, and texture radiomic features from each MRI sequence using PyRadiomics (Table S2), yielding 2,264 features after concatenating the DWI and T2WI features. Nine features were selected using the least absolute shrinkage and selection operator (LASSO) procedure (Table S3). Correlations among the selected features were examined using Spearman analysis and hierarchical clustering. The

support vector machine (SVM) yielded the highest held-out AUC of 0.833 (Table S4). Some linear models yielded higher AUCs than nonlinear models, which may reflect the physical meaning and lower dimensionality of the selected handcrafted imaging features. Waterfall plots depict Radiology Signature scores by HRD status (Figure 2). A Radio-Clinical model was subsequently constructed using the same split as the Radiology model, enabling a common-patient comparison of the incremental clinical variables.

Pathology Signature

First, we divided the WSI into 512×512 patches, and only patches with an overlap of over 80% with the region of interest (ROI) were used for the subsequent analysis. Patch-level tumor-region prediction was performed based on the ResNet50 model. The ResNet50 model demonstrated high efficacy in distinguishing tumor regions from non-tumor regions, with an AUC of 0.936 (95% CI 0.935–0.936) in the test cohort (Figure 3 & Table S5). Subsequently, the patch labels and corresponding probabilities were applied to generate 106 pathological WSI-level features using the PLH and BoW pipelines (Table S6). Then WSI-level features were selected by LASSO regression. Finally, 42 features were applied for model construction by a series of algorithms (Table S7), including logistic regression (LR), support vector machine (SVM), random forest, extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM), to predict HRD status for each case. The LightGBM model (Pathology Signature) achieved an AUC of 0.815 in external validation. Using a fixed cutoff of 0.395, applied unchanged to both cohorts, the model identified 72 of 72 HRD cases (100%) in the training cohort and 15 of 18 HRD cases (83.3%) in the external test cohort. At the same cutoff, 243 of 313 HRP cases (77.6%) and 84 of 126 HRP cases (66.7%) were correctly classified as HRP in the training and test cohorts, respectively (Figure 3 & Table S8). Among the 11 patients who received neoadjuvant androgen deprivation therapy (ADT) in the test cohort (all actual HRP), four cases were predicted as HRD and seven cases were predicted as HRP. Treatment-associated pathological changes may affect the features used for HRD prediction.

Multimodal Signatures

In the fixed 85-patient Patho-Radiology cohort (68 training and 17 held-out patients), the Naïve Bayes model yielded a held-out AUC of 0.900; the held-out AUCs for logistic regression, KNN, and LightGBM were 0.867, 0.833, and 0.867, respectively (Figure 4 & Table S9). The Naïve Bayes model was selected as the Patho-Radiology Signature. In addition, we summarized the reported performance of the Patho-Clinical, Patho-Radiology, and Patho-Radio-Clinical models (Figure 4). Sensitivity and negative predictive value across training-score percentile cutoffs were also assessed for the Pathology, Radiology, and Patho-Radiology Signatures (Table S10). Because these three multimodal models were assessed under different validation schemes, their reported AUCs were not interpreted as head-to-head evidence. By contrast, a direct same-patient comparison of the Pathology, Radiology, and Patho-Radiology models was performed within the fixed 85-patient Patho-Radiology cohort. In the shared 17-patient held-out set, the AUCs were 0.867 for Pathology, 0.833 for Radiology, and 0.900 for Patho-Radiology. The corresponding paired DeLong tests did not show significant differences between Patho-Radiology and Pathology (P=0.795) or Radiology (P=0.671), indicating a favorable numerical trend rather than evidence of superiority (Figure S4 & Table S11). Across the evaluated algorithms, adding clinical variables was associated with lower held-out AUCs in most Pathology- and Radiology-based models (8/10 and 7/10, respectively), with a similar trend observed for Patho-Radiology models, suggesting limited incremental predictive value of the available clinical variables.

Signature score distributions differed across models (Figure 4C). The individual Clinical, Pathology, and Radiology scores were weakly correlated (Spearman r=0.03–0.19), whereas Patho-Radiology was modestly correlated with Pathology and Radiology (r=0.31 and 0.35; Figure 4D). Patient-level plots showed discordant Pathology and Radiology classifications and overlapping score patterns among correctly and incorrectly classified Patho-Radiology cases (Figures 4E-F).

Molecular characterization of patients with actual and Pathology-predicted HRD

The Prediction Analysis of Microarray 50-gene (PAM50) molecular subtypes, genetic alterations, and RNA expression of HRR-associated genes in patients with actual HRD and Pathology Signature-predicted HRD are summarized (Figure 5). For the PAM50 classification, the basal-like, Luminal A and luminal B subtypes were observed in 18.9%, 37.8% and 43.3% of cases with actual HRD and in 21.1%, 36.9%, and 42.0% of cases predicted as HRD by the Pathology Signature, respectively. There was no significant difference in the distribution of the different subtypes between the actual and predicted HRD patients (P=0.95). The ETS fusion was the most prevalent genetic alteration, followed by the FOXA1 mutation, SPOP mutation and RB1 deletion. In patients with actual HRD, the ETS gene fusion, RB1 deletion, HDAC2 deletion, FOXA1 mutation, and ROS1 deletion were the most common alterations (37.8%, 26.7%, 21.1%, 18.9%, and 18.9%, respectively), whereas these alterations were observed in 30.7%, 18.7%, 16.7%, 24.7%, and 14.0% of cases predicted as HRD, respectively. Both the actual and predicted HRD cases exhibited higher incidences of RB1 deletion (predicted HRD vs. predicted HRP, P<0.001; actual HRD vs. actual HRP, P<0.001), HDAC2 deletion (predicted HRD vs. predicted HRP, P=0.002; actual HRD vs. actual HRP, P<0.001), ROS1 deletion (predicted HRD vs. predicted HRP, P<0.001; actual HRD vs. actual HRP, P<0.001), and PMS2 mutations (predicted HRD vs. predicted HRP, P<0.001; actual HRD vs. actual HRP, P=0.003) when compared to their actual and predicted HRP counterparts. There was no significant difference in the frequency of genetic alterations between the predicted and actual patients with HRD (P=0.37). The predicted and actual HRD cases had similar RNA expression levels for all the genes. In conclusion, there was no significant difference in the molecular subtypes, genetic alterations, and RNA expression of HRR-associated genes in patients with actual and predicted HRD.

Interpretability of the Pathology Signature

To improve the interpretability of the Pathology Signature, the Grad-CAM method was applied to illustrate the pathological characteristics of the images at the patch level (Figure 6). After viewing all the Grad-CAM figures, two pathologists discussed and summarized four HRD-associated pathological features, including high-grade tumors (enlarged nuclear-to-cytoplasmic ratio, vacuolated appearance of cells, presence of pathological mitotic figures, poor formation of glandular follicles arranged in sheets, and small nests), fibrosis of the stroma (increased proliferation of fibroblasts with significant collagen degeneration observed around high-grade tumor cells), inflammatory cell infiltration (intensive infiltration of lymphoplasmacytic cells, usually observed around the tumor), and cribriform structures (presence of irregular glandular spaces resembling a sieve or honeycomb pattern), which the pathologists considered potentially associated with HRD. The Grad-CAM maps also highlighted regions assigned a higher likelihood in patches at different magnifications (e.g. ×12.5, ×20, ×50, ×100, and ×200; Figure S5). These observations were qualitative and were not evaluated against quantitative localization annotations.

To evaluate whether the features of the Pathology Signature could be understood and identified by pathologists, nine pathologists with different levels of clinical experience were included in the following analysis: specialized urologic pathologists (pathologists 1 to 3), attending general pathologists (pathologists 4 to 6), and junior general pathologists (pathologists 7 to 9). First, nine pathologists were trained by a senior pathologist to understand the HRD-associated pathological features using the selected Grad-CAM figures. Furthermore, the nine pathologists were asked to classify 20 WSIs (10 WSIs with the highest Pathology Signature scores and 10 WSIs with the lowest Pathology Signature scores) into the HRD or HRP group independently. After four weeks, the pathologists were informed about the predicted Pathology Signature score of each WSI and were asked to reclassify these WSIs. Accordingly, each experience group contributed 60 individual readings under each reading condition (20 WSIs evaluated by three pathologists).

While there was no significant difference in the AUC between specialized urologic and attending general pathologists (p = 0.59), a statistically significant difference was observed when compared with junior general pathologists (p = 0.01). Specialized urologic pathologists selected "very certain" more frequently than junior general pathologists (26/60 vs. 14/60 readings, p = 0.02). After being informed about the Pathology Signature score, most pathologists showed higher AUC and accuracy. The perceived certainty of the

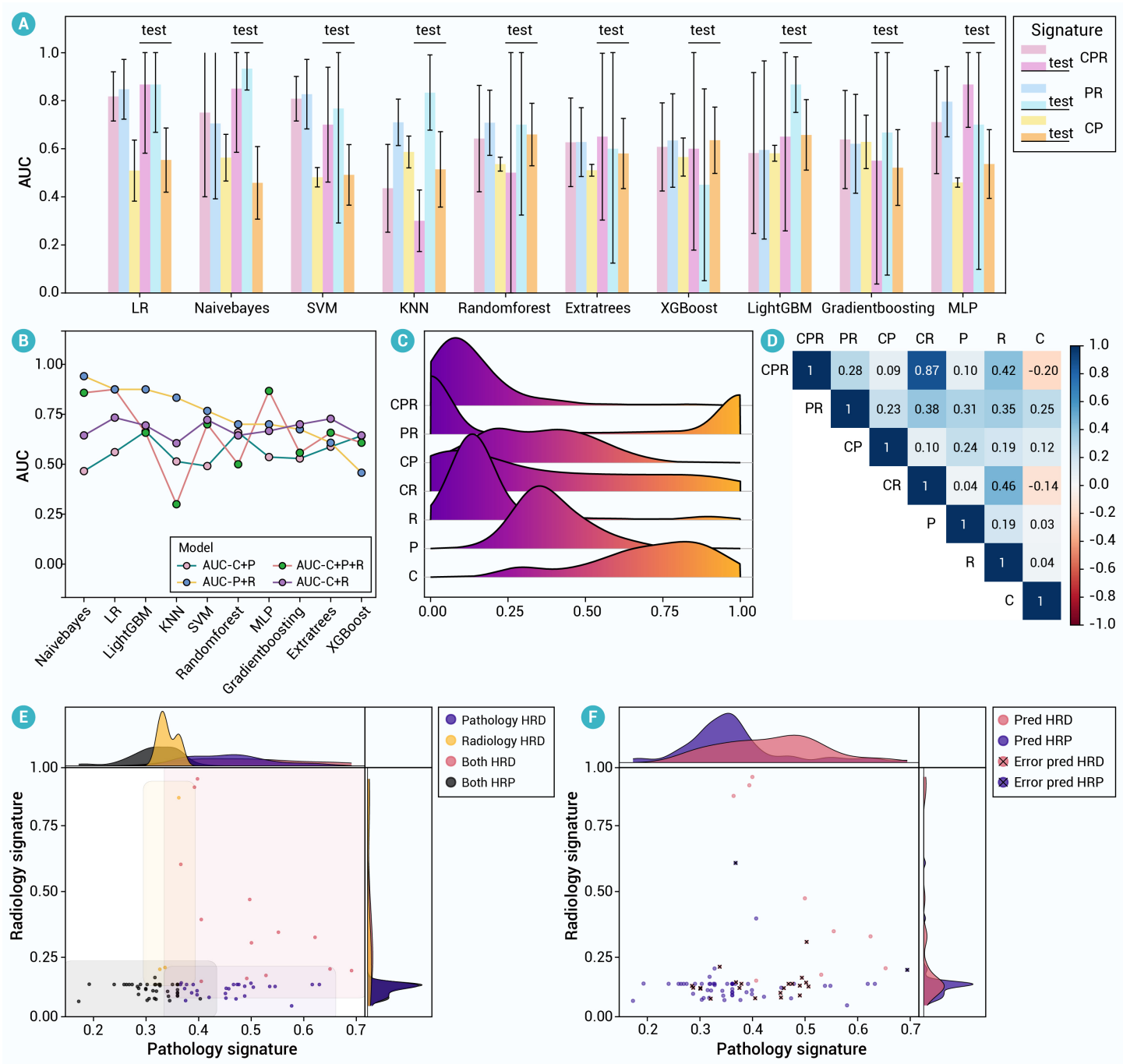


Figure 4. Predictive performance of multimodal models and associations among individual modality signatures (A) Training- and held-out test-set AUCs of the Patho-Clinical, Patho-Radiology, and Patho-Radio-Clinical models across algorithms; error bars represent DeLong 95% confidence intervals calculated from patient-level predictions in each dataset. (B) Test-set AUCs across ten algorithms for the Patho-Clinical, Radio-Clinical, Patho-Radiology, and Patho-Radio-Clinical models. (C) Pooled training-plus-test predicted-score distributions for seven representative signatures: Clinical (C), Pathology (P), Radiology (R), Patho-Clinical (CP), Radio-Clinical (CR), Patho-Radiology (PR), and Patho-Radio-Clinical (CPR). (D) Upper-triangular Spearman correlation matrix for the seven signature scores among the 85 patients with complete scores. (E) Pathology and Radiology Signature scores of individual patients, classified according to predictions from either, both, or neither modality. (F) Pathology and Radiology Signature scores showing correctly and incorrectly classified HRD and HRP cases by the Patho-Radiology model. AUC=area under the curve. HRD=homologous recombination deficiency. HRP=homologous recombination proficient.

pathologists' classifications (defined as "very certain", "fairly certain", "relatively uncertain") also improved after the pathologists were informed about the Pathology Signature score across the three groups of pathologists. However, based on repeated measures analysis of variance (ANOVA), only junior general pathologists showed a statistically significant improvement (p=0.04). In contrast, specialized urologic and attending general pathologists did not demonstrate significant changes (p=0.33 and p=0.93, respectively). Although preliminary, these results suggested that the pathological features could be understood by pathologists and helped address the interpretability of the Pathology Signature. Because the 20 cases were selected from the score extremes and readers were evaluated before and after score disclosure, this should be interpreted as an exploratory reader study rather than an independent clinical validation. This deliberate score enrichment was intended to increase morphological contrast between the model-defined groups, while reuse of the identical case set before and after score disclosure provided a within-reader paired comparison; neither design feature was intended to estimate performance in a routine case mix.

Association between HRD and biochemical recurrence (BCR)

Among 612 patients with complete BCR follow-up data, the 3-year Kaplan–Meier BCR-free survival estimate was 79.5% (95% CI 71.7–88.2) in

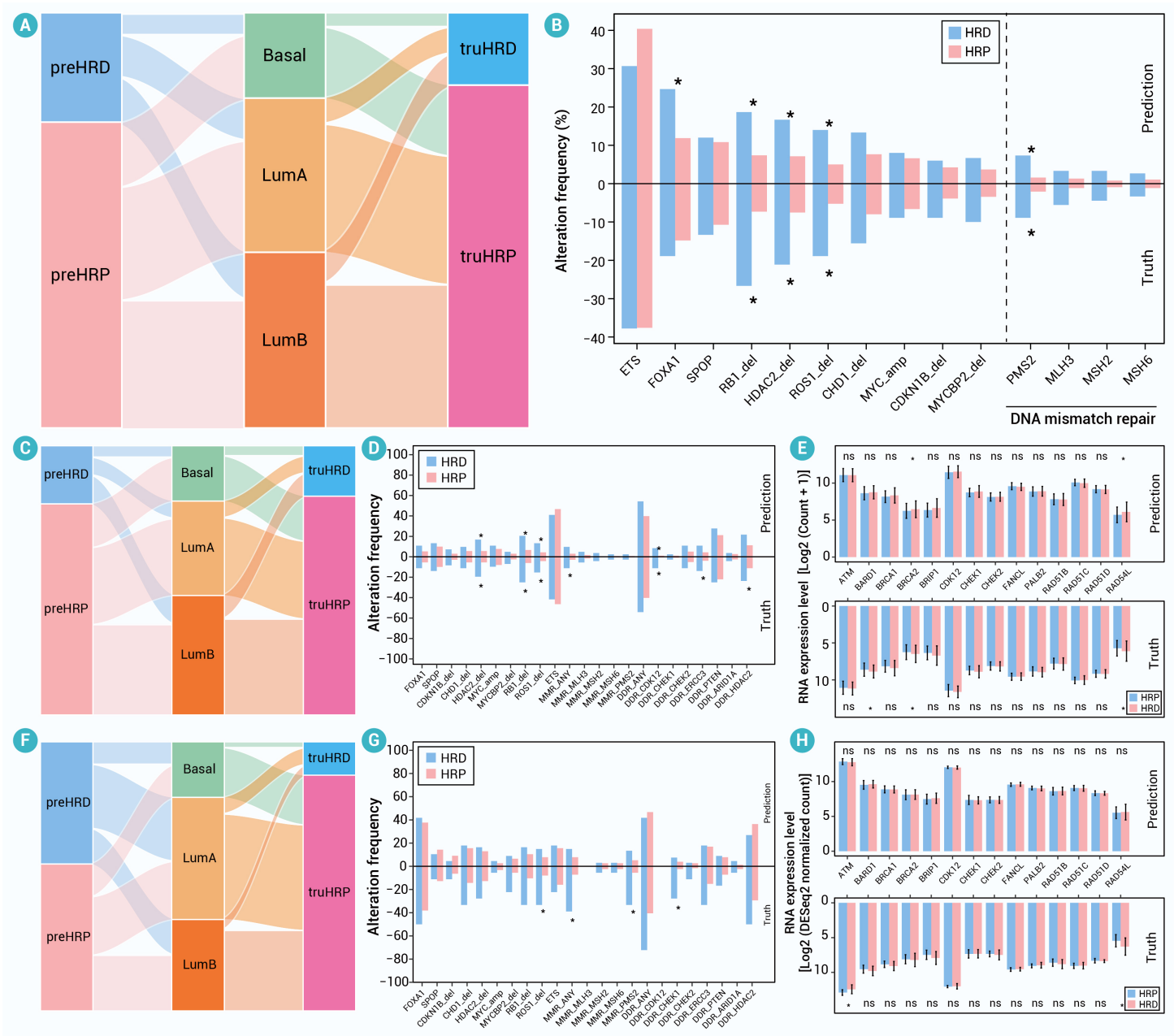


Figure 5. Molecular Characterization of patients with actual and Pathology Signature-predicted HRD Status. Molecular subtypes for both actual and predicted HRD and HRP cases are exhibited in the overall cohort (A), CPGEA cohort (C), and TCGA-PRAD cohort (F). Genetic alterations for these cases are depicted in the overall cohort (B), CPGEA (D), and TCGA-PRAD (G). RNA expression for both actual and predicted HRD and HRP cases is shown in CPGEA cohort (E) and TCGA-PRAD cohort (H). HRD=homologous recombination deficiency. HRP=homologous recombination proficient. CPGEA=Chinese Prostate Cancer Genetic and Epigenetic Atlas. TCGA-PRAD=The Cancer Genome Atlas Prostate Adenocarcinoma. RNA=ribonucleic acid.

patients with actual HRD and 76.0% (95% CI 71.9–80.3) in patients with HRP. No significant difference was observed between the groups (two-sided log-rank P=0.341). The HRD status predicted by the Clinical Signature was reported to stratify patients' risk of BCR in the training and test cohorts. However, the HRD status predicted by the Pathology Signature and Radiology Signature did not stratify the risk of BCR (Figure S6). The observed BCR event proportion was 146/612 (23.9%) overall, including 22/113 (19.5%) in patients with actual HRD and 124/499 (24.8%) in patients with HRP. The signature-specific analyzable sample/event counts in the training and test cohorts were 418/57 and 194/89 for Clinical, 325/40 and 136/65 for Pathology, and 84/37 and 23/9 for Radiology, respectively. The especially small Radiology test set (23 patients and nine events) produced imprecise estimates. Accordingly, the non-significant Pathology and Radiology results do not establish absence of prognostic association, and the Clinical finding should be considered hypothesis-generating.

DISCUSSION

PARP inhibitors have been approved by the United States Food and Drug Administration (FDA) for the treatment of patients with advanced PCa and HRD. Thus, predicting HRD is crucial for the treatment of patients with advanced PCa; however, current prediction methods based on genetic testing are usually time-consuming, commonly inaccessible, and very expensive.[22] For instance, the FDA approved three assays for the detection of HRD in PCa (BRACAnalysis CDx testing blood,[23] FoundationOne CDx testing tissue,[24] and FoundationOne Liquid CDx testing plasma).[25] but these tests are commonly inaccessible for patients in underdeveloped regions. Additionally, the cost of these assays is as high as $4,800 to $5,800 per test. Patients with HRD are relatively infrequently observed among PCa patients and require genetic tests of approximately ten patients to identify one patient with HRD, which means it costs $48,000 to $58,000 to identify one patient with HRD.[8] These logistic and financial challenges discouraged patients from undergoing genetic testing, resulting in missed opportunities for the accurate application

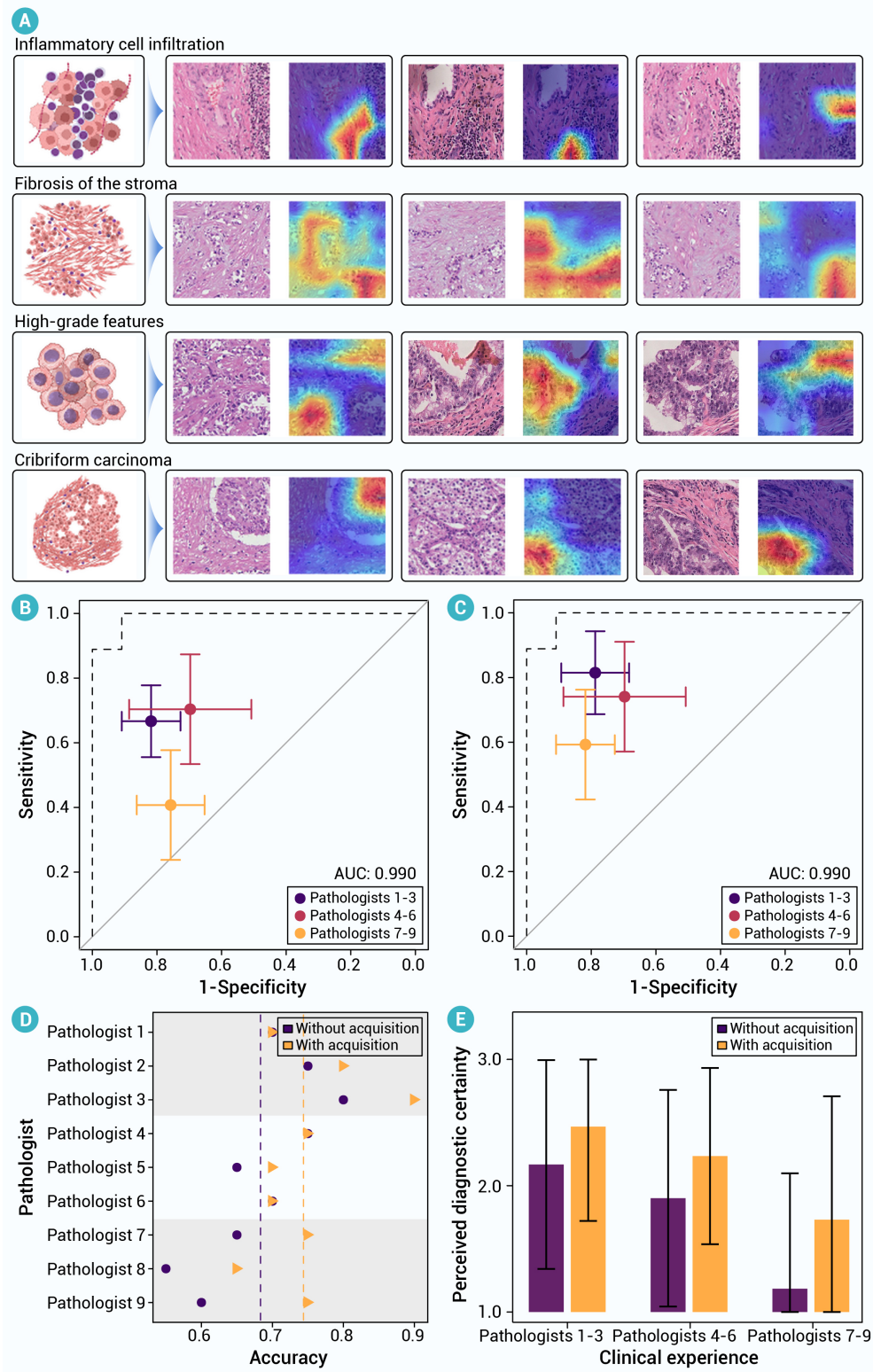


Figure 6. Visualization and interpretability of the Pathology Signature (A) Four pathological features including high-grade tumors, fibrosis of the stroma, inflammatory cell infiltration, and cribriform structures were identified by pathologists to be associated with HRD status in Gradient-weighted Class Activation Mapping (Grad-CAM). (B) Performance of the Pathology Signature and pathologist groups before disclosure of the signature score. Points show mean specificity and sensitivity; horizontal and vertical error bars indicate SD across three pathologists. (C) Corresponding performance after score disclosure. (D) Individual pathologist accuracy before and after score disclosure. Dashed lines indicate the mean accuracy across all nine pathologists. (E) Mean diagnostic certainty, scored from 1 to 3. Error bars indicate SD across 60 ratings per group and condition (20 cases × 3 pathologists). HRD=homologous recombination deficiency. AUC=area under the curve.

of PARP inhibitors in PCa patients with HRD.

AI has increasingly been integrated into cancer management. The U.S. Food and Drug Administration (FDA) authorized Paige Prostate to assist pathologists in detecting cancer in prostate needle biopsy specimens.[26] In this study, we developed an AI-based HRD status prediction tool for patients with PCa using multimodal data including pathological, radiological, and clinical features. The Pathology Signature, Radiology Signature, and Patho-Radiology Signature showed predictive performance for the HRD status in PCa patients. Our findings suggest that patients with a higher probability of HRD may be prescreened by an AI-based algorithm using WSIs and MRI images. This workflow can use routinely acquired data when the relevant modalities are already available. Pathological slides and MRI images were routinely collected. This enables retrospective model development. Once implemented, inference may be faster than confirmatory genetic testing, but turnaround time and cost-effectiveness were not evaluated in this study. In clinical settings, this prediction tool could be evaluated as a prescreening method to prioritize cases with a higher predicted probability of HRD for confirmatory testing.

In single-modal analyses, the Pathology and Radiology Signatures showed predictive signal, whereas the Clinical Signature showed limited discrimination (AUC=0.560), consistent with previous reports linking HRD more closely to molecular and image-derived features than to routine clinical variables.[16,17,27–32] In the fixed Patho-Radiology validation set, the combined model showed a numerically higher AUC (0.900, compared with 0.867 for Pathology and 0.833 for Radiology), but the paired DeLong comparisons were not significant. Thus, the present data suggest a favorable multimodal trend rather than establish superiority. Naïve Bayes may have performed well because its conditional-independence structure is relatively parsimonious for a small complete-case dataset; this explanation should be tested in larger paired cohorts. The Patho-Clinical and Radio-Clinical analyses likewise showed no clear improvement over Pathology and Radiology alone, respectively, suggesting limited incremental information from the available clinical variables. Training AUCs of 1.000 for several flexible models, coupled with substantially lower held-out AUCs, indicate potential overfitting or high variance; interpretation therefore prioritized held-out estimates and their confidence intervals.

In analyzing the pathological data, previous studies suggested that a weakly supervised approach using only slide-level tumor/non-tumor labels can yield satisfactory results.[33] Initially, we attempted to apply weakly supervised methods; however, we found that the use of non-annotated slides failed to yield satisfactory results. We attributed the differences between our study and previous studies using weakly supervised methods to the nature of PCa, in which tumor heterogeneity played a significant role. Unlike tumors with distinct boundaries between the tumor and normal tissues, cancer, normal, and prostate stromal tissues are intertwined in PCa. The tumor proportion in one slide varied from 5% to 95% among different slides.[34] Without defining the

tumor region, it was difficult for the model to capture the features of the tumor. Therefore, in this study, we employed annotation of the tumor region, defined as the area with the major proportion of tissue consisting of tumor cells and intratumoral stromal tissue. We did not strictly annotate every region of the adenocarcinoma at the cellular level because we believe that the stromal structures within the tumor might be relevant to molecular changes. In particular, the pathological grade of PCa is closely related to tissue structure rather than the characteristics of individual cancer cells, which is different from many other malignancies.[35] Annotating only the cancerous region, but not the stromal region of PCa, would result in the loss of important structural information. In contrast, for cancer types with pathology slides predominantly consisting of tumor tissue, weakly supervised methods can improve the prediction.[36] Our findings demonstrated that the supervised approach may be more applicable for predicting molecular alterations in prostate cancer.

To enhance interpretability, a Grad-CAM plot was used to analyze the pathology images. Pathologists have identified features including high-grade tumors, fibrosis of the stroma, inflammatory cell infiltration, and cribriform structures. This is in accordance with a preprint paper by Jakob et al., which identified that high grade, fibrosis, and lymphocytic infiltration were associated with HRD in different types of tumors.[37] Hemorrhage was associated with HRD in that study, but this association was not confirmed in this study because there were very few hemorrhages in PCa. In addition, cribriform structures were associated with HRD in this study but not in a previous study. We speculate that this is because cribriform structures are more often observed in high-grade PCa than in other malignancies. To interpret the Patho-Radiology Signature based on the Naïve Bayes algorithm, a probability density function plot was performed. The mean and variance of the Gaussian distribution for each feature were illustrated (Figures S7-S8). Beyond model interpretability, calibration and threshold-dependent net benefit should also be considered when assessing the potential of these models for HRD prescreening (Figure S9).

The main limitations were the small number of modality-complete cases and the lack of an independent MRI-rich external cohort. Accordingly, the radiology and multimodal estimates should be interpreted as exploratory, and prospective validation in a larger paired cohort is needed. The retrospective reader study and qualitative Grad-CAM assessment were also preliminary. The score-extreme design can increase morphological contrast for a proof-of-concept interpretability assessment, but it enriches the case spectrum. Together with only 20 WSIs and three readers per experience level, this limits the precision and generalizability of comparisons among specialization groups; a larger reader study using a representative case mix and prespecified endpoints is needed. The apparent BCR stratification by the Clinical Signature, despite its limited discrimination for HRD (AUC=0.560), may reflect conventional clinicopathologic prognostic information rather than HRD-specific biology. Therefore, the negative Pathology and Radiology findings are inconclusive, and none of the signatures should be considered validated prognostic tools without a prospectively powered external cohort. To address these limitations, we are currently collecting a larger prospective cohort with more modality-complete cases, in which future interpretability analyses will incorporate prespecified histological definitions, blinded multi-pathologist annotations, interobserver assessment and adjudication, and quantitative spatial validation of Grad-CAM-highlighted regions.

CONCLUSION

We illustrated that the Pathology Signature, Radiology Signature, and Patho-Radiology Signature showed potential in predicting HRD status in PCa patients. The framework of this study may be adapted to other malignancies and highlights the promising applications of AI for HRD prescreening in cancer management. In clinical settings, AI-based prediction tools may serve as prescreening tools to reduce the number of genetic analyses required. Overall, these findings support the feasibility of multimodal HRD prescreening and suggest a favorable trend that requires prospective validation in a larger paired cohort.

REFERENCES

1. Siegel R.L., Miller K.D., Wagle N.S., et al. (2023). Cancer statistics, 2023. *CA Cancer J. Clin.* **73**:17–48. DOI:10.3322/caac.21763
2. Mitchell T. and Neal D.E. (2015). The genomic evolution of human prostate cancer. *Br. J. Cancer* **113**:193–198. DOI:10.1038/bjc.2015.234
3. Nguyen L., Martens J.W.M., Van Hoeck A., et al. (2020). Pan-cancer landscape of homologous recombination deficiency. *Nat. Commun.* **11**:5584. DOI:10.1038/s41467-020-19406-4
4. Hoeijmakers J.H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* **411**:366–374. DOI:10.1038/35077232
5. Groelly F.J., Fawkes M., Dagg R.A., et al. (2023). Targeting DNA damage response pathways in cancer. *Nat. Rev. Cancer* **23**:78–94. DOI:10.1038/s41568-022-00535-5
6. ESMO. (2019). ESMO 2019: PROfound: Phase 3 study of olaparib vs. enzalutamide or abiraterone for metastatic castration-resistant prostate cancer with homologous recombination repair gene alterations. *UroToday.* https://www.urotoday.com/conference-highlights/esmo-2019/esmo-2019-prostate-cancer/115401-esmo-2019-profound-phase-3-study-of-olaparib-vs-enzalutamide-or-abiraterone-for-metastatic-castration-resistant-prostate-cancer-with-homologous-recombination-repair-gene-alterations.html (Accessed June 3, 2023)
7. de Luca X.M., Newell F., Kazakoff S.H., et al. (2020). Using whole-genome sequencing data to derive the homologous recombination deficiency scores. *npj Breast Cancer* **6**:33. DOI:10.1038/s41523-020-0172-0
8. Sztupinszki Z., Diossy M., Krzystanek M., et al. (2020). Detection of molecular signatures of homologous recombination deficiency in prostate cancer with or without BRCA1/2 mutations. *Clin. Cancer Res.* **26**:2673–2680. DOI:10.1158/1078-0432.CCR-19-2135
9. Coudray N. and Tsirigos A. (2020). Deep learning links histology, molecular signatures and prognosis in cancer. *Nat. Cancer* **1**:755–757. DOI:10.1038/s43018-020-0099-2
10. Kather J.N., Pearson A.T., Halama N., et al. (2019). Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **25**:1054–1056. DOI:10.1038/s41591-019-0462-y
11. Yamashita R., Long J., Longacre T., et al. (2021). Deep learning model for the prediction of microsatellite instability in colorectal cancer: A diagnostic study. *Lancet Oncol.* **22**:132–141. DOI:10.1016/S1470-2045(20)30535-0
12. Cifci D., Foersch S. and Kather J.N. (2022). Artificial intelligence to identify genetic alterations in conventional histopathology. *J. Pathol.* **257**:430–444. DOI:10.1002/path.5898
13. Dadhania V., Gonzalez D., Yousif M., et al. (2022). Leveraging artificial intelligence to predict ERG gene fusion status in prostate cancer. *BMC Cancer* **22**:494. DOI:10.1186/s12885-022-09559-4
14. Schaumberg A.J., Rubin M.A. and Fuchs T.J. (2016). H&E-stained whole slide image deep learning predicts SPOP mutation state in prostate cancer. *bioRxiv.* DOI:10.1101/064279.
15. Houlahan K.E., Salmasi A., Sadun T.Y., et al. (2019). Molecular hallmarks of multiparametric magnetic resonance imaging visibility in prostate cancer. *Eur. Urol.* **76**:18–23. DOI:10.1016/j.eururo.2018.12.036
16. Boehm K.M., Aherne E.A., Ellenson L., et al. (2022). Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. *Nat. Cancer* **3**:723–733. DOI:10.1038/s43018-022-00388-9
17. Vanguri R.S., Luo J., Aukerman A.T., et al. (2022). Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat. Cancer* **3**:1151–1164. DOI:10.1038/s43018-022-00416-8
18. Fremond S., Andani S., Barkey Wolf J., et al. (2023). Interpretable deep learning model to predict the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images: A combined analysis of the PORTEC randomised trials and clinical cohorts. *Lancet Digit. Health* **5**:e71–e82. DOI:10.1016/S2589-7500(22)00210-2
19. Li J., Xu C., Lee H.J., et al. (2020). A genomic and epigenomic atlas of prostate cancer in Asian populations. *Nature* **580**:93–99. DOI:10.1038/s41586-020-2135-x
20. Cancer Genome Atlas Research Network. (2015). The molecular taxonomy of primary prostate cancer. *Cell* **163**:1011–1025. DOI:10.1016/j.cell.2015.10.025
21. Clarke N., Wiechno P., Alekseev B., et al. (2018). Olaparib combined with abiraterone in patients with metastatic castration-resistant prostate cancer: A randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Oncol.* **19**:975–986. DOI:10.1016/S1470-2045(18)30365-6
22. Wu K., Liang J., Shao Y., et al. (2021). Evaluation of the efficacy of PARP inhibitors in metastatic castration-resistant prostate cancer: A systematic review and meta-analysis. *Front. Pharmacol.* **12**:777663. DOI:10.3389/fphar.2021.777663
23. De Bono J.S., Hussain M., Thiery-Vuillemin A., et al. (2017). PROfound: A randomized Phase III trial evaluating olaparib in patients with metastatic castration-resistant prostate cancer and a deleterious homologous recombination DNA repair aberration. *J. Clin. Oncol.* **35**:TPS5091–TPS5091. DOI:10.1200/JCO.2017.35.15_suppl.TPS5091
24. Milbury C.A., Creeden J., Yip W.-K., et al. (2022). Clinical and analytical validation of FoundationOne®CDx, a comprehensive genomic profiling assay for solid tumors.

PLoS One **17**:e0264138. DOI:10.1371/journal.pone.0264138
25. Woodhouse R., Li M., Hughes J., et al. (2020). Clinical and analytical validation of FoundationOne Liquid CDx, a novel 324-gene cfDNA-based comprehensive genomic profiling assay for cancers of solid tumor origin. *PLoS One* **15**:e0237802. DOI:10.1371/journal.pone.0237802
26. U.S. Food and Drug Administration. (2021). Device classification under Section 513(f)(2) (De Novo): Paige Prostate (DEN200080). *FDA*. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/denovo.cfm?id=DEN200080 (Accessed October 15, 2026)
27. Feng C., Zhang Y., Wu F., et al. (2023). Relationship between homologous recombination deficiency and clinical features of breast cancer based on genomic scar score. *Breast* **69**:392–400. DOI:10.1016/j.breast.2023.04.002
28. Fu Y., Jung A.W., Torne R.V., et al. (2020). Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer* **1**:800–810. DOI:10.1038/s43018-020-0085-8
29. Kim M., Jung S.Y., Park J.E., et al. (2020). Diffusion- and perfusion-weighted MRI radiomics model may predict isocitrate dehydrogenase (IDH) mutation and tumor aggressiveness in diffuse lower grade glioma. *Eur. Radiol.* **30**:2142–2151. DOI:10.1007/s00330-019-06548-3
30. Zhang J., Zhao X., Zhao Y., et al. (2020). Value of pre-therapy 18F-FDG PET/CT radiomics in predicting EGFR mutation status in patients with non-small cell lung cancer. *Eur. J. Nucl. Med. Mol. Imaging* **47**:1137–1146. DOI:10.1007/s00259-019-04592-1
31. Sammut S.-J., Crispin-Ortuzar M., Chin S.-F., et al. (2022). Multi-omic machine learning predictor of breast cancer therapy response. *Nature* **601**:623–629. DOI:10.1038/s41586-021-04278-5
32. Lipkova J., Chen R.J., Chen B., et al. (2022). Artificial intelligence for multimodal data integration in oncology. *Cancer Cell* **40**:1095–1110. DOI:10.1016/j.ccell.2022.09.012
33. Muti H.S., Heij L.R., Keller G., et al. (2021). Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: A retrospective multicentre cohort study. *Lancet Digit. Health* **3**:e654–e664. DOI:10.1016/S2589-7500(21)00133-3
34. Mohanty S.K., Lobo A. and Cheng L. (2023). The 2022 revision of the World Health Organization classification of tumors of the urinary system and male genital organs: Advances and challenges. *Hum. Pathol.* **136**:123–143. DOI:10.1016/j.humpath.2022.08.006
35. van Leenders G.J.L.H., van der Kwast T.H., Grignon D.J., et al. (2020). The 2019 International Society of Urological Pathology (ISUP) consensus conference on grading of prostatic carcinoma. *Am. J. Surg. Pathol.* **44**:e87–e99. DOI:10.1097/PAS.0000000000001497
36. Bilal M., Raza S.E.A., Azam A., et al. (2021). Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: A retrospective study. *Lancet Digit. Health* **3**:e763–e772. DOI:10.1016/S2589-7500(21)00180-1
37. Loeffler C.M.L., El Nahhas O.S.M., Muti H.S., et al. (2023). Direct prediction of homologous recombination deficiency from routine histology in ten different tumor types with attention-based multiple instance learning: A development and validation study. *medRxiv* 2023.03.08.23286975. DOI:10.1101/2023.03.08.23286975.

FUNDING AND ACKNOWLEDGMENTS

This study was supported by grants from the National Natural Science Foundation of China (NSFC) (grant number: 82272905), Rising Star Program of Shanghai Science and Technology Commission (grant number: 21QA1411500), and the Shanghai Action Plan for Technological Innovation Grant (No. 22ZR1478000) and Guhai Project of CH Hospital. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization: R.C., G.X., and L.W. Methodology: Z.S., Q.Z., N.T., Y.Z., W.H.Z., and L.D. Investigation: L.D., Z.S., Q.Z., N.T., Y.Z., W.H.Z., Y.W., M.Q., Z.D., H.Y., Y.M.L., X.W., Y.Y., H.W., L.Z., Y.G., W.Z., and Y.C.L. Software: Z.S., Q.Z., N.T., Y.Z., W.H.Z., L.D., Y.W., M.Q., Z.D., and H.Y. Visualization: L.D., W.H.Z., Y.W., M.Q., Z.D., H.Y., H.W., L.Z., Y.G., W.Z., and Y.C.L. Supervision: R.C., G.X., and L.W. Writing—original draft: Z.S., L.D., Q.Z., N.T., Y.Z., and W.H.Z. Writing—review and editing: Y.W., L.D., M.Q., Z.D., H.Y., H.W., L.Z., Y.G., W.Z., Y.C.L., R.C., G.X., and L.W. Resources: Z.S., Q.Z., N.T., Y.Z., Y.M.L., X.W., and Y.Y. Validation: R.C., G.X., and L.W. Project administration: R.C., G.X., and L.W. Funding acquisition: R.C., G.X., and L.W. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ETHICAL STATEMENT AND PATIENT CONSENT

This study complied with relevant ethical regulations. The study protocols were approved by the Institutional Review Board of Shanghai Changhai Hospital (approval number: CHEC2022-151). This study was registered with the Chinese Clinical Trial Registry (ChiCTR2200064329). Given the retrospective nature of this study, the requirement for individual informed consent was waived with explicit approval granted by the Institutional Review Board of Shanghai Changhai Hospital. Participants were not compensated.

DATA AND CODE AVAILABILITY

This study used the TCGA-PRAD and CPGEA datasets. Complete intermediate files generated during data processing and analysis are available from the corresponding authors upon request. The source code used for data preprocessing, model development, and evaluation will be made publicly available through the MMDLPC GitHub repository (https://github.com/Gene1343/MMDLPC) upon acceptance. The repository will also provide documentation and the shareable supporting files.

SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.59717/j.xinn-oncol.2026.100043