EBioMedicine 68 (2021) 103395

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

Integration of clinicopathologic identification and deep transferrable image feature representation improves predictions of lymph node metastasis in prostate cancer

Ying Hou[a,1], Jie Bao[c,1], Yang Song[d], Mei-Ling Bao[b], Ke-Wen Jiang[a], Jing Zhang[a], Guang Yang[d], Chun-Hong Hu[c], Hai-Bin Shi[a], Xi-Ming Wang[c,*], Yu-Dong Zhang[a,*]

[a] Department of Radiology, The First Affiliated Hospital of Nanjing Medical University; Nanjing, Jiangsu Province, PR China
[b] Department of Pathology, The First Affiliated Hospital of Nanjing Medical University; Nanjing, Jiangsu Province, PR China
[c] Department of Radiology, The First Affiliated Hospital of Soochow University, Suzhou 215006, PR China
[d] Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai, PR China

ARTICLE INFO

Article History:
Received 14 December 2020
Revised 28 April 2021
Accepted 28 April 2021
Available online xxx

Keywords:
Pelvic lymph node metastasis
Prostate cancer
Radiomics
Deep learning

ABSTRACT

Background: Accurate identification of pelvic lymph node metastasis (PLNM) in patients with prostate cancer (PCa) is crucial for determining appropriate treatment options. Here, we built a PLNM-Risk calculator to obtain a precisely informed decision about whether to perform extended pelvic lymph node dissection (ePLND).

Methods: The PLNM-Risk calculator was developed in 280 patients and verified internally in 71 patients and externally in 50 patients by integrating a set of radiologists' interpretations, clinicopathological factors and newly refined imaging indicators from MR images with radiomics machine learning and deep transfer learning algorithms. Its clinical applicability was compared with Briganti and Memorial Sloan Kettering Cancer Center (MSKCC) nomograms.

Findings: The PLNM-Risk achieved good diagnostic discrimination with areas under the receiver operating characteristic curve (AUCs) of 0.93 (95% CI, 0.90-0.96), 0.92 (95% CI, 0.84-0.97) and 0.76 (95% CI, 0.62-0.87) in the training/validation, internal test and external test cohorts, respectively. If the number of ePLNDs missed was controlled at $< 2\%$, PLNM-Risk provided both a higher number of ePLNDs spared (PLNM-Risk 59.6% vs MSKCC 44.9% vs Briganti 38.9%) and a lower number of false positives (PLNM-Risk 59.3% vs MSKCC 70.1% and Briganti 72.7%). In follow-up, patients stratified by the PLNM-Risk calculator showed significantly different biochemical recurrence rates after surgery.

Interpretation: The PLNM-Risk calculator offers a noninvasive clinical biomarker to predict PLNM for patients with PCa. It shows improved accuracy of diagnosis support and reduced overtreatment burdens for patients with findings suggestive of PCa.

Funding: This work was supported by the Key Research and Development Program of Jiangsu Province (BE2017756) and the Suzhou Science and Technology Bureau-Science and Technology Demonstration Project (SS201808).

1. Introduction

Prostate cancer (PCa) is one of the most common malignancies and the second leading cause of cancer-related mortality in Western men [1]. Pelvic lymph node metastasis (PLNM), accounting for ~15% of all newly diagnosed PCa patients, is an important prognostic factor that connects to biochemical recurrence (BCR) and distant metastases after curative treatment [2,3]. Therefore, accurate pre-treatment identification of PLNM of localized PCa would have a significant impact on clinical decision making, treatment planning and prediction of outcomes for patients [4].

Pelvic lymph node dissection (PLND), with external and obturator iliac lymph nodes dissected, or extended PLND (ePLND), with external, obturator, internal iliac and presacral lymph nodes dissected, is generally recommended for high-risk PCa patients who are

* Corresponding authors.
E-mail addresses: njmu_hy@163.com (Y. Hou), baojie7346@sina.com (J. Bao), ysong@phy.ecnu.edu.cn (Y. Song), njmu_bml@163.com (M.-L. Bao), cathyjiang_21@126.com (K.-W. Jiang), njmu_zj@163.com (J. Zhang), gyang@phy.ecnu.edu.cn (G. Yang), hch5305@163.com (C.-H. Hu), hbshi346@163.com (H.-B. Shi), wangximing1998@163.com (X.-M. Wang), njmu_zyd@163.com (Y.-D. Zhang).
[1] Ying Hou and Jie Bao contribute equally to this work.

https://doi.org/10.1016/j.ebiom.2021.103395

Research in Context

Evidence before this study

We searched the online database PubMed for the query "((lymph node invasion [Title/Abstract] OR lymph node metastasis [Title/Abstract]) AND (prostate cancer [Title/Abstract]))". We did not use a date restriction, but we limited the search to English. On Dec. 30, 2019, this search generated 956 results. Most focused on clinicopathological parameters for the prediction of lymph node metastasis, and there is still no study integrating radiomics and/or deep learning into clinicopathological parameters. Our study included a large dataset and integrated a set of radiologists' interpretations, clinicopathological factors and newly refined imaging indicators from MR images with radiomics machine learning and deep learning algorithms. Radical prostatectomy specimens were used as reference standards.

Added value of this study

We proposed and validated a pelvic lymph node metastasis (PLNM) risk calculator to aid in extended pelvic lymph node dissection (ePLND) decisions in prostate cancer patients. The results showed that our PLNM-Risk calculator can help to spare more ePLNDs while at a lower cost of missing lymph node metastasis than the established Briganti and MSKCC nomograms.

Implications of all the available evidence

Imaging biomarkers using computational radiomics and deep learning approaches play an incremental role in the preoperative assessment of pelvic lymph node metastasis risk in patients with prostate cancer. Our pelvic lymph node metastasis risk calculator provides improved accuracy of diagnosis support for the management of patients with findings suggestive of prostate cancer and may assist clinicians in patient management.

undergoing radical prostatectomy [4]. However, there is no ideal preoperative tool available to select candidates for PLND or ePLND due to the variable sensitivity of 27%-100% and application limitations. For example, the use of superparamagnetic iron oxide or sentinel lymph node techniques remains in the research field, with no clinical routine in the field of PCa application [5]. In the last few decades, risk assessment tools, including the Briganti score, Memorial Sloan Kettering Cancer Center (MSKCC) nomogram and Partin tables, were proposed for the assessment of PLNM risk [6–8], while showing moderate predictive accuracy on internal and external validations [9–12]. Multiparametric magnetic resonance imaging (mpMRI) has been widely used to preoperatively characterize PCa and determine the tumour and nodal stage [13]. Lymph nodes over 8 mm in the short-axis dimension on T_2-weighted imaging (T_2WI) and high signal intensity on diffusion weighted imaging (DWI) are recognized as suspicious for malignancy. However, the performance of direct MRI-based nodal staging of PCa is relatively poor, with a sensitivity of 40-60% [14]. Moreover, imaging interpretation of nodal stage requires the expertise of experienced radiologists; accordingly, inconsistencies exist between readers with varying levels of experience, thus leaving room for improvement. Recently, qualitative and quantitative measurements on high-resolution T_2-weighted imaging (T_2WI), diffusion-weighted imaging (DWI) and dynamic contrast-enhanced MRI have been regarded as promising predictors of PLNM [15,16]. A clinically available tool should leverage the integration of all data representations to enable detailed risk assessment in PCa, but this is lacking. Several studies have attempted to develop predictive models for PLNM by incorporating clinical/demographic and imaging indicators [17–19]. Others have carried out the application of machine learning or deep learning approaches, yielding higher accuracy than the Briganti and MSKCC nomograms [20]. Translating multimodal high-dimensional observations into a clinically interpretable signature is a potential alternative for improved prognostic and predictive accuracy of PCa. However, there are challenges associated with such analyses. First, there are always multimodal data generated in different clinical workflows and procedures, and effective approaches for integrating these data information are lacking. Second, although high-throughput deep learning networks have matured to a point that enables detailed discoveries of diseases in task-specific programs, the limited cohort size and high dimensionality of the data increase the possibility of false-positive discoveries and overfitting.

Therefore, the purpose of this study is to design and validate a new risk assessment tool, termed PLNM-Risk, that integrates clinical data, prostate biopsies, radiologists' identifications, high-throughput radiomics and deep learning imaging features from mpMRI into a more optimized intuitive model for PLNM prediction in patients with localized PCa.

2. Methods

2.1. Ethical information and study cohorts

This retrospective study was approved by the local Research Ethics Board of The First Affiliated Hospital of Nanjing Medical University (protocol 2016-SRFA-093), and informed patient consent was waived. All procedures performed in studies involving human participants were in accordance with the 1964 Helsinki declaration and its later amendments.

A total of 1843 patients admitted to prostate mpMRI at two tertiary care medical centres (centre 1: The First Affiliated Hospital of Nanjing Medical University, time interval: between Sep 2012 and Jun 2019; centre 2: The First Affiliated Hospital of Soochow University, time interval: between Jan 2016 and Dec 2019) with biopsy-confirmed PCa were retrospectively screened. Among them, 401 patients with pathologically proven PCa (centre 1, n = 351; centre 2, n = 50) who underwent both radical prostatectomy (RP) and ePLND treatment were finally included. All patients had no history of previous surgery, radiotherapy or adjuvant therapies for PCa before mpMRI. A flow diagram of patient selection with inclusion and exclusion criteria is provided in supplementary **Fig. S1**.

Data from centre 1 were randomly split into training/validation (n = 280) and test (n = 71) groups for model development and internal testing, respectively. Data from centre 2 (n = 50) were used for external tests.

2.2. Preoperative clinical characteristics and histological data

Clinical and biopsy characteristics included age, serum prostate-specific antigen (PSA), PSA density (PSAD), clinical T stage ($\leq$ T1c, T2a-c, $\geq$ T3), biopsy Gleason score ($\leq$ 3+3, 3+4, 4+3, and $\geq$ 4+4), percentage of positive cores, and perineural invasion (absent or present). The surgical findings included Gleason score ($\leq$ 3+3, 3+4, 4+3, and $\geq$ 4+4), surgical margin (negative vs positive), extracapsular extension (ECE) (negative vs positive), seminal vesicle invasion (SVI) (negative vs positive), and PLNM (negative vs positive). Histopathological results of ePLND, including the total number of resected lymph nodes and total number of positive nodes, were recorded.

Transrectal ultrasound (TRUS)-guided cognitive biopsy and/or targeted TRUS/MRI-fusion biopsy were performed followed by a standard 13-core systematic biopsy after mpMRI scans. RP with ePLND was performed in high-risk patients classified by the EAU risk group classification criteria. All biopsies and surgical specimens were prepared and examined by two pathologists who had 10 years of

experience in urologic pathology according to the ISUP 2005 and 2014 recommendations [21,22].

2.3. Follow-up

The first postoperative visit was 6 weeks after RP and ePLND, and then patients were consistently followed up at intervals of 3 to 6 months based on PSA. The time of biochemical recurrence (BCR) was recorded. Patients were censored in case of emigration, or on 30th Jul 2020, whichever came first. PCa BCR was defined as three successive rises in PSA level of >0.1 ng/ml at least 6 weeks postoperatively with final PSA >0.2 ng/ml, or administration of secondary therapy for evidence of detectable PSA >0.1 ng/ml at least 6 weeks postoperatively, or PSA >0.4 ng/ml at least 6 weeks postoperatively, referred to criteria previously reported [23, 24].

2.4. Prostate mp-MRI examination and sequence pre-process

All images were acquired prior to prostatic biopsy on two 3.0T MR scanners (Verio and Skyra; Siemens, Erlangen, Germany) at the two institutes with a pelvic phased-array coil. The scan protocol was a combination of transverse T_1WI, transverse, coronal, and sagittal T_2WI, and transverse DWI. The apparent diffusion coefficient (ADC) value was measured by expending a monoexponential fitting model. The details of the MRI parameters are summarized in **Table S1**.

The elastic transform from DWI and ADC to high-resolution T_2WI was estimated by an Elastix software package (v. 4.10) referring to the suggested parameter file "*par0001bspline16*". The parameter configuration for the registration was set according to Klein et al.'s work [25]. The aligned DWI and ADC had the same resolution, field of view and orientation as T_2WI. Before image analysis, T_2WI, DWI and ADC images within and between the two sites were normalized by remapping their histogram to fit within $\mu \pm 3\sigma$ (μ: mean gray-level within the VOI; σ: gray-level standard deviation).

2.5. Radiologists' interpretation

All images were retrospectively interpreted using the Prostate Imaging and Reporting and Data System (PI-RADS) version 2.1 [26] by four genitourinary radiologists at two institutions (reader 1 [Z.Y.] with 10 years of experience and reader 2 [Z.J.] with 15 years of experience in prostatic MRI in centre 1; reader 3 [B.J.] with 5 years of experience and reader 4 [W.X.] with 10 years of experience in prostatic MRI in centre 2). All readers were blinded to the clinicopathological information. In each patient, radiologists first identified a leading cancer lesion, which had a higher PI-RADS score or larger diameter if the score was the same. Then, the following imaging features related to the leading cancer lesion were assessed: (i) prostate volume; (ii) zone of lesion originated (peripheral zone [PZ] or transitional zone [TZ]); (iii) lesion shape (regular or irregular); (iv) lesion margin (well-defined or ill-defined); (v) lesion max diameter; (vi) lesion volumetric mean ADC value; (vii) PI-RADS score (PI-RADS 1-5) [27,28]; (viii) MRI T-stage ($\leq$ T1c, T2a, T2b, $\geq$ T2c); and (ix) MRI-based assessments for ECE, SVI, and PLNM (absent or present). The definition of MRI-based assessments was described according to criteria previously reported [20,29]. All cases were interpreted individually first and then reviewed in consensus by the two readers 4 weeks after individual evaluation. Individual readings were used to evaluate interobserver agreement, and a consensus review was used for model development. Any disagreement in consensus review was discussed until a final standard consensus was generated.

2.6. Radiomics machine learning and deep learning

PCa lesions were manually segmented by two genitourinary fellows (Reader 5 [H.Y.] and Reader 6 [J.K. W]) independently. The entire volume of interest (VOI) of the tumour was drawn based on radiological-pathological correlation slice by slice. For patients with multiple tumour lesions, only the leading cancer lesion was analysed. All sequences and the corresponding VOIs were resampled to an inner resolution of 0.5×0.5 mm^2 by Bicubic methods. To ensure the image quality and accuracy of the VOI, the presence of deformation was rechecked by radiologists at the same time.

Radiomics features, including shapes (14), histograms (18), textures (75), and wavelets (744), were extracted from T_2WI, DWI with b = 1500 s/mm^2, and ADC by PyRadiomics [30]. Finally, a total of 2553 features were obtained. Each feature was normalized to a similar scale to avoid scale effects. The feature matrix was normalized by Z-score normalization, which maps each feature with a mean of zero and a standard deviation of one. The formula was described as follows:$z - score = \frac{x_i - \bar{x}}{\sigma}$, where $\bar{x}$ is the mean value of the feature dedicated and σ is the standard deviation of all mean values. The Pearson product-moment correlation coefficient (PCC) was estimated between each pair of features, and random features were removed if the value was larger than 0.85. Additional tumour volumes (in mm^3, the default of PyRadiomics) in all sequences were excluded from the assessment.

The entire-volumetric radiomics features focus on only the inner regions of PCa. We further proposed a deep transfer learning representation (DTLR) approach to determine the interactive effect between tumours and tumour-related regions. The tumour-related region is a 5-mm extended region around the tumour, which can be automatically determined using erosion and dilation algorithms as described in our previous study [31]. DTLR features were measured on the combined tumour and tumour-related regions using a dedicated image embedding toolbox package (https://github.com/biolab) through five pre-trained deep neural networks using ImageNet data, i.e., DeepLoc, Inception v3, SqueenzeNet, VGG-16 and VGG-19 as embedders [32]. Just as the human visual cortex can adapt to the analysis of many scenes and images, we assumed that the proposed image embedders pretrained on a sufficiently large number of diverse images may infer useful features from a broad range of new image sets. The idea is that deep transfer learning stores the knowledge obtained from one problem in a trained model and applies it to another problem, which may be quite different. For each case, the axis slice with the largest area of the tumour cross section was selected from axial T_2WI, high-b DWI and ADC images. To obtain the representative imaging features of the target lesion, we used hand-cropped ROI as an attention to gate model training (i.e., regions around the PCa) in the centre slice of the MRI scan. For image embedding, we used the penultimate layer of the model to produce a total of 37,809 new feature vectors, serving as another set of deep learning-derived imaging features in parallel to radiomics features for the prediction of PLNM. The detailed parameters of each embedder are summarized in the supplementary data (**Table S2**).

Random forest classifiers, which combined the concepts of feature selection and step model training, were used to build radiomics and deep learning signatures for predicting PLNM. Cross-validation was applied on the training cohort to optimize the hyperparameters of each method, and one standard error was used to determine the number of features. In the random forest algorithm, each tree is developed from a bootstrap sample from the training data. When developing individual trees, an arbitrary subset of attributes is drawn from which the best attribute for the split is selected. The final model is based on the majority vote from individually developed trees in the forest. Regarding feature selection, we assessed the features using a mean decrease Gini index (MDGI). The MDGI represents the importance of individual features for correctly classifying a residue into linker and nonlinker regions. The MDGI was calculated by classifying randomly selected linker features and nonlinker features, and the mean MDGI was calculated as the averaged MDGI over 100 trials. The mean MDGI z-score of each feature was calculated as *MDGI*

$z-score = \frac{x_i-\bar{x}}{\sigma}$, where $\bar{x}$ is the mean MDGI of the feature dedicated and σ is the standard deviation of all mean MDGI. Vector elements with MDGI Z-Score larger than 2 were selected as optimum feature candidates. The direct output values of the random forest classifiers do not show probabilities of PLNM positivity of PCa; therefore, we converted their output values to probabilities (*Pi*) by applying a sigmoid function as follows: $P_i = 1/(1+exp(-x_i))$, where x_i is the classifier output value. The value of P_i indicates the probability that the observation is PLNM (+). Thus, possibilities termed the Rad score and DTLR score were obtained.

2.7. Development, performance, and validation of PLNM-Risk model

Clinicopathological variables included age, PSA, PSAD, clinical T stage, biopsy Gleason score, percentage of positive cores and perineural invasion. Radiological measurements included the prostate volume, zone of lesion origin, shape, margin, maximum diameter, volumetric ADC, PI-RADS score, MRI T-stage, MRI-based ECE, SVI and PLNM. Finally, eighteen clinical, biopsy and radiological variables as well as the Rad score and DTLR score were integrated into a PLNM-Risk model for predicting the nodal status of PCa patients.

We used an open-source AutoGluon platform for PLNM-Risk model development and validation. By processing raw CSV input into predictions for test data, the AutoGluon framework employs a novel form of multilayer stack ensembling. The first layer of AutoGluon has n types of base learners, including extremely randomized trees, k-nearest neighbours, gradient boosting machines, random forests and a tabular neural network, whose outputs are concatenated and then fed into the next layer, which itself consists of multiple stacker models. These stackers then act as base models to an additional layer. It merely employs random search for hyperparameter tuning, model selection, ensembling, feature engineering, data pre-processing, data splitting, etc., thus allowing us to implement all strategies for hyperparameter tuning, feature selection, model selection and ensemble. The entire flowchart of PLNM-Risk model development is shown in **Fig. 1**.

2.8. Statistical analysis

The interreader agreements for the measurable variables, such as tumour diameter and ADC, were analysed using the intraclass correlation coefficient test, and the MRI-derived categorical variables were measured by Kappa identity test analysis.

The detailed differences of clinical, biopsy and radiological factors, Rad and DTLR score of PLNM-absent and PLNM-present were compared by T-test or Mann-Whitney U test. The discrimination performance of predictive models was quantified by the area under the ROC curve (AUC) value in the primary training data and internally validated in the independent test data. Diagnostic sensitivity, specificity and accuracy were calculated at a cut-off point that maximized the value of the Youden index. True-positive and false-positive rates, weighted by the odds of the selected threshold probability of risk, were assessed to evaluate the clinical usefulness and net benefits of the PLNM-Risk model. BCR-free survival was defined as the time interval from the date of surgery to the date of BCR. Survival curves were generated with the Kaplan-Meier method and compared by two-sided log-rank tests. The statistical analysis was conducted with an R package (version 3.3.4; Project for Statistical Computing; http://www.Rproject.org). The reported statistical significance levels were all two-sided, with statistical significance set at 0.05.

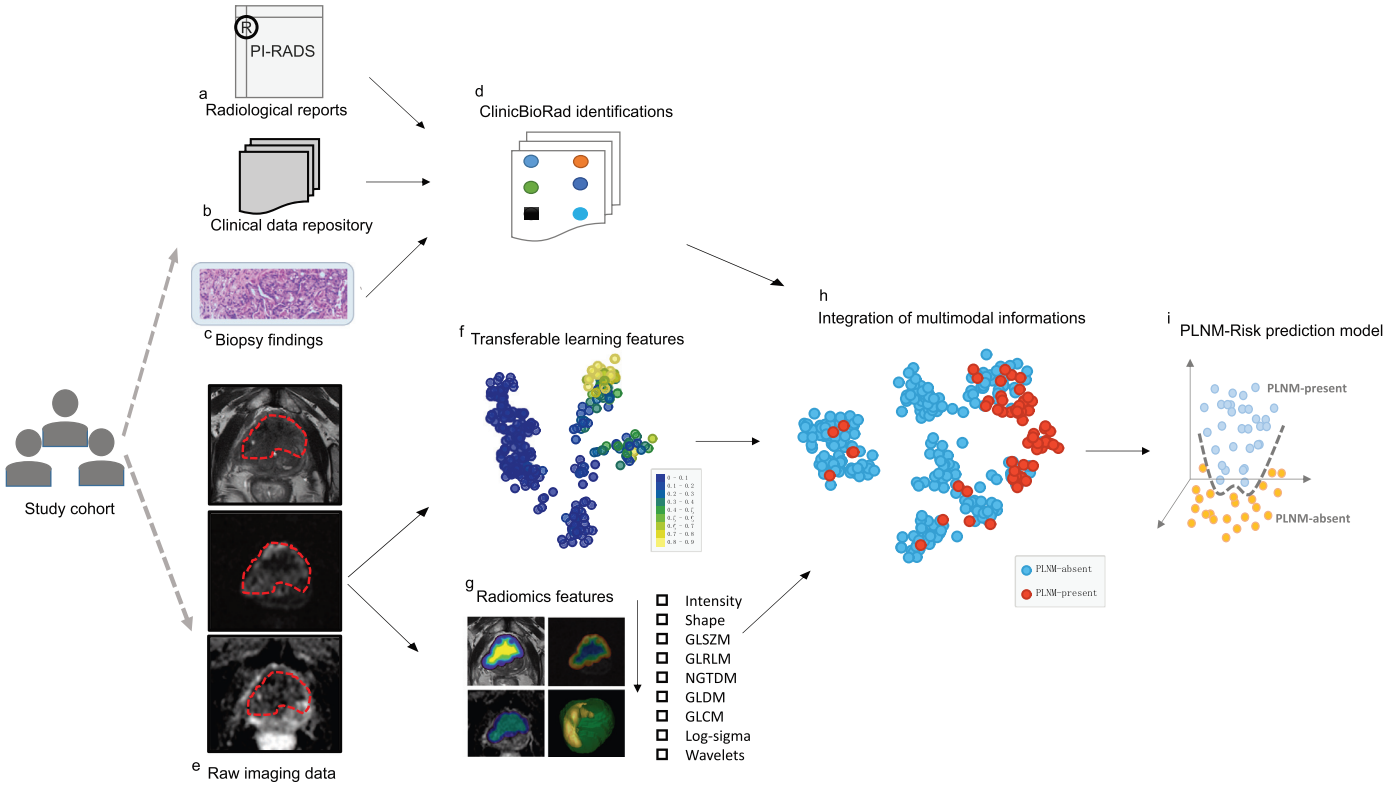


Fig. 1. Flowcharts of PLNM-Risk development. a–d, Radiologic characteristics, clinical data repository and biopsies within the primary cohort of study, in response to PLNM status, were interpreted by a panel of experts and encoded into a dedicative ClinicBioRad identification. **e-g** Individual image data within the cohort of study are embedded with training-sparing deep image embedders and a radiomics toolbox to extract various image feature representations. This produces deep imaging feature measurements with auto machine learning algorithms, resulting in complex network representations of PLNM. **h,i,** All translated datasets are then fed into the auto machine frameworks for predictive modelling of the outcome of interest.

PLNM = pelvic lymph node metastasis.

2.9. Role of funding source

The Key Research and Development Program of Jiangsu Province and Suzhou Science and Technology Bureau-Science and Technology Demonstration Project had neither a role in the study design, data collection, analyses or data interpretation nor in the writing of the report. The corresponding authors had full access to all of the data in the study and had final responsibility for the decision to submit the publication.

3. Results

3.1. Baseline characteristics

Out of all patients included, the presence of histological PLNM was diagnosed in the explanted tissue of 70/401 patients (17.5%), with 64/351 (18.2%) in centre 1 and 6/50 (12%) in centre 2. Detailed baseline characteristics of the patients are summarized in **Table 1**.

Table 1
The baseline characteristics of PCa patients between PLNM (-) and PLNM (+) in two medical centers.

Variable	Center 1			Center 2		
	PLNM (-) n = 287	PLNM (+) n = 64	*p*	PLNM (-) n = 44	PLNM (+) n = 6	*p*
Age (y), mean (std)	69.5 (6.6)	67.5 (7.2)	0.032†	70.1 (6.3)	66.2 (7.4)	0.160†
PSA (ng/mL), mean (range)	29.9 (1.4-591.0)	75.6 (3.9-676.0)	0.000†	31.2 (3.3-100.0)	77.6 (22.6-200.1)	0.155†
PSAD (ng/mL/cc), mean (range)	0.8 (0.1-9.2)	1.6 (0.1-16.3)	0.008†	0.8 (0.1-3.3)	1.8 (0.5-5.3)	0.008†
Clinical T-stage			0.031*			0.328*
T1c or less	180/287 (62.7)	32/64 (50.0)		29/44 (65.9)	3/6 (50.0)	
T2a-c	86/287 (30.0)	30/64 (46.9)		13/44 (29.5)	2/6 (33.3)	
T3 or more	21/287 (7.3	2/64 (3.1)		2/44 (4.5)	1/6 (16.7)	
Prostate volume (cm^3), mean (range)	39.0(11.2-189.6)	50.4 (11.2-117.7)	0.000†	43.8 (15.7-111.1)	46.5(36.4-61.0)	0.753†
Tumor shape			0.000*			0.650*
Regular	138/287(48.1)	6/64 (9.4)		16/44 (36.4)	1/6 (16.7)	
Irregular	149/287 (51.9)	58/64 (90.6)		28/44 (63.6)	5/6 (83.3)	
Zone of tumor origin			0.048*			1.000*
TZ	91/287(31.7)	12/64 (18.8)		11/44(25.0)	1/6 (16.7)	
PZ	196/287 (68.3)	52/64 (81.2)		33/44 (75.0)	5/6 (83.3)	
Tumor margin			0.000*			1.000*
Well-defined, n (%)	126/287 (43.9)	5/64 (7.8)		8/44 (18.2)	1/6 (16.7)	
Ill-defined, n (%)	161/287 (56.1)	59/64 (92.2)		36/44 (81.8)	5/6 (83.3)	
Tumor max diameter (cm), mean (range)	1.7 (0.4-5.9)	3.1 (0.7-6.3)	0.001†	2.1 (0.5-5.4)	3.1 (1.4-4.1)	0.055†
Mean ADC ($\times 10^{-3}$ s/mm^2), mean (std)	0.8 (0.2)	0.8 (0.2)	0.042†	1.0(0.2)	0.9 (0.1)	0.829†
PI-RADS score			0.000*			1.000*
1-2	19/287 (6.6)	0/64 (0.0)		3/44 (6.8)	0/6 (0.0)	
3	40/287 (13.9)	0/64 (0.0)		2/44 (4.5)	0/6 (0.0)	
4	95/287 (33.1)	4/64 (6.3)		8/44 (18.2)	1/6 (16.7)	
5	133/287 (46.3)	60/64 (93.7)		31/44 (70.5)	5/6 (83.3)	
MRI-based stage			0.000*			1.000*
T1c or less	42/287 (14.6)	0/64 (0.0)		2/44 (4.5)	0/6 (0.0)	
T2a	98/287 (34.1)	2/64 (3.1)		14/44 (31.8)	2/6 (33.3)	
T2b	24/287 (8.4)	5/64 (7.8)		4/44 (9.1)	0/6 (0.0)	
T2c or more	123/287 (42.8)	57/64 (89.1)		24/44 (54.5)	4/6 (66.7)	
MRI-ECE+	90/287 (31.4)	55/64 (85.9)	0.000*	15/44 (34.1)	3/6 (50.0)	0.654*
MRI-SVI+	29/287 (10.1)	43/64 (67.2)	0.000*	9/44 (20.5)	2/6 (33.3)	0.601*
MRI-LNI+	15/287 (5.2)	36/64 (56.3)	0.000*	10/44 (22.7)	2/6 (33.3)	0.621*
Biopsy findings			0.000*			0.504*
GS 3+3	70/287 (24.4)	3/64 (4.7)		0/44 (0.0)	0/6 (0.0)	
GS 3+4	63/287 (22.0)	5/64 (7.8)		9/44 (20.5)	0/6 (0.0)	
GS 4+3	77/287 (26.8)	21/64 (32.8)		11/44 (25.0)	1/6 (16.7)	
GS ≥ 4+4	77/287 (26.8)	35/64 (54.7)		24/44 (54.5)	5/6 (83.3)	
Percentage of positive cores, median (range)	0.4 (0.0-1.0)	0.7 (0.2-1.0)	0.000†	0.5(0.1-1.0)	0.8 (0.6-1.0)	0.008†
Perineural invasion+	45/287 (15.7)	28/64 (43.8)	0.000*	8/44 (18.2)	1/6 (16.7)	1.000*
Surgical findings						
GS 3+3	33/287 (11.5)	0/64 (0.0)		0/44 (0.0)	0/6 (0.0)	
GS 3+4	87/287 (30.3)	4/64 (6.3)		8/44 (18.2)	0/6 (0.0)	
GS 4+3	97/287 (33.8)	17/64 (26.6)		17/44 (38.6)	0/6 (0.0)	
GS ≥ 4+4	71/287 (24.7)	43/64 (67.2)		19/44 (53.2)	6/6 (100.0)	
ECE+	76/287 (33.0)	44/64 (68.8)		13/44 (29.5)	4/6 (66.7)	
SVI+	31/287 (13.5)	43/64 (67.2)		5/44 (11.4)	2/6 (33.3)	
SM+	91/287 (39.6)	48/64 (75.0)		15/44 (34.1)	4/6 (66.7)	
LN+	0/287 (0.0)	64/64 (100.0)		0/44 (0.0)	6/6 (100.0)	
No. of nodes dissected	2758	798		238	71	
No. of positive nodes	0	248		0	19	

Note. -Unless indicated otherwise, data are number of tumors, with percentages in parentheses. ADC = apparent diffusion coefficient; PSA = prostate serum antigen. PSAD = prostate serum antigen density. PI-RADS= Prostate Imaging and Reporting and Data System version 2.1; ECE = extracapsular extension. SVI = seminal vesicle invasion. LN = lymph node; PLNM = pelvic lymph node metastasis.GS= Gleason Score. RP=radical prostatectomy. SM=surgical margin.

† Independent sample t test.

* Chi-Square test.

3.2. Development, performance, and validation of PLNM-Risk

The dynamic performance tuning of two types of imaging signature models (Rad score and DTLR score) is described in supplementary **Fig. S2,** and the feature expression maps for the top radiomic features on T2WI and ADC images are shown in supplementary **Fig. S3**. The stacked ensemble variable importance and model importance are plotted in **Fig. 2** and show that the Rad score, MRI-based SVI assessment and DTLR score are the three most important predictors of PLNM. The *weighted_ensemble_k0_11 model*, with a stacked ensemble of 10 base models, achieves the best performance with respect to the prediction of outcome interest.

The resulting PLNM-Risk model (*weighted_ensemble_k0_11*) produced an area under the receiver operating characteristic curve (AUC) of 0.93 (95% CI, 0.90-0.96) in the training group, 0.92 (95% CI, 0.84-0.97) in the internal test group and 0.76 (95% CI, 0.62-0.87) in the external test group (**Fig. 3a**). The predicted PLNM-Risk score showed a significant difference (independent-samples T-test, $p < 0.001$) between the PLNM-absent and PLNM-present groups. Using an optimal threshold (PLNM-Risk > 0.131) that maximizes the Youden index of the ROC analysis in training data, the model resulted in a sensitivity of 92.2%, 84.6% and 50.0% with specificity of 81.2%, 84.5% and 84.0% in predicting PLNM in training, internal and external test sets, respectively (**Fig. 3b**). Additionally, the PLNM-Risk model demonstrated better performance than *Rad* score, *DTLR* score, *Rad + DTLR*, a *ClinicBioRad* model (incorporating clinical factors, biopsy findings and radiologist' interpretations), *ClinicBioRad + Rad* and *ClinicBioRad + DTLR* based on a single modality or multimodalities in the training and combined test cohorts (**Fig. 3c**).

3.3. Clinical implication of PLNM-Risk

The true and false positive rates of PLNM-Risk in comparison with established MSKCC and Briganti scores for stratifying PLNM risk are plotted in **Fig. 4**. For internal-tested data, PLNM-Risk resulted in similar true positive (TP) rates while achieving notably lower false positive (FP) rates at threshold probabilities of PLNM < 20% compared to MSKCC and Briganti scores. For externally tested data, the PLNM-Risk resulted in lower TP rates and lower FP rates compared to MSKCC and Briganti scores. According to the European Association of Urology (EAU) and National Comprehensive Cancer Network (NCCN) guidelines for the treatment of PCa, if the number of ePLNDs missed for risk assessment is controlled at < 2% (PLNM-Risk, 1.7% vs MSKCC 2.2% and Briganti 1.9%), the PLNM-Risk can provide both a higher number of ePLNDs spared (PLNM-Risk: 239/401 [59.6%] vs MSKCC: 180/401 [44.9%] vs Briganti 156/401[38.9%]) and a lower number of false positives (PLNM-Risk, 96/162 [59.3%] vs MSKCC, 155/221 [70.1%] and Briganti, 178/245 [72.7%]) compared with MSKCC and Briganti scores (supplementary **Tables S3,S4)**.

3.4. Prognostic aspects of PLNM-Risk for biochemical recurrence

As of 30th Jul 2020, 331/401 (82.5%) patients who had completed 3-year BCR follow-up after the surgery were enrolled. The overall recurrence rate was 34.7% (115/331), with 36.4% (104/286) in centre 1 and 24.4% (11/45) in centre 2. The median BCR-free survival time of the patients was 5.8 (range, 1.4-40.7) months: 5.2 (1.4-40.7) months for those with PLNM and 6.2 (1.4-38.1) months for those without PLNM (log-rank test, $p < 0.001$). Similar results were observed in the PLNM-Risk model: The median BCR-free survival time was 4.3 (1.4-28.8) months for patients with high PLNM-Risk scores and 8.0 (1.4-40.7) months for those with low PLNM-Risk scores (log-rank test, p < 0.001) in the whole cohort. BCR-free survival of the separate training, internal and external validation cohorts is shown in **Fig. 5**.

4. Discussion

In this study, we developed and validated a collaborative model that enables the integration of clinical data, biopsies, experts' knowledges, radiomics and transferable deep learning image features into

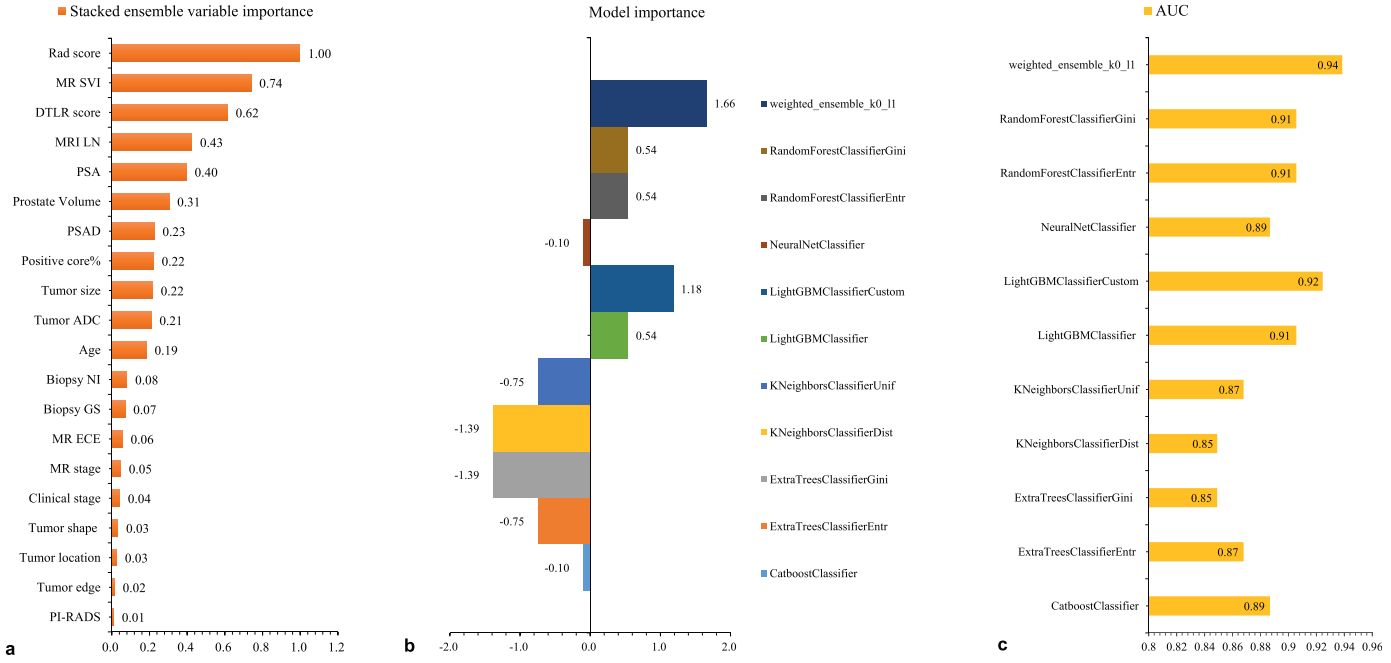
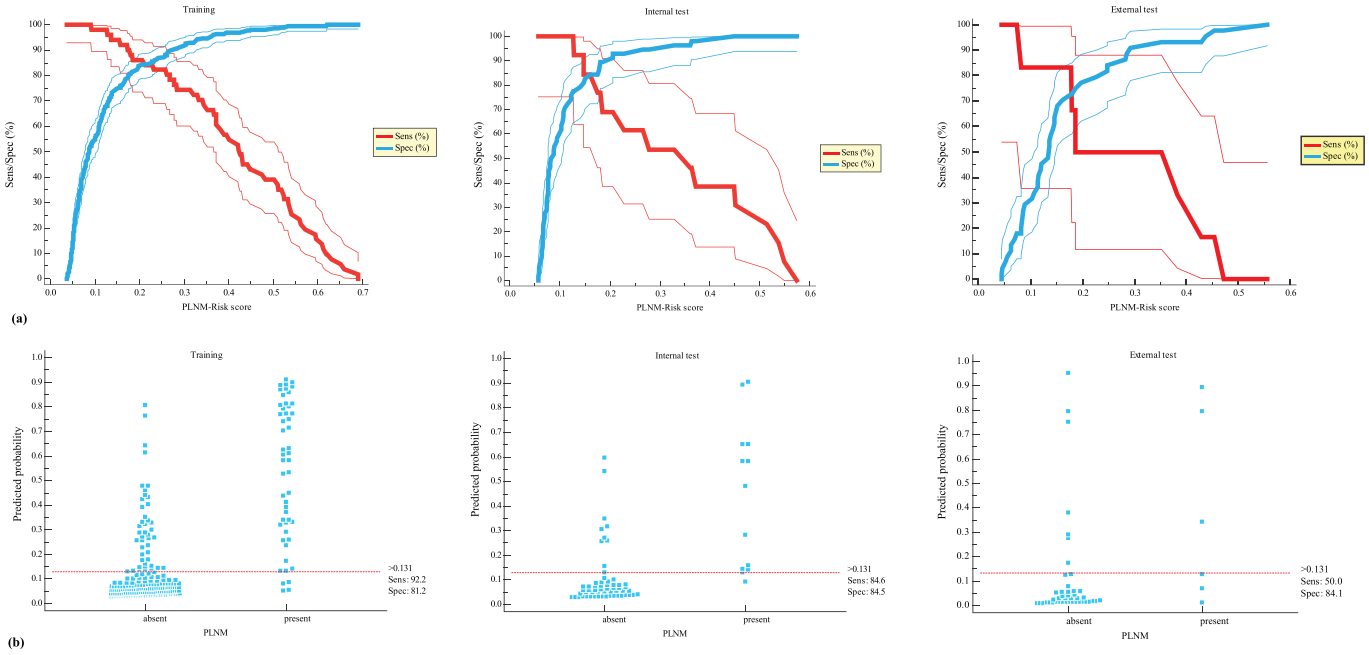


Fig. 2. Importance (a,b) of multimodal variables and performance (c) of base machine learning models for PLNM-Risk. Stacked ensemble variable importance is generated by averaging the importance from ten base models. Model importance is generated by: $importance = \frac{x_i - \overline{x}}{\sigma}$, where x_i is the AUC of each model dedicated; and $\overline{x}$ and σ are the mean and standard deviation of all AUCs, respectively.

GS = Gleason score; NI = perineural invasion; DTLR = deep transferable learning image representation; ECE = extracapsular extension; SVI = seminal vesicle invasion; LN = lymph node; PI-RADS = Prostate Imaging and Reporting and Data System; PSA = prostate serum antigen; PSAD = PSA density; ADC = apparent diffusion coefficient; PLNM = pelvic lymph node metastasis; AUC = area under the receiver operating characteristic curve.



Diagnostic performance of predictive models for N-staging

Model	*Training/Validation (n = 280)*			*Internal Test (n = 71)*			*External Test (n = 50)*		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
Rad score	0.86 (0.81-0.90)	44/51 (0.863)	167/229 (0.730)	0.83 (0.72-0.91)	11/13 (0.846)	32/58 (0.552)	0.78 (0.65-0.89)	4/6 (0.667)	32/44 (0.727)
DTLR score	0.86 (0.81-0.89)	36/51 (0.706)	200/229 (0.873)	0.84 (0.74-0.92)	6/13 (0.462)	51/58 (0.880)	0.78 (0.64-0.88)	3/6 (0.500)	38/44 (0.864)
Rad +DTLR	0.89 (0.84-0.92)	38/51 (0.745)	198/229 (0.865)	0.85 (0.75-0.93)	8/13 (0.615)	53/58 (0.907)	0.77 (0.62-0.87)	3/6 (0.500)	38/44 (0.864)
ClinicBioRad	0.90 (0.86-0.93)	43/51 (0.843)	192/229 (0.838)	0.88 (0.78-0.94)	10/13 (0.769)	48/58 (0.828)	0.65 (0.50-0.78)	4/6 (0.667)	29/44 (0.659)
ClinicBioRad+ Rad	0.92 (0.88-0.95)	44/51 (0.863)	197/229 (0.860)	0.93 (0.84-0.98)	10/13 (0.769)	49/58 (0.845)	0.72 (0.57-0.83)	4/6 (0.667)	32/44 (0.727)
ClinicBioRad+DTLR	0.92 (0.88-0.95)	44/51 (0.863)	195/229 (0.852)	0.91 (0.82-0.97)	10/13 (0.769)	50/58 (0.862)	0.71 (0.57-0.83)	4/6 (0.667)	31/44 (0.705)
PLNM-Risk	0.93 (0.90-0.96)	47/51 (0.922)	186/229 (0.812)	0.92 (0.84-0.97)	11/13 (0.846)	49/58 (0.845)	0.76 (0.62-0.87)	3/6 (0.500)	37/44 (0.841)

Note. PLNM = pelvic lymph node metastasis; Rad = radiomics; DTLR = deep transfer learning feature representation; AUC = area under the receiver operating characteristic curve; Sens = sensitivity; Spec = specificity.

(c)

Fig. 3. Performance of PLNM-Risk in primary training, internal testing and external validation datasets. (a) Sens and Spec curves are created by sweeping a threshold over the predicted probability, and the hidden lines represent the 95% confidence interval. (b) Distributions of predicted results are plotted in the PLNM-absent vs PLNM-present group; using the optimal risk threshold (0.131), the sensitivity and specificity of the PLNM-Risk model are determined. (c) The performance of the models trained with single- and multi-modality integrations is compared, in which PLNM-Risk is superior to any other model for performance in the whole internal and external test cohort.

PLNM = pelvic lymph node metastasis; Sens = sensitivity; Spec = specificity.

an interpretable tool to improve the prediction of PLNM in localized PCa. Unlike most approaches focusing on the task of combined algorithm selection and hyperparameter optimization [33,34], our study used an auto meta-learning framework, allowing us to combine the aggregated predictions of base models as its features and to exploit interactions between base models that offer enhanced predictive accuracy. This deep multimodal interaction and data mining approach might provide a new alternative for solving similar medical tasks. Second, our deep image embedding does not need training on a closely related set of images. This is functional for avoiding potential model overfitting when the cohort is limited and thus contributes important methodology to address the problem of high-dimensional data classification in the face of a limited number of samples. Finally, our results on a cohort of 401 patients with ePLND from two tertiary care medical centres show promise of the PLNM-Risk model and potential utilities of this tool for therapeutic decision making. Additionally, the results of the follow-up of BCR show a favourable prognostic aspect of PLNM-Risk for disease progression assessment.

In our previous single-centre cohort study [20], we presented in-house traditional machine learning methods by integrating clinical factors and radiologists' interpreted imaging features to predict PLNM. The results were promising for precision risk assessment of PCa, which is consistent with the studies of Wang et al. and Brembilla et al. [16,35]. In the current work, distinct facets of our results deserve attention. First, our results showed that the newly built imaging hallmarks, i.e., Rad score and DTLR score, generated from high-throughput imaging features, were workable for the nodal staging of PCa. Second, combining the Rad score and DTLR score with conventional clinical factors and radiologists' interpreted imaging features improved the prediction performance and showed an incremental role to radiologists' interpretation for PLNM prediction. Among 2,000 radiomics features, factors related to tumour gray level intensity (e.g., wavelet coarseness and emphasis) on high-b DWI and ADC and tumour heterogeneity on T_2WI (e.g., complexity and entropy) were identified as the top predictors for PLNM. We assumed that gray-level-intensity features on DWI and ADC might reflect the intratumour cellularity that connects to the PCa Gleason grade.

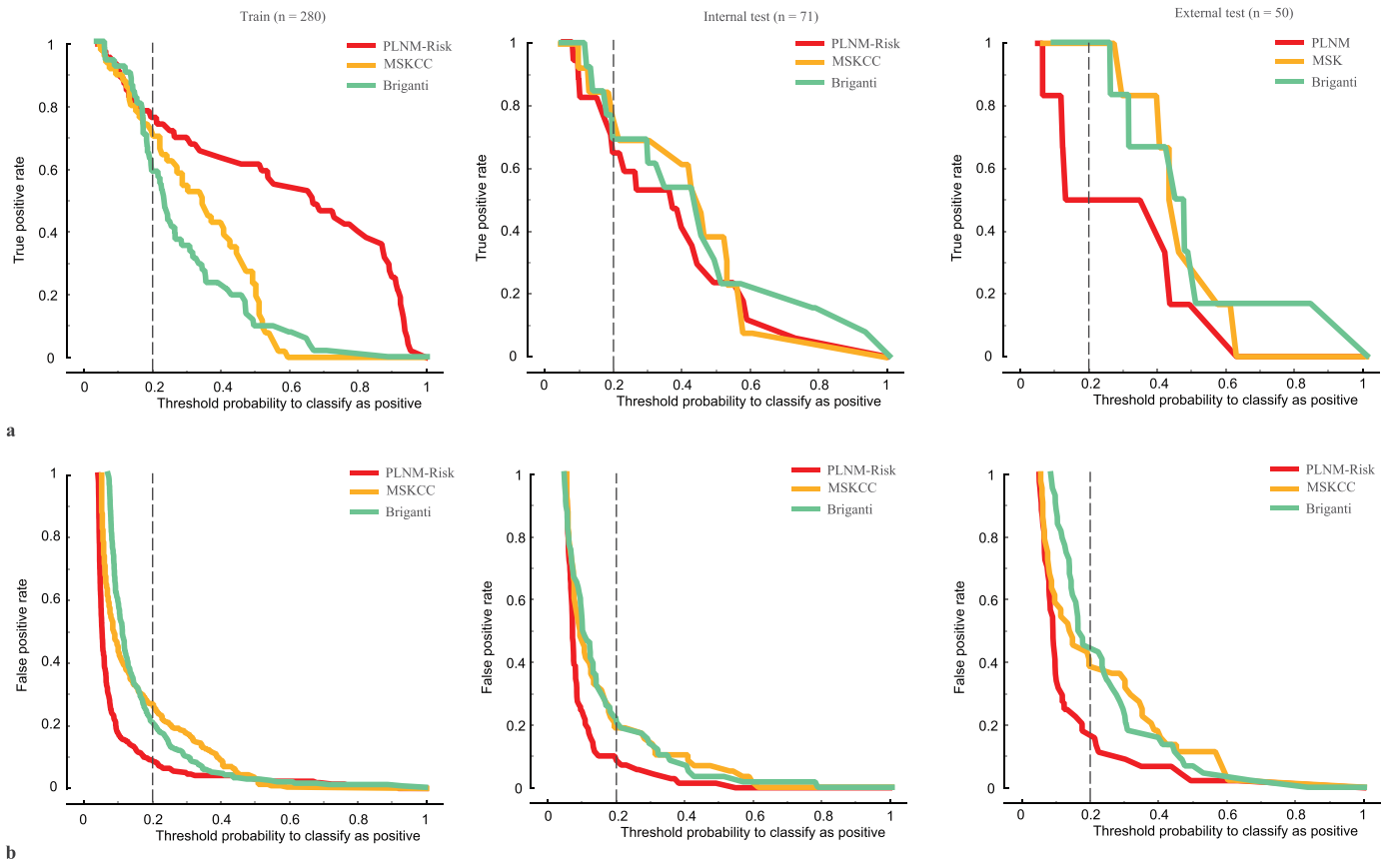


Fig. 4. Assistant role of the PLNM-Risk model in selecting patients who are candidates for ePLND. Plots show true positive rates **(a)** and false negative rates **(b)** of MSKCC, Briganti and PLNM-Risk for the prediction of PLNM in the training, internal test and external test cohorts, respectively. For the internal test cohort, PLNM-Risk resulted in a similar true positive rate but achieved a notably lower false positive rate at threshold probabilities of PLNM < 20% (dashed line) compared to MSKCC and Briganti score. For the external test cohort, PLNM-Risk resulted in a lower true positive rate and lower false positive rate than MSKCC and Briganti scores. This might be caused by sample bias, as there were only 6/50 PLNM-positive cases in the external cohort.

PLNM = pelvic lymph node metastasis; MSKCC = Memorial Sloan Kettering Cancer Center.

Volumetric complexity and entropy measured on high-resolution T_2WI might reflect intratumour heterogeneity, which is an important predictor of disease aggressiveness. Our findings are partly consistent with those of previous studies, in which quantitative measurements on mpMRI could be helpful for the prediction of PCa Gleason grade and disease aggressiveness [36–38]. Third, we set up separate internal and external validation sets. To some extent, the accuracy, repeatability and generalization of our PLNM-Risk model were proven. Last but not least, instead of comparing different traditional machine learning methods in the previous study, we used the new Auto Gluon platform to generate an ensemble, hyperparameter optimized classifier, which allows us to provide robust performance results. Due to the powerful AutoGluon that allows multimodal interaction and a multialgorithm stack ensemble, our PLNM-Risk model, integrating 20 multimodal factors and embedding 10 base models, was superior to any other existing clinical nomogram. This data mining approach might be helpful and inspiring to solve similar medical problems. Our findings are consistent with those of previous studies, which used similar computational approaches for breast cancer assessment [39,40].

Additionally, we made a head-to-head comparison of PLNM-Risk with the MSKCC and Briganti nomograms. Patients with a risk threshold higher than 5% estimated by the MSKCC or Briganti nomograms are candidates for ePLND on the basis of the EAU guidelines [41,42]. The adoption of a 5% threshold at the Briganti nomogram in the original cohort resulted in sparing 65% of ePLNDs at the cost of only 1.5% PLNM missed [6]. On the following external validation, using a 7% cut-off, the Briganti nomogram had an AUC of 0.79 and resulted in 56-70% ePLNDs spared at the cost of 1.5-2.6% PLNM missed [43–45]. In our cohort, with the PLNM missing rate controlled less than 2%, the PLNM-Risk resulted in a higher number of ePLNDs spared at lower cost of number of PLNMs missed than the MSKCC and Briganti nomograms. Therefore, our PLNM-Risk, using an optimal threshold of 8%, can result in a higher number of ePLNDs spared at the cost of fewer PLNMs missed. Our preliminary results also showed that the PLNM-Risk even revealed a potential role in predicting the preoperative disease progression risk. We found that BCR-free survival of PCa patients after RP differed significantly in subgroups stratified by PLNM-Risk.

However, as shown by the results in the external validation cohort, the PLNM-Risk might not be able to completely replace the current MSKCC or Briganti nomogram, the accuracy of which has been internally and externally validated. This may be caused by the small external test cohort of patients and the difference in PLNM distributions in positive and negative data. In addition, MRI-derived characteristics of patients in two hospitals were evaluated by radiologists at their respective sites (reader 1 and reader 2 evaluating the data from Centre 1 and reader 3 and reader 4 evaluating the data from Centre 2). It also requires the expertise of experienced radiologists, and there is inconsistency across readers of varying experience. In addition, subjective radiologists' interpretations lack standardization, which may be another important factor exerting an influence on external validation. Therefore, it is further necessary to test the generalization of PLNM-Risk in more external cohorts.

Our study had several limitations. First, because of its retrospective character, a large number of patients who did not undergo

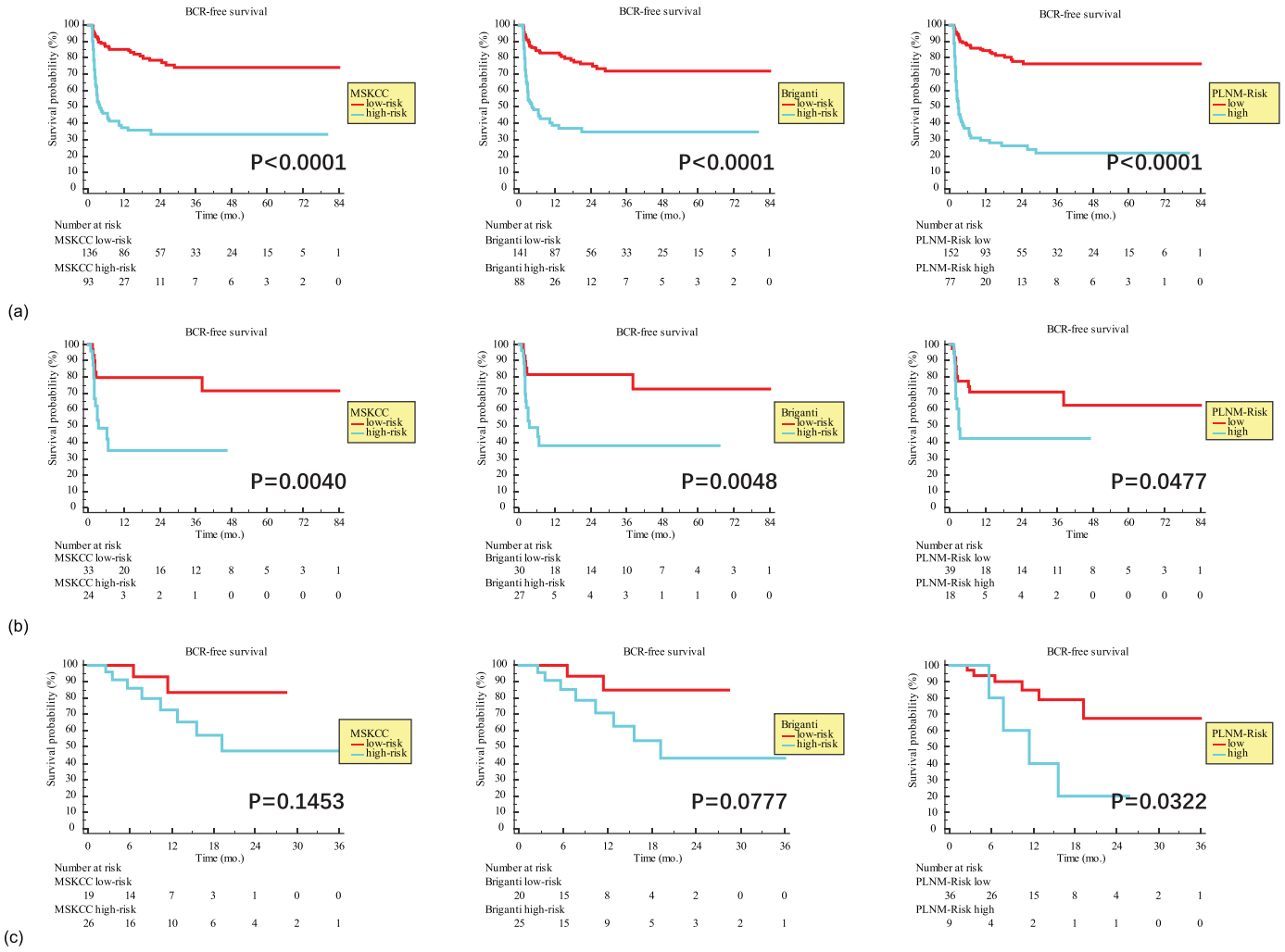


Fig. 5. Kaplan-Meier survival curves of biochemical recurrence (BCR) according to MSKCC, Briganti and PLNM-Risk in the training (a), internal test (b) and external test cohorts (c). The results reflect significantly different BCR-free survival in the subgroups stratified by MSKCC, Briganti and PLNM-Risk scores, implying prognostic relevance for the short- and long-term management of patients. P values were obtained from the log-rank tests.

BCR = biochemical recurrence; PLNM = pelvic lymph node metastasis; MSKCC = Memorial Sloan Kettering Cancer Center.

ePLND were excluded from the study, and selection bias and treatment procedure might influence model development. Second, the model was tested externally on a relatively small group of patients, and validation on a larger number of patients in multiple centres is needed. Third, prospective application of an automatic decision support tool in a clinical scenario involves a series of challenges, such as individual variations in the pathological manifestations of disease and technical variations in imaging devices, parameter setting and image processing. Therefore, in its current state, our PLNM-Risk might not be routinely available.

In summary, we proposed a PLNM-Risk tool for the pretreatment prediction of PLNM in PCa. Our end-to-end model, maintaining the exploratory nature of multimodal high-dimensional database integration and multialgorithm embedding, provides strong predictive capabilities and prognostic accuracy in internal and external validation. The interpretability of PLNM risk is particularly important for building trustworthy autodecision tools for clinical applications. Adoption of this tool would allow 59.6% of ePLND procedures to be spared at the cost of missing only 1.7% of PLNM cases, which is better than the current MKSCC and Briganti nomograms. Therefore, our automachine intelligence approach would provide a great alternative to improve routine procedures in patient management.

Data sharing section

The data supporting the findings of this study are available within the article and/or the supplementary materials. Created datasets and code are not publicly available. The data not shown can be available from the corresponding author Xi-Ming Wang or Yu-Dong Zhang upon reasonable requests.

CRediT authorship contribution statement

Ying Hou: Conceptualization, Data curtion, Formal analysis, Writing – original draft. **Jie Bao:** Conceptualization, Data curtion, Formal analysis, Writing – original draft. **Yang Song:** Formal analysis, Writing – original draft. **Mei-Ling Bao:** Formal analysis, Writing – original draft. **Ke-Wen Jiang:** Data curtion, Formal analysis, Writing – original draft. **Jing Zhang:** Data curtion, Formal analysis, Writing – original draft. **Guang Yang:** Formal analysis, Writing – original draft. **Chun-Hong Hu:** Data curtion, Formal analysis, Writing – original draft. **Hai-Bin Shi:** Formal analysis, Writing – original draft. **Xi-Ming Wang:** Data curtion, Formal analysis, Project administration, Writing – original draft. **Yu-Dong Zhang:** Data curtion, Formal analysis, Project administration, Writing – original draft.

Declaration of Competing Interest

The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflicts of interest with respect to this manuscript.

CRediT authorship contribution statement

Ying Hou: Conceptualization, Data curtion, Formal analysis, Writing – original draft. **Jie Bao:** Conceptualization, Data curtion, Formal analysis, Writing – original draft. **Yang Song:** Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Mei-Ling Bao:** Formal analysis, Writing – original draft. **Ke-Wen Jiang:** Data curtion, Formal analysis, Writing – original draft. **Jing Zhang:** Data curtion, Formal analysis, Writing – original draft. **Guang Yang:** Formal analysis, Methodology, Software, Writing – original draft. **Chun-Hong Hu:** Data curtion, Formal analysis, Writing – original draft. **Hai-Bin Shi:** Formal analysis, Writing – original draft. **Xi-Ming Wang:** Data curtion, Formal analysis, Project administration, Writing – original draft. **Yu-Dong Zhang:** Data curtion, Formal analysis, Project administration, Writing – original draft.

Acknowledgment

Contract grant sponsor: Key Research and Development Program of Jiangsu Province; contract grant number: BE2017756 (to Y.D.Z.); Contract grant sponsor: Suzhou Science and Technology Bureau-Science and Technology Demonstration Project; contract grant number: SS201808 (to X.M.W)

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2021.103395.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. 2018;68(1):7-30.
[2] Wilczak W, Wittmer C, Clauditz T, Minner S, Steurer S, Buscheck F, et al. Marked prognostic impact of minimal lymphatic tumor spread in prostate cancer. Eur Urol 2018;74(3):376–86.
[3] von Bodman C, Godoy G, Chade DC, Cronin A, Tafe LJ, Fine SW, et al. Predicting biochemical recurrence-free survival for patients with positive pelvic lymph nodes at radical prostatectomy. J Urol 2010;184(1):143–8.
[4] Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent. Eur Urol 2017;71(4):618–29.
[5] Muteganya R, Goldman S, Aoun F, Roumeguere T, Albisinni S. Current imaging techniques for lymph node staging in prostate cancer: a review. Front Surg 2018;5:74.
[6] Briganti A, Larcher A, Abdollah F, Capitanio U, Gallina A, Suardi N, et al. Updated nomogram predicting lymph node invasion in patients with prostate cancer undergoing extended pelvic lymph node dissection: the essential importance of percentage of positive cores. Eur Urol 2012;61(3):480–7.
[7] Memorial Sloan Kettering Cancer Center. Dynamic prostate cancer nomogram: coefficients. cited; Available from: https://wwwmskccorg/nomograms/prostate/pre-op/coefficients. Last Updated: January 14, 2020
[8] Tosoian JJ, Chappidi M, Feng Z, Humphreys EB, Han M, Pavlovich CP, et al. Prediction of pathological stage based on clinical stage, serum prostate-specific antigen, and biopsy gleason score: partin tables in the contemporary era. BJU Int 2017;119(5):676–83.
[9] Hueting TA, Cornel EB, Somford DM, Jansen H, van Basten JA, Pleijhuis RG, et al. External validation of models predicting the probability of lymph node involvement in prostate cancer patients. Eur Urol Oncol 2018;1(5):411–7.
[10] Grivas N, Wit E, Tillier C, van Muilekom E, Pos F, Winter A, et al. Validation and head-to-head comparison of three nomograms predicting probability of lymph node invasion of prostate cancer in patients undergoing extended and/or sentinel lymph node dissection. Eur J Nucl Med Mol Imaging 2017;44(13):2213–26.
[11] Nason GJ, O'Connor EM, MacMahon D, Moss B, Considine SW, Cahill A, et al. Comparison of nomograms predicting lymph node invasion in patients undergoing radical prostatectomy for prostate cancer. Ir J Med Sci 2018;187(1):33–7.
[12] Cimino S, Reale G, Castelli T, Favilla V, Giardina R, Russo GI, et al. Comparison between Briganti, Partin and MSKCC tools in predicting positive lymph nodes in prostate cancer: a systematic review and meta-analysis. Scand J Urol 2017;51(5):345–50.
[13] Yakar D, Debats OA, Bomers JG, Schouten MG, Vos PC, van Lin E, et al. Predictive value of MRI in the localization, staging, volume estimation, assessment of aggressiveness, and guidance of radiotherapy and biopsies in prostate cancer. J Magn Reson Imaging JMRI 2012;35(1):20–31.
[14] Woo S, Suh CH, Kim SY, Cho JY, Kim SH. The diagnostic performance of mri for detection of lymph node metastasis in bladder and prostate cancer: an updated systematic review and diagnostic meta-analysis. AJR Am J Roentgenol 2018;210(3):W95–w109.
[15] Porpiglia F, Manfredi M, Mele F, Bertolo R. Indication to pelvic lymph nodes dissection for prostate cancer: the role of multiparametric magnetic resonance imaging when the risk of lymph nodes invasion according to Briganti updated nomogram is <5. 2018;21(1):85-91.
[16] Wang L, Hricak H, Kattan MW, Schwartz LH, Eberhardt SC, Chen HN, et al. Combined endorectal and phased-array MRI in the prediction of pelvic lymph node metastasis in prostate cancer. AJR Am J Roentgenol 2006;186(3):743–8.
[17] Rayn KN, Bloom JB, Gold SA, Hale GR, Baiocco JA, Mehralivand S, et al. Added value of multiparametric magnetic resonance imaging to clinical nomograms for predicting adverse pathology in prostate cancer. J Urol 2018;200(5):1041–7.
[18] Alberts AR, Roobol MJ, Verbeek JFM, Schoots IG, Chiu PK, Osses DF, et al. Prediction of high-grade prostate cancer following multiparametric magnetic resonance imaging: improving the rotterdam European randomized study of screening for prostate cancer risk calculators. Eur Urol 2019;75(2):310–8.
[19] Morlacco A, Sharma V, Viers BR, Rangel LJ, Carlson RE, Froemming AT, et al. The incremental role of magnetic resonance imaging for prostate cancer staging before radical prostatectomy. Eur Urol 2017;71(5):701–4.
[20] Hou Y, Bao ML, Wu CJ, Zhang J, Zhang YD, Shi HB, et al. A machine learning-assisted decision-support model to better identify patients with prostate cancer requiring an extended pelvic lymph node dissection. BJU Int 2019;124(6):972–83.
[21] Epstein JI, Allsbrook Jr. WC, Amin MB, Egevad LL. The 2005 International society of urological pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma. Am J Surg Pathol 2005;29(9):1228–42.
[22] Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. The 2014 international society of urological pathology (ISUP) Consensus conference on gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. A J Surg Pathol 2016;40(2):244–52.
[23] Cookson MS, Aus G, Burnett AL, Canby-Hagino ED, D'Amico AV, Dmochowski RR, et al. Variation in the definition of biochemical recurrence in patients treated for localized prostate cancer: the american urological association prostate guidelines for localized prostate cancer update panel report and recommendations for a standard in the reporting of surgical outcomes. J Urol 2007;177(2):540–5.
[24] Brockman JA, Alanee S, Vickers AJ, Scardino PT, Wood DP, Kibel AS, et al. Nomogram predicting prostate cancer-specific mortality for men with biochemical recurrence after radical prostatectomy. Eur Urol 2015;67(6):1160–7.
[25] Klein S, Staring M Fau - Murphy K, Murphy K Fau - Viergever MA, Viergever Ma Fau - Pluim JPW, Pluim JP. Elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging 2010;29(1):196–205.
[26] Padhani AR, Weinreb J, Rosenkrantz AB, Villeirs G, Turkbey B, Barentsz J. Prostate imaging-reporting and data system steering committee: PI-RADS v2 status update and future directions. Eur Urol 2019;75(3):385–96.
[27] Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS prostate imaging - reporting and data system: 2015, version 2. Eur Urol 2016;69(1):16–40.
[28] Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, 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(3):340–51.
[29] Zhang YD, Wu CJ, Bao ML, Li H, Wang XN, Liu XS, et al. MR-based prognostic nomogram for prostate cancer after radical prostatectomy. J Magn Reson Imaging JMRI 2017;45(2):586–96.
[30] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77(21):e104.
[31] Xu X, Zhang HL, Liu QP, Sun SW, Zhang J, Zhu FP, et al. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. J Hepatol 2019;70(6):1133–44.
[32] Wallis TSA, Funke CM, Ecker AS, Gatys LA, Wichmann FA, Bethge M. A parametric texture model based on deep convolutional features closely matches texture appearance for humans. J Vis 2017;17(12):5.
[33] Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J clin Oncol Off J Am Soc Clin Oncol 2016;34(18):2157–64.
[34] Wang T, Gao T, Yang J, Yan X, Wang Y, Zhou X, et al. Preoperative prediction of pelvic lymph nodes metastasis in early-stage cervical cancer using radiomics nomogram developed based on T2-weighted MRI and diffusion-weighted imaging. Eur J Radiol 2019;114:128–35.
[35] Brembilla G, Dell'Oglio P, Stabile A, Ambrosi A, Cristel G, Brunetti L, et al. Preoperative multiparametric MRI of the prostate for the prediction of lymph node metastases in prostate cancer patients treated with extended pelvic lymph node dissection. Eur Radiol 2018;28(5):1969–76.
[36] Turkbey B, Shah VP, Pang Y, Bernardo M, Xu S, Kruecker J, et al. Is apparent diffusion coefficient associated with clinical risk scores for prostate cancers that are visible on 3-T MR images? Radiology 2011;258(2):488–95.

[37] Zhang YD, Wang Q, Wu CJ, Wang XN, Zhang J, Liu H, et al. The histogram analysis of diffusion-weighted intravoxel incoherent motion (IVIM) imaging for differentiating the gleason grade of prostate cancer. Eur Radiol 2015;25(4):994–1004.
[38] Fehr D, Veeraraghavan H, Wibmer A, Gondo T, Matsumoto K, Vargas HA, et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. Proc Natl Acad Sci U S A 2015;112(46):E6265–73.
[39] He T, Puppala M, Ezeana CF, Huang YS, Chou PH, Yu X, et al. A deep learning-based decision support tool for precision risk assessment of breast cancer. JCO Clin Cancer Inform 2019;3:1–12.
[40] Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nat Commun 2020;11(1):1236.
[41] Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent. Eur Urol 2017;71(4):618–29.
[42] National Comprehensive Cancer Network. NCCN guidelines https://www.nccn.org/professionals/physician_gls/f_guidelines.asp.
[43] Gandaglia G, Ploussard G, Valerio M, Mattei A, Fiori C, Fossati N, et al. A novel nomogram to identify candidates for extended pelvic lymph node dissection among patients with clinically localized prostate cancer diagnosed with magnetic resonance imaging-targeted and systematic biopsies. Eur Urol 2019;75(3):506–14.
[44] Gandaglia G, Martini A, Ploussard G, Fossati N, Stabile A, De Visschere P, et al. External validation of the 2019 Briganti nomogram for the identification of prostate cancer patients who should be considered for an extended pelvic lymph node dissection. Eur Urol 2020.
[45] Gandaglia G, Fossati N, Zaffuto E, Bandini M, Dell'Oglio P, Bravi CA, et al. Development and internal validation of a novel model to identify the candidates for extended pelvic lymph node dissection in prostate cancer. Eur Urol 2017;72(4):632–40.