

# Artificial Intelligence Is a Promising Prospect for the Detection of Prostate Cancer Extracapsular Extension With Mp-mri: A Two-center Comparative Study

**Ying Hou**

the first affiliated hospital of Nanjing Medical University

**Yi-Hong Zhang**

East China Normal University

**Jie Bao**

the first affiliated hospital of Soochow University

**Mei-Ling Bao**

The First Affiliated Hospital of Nanjing Medical University

**Guang Yang**

East China Normal University

**Hai-Bin Shi**

the first appiliated hospital of Nanjing Medical University

**Yang Song**

East China Normal University

**yudong zhang** (✉ [njmu\\_zyd@163.com](mailto:njmu_zyd@163.com))

The first affiliated hospital of Nanjing Medicial University

---

## Research Article

**Keywords:** prostate cancer, extracapsular extension, deep learning, artificial intelligence

**Posted Date:** March 19th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-298296/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Purpose:** A balance between preserving urinary continence and achievement of negative margins is of clinical relevance while implementary difficulty. Preoperatively accurate detection of prostate cancer (PCa) extracapsular extension (ECE) is thus crucial for determining appropriate treatment options. We aimed to develop and clinically validate an artificial intelligence (AI)-assisted tool for the detection of ECE in patients with PCa using multiparametric MRI.

**Methods:** 849 patients with localized PCa underwent multiparametric MRI before radical prostatectomy were retrospectively included from two medical centers. The AI tool was built on a ResNeXt network embedded with a spatial attention map of experts' prior knowledges (PAGNet) from 596 training data sets. The tool was validated in 150 internal and 103 external data sets, respectively; and its clinical applicability was compared with expert-based interpretation and AI-expert interaction.

**Results:** An index PAGNet model using a single-slice image yielded the highest areas under the receiver operating characteristic curve (AUC) of 0.857 (95% confidence interval [CI], 0.827-0.884), 0.807 (95% CI, 0.735-0.867) and 0.728 (95% CI, 0.631-0.811) in the training, internal test and external test cohorts, compared to the conventional ResNeXt networks. For experts, the inter-reader agreement was observed in only 437/849 (51.5%) patients with a Kappa value 0.343. And the performance of two experts (AUC, 0.632 to 0.741 vs 0.715 to 0.857) was lower (paired comparison, all p values < 0.05) than that of AI assessment. When expert' interpretations were adjusted by the AI assessments, the performance of both two experts was improved.

**Conclusion:** Our AI tool, showing improved accuracy, offers a promising alternative to human experts for imaging staging of PCa ECE using multiparametric MRI.

## Introduction

Preoperative staging of prostate cancer (PCa) is critical for guiding the treatment selection of patients and preventing both under and over treatment[1]. The presence of extracapsular extension (ECE), that is, T3a stage, accounting for one-third of all PCa patients primarily diagnosed[2, 3], is associated with higher rates of positive surgical margins and early biochemical recurrence after radical prostatectomy[4]. Resection of the neurovascular bundle (NVB) is recommended in these diseases with the aim of decreasing positive surgical margins, which may substantially affect urinary continence and sexual potency[5, 6]. To date, the balance between preserving urinary continence and the achievement of negative margins for radical prostatectomy remains a challenge, preoperatively accurate detection of ECE would thus have a significant impact on treatment planning and prediction of outcomes in patients with PCa.

Historically, digital rectal examination (DRE) has been the principal approach for clinical T-staging of PCa[7]. Clinical staging and risk stratification models combining the prostate specific antigen (PSA) and Gleason score of the prostate biopsy with DRE-derived clinical T-stage have been designed to obtain more

accurate predictions of PCa aggressiveness, disease mortality, and biochemical recurrence[8-10]. However, DRE is generally believed to be a rather subjective test with potential inter-observer variability and is at risk of underestimating the extent of anteriorly located tumors[11]. In the last few decades, multiparametric magnetic resonance imaging (mpMRI) has been widely used to characterize PCa preoperatively and determine the clinical stage. While, the use of MRI instead of DRE leads to a significant upstaging of clinical T-stage and risk grouping[12-15]. In addition, despite considerable efforts such as alternative high-resolution imaging and new grading approaches, the diagnostic accuracy of MRI for T3a-staging revealed a poor and heterogeneous sensitivity of 30%–70%[16-19]. The heterogeneity of MRI in PCa T3a-staging may be caused by the fact that there are no standard criteria for evaluation[20]. , the high level of expertise required for radiologists with the aim of accurate interpretation and interobserver variability limit its consistency and availability[21].

Recently, artificial intelligence (AI), particularly deep learning (DL), has been proposed as a promising solution to many medical imaging tasks involving organ segmentation, lesion detection, and disease classification[22-25]. AI does not rely on predefined representations of low-level visual features within images that were required in early machine learning approaches. Instead, DL can learn to discover task-specific features such as anatomic localization, tumor contacting, neurovascular bundles, or direct evidence of abnormalities in periprostatic adipose tissue, which are the footstones for the imaging detection and staging of PCa. With a sufficient supply of expertly labeled examples, an appropriately designed model can learn to emulate the judgments of expert clinicians who provide the labels.

Therefore, in this study, we hypothesized that an AI-assisted tool trained from a large dataset of high-quality labels would produce automated ECE-staging capable of emulating the diagnostic acumen of a team of experienced radiologists. We further hypothesized that when the assessment of the model is provided to radiologists, their performance in ECE staging of PCa with mpMRI would be improved. We verified our hypothesis by building a ResNeXt-based deep classification and detection model embedded with a spatial attention map of the prior knowledge of the radiologists for an imaging interpretation of ECE in patients with PCa[26]. We then validated the model by comparing it with expert interpretation on two independent cohorts from two tertiary care medical centers with detailed outcome information.

## **Material And Methods**

### **Patients**

This was a retrospective study involving routine care at two tertiary care medical centers. Ethics committee approval was granted by the local institutional ethics review board (protocol 2016-SRFA-093), with a waiver of written informed consent. All procedures conducted in the studies involving human participants were in accord with the 1964 Helsinki Declaration and its later amendments.

The two primary cohorts comprised an evaluation of the local database for the medical records to identify patients with pathologically confirmed PCa. The inclusion criteria were as follows: i) PCa with radical prostatectomy and ii) standard prostatic mpMRI exam within 4 weeks prior to surgical

intervention. Patients without radical prostatectomy or with histories of previous surgeries or adjuvant therapies for PCa (interventions for benign prostatic hyperplasia or bladder outflow obstruction were deemed acceptable) were excluded. Finally, a total of 746 consecutive patients between January 2015 and June 2019 from Center 1 and 103 PCa patients between January 2017 and December 2019 from Center 2 who underwent standard prostate mp-MRI and radical prostatectomy were enrolled. The patient enrollment procedures are summarized in the supplementary data (**Fig. S1**).

Clinical variables included the age, PSA level, PSA density, biopsy Gleason score, number of positive cores and perineural invasion. Histopathological outcomes such as surgical Gleason score, positive surgical margin, presence of histological ECE, and presence of histological seminal vesicle invasion were also determined. All biopsies and surgical specimens were prepared and examined by two pathologists who had 10-yr experience in urologic pathology according to the ISUP 2005 recommendations. Histopathological ECE, referring to the tumor breaking through the prostatic capsule into periprostatic fat, was the primary clinical endpoint of this study.

Patients included in the Center 1 dataset were randomly split into training (n = 596) and test (n = 150) groups for model development and internal validation, respectively. A cohort of 103 patients from Center 2 dataset was used for external validation.

### **Image Acquisition and Analysis**

Patients in two academic institutions underwent a pelvic phased-array prostatic mpMRI examination on a same type of 3.0 T MR scanner (Skyra; Siemens Healthcare, Erlangen, Germany). The scanning protocols are a combination of transverse T1-weighted, transverse, coronal, and sagittal T<sub>2</sub>-weighted imaging (T<sub>2</sub>WI) and transverse DWI sequences. The apparent diffusion coefficient (ADC) was measured using DWI with a mono-exponential fitting model. The scanner types and imaging parameters are summarized in Supplementary Materials (**Table S1**).

All images were retrospectively interpreted based on the guidelines of ESUR by two genitourinary radiologists at two institutions (reader 1, 15 years of experience with prostate MRI; reader 2, 10 years of experience with prostatic MRI) who were blinded to the pathological results and all clinical information. Staging assessment with mpMRI was performed using the ECE grading system introduced by Mehralivand et al.[17]. Imaging diagnosis of ECE is based on a three-tier grading approach using capsular contact length (CCL) of 15 mm or greater, capsular irregularity or bulge, and frank breach of the capsule: i) grade 0, no suspicion of pathological ECE, ii) grade 1, either CCL of 15 mm or greater or capsular irregularity or bulge, iii) grade 2, both CCL of 15 mm or greater and capsular irregularity or bulge, and iv) grade 3, frank ECE visible at mpMRI.

### **Construction of Deep Learning Networks**

**Image annotation and preprocess:** Segmentation of prostate and PCa was performed with an in-house software (Oncology Imaging Analysis version 2; Shanghai Key Laboratory of MR, ECNU, Shanghai, China)

by two experienced genitourinary radiologists. A prior attention was generated according to the attention of the prostate and PCa. Diffusion related sequences were aligned onto T2WI and all images were resampled to an inner-resolution of  $0.5 \times 0.5 \text{ mm}^2$ . Then the patch with a size of  $200 \times 200$  were cropped and normalized by Z-score to make the scale similar before importing into the model .. The detail of image annotation and preprocess were described in **Supplementary Section 1-2**.

**Architecture of Network:** A two-denominational ResNeXt, which was proved to be an effective CNN model, with a convolutional block attention module (CBAM) was used to analyze the mpMRI images with labels provided by the concatenated use of high-resolution T2WI, high-b value ( $1500 \text{ s}^2/\text{mm}$ ) DWI, and ADC[26]. The output of the model was the prediction of the ECE. In each training dataset, a single leading slice image with the largest cross-section of the tumor was used for model development. To guide the ResNeXt network to emulate the judgments of experts who provided the labels of the targeted lesion, we introduced a prior-attention guide (PAGNet) unit by inputting the attention map into CBAM[27]. The attention map was generated based on the annotations of the whole prostate and tumor lesion, and a high computational value in the attention map denoted a deserved-focusing region. Ensemble learning with 5-fold cross validation was used during the training stage, and in the inference stage, the average prediction of five independent models was treated as the final prediction of ECE risk. Details of attention map generation, network architecture, and analysis are described in **Supplementary Sections 3–6 and Fig. 1**.

**Postprocess:** Considering that, for each patient, the tumor can involve several imaging slices while the ECE may involve only parts of the involved imaging slices, we thus proposed two analysis approaches to postprocess the predicted outputs. One is a single-slice (SS) based prediction that is derived from a preset leading-slice image. The other is multi-slice (MS) based prediction, which is derived from images with entire tumoral coverage, among which the highest predicted result was used as the final MS prediction.

### **Integration of PAGNet and Clinical Identifications**

Finally, we evaluated the integrative effects of clinical factors on DL networks to improve the diagnostic performance. The PSA, age, biopsy Gleason score, percentage of positive cores, and biopsy perineural invasion were added to the PAGNet model, namely, PAGNet+C, in which clinical information was directly added to the penultimate layer of the fully connected (FC) layer of PAGNet by increasing the number of neurons.

### **Performance of Deep Diagnostic Model**

To evaluate the performance and clinical applicability of the deep diagnostic model, all data assessments were conducted independently based on AI, human experts, and expert-AI interaction. For expert-AI interaction, the expert score is upgraded when a positive assessment by the AI was determined, whereas the highest score of 3 remained unchanged even with a positive assessment by the AI. Conversely, the expert score is downgraded if a negative assessment is determined by the AI, and the

lowest score of 0 remains unchanged with negative findings by the AI. To assess the effect of pathological variants on the performance, the assessments were conducted specifically in groups stratified by lesion size, D'Amico risk group[9], and PI-RADS score[28].

## Statistical Analysis

Inter-reader variability was evaluated using inter-reader agreement and Cohen's kappa. Model performance was typically evaluated against a "ground truth" with histopathological manifestations using a receiver operating characteristic (ROC) analysis. An inter-method comparison between expert, AI, and expert-AI interaction was applied using a summary ROC (SROC) curve through a Bayesian meta-analysis, which allows an assessment of the independent and pooled performance of all methods. For each comparison, contingency tables were used to present the results and calculate the diagnostic accuracy. The unit of assessment for the contingency table for the assessment of accuracy was one patient. Performance characteristics such as the area under the ROC curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy were also reported.

Second, the clinical usefulness and net benefits of the models were assessed using a decision curve analysis (DCA). The DCA estimates the net benefit of a model based on the difference between the numbers of true positives and false positives, weighted by the odds of the selected threshold probability of risk. SROC was estimated using Stata 15, DCA was estimated using R, and other statistical values were estimated using Python with scipy (v1.4.1) and the scikit-learn package (v.0.22). The reported statistical significance levels were all two-sided, with statistical significance set at 0.05.

## Results

### Baseline characteristics

Of all patients included, histopathological ECE was diagnosed in the explanted tissue of 151/596 patients (25.3%) in the training group, 40/150 (26.7%) in the internal validation group, and 33/103 (32.0%) in the external validation group. The details of the baseline characteristics are summarized in **Table 1**. Preoperative PSA, percentage of positive cores, biopsy Gleason Score, and perineural invasion differed significantly between the groups with and without ECE (all  $p$ -values < 0.001). Age was not significantly different between the two groups ( $p = 0.964$ ).

### Comparison of Deep Network Models

To determine the impacts of the two post-processing approaches on the prediction, the performance of SS-based versus MS-based assessments is shown in **Fig. 2a**. In the training group, all SS-based assessments were superior to the corresponding MS-based assessments (all  $p$  values < 0.028). In the internal validation group, SS-PAGNet and SS-PAGNet+C were superior to the corresponding MS-based networks (all  $p$  values < 0.019), whereas MS-ResNeXt was superior to SS-ResNeXt ( $p = 0.017$ ). In the external validation group, SS-ResNeXt and SS-PAGNet were superior to the corresponding MS networks

(all  $p$  values < 0.046), and MS-PAGNet+C was superior to the corresponding SS network ( $p = 0.012$ ). In internal and external validation, the SS-PAGNet, requiring minimal post-processing, advanced with the attention-gated units and achieved a better performance than any other SS or MS-based networks and was therefore selected as the index model for clinical application.

To illustrate the robustness of the index PAGNet model, deep generative features from the last penultimate layer of each network were extracted and plotted based on a t-distributed stochastic neighbor embedding (t-SNE) analysis (**Fig. 2b**). The features of index PAGNet showed better intraclass aggregation and interclass separation than those of the ResNeXt networks. To illustrate the interpretability, the model was visualized by a Gradient-weighted Class Activation Mapping (Grad-CAM), which provided an activation map at the end of block 4. The high activation region of the visual Grad-CAM map was the major contributor to the prediction. To highlight the advancement of interpretable PAGNet, two representative clinical cases are shown in **Fig. 2c**.

### **Performance and clinical application of deep learning models**

The performances of ECE prediction between AI, expert, and expert-AI interaction are summarized in a confusion matrix, as shown in **Fig. 3**, in which the true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs) of each diagnostic approach are compared. Regarding the expert-based approach for ECE interpretation, the inter-reader agreement for ECE staging was observed in 437/849 (51.5%) observations, with a Kappa value of 0.343. And the performance of both two experts (AUC, 0.632 to 0.741 vs 0.715 to 0.857) is significantly (paired comparison of ROC, all  $p$ -value < 0.05) lower than that of any of the AI assessments in the three cohorts. When expert-AI interaction was performed, by which the expert's interpretation was modified by AI assessment, the performance of the two experts was significantly improved. To provide a more complete picture to illustrate the assistant role of AI to experts, the independent and integrated effects of the experts, AI, and expert-AI interaction were evaluated using SROC curves and forest plots with a Bayesian meta-analysis (**Fig. 4**).

### **Clinical implications of Deep Network Models**

The benefit derived from applying the index PAGNet model in clinical practice, according to the decision curve method, is depicted in **Fig. S2**. The PAGNet-derived probability demonstrated an improved clinical risk prediction against threshold probabilities of  $ECE \leq 60\%$ . The graph demonstrated better clinical risk prediction when using the PAGNet or expert-AI interaction approach as compared to expert-based grading.

Additionally, we evaluated the clinical application of PAGNet for ECE staging in different clinicopathologic manifestations such as the tumor size, D'Amico risk group, and PI-RADS score, a pretreatment stratification of which might have a significant impact on the clinical decision making. Compared to experts and other DL models, the index PAGNet model showed a higher NPV in terms of tumor size < 1.5 cm low D'Amico risk, and PI-RADS 3 lesions, and showed a higher PPV in terms of tumor size  $\geq 1.5$  cm, intermediate/high D'Amico risk, and PI-RADS 4–5 lesions (**Fig. 5**).

## Discussion

Tailoring the most suitable surgical approach for patients with PCa in terms of nerve sparing, urologists are required to balance the risk of ECE versus benefits from NVB preservation before the RP is delivered. Expert-based assessment of the ECE stage using DRE and MRI is highly heterogeneous[16]. Recently, AI-assisted diagnosis of prostate diseases using mpMRI has attracted increasing attention and has shown promising prospects[24, 29]. In this study, we developed and validated an AI-assisted tool to preoperatively assess ECE stage of localized PCa using mpMRI. Our study contributes important methodology accompanied with model interpretability to address a critical question for clinical tumor staging of PCa. Our results on a cohort of 849 patients with RP from two tertiary care medical centers show promises of deep diagnostic model for ECE staging and potential utilities of this tool for improving performance and reducing inter-reader variability.

Our study has several innovations compared with previous relevant researches. First, to our knowledge, this is the first study to apply an automatic AI tool in ECE staging in patients with PCa. The view was expressed that imaging detection of ECE remains a challenge for that we are unable to detect microscopic ECE at histopathology and DL algorithm can provide potential improvement through training high-throughput-derived imaging features[30]. Our results revealed that our AI tool is capable of discriminating ECE in a quantitative and objective manner, and performs better than binary mpMRI interpretation or objective scoring scheme[31-33]. Second, in our approach, the proposed model generated a prior-attention probability map by gating the networks to learn potentially useful features across the boundaries between the prostate and tumor, thus making our model more robust and interpretable compared to the traditional black-box learning approaches. Third, from the perspective of Grad-CAM, our tool can not only make the diagnosis of the ECE stage, but also provide a predicted region that is highly suspicious of ECE. This is a significant advancement compared to traditional predictive nomograms that only provide binary classification or prediction [2, 3, 34]. This approach might be more applicable to radiologists in real-world clinical scenarios.

Our results have several clinical implications. First, from a clinical perspective, ECE most likely occurs in the pericapsular regions of the leading imaging slice, the accuracy of which needs to be carefully clarified. Taking this into account, two analysis strategies, i.e., SS versus MS, were proposed to optimize the predictions of our networks. The results revealed that the SS-based analysis performed better than the corresponding MS-based approach. This implies that, to a certain extent, an overprediction occurs based on the MS analysis, which leads to a false-positive prediction in non-leading imaging slices. This finding is partly consistent with our primary assumption that the pericapsular region on the leading imaging slice is the most suspicious location for the occurrence of ECE. Focusing the attention learning on the leading slice can provide a more effective assessment of ECE staging. Second, the results of previous studies have revealed the critical roles of clinical characteristics such as PSA, PSAD, and biopsy findings for ECE prediction. Unfortunately, adding these factors did not contribute significantly to the improvement in the model performance. This is contrary to most other studies, supposing that an improved diagnostic accuracy of combining MRI with clinical indications [35, 36]. This may be caused by the fact that the



hidden FC features of our networks are significantly larger than embedded clinical factors, and data in a training cohort are relative smaller than the deep layer features. Therefore, it is difficult for deep networks to extract critical information from these limited sparse clinical features, that is, the curse of dimensionality. Third, although the expert-based ECE grading system used in our study has potential advances against the traditional non-standardized reporting method [19], we did observe large inter-observer variances in ECE grade interpretation. The intrareader agreement is fair in our two cohorts, which varies significantly from that of Park et al.[19]. The positive rates in each interpreted ECE grade in our cohorts are comparable to that of Mehralivand et al.[17] but markedly lower than that of Park et al[19]. In addition, we conduct a head-to-head comparison of performance between AI, expert, and expert-AI interaction. We did demonstrate that an AI-based assessment has a higher accuracy compared to expert-based ECE grading, and results from expert-AI interaction show that our AI could be of great assistance to radiologists in improving the diagnostic performance. Fourth, we further elaborated a subgroup analysis of PAGNet, the results of which support the idea that our AI achieved higher PPVs in tumor with larger size, higher D'Amico risk score, and higher PI-RADS score. Therefore, personalized surgical treatment of patients with PCa is feasible when MRI-derived ECE risk and other risk-based approaches are combined.

Although encouraging results were obtained in our preliminary work, several limitations warrant mention. First, the DL model was trained on single-center data, and although the test data originated from two medical centers, the cohort size was still limited for our data-driven approach. In addition, in the external test cohort, the diagnostic performance of any AI-based method decreased markedly compared with that in the training and internal test cohorts. Currently, the wish of multi-center application of AI-based approaches may be challenged by the sample size, study cohorts, and distribution differences. This may be overcome by increasing the data samples from external sites, which is one of our ongoing works. Second, a prospective multicenter controlled experiment is needed to validate the model in clinical scenarios before it can be made routinely available.

In conclusion, we proposed an AI-assisted tool embedded with a spatial attention map of the experts' prior-knowledge for ECE staging using mpMRI. The tool performed better than expert-based interpretation and provided assistant role to radiologists. The interpretability of our AI-based approach is particularly imperative towards building trustable auto classification and detection tool for clinical application and facilitating a streamlined patients management process.

## Declarations

### Funding information:

Contract grant sponsor: Key research and development program of Jiangsu Province; contract grant number: BE2017756 (to Y.D.Z.). The key Project of National Natural Science Foundation of China; contract grant number: 61731009 (to G. Y.). Open Project from Shanghai Key Laboratory of Magnetic Resonance; Contract grant number: N2019001 (to G. Y.)

### Conflict of interest:

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

### **Availability of data and material:**

The imaging studies and clinical data used for algorithm development are not publicly available, because they contain private patient health information. Interested users may request access to these data, where institutional approvals along with signed data use agreements and/or material transfer agreements may be needed/negotiated. Derived result data supporting the findings of this study are available upon reasonable requests.

### **Authors' contributions**

Y.D.Z. and Y.S. conceived, designed and supervised the project; Y.H., Y.H.Z, J.B., M.B. H.S. and G.Y. collected and pre-processed all data and performed the research; Y.H. and Y.H.Z performed imaging data annotation and clinical data review; Y.D.Z. and Y.S. proposed the model; Y.H. and Y.H.Z drafted the paper; all authors reviewed, edited and approved the final version of article.

### **Ethics approval and Consent to participate:**

This study was retrospective and 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.

### **Consent for publication:**

Not applicable

### **Correspondence to:**

Yu-Dong Zhang, MD, Ph.D.

Department of Radiology, the First Affiliated Hospital with Nanjing Medical University, No. 300, Guangzhou Road, Nanjing, Jiangsu Province, China, 210029

Tel: +86 158-0515-1704

E-mail: njmu\_zyd@163.com

Yang Song, Ph.D.

Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai, China. 3663 N. Zhongshan Rd., Shanghai, China 200062

Tel: +86-021-62233873

E-mail: [songyangmri@gmail.com](mailto:songyangmri@gmail.com)

## References

1. Mottet N, van den Bergh RCN, Briers E, Van den Broeck T, Cumberbatch MG, De Santis M, et al. EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer-2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *European urology*. 2020. doi:10.1016/j.eururo.2020.09.042.
2. Gandaglia G, Ploussard G, Valerio M, Mattei A, Fiori C, Roumiguié M, et al. The Key Combined Value of Multiparametric Magnetic Resonance Imaging, and Magnetic Resonance Imaging-targeted and Concomitant Systematic Biopsies for the Prediction of Adverse Pathological Features in Prostate Cancer Patients Undergoing Radical Prostatectomy. *European urology*. 2020;77:733-41. doi:10.1016/j.eururo.2019.09.005.
3. Diamand R, Ploussard G, Roumiguié M, Oderda M, Benamran D, Fiard G, et al. External Validation of a Multiparametric Magnetic Resonance Imaging-based Nomogram for the Prediction of Extracapsular Extension and Seminal Vesicle Invasion in Prostate Cancer Patients Undergoing Radical Prostatectomy. *European urology*. 2020. doi:10.1016/j.eururo.2020.09.037.
4. Jeong BC, Chalfin HJ, Lee SB, Feng Z, Epstein JI, Trock BJ, et al. The relationship between the extent of extraprostatic extension and survival following radical prostatectomy. *European urology*. 2015;67:342-6. doi:10.1016/j.eururo.2014.06.015.
5. Walz J, Epstein JI, Ganzer R, Graefen M, Guazzoni G, Kaouk J, et al. A Critical Analysis of the Current Knowledge of Surgical Anatomy of the Prostate Related to Optimisation of Cancer Control and Preservation of Continence and Erection in Candidates for Radical Prostatectomy: An Update. *European urology*. 2016;70:301-11. doi:10.1016/j.eururo.2016.01.026.
6. Nguyen LN, Head L, Witiuk K, Punjani N, Mallick R, Clossen S, et al. The Risks and Benefits of Cavernous Neurovascular Bundle Sparing during Radical Prostatectomy: A Systematic Review and Meta-Analysis. *The Journal of urology*. 2017;198:760-9. doi:10.1016/j.juro.2017.02.3344.
7. Borkenhagen JF, Eastwood D, Kilari D, See WA, Van Wickle JD, Lawton CA, et al. Digital Rectal Examination Remains a Key Prognostic Tool for Prostate Cancer: A National Cancer Database Review. *Journal of the National Comprehensive Cancer Network : JNCCN*. 2019;17:829-37. doi:10.6004/jnccn.2018.7278.
8. Graefen M. [Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer—a multi-institutional update]. *Aktuelle Urologie*. 2004;35:377-8. doi:10.1055/s-2004-834369.
9. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *Jama*. 1998;280:969-74. doi:10.1001/jama.280.11.969.

10. Eifler JB, Feng Z, Lin BM, Partin MT, Humphreys EB, Han M, et al. An updated prostate cancer staging nomogram (Partin tables) based on cases from 2006 to 2011. *BJU international*. 2013;111:22-9. doi:10.1111/j.1464-410X.2012.11324.x.
11. Gosselaar C, Kranse R, Roobol MJ, Roemeling S, Schröder FH. The interobserver variability of digital rectal examination in a large randomized trial for the screening of prostate cancer. *The Prostate*. 2008;68:985-93. doi:10.1002/pros.20759.
12. Kam J, Yuminaga Y, Koschel S, Aluwihare K, Sutherland T, Skinner S, et al. Evaluation of the accuracy of multiparametric MRI for predicting prostate cancer pathology and tumour staging in the real world: an multicentre study. 2019;124:297-301. doi:10.1111/bju.14696.
13. Muglia VF, Westphalen AC, Wang ZJ, Kurhanewicz J, Carroll PR, Coakley FV. Endorectal MRI of prostate cancer: incremental prognostic importance of gross locally advanced disease. *AJR American journal of roentgenology*. 2011;197:1369-74. doi:10.2214/ajr.11.6425.
14. 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. *European urology*. 2017;71:701-4. doi:10.1016/j.eururo.2016.08.015.
15. Soeterik TFW, van Melick HHE, Dijkstra LM, Biesma DH, Witjes JA, van Basten JA. Multiparametric Magnetic Resonance Imaging Should Be Preferred Over Digital Rectal Examination for Prostate Cancer Local Staging and Disease Risk Classification. *Urology*. 2020. doi:10.1016/j.urology.2020.08.089.
16. de Rooij M, Hamoen EH, Witjes JA, Barentsz JO, Rovers MM. Accuracy of Magnetic Resonance Imaging for Local Staging of Prostate Cancer: A Diagnostic Meta-analysis. *European urology*. 2016;70:233-45. doi:10.1016/j.eururo.2015.07.029.
17. Mehravand S, Shih JH, Harmon S. A Grading System for the Assessment of Risk of Extraprostatic Extension of Prostate Cancer at Multiparametric MRI. 2019;290:709-19. doi:10.1148/radiol.2018181278.
18. Boesen L, Chabanova E, Løgager V, Balslev I, Mikines K, Thomsen HS. Prostate cancer staging with extracapsular extension risk scoring using multiparametric MRI: a correlation with histopathology. *European radiology*. 2015;25:1776-85. doi:10.1007/s00330-014-3543-9.
19. Park KJ, Kim MH, Kim JK. Extraprostatic Tumor Extension: Comparison of Preoperative Multiparametric MRI Criteria and Histopathologic Correlation after Radical Prostatectomy. 2020;296:87-95. doi:10.1148/radiol.2020192133.
20. Eberhardt SC. Local Staging of Prostate Cancer with MRI: A Need for Standardization. *Radiology*. 2019;290:720-1. doi:10.1148/radiol.2019182943.
21. Fütterer JJ, Engelbrecht MR, Huisman HJ, Jager GJ, Hulsbergen-van De Kaa CA, Witjes JA, et al. Staging prostate cancer with dynamic contrast-enhanced endorectal MR imaging prior to radical prostatectomy: experienced versus less experienced readers. *Radiology*. 2005;237:541-9. doi:10.1148/radiol.2372041724.

22. Hesamian MH, Jia W, He X, Kennedy P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of digital imaging*. 2019;32:582-96. doi:10.1007/s10278-019-00227-x.
23. Zhang M, Young GS, Chen H, Li J, Qin L, McFaline-Figueroa JR, et al. Deep-Learning Detection of Cancer Metastases to the Brain on MRI. 2020;52:1227-36. doi:10.1002/jmri.27129.
24. Schelb P, Kohl S. Classification of Cancer at Prostate MRI: Deep Learning versus Clinical PI-RADS Assessment. 2019;293:607-17. doi:10.1148/radiol.2019190938.
25. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep Learning in Medical Image Analysis. *Advances in experimental medicine and biology*. 2020;1213:3-21. doi:10.1007/978-3-030-33128-3\_1.
26. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated Residual Transformations for Deep Neural Networks. 2016.
27. Woo S, Park J, Lee J-Y, Kweon IS. CBAM: Convolutional Block Attention Module. 2018.
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. *European urology*. 2019;76:340-51. doi:10.1016/j.eururo.2019.02.033.
29. Alkadi R, Taher F, El-Baz A, Werghi N. A Deep Learning-Based Approach for the Detection and Localization of Prostate Cancer in T2 Magnetic Resonance Images. *Journal of digital imaging*. 2019;32:793-807. doi:10.1007/s10278-018-0160-1.
30. Padhani AR, Petralia G. Radiologists Should Integrate Clinical Risk Factors with MRI Findings for Meaningful Prostate Cancer Staging. 2020;296:96-7. doi:10.1148/radiol.2020201082.
31. Zanelli E, Giannarini G, Cereser L, Zuiani C, Como G, Pizzolitto S, et al. Head-to-head comparison between multiparametric MRI, the partin tables, memorial sloan kettering cancer center nomogram, and CAPRA score in predicting extraprostatic cancer in patients undergoing radical prostatectomy. *Journal of magnetic resonance imaging : JMRI*. 2019;50:1604-13. doi:10.1002/jmri.26743.
32. Kim W, Kim CK. Evaluation of extracapsular extension in prostate cancer using qualitative and quantitative multiparametric MRI. 2017;45:1760-70. doi:10.1002/jmri.25515.
33. Muehlematter UJ, Burger IA, Becker AS, Schawkat K, Hötter AM, Reiner CS, et al. Diagnostic Accuracy of Multiparametric MRI versus (68)Ga-PSMA-11 PET/MRI for Extracapsular Extension and Seminal Vesicle Invasion in Patients with Prostate Cancer. 2019;293:350-8. doi:10.1148/radiol.2019190687.
34. Tosco L, De Coster G, Roumeguère T, Everaerts W, Quackels T, Dekuyper P, et al. Development and External Validation of Nomograms To Predict Adverse Pathological Characteristics After Robotic Prostatectomy: Results of a Prospective, Multi-institutional, Nationwide series. *European urology oncology*. 2018;1:338-45. doi:10.1016/j.euo.2018.04.008.
35. Rayn KN, Bloom JB, Gold SA, Hale GR, Baiocco JA, Mehravand S, et al. Added Value of Multiparametric Magnetic Resonance Imaging to Clinical Nomograms for Predicting Adverse Pathology in Prostate Cancer. *The Journal of urology*. 2018;200:1041-7. doi:10.1016/j.juro.2018.05.094.

36. Chen Y, Yu W, Fan Y, Zhou L, Yang Y, Wang H, et al. Development and comparison of a Chinese nomogram adding multi-parametric MRI information for predicting extracapsular extension of prostate cancer. *Oncotarget*. 2017;8:22095-103. doi:10.18632/oncotarget.11559.

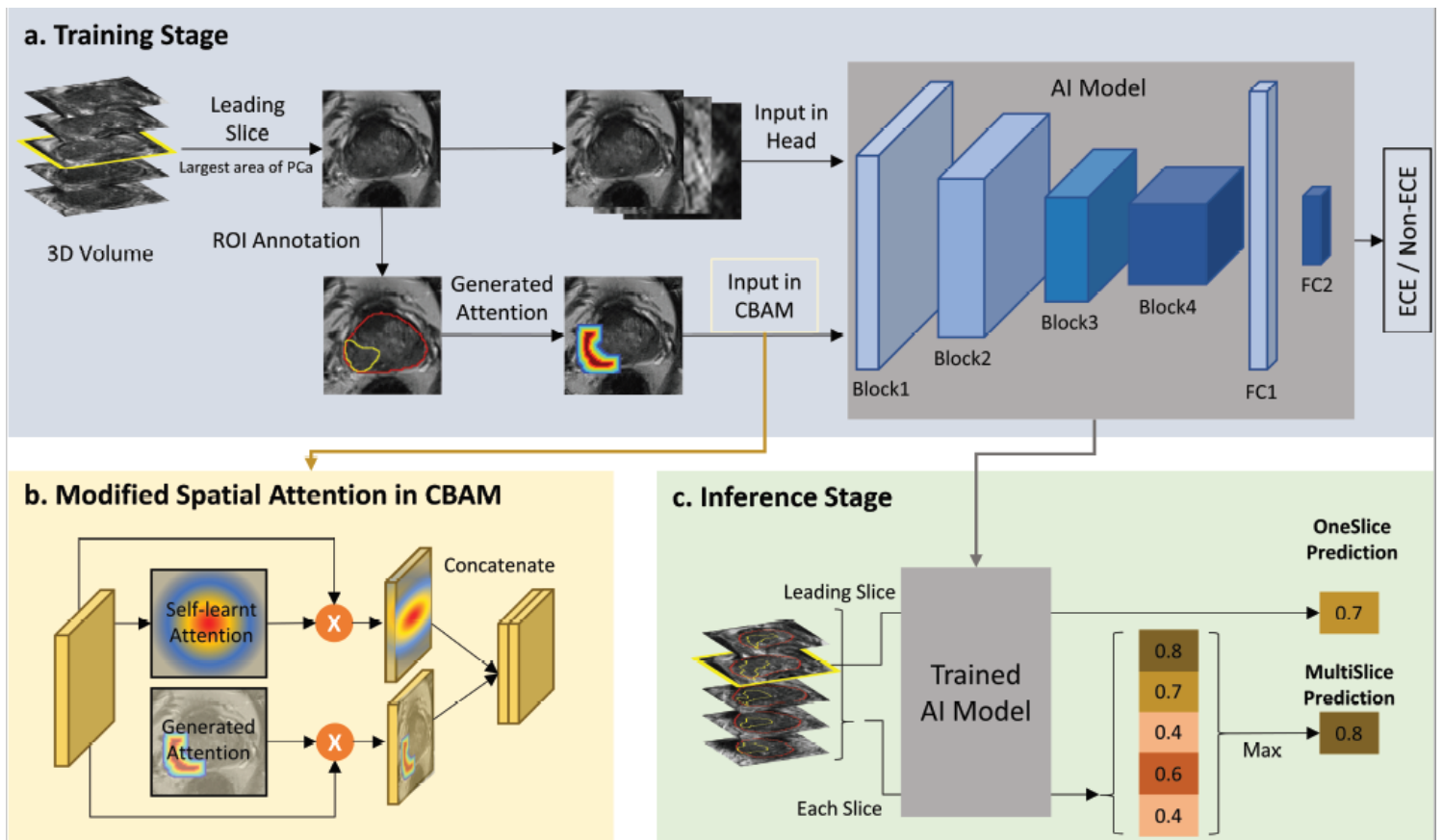
## Table

<b>Table 1: The baseline characteristics of the patients in the training, internal and external validation sets</b>			
<b>Variable</b>	<b>Training (n = 596)</b>	<b>Internal validation (n = 150)</b>	<b>External validation (n = 103)</b>
<b>Age (y)</b>	69.2±7.1 (42-86)	69.2±6.9 (48-83)	70.2±6.8 (52-87)
<b>PSA (ng/mL)</b>	28.2±47.4 [0.7-676.0]	30.7±36.5 (0.8-214.4)	31.4±35.5 (3.0-201.4)
<b>D'Amico risk group</b>			
Low risk	77/596 (12.9%)	17/150 (11.3%)	6/103 (5.8%)
Intermediate risk	227/596 (38.1%)	57/150 (38.0%)	33/103 (32.0%)
High risk	292/596 (49.0%)	76/150 (50.7%)	64/103 (62.1%)
<b>Tumor diameter</b>	1.8±1.1 (0.4-6.3)	1.8±1.1 (0.5-5.4)	2.2±1.0 (0.5-5.4)
<b>PI-RADS score</b>			
1-2	40/596 (6.7%)	12/150 (8.0%)	12/103 (11.7%)
3	89/596 (14.9%)	24/150 (16.0%)	8/103 (7.8%)
4	191/596 (32.0%)	47/150 (31.3%)	14/103 (13.6%)
5	276/596 (46.3%)	67/150 (44.7%)	69/103 (67.0%)
<b>MRI-based ECE grade</b>			
<b>Reader 1   reader 2</b>			
0	196/596 (32.9%)   215/596 (36.1%)	50/150 (33.3%)   48/150 (32.0%)	28/103 (27.2%)   25/103 (24.3%)
1	150/596 (25.2%)   179/596 (30.0%)	31/150 (20.7%)   50/150 (33.3%)	29/103 (28.2%)   31/103 (30.1%)
2	130/596 (21.8%)   103/596 (17.3%)	34/150 (22.7%)   23/150 (15.3%)	26/103 (25.2%)   26/103 (25.2%)
3	120/596 (20.1%)   99/596 (16.6%)	35/150 (23.3%)   29/150 (19.3%)	20/103 (19.4%)   21/103 (20.4%)
<b>Biopsy Gleason Score</b>			
GS 3+3	151/596 (25.3%)	41/150 (27.3%)	12/103 (11.7%)
GS 3+4	128/596 (21.5%)	34/150 (22.7%)	19/103 (18.4%)
GS 4+3	149/596 (25.0%)	43/150 (28.7%)	26/103 (25.2%)
GS ≥ 4+4	168/596 (28.2%)	32/150 (21.3%)	46/103 (44.7%)

<b>Percentage of positive cores</b>	0.4±0.3 (0.1-1.0)	0.4±0.3 (0.1-1.0)	0.5±0.3 (0.1-1.0)
<b>Perineural invasion</b>			
present	92/596 (15.4%)	17/150 (11.3%)	9/103 (8.7%)
absent	504/596 (84.6%)	133/150 (88.7%)	94/103 (91.3%)
<b>Surgical Gleason Score</b>			
GS 3+3	85/596 (14.3%)	20/150 (13.3%)	6/103 (5.8%)
GS 3+4	167/596 (28.0%)	41/150 (27.3%)	21/103 (20.4%)
GS 4+3	180/596 (30.2%)	48/150 (32.0%)	34/103 (33.0%)
GS ≥ 4+4	164/596 (27.5%)	41/150 (27.3%)	42/103 (40.7%)
<b>Pathological ECE</b>			
present	151/596 (25.3)	40/150 (26.7%)	33/103 (32.0%)
absent	445/596 (74.7)	110/150 (73.3%)	70/103 (68.0%)
<b>Pathological SVI</b>			
present	101/596 (16.9%)	26/150 (17.3%)	15/103 (14.6%)
absent	495/596 (83.1%)	124/150 (82.7%)	88/103 (85.4%)
<b>Pathological SM</b>			
present	265/596 (44.5%)	66/150 (44.0%)	30/103 (29.1%)
absent	331/596 (55.5%)	84/150 (56.0%)	73/103 (70.9%)
<b>Note:</b> Unless indicated otherwise, data are number of tumors, with percentages in parentheses. PSA = prostate serum antigen. PI-RADS= Prostate Imaging and Reporting and Data System version 2.1; ECE = extracapsular extension. SVI = seminal vesicle invasion. GS= Gleason Score. RP=radical prostatectomy. SM=surgical margin.			

## Figures





**Figure 1**

Overview of the model development. (a) Training stage: single leading-slice image, determined with the largest area of the tumor on T2WI, DWI, and ADC, respectively, formed as the input sets of deep networks. A prior-attention map, indicating the boundary features of the prostate and tumor with expert-provided labels, formed as the principle components of CBAM. (b) Modified CBAM: self-learned attention of the CBAM and the prior-attention map, serving as additional attention-gated factor, are implemented into the deep layers of backbone networks. (c) Inference Stage: Innovatively, for each patient, our deep network model outputs a single-slice prediction and multi-slice predictions. The output of a single-slice prediction is derived from the preset leading-slice image, and the final output of multi-slice predictions is derived by computing the highest probability of all tumor-involved images. ROI, region of interest; CBAM, convolutional block attention module; AI, artificial intelligence; ECE, extracapsular extension

Training Data				
Variable	AUC	SE	95% CIs	P
ResNeXt-OneSlice	0.818	0.020	0.779 to 0.856	0.019
ResNeXt-MultiSlice	0.799	0.020	0.760 to 0.839	
Variable	AUC	SE	95% CIs	P
PAGNet-OneSlice	0.857	0.017	0.823 to 0.891	0.028
PAGNet-MultiSlice	0.829	0.019	0.793 to 0.865	
Variable	AUC	SE	95% CIs	P
PAGNet+C-OneSlice	0.853	0.018	0.819 to 0.887	0.007
PAGNet+C-MultiSlice	0.846	0.018	0.811 to 0.880	

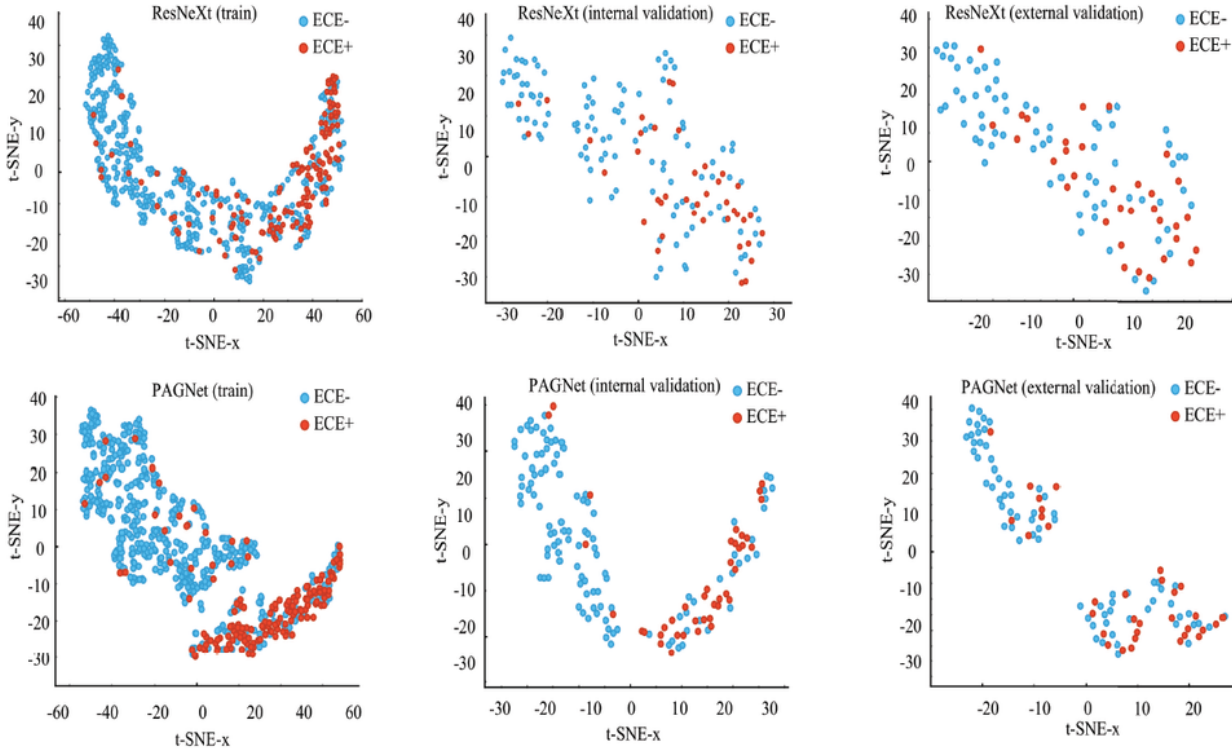
  

Internal Validation Data				
Variable	AUC	SE	95% CIs	P
ResNeXt-OneSlice	0.746	0.045	0.658 to 0.833	0.017
ResNeXt-MultiSlice	0.762	0.044	0.676 to 0.849	
Variable	AUC	SE	95% CIs	P
PAGNet-OneSlice	0.807	0.038	0.732 to 0.882	0.003
PAGNet-MultiSlice	0.805	0.038	0.731 to 0.878	
Variable	AUC	SE	95% CIs	P
PAGNet+C-OneSlice	0.794	0.040	0.715 to 0.872	0.019
PAGNet+C-MultiSlice	0.775	0.041	0.694 to 0.855	

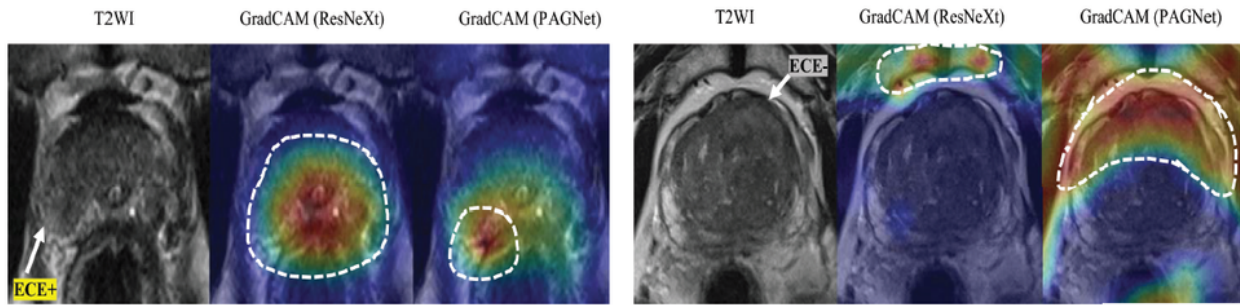
  

External Validation Data				
Variable	AUC	SE	95% CIs	P
ResNeXt-OneSlice	0.723	0.054	0.618 to 0.828	0.046
ResNeXt-MultiSlice	0.677	0.060	0.559 to 0.796	
Variable	AUC	SE	95% CIs	P
PAGNet-OneSlice	0.728	0.052	0.626 to 0.830	0.025
PAGNet-MultiSlice	0.703	0.060	0.586 to 0.820	
Variable	AUC	SE	95% CIs	P
PAGNet+C-OneSlice	0.715	0.053	0.613 to 0.818	0.012
PAGNet+C-MultiSlice	0.727	0.055	0.619 to 0.836	

a.



b.



c.

**Figure 2**

Graphs show (a) the diagnostic performance of SS and MS ResNeXt, PAGNet, and PAGNet+C, (b) t-SNE of the features from the penultimate layer of ResNeXt and PAGNet, and (c) the Grad-CAM of ResNeXt and PAGNet. (a) SS-PAGNet achieved a better performance than any other SS or MS network; the yellow and green colors highlight the best performance of the dedicated methods. (b) The features of PAGNet showed better intra-class aggregation and inter-class separation than that of ResNeXt. (c) The highly

activated region (dotted lines) on the Grad-CAM map represents the attention-focused region. For the ECE-present case (left), PAGNet gates the learning on the right peripheral zone of the prostate, which is the region that undergoes histopathological ECE, whereas ResNeXt focuses on the entire-gland regions. For the ECE-absent case (right), PAGNet gates attention learning on the anterior peri-capsular region, which is the most suspicious region for ECE, whereas ResNeXt gates attention to the region outside the prostate. To some extent, this proves the effectiveness of the expert-gated attention units on the maturity of the model. ECE, extracapsular extension; SS, single slice; MS, multi slice; AUC, area under the curve; SE, standard error; CI, Confidence interval; t-SNE, t-distributed stochastic neighbor embedding; Grad-CAM, gradient-weighted class activation mapping.

Diagnosis	Cut-off	Training					
		AUC	TP	FP	TN	FN	
ResNeXt	0.835	0.818 (0.779-0.856)	112/199 (56.3%)	87/199 (43.7%)	358/397 (90.2%)	39/397 (9.8%)	
PAGNet	0.8	<b>0.857 (0.823-0.891)</b>	<b>108/163 (66.3%)</b>	<b>55/163 (33.7%)</b>	390/433 (90.1%)	43/433 (9.9%)	
PAGNet+C	0.504	0.853 (0.819-0.887)	126/223 (56.5%)	97/223 (43.5%)	<b>348/373 (93.3%)</b>	<b>25/373 (6.7%)</b>	
Reader 1	>=2	0.741 (0.697-0.786)	104/250 (41.6%)	146/250 (58.4%)	299/346 (86.4%)	47/346 (13.6%)	
Reader 2	>=2	0.674 (0.625-0.722)	82/202 (40.6%)	120/202 (59.4%)	325/394 (82.5%)	69/394 (17.5%)	
PAGNet+Reader 1	>=2	0.809 (0.768-0.850)	108/187 (57.8%)	79/187 (42.2%)	366/409 (89.5%)	43/409 (10.5%)	
PAGNet+Reader 2	>=1	0.795 (0.754-0.836)	123/262 (46.9%)	139/262 (53.1%)	306/334 (91.6%)	28/334 (8.4%)	

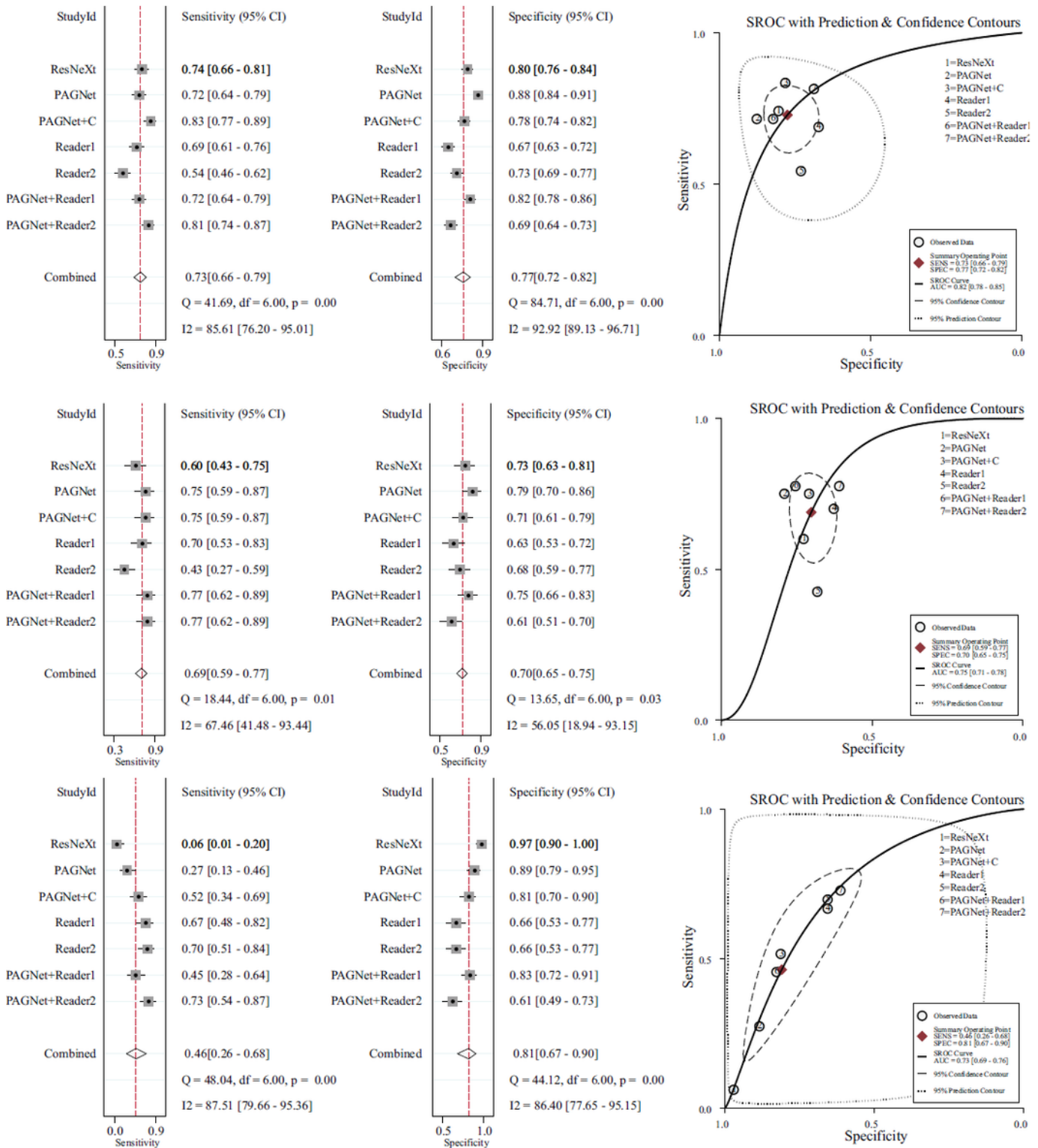
Diagnosis	Cut-off	Internal Validation					
		AUC	TP	FP	TN	FN	
ResNeXt	0.835	0.746 (0.658-0.833)	24/54 (44.4%)	30/54 (55.6%)	80/96 (83.3%)	16/96 (16.7%)	
PAGNet	0.8	<b>0.807 (0.732-0.882)</b>	<b>30/53 (56.6%)</b>	<b>23/53 (43.4%)</b>	87/97 (89.7%)	10/97 (10.3%)	
PAGNet+C	0.504	0.794 (0.715-0.872)	30/62 (48.4%)	32/62 (51.6%)	78/88 (88.6%)	10/88 (11.4%)	
Reader 1	>=2	0.725 (0.638-0.811)	28/69 (40.6%)	41/69 (59.4%)	69/81 (85.2%)	12/81 (14.8%)	
Reader 2	>=2	0.632 (0.541-0.723)	17/52 (32.7%)	35/52 (67.3%)	75/98 (76.5%)	23/98 (23.5%)	
PAGNet+Reader 1	>=2	0.798 (0.719-0.877)	31/58 (53.4%)	27/58 (46.6%)	<b>83/92 (90.2%)</b>	<b>9/92 (9.8%)</b>	
PAGNet+Reader 2	>=1	0.733 (0.645-0.822)	31/74 (41.9%)	43/74 (58.1%)	67/76 (88.2%)	9/76 (11.8%)	

Diagnosis	Cut-off	External Validation					
		AUC	TP	FP	TN	FN	
ResNeXt	0.835	0.723 (0.618-0.828)	<b>2/3 (66.7%)</b>	<b>1/3 (33.3%)</b>	69/100 (69.0%)	31/100 (31.0%)	
PAGNet	0.8	<b>0.728 (0.626-0.830)</b>	9/17 (52.9%)	8/17 (47.1%)	62/86 (72.1%)	24/86 (27.9%)	
PAGNet+C	0.504	0.715 (0.613-0.818)	17/30 (56.7%)	13/30 (43.3%)	57/73 (78.1%)	16/73 (21.9%)	
Reader 1	>=2	0.694 (0.588-0.799)	22/46 (47.8%)	24/46 (52.2%)	46/57 (80.7%)	11/57 (19.3%)	
Reader 2	>=2	0.712 (0.610-0.815)	23/47 (48.9%)	24/47 (51.1%)	46/56 (82.1%)	10/56 (17.9%)	
PAGNet+Reader 1	>=2	0.688 (0.581-0.795)	15/27 (55.6%)	12/27 (44.4%)	58/76 (76.3%)	18/76 (23.7%)	
PAGNet+Reader 2	>=1	0.702 (0.597-0.807)	24/51 (47.1%)	27/51 (52.9%)	<b>43/52 (82.7%)</b>	<b>9/52 (17.3%)</b>	

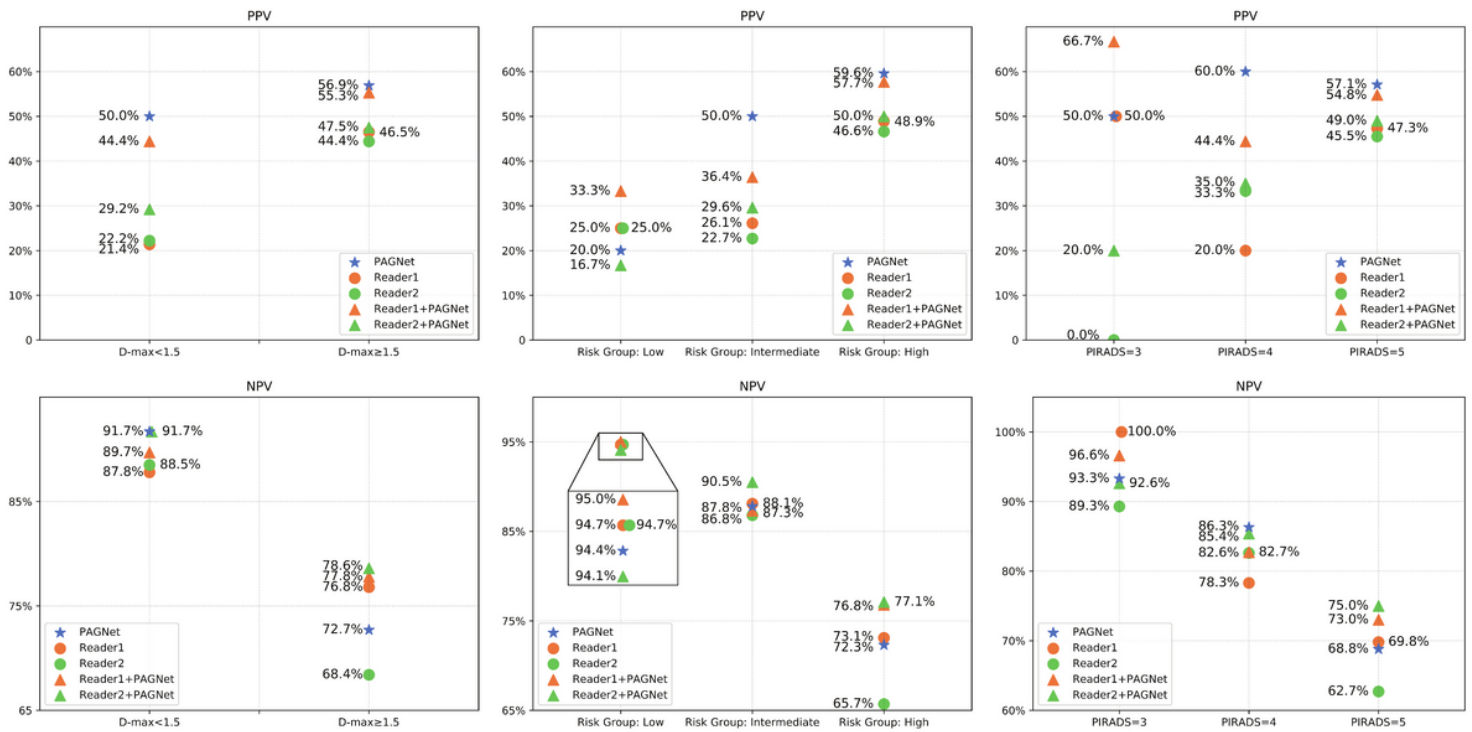
**Figure 3**

Diagnostic performance of different methods for ECE staging in the training, internal, and external validation cohorts. Yellow color highlights the best performance of the dedicated methods. AUC, area under the curve; TP, true predictive; FP, false positive; TN, true negative; FN, false negative. Optimal cutoffs were selected using the Youden J-index.



**Figure 4**

The sensitivity, specificity and summary receiver operating characteristic (SROC) curves of the AI, experts, and expert-AI interaction for ECE staging in the training, internal and external validation cohorts. Plots are individual and combined sensitivity, specificity and area under SROC curves of the diagnostic methods using a Meta regression analysis, by which the integrated effects of AI and experts are evaluated and visualized.



**Figure 5**

Positive predictive value and negative predictive value of AI, expert and expert-AI interaction for ECE staging for patients with different tumor size, D'Amico risk grade, and PI-RADS score in combined internal and external validation cohort. PPV, positive predictive value; NPV, negative predictive value.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig.S1.eps](#)
- [Fig.S2.eps](#)
- [Fig.S3.eps](#)
- [Fig.S4.eps](#)
- [SupplementarySection.docx](#)
- [tableS1.docx](#)