

CRITICAL REVIEW

Open Access



Prediction of extracapsular extension of prostate cancer by MRI radiomic signature: a systematic review

Adalgisa Guerra^{1*} , Helen Wang², Matthew R. Orton³, Marianna Konidari³, Nickolas K. Papanikolaou³, Dow Mu Koh³, Helena Donato^{4,5} and Filipe Caseiro Alves⁵

Abstract

The objective of this review is to survey radiomics signatures for detecting pathological extracapsular extension (pECE) on magnetic resonance imaging (MRI) in patients with prostate cancer (PCa) who underwent prostatectomy. Scientific Literature databases were used to search studies published from January 2007 to October 2023. All studies related to PCa MRI staging and using radiomics signatures to detect pECE after prostatectomy were included.

Systematic review was performed according to Preferred Reporting Items for Systematic Review and Meta-analyses (PRISMA). The risk of bias and certainty of the evidence was assessed using QUADAS-2 and the radiomics quality score. From 1247 article titles screened, 16 reports were assessed for eligibility, and 11 studies were included in this systematic review. All used a retrospective study design and most of them used 3 T MRI. Only two studies were performed in more than one institution. The highest AUC of a model using only radiomics features was 0.85, for the test validation. The AUC for best model performance (radiomics associated with clinical/semantic features) varied from 0.72–0.92 and 0.69–0.89 for the training and validation group, respectively. Combined models performed better than radiomics signatures alone for detecting ECE. Most of the studies showed a low to medium risk of bias. After thorough analysis, we found no strong evidence supporting the clinical use of radiomics signatures for identifying extracapsular extension (ECE) in pre-surgery PCa patients. Future studies should adopt prospective multicentre approaches using large public datasets and combined models for detecting ECE.

Critical relevant statement The use of radiomics algorithms, with clinical and AI integration, in predicting extracapsular extension, could lead to the development of more accurate predictive models, which could help improve surgical planning and lead to better outcomes for prostate cancer patients.

Protocol of systematic review registration PROSPERO CRD42021272088. Published: <https://doi.org/10.1136/bmjopen-2021-052342>.

Key Points

- Radiomics can extract diagnostic features from MRI to enhance prostate cancer diagnosis performance.
- The combined models performed better than radiomics signatures alone for detecting extracapsular extension.
- Radiomics are not yet reliable for extracapsular detection in PCa patients.

Keywords Systematic review, Radiomics, Machine learning, Prostate cancer, Extracapsular extension

*Correspondence:

Adalgisa Guerra

gisaguerra@gmail.com

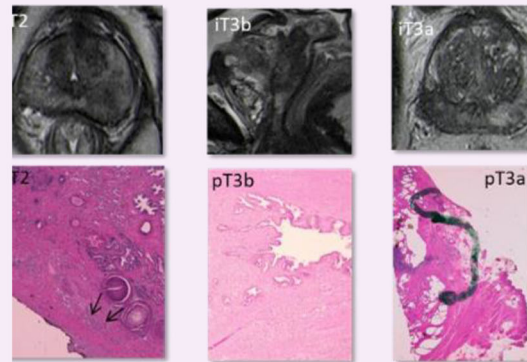
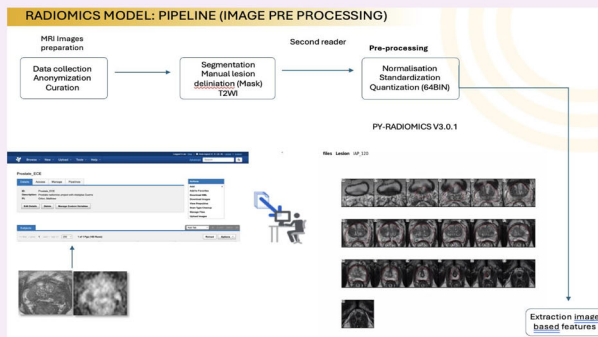
Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Graphical Abstract

Prediction of extra-capsular extension of prostate cancer by MRI radiomic signature: a systematic review


 EUROPEAN SOCIETY OF RADIOLOGY


The use of radiomics algorithms in predicting extra-capsular extension could potentially lead to the development of more accurate predictive models, which in turn could help to improve surgical planning and ultimately lead to better outcomes for prostate cancer patients.



**Insights Imaging (2024) Guerra A, Wang H, Orton MR et al.
DOI: 10.1186/s13244-024-01776-8**

Background

Prostate cancer (PCa) is a common cancer in Europe and all over the world, with around 6600 new cases diagnosed each year [1]. Radical prostatectomy is widely recognised as the standard surgical treatment for early-stage PCa. The detection of extracapsular extension (ECE) is fundamental for planning the surgical approach because it can lead to high rates of positive surgical margins, recurrence, and decreased survival [2–7]. Nomograms, such as D’Amico or CAPRA, are often used to predict the risk of advanced disease [8, 9]. Magnetic resonance imaging (MRI) has been shown to improve accuracy in predicting ECE, but there is high inter-reader variability related to the semantic features interpretation on MRI which is not consensual among the authors [10–12]. The high-quality MRI acquisition protocol and the high experience of the readers could help to improve the accuracy of MRI [10–12].

Radiomics can help extract different features from medical images using data-characterisation algorithms, improving diagnostic performance in PCa, as well as the reproducibility of the MRI examinations. Artificial intelligence (AI) and machine learning (ML) can help apply radiomics in everyday practice. However, clinically accepted and validated algorithms have not been established [13–15].

This systematic review aims to summarise evidence on using radiomics algorithms to predict pathological extracapsular extension (pECE) in PCa patients to aid surgical planning and improve outcomes.

Methods

This systematic review follows the guidelines for Preferred Reporting Items for Systematic Reviews and Meta-analyses and the protocol was registered with PROSPERO (CRD42020215671) and published in *BMJ Open* [16].

Eligibility criteria

This article reviewed manuscripts involving adult PCa patients who had a presurgical prostate biopsy indicating a Gleason score equal to or greater than 6 and underwent MRI before their surgery. Only studies using 1.5-T or 3-T MRI scanners and no prior treatment were included.

The primary outcome was pathologic local staging after surgery, with the goal of identifying imaging and clinical predictors of extracapsular extension on pathology specimens (pECE). The eligible studies were required to be retrospective or prospective cohort studies or randomised controlled trials that included prognostic factor analysis. Furthermore, these studies needed to have been published in peer-reviewed journals.

Studies were included if:

- Information regarding PCa MRI staging and pathological PCa staging was available in the published report.
- MRI images and radiomics signatures were used to detect pECE after prostatectomy.

Studies were excluded if:

- The AI/ML predictive models were built with another main predictive endpoint, such as localisation, segmentation, recurrence, prognosis, or lymph node metastasis, and characterisation of PCa without reference to the pathologic PCa staging endpoint. The paper was included in the analysis if the authors built different models with different endpoints but included the pathologic PCa staging endpoint.
- Studies with only MRI image characteristics, interpretative MRI semantic features, or combine feature for any signature without combining the radiomics feature extraction signature, were excluded.
- Cross-sectional studies, case series, case reports, case-control studies, systematic reviews, conference proceedings, and master's or PhD theses were excluded.

Search strategy

We conducted a comprehensive search across six electronic databases, namely CINAHL, EMBASE, CENTRAL (Cochrane Central Register of Controlled Trials via Wiley Online Library), PubMed, Web of Science Core Collection, and for grey literature, OpenGrey and Grey Literature Network Service. Furthermore, we manually searched through the reference lists of all included studies and previously published systematic reviews of MRI staging of PCa.

The search strategy was developed by a medical librarian with expertise in systematic reviews. The search terms were customised to the specific requirements of each database. Keywords (“Prostate neoplasm”, “Machine learning”, “Artificial intelligence” “Radiomics”, “Deep Learning”, “Staging” and “Magnetic Resonance Imaging”) or subject headings specific to each database (e.g., MeSH) were used along with Boolean operators ‘OR’ and ‘AND’ to combine the search terms effectively. The search strategy is detailed in the published protocol [16].

Study selection

There were no restrictions applied, and only studies published in the English language were included. The search was conducted in each database from January 2007 to October 2023.

Following title/abstract screening, the full texts of potentially relevant studies were evaluated. In cases where

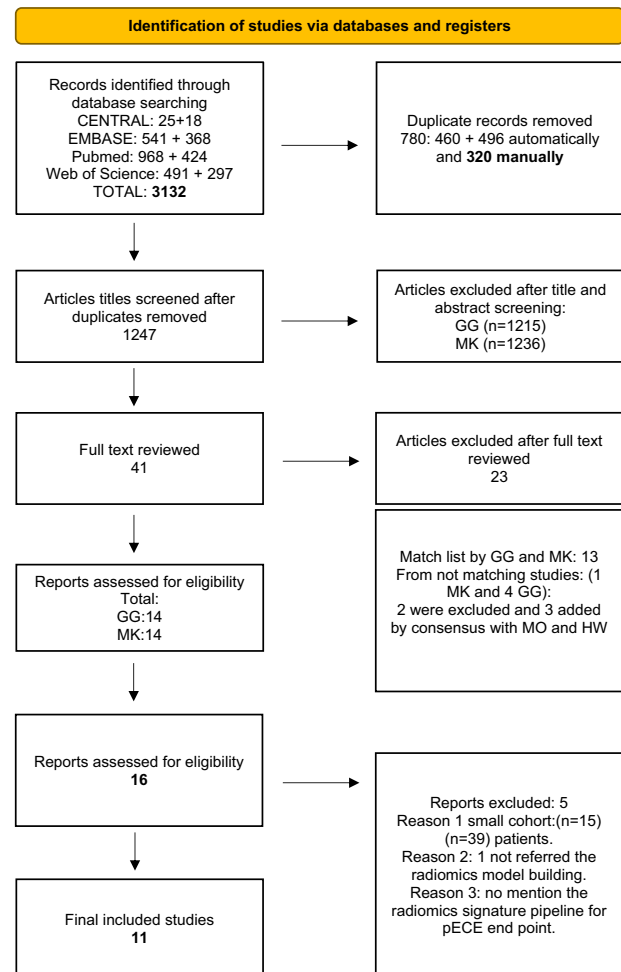


Fig. 1 PRISMA flow diagram of the study selection

a consensus was not reached between the two reviewers (A.G. and H.W.), a third reviewer (M.K.) was consulted. Additionally, the reference lists of the studies chosen for inclusion were examined for any other relevant studies. The data collection process is illustrated in Fig. 1.

Data collection process

Data extraction: Two reviewers (A.G. and M.K.) independently extracted the following data from the included studies. In cases of disagreement between the two reviewers, a consensus was reached through discussion. If necessary, two additional expert reviewers (M.O. and H.W.) were consulted. The extracted data were broadly categorised into patient and study characteristics, radiologist details, type of feature extraction (agnostic if extracted by computation algorithms), semantic (interpreted by a radiologist), model characteristics, and predictive performance. Sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) were extracted in the training and validation groups, with 95% confidence intervals where available. The radiomics and

integrated models were compared, and the best predictive performances were registered.

Risk of bias applicability

The risk of bias in individual studies was assessed by three reviewers (A.G., H.W., and M.K.). Since we included diverse types of studies, we used different tools to assess the risk of bias depending on the characteristics of the studies. Data from these studies were extracted, tabulated, and then reviewed for risk of bias and applicability using the Quality Assessment of Diagnosis Accuracy Studies version 2 (QUADAS-2) tool [17]. This tool covers four sources of bias: (1) patient selection, (2) index test, (3) standard domain, and (4) flow and timing bias. For each one, the risk of bias was assessed as high risk, unclear risk, or low risk, depending on the information offered by the study. The review authors used the signalling QUADAS-2 question information to judge the risk of bias. If all signalling questions for a domain were answered 'yes' then the risk of bias was considered 'low'. If any signalling question was answered 'no' this flagged the potential for bias. The 'unclear' category was used only when insufficient data were reported to permit a judgement.

Because QUADAS-2 sometimes does not accommodate the niche of terminology encountered in AI studies, we also added a radiomics quality score (RQS) proposed by Lambin et al to this systematic review [17, 18]. Studies with a high risk of bias and low applicability were excluded. A narrative synthesis was conducted, acknowledging the risk of bias and the strength and consistency of significant associations.

Synthesis of results

Due to differences in AI system applications, study designs, algorithms, patient cohorts, evaluation strategies, and performance metrics, narrative synthesis was chosen instead of meta-analysis. Meta-analysis could be not recommended for studies of diagnostic test accuracy that have significant differences in patient cohorts and test settings, as it would produce biased results.

Results

Studies characteristics (Table 1)

The eleven final included studies (corresponding to 0.009% from a total of 1247 screened papers), were published between 2019 and 2023, used a retrospective study design and were mainly (8) from China, two from Italy, and one from Norway. All the studies described a model based on radiomics extracted features, either alone [19–22], combined with clinical features [23–27], or in combination (integrated model with semantic interpretative features, plus agnostic radiomics features associated with clinical features to predict ECE in histopathological specimen analysis [28, 29]. All but three studies used a 3 T field strength and the total number of patients included in the models ranged between 62 to 284.

Only two studies [23, 30] were performed in more than one institution. The lesion segmentation and feature interpretation were undertaken by more than one radiologist, and the inter-agreement ratio was evaluated in all studies, except Losnegård et al [28]. A recent study [29] compared three individual models: radiomics, clinical and the assessment of ECE on MRI done by four radiologists: semantic model) and compared them to a combined model with all relevant features from the three models. Three studies also referred to other endpoints, such as the positivity of surgical margins [24, 26], lymph node metastases and tumour aggressivity [19, 26]. One study also built a radiomics model to predict PCa prognostic biological biomarkers [24].

Radiomics characteristics (Table 2 and Table-S1)

Model performance

All studies reviewed focus on developing radiomics models based on agnostic features extracted from T2WI (T2-weighted imaging) and ADC (apparent diffusion coefficient) of the manual segmented tumoural region. DCE images were also used in four publications [20, 21, 24, 28]. They compared different signatures composed by IFs (imaging features) extracted from T2WI and ADC maps independently and from the two modalities, and they tested them for prediction presence vs. absence of pECE.

For each MRI sequence, shape features (size and sphericity) and texture features (GLCM, GLRLM, GLSZM, NGLDM) were the most common features with discriminative importance and in the majority of cases were for those extracted from T2WI. The exception was Fan et al [24], where the most relevant feature was extracted from DCE. The radiomics features derived from histograms were not so relevant as the features previously mentioned. The coefficients for the calculation of the selected radiomics features were different between the studies and the authors did not find a common stable radiomics feature which could be the dominant impact factor for pECE across them. The image processing and feature selection methods were very heterogenous between the studies. Matlab, original Pyradiomics and Laplacian of Gaussian (LoG) and wavelet-filters were used for images extraction. Most researchers compared radiomics with clinical and combined models (radiomics + clinical features). In these cases, the combined models achieved the best performance (AUC: 0.92) [25], 0.72 [26], 0.95 [24], 0.72 [23], 0.76 [27] and 0.89 [29].

The highest AUC of a model using only radiomics features (tumoural region) to predict ECE was 0.93 for the training group and 0.85, for the test validation [24]. This was followed by Xu et al [25] (AUC 0.91), Yang et al [29] (AUC 0.86) and Cuocolo et al [22] (AUC 0.83 and 0.80/0.73, in training set and two external validation sets, respectively).

Ma et al [20, 21] built a radiomics signature in the peritumoural region (capsule and periprostatic fat) and

Table 1 General characteristics of the studies

Author	Year	Country	Semantic features ^a	ROI (radiomics features)	Clinical features	Combine models	Type of study	Total patients	Training group	Validation group	Number of institutions	Readers	External validation	Field strength
Ma et al [20]	2019	China	No	Manual extracapsular tissue	No	No	Retrospective	119	74	45 (internal validation)	1	2	No	3 T
Ma et al [21]	2019	China	No	Capsule and extracapsular tissue	No	No	Retrospective	210	143	67 (internal validation)	1	3	No	3 T
Losnegard et al [28]	2020	Norway	Yes	8:2 (lesion) and 6 (whole prostate)	MSKCC nomogram	Yes	Retrospective	228	Not available	Not available	1	1	No	1.5 T Endo coil
Xu et al [25]	2020	China	No	Lesion on DWI	tPSA, fPSA, PSAf/ tPSA, GS	Yes	Retrospective	95	82	33 (internal validation)	1	2	No	3 T
Bai H. et al [23]	2021	China	No	Peri-tumoural (automatically) & manual intratumoural	tPSA, GS (prostate biopsies)	Yes	Retrospective	284	158	68 (internal validation) 58 (external validation)	2	2	Yes	3 T
Cuocolo R. et al [22]	2021	Italy	No	Manual on lesion	No	No	Retrospective	193	104 (site 1)	43 (site 2), 46 (site 3)	3	3	Yes	1.5 & 3 T
Damascelli [19]	2021	Italy	No	Manual on lesion	No	No	Retrospective	62	Not available	No	1	2	No	1.5 T Endo coil
He et al [26]	2021	China	No	Lesion on ADC and T2WI	Age, tPSA, fPSA, (fPSA/fPSA), GS biopsies, % positive biopsies, cores	Yes	Retrospective	273	192	81 (internal validation)	1	3	No	3 T
Fan et al [24]	2022	China	No	Manual on lesion on DWI	Age, PSA, WBC, RBC, haemoglobin, lymphocyte, platelet, albumin, ALP, PLR, fibrinogen	Yes	Retrospective	232	185	47 (internal validation)	1	2	No	3 T
Yang Liu [27]	2023	China	No	Lesion on T2WI	Age, Gleason score P504s	Yes	Retrospective	86	no	no	1	2	No	3 T
Liqin Yang et al [29]	2023	China	Yes	Lesion on T2, ADC and DWI	Age PSA, lesion localisation GS biopsies	Yes	Retrospective	392	274	118	1	4	No	3 T

tPSA prostate-specific antigen total, fPSA prostate-specific antigen free, GS Gleason score, WBC white blood cell, RBC red blood cell, ALP alpha fetoprotein, PLR platelet-to-lymphocyte ratio, MSKCC Memorial Sloan Kettering Cancer Center, 1.5 T tesla, 3 T tesla, T2WI weighted image, DWI diffusion-weighted imaging
^a Interpreted by radiologist

Table 2 Radiomics characteristics

Author	Vendors	Input sequences	Vendor harmonisation	Image processing	Radiomics extraction	Algorithm	Evaluation strategy	SN Training group ^a	SP Training group ^a	AUC Training group ^a	95% CI Training group ^a	SN Validation group ^b	SP Validation group ^b	AUC Validation group ^b	95% CI Validation group ^b	AUC (radiomics signature alone)	95% CI (radiomics signature alone)
Bai H. et al [23]	2/Siemens. 2/GE. 2/United Imaging Health care	T2WI, ADC	Z-score normalisation	Yes	Pyradiomics	10-folds CV logistic regression with LASSO	Randomly internal validation (30%) external validation.	0.783	0.723	0.723	Not reported	0.68	0.667	0.692	Not reported	0.7	Not reported
Cuocolo R. et al [22]	2/Philips. 1/Siemens	T2WI, ADC	Z-score normalisation	Yes	Pyradiomics	10-fold cross-validation SVM	External validation	0.83	0.83	0.83	Not reported	0.79 and 0.74	0.80 and 0.73	0.80 and 0.73	Not reported	n/a	Not reported
Damascelli [19]	1/Philips	T2WI, ADC	Not required	Yes	3D slicer v.4.10.2	unsupervised hierarchical clustering, random Forest	No validation	0.88	0.89	0.88 (accuracy)	0.82–0.9	n/a	n/a	n/a	0.88 (training)	0.82–0.91	
Fan et al [24]	3/GE. 2/Siemens	T2, DWI, DCE	Not required	No	Pyradiomics	random Forest	Internal validation	0.85	0.88	0.95	0.71	0.9	0.85	0.85	0.93 (training) 0.84 (validation)	0.825	0.724
He et al [26]	1/Siemens	T2WI, ADC	Not required	Yes	Pyradiomics	10-folds CV logistic regression with LASSO	Internal validation	Not reported	Not reported	Not reported	0.727	0.688	0.728	0.821	0.625	0.724	
Losnegard [28]	1/Siemens	T2WI, T1WI, DWI, DCE	Not required	Yes	Matlab	10-folds CV random forest	No validation	0.79	0.57	0.75	SD 0.01	n/a	n/a	n/a	0.75	SD 0.06	
Ma et al [20]	1/Philips. 1/GE	T2, DWI, DCE	Colletwet normalisation	No	Matlab	10-folds CV logistic regression with LASSO	Internal validation	0.906	Not reported	Not reported	0.847, 0.948	0.821	Not reported	Not reported	0.821	0.726, 0.894	
Ma et al [21]	1/Philips. 1/GE	T2, DWI, DCE	No	No	Matlab	10-folds CV logistic regression with LASSO	Internal validation	0.768	0.919	0.902	0.840, 0.945	0.75	0.914	0.883	95% CI: 0.781, 0.949	0.883	0.781, 0.949
Xu et al [25]	2/Siemens. 2/GE. 2/United Imaging Health care	T2WI, ADC	Not required	Yes	Pyradiomics	10-folds CV logistic regression with LASSO	Randomly internal validation (30%), external validation.	0.829	0.894	0.91	0.861–0.978	0.714	0.895	0.865	0.738–0.992	0.865	0.738–0.992
Liqin Yang et al [29]	1/Siemens /Skyra	T2WI, ADC, DWI	Not required	Yes	Pyradiomics /FEature explorer: FAEx/040	LASSO	Randomly internal validation	0.817	0.818	0.897	0.861–0.934	0.90	0.73	0.89	0.861–0.934 0.837–0.950	0.882 training 0.835 validation	0.844–0.921 0.763–0.908
Yang Liu [27]	1/GE discovery	T2WI	Not required	No	Pyradiomics. Slicer radiomics	LASSO	No validation	0.952	0.523	0.763	0.662–0.864	n/a	n/a	n/a	0.861–0.934 0.837–0.950	0.739	0.631–0.846

SN sensitivity, SP specificity, AUC area under the curve, GE general electrics, CV cross-validation, SVM support vector machine, LASSO regression analysis method

^a Best results model

^b In case of combine models (radiomics + clinical + Mehrlivand grade)

compared it with the radiologist's interpretation. Pairwise comparisons showed that the radiomics signature was more accurate than the radiologist's interpretation. The accuracies (90% and 88%, respectively, in the training and validation groups) were much higher than that achieved directly by the radiologists (AUCs 0.685–0.755 in the training cohort and 0.600–0.697 in the validation cohort). This study is aligned with Yang et al study [29], where the radiomics signature is superior to radiologist interpretation (AUC 0.88 and 0.835, training and validation groups vs. AUC 0.746 and 0.774 training and validation groups) respectively.

Bai et al compare intra and peri-tumoural (PT) single radiomics signatures and achieve the best predictive value AUC: 0.70., in the PT region extracted from the ADC map. In this study the PT were automatically derived through 3D dilatation and was extracted from ADC map.

Only two authors [28, 29] built a combined model with clinical and semantic interpretative MRI features using Mehralivand's proposed EPE-grade criteria [31]. They compared the radiologist's interpretation and the Memorial Sloan Kettering Cancer Center (MSKCC) nomogram with the radiomics signature and the combined models. The AUC of the radiologist's interpretative model was similar in both studies in the training group (AUC of 0.74) [28, 29] and 0.77 in the validation group in the study of Yang et al [29]. In relation to radiomics model, the AUC was in both studies, AUC 0.75 [28] and 0.88 [29], respectively.

The combination of radiomics, radiology plus clinical interpretation performed statistically better (AUC 0.89; $p < 0.05$) than clinical model (AUC 0.74) and semantic model (AUC 0.77) but not statistically significantly (p -value 0.167) than the radiomics alone (AUC 0.835) [29].

Risk of bias assessment (Table 3)

The review authors used the QUADAS-2 and RQS methods to judge the risk of bias.

QUADAS-2

Patient selection: Only Damascelli et al [19] study was deemed to have a high risk of bias as the case selection process was unclear due to insufficient description.

The cases of Cuocolo et al [22], Losnegård et al [28] and Yang Liu [27] were considered unclear because they did not mention whether they excluded patients with any type of treatment before radical prostatectomy.

Index test (MRI Images): All patients underwent adequate and identical institutional MRI protocol. Manual segmentation of the lesions was reproducible in all the studies except one [28], in which only one radiologist undertook the lesions' segmentation, and one study [26] did not exclude poor-quality images.

Reference standard: The risk of bias for the reference standard (presence of ECE in the specimen) was low in all of the included studies.

Flow and timing: Except for four studies [19, 23, 24, 27], which did not mention the time between MRI and prostatectomy, all the included studies were consistent in using appropriate reference standards, for patients and maintaining appropriate intervals between MRI and obtaining histopathology.

Radiomics quality score

Cuocolo et al [22] study had the maximum RQS of 20 points, and Damascelli et al [19], Losnegård et al [28] and Yang Liu had the lowest scores of 10, 9 and 8 points, respectively. The main reason for this was the absence of model validation. No study was prospective or presented a phantom study on all scanners or imaging analyses at multiple time points. Only one study demonstrated open science data [22]. A cost-effectiveness analysis or biological correlation was not performed in all the studies.

Discussion

This systematic review found ten studies that aimed to predict pECE in PCa patients using radiomics signatures. Most of these studies had limited sample sizes and used data from a single centre, and four used a single MR scanner, which restricts the generalisability of their models. All the models utilised textural feature extraction, but the most significant textural features varied among the studies. The majority of the significant features were extracted from T2WI.

The Damascelli et al study being referred to here has a high risk of bias, did not perform external validation, and its patient sample size of only 62 patients was not considered adequate for robust conclusions [19]. Cuocolo et al achieved an accuracy of 83% in the training group using only ROIs of intraprostatic lesions to predict pECE [22].

Ma et al did two complementary studies comparing the radiomics model built from the first study [20] with a semantic interpretative model MRI EPE grade, in the second study [21]. They found that the radiomics model achieved higher accuracy compared to the performance of radiologists as described in the results. The low accuracy of the radiologists may be due to the difficulty in determining macroscopic ECE involvement using limited visual interpretive findings. The radiomics model has a low risk of bias, as assessed by the QUADAS tool and RQS scale, but its MATLAB feature generation approach is not open-source and uses non-standard techniques, making it difficult to replicate. The study performed an internal training and validation split of 2:1 but, did not specify which dataset was used for feature selection, which may have affected the results. The model has not been independently evaluated at another institution, and further studies are needed to validate its performance. Nevertheless, this study suggests that it is possible to use peri-tumoural regions to create radiomics signatures for predicting ECE [21].

Table 3 Risk of bias: QUADAS-2 and radiomics quality scores

Risk of bias	Xu [25]	MA [20]	MA [21]	Losnegard [28]	FAN [24]	He [26]	Damascelli [19]	Cuocolo [22]	Bai [23]	Li Qin Yang [29]	Yang Liu [27]
QUADAS-2											
Patient selection	Low	Low	Low	Unclear	Low	Low	High	Unclear	Low	Low	Unclear
Index test (MRI)	Low	Low	Low	High	Low	Unclear	Low	Low	Low	Low	Unclear
Reference standard (pECE)	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low	Unclear
Flow and timing	Low	Low	Low	Low	Unclear	Low	Unclear	Low	Unclear	Low	Low
Global	Low	Low	Low	High	Unclear	Unclear	High	Unclear	Unclear	Low	Unclear
Radiomics quality score (RQS)											
Image protocol quality	2	2	2	2	2	2	2	2	2	2	1
Multiple segmentations	1	1	1	1	1	1	1	1	1	1	1
Phantom study on all scanners	0	0	0	0	0	0	0	0	0	0	0
Imaging at multiple time points	0	0	0	0	0	0	0	0	0	0	0
Feature reduction or adjustment for multiple testing	3	3	3	3	3	3	3	3	3	3	3
Multivariable analysis with non-radiomics features	1	0	0	1	1	1	0	0	1	1	1
Detect and discuss biological correlates	0	0	0	0	0	0	0	0	0	0	0
Cut-off analyses	1	1	1	1	1	1	1	1	1	1	1
Discrimination statistics	2	2	2	2	2	2	2	2	2	2	1
Calibration statistics	2	2	2	0	2	2	2	2	1	2	1
Prospective study registered in a trial database	0	0	0	0	0	0	0	0	0	0	0
Validation	2	2	2	-5	2	2	-5	4	3	2	-5
Comparison to 'gold standard' (pECE)	2	2	2	2	2	2	2	2	2	2	2
Potential clinical utility	2	2	2	2	2	2	2	2	2	2	2
Cost-effectiveness analysis	0	0	0	0	0	0	0	0	0	0	0
Open science and data	0	0	0	0	0	0	0	1	0	0	0
Total	18	17	17	9	18	18	10	20	16	18	8

pECE extracapsular extension in pathology, MRI magnetic resonance imaging

In relation to the remaining studies [23–27], clinical features were used to construct combined models in addition to the radiomics features. The dominant clinical predictors were the serum PSA and Gleason Score (GS) of the biopsy. They reported better results in combined models with clinical and radiomics variables compared with models using radiomics features alone, with values that varied between 72% and 95% and with moderate risk of bias in the QUADAS-2 evaluation and RQS between 16 and 18 points. From these, Fan et al [24] study had the best accuracy, and the textural extraction features were derived from the ROI drawn on DWI. However, it is a single institution study using two different scanners with an internal validation dataset.

Two studies [28, 29], created a combined model that associated semantic features from the radiologist's interpretation with the radiomics model and clinical features. In the study [28] the AUC of radiomics model is almost the same than semantic interpretative model (AUC 0.75 vs. AUC 0.74, respectively). However, the model was executed in a limited way, without a validation group, and the risk of bias was very high (high in QUADAS and only 9 in RQS). In the other [29], more recent research, the authors proved that a combined model achieved the best AUC in the validation group compared to the other models. However, this study should be conducted to utilise external cohorts (from different institutions) to validate the robustness of radiomics and combined models to detect ECE.

While all the studies used one or more feature selection strategies, in order to reduce the overfitting. In addition, the use of different feature sets in different studies led to a lack of consistency in the features present in the final models, as previously mentioned, precluding any attempt to synergistically analyse the relevant radiomics features for predicting pECE across studies.

Future radiomics studies must ensure high-quality data collection and standardisation of radiomics features across different institutions and imaging protocols. The IBSI (Image Biomarker Standardisation Initiative) seeks to provide image biomarker nomenclature and definitions, benchmark datasets, and benchmark values to verify image processing and image biomarker calculations, as well as reporting guidelines, for high-throughput image analysis.

By addressing these concerns, future radiomics studies can enhance the reliability and clinical utility of radiomics signatures in detecting ECE in patients with PCa before surgery.

This review had certain limitations. Firstly, although our search strategy was comprehensive, there could be studies that were published between our search period and the publication of this review. Secondly, this systematic review only focused on radiomics signatures and did not analyse other AI methods, semantic interpretative scores or nomograms to detect ECE.

Finally, this review would have benefited from a quantitative synthesis or meta-analysis of the analysed articles, but unfortunately was not possible as key statistical data and the dominant features were not reported for this small sample of studies.

Conclusion

Non-imaging biomarkers such as PSA and GS have shown promise in predicting ECE in PCa. When combined with MRI data, Radiomics signatures could enhance accuracy in predicting ECE. However, current evidence lacks robustness to support the clinical use of radiomics signatures for ECE detection pre-surgery. Future radiomics studies need prospective testing in multicentre settings with large datasets, including external validation cohorts, to enhance reliability and clinical utility in detecting ECE.

Abbreviations

AI	Artificial intelligence
ECE	Extracapsular extension
ML	Machine learning
PCa	Prostate cancer
pECE	Extracapsular extension in the prostatectomy specimen
QUADAS-2	Quality Assessment of Diagnosis Accuracy Studies version 2
RQS	Radiomics quality score

Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1186/s13244-024-01776-8>.

ELECTRONIC SUPPLEMENTARY MATERIAL

Authors contributions

A.G.: study design, data collection, risk of bias and data analysis, manuscript elaboration. H.W.: data collection, risk of bias, data analysis and manuscript elaboration. M.K.: data collection. M.O.: data analysis. N.P.: review and editing. D.K.: review and editing. H.D.: search strategy and manage the study records and review. F.C.A.: review. All the authors read and agreed with the final version of the manuscript. Adalgisa Guerra is the guarantor of this article.

Funding

This project represents independent research funded by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, London, and by the Royal Marsden Cancer Charity [and Cancer Research UK (CRUK) National Cancer Imaging Trials Accelerator (NCITA) and is also supported by a PhD student scholarship supported by Luz Saúde Clinical Research and Innovation Programme. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Data availability

This is a systematic review based on previously published data, therefore no participant recruitment takes place.

Declarations

Ethics approval and consent to participate

The involvement of participants in the recruitment and dissemination of results is not applicable.

Consent for publication

Not required.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Radiology, Hospital da Luz Lisbon, Lisboa, Portugal. ²Royal Surrey County Hospital HSH Foundation Trust. Royal Marsden Hospital NHS Foundation Trust, London, England. ³Royal Marsden Hospital NHS Foundation Trust, London, England. ⁴Documentation and Scientific Information Service, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal. ⁵Faculty of Medicine, University of Coimbra, Coimbra, Portugal.

Received: 8 January 2024 Accepted: 10 July 2024

Published online: 26 August 2024

References

- Torre LA, Bray F, Siegel RL et al (2015) Global cancer statistics 2012. *CA Cancer J Clin* 65:87–108. <https://doi.org/10.3322/caac.21262>
- Johnson MT, Ramsey ML, Ebel JJ et al (2014) Do robotic prostatectomy positive surgical margins occur in the same location as extraprostatic extension? *World J Urol* 32:761–767. <https://doi.org/10.1007/s00345-013-1149-5>
- Ficarra V, Novara G, Secco S et al (2009) Predictors of positive surgical margins after laparoscopic robot assisted radical prostatectomy. *J Urol* 182:2682–2688. <https://doi.org/10.1016/j.juro.2009.08.037>
- Hull GW, Rabbani F, Abbas F et al (2002) Cancer control with radical prostatectomy alone in 1,000 consecutive patients. *J Urol* 167:528–534. [https://doi.org/10.1016/S0022-5347\(01\)69079-7](https://doi.org/10.1016/S0022-5347(01)69079-7)
- Mikel Hubanks J, Boorjian SA, Frank I et al (2014) The presence of extracapsular extension is associated with an increased risk of death from prostate cancer after radical prostatectomy for patients with seminal vesicle invasion and negative lymph nodes. *Urologic Oncol* 32:21–27. <https://doi.org/10.1016/j.urolonc.2012.09.002>
- Tollefson MK, Karnes RJ, Rangel LJ et al (2013) The impact of clinical stage on prostate cancer survival following radical prostatectomy. *J Urol* 189:1707–1712. <https://doi.org/10.1016/j.juro.2012.11.065>
- Wheeler TM, Dilliglugil Ö, Kattan MW et al (1998) Clinical and pathological significance of the level and extent of capsular invasion in clinical stage T1–2 prostate cancer. *Hum Pathol* 29:856–862. [https://doi.org/10.1016/S0046-8177\(98\)90457-9](https://doi.org/10.1016/S0046-8177(98)90457-9)
- Reisæter LAR, Fütterer JJ, Losnegård A et al (2018) Optimising preoperative risk stratification tools for prostate cancer using mpMRI. *Eur Radiol* 28:1016–1026. <https://doi.org/10.1007/s00330-017-5031-5>
- D'Amico AV (1998) Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 280:969. <https://doi.org/10.1001/jama.280.11.969>
- Chun FKH, Steuber T, Erbersdobler A et al (2006) Development and internal validation of a nomogram predicting the probability of prostate cancer Gleason sum upgrading between biopsy and radical prostatectomy pathology. *Eur Urol* 49:820–826. <https://doi.org/10.1016/j.eururo.2005.11.007>
- Hoeks CMA, Barentsz JO, Hambroek T et al (2011) Prostate cancer: multiparametric MR imaging for detection, localization, and staging. *Radiology* 261:46–66. <https://doi.org/10.1148/radiol.11091822>
- de Rooij M, Israël B, Tummers M et al (2020) ESUR/ESUI consensus statements on multi-parametric MRI for the detection of clinically significant prostate cancer: quality requirements for image acquisition, interpretation and radiologists' training. *Eur Radiol* 30:5404–5416. <https://doi.org/10.1007/s00330-020-06929-z>
- Stoyanova R, Takhar M, Tschudi Y et al (2016) Prostate cancer radiomics and the promise of radiogenomics. *Transl Cancer Res* 5:432–447. <https://doi.org/10.21037/tcr.2016.06.20>
- Toivonen J, Perez IM, Movahedi P et al (2019) Radiomics and machine learning of multisequence multiparametric prostate MRI: towards improved non-invasive prostate cancer characterization. *PLoS One* 14:e0217702. <https://doi.org/10.1371/journal.pone.0217702>
- Syer T, Mehta P, Antonelli M et al (2021) Artificial intelligence compared to radiologists for the initial diagnosis of prostate cancer on magnetic resonance imaging: a systematic review and recommendations for future studies. *Cancers* 13:3318. <https://doi.org/10.3390/cancers13133318>
- Guerra A, Negrão E, Papanikolaou N, Donato H (2022) Machine learning in predicting extracapsular extension (ECE) of prostate cancer with MRI: a protocol for a systematic literature review. *BMJ Open* 12:e052342. <https://doi.org/10.1136/bmjopen-2021-052342>
- Jayakumar S, Sounderajah V, Normahani P et al (2022) Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research study. *NPJ Digit Med* 5:11. <https://doi.org/10.1038/s41746-021-00544-y>
- Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14:749–762. <https://doi.org/10.1038/nrclinonc.2017.141>
- Damascellì A, Gallivanone F, Cristel G et al (2021) Advanced imaging analysis in prostate MRI: building a radiomic signature to predict tumor aggressiveness. *Diagnostics* 11:1–13. <https://doi.org/10.3390/diagnostics11040594>
- Ma S, Xie H, Wang H et al (2020) Preoperative prediction of extracapsular extension: radiomics signature based on magnetic resonance imaging to stage prostate cancer. *Mol Imaging Biol* 22:711–721. <https://doi.org/10.1007/s11307-019-01405-7>
- Ma S, Xie H, Wang H et al (2019) MRI-based radiomics signature for the preoperative prediction of extracapsular extension of prostate cancer. *J Magn Reson Imaging* 50:1914–1925. <https://doi.org/10.1002/jmri.26777>
- Cuocolo R, Stanzione A, Faletti R et al (2021) MRI index lesion radiomics and machine learning for detection of extraprostatic extension of disease: a multicenter study. *Eur Radiol* 31:7575–7583. <https://doi.org/10.1007/s00330-021-07856-3>
- Bai H, Xia W, Ji X et al (2021) Multiparametric magnetic resonance imaging-based peritumoral radiomics for preoperative prediction of the presence of extracapsular extension with prostate cancer. *J Magn Reson Imaging* 54:1222–1230. <https://doi.org/10.1002/jmri.27678>
- Fan X, Xie N, Chen J et al (2022) Multiparametric MRI and machine learning based radiomic models for preoperative prediction of multiple biological characteristics in prostate cancer. *Front Oncol* 12:1–12. <https://doi.org/10.3389/fonc.2022.839621>
- Xu L, Zhang G, Zhao L et al (2020) Radiomics based on multiparametric magnetic resonance imaging to predict extraprostatic extension of prostate cancer. *Front Oncol* 10:1–9. <https://doi.org/10.3389/fonc.2020.00940>
- He D, Wang X, Fu C et al (2021) MRI-based radiomics models to assess prostate cancer, extracapsular extension and positive surgical margins. *Cancer Imaging* 21:46. <https://doi.org/10.1186/s40644-021-00414-6>
- Liu Y (2023) Comparison of magnetic resonance imaging-based radiomics features with nomogram for prediction of prostate cancer invasion. *Int J Gen Med* 16:3043–3051. <https://doi.org/10.2147/ijgm.s419039>
- Losnegård A, Reisæter LAR, Halvorsen OJ et al (2020) Magnetic resonance radiomics for prediction of extraprostatic extension in non-favorable intermediate-and high-risk prostate cancer patients. *Acta Radiol* 61:1570–1579. <https://doi.org/10.1177/0284185120905066>
- Yang L, Jin P, Qian J et al (2023) Value of a combined magnetic resonance imaging-based radiomics-clinical model for predicting extracapsular extension in prostate cancer: a preliminary study. *Transl Cancer Res* 12:1787–1801. <https://doi.org/10.21037/tcr-22-2750>
- Cuocolo R, Cipullo MB, Stanzione A et al (2019) Machine learning applications in prostate cancer magnetic resonance imaging. *Eur Radiol Exp* 3:e35. <https://doi.org/10.1186/s41747-019-0109-2>
- Mehralivand S, Shih JH, Harmon S et al (2019) A grading system for the assessment of risk of extraprostatic extension of prostate cancer at multiparametric MRI. *Radiology* 290:709–719. <https://doi.org/10.1148/radiol.2018181278>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.