

Review

# Radiomic and Genomic Machine Learning Method Performance for Prostate Cancer Diagnosis: Systematic Literature Review

Rossana Castaldo<sup>1</sup>, PhD; Carlo Cavaliere<sup>1</sup>, PhD; Andrea Soricelli<sup>1</sup>, PhD; Marco Salvatore<sup>1</sup>, PhD; Leandro Pecchia<sup>2\*</sup>, PhD; Monica Franzese<sup>1\*</sup>, PhD

<sup>1</sup>IRCCS SDN, Naples, Italy

<sup>2</sup>University of Warwick, Coventry, United Kingdom

\*these authors contributed equally

**Corresponding Author:**

Carlo Cavaliere, PhD

IRCCS SDN

113 Via E Gianturco

Naples, 80143

Italy

Phone: 39 3470563424

Email: [carlo.cavaliere@synlab.it](mailto:carlo.cavaliere@synlab.it)

## Abstract

**Background:** Machine learning algorithms have been drawing attention at the joining of pathology and radiology in prostate cancer research. However, due to their algorithmic learning complexity and the variability of their architecture, there is an ongoing need to analyze their performance.

**Objective:** This study assesses the source of heterogeneity and the performance of machine learning applied to radiomic, genomic, and clinical biomarkers for the diagnosis of prostate cancer. One research focus of this study was on clearly identifying problems and issues related to the implementation of machine learning in clinical studies.

**Methods:** Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol, 816 titles were identified from the PubMed, Scopus, and OvidSP databases. Studies that used machine learning to detect prostate cancer and provided performance measures were included in our analysis. The quality of the eligible studies was assessed using the QUADAS-2 (quality assessment of diagnostic accuracy studies–version 2) tool. The hierarchical multivariate model was applied to the pooled data in a meta-analysis. To investigate the heterogeneity among studies,  $I^2$  statistics were performed along with visual evaluation of coupled forest plots. Due to the internal heterogeneity among machine learning algorithms, subgroup analysis was carried out to investigate the diagnostic capability of machine learning systems in clinical practice.

**Results:** In the final analysis, 37 studies were included, of which 29 entered the meta-analysis pooling. The analysis of machine learning methods to detect prostate cancer reveals the limited usage of the methods and the lack of standards that hinder the implementation of machine learning in clinical applications.

**Conclusions:** The performance of machine learning for diagnosis of prostate cancer was considered satisfactory for several studies investigating the multiparametric magnetic resonance imaging and urine biomarkers; however, given the limitations indicated in our study, further studies are warranted to extend the potential use of machine learning to clinical settings. Recommendations on the use of machine learning techniques were also provided to help researchers to design robust studies to facilitate evidence generation from the use of radiomic and genomic biomarkers.

(*J Med Internet Res* 2021;23(4):e22394) doi: [10.2196/22394](https://doi.org/10.2196/22394)

**KEYWORDS**

prostate cancer; machine learning; systematic review; meta-analysis; diagnosis; imaging; radiomics; genomics; clinical; biomarkers

## Introduction

Prostate cancer (PCa) is the second most diagnosed cancer worldwide in men [1,2]. To guarantee cancer-specific survival, early detection of PCa is essential at a treatable stage. The most common method to diagnose PCa is via transrectal ultrasonography (TRUS) [3]. The rapid development of medical imaging techniques and modalities has demonstrated great value in the screening, diagnosis, treatment response measurement, and prognosis evaluation of PCa. In particular, radiomic investigation, defined as computationally extracting quantitative image features for the characterization of disease patterns [4], has been intensively applied to tumor detection, localization, staging, aggressiveness assessment, treatment decision-making assistance, and patient follow-up in PCa [5].

More recently, multiparametric magnetic resonance imaging (mpMRI) has been demonstrated to be a better radiomic biomarker than systematic TRUS biopsy, achieving high diagnostic accuracy and becoming a clinical routine investigation for suspected PCa patients [6,7]. The second version of the Prostate Imaging Reporting and Data System (PI-RADS-V2) was updated in regard to minimum technical acquisition parameters and image interpretation [8]. It describes a standard prostate mpMRI protocol that combines anatomical T2-weighted images with functional sequences, that is, diffusion-weighted imaging (DWI) or dynamic contrast-enhanced (DCE) sequences.

Alongside radiomic investigation, there are numerous Food and Drug Administration–approved genomic biomarkers underlying the biomolecular functions most strongly associated with clinical outcomes. In fact, a major focus of personalized medicine has been the biomolecular characterization of tumors by integrating genomics into clinical oncology to identify unique druggable targets and generate higher-order tumor classification methods that can support clinical treatment decisions [9]. They are mainly used to decide whether biopsy screening is necessary and whether patients require primary treatment (such as radical prostatectomy or radiation therapy) [10]. The combination of biopsy screening and evaluation of the Gleason score still remains the most widely accepted grading system in the evaluation of prostatic adenocarcinoma [11]. The Gleason grading system is based on a morphologic continuum of architectural dedifferentiation and is directly correlated with response to therapy and mortality rate. However, novel biomarker tests that can potentially detect PCa from blood, urine, tissue, and semen samples continue to be investigated. Prostate-specific antigen (PSA) is the most commonly used biomarker for the management of PCa [12]. Increased PSA density has been shown to be associated with increased risk of PCa compared to healthy or benign prostatic hyperplasia patients [13]. The Prostate Health Index and 4Kscore utilize isoforms of PSA and its precursors to help risk-stratify patients with an abnormal PSA level. In addition, microRNAs have an important role during tumor progression, and their combination with PSA serum can improve prediction of PCa status [14-16]. Other proposed biomarkers that belong to various classes of biological compounds, including proteins and metabolites, have shown to be noninvasive methods with high diagnostic potential [17].

Over the last decade, the landscape for PCa detection tools has expanded to include novel biomarkers, clinical information, genomic assays, and noninvasive imaging tests. The prospect of detecting PCa using readily available clinical and demographic health information is a potentially innovative part of improving screening practices [18].

In this scenario, machine learning (ML) is helping researchers in identifying and discovering new biomarkers to detect PCa. ML is a branch of artificial intelligence (AI) techniques based on the development and training of algorithms by learning from data and the performance of predictions. ML methods are able to improve and learn over time in a more efficient way than classical statistical approaches [19]. Therefore, ML has been widely used in radiology and recently in the field of bioinformatics [6,20]. A recent field of ML, deep learning (DL), is based on artificial neural networks, which offer superior problem-solving capabilities applied to large heterogenous data sets [20,21]. Specifically, ML allows the integration or combination of different layers of data, such as those from medical images, laboratory results, clinical outcomes, biomarkers, and other biological features, for better prognostication and stratification of patients toward personalized medicine [22,23]. However, the accuracy of such algorithms can be highly impacted by the complex workflows adopted to develop and generalize such ML algorithms [24,25]. High heterogeneity is expected, as ML problems are usually regarded as black boxes, and the consideration of all possible risk factors and transformation is tremendously difficult [26,27]. Moreover, there are no clear guidelines on how to develop ML approaches for medical studies.

Therefore, this study aimed to suggest an integrated estimate of the accuracy for use of ML algorithms in detecting PCa through a systematic review and meta-analysis of the available studies. Due to the internal heterogeneity of ML algorithms, subgroup analyses helped in investigating the diagnostic capability of ML systems and highlighting the sources of bias and common pitfalls to avoid in order to assure reproducibility among studies. Subgroup analyses were mainly based on the model choice, model development, and validation methods to identify potential covariates that could influence the diagnostic performance of ML.

This review helps to support ML studies in rising up the pyramid of evidence. In fact, we identify and discuss recurrent factors that hinder the uptake of these studies in clinical settings.

To the best of the authors' knowledge, there are no systematic review and meta-analysis studies evaluating the performance and estimating the current status of existing approaches on PCa detection. Therefore, this study aims to fill the gap in the existing literature and gather recommendations on ML model development to achieve robust results to automatically detect PCa.

## Methods

We conducted and reported this meta-analysis in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [28]. Two researchers

(RC and MF), who were blinded to the articles' author information, conducted the study inclusion, data extraction, and assessment of the risk of bias independently. A third author (CC) was consulted in case of disagreements.

## Search Strategy

The PubMed, Scopus, and OvidSP (ie, Embase) databases were searched to identify studies evaluating the accuracy of radiomic, clinical, and genomic biomarkers in the diagnosis of PCa. The following criteria were used to limit the research: papers published in the last 5 years (from 2015 to 2020) to guarantee homogeneity among radiomic studies, as the new protocol (PI-RADS) for mpMRI was updated in 2015 [8]; study on adult humans (ie, not animals); language (English); and full-text publications. The search took place on February 24, 2020. The reference lists of the included studies were checked, and the authors were contacted if required. The search strategy and queries for each search database are presented in Table S1 in [Multimedia Appendix 1](#).

An author (RC) retrieved the initial search results and removed duplicates via Excel (Microsoft). Subsequently, another author (MF) manually searched for and removed any remaining duplicates. Finally, RC and MF independently screened the studies by title, abstract, and keywords, after which the full texts of the selected studies were assessed by inclusion and exclusion criteria. The main considerations for study inclusion were if machine learning was fully applied in distinguishing individuals or lesions with clinically diagnosed PCa from controls and if the study assessed the accuracy of such applications. Detailed inclusion and exclusion criteria are reported in Table S2 in [Multimedia Appendix 1](#).

## Data Extraction and Outcomes of Interest

After the evaluation was completed, two authors extracted the following information from the selected literature: literature data—the first author, publication date, study population, number of patients, study design, and data collection; basic research information—age, Gleason score, and PSA level, where possible; information regarding the reference standard used in individual studies; definitions of positive and negative PCa (PCa positive and control) and methodologies to distinguish individuals or lesions with PCa from the control group; specific methodologies to process and classify data for use in machine learning algorithms; and the sensitivity, specificity, and, if available, true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) rates.

The authors independently graded the quality of the eligible studies using the quality assessment of diagnostic accuracy studies—version 2 (QUADAS-2) tool [29]. The full process is provided in the supplementary materials in [Multimedia Appendix 1](#).

## Meta-analysis Paper Inclusion Criteria and Subgroup Analysis

For radiomic analysis, due to the very low number of included studies investigating central gland and transition zone (TZ) prostate tumors, only studies investigating the peripheral zone were included in the meta-analysis. This was also due to the

fact that central gland and TZ prostate tumors have significantly different quantitative imaging signatures [30], and they could have highly biased the final results.

Due to the low number of studies employing 3D volumes of interest (VOIs) to extract quantitative features, only studies delineating 2D regions of interest (ROIs) were included in the meta-analysis to reduce the risk of bias. This was mainly due to the fact that significant differences were found between prediction performance when using 3D VOIs and that when using 2D ROIs [31]. If studies investigated several diagnostic imaging techniques via ML, only classification models using mpMRI sequences were included in the meta-analysis.

To reduce heterogeneity among the selected studies, subgroup analyses were carried out for radiomic and genomic studies due to their intrinsic differences in data acquisition, analysis, and feature extraction. Radiomic subgroup analyses helped to investigate the role of the mpMRI biomarker in detecting PCa via ML, whereas genomic subgroup analyses were carried out to understand the role of genomic biomarkers in detecting PCa via ML.

Several covariates suitable for subgroup analysis were identified during the review process where the individual peculiarities of the studies, which may affect the outcome, were investigated.

The included studies were investigated if they explored a patient- or lesion-based model, validation approaches (cross-validation, hold-out approach or external validation, or no validation), ML algorithms (regression-based model, tree-based model, or deep learning algorithms), whether the studies used a DL or ML approach, or whether the employed data set was balanced or unbalanced. For genomic studies, the use of different specimens (ie, urine, serum, semen, and tissue) was also investigated in a subgroup analysis. One study [17] investigated both urine and serum specimens separately; therefore, ML performance was included for both predictors in the meta-analysis.

In case a study investigated multiple ML algorithms, only the method achieving the highest area under the curve (AUC) was included in the meta-analysis, as AUC is a good estimator of ML performance.

## Statistical Analysis and Software Tools

This meta-analysis was conducted via the Open Meta-Analyst Software tool, and statistical significance was expressed with 95% CIs. Pooled estimates for sensitivity and specificity with the corresponding 95% CIs were used to determine the accuracy of machine learning for detecting PCa in radiomic and genomic studies. From these data, we generated a hierarchical summary receiver operating characteristic curve (HSROC) and coupled forest plots by random-effects model. Heterogeneity among studies was assessed by calculation of the inconsistency index ( $I^2$ ) and evaluation of the Cochran  $\chi^2$  test ( $Q$  test). An  $I^2$  of  $\geq 50\%$  and  $P < .001$  indicated substantial between-study heterogeneity. The TP/FP/TN/FN values were extracted or calculated from each independent study. A correction factor of 0.5 was added if any of the TP/FP/TN/FN rates reported a value of 0, in order to prevent zero cell count problem [32].

In our meta-analysis, a multivariate random-effects model was used to consider both within- and between-subject variability and threshold effects [33]. The HSROC curve was specified by pooled sensitivity and specificity point. Attempts were made to resolve the heterogeneity by performing a subgroup analysis [34].

## Results

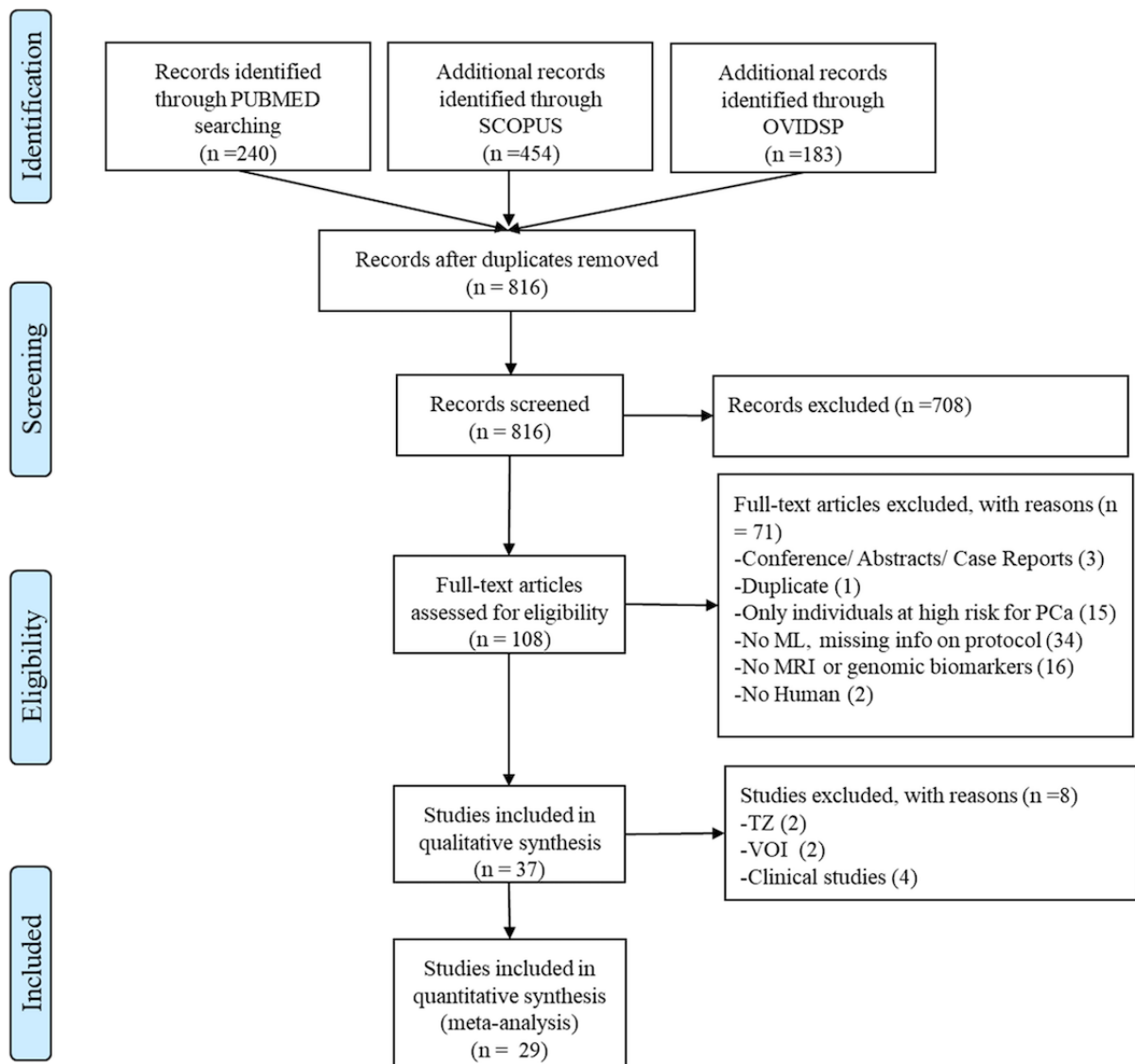
### Literature Search

According to the search strategy described above, 877 titles were identified in PubMed, Scopus, and OvidSP. After removing

duplicates, 816 titles were considered. Of these, 708 were excluded after reading of the abstracts because they did not meet the inclusion criteria. From the remaining 108 full-text articles, 71 were removed due to the exclusion criteria. Finally, 37 full texts were included in the qualitative analysis, and 29 studies were considered appropriate for inclusion in the meta-analysis. A flowchart of the literature search is shown in Figure 1.

The distribution of the risk of bias evaluated via the QUADAS-2 tool for the included studies is presented in the supplementary materials (Figure S1 in Multimedia Appendix 1).

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of literature search: included/excluded titles, abstracts, and full papers. ML: machine learning; MRI: magnetic resonance imaging; PCa: prostate cancer; TZ: transition zone; VOI: volume of interest.



### Characteristics of the Included Studies

The publication years ranged from 2015 to 2020 to guarantee homogeneity among radiomic studies, as the new PI-RADS

was updated in 2015 [8]. All patients were diagnosed with PCa by biopsy. The main characteristics of the studies are reported in Table 1. The extracted raw data are presented in Tables S3 and S4 in Multimedia Appendix 1.

**Table 1.** Characteristics of 37 studies included in the systematic review.

Characteristics	Studies, n	Patients (average over the number of studies), n
<b>Study type</b>		
Prospective	8	2210 (276.25)
Retrospective	29	6414 (221.17)
<b>Data set type</b>		
Private data set	33	7760 (235.15)
Public database (SPIE-AAPM-NCI <sup>a</sup> PROSTATEx challenge)	2	399 (199.5)
Mixed (private and public) data set	2	465 (232.5)
<b>Classification algorithms</b>		
Random forest	4	1621(405.25)
Regression-based models	20	4678 (233.9)
Partial least squares discriminant analysis (PLS-DA)	2	180 (90)
Linear discriminant analysis (LDA)	1	53
Support vector machine (SVM)	2	65 (32.5)
Classification and regression tree (CART)	1	67
Artificial neural networks (ANNs)	2	1012 (506)
Deep neural networks (DNNs)	1	195
Convolutional neural networks (CNNs)	3	696 (232)
Deep learning: SNCSAE <sup>b</sup>	1	57
<b>Predictor type</b>		
Multiparametric MRI <sup>c</sup>	20	5058 (252.9)
<b>Genetic or molecular biomarker</b>		
Urine	6	930 (155)
Serum	3	901 (300.3)
Semen	2	108 (54)
Tissue	2	800 (400)
Clinical data	4	2812 (703)
<b>Validation method</b>		
Internal validation	29	6540 (225.52)
External validation	3	1380 (460)
Internal and external validation	1	364
Unknown	5	704 (140.8)

<sup>a</sup>SPIE-AAPM-NCI: International Society for Optics and Photonics–American Association of Physicists in Medicine–National Cancer Institute.

<sup>b</sup>SNCSAE: stacked nonnegativity constraint sparse autoencoders.

<sup>c</sup>MRI: magnetic resonance imaging.

## Quantitative Analysis (Meta-analysis)

Of the final 37 papers, 29 were considered for the meta-analysis. Eight studies were excluded to reduce heterogeneity among the studies. Of those, 2 studies were excluded because they extracted radiomic features from VOIs [35,36], and 2 studies [37,38] were excluded because they only focused on detecting TZ tumors. Due to the low number of studies investigating TZ tumors, a comparative assessment of the results for the peripheral zone, central gland, and TZ was not possible.

Studies [18,39-41] employing only clinical information were excluded because a minimum sample of 5 studies is recommended for a meta-analysis [34,42]. In fact, 5 or more studies are needed to reasonably achieve power from random-effects meta-analyses [43].

## Radiomic

All the included studies for the radiomic analysis are reported in Table 2. A total of 4438 independent samples were inspected

from 16 studies with sensitivity and specificity ranging from 0.62 to 0.99 and 0.51 to 0.98, respectively.

Multivariate meta-analysis via the HSROC model was assessed for all the studies (Figure S2 in [Multimedia Appendix 1](#)). The pooled sensitivity and specificity were 0.815 (95% CI 0.410-0.999) and 0.828 (95% CI 0.424-0.999), respectively.

The calculated heterogeneity values for pooled sensitivity and specificity were 84% and 79% ( $P<.001$ ), respectively; therefore, a random-effects model was adopted to generate coupled forest plots (Figure S3 in [Multimedia Appendix 1](#)).

### Subgroup Analysis

To resolve the heterogeneity, subgroup analysis was conducted for different covariates. The subgroup analysis per model-based covariate is shown in [Figure 2](#). Subgroup 1 included the studies that employed a lesion-based ML approach. Those studies [44-54] employed multiple lesions for each patient enrolled in the study. Subgroup 2 gathered those studies [55-59] that enrolled two distinct groups (PCa and controls) and employed a patient-based ML approach. The heterogeneity in the subgroups was greater than 70% (subgroup 1:  $P<.001$ , subgroup 2:  $P=.002$ ).

**Table 2.** Accuracy measures of radiomic studies for the systematic review.

Study, year	Model basis <sup>a</sup>	Patients, n	Total sample (PCa+, PCa-) <sup>b</sup>	Crossval <sup>c</sup> /split/none	ML <sup>d</sup> methods <sup>e</sup>	TP, <sup>f</sup> n	FN, <sup>g</sup> n	FP, <sup>h</sup> n	TN, <sup>i</sup> n	Sen <sup>j</sup> (lower-upper)	Spe <sup>k</sup> (lower-upper)
Zhao, 2015 [44]	LB	71	238 (92, 146)	120 (60, 60)	ANN	57	35	16	130	0.620 (0.517-0.713)	0.890 (0.829-0.932)
Valerio, 2016 [45]	LB	53	106 (53, 53)	None	LDA	51	2	1	53	0.962 (0.861-0.991)	0.981 (0.880-0.997)
Lay, 2017 [46]	LB	224	410 (123, 287)	Crossval	RF	109	14	57	230	0.886 (0.817-0.931)	0.801 (0.751-0.844)
Reda, 2017 [47]	LB	18	53 (26, 27)	Crossval	SNCSAE	26	1	1	27	0.963 (0.779-0.995)	0.964 (0.786-0.995)
Starobinets, 2017 [48]	LB	169	509 (291, 218)	Crossval	LR	264	27	24	194	0.907 (0.868-0.936)	0.890 (0.841-0.925)
Wang, 2017 [36]	PB	172	172 (79, 93)	Crossval	DCNN	55	24	15	78	0.696 (0.587-0.787)	0.839 (0.750-0.900)
Le, 2017 [52]	LB	364	913 (463, 450)	275 (139, 135)	multi-modal CNN	125	14	6	129	0.899 (0.837-0.939)	0.956 (0.905-0.980)
Kwon, 2018 [49]	LB	204	191 (36, 155)	Crossval	LASSO LR	35	5	9	90	0.875 (0.733-0.947)	0.909 (0.834-0.952)
Song, 2018 [50]	LB	195	547 (261, 286)	55 (23, 32)	DNN	20	3	3	29	0.870 (0.665-0.957)	0.906 (0.746-0.969)
Chen, 2019 [56]	PB	381	381 (182, 199)	155 (55, 60)	LR	55	1	1	59	0.982 (0.884-0.997)	0.983 (0.891-0.998)
Devine, 2019 [51]	LB	65	97 (81, 16)	Crossval	LR	61	20	2	14	0.753 (0.648-0.835)	0.875 (0.614-0.969)
Gholizadeh, 2019 [54]	LB	11	297 (161, 136)	Crossval	SVM	161	1	9	127	0.994 (0.958-0.999)	0.934 (0.878-0.965)
Ma, 2019 [58]	PB	81	81 (44, 37)	None	LR	42	2	5	32	0.955 (0.836-0.989)	0.865 (0.714-0.943)
Mazaheri, 2019 [53]	LB	67	170 (102, 68)	91 (52, 39)	CART	51	1	19	20	0.981 (0.876-0.997)	0.513 (0.360-0.664)
Qi, 2019 [57]	PB	199	199 (85, 114)	66 (28, 38)	LR	23	5	3	35	0.821 (0.636-0.924)	0.921 (0.782-0.974)
Zhang, 2019 [59]	PB	140	140 (60, 80)	Crossval	RF	14	6	5	22	0.700 (0.473-0.859)	0.815 (0.625-0.921)

<sup>a</sup>LB: lesion-based model; PB: patient-based model.

<sup>b</sup>PCa: prostate cancer.

<sup>c</sup>Crossval: cross-validation techniques.

<sup>d</sup>ML: machine learning.

<sup>e</sup>ANN: artificial neural networks; LDA: linear discriminant analysis; RF: random forest; SNCSAE: stacked nonnegativity constraint sparse autoencoders; LR: logistic regression; DCNN: deep convolutional neural networks; LASSO: least absolute shrinkage and selection operator; DNN: deep neural networks; SVM: support vector machine; CART: classification and regression tree.

<sup>f</sup>TP: true-positive.

<sup>g</sup>FN: false-negative.

<sup>h</sup>FP: false-positive.

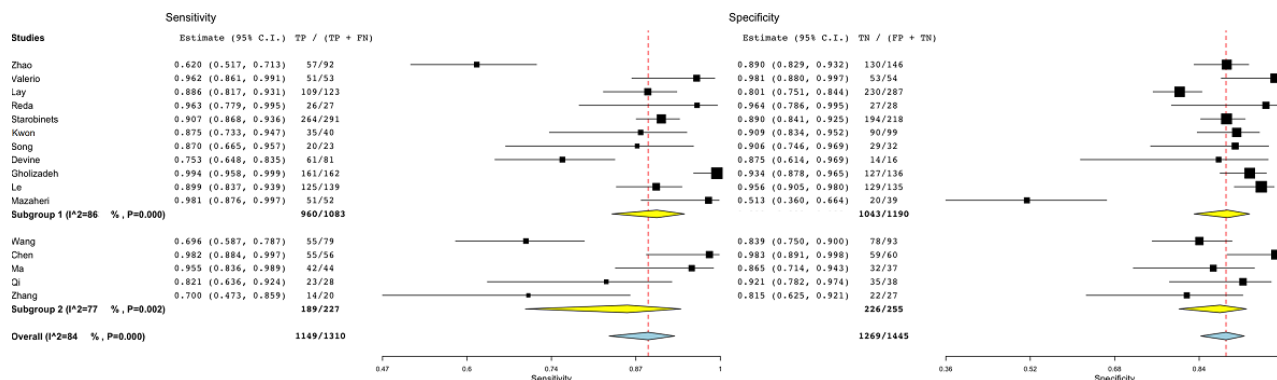
<sup>i</sup>TN: true-negative.

<sup>j</sup>Sen: sensitivity.

<sup>k</sup>Spe: specificity.

Figure 3 shows the subgroup analysis among studies that employed internal cross-validation techniques (subgroup 1) [44,50,52,53,56,57], split validation approaches (subgroup 2) [46-49,51,54,55,59], and no validation (subgroup 3) [45,58]. The heterogeneity for subgroups 1 and 2 was around 80% ( $P<.001$ ).

**Figure 2.** Subgroup analysis for the model-based covariate in radiomic studies. Subgroup 1: lesion-based models; subgroup 2: patient-based models. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



**Figure 3.** Subgroup analysis for the validation covariate in radiomic studies. Subgroup 1: internal cross-validation; subgroup 2: hold-out approach or external validation; subgroup 3: no validation. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.

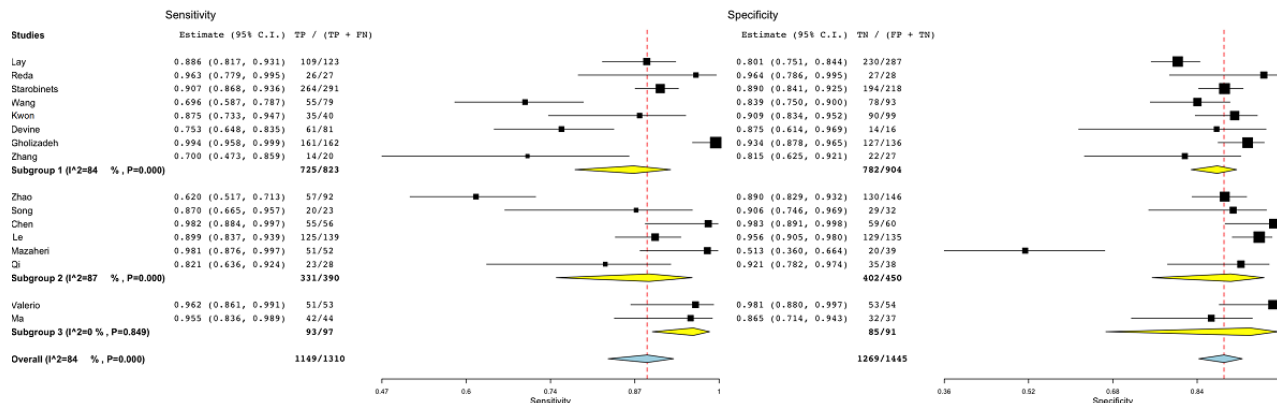


Figure 4 shows the subgroup analysis for regression-based models (subgroup 1) [45,48,49,51,56-58], tree-based models (subgroup 2) [46,53,59], and DL methods (subgroup 3) [44,47,50,52,55]. One study was not included [54], as it was the only study employing a support vector machine model. The heterogeneity among groups oscillated between 74% and 86% (subgroup 1:  $P=.001$ , subgroup 2:  $P=.01$ , subgroup 3:  $P<.001$ ).

The results of the subgroup analysis to discriminate among machine and deep learning methods are reported in Figure 5. Subgroup 1 included the studies [45,46,48,49,51,53,56-59] employing ML methods, whereas subgroup 2 comprised the studies [44,47,50,52,55] employing DL methods (based on artificial neural networks) such as convolutional neural networks and deep neural networks. The  $I^2$  statistics for subgroups 1 and 2 were 76% and 86% ( $P<.001$ ), respectively.

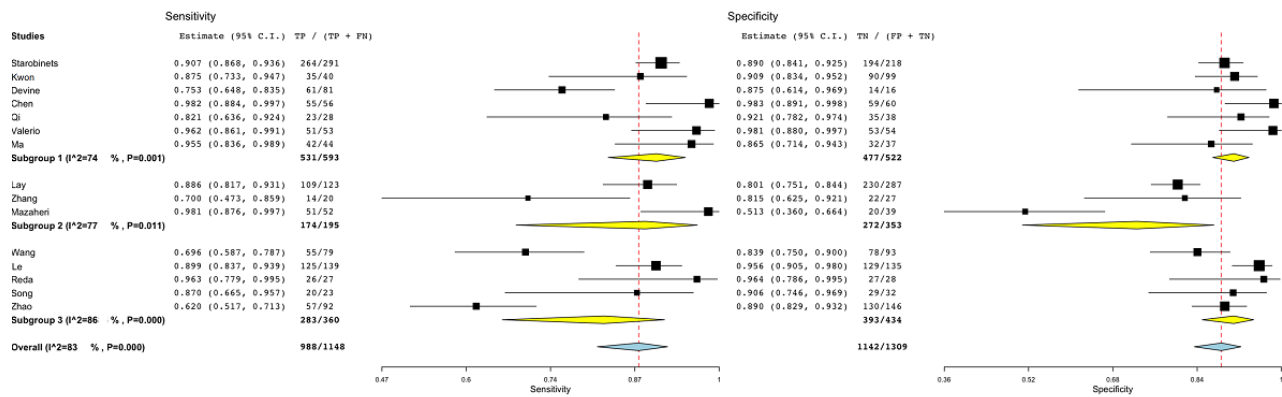
Figure 6 shows the subgroup analysis based on whether the studies employed a balanced or unbalanced data set. A data set was defined as unbalanced if it had more than 30% of the total observations in one specific class rather than the other (PCa and controls) and did not apply any correction on performance (eg, synthetic minority oversampling technique [SMOTE] or voting techniques). The heterogeneity of subgroup 1 [36,44,51,53] was around 58% ( $P=.005$ ). As a result, among the several covariates, the imbalance covariate was the only one by which the heterogeneity could be partially resolved.

Therefore, Devine et al [51], Wang et al [36], Mazaheri et al [53], and Zhao et al [44] were excluded from the coupled forest plot (Figure 7).

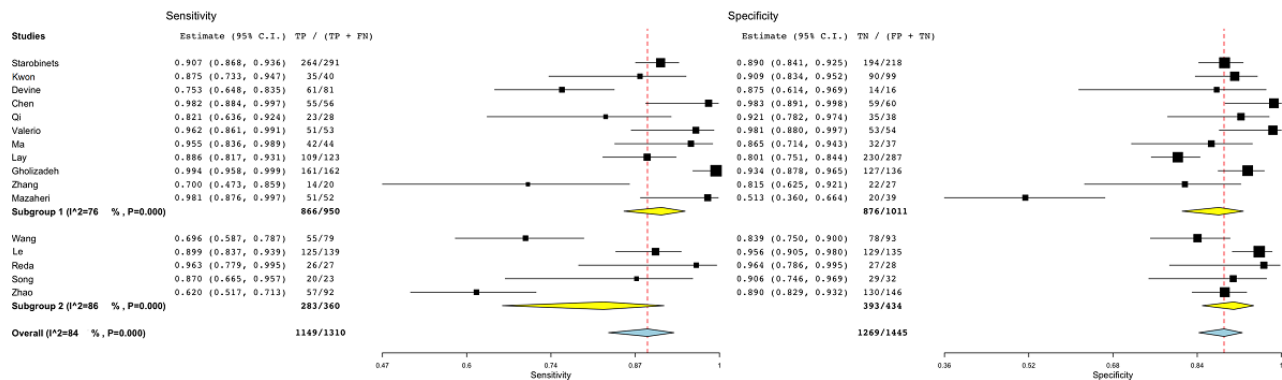
Figure 8 shows the HSROC curve for the studies employing balanced data sets to automatically detect PCa. The pooled sensitivity and specificity were 0.808 (95% CI 0.38-0.999) and 0.831 (95% CI 0.41-0.999), respectively.



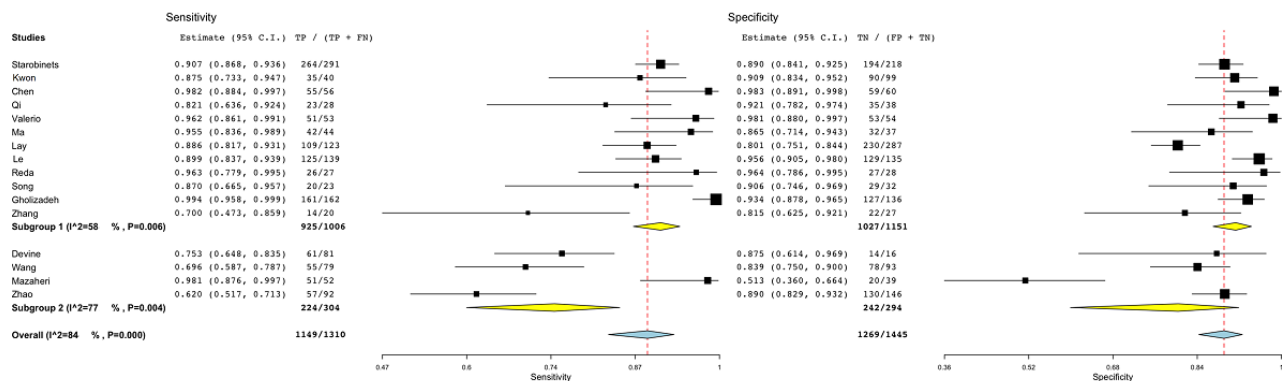
**Figure 4.** Subgroup analysis for the machine learning algorithm covariate in radiomic studies. Subgroup 1: regression-based models; subgroup 2: tree-based models; subgroup 3: deep learning methods. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



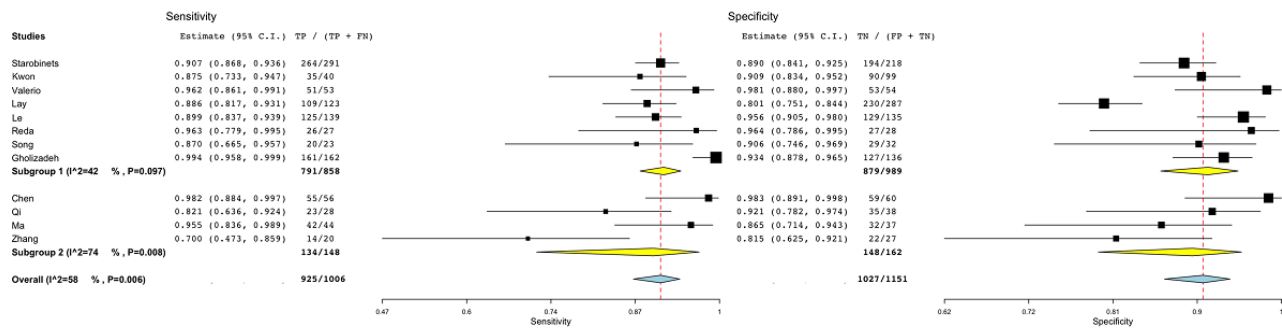
**Figure 5.** Subgroup analysis for the machine learning or deep learning covariate in radiomic studies. Subgroup 1: machine learning-based models; subgroup 2: deep learning methods. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



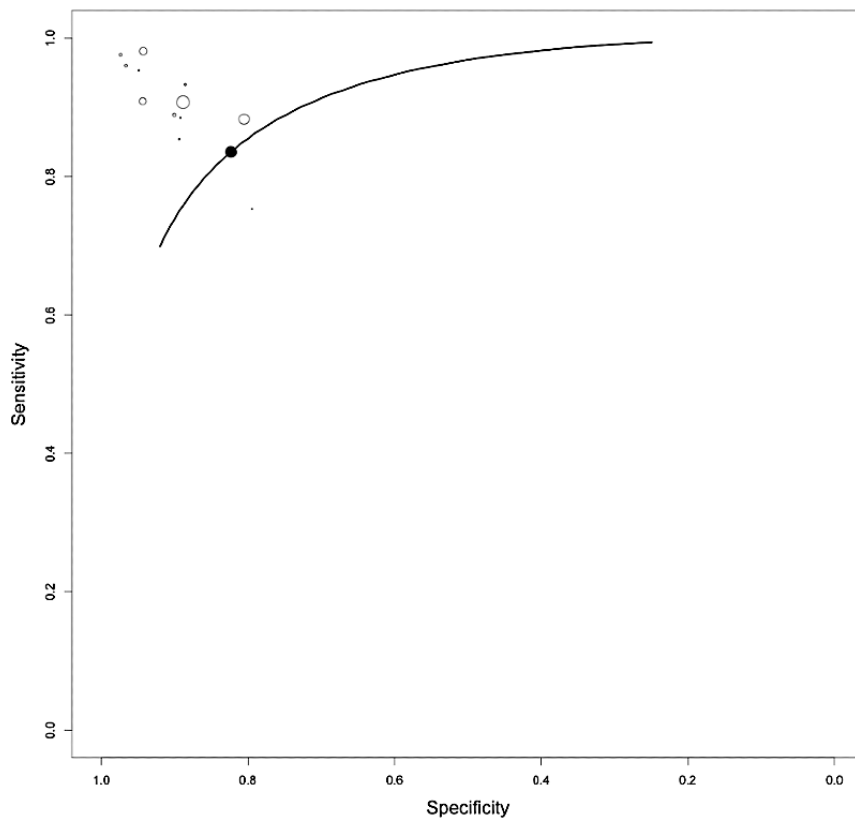
**Figure 6.** Subgroup analysis for the imbalance covariate in radiomic studies. Subgroup 1: balanced data sets; subgroup 2: unbalanced data sets. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



**Figure 7.** Subgroup analysis for the model-based covariate in a subset of radiomic studies. Subgroup 1: lesion-based models; subgroup 2: patient-based models. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



**Figure 8.** Overall hierarchical summary receiver operating characteristic curve (HSROC) for a subset of radiomic studies. HSROC was calculated for radiomic studies with low heterogeneity, excluding 4 studies [36,44,51,53].



**Genomic**

All the included studies for the genomic analysis are reported in Table 3. A total of 3221 independent samples were inspected from 14 studies and included in the meta-analysis, with sensitivity and specificity ranging from 0.67 to 0.95 and 0.15 to 0.97, respectively.

An HSROC model was assessed for all genomic studies (Figure S4 in Multimedia Appendix 1). The pooled sensitivity and specificity were 0.883 (95% CI 0.541-0.999) and 0.734 (95% CI 0.330-0.999), respectively.

The calculated heterogeneity values for the pooled sensitivity and specificity were 73% and 92% (P<.001), respectively;

therefore, a random-effects model was adopted to generate the coupled forest plots (Figure S5 in Multimedia Appendix 1).

**Subgroup Analysis**

To resolve this heterogeneity, subgroup analyses were conducted for several covariates. The subgroup analysis for model-based covariates is shown in Figure 9. Subgroup 1 included the studies [60,61] that used malignant lesions and benign-adjacent tissue from PCa patients. Subgroup 2 gathered those studies [15-17,62-69] that enrolled two distinct groups (PCa and controls) and employed a patient-based ML approach. The heterogeneity for subgroup 1 was greater than 80%, whereas for subgroup 2 it was around 60%. However, subgroup 2 only included 2 studies.

**Table 3.** Accuracy measures of genomic studies for the systematic review.<sup>a</sup>

Study, year	Model basis <sup>b</sup>	Predictor	Patients, n	Total sample (PCa+, PCa-) <sup>c</sup>	Crossval <sup>d</sup> /split/none	TP, <sup>e</sup> n	FN, <sup>f</sup> n	FP, <sup>g</sup> n	TN, <sup>h</sup> n	Sen <sup>i</sup> (lower-upper)	Spe <sup>j</sup> (lower-upper)
Donovan, 2015 [62]	PB	Urine	195	195 (89, 106)	None	80	9	84	22	0.899 (0.817-0.947)	0.208 (0.141-0.295)
Roberts, 2015 [16]	PB	Semen	66	66 (12, 54)	Crossval	11	1	32	20	0.917 (0.587-0.988)	0.385 (0.263-0.522)
Zhang, 2015 [63]	PB	Serum	580	580 (180, 400)	320 (120, 200)	84	36	5	195	0.7 (0.612-0.775)	0.975 (0.941-0.99)
Mengual, 2016 [64]	PB	Urine	224	224 (15, 73)	Crossval	116	35	12	61	0.768 (0.694-0.829)	0.836 (0.732-0.904)
Salido-Guadarrama, 2016 [15]	PB	Urine	143	143 (73, 70)	None	60	13	13	57	0.822 (0.717-0.894)	0.814 (0.706-0.889)
Dereziński, 2017 [17]	PB	Serum	89	89 (49, 40)	34 (19, 15)	13	6	0	15	0.675 (0.449-0.841)	0.969 (0.65-0.998)
Dereziński, 2017a [17]	PB	Urine	89	89 (49, 40)	34 (19, 15)	17	2	4	11	0.895 (0.663-0.974)	0.733 (0.467-0.896)
Kirby, 2017 [60]	LB	Tissue	101	398 (286, 112)	262 (213, 49)	180	33	4	45	0.845 (0.79-0.888)	0.918 (0.802-0.969)
Barceló, 2018 [65]	PB	Semen	42	42 (34, 18)	None	22	2	5	13	0.917 (0.721-0.979)	0.722 (0.481-0.879)
Amante, 2019 [66]	PB	Urine	91	91 (43, 48)	Crossval	40	3	5	43	0.93 (0.805-0.977)	0.896 (0.773-0.956)
Brikun, 2019 [67]	PB	Urine	94	94 (42, 52)	29 (13, 16)	12	1	5	11	0.923 (0.609-0.989)	0.687 (0.433-0.864)
Gao, 2019 [69]	PB	Urine	183	183 (108, 75)	77 (55, 22)	48	7	5	17	0.873 (0.756-0.938)	0.773 (0.556-0.902)
Patel, 2019 [61]	LB	Tissue	699	795 (699, 96)	242 (212, 30)	199	13	2	28	0.939 (0.897-0.964)	0.933 (0.769-0.983)
Santotoribio, 2019 [68]	PB	Serum	232	232 (32, 200)	None	30	2	58	142	0.937 (0.782-0.984)	0.71 (0.643-0.769)

<sup>a</sup>All studies employed regression-based models.

<sup>b</sup>LB: lesion-based model; PB: patient-based model.

<sup>c</sup>PCa: prostate cancer.

<sup>d</sup>Crossval: cross-validation techniques.

<sup>e</sup>TP: true-positive.

<sup>f</sup>FN: false-negative.

<sup>g</sup>FP: false-positive.

<sup>h</sup>TN: true-negative.

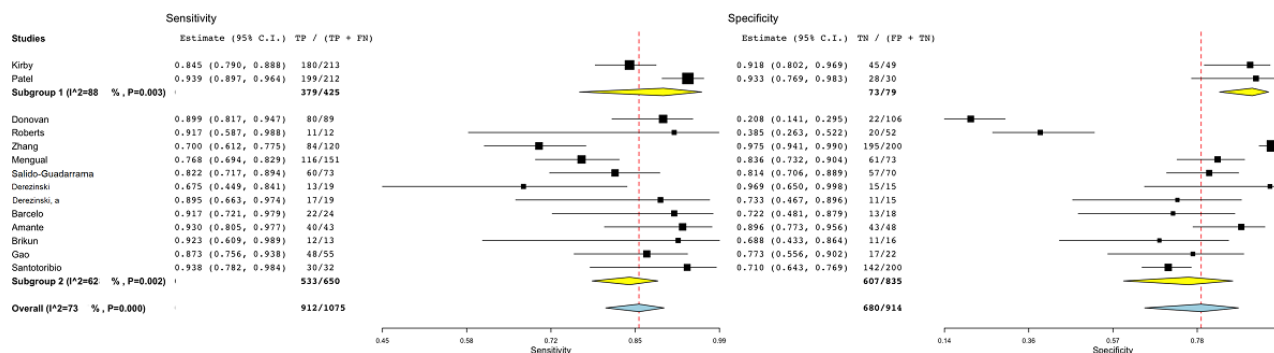
<sup>i</sup>Sen: sensitivity.

<sup>j</sup>Spe: specificity.

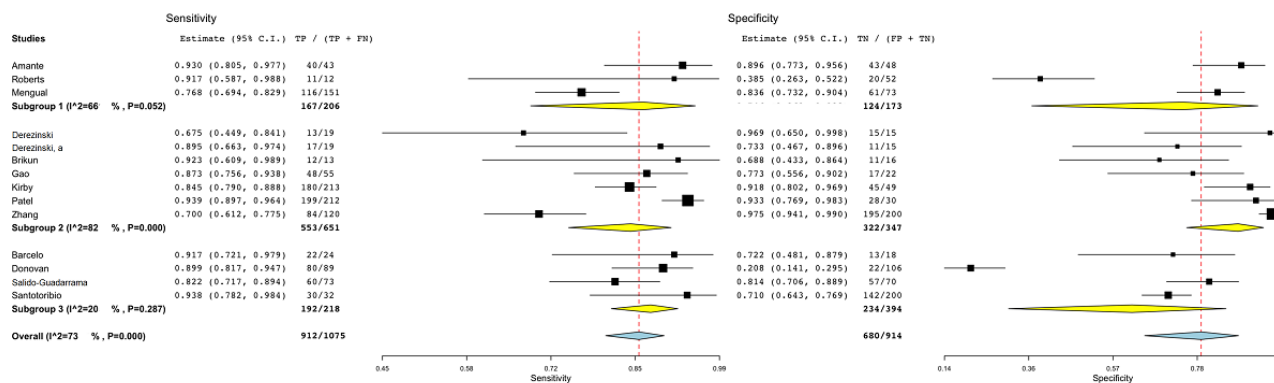
The subgroup analysis among studies that employed internal cross-validation techniques (subgroup 1) [16,64,66], split validation approaches (subgroup 2) [17,60,61,63,67,69], and

no validation (subgroup 3) [15,62,65,68] is shown in [Figure 10](#). For subgroups 1 and 2, the heterogeneity was greater than 50%. In subgroup 3, the heterogeneity was around 20%.

**Figure 9.** Subgroup analysis for the model-based covariate in genomic studies. Subgroup 1: lesion-based models; subgroup 2: patient-based models. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



**Figure 10.** Subgroup analysis for the validation covariate in genomic studies. Subgroup 1: internal cross-validation; subgroup 2: hold-out approach or external validation; subgroup 3: no validation. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



A subgroup analysis was also carried out based on the specimen used by the genomic studies (ie, urine [15,17,62,64,66,67,69], semen [16,65], serum [17,63,68], and tissue [60,61] biomarkers). The subgroup of studies investigating urine biomarkers to automatically detect PCa presented a lower heterogeneity than studies employing tissue and serum biomarkers and included more than 5 studies (Figure 11).

An inspection of ML algorithms among genomic studies was not possible because all the included studies employed a regression-based model (Table S4 in Multimedia Appendix 1).

Finally, the effect of using balanced or highly unbalanced data sets in ML approaches was investigated (Figure 12). Seven studies were included in subgroup 2, as they employed highly unbalanced data sets. The heterogeneity of subgroup 1 was around 36%, whereas subgroup 2 showed a high heterogeneity ( $I^2=84%$ ,  $P<.001$ ).

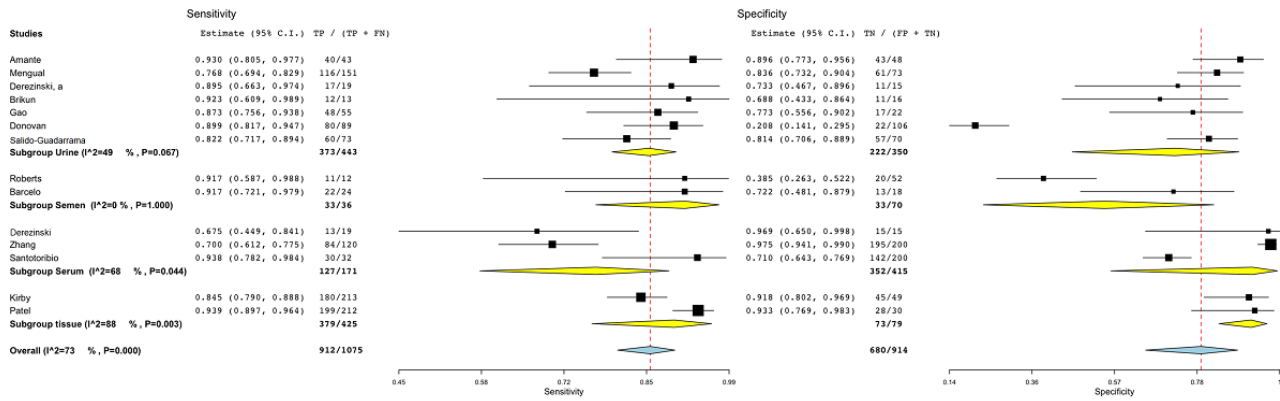
As a result, among several covariates, the imbalance covariate was the only one by which the heterogeneity could be partially resolved for more than 5 studies.

By inspecting Figure 12, Donovan et al [62] presented a very low value for specificity; this was due to the fact that they fixed the sensitivity threshold value at 90%.

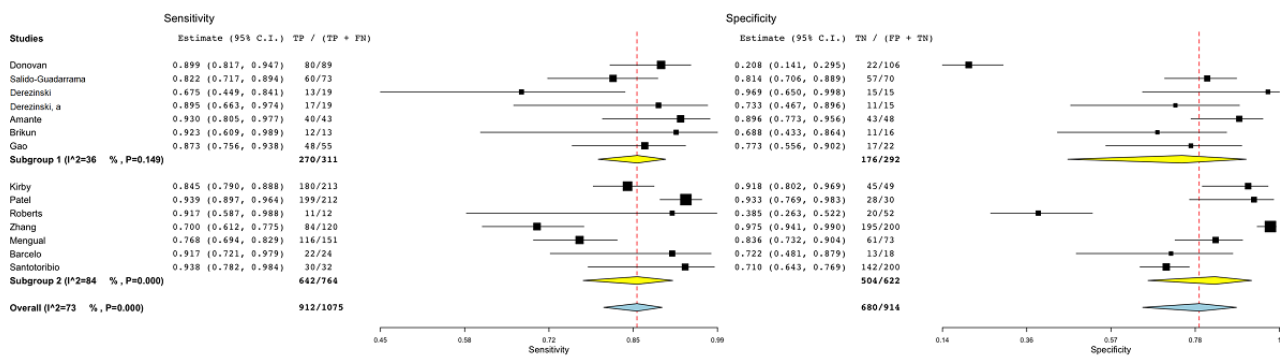
Five studies employing urine specimens and balanced data sets showed a very low heterogeneity (Figure 13) [15,17,66,67,69].

The HSROC curve for the studies employing balanced data sets to automatically detect PCa via urine biomarkers is shown in Figure 14. The pooled sensitivity and specificity were 0.812 (95% CI 0.577-0.999) and 0.8101 (95% CI 0.544-0.999), respectively.

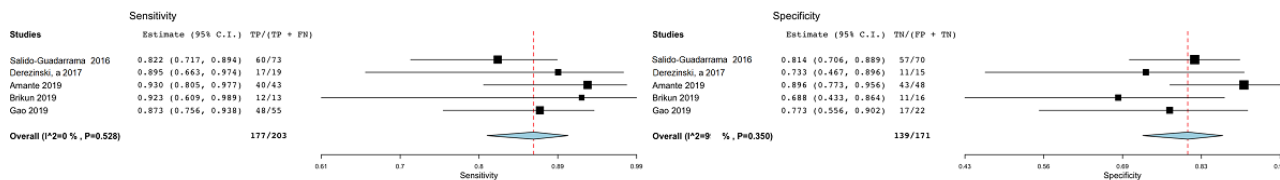
**Figure 11.** Subgroup analysis for the predictor covariate in genomic studies. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



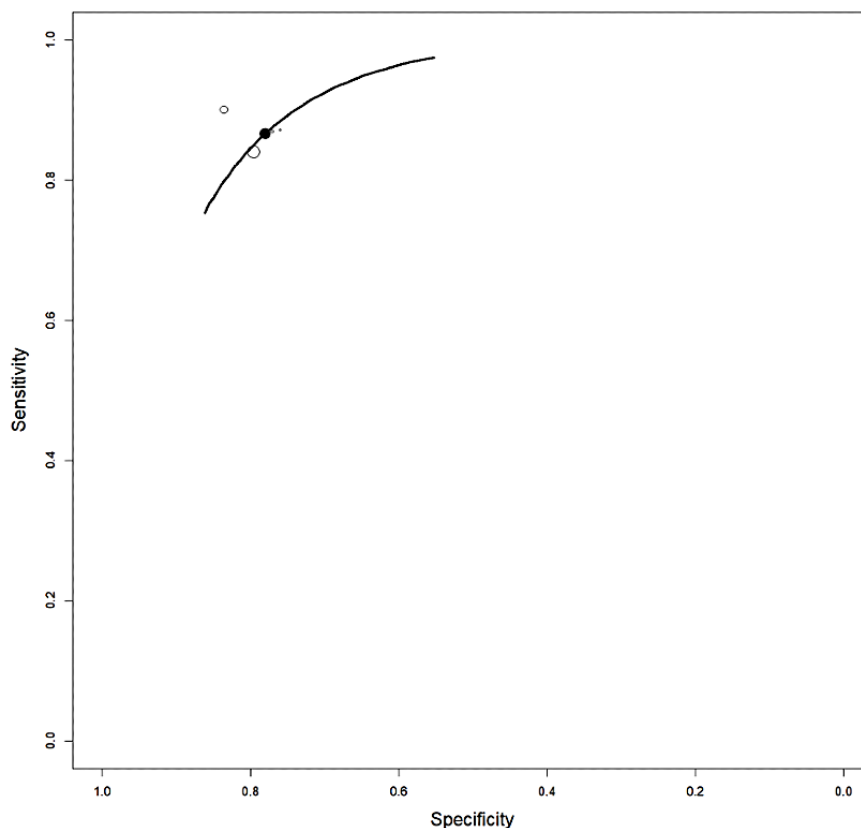
**Figure 12.** Subgroup analysis for the imbalance covariate in genomic studies. Subgroup 1: balanced data sets; subgroup 2: unbalanced data sets. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



**Figure 13.** Coupled forest plots for balanced studies. The included studies investigated urine specimens. FN: false-negative; FP: false-positive; TN: true-negative; TP: true-positive.



**Figure 14.** Hierarchical summary receiver operating characteristic curve (HSROC) for a subset of genomic studies. HSROC was calculated for genomic studies with low heterogeneity [15,17,66,67,69].



## Discussion

### Principal Findings

This paper presents the results of a systematic literature review with meta-analysis of articles investigating machine learning algorithms to detect PCa via radiomic or genomic analysis. One research focus of this study was on clearly evaluating how the implementation of different ML approaches impacts the clinical results. At this stage, due to the high heterogeneity of methods and tools employed in the existing literature, no clear clinical relevance on the use of ML for PCa can be drawn from this study. This review shows that ML has helped to improve the diagnostic performance of the detection of PCa, but challenges still remain for clinical applicability of such methods, and more research is needed. The presented literature aims to help in building an ML system that is robust and computationally efficient to assist clinicians in the diagnosis of PCa via radiomic and genomic biomarkers.

In this review, 37 studies were shortlisted, and 29 studies were included in a meta-analysis. All patients were diagnosed with PCa by biopsy. However, not all the included studies reported full information on the methods used to carry out biopsy (eg, direct MRI-guided, cognitive fusion, or MRI-TRUS fusion biopsy).

In the radiomic and genomic meta-analysis, 16 and 14 studies were included, respectively. Heterogeneity among radiomic and genomic studies was 84% and 73%, respectively. This was expected, as ML methods are usually regarded as black boxes, and the consideration of all possible transformations is onerous.

Moreover, there are no clear guidelines on how to develop AI approaches for medical studies, even though a few recommendations have been summarized by Foster et al [24] and Chen et al [25]. Another font of heterogeneity in radiomic studies may be due to the inclusion of PI-RADS score 3 and Gleason score 3+3 lesions, which are equivocal and should be disregarded in classification processes.

To partially solve the heterogeneity for the included studies, subgroup analyses were conducted based on several covariates. In the field of ML, applications where repeated measures or records have been captured on each subject can affect the overall performance. In most studies, the main aim is to predict if a given subject is “sick” or a “control” subject. In these applications, each subject has a single label type (eg, “sick” or control case). Nonetheless, there are other classification problems where each subject can have multiple labels. For instance, multiple lesions can be extracted from the same subject, and the control part can be represented by the benign-adjacent prostate lesion. It has been demonstrated that this phenomenon, known as identity confounding, can cause discrepancy in classification performance [70,71]. Therefore, the studies included in the meta-analysis were investigated to determine whether they explored patient- or lesion-based models. A patient-based model could be defined as a model that is developed and assessed in a “subject-wise” fashion, where all the records of each subject are considered as a group in the training and testing set and when assessing the model performance; conversely, a lesion-based model could be defined as a model that is developed and assessed in a “record-wise” fashion, where each measurement or record contributes to both

the training and test sets and when assessing the model performance [70].

In both radiomic and genomic studies, patient-based models presented lower heterogeneity and performance than lesion-based models; this could be due to the fact that lesion-based models employed a bigger size sample, but the models may be overfit due to repeated measures.

A second important covariate to examine in ML problems is the data set construction. In particular, the data set is usually divided into training and testing sets in order to reduce overfitting problems [70,71]. The training set is often further split into a training set and a validation set, which is used to update model parameters. At least one procedure of internal or external validation is required in ML approaches. Cross-validation techniques are preferred if availability of data is not a problem. It is also strongly suggested to retrain on a subset of data or use an independent data set for external testing. Therefore, “validation approach” was used as a covariate in subgroup analysis. Validation approaches were divided into cross-validation, hold-out approach (split) or external validation, and no validation. In both radiomic and genomic analysis, studies employing cross-validation techniques and hold-out approaches had very high heterogeneity and similar performances among them. High heterogeneity may be due to the different cross-validation techniques used (eg, bootstrapping [16,40,52], Monte Carlo cross-validation [17]) or the choice of number of folders used in cross-validation methods; if an external data set was used [52,60,61,63], differences in the study protocols may have increased the bias among studies. Moreover, few studies in radiomic [50,53,57,59] and genomic [17,67] analysis employed both cross-validation and external testing. Studies employing no validation showed very low heterogeneity (only 2 studies in radiomic analysis), which may be due to the absence of other confounding variables, and high performances may be due to overfitting problems. A lower specificity was only noted in genomic analysis; this was due to Donovan et al [62], which used a fixed threshold for sensitivity at 90%.

Different ML approaches were also investigated among radiomic studies as a possible covariate factor. There were no relevant differences in heterogeneity or performance among subgroups (Figure 4). All genomic studies employed regression-based models. In fact, one limitation of the genomic studies was that none of the selected studies explored the potential of ML techniques at full capacity. Subgroup analysis was also conducted among radiomic studies employing ML or DL (ie, based on artificial neural networks) approaches. As expected, heterogeneity among DL studies was higher than among the studies employing other ML approaches to detect PCa. This could be mainly due to the high complexity of DL methods and hyperparameters. Moreover, DL approaches showed lower performance due to the small sample sizes used; they need large volumes of data to automatically identify patterns and achieve high performance.

The imbalance covariate was crucial in this study. Unbalanced and small data sets are very common in the medical field, and ML algorithms tend to produce unsatisfactory classifiers when handled with imbalanced data sets. Therefore, several techniques

to overcome this problem have been proposed over time [72]. In this review, none of the studies included in the subgroup of unbalanced data sets had used any techniques to overcome the problem. Only one study [56] used SMOTE, but it did not employ a highly unbalanced data set.

For radiomic studies, after excluding studies that employed highly unbalanced data sets, the heterogeneity was less than 50%. The final pooled sensitivity and specificity for the use of mpMRI were 0.808 (95% CI 0.38-0.999) and 0.831 (95% CI 0.41-0.999), respectively.

For genomic studies, the heterogeneity dropped to 36% and reached a value close to zero when Donovan et al [62] was excluded because they fixed a threshold of 90% for sensitivity. The final pooled sensitivity and specificity were 0.812 (95% CI 0.577-0.999) and 0.8101 (95% CI 0.544-0.999), respectively. The predictor used to estimate the final pooled sensitivity and specificity was urine specimen.

Only 4 studies [18,39-41] investigating clinically based models were identified through the search. All the included studies adopted internal validation techniques (3 cross-validation [39-41] and 1 internal split validation [18]). Two studies [40,41] employed regression-based models, one [39] employed a tree-based model, and lastly, one employed a DL approach [18]. Heterogeneity was very high among them ( $I^2=96%$ ,  $P=.01$ ) due to different sample sizes and diversity of predictors. However, contributions from genomic and imaging biomarkers should be considered to improve the overall performance of the clinically based diagnostic models.

Comparison among genomic and radiomic studies was not possible because they describe two different but complementary prospective approaches to the disease. However, the pooled sensitivity and specificity for both mpMRI and urine biomarkers were around 80%, showing them to be promising biomarkers in the detection of PCa via ML in clinical practice. The use of mpMRI has shown great diagnostic potential [73]; however, its analysis and interpretation are quite challenging, and there is not a consensus on how to optimally extract significant information. On the other side, genomic analyses have significantly increased our understanding of PCa and greatly improved patient risk classification, thus impacting treatment decision making. Therefore, a new prospective approach is the integration of radiomic and genomic signatures, commonly known as radiogenomics [74-76], in order to improve the overall performance of diagnostic tools to automatically detect PCa. In the existing literature, only a few studies have investigated “radiophenotypes” to complement existing validated clinical and genomic risk stratification biomarkers [77-79].

In this scenario, a typical ML postprocessing pipeline for radiomic and genomic analysis to automatically detect PCa may be constituted of a few crucial steps. In the case of radiomic studies, a common pipeline may be constituted of (1) examination of mpMRI; (2) image segmentation through the delineation of ROIs or VOIs, which can include whole gland volume, a specific zone, and one or multiple lesions, which should be explicitly specified in the manuscript; (3) image preprocessing; (4) filtering; (5) feature extraction; (6) integration

of radiomic data with clinical data, genomic data, or both; (7) feature selection in relation to the target class; and (8) algorithm training, validation, and testing. Alternatively, a DL approach would only require the examination of the images and annotation of the ROIs or VOIs of the whole image, according to the desired classification output.

The image processing pipeline should be carefully described in the manuscripts, and the spatial coregistration of DWIs is a critical factor in the correct analysis of diffusion tensor imaging data, which has often been used as a predictor of PCa diagnosis. Moreover, the use of endorectal coil can cause high deformation of the prostate compared with other coils and may not provide adequate MR image quality [80]. Therefore, further processing of the images should also be considered, especially when the study is multicenter and different protocols have been adopted.

Due to the high heterogeneity of genomic studies, a standard pipeline configuration could be structured into (1) missing value management; (2) filtering to remove low-variance features; (3) data normalization due to data coming from heterogeneous formats; (4) a feature selection step to remove irrelevant features due to the high dimension of data; (5) dealing with class imbalance distribution present in this type of large-scale data set; and (6) algorithm training, validation, and testing. Alternatively, a DL approach would handle filtering and feature selection to generate handcrafted features. Deep learning is a powerful tool to integrate different “omics” and increase the computational power of diagnostic tools.

Further general recommendations on how to avoid bias and pitfalls in applying ML to medical problems are as follows: (1) in the case of multicenter studies, it is recommended to use batch effect approaches to prevent any bias due to different study protocols and feature normalization procedures to reduce within-subject bias [81]; and (2) for classifier performance, it is necessary to report if any threshold has been used to identify sensitivity and specificity and whether the performance was reported on patient-based or lesion-based data sets.

### Limitations

Our study presents several limitations. Some variability still remains due to the actual thresholds between studies. However, the multiple hierarchical model accounts for between- and

within-subject variability among studies, including threshold effects. Another factor that could have affected the heterogeneity among studies is the use of different predictors among radiomic and genomic studies. Moreover, several studies reported little or incomplete information on the parameters used to develop ML models. Therefore, the number of parameters that are estimated by each technique was not investigated as a possible source of heterogeneity among studies. Additional heterogeneity in the observed results is due to the variability of calibration differences between equipment and differences between readers or observers, as well as variation in the implementation of tests. Another possible bias may be due to the preprocessing techniques on the extracted data and feature selection and feature normalization methods.

We limited the search to English-only studies; although this is common in systematic reviews, this exclusion criterion could have reduced the generalizability of the findings. However, the extent and effects of language bias have recently diminished because of a shift toward publication of studies in English [82]. At this stage, we also excluded PCa risk stratification studies to reduce bias and heterogeneity among studies, but further investigation on the use of ML methods to assess risk stratification biomarkers could give a comparative perspective on the treatment selection.

Finally, publication bias was not assessed in our analysis, as there are currently no statistically adequate models in the field of meta-analysis of diagnostic test accuracy [29].

### Conclusion

ML has shown its potential to empower clinicians in the detection of prostate cancer. The accuracy of ML algorithms for diagnosis of PCa was considered acceptable, in terms of heterogeneity, for 12 radiomic studies investigating mpMRI and 5 genomic studies using urine biomarkers.

However, given the limitations indicated in our study, further well-designed studies are warranted to extend the potential use of ML algorithms to clinical settings. Recommendations on the use of these techniques were also provided to help researchers to design robust studies aiming to identify radiomic and genomic biomarkers to detect cancer.

---

### Acknowledgments

This work was supported by “Progetti di Ricerca Corrente” funded by the Italian Ministry of Health.

---

### Authors' Contributions

RC, MF, and CC collected the data. All authors contributed to project development, data analysis, and the writing and editing of the manuscript.

---

### Conflicts of Interest

None declared.

---

### Multimedia Appendix 1

Supplementary material.

[\[DOCX File, 984 KB-Multimedia Appendix 1\]](#)



## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018 Nov;68(6):394-424 [FREE Full text] [doi: [10.3322/caac.21492](https://doi.org/10.3322/caac.21492)] [Medline: [30207593](https://pubmed.ncbi.nlm.nih.gov/30207593/)]
2. Grönberg H. Prostate cancer epidemiology. *Lancet* 2003 Mar 08;361(9360):859-864. [doi: [10.1016/S0140-6736\(03\)12713-4](https://doi.org/10.1016/S0140-6736(03)12713-4)] [Medline: [12642065](https://pubmed.ncbi.nlm.nih.gov/12642065/)]
3. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol* 2017 Apr;71(4):618-629. [doi: [10.1016/j.eururo.2016.08.003](https://doi.org/10.1016/j.eururo.2016.08.003)] [Medline: [27568654](https://pubmed.ncbi.nlm.nih.gov/27568654/)]
4. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp* 2018 Nov 14;2(1):36 [FREE Full text] [doi: [10.1186/s41747-018-0068-z](https://doi.org/10.1186/s41747-018-0068-z)] [Medline: [30426318](https://pubmed.ncbi.nlm.nih.gov/30426318/)]
5. Yao S, Jiang H, Song B. Radiomics in prostate cancer: basic concepts and current state-of-the-art. *Chin J Acad Radiol* 2019 Nov 12;2(3-4):47-55. [doi: [10.1007/s42058-019-00020-3](https://doi.org/10.1007/s42058-019-00020-3)]
6. Cuocolo R, Cipullo MB, Stanzione A, Ugga L, Romeo V, Radice L, et al. Machine learning applications in prostate cancer magnetic resonance imaging. *Eur Radiol Exp* 2019 Aug 07;3(1):35 [FREE Full text] [doi: [10.1186/s41747-019-0109-2](https://doi.org/10.1186/s41747-019-0109-2)] [Medline: [31392526](https://pubmed.ncbi.nlm.nih.gov/31392526/)]
7. Cantiello F, Russo GI, Kaufmann S, Cacciamani G, Crocero F, Ferro M, et al. Role of multiparametric magnetic resonance imaging for patients under active surveillance for prostate cancer: a systematic review with diagnostic meta-analysis. *Prostate Cancer Prostatic Dis* 2019 May;22(2):206-220. [doi: [10.1038/s41391-018-0113-2](https://doi.org/10.1038/s41391-018-0113-2)] [Medline: [30487646](https://pubmed.ncbi.nlm.nih.gov/30487646/)]
8. Weinreb J, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *Eur Urol* 2016 Jan;69(1):16-40 [FREE Full text] [doi: [10.1016/j.eururo.2015.08.052](https://doi.org/10.1016/j.eururo.2015.08.052)] [Medline: [26427566](https://pubmed.ncbi.nlm.nih.gov/26427566/)]
9. Chin L, Andersen JN, Futreal PA. Cancer genomics: from discovery science to personalized medicine. *Nat Med* 2011 Mar;17(3):297-303. [doi: [10.1038/nm.2323](https://doi.org/10.1038/nm.2323)] [Medline: [21383744](https://pubmed.ncbi.nlm.nih.gov/21383744/)]
10. Kornberg Z, Cooperberg MR, Spratt DE, Feng FY. Genomic biomarkers in prostate cancer. *Transl Androl Urol* 2018 Jun;7(3):459-471 [FREE Full text] [doi: [10.21037/tau.2018.06.02](https://doi.org/10.21037/tau.2018.06.02)] [Medline: [30050804](https://pubmed.ncbi.nlm.nih.gov/30050804/)]
11. Gleason DF. Classification of prostatic carcinomas. *Cancer Chemother Rep* 1966 Mar;50(3):125-128. [Medline: [5948714](https://pubmed.ncbi.nlm.nih.gov/5948714/)]
12. Jiandani D, Randhawa A, Brown RE, Hamilton R, Matthew AG, Kuk JL, et al. The effect of bicycling on PSA levels: a systematic review and meta-analysis. *Prostate Cancer Prostatic Dis* 2015 Sep;18(3):208-212. [doi: [10.1038/pcan.2015.16](https://doi.org/10.1038/pcan.2015.16)] [Medline: [25939515](https://pubmed.ncbi.nlm.nih.gov/25939515/)]
13. Kundu SD, Roehl KA, Yu X, Antenor JAV, Suarez BK, Catalona WJ. Prostate specific antigen density correlates with features of prostate cancer aggressiveness. *J Urol* 2007 Feb;177(2):505-509. [doi: [10.1016/j.juro.2006.09.039](https://doi.org/10.1016/j.juro.2006.09.039)] [Medline: [17222621](https://pubmed.ncbi.nlm.nih.gov/17222621/)]
14. Fang Y, Gao W. Roles of microRNAs during prostatic tumorigenesis and tumor progression. *Oncogene* 2014 Jan 09;33(2):135-147. [doi: [10.1038/onc.2013.54](https://doi.org/10.1038/onc.2013.54)] [Medline: [23455326](https://pubmed.ncbi.nlm.nih.gov/23455326/)]
15. Salido-Guadarrama AI, Morales-Montor JG, Rangel-Escareño C, Langley E, Peralta-Zaragoza O, Cruz Colin JL, et al. Urinary microRNA-based signature improves accuracy of detection of clinically relevant prostate cancer within the prostate-specific antigen grey zone. *Mol Med Rep* 2016 Jun;13(6):4549-4560 [FREE Full text] [doi: [10.3892/mmr.2016.5095](https://doi.org/10.3892/mmr.2016.5095)] [Medline: [27081843](https://pubmed.ncbi.nlm.nih.gov/27081843/)]
16. Roberts MJ, Chow CWK, Schirra HJ, Richards R, Buck M, Selth LA, et al. Diagnostic performance of expression of PCA3, Hepsin and miR biomarkers in ejaculate in combination with serum PSA for the detection of prostate cancer. *Prostate* 2015 Apr 01;75(5):539-549. [doi: [10.1002/pros.22942](https://doi.org/10.1002/pros.22942)] [Medline: [25597828](https://pubmed.ncbi.nlm.nih.gov/25597828/)]
17. Dereziński P, Klupczynska A, Sawicki W, Pałka JA, Kokot ZJ. Amino Acid Profiles of Serum and Urine in Search for Prostate Cancer Biomarkers: a Pilot Study. *Int J Med Sci* 2017;14(1):1-12 [FREE Full text] [doi: [10.7150/ijms.15783](https://doi.org/10.7150/ijms.15783)] [Medline: [28138303](https://pubmed.ncbi.nlm.nih.gov/28138303/)]
18. Roffman DA, Hart GR, Leapman MS, Yu JB, Guo FL, Ali I, et al. Development and Validation of a Multiparameterized Artificial Neural Network for Prostate Cancer Risk Prediction and Stratification. *JCO Clin Cancer Inform* 2018 Dec;2:1-10 [FREE Full text] [doi: [10.1200/CCI.17.00119](https://doi.org/10.1200/CCI.17.00119)] [Medline: [30652591](https://pubmed.ncbi.nlm.nih.gov/30652591/)]
19. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *J Arthroplasty* 2018 Aug;33(8):2358-2361. [doi: [10.1016/j.arth.2018.02.067](https://doi.org/10.1016/j.arth.2018.02.067)] [Medline: [29656964](https://pubmed.ncbi.nlm.nih.gov/29656964/)]
20. Goldenberg SL, Nir G, Salcudean SE. A new era: artificial intelligence and machine learning in prostate cancer. *Nat Rev Urol* 2019 Jul;16(7):391-403. [doi: [10.1038/s41585-019-0193-3](https://doi.org/10.1038/s41585-019-0193-3)] [Medline: [31092914](https://pubmed.ncbi.nlm.nih.gov/31092914/)]
21. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. *NPJ Digit Med* 2020;3:126 [FREE Full text] [doi: [10.1038/s41746-020-00333-z](https://doi.org/10.1038/s41746-020-00333-z)] [Medline: [33043150](https://pubmed.ncbi.nlm.nih.gov/33043150/)]
22. Trivizakis E, Papadakis G, Souglakos I, Papanikolaou N, Koumakis L, Spandidos D, et al. Artificial intelligence radiogenomics for advancing precision and effectiveness in oncologic care (Review). *Int J Oncol* 2020 Jul;57(1):43-53 [FREE Full text] [doi: [10.3892/ijo.2020.5063](https://doi.org/10.3892/ijo.2020.5063)] [Medline: [32467997](https://pubmed.ncbi.nlm.nih.gov/32467997/)]

23. Zitnik M, Nguyen F, Wang B, Leskovec J, Goldenberg A, Hoffman MM. Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. *Inf Fusion* 2019 Oct;50:71-91 [FREE Full text] [doi: [10.1016/j.inffus.2018.09.012](https://doi.org/10.1016/j.inffus.2018.09.012)] [Medline: [30467459](https://pubmed.ncbi.nlm.nih.gov/30467459/)]
24. Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research - commentary. *Biomed Eng Online* 2014 Jul 05;13:94 [FREE Full text] [doi: [10.1186/1475-925X-13-94](https://doi.org/10.1186/1475-925X-13-94)] [Medline: [24998888](https://pubmed.ncbi.nlm.nih.gov/24998888/)]
25. Chen PC, Liu Y, Peng L. How to develop machine learning models for healthcare. *Nat Mater* 2019 May;18(5):410-414. [doi: [10.1038/s41563-019-0345-0](https://doi.org/10.1038/s41563-019-0345-0)] [Medline: [31000806](https://pubmed.ncbi.nlm.nih.gov/31000806/)]
26. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. *J Med Internet Res* 2016 Dec 16;18(12):e323 [FREE Full text] [doi: [10.2196/jmir.5870](https://doi.org/10.2196/jmir.5870)] [Medline: [27986644](https://pubmed.ncbi.nlm.nih.gov/27986644/)]
27. Sandhu S, Lin AL, Brajer N, Sperling J, Ratliff W, Bedoya AD, et al. Integrating a Machine Learning System Into Clinical Workflows: Qualitative Study. *J Med Internet Res* 2020 Nov 19;22(11):e22421 [FREE Full text] [doi: [10.2196/22421](https://doi.org/10.2196/22421)] [Medline: [33211015](https://pubmed.ncbi.nlm.nih.gov/33211015/)]
28. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015 Jan 01;4:1 [FREE Full text] [doi: [10.1186/2046-4053-4-1](https://doi.org/10.1186/2046-4053-4-1)] [Medline: [25554246](https://pubmed.ncbi.nlm.nih.gov/25554246/)]
29. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)] [Medline: [22007046](https://pubmed.ncbi.nlm.nih.gov/22007046/)]
30. Viswanath SE, Bloch NB, Chappelw JC, Toth R, Rofsky NM, Genega EM, et al. Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 Tesla endorectal, in vivo T2-weighted MR imagery. *J Magn Reson Imaging* 2012 Jul;36(1):213-224 [FREE Full text] [doi: [10.1002/jmri.23618](https://doi.org/10.1002/jmri.23618)] [Medline: [22337003](https://pubmed.ncbi.nlm.nih.gov/22337003/)]
31. Monti S, Brancato V, Di Costanzo G, Basso L, Puglia M, Ragozzino A, et al. Multiparametric MRI for Prostate Cancer Detection: New Insights into the Combined Use of a Radiomic Approach with Advanced Acquisition Protocol. *Cancers* 2020 Feb 07;12(2):390. [doi: [10.3390/cancers12020390](https://doi.org/10.3390/cancers12020390)]
32. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* 2004 May 15;23(9):1351-1375. [doi: [10.1002/sim.1761](https://doi.org/10.1002/sim.1761)] [Medline: [15116347](https://pubmed.ncbi.nlm.nih.gov/15116347/)]
33. Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part II. *Statistical Methods of Meta-Analysis*. *Korean J Radiol* 2015;16(6):1188-1196 [FREE Full text] [doi: [10.3348/kjr.2015.16.6.1188](https://doi.org/10.3348/kjr.2015.16.6.1188)] [Medline: [26576107](https://pubmed.ncbi.nlm.nih.gov/26576107/)]
34. Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: A tutorial. *Clinical Epidemiology and Global Health* 2019 Jun;7(2):192-198. [doi: [10.1016/j.cegh.2018.05.005](https://doi.org/10.1016/j.cegh.2018.05.005)]
35. Bonekamp D, Kohl S, Wiesenfarth M, Schelb P, Radtke JP, Götz M, et al. Radiomic Machine Learning for Characterization of Prostate Lesions with MRI: Comparison to ADC Values. *Radiology* 2018 Oct;289(1):128-137. [doi: [10.1148/radiol.2018173064](https://doi.org/10.1148/radiol.2018173064)] [Medline: [30063191](https://pubmed.ncbi.nlm.nih.gov/30063191/)]
36. Wang J, Wu C, Bao M, Zhang J, Wang X, Zhang Y. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol* 2017 Oct;27(10):4082-4090. [doi: [10.1007/s00330-017-4800-5](https://doi.org/10.1007/s00330-017-4800-5)] [Medline: [28374077](https://pubmed.ncbi.nlm.nih.gov/28374077/)]
37. Wu M, Krishna S, Thornhill RE, Flood TA, McInnes MD, Schieda N. Transition zone prostate cancer: Logistic regression and machine-learning models of quantitative ADC, shape and texture features are highly accurate for diagnosis. *J Magn Reson Imaging* 2019 Sep;50(3):940-950. [doi: [10.1002/jmri.26674](https://doi.org/10.1002/jmri.26674)] [Medline: [30701625](https://pubmed.ncbi.nlm.nih.gov/30701625/)]
38. Yang X, Liu C, Wang Z, Yang J, Min HL, Wang L, et al. Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI. *Med Image Anal* 2017 Dec;42:212-227. [doi: [10.1016/j.media.2017.08.006](https://doi.org/10.1016/j.media.2017.08.006)] [Medline: [28850876](https://pubmed.ncbi.nlm.nih.gov/28850876/)]
39. Xiao L, Chen P, Gou Z, Li Y, Li M, Xiang L, et al. Prostate cancer prediction using the random forest algorithm that takes into account transrectal ultrasound findings, age, and serum levels of prostate-specific antigen. *Asian J Androl* 2017;19(5):586-590 [FREE Full text] [doi: [10.4103/1008-682X.186884](https://doi.org/10.4103/1008-682X.186884)] [Medline: [27586028](https://pubmed.ncbi.nlm.nih.gov/27586028/)]
40. Nowroozi MR, Amini S, Kasaeian A, Zavarehei MJ, Eshraghian MR, Ayati M. Development, validation and comparison of two nomograms predicting prostate cancer at initial 12-core biopsy. *Asia Pac J Clin Oncol* 2016 Jun;12(2):e289-e297. [doi: [10.1111/ajco.12186](https://doi.org/10.1111/ajco.12186)] [Medline: [24684767](https://pubmed.ncbi.nlm.nih.gov/24684767/)]
41. Huang C, Song G, Wang H, Ji G, Li J, Chen Y, et al. MultiParametric Magnetic Resonance Imaging-Based Nomogram for Predicting Prostate Cancer and Clinically Significant Prostate Cancer in Men Undergoing Repeat Prostate Biopsy. *Biomed Res Int* 2018;2018:6368309 [FREE Full text] [doi: [10.1155/2018/6368309](https://doi.org/10.1155/2018/6368309)] [Medline: [30276213](https://pubmed.ncbi.nlm.nih.gov/30276213/)]
42. Macaskill P. *Cochrane handbook for systematic reviews of diagnostic test accuracy*, Version 0.9.0. London: The Cochrane Collaboration; 2010. URL: <https://methods.cochrane.org/sdt/> [accessed 2019-02-20]
43. Jackson D, Turner R. Power analysis for random-effects meta-analysis. *Res Synth Methods* 2017 Sep;8(3):290-302 [FREE Full text] [doi: [10.1002/jrsm.1240](https://doi.org/10.1002/jrsm.1240)] [Medline: [28378395](https://pubmed.ncbi.nlm.nih.gov/28378395/)]

44. Zhao K, Wang C, Hu J, Yang X, Wang H, Li F, et al. Prostate cancer identification: quantitative analysis of T2-weighted MR images based on a back propagation artificial neural network model. *Sci China Life Sci* 2015 Jul;58(7):666-673. [doi: [10.1007/s11427-015-4876-6](https://doi.org/10.1007/s11427-015-4876-6)] [Medline: [26025283](https://pubmed.ncbi.nlm.nih.gov/26025283/)]
45. Valerio M, Zini C, Fierro D, Giura F, Colarieti A, Giuliani A, et al. 3T multiparametric MRI of the prostate: Does intravoxel incoherent motion diffusion imaging have a role in the detection and stratification of prostate cancer in the peripheral zone? *Eur J Radiol* 2016 Apr;85(4):790-794. [doi: [10.1016/j.ejrad.2016.01.006](https://doi.org/10.1016/j.ejrad.2016.01.006)] [Medline: [26971425](https://pubmed.ncbi.nlm.nih.gov/26971425/)]
46. Lay N, Tsehay Y, Greer MD, Turkbey B, Kwak JT, Choyke PL, et al. Detection of prostate cancer in multiparametric MRI using random forest with instance weighting. *J Med Imaging (Bellingham)* 2017 Apr;4(2):024506 [FREE Full text] [doi: [10.1117/1.JMI.4.2.024506](https://doi.org/10.1117/1.JMI.4.2.024506)] [Medline: [28630883](https://pubmed.ncbi.nlm.nih.gov/28630883/)]
47. Reda I, Shalaby A, Elmogy M, Elfotouh AA, Khalifa F, El-Ghar MA, et al. A comprehensive non-invasive framework for diagnosing prostate cancer. *Comput Biol Med* 2017 Feb 01;81:148-158. [doi: [10.1016/j.combiomed.2016.12.010](https://doi.org/10.1016/j.combiomed.2016.12.010)] [Medline: [28063376](https://pubmed.ncbi.nlm.nih.gov/28063376/)]
48. Starobinets O, Simko JP, Kuchinsky K, Kornak J, Carroll PR, Greene KL, et al. Characterization and stratification of prostate lesions based on comprehensive multiparametric MRI using detailed whole-mount histopathology as a reference standard. *NMR Biomed* 2017 Dec;30(12):e3796. [doi: [10.1002/nbm.3796](https://doi.org/10.1002/nbm.3796)] [Medline: [28961382](https://pubmed.ncbi.nlm.nih.gov/28961382/)]
49. Kwon D, Reis IM, Breto AL, Tschudi Y, Gautney N, Zavala-Romero O, et al. Classification of suspicious lesions on prostate multiparametric MRI using machine learning. *J Med Imaging (Bellingham)* 2018 Jul;5(3):034502 [FREE Full text] [doi: [10.1117/1.JMI.5.3.034502](https://doi.org/10.1117/1.JMI.5.3.034502)] [Medline: [30840719](https://pubmed.ncbi.nlm.nih.gov/30840719/)]
50. Song Y, Zhang Y, Yan X, Liu H, Zhou M, Hu B, et al. Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. *J Magn Reson Imaging* 2018 Dec;48(6):1570-1577. [doi: [10.1002/jmri.26047](https://doi.org/10.1002/jmri.26047)] [Medline: [29659067](https://pubmed.ncbi.nlm.nih.gov/29659067/)]
51. Devine W, Giganti F, Johnston EW, Sidhu HS, Panagiotaki E, Punwani S, et al. Simplified Luminal Water Imaging for the Detection of Prostate Cancer From Multiecho T MR Images. *J Magn Reson Imaging* 2019 Sep;50(3):910-917 [FREE Full text] [doi: [10.1002/jmri.26608](https://doi.org/10.1002/jmri.26608)] [Medline: [30566264](https://pubmed.ncbi.nlm.nih.gov/30566264/)]
52. Le MH, Chen J, Wang L, Wang Z, Liu W, Cheng KT, et al. Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks. *Phys Med Biol* 2017 Jul 24;62(16):6497-6514. [doi: [10.1088/1361-6560/aa7731](https://doi.org/10.1088/1361-6560/aa7731)] [Medline: [28582269](https://pubmed.ncbi.nlm.nih.gov/28582269/)]
53. Mazaheri Y, Shukla-Dave A, Goldman DA, Moskowitz CS, Takeda T, Reuter VE, et al. Characterization of prostate cancer with MR spectroscopic imaging and diffusion-weighted imaging at 3 Tesla. *Magn Reson Imaging* 2019 Jan;55:93-102 [FREE Full text] [doi: [10.1016/j.mri.2018.08.025](https://doi.org/10.1016/j.mri.2018.08.025)] [Medline: [30176373](https://pubmed.ncbi.nlm.nih.gov/30176373/)]
54. Gholizadeh N, Greer PB, Simpson J, Denham J, Lau P, Dowling J, et al. Characterization of prostate cancer using diffusion tensor imaging: A new perspective. *Eur J Radiol* 2019 Jan;110:112-120. [doi: [10.1016/j.ejrad.2018.11.026](https://doi.org/10.1016/j.ejrad.2018.11.026)] [Medline: [30599846](https://pubmed.ncbi.nlm.nih.gov/30599846/)]
55. Wang X, Yang W, Weinreb J, Han J, Li Q, Kong X, et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning. *Sci Rep* 2017 Nov 13;7(1):15415 [FREE Full text] [doi: [10.1038/s41598-017-15720-y](https://doi.org/10.1038/s41598-017-15720-y)] [Medline: [29133818](https://pubmed.ncbi.nlm.nih.gov/29133818/)]
56. Chen T, Li M, Gu Y, Zhang Y, Yang S, Wei C, et al. Prostate Cancer Differentiation and Aggressiveness: Assessment With a Radiomic-Based Model vs. PI-RADS v2. *J Magn Reson Imaging* 2019 Mar;49(3):875-884 [FREE Full text] [doi: [10.1002/jmri.26243](https://doi.org/10.1002/jmri.26243)] [Medline: [30230108](https://pubmed.ncbi.nlm.nih.gov/30230108/)]
57. Qi Y, Zhang S, Wei J, Zhang G, Lei J, Yan W, et al. Multiparametric MRI-Based Radiomics for Prostate Cancer Screening With PSA in 4-10 ng/mL to Reduce Unnecessary Biopsies. *J Magn Reson Imaging* 2020 Jun;51(6):1890-1899. [doi: [10.1002/jmri.27008](https://doi.org/10.1002/jmri.27008)] [Medline: [31808980](https://pubmed.ncbi.nlm.nih.gov/31808980/)]
58. Ma X, Lv K, Sheng J, Yu Y, Pang P, Xu M, et al. Application evaluation of DCE-MRI combined with quantitative analysis of DWI for the diagnosis of prostate cancer. *Oncol Lett* 2019 Mar;17(3):3077-3084 [FREE Full text] [doi: [10.3892/ol.2019.9988](https://doi.org/10.3892/ol.2019.9988)] [Medline: [30867737](https://pubmed.ncbi.nlm.nih.gov/30867737/)]
59. Zhang S. Radiomics analysis potentially reduces over-diagnosis of prostate cancer with PSA levels of 4-10 ng/ml based on DWI data. 2019 Presented at: Medical Imaging 2019: Computer-Aided Diagnosis; 2019; San Diego, California, United States. [doi: [10.1117/12.2511497](https://doi.org/10.1117/12.2511497)]
60. Kirby MK, Ramaker RC, Roberts BS, Lasseigne BN, Gunther DS, Burwell TC, et al. Genome-wide DNA methylation measurements in prostate tissues uncovers novel prostate cancer diagnostic biomarkers and transcription factor binding patterns. *BMC Cancer* 2017 Apr 17;17(1):273 [FREE Full text] [doi: [10.1186/s12885-017-3252-2](https://doi.org/10.1186/s12885-017-3252-2)] [Medline: [28412973](https://pubmed.ncbi.nlm.nih.gov/28412973/)]
61. Patel PG, Wessel T, Kawashima A, Okello JBA, Jamaspishvili T, Guérard KP, et al. A three-gene DNA methylation biomarker accurately classifies early stage prostate cancer. *Prostate* 2019 Oct;79(14):1705-1714. [doi: [10.1002/pros.23895](https://doi.org/10.1002/pros.23895)] [Medline: [31433512](https://pubmed.ncbi.nlm.nih.gov/31433512/)]
62. Donovan MJ, Noerholm M, Bentink S, Belzer S, Skog J, O'Neill V, et al. A molecular signature of PCA3 and ERG exosomal RNA from non-DRE urine is predictive of initial prostate biopsy result. *Prostate Cancer Prostatic Dis* 2015 Dec;18(4):370-375. [doi: [10.1038/pcan.2015.40](https://doi.org/10.1038/pcan.2015.40)] [Medline: [26345389](https://pubmed.ncbi.nlm.nih.gov/26345389/)]
63. Zhang W, Zheng X, Wang X. Oxidative stress measured by thioredoxin reductase level as potential biomarker for prostate cancer. *Am J Cancer Res* 2015;5(9):2788-2798 [FREE Full text] [Medline: [26609484](https://pubmed.ncbi.nlm.nih.gov/26609484/)]

64. Mengual L, Lozano JJ, Ingelmo-Torres M, Izquierdo L, Musquera M, Ribal MJ, et al. Using gene expression from urine sediment to diagnose prostate cancer: development of a new multiplex mRNA urine test and validation of current biomarkers. *BMC Cancer* 2016 Feb 09;16:76 [FREE Full text] [doi: [10.1186/s12885-016-2127-2](https://doi.org/10.1186/s12885-016-2127-2)] [Medline: [26856686](https://pubmed.ncbi.nlm.nih.gov/26856686/)]
65. Barceló M, Castells M, Bassas L, Vigués F, Larriba S. Semen miRNAs Contained in Exosomes as Non-Invasive Biomarkers for Prostate Cancer Diagnosis. *Sci Rep* 2019 Sep 24;9(1):13772 [FREE Full text] [doi: [10.1038/s41598-019-50172-6](https://doi.org/10.1038/s41598-019-50172-6)] [Medline: [31551516](https://pubmed.ncbi.nlm.nih.gov/31551516/)]
66. Amante, Salomone, Alladio, Vincenti, Porphiglia, Bro. Untargeted Metabolomic Profile for the Detection of Prostate Carcinoma-Preliminary Results from PARAFAC2 and PLS-DA Models. *Molecules* 2019 Aug 22;24(17):3063 [FREE Full text] [doi: [10.3390/molecules24173063](https://doi.org/10.3390/molecules24173063)] [Medline: [31443574](https://pubmed.ncbi.nlm.nih.gov/31443574/)]
67. Brikun I, Nusskern D, Decatus A, Harvey E, Li L, Freije D. A panel of DNA methylation markers for the detection of prostate cancer from FV and DRE urine DNA. *Clin Epigenetics* 2018;10:91 [FREE Full text] [doi: [10.1186/s13148-018-0524-x](https://doi.org/10.1186/s13148-018-0524-x)] [Medline: [29988684](https://pubmed.ncbi.nlm.nih.gov/29988684/)]
68. Santotoribio JD, Jimenez-Romero ME. Serum biomarkers of inflammation for diagnosis of prostate cancer in patients with nonspecific elevations of serum prostate specific antigen levels. *Transl. Cancer Res* 2019 Feb;8(1):273-278. [doi: [10.21037/tcr.2019.01.31](https://doi.org/10.21037/tcr.2019.01.31)]
69. Gao Q, Su X, Annabi MH, Schreiter BR, Prince T, Ackerman A, et al. Application of Urinary Volatile Organic Compounds (VOCs) for the Diagnosis of Prostate Cancer. *Clin Genitourin Cancer* 2019 Jun;17(3):183-190. [doi: [10.1016/j.clgc.2019.02.003](https://doi.org/10.1016/j.clgc.2019.02.003)] [Medline: [30853355](https://pubmed.ncbi.nlm.nih.gov/30853355/)]
70. Chaibub Neto E, Pratap A, Perumal TM, Tummalacherla M, Snyder P, Bot BM, et al. Detecting the impact of subject characteristics on machine learning-based diagnostic applications. *NPJ Digit Med* 2019;2:99 [FREE Full text] [doi: [10.1038/s41746-019-0178-x](https://doi.org/10.1038/s41746-019-0178-x)] [Medline: [31633058](https://pubmed.ncbi.nlm.nih.gov/31633058/)]
71. Little MA, Varoquaux G, Saeb S, Lonini L, Jayaraman A, Mohr DC, et al. Using and understanding cross-validation strategies. Perspectives on Saeb et al. *Gigascience* 2017 May 01;6(5):1-6 [FREE Full text] [doi: [10.1093/gigascience/gix020](https://doi.org/10.1093/gigascience/gix020)] [Medline: [28327989](https://pubmed.ncbi.nlm.nih.gov/28327989/)]
72. Ganganwar V. An overview of classification algorithms for imbalanced datasets. *International Journal of Emerging Technology and Advanced Engineering* 2012;2(4):42-47.
73. Sathianathen NJ, Butaney M, Bongiorno C, Konety BR, Bolton DM, Lawrentschuk N. Accuracy of the magnetic resonance imaging pathway in the detection of prostate cancer: a systematic review and meta-analysis. *Prostate Cancer Prostatic Dis* 2019 Mar;22(1):39-48. [doi: [10.1038/s41391-018-0075-4](https://doi.org/10.1038/s41391-018-0075-4)] [Medline: [30108376](https://pubmed.ncbi.nlm.nih.gov/30108376/)]
74. Incoronato M, Aiello M, Infante T, Cavaliere C, Grimaldi A, Mirabelli P, et al. Radiogenomic Analysis of Oncological Data: A Technical Survey. *Int J Mol Sci* 2017 Apr 12;18(4):805 [FREE Full text] [doi: [10.3390/ijms18040805](https://doi.org/10.3390/ijms18040805)] [Medline: [28417933](https://pubmed.ncbi.nlm.nih.gov/28417933/)]
75. Zanfardino M, Franzese M, Pane K, Cavaliere C, Monti S, Esposito G, et al. Bringing radiomics into a multi-omics framework for a comprehensive genotype-phenotype characterization of oncological diseases. *J Transl Med* 2019 Oct 07;17(1):337 [FREE Full text] [doi: [10.1186/s12967-019-2073-2](https://doi.org/10.1186/s12967-019-2073-2)] [Medline: [31590671](https://pubmed.ncbi.nlm.nih.gov/31590671/)]
76. Zanfardino M, Pane K, Mirabelli P, Salvatore M, Franzese M. TCGA-TCIA Impact on Radiogenomics Cancer Research: A Systematic Review. *Int J Mol Sci* 2019 Nov 29;20(23):6033 [FREE Full text] [doi: [10.3390/ijms20236033](https://doi.org/10.3390/ijms20236033)] [Medline: [31795520](https://pubmed.ncbi.nlm.nih.gov/31795520/)]
77. Stoyanova R, Pollack A, Lynne C, Jorda M, Erho N, Lam LL, et al. Using radiogenomics to characterize MRI-guided prostate cancer biopsy heterogeneity. *JCO* 2015 Mar 01;33(7\_suppl):25-25. [doi: [10.1200/jco.2015.33.7\\_suppl.25](https://doi.org/10.1200/jco.2015.33.7_suppl.25)]
78. Jamaspishvili T, Berman DM, Ross AE, Scher HI, De Marzo AM, Squire JA, et al. Clinical implications of PTEN loss in prostate cancer. *Nat Rev Urol* 2018 Apr;15(4):222-234 [FREE Full text] [doi: [10.1038/nrurol.2018.9](https://doi.org/10.1038/nrurol.2018.9)] [Medline: [29460925](https://pubmed.ncbi.nlm.nih.gov/29460925/)]
79. VanderWeele DJ, McCann S, Fan X, Antic T, Jiang Y, Oto A. Radiogenomics of prostate cancer: Association between quantitative multiparametric MRI features and PTEN. *JCO* 2015 Mar 01;33(7\_suppl):126-126. [doi: [10.1200/jco.2015.33.7\\_suppl.126](https://doi.org/10.1200/jco.2015.33.7_suppl.126)]
80. Martin GV, Kudchadker RJ, Bruno TL, Frank SJ, Wang J. Comparison of prostate distortion by inflatable and rigid endorectal MRI coils in permanent prostate brachytherapy imaging. *Brachytherapy* 2018;17(2):298-305. [doi: [10.1016/j.brachy.2017.09.014](https://doi.org/10.1016/j.brachy.2017.09.014)] [Medline: [29169971](https://pubmed.ncbi.nlm.nih.gov/29169971/)]
81. Castaldo R, Pane K, Nicolai E, Salvatore M, Franzese M. The Impact of Normalization Approaches to Automatically Detect Radiogenomic Phenotypes Characterizing Breast Cancer Receptors Status. *Cancers (Basel)* 2020 Feb 24;12(2):518 [FREE Full text] [doi: [10.3390/cancers12020518](https://doi.org/10.3390/cancers12020518)] [Medline: [32102334](https://pubmed.ncbi.nlm.nih.gov/32102334/)]
82. Morrison A, Polisen J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care* 2012 Apr;28(2):138-144. [doi: [10.1017/S0266462312000086](https://doi.org/10.1017/S0266462312000086)] [Medline: [22559755](https://pubmed.ncbi.nlm.nih.gov/22559755/)]

## Abbreviations

**AI:** artificial intelligence

**AUC:** area under the curve

**DCE:** dynamic contrast-enhanced  
**DL:** deep learning  
**DWI:** diffusion-weighted imaging  
**FN:** false-negative  
**FP:** false-positive  
**HSROC:** hierarchical summary receiver operating characteristic curve  
**ML:** machine learning  
**mpMRI:** multiparametric magnetic resonance imaging  
**PCa:** prostate cancer  
**PI-RADS-V2:** Prostate Imaging Reporting and Data System  
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses  
**PSA:** prostate-specific antigen  
**QUADAS-2:** quality assessment of diagnostic accuracy studies–version 2  
**ROI:** region of interest  
**SMOTE:** synthetic minority oversampling technique  
**TN:** true-negative  
**TP:** true-positive  
**TRUS:** transrectal ultrasonography  
**TZ:** transition zone  
**VOI:** volume of interest

*Edited by G Eysenbach; submitted 10.07.20; peer-reviewed by G Fico, K Kowalewski, A Drory, DA Banerjee; comments to author 17.11.20; revised version received 26.11.20; accepted 17.01.21; published 01.04.21*

*Please cite as:*

Castaldo R, Cavaliere C, Soricelli A, Salvatore M, Pecchia L, Franzese M  
*Radiomic and Genomic Machine Learning Method Performance for Prostate Cancer Diagnosis: Systematic Literature Review*  
*J Med Internet Res* 2021;23(4):e22394  
URL: <https://www.jmir.org/2021/4/e22394>  
doi: [10.2196/22394](https://doi.org/10.2196/22394)  
PMID: [33792552](https://pubmed.ncbi.nlm.nih.gov/33792552/)

©Rossana Castaldo, Carlo Cavaliere, Andrea Soricelli, Marco Salvatore, Leandro Pecchia, Monica Franzese. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 01.04.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.