Insights into Imaging
EUROPEAN SOCIETY OF RADIOLOGY

ORIGINAL ARTICLE

Open Access

Automatic segmentation of prostate zonal anatomy on MRI: a systematic review of the literature

Carine Wu[1,2*], Sarah Montagne[1,2,3,4], Dimitri Hamzaoui[5], Nicholas Ayache[5], Hervé Delingette[5] and Raphaële Renard-Penna[1,2,3,4]

Abstract

Objectives: Accurate zonal segmentation of prostate boundaries on MRI is a critical prerequisite for automated prostate cancer detection based on PI-RADS. Many articles have been published describing deep learning methods offering great promise for fast and accurate segmentation of prostate zonal anatomy. The objective of this review was to provide a detailed analysis and comparison of applicability and efficiency of the published methods for automatic segmentation of prostate zonal anatomy by systematically reviewing the current literature.

Methods: A Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) was conducted until June 30, 2021, using PubMed, ScienceDirect, Web of Science and EMBase databases. Risk of bias and applicability based on Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) criteria adjusted with Checklist for Artificial Intelligence in Medical Imaging (CLAIM) were assessed.

Results: A total of 458 articles were identified, and 33 were included and reviewed. Only 2 articles had a low risk of bias for all four QUADAS-2 domains. In the remaining, insufficient details about database constitution and segmentation protocol provided sources of bias (inclusion criteria, MRI acquisition, ground truth). Eighteen different types of terminology for prostate zone segmentation were found, while 4 anatomic zones are described on MRI. Only 2 authors used a blinded reading, and 4 assessed inter-observer variability.

Conclusions: Our review identified numerous methodological flaws and underlined biases precluding us from performing quantitative analysis for this review. This implies low robustness and low applicability in clinical practice of the evaluated methods. Actually, there is not yet consensus on quality criteria for database constitution and zonal segmentation methodology.

Key points

- Several limitations exist with current methods of automatic prostate segmentation.
- There is wide variability of databases, imaging tasks and assessment criteria.
- No preferred methodology for prostate zonal segmentation methodology and terminology was used.
- Vast majority of papers share common methodological flaws discussed in this review.

*Correspondence: carine.wu@aphp.fr

[2] Academic Department of Radiology, Hôpital Tenon, Assistance Publique des Hôpitaux de Paris, 4 Rue de La Chine, 75020 Paris, France
Full list of author information is available at the end of the article

Springer Open

© The Author(s) 2022. **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/.

Keywords: Artificial intelligence, Deep learning, Magnetic resonance imaging, Prostate cancer

Introduction

Magnetic resonance imaging (MRI) is the first imaging choice for detecting and localizing prostate cancer [1, 2], based on the Prostate Imaging Reporting and Data System (PI-RADS) scoring system [3] and depending on zonal anatomy. Zonal segmentation of the prostate plays a crucial role for prostate cancer detection as the PI-RADS score differs depending on the areas studied, based on diffusion-weighted imaging (DWI) for peripheral zone lesions and T2-weighted (T2W) imaging for transitional zone lesions, but also for multiple clinical application such as reproducible prostate volume and Prostate Specific Antigen (PSA) density evaluation [4], MRI-ultrasound fusion biopsy, radiotherapy, or focal planning.

Zonal segmentation of the prostate is usually performed manually on T2W images by contouring the prostate in a slice-by-slice manner. It is extremely time-consuming, tedious, and prone to inter and intra-observer variability due to the subjective human interpretation of organ boundaries and large variability in prostate anatomy and gland intensity heterogeneity across patients [5]. There is a real need to develop automatic methods to accelerate the whole process and offer robust and accurate prostate segmentation.

Automatic zonal segmentation of the prostate is a challenging task for multiple reasons. Prostate gland is subject to large morphological variation, intra-prostatic heterogeneity, and poor contrast with adjacent tissues, making delineation of prostatic zonal contours laborious. Multi-institutional applicability can be difficult to evaluate as there is a wide technically induced variability in the image acquisition, as MRI signal intensity is not standardized and image characteristics are strongly influenced by acquisition protocol, field strength, scanner type, coil type, etc. [6].

Finally, the performances of an automated segmentation method depend in part on the database (heterogeneity of the data used, knowledge of possible selection biases), quality of ground truth (manual delineation of the prostate performed by human experts), training time and hardware requirements. First commonly used methods were based on machine learning methods, such as atlas-based registration models in which several reference images with corresponding labels are registered and deformed onto the target image [7, 8] or C-means clustering models [9, 10]. Most common methods described after 2017 are based on deep learning with convolutional neural networks (CNN) allowing automatic extraction of features and semantic image segmentation. Common architectures such as U-net [11], V-net or ResNet have been extensively used. Modification and fine tuning of existing models, by either combining multiple U-nets [12–14], adding attention modules such as squeeze and excitation [15], feature pyramid attention [16], adding blocks [17], transition layers or up-sampling strategies [18], allowed either improving accuracy of classical CNN or obtaining same accuracy with reduced memory and storage requirements.

The primary objective of this review was to provide a detailed analysis and comparison of applicability and efficiency of the published methods for automatic segmentation of prostate zonal anatomy by systematically outlining, analyzing, and categorizing the relevant publications in the field to date. We also aimed to identify methodological flaws and biases to demonstrate the need for a consensus on quality criteria for database constitution and prostate zonal segmentation methodology.

Materials and methods

This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (PRISMA) [19]. The methods for performing this systematic review were registered on PROSPERO [20] database (registration number CD42021265371), and were agreed by all authors before the start of the review process to avoid bias. This study was exempt from ethical approval at our institution because the analysis involved only de-identified data.

Data sources and search

Medical literature published in the English language published until 30 June 2021 was searched in multiple databases (Medline, Science direct, Embase and Web of Science) using the following terms:

(prostatic OR prostate) AND (automated OR automatic) AND (segmentation OR segmented) AND (zone OR zonal) AND (\"magnetic resonance\" OR mri OR \"magnetic resonance\"OR mri OR mr) AND (\"artificial intelligence\" OR \"deep learning\" OR \"machine learning\") and all possible combinations.

No beginning date was applied.

Study selection

Full-text selection was independently performed by two radiologists, one experimented radiologist specialized in uroradiology and prostate imaging (S.M., 5 years in

prostate imaging, with more than 1000 cases of prostate MRI per year) and one radiology fellow specialized in uroradiology and prostate imaging (C.W., 1 year in prostate imaging, with more than 1000 cases of prostate MRI per year). A third experimented professor of radiology specialized in prostate imaging (R.R.P., 15 years in prostate imaging, with more than 1000 cases of prostate MRI per year) intervened in case of disagreement. We summarized search strategy details for each database in Fig. 1.

We imported all articles retrieved into the reference manager Zotero and removed all duplicates. The same two radiologists (C.W., S.M.) then independently and manually screened titles and abstracts of the resultant database to ensure relevance. Articles that were obviously out of the scope of the research topic were excluded at this stage. Subsequently, all the remaining articles full texts were retrieved and read, applying inclusion and exclusion criteria (explained below) with conflicts resolved by consensus with the third reviewer. Reference lists of these relevant articles were also reviewed for possible papers missed in the primary search, and those papers were screened using the same initial inclusion and exclusion criteria.

Selection criteria

Inclusion criteria

Articles were included if they were original articles, used machine learning or deep learning algorithms and aimed to segment prostate human MRI images by zonal anatomy, using a fully automated method with manual segmentation as ground truth.

Exclusion criteria

Articles were excluded if they were commentaries, editorials, letters, case reports or abstracts. Were also excluded articles with semi-automated segmentation methods, no description of segmentation method, segmentation of the whole gland (WG), or prostate cancer without zonal anatomy, absence of similarity metrics or of evaluation against ground truth segmentations.

Data collection and extraction process

The qualifying papers were then reviewed, and various data of the studies were extracted and tabulated prior to analysis (Table 1).

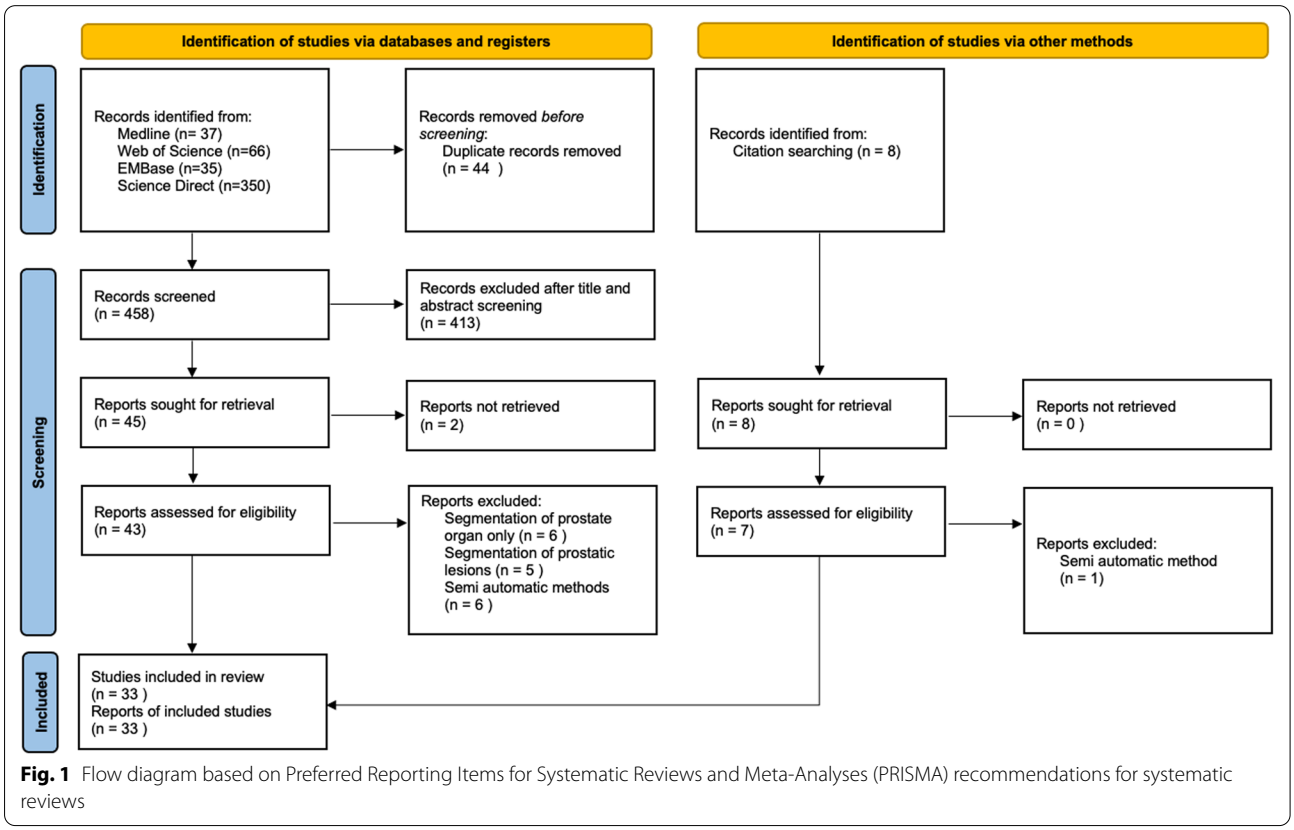

Fig. 1 Flow diagram based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations for systematic reviews

Table 1 Data extraction

Sources	Patients	Data	Flow and timing	Reference standard	Test
Scientific database	Public or in-house database	Vendor	Cross-validation	Type of annotation	Validation or test on external data
Title	Eligibility criteria: inclusion and exclusion criteria	Field	Splitting in training, validation and test set	Annotation tool *if used*	Performance metrics
Authors	Sample size	Array		Number of annotators	Results based on DSC
Year of publication	Ethic consent	Field of view		Ground truth segmentation and rationale	
Journal name	Presence of benign prostate hypertrophia	Pre-processing		Measurements of inter- and intra-rater variability *if any*	
	Presence of prostate cancer	Post-processing		Type of annotators	
	Percentage of prostate cancer	Number of vendors		Experience of annotators	
	Uni or multicentric	Slice thickness			
	Prospective or retrospective	Type of slice and sequence			
		Cross-validation			

DSC dice similarity coefficient

Assessment of methodological quality

The two same radiologists (C.W., S.M.) independently assessed and extracted data from each of the included articles, using the Quality Assessment of Diagnostic Accuracy Studies tool-2 (QUADAS-2) framework [21] adjusted with topics from the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) [22] to evaluate the risk of bias and applicability for each selected study, with conflicts resolved by consensus with the third reviewer.

Extracted data were tabulated, synthesized, and evaluated for methodological flaws and applicability of the proposed techniques.

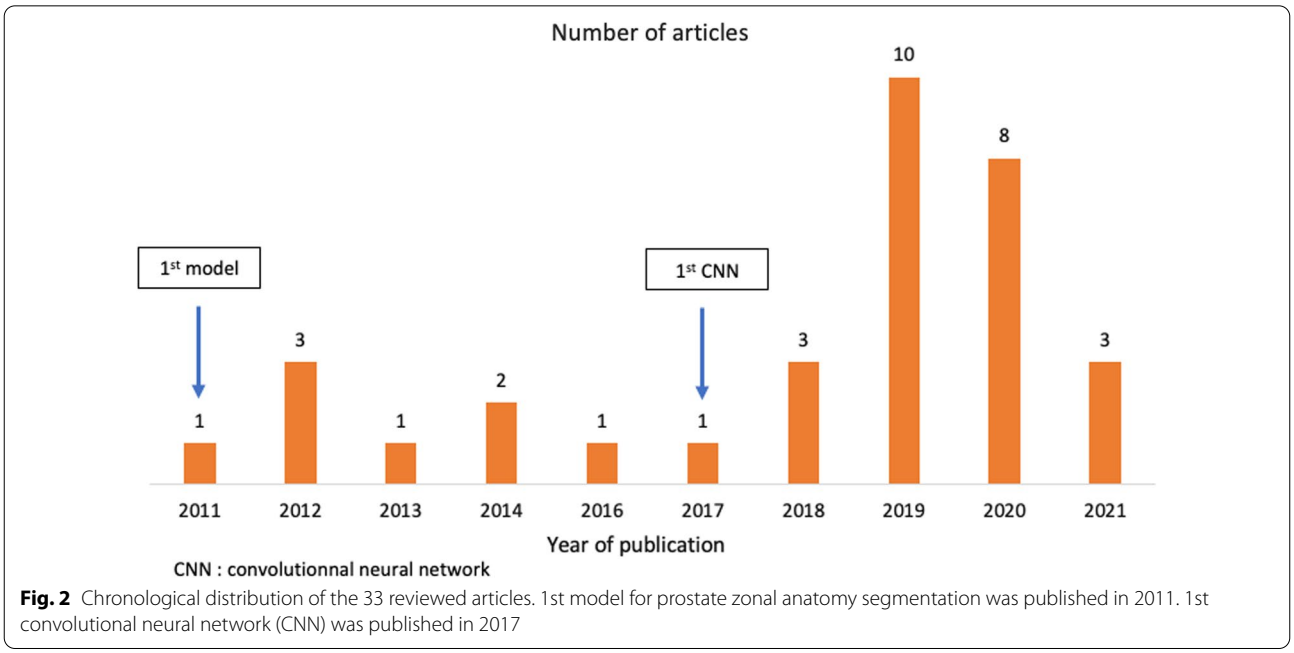

Fig. 2 Chronological distribution of the 33 reviewed articles. 1st model for prostate zonal anatomy segmentation was published in 2011. 1st convolutional neural network (CNN) was published in 2017

Results

After removing duplicates, 458 articles were remaining. Final consensus was reached yielding a total of 33 articles [6–10, 12–18, 23–43] (Figs. 1, 2).

Datasets

Training, validation, and test sets

All articles used retrospective datasets.

Wide heterogeneity in training, validation and test datasets was found (Table 2).

Performance testing of the algorithms can be done on same source than for the development or use different source of data, and based on either public data, private data or a combination of both. Public data were used in 15/33 articles for testing. Only 7 studies [6, 9, 14, 30, 33, 36, 37] used both private and public data for testing, allowing better generalizability of their algorithms. None of them used prospective data for validation and testing.

Most used public datasets were PROSTATEx [44], NCI-ISBI 2013 [45] and PROMISE12 [46] (Additional file 1: Table S1).

Eight authors applied cross-validation, using a subset of available dataset as training set, while the remaining data constituted the test set to evaluate the segmentation performance and accuracy. Nine reported using cross-validation for testing, averaging the results from the different rounds, hence adding bias.

Technique

We identified major technical differences in datasets regarding the number of vendors, field strength, type of coils, sequences, slice thickness, field of view (FOV) and input data used for automatic segmentation (Table 3). Less than half (14/33) studies used more than one type of vendors and 7/33 used both 1.5T and 3T MRI machines. More than 2/3 (24/33) used mono-modal input, mainly T2-weighted planes, in combination with apparent diffusion coefficient (ADC) map in one study [13] or with multiparametric and multi-incidence MR images in another [9]. The slice thickness of T2-weighted axial planes was consistent with the PI-RADS v2.1 recommendations in 13/33 studies (≤ 3 mm), which was not the case for the public data base PROSTATEx (3.6 mm). Only 7 studies provided sequence details (type of sequence, slice thickness, FOV) used for ground truth manual segmentation.

Zonal anatomy

We found 18 different types of very heterogeneous and unclear terminologies of zonal anatomy (Fig. 3, Additional file 1: Fig. S1). Out the 33 articles reviewed, less than 1/4 (8/33) [23, 25, 32, 34, 36, 37, 40, 43] provided precise terminology and segmentation protocol. Frequently the inappropriate term "central gland" (CG) was used, with ambiguous definition of central zone (CZ) and anterior fibro-muscular stroma (AFMS) alternatively included in peripheral zone (PZ) or transition zone (TZ), or mainly not described at all. Two studies mis-used the term "central zone" to refer to the "central gland" [27, 39].

Ground truth

Manual delineation of the prostate gland performed by human experts was used to generate ground truth (Table 4).

Annotation tool

Twenty studies (61%) reported using manual contouring, while a third (11/33) reported using annotation tools. One team [31] specified that the radiologist did not delineate zones on all slices but relied on interpolation performed by their annotation tools. Two studies [32, 33] did not provide any information.

Qualifications of annotators

Most studies (27/33, 81%) reported a radiologist or a radiation oncologist as human expert. In 3 papers, no detail was provided on annotators qualification, although one [15] specified using an "expert" reader. Definition of an "expert" reader was mostly unclear with no specification of number of MRI they interpreted, for example [10, 15, 26, 31, 34, 39].

Number of readers

Number of readers and their experience are described in Table 4. Number of readers was not available in two studies. While 2/3 of teams (22/33) reported using more than one reader, with splitted, stratified or blinded reading approaches, 7 did not provide information on reading approach.

Intra and inter-rater variability

Inter-rater variability for annotations was rated in only 4 studies [7, 10, 23, 39]. Some studies used alternative techniques to approach better homogeneity of ground truth. In [13], the four radiologists met for a training session and together segmented two example patients to achieve a similar methodology for the rest of the dataset, using only experienced radiologists. In [6], the contours segmented by three radiologists were cross-checked and reviewed by two radiation oncologists, resulting in better homogeneity of ground truth. In [18], the initial prostate masks were drawn by two students who were trained in segmenting prostate zones.

Risk of bias and quality assessment

The detailed results are presented in Fig. 4 and Additional file 1: Table S2.

Table 2 Overview of types of databases used with training, validation and test sets distribution

First author, year of publication	Inclusion criteria	Presence of PCa	Number of patients (total)	Training			Validation		Test		
				Total	Public data	In-house data	Cross-validation	Validation data	Total	Public data	In-house data
Cuocolo et al. [43]	✓	✓	204	79	79[(A)]	0	Fivefold	20	105	105[(A)]	0
Bardis et al. [42]	✓	✓	242	146	0	145	0	48	48	0	48
Lai et al. [41]	✓	✓	115	80	80[(A)]	0	Fivefold	20	15	15[(A)]	0
Nai et al. [18]	✓	✓	160	120	120[(A)]	0	0	20	20	20[(A)]	0
Sanford et al. [40]	✓	✓	1054	518 + 162[§]	0	680	0	130 + 42[§]	202	0	202
Aldoj et al. [39]	✓	✓	188	106	106[(A)]	0	Fourfold	35	47	20[(A)]	0
Zavala-Romero et al. [6]	✓	✓	550	198 or 297	297[(A)]	198	0	0	variable	33[(A)]	22
Lee et al. [38]	✓	✓	330	260 *(for WG)* or 162 *(for TZ)*	0	260 *(for WG)* or 162 *(for TZ)*	0	0	70 *(for WG)* or 50 *(for TZ)*	0	50
Liu et al. [37]	✓	✓	351	218	218[(A)]	0	0	45	92	45[(A)]	47
Qin et al. [36]	×	✓	240	162 + 45	45 [(B)]	162	0	0	33	15 [(B)]	18
Motamed et al. [35]	×	Unknown	681	291 *(source)* + variable *(target)*	0	406	0	97	145 *(source)* + 33 *(target)*	0	178
Zabihollahy et al. [13]	✓	✓	225	80	0	80	0	20	125	0	125
Padgett et al. [8]	✓	✓	61	Variable	0	Variable	0	0	Variable	0	1
Rundo et al. [15][1]	×	✓	80	Variable	Variable[(B)]	Variable	0	0	Variable	Variable[(B)]	Variable
Meyer et al. [34]	✓	✓	98	58	58[(A)]	0	Fourfold	20	20	20[(A)]	0
Liu et al. [16]	✓	✓	359	200	200[(A)]	0	Fivefold	50	110	63[(A)]	46
Rundo et al. [33][2]	×	✓	40 [†]	Variable [†]	[†(C)]	Variable	0	0	Variable	0	Variable
Hambarde et al. [32]	×	Unknown	52	42	0	42	0	0	10	0	10
Jensen et al. [31]	✓	✓	40	32	0	32	Fivefold	2	8	0	8
Khan et al. [17]	×	✓	80	35	35[(B)]	0	0	15	30	30[(B)]	0
Cheng et al. [30]	×*	✓	225	116 +/− [†]	8[(A)]	108	0	0	Variable	Variable[(A+C)]	27
Zhu et al. [12]	✓	✓	163	76	0	76	0	36	51	0	51
Mooij et al. [29]	0	Unknown	53	36	0	36	Fivefold	9	8	0	8
Can et al. [28]	0	Unknown	29	12	12[(B)]	0	0	7	10	10[(B)]	0
Clark et al. [14]	×*	✓	154	115	78[(C)]	37	0	0	38	12[(C)]	26
Chilali et al. [9]	✓	✓	55	30	30 *(Prostatlas)*	0	0	0	25	13[(C)]	12

Table 2 (continued)

First author, year of publication	Inclusion criteria	Presence of PCa	Number of patients (total)	Training			Validation		Test		
				Total	Public data	In-house data	Cross-validation	Validation data	Total	Public data	In-house data
Makni et al. [10]	✓	✓	31	? *(simulated images)*	0	0	0	0	31	0	31
Chi et al. [27]	✓	Unknown	8	4	0	4	0	0	4	0	4
Toth et al. [26]	✓	✓	40	Variable	0	Variable	0	0	Variable	0	Variable
Litjens et al. [7]	×	Unknown	48	48	0	47	0	0	1	0	1
Moschidis and Graham [25]	✓	×	22	Variable	0	Variable	0	0	Variable	0	Variable
Yin et al. [24]	×	✓	522 *(images)*	261 *(images)*	0	261 *(images)*	Fivefold	52 *(images)*	261 *(images)*	0	261 *(images)*
Makni et al. [23]	✓	✓	31	?	0	0	0	0	31	0	31

*Not specified for in-house data

† +/− 50 patients from PROMISE12 dataset used for pre-training of WG segmentation

§ Pre-training data + data for transfer learning

(A) Public data used is PROSTATE-X

(B) Public data used is NCI-ISBI

(C) Public data used is PROMISE12

[1] Rundo et al., USE-Net: incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets [30]

[2] Rundo et al., CNN-based Prostate Zonal Segmentation on T2-weighted MR Images: A Cross-dataset Study[27]

PCa prostate cancer, *WG* whole gland, *TZ* transition zone

Table 3 Input MRI parameters (number of vendors, type of field, type of coil, input sequences)

First author, year of publication	Vendors				Type of coil		Field strenght		Data input		
	Number of vendors	Philips	GE	Siemens	ERC	SC	1.5T	3T	Mono-parametric	Sequence	Slice thickness*
Cuocolo et al. [43]	1	–	–	✓	–	✓	–	✓	✓	Axial T2W	3.6
Bardis et al. [42]	2	✓	–	✓	–	✓	–	✓	✓	Axial T2W	3.0
Lai et al. [41]	1	–	–	✓	–	✓	–	✓	×	Axial T2W + DWI + ADC	3.6
Nai et al. [18]	1	–	–	✓	–	✓	–	✓	×	Axial T2W + DWI + ADC	3.6
Sanford et al. [40]	3	✓	✓	✓	✓	✓	✓	✓	✓	Axial T2W	3.0
Aldoj et al. [39]	1	–	–	✓	–	✓	–	✓	✓	Axial T2W	3.6
Zavala-Romero et al. [6]	2	–	✓	✓	–	✓	–	✓	✓	3 planes T2W	3.6
Lee et al. [38]	1	–	–	✓	–	✓	–	✓	✓	Axial + sagittal T2W	3.0
Liu et al. [37]	1	–	–	✓	–	✓	–	✓	✓	Axial T2W	3.0–3.6
Qin et al. [36]	at least 2	✓	?	✓	✓	✓	✓	✓	×	Axial T2W + ADC	3.6
Motamed et al. [35]	2	✓	–	✓	?	?	?	?	✓	DWI	3.0
Zabihollahy et al. [13]	1	–	✓	–	–	✓	–	✓	×	Axial T2W + ADC	3.0–4.0
Padgett et al. [8]	2	–	✓	✓	?	?	–	✓	✓	Axial T2W	2.5
Rundo et al. [15][1]	2	✓	–	✓	–	✓	–	✓	✓	Axial T2W	1.25–4.0
Meyer et al. [34]	1	–	–	✓	–	✓	–	✓	✓	Axial T2W	3.0
Liu et al. [16]	1	–	–	✓	–	✓	–	✓	✓	Axial T2W	3.6
Rundo et al. [33][2]	2	✓	–	✓	–	✓	–	✓	✓	Axial T2W	1.25–3.0
Hambarde et al. [32]	1	✓	–	–	?	?	✓	–	✓	Axial T2W	5.0
Jensen et al. [31]	2	–	✓	✓	✓	✓	✓	✓	✓	Axial T2W	1.5–3.0
Khan et al. [17]	2	✓	–	✓	✓	✓	✓	✓	✓	Axial T2W	3.0–4.0
Cheng et al. [30]	multiple	?	?	✓	✓	✓	?	✓	✓	Axial T2W	3.0
Zhu et al. [12]	1	✓	–	–	–	✓	–	✓	×	Axial T2W + DWI	4.0
Mooij et al. [29]	?	?	?	?	?	?	?	?	✓	3D T2W	3.6
Can et al. [28]	2	✓	–	✓	✓	✓	✓	✓	✓	Axial T2W	3.0–4.0
Clark et al. [14]	multiple	✓	?	?	✓	✓	✓	✓	✓	DWI	?
Chilali et al. [9]	3	✓	✓	✓	✓	✓	✓	✓	✓	Axial T2W	3.0–4.0
Makni et al. [10]	1	✓	–	–	?	?	✓	–	×	Axial T2W + DWI + CE	1.25
Chi et al. [27]	1	–	–	✓	–	✓	–	✓	×	Axial T2W + ADC	3.3–3.75
Toth et al. [26]	?	?	?	?	✓	–	–	✓	✓	Axial T2W	3.0
Litjens et al. [7]	?	–	–	–	?	?	?	?	×	Axial T2W + ADC	4.0

Table 3 (continued)

First author, year of publication	Vendors				Type of coil		Field strenght		Data input		
	Number of vendors	Philips	GE	Siemens	ERC	SC	1.5T	3T	Mono-parametric	Sequence	Slice thickness*
Moschidis and Graham [25]	1	✓	–	–	–	✓	✓	–	✓	3D T2W	?
Yin et al. [24]	1	✓	–	–	✓	✓	–	✓	✓	Axial T2W	3.0
Makni et al. [23]	1	✓	–	–	–	✓	✓	–	×	Axial T2W + DWI + CE	2.5

ERC endorectal coil, *SC* Surface coil, *T2W* T2-weighted, *DWI* diffusion-weighted imaging, *ADC* apparent diffusion coefficient, CE contrast-enhanced

*Slice thickness in mm, when axial T2W slices used

?Not reported

[1] Rundo et al., USE-Net: incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets [30]

[2] Rundo et al., CNN-based Prostate Zonal Segmentation on T2-weighted MR Images: A Cross-dataset Study[27]

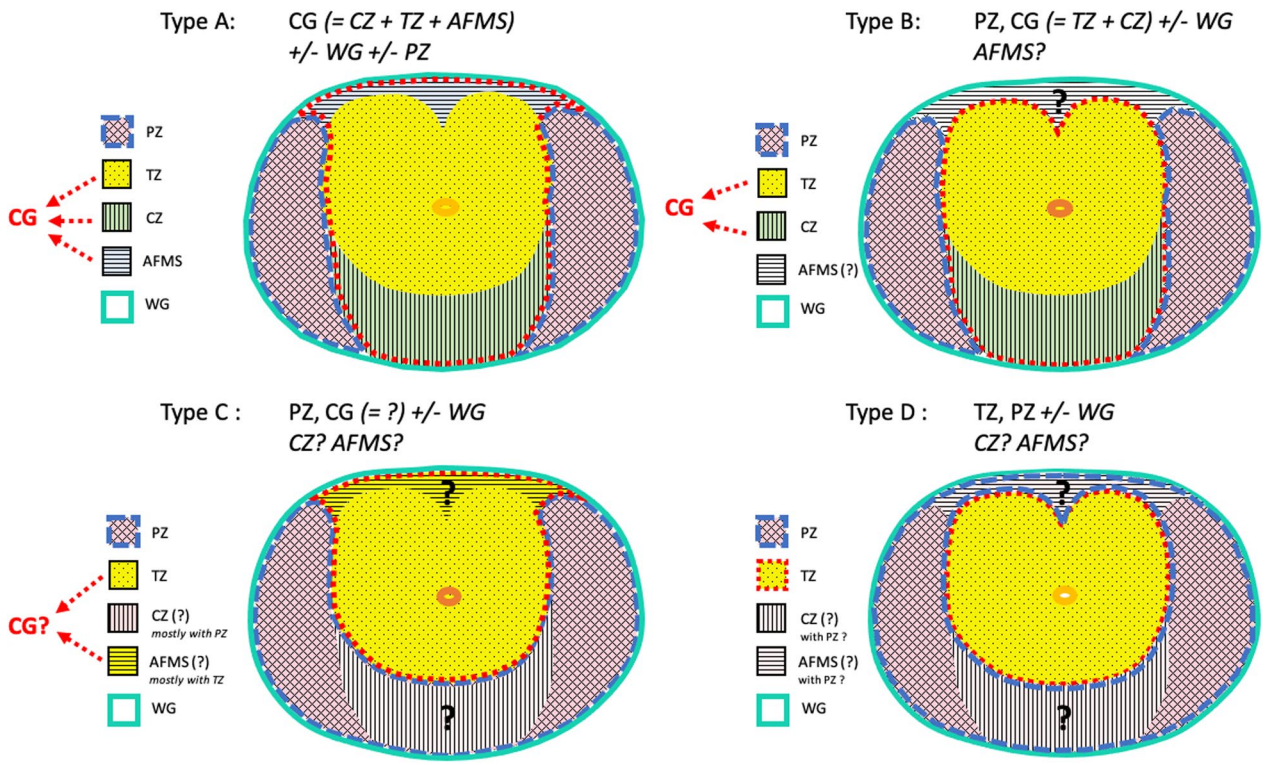

Fig. 3 Schematic of the four major types of protocol of zonal segmentation. Type A: articles for which "central gland" included CZ, TZ and AFMS. Type B: articles for which "central gland" included TZ and CZ. No details for AFMS. Type C: articles which did not provide details for AFMS, CZ or CG. CZ seemed to be mostly segmented PZ, while AFMS seemed to be mostly segmented with TZ, usually called "CG". Type D: articles which did not provide details for AFMS or CZ. CZ and AFMS seemed to be mostly segmented with PZ. *CZ* central zone, *TZ* transition zone, *AFMS* anterior fibro-muscular stroma, *PZ* peripheral zone, *CG* central gland

Regarding patient selection, we considered a low risk of bias if there were clear data inclusion and exclusion criteria, inclusion of patients with and without PCa. Models were considered less applicable if datasets were composed of only one type of scanners or if no information was specified.

For reference standard, number of readers and type of reading for ground truth segmentation were reviewed.

Clear partitioning of the database (into training, validation, and test sets) was needed to waive risk of bias for flow and timing. Some articles used cross-validation methods without keeping a clear independent test dataset [6–8, 15, 25, 26, 30, 33, 36].

Overall, all 33 included studies were judged to have a low risk of bias in the domain "index test" and 22 of 33 (67%) of the studies were judged to have a low risk of bias considering "flow and timing". However, only 1/4 of the studies (8/33) were judged to have a low risk of bias in the domain "patient selection", 1/3 (10/33) in the domain "reference standard". Only 2 articles were judged to have a low risk of bias in all four domains.

AI methodology

Before 2017, authors mostly used machine learning-based methods for automatic segmentation of prostatic zones. After 2017, almost all publications were based on deep learning with convolutional neural networks (CNN) (72%, 24/33). Common architectures such as U-net [11] have been extensively used, with modification and fine tuning of existing models, allowing either improved accuracy of classical networks or reduced memory and storage requirements.

Dice coefficient (DSC) and Hausdorff distance [47] were commonly used metrics. Almost all authors found inferior results for PZ than WG, CG or TZ segmentation, attributing this to the more complex shape and structure of PZ, especially within the anterior bundles. Eleven authors subsequently stratified their DSC results based on prostate height, with various methods:in three equal parts [13], in 25% apex, 50% mid gland and 25% base [39] in 30%, 40% and 30%, respectively [31]. Five authors did not provide any details on how they divided the volume.

These results as well as the remaining metrics are summarized in Table 5.

Table 4 Type of ground truth segmentation

First author, year of publication	Annotation		Annotators			
	Type	Tool	Qualification	Number	Type of reading	Experience*
Cuocolo et al. [43]	Software	itk-SNAP	(A)	4	Splitted[¶]	2 to 5
Bardis et al. [42]	Software	In house	(A)	12	Stratified	10
Lai et al. [41]	Manual	–	(A)	1	–	10
Nai et al. [18]	Software	MITK	Medical physicist	4	Stratified	2 to 10
Sanford et al. [40]	Software	pseg	(A)	1	–	10
Aldoj et al. [39]	Manual	–	(A)	1	–[Ŧ]	"Expert"
Zavala-Romero et al. [6]	Manual	–	(A) + (B)	3	Stratified	10
Lee et al. [38]	Manual	–	(A)	2	?	4
Liu et al. [37]	Software	Osirix	(C) + (A)	More than 2	Stratified	10 and 19
Qin et al. [36]	Manual	–	(A)[†]	?	–	?
Motamed et al. [35]	Manual	–	(A)	2	?	4 and 6
Zabihollahy et al. [13]	Software	itk-SNAP	(A)	4	Splitted	5 and 14
Padgett et al. [8]	Manual	–	(B)	2	Blinded[§]	10 and 26
Rundo et al. [15][1]	Manual	–	?	Multiple	?	"Expert"
Meyer et al. [34]	Software	3DSLICER	Medical student + urologist + (A)	4	Stratified	"Expert"
Liu et al. [16]	Software	Osirix	(C) + (A)	7	Stratified	10–15
Rundo et al. [33][2]	?	?	(A)	Multiple	?	?
Hambarde et al. [32]	?	?	(A)	Multiple	?	?
Jensen et al. [31]	Software	?	(A)	1	–	"Expert"
Khan et al. [17]	Manual	–	(A)	3	?	?
Cheng et al. [30]	Software	pseg	(A)	1	–	10
Zhu et al. [12]	Manual	–	?	2	?	More than 5
Mooij et al. [29]	Manual	–	?	?	?	?
Can et al. [28]	Manual	–	(A)	3	?	?
Clark et al. [14]	Manual	–	(A)	1	?	?
Chilali et al. [9]	Manual	–	(A)	1	–	15
Makni et al. [10]	Manual	–	(A)	3	Blinded	"Expert"
Chi et al. [27]	Manual	–	(A)	1	–	5
Toth et al. [26]	Software	3DSLICER	(A)	1	–	"Expert"
Litjens et al. [7]	Manual	–	(A)	3	?	?
Moschidis and Graham [25]	Manual	–	(A)	2	?	?
Yin et al. [24]	Manual	–	"Radiologist-trained operators"	2	Splitted	?
Makni et al. [23]	Manual	–	(A)	3	Blinded	4, 6 and 9

(A) Radiologist

(B) Radiation oncologist

(C) Research fellow

? Data not reported

*Experience of reader(s), in years

[†] Unclear for PROMM, in-house data

[¶] Splitted but consensus per binome resident-senior

[Ŧ] Only one reading for ground truth segmentation but evaluation of intra and inter observator variability on some masks

[§] Measure of inter- observator variability for 10 masks

Splitted: database is divided such as each set of images is read only once, resulting in an equivalent of single reader

Stratified: first reading (mostly by a less experienced reader) subsequently corrected by a more experience reader

Blinded: blinded reading by at least 2 readers

[1] Rundo et al., USE-Net: incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets [30]

[2] Rundo et al., CNN-based Prostate Zonal Segmentation on T2-weighted MR Images: A Cross-dataset Study[27]

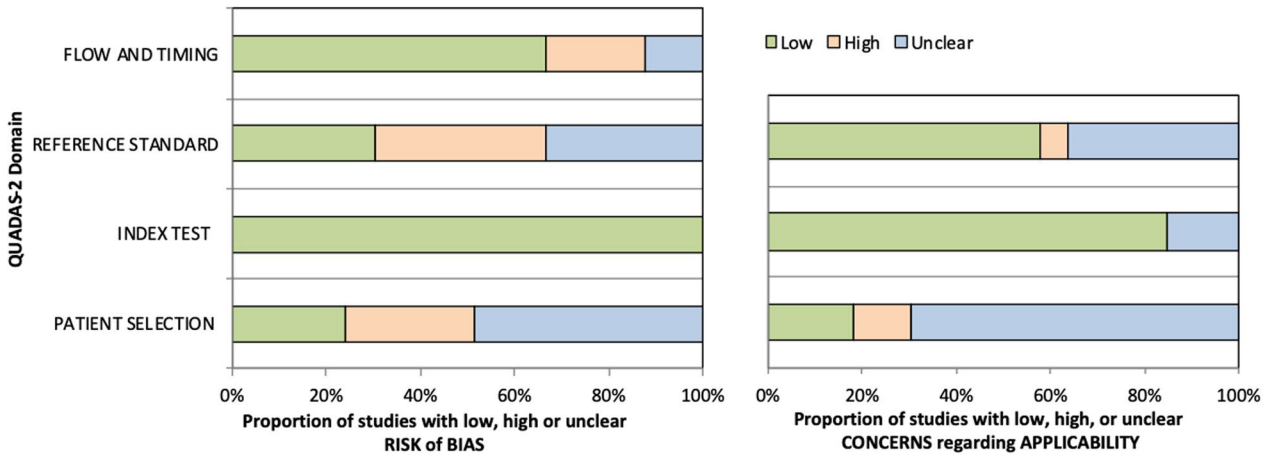

Fig. 4 Stacked bar charts showing results of quality assessment for risk of bias and applicability of included studies. QUADAS-2 scores for methodologic study quality are expressed as the percentage of studies that met each criterion. For each quality domain, the proportion of included studies that were determined to have low, high, or unclear risk of bias and/or concerns regarding applicability is displayed in green, orange, and blue, respectively. QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies 2

Discussion

Our systematic review highlights the high prevalence of deficiencies in methodology in the literature on automatic segmentation of prostate gland on MRI.

Since 2011, 33 studies proposed new or fine-tuned existing approaches for automatic prostatic zonal segmentation. Many studies are hampered by issues with limitation of the dataset used in the model, methodological mistakes, poor reproducibility, and biases in study design. Most studies focused on achieving the best accuracy for their algorithms, sometimes putting aside validity and applicability in clinical practice. Indeed, only two articles presented with an overall low risk of bias.

The common limitations concerned datasets used for the model development, definition of the ground truth for evaluation of the model and strategies used for model evaluation.

Regarding the datasets used, some are private, and some are public open source. For private databases, advanced technical characteristics of images (e.g., imaging sequence, field of view, noise) used and patient's inclusion and exclusion criteria were poorly or not described. Most databases lacked representability of patients' variability as prostate volume, prostate tissue heterogeneity, prostatic pathology as PCa or benign hypertropia. Open-source prostate MRI databases also have several limitations such as selection bias, limited annotations, low-resolution images, unclear terminology, lack of demographic statistics and of precise histologic data.

This can have a direct impact on the generalizability of the model developed. Indeed, it has been shown for example that prostate morphological differences contribute to segmentation variability: Montagne et al. [48], showed that the smaller the prostate volume was, the higher the variability was; several authors [18, 39, 43] found poorer performance of their model applied on special cases such as history of trans-urethral-resection of prostate (TURP), while most databases lacked representativity of patients variability.

Even though it is tedious and time-consuming, reference segmentation should require at least two trained readers because inter- and intra-rater variability can be significant. Quality of images (slice thickness, partial volume artifacts), apex or base location [48, 49] or prostate morphological differences [48] have been shown to decrease accuracy of segmentation. Meyer et al. [34] showed that training on segmentation obtained by a single reader introduced bias into the training data. Indeed, performance was higher when obtained from the expert who created the training data in comparison with evaluation against other expert segmentation. Aldoj et al. [39] emphasized the need for finely annotated sets as they improved overall performances of their algorithms, showing the greater importance of well annotated databases compared to large and coarsely annotated databases.

Quality of the resulting auto segmentation is evaluated against the corresponding reference segmentation, so called the ground truth. The main approach is manual delineation of the prostate zones performed by human experts. We found a great heterogeneity on the segmentation protocols and terminology used. Eighteen different types of prostate delineation were found; each

Table 5 Overview of segmentation methods with performance based on DSC. Number of articles reporting stratification by gland height, and reporting pre- or post-processing steps

First author, year of publication	Type	DSC results[†]				Stratification by gland height	Pre-processing details	Post-processing details
		WG	TZ	PZ	CG			
Cuocolo et al. [43]	CNN	0.9063*	–	0.7142*	0.8692*	×	✓	×
Bardis et al. [42]	CNN	0.94	0.91	0.774	–	×	✓	×
Lai et al. [41]	CNN	–	0.93	0.7004	–	×	✓	×
Nai et al. [18]	CNN	0.89*	–	0.712*	0.856*	✓	✓	×
Sanford et al. [40]	CNN	0.915	0.89	–	–	×	✓	×
Aldoj et al. [39]	CNN	0.921*	–	0.781*	0.895*	✓	✓	×
Zavala-Romero et al. [6]	CNN	0.825[a] 0.892[b]	–	0.788 [a] 0.811 [b]	–	×	✓	✓
Lee et al. [38]	CNN	0.8712	0.7648	–	–	×	✓	×
Liu et al. [37]	CNN	–	0.89*[c] 0.87*[d]	0.80*[c] 0.79*[d]	–	✓	✓	×
Qin et al. [36]	CNN	–	–	0.806	0.901	×	✓	✓
Motamed et al. [35]	CNN	0.89[e] 0.85[f]	0.86[e] 0.84[f]	–	–	×	×	✓
Zabihollahy et al. [13]	CNN	0.9533[g] 0.9209[h]	–	0.8678[g] 0.861[h]	0.9375[g] 0.8989[h]	✓	✓	✓
Padgett et al. [8]	Atlas	0.83*	0.75*	0.59*	–	✓	×	×
Rundo et al. [15][1]	CNN	–	–	0.919[i] 0.831[j] 0.801[k]	0.871[i] 0.886[j] 0.937[k]	×	✓	✓
Meyer et al. [34]	CNN	–	0.876	0.798	–	×	✓	✓
Liu et al. [16]	CNN	–	0.86[c] 0.79[d]	0.74[c] 0.74[d]	–	✓	✓	×
Rundo et al. [33][2]	CNN	–	–	0.91* (with pre-training)	0.85* (with pre-training)	×	✓	✓
Hambarde et al. [32]	CNN	–	–	0.8733	–	×	✓	×
Jensen et al. [31]	CNN	–	–	0.692	0.794	✓	✓	✓
Khan et al. [17]	CNN	–	–	0.703*	0.88*	×	×	×
Cheng et al. [30]	CNN	0.9235*	–	–	0.9006*	✓	✓	✓
Zhu et al. [12]	CNN	0.927	–	0.793	–	✓	✓	×
Mooij et al. [29]	CNN	–	0.85*	0.6*	–	×	✓	×
Can et al. [28]	CNN	–	–	0.722*	0.89*	×	×	×
Clark et al. [14]	CNN	0.886[c] 0.862[d]	0.847[c]	–	–	×	✓	×
Chilali et al. [9]	C means + Atlas	0.9478	0.7023	0.62	–	×	✓	×
Makni et al. [10]	C means	–	0.88	0.78	–	×	✓	×
Chi et al. [27]	Gaussian model	0.8	–	0.53	0.83	×	×	×
Toth et al. [26]	Active appearance model	0.81	–	0.68[l] 0.60[m]	0.79[l] 0.72[m]	✓	✓	×
Litjens et al. [7]	Atlas	–	–	0.75	0.8	×	×	×
Moschidis and Graham [25]	Random Forrest + Graph Cuts	–	–	–	–	×	✓	×
Yin et al. [24]	Graph Cuts	–	–	–	0.81	×	×	×
Makni et al. [23]	C means	–	–	0.76[l]	0.87[l]	✓[¶]	×	×

CNN convolutional neural network

[†] Dice similarity coefficient (DSC) for whole gland (WG), transition zone (TZ), peripheral zone (PZ) or central gland (CG) (*means*)

*Best results if several models were tested

[¶] no Dice Similary Coefficien (DSC) provided

[a,b] Trained on combined datasets and, respectively, tested on GE[a] or Siemens[b] dataset

Table 5 (continued)

[c,d] Respectively for testing on internal[c] or external[d] data

[e,f] Respectively for source[e] or target[f] with 115 patients for training (best results)

[g,h] Respectively for T2-weighted[g] and apparent diffusion coefficient (ADC) map[h]

[i,j,k] Trained on combined datasets and, respectively, tested on dataset #1[i], #2[j] or #3[k]

[l,m] Using pre-segmented whole gland (WG)[l], or with whole process[m]

[1] Rundo et al., USE-Net: incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets[30]

[2] Rundo et al., CNN-based Prostate Zonal Segmentation on T2-weighted MR Images: A Cross-dataset Study[27]

anatomical zone was segmented directly or obtained by subtraction from one region to another (resulting in CZ, AFMS and PZ, which can be obtained either by delineation or by subtraction of WG and TZ). Terminology used was extremely variable from one study to another and did not always respect the one used and referenced in the PI-RADS [3, 50] (for example, use of "central gland" instead of CZ or TZ).

Number of readers, level of expertise, inter- and intra-variability evaluation were mostly absent, limiting the generalizability of the developed models due to inter-observer variability. Only 2/33 studies [10, 23] used blinded reading for ground truth. Nonetheless, prostate segmentation is a very challenging task. The prostate gland usually has fuzzy boundaries. Pixel intensities are heterogeneous both inside and outside the prostate, and contrasts and pixel intensities are very similar for prostate and non-prostate regions. The manual delineation of the prostate zones is therefore limited by the subjective interpretation of the organ boundaries. Becker et al. [49] found in a multi-reader study a higher variability at the extreme part of the gland (apex and base) and for the TZ delineation. Similar results were found by Padgett et al. [8] who found a difference of DSC from 0.88 to 0.81 for WG and TZ. Meyer et al. [34] showed that training on segmentation obtained by a single reader introduced bias into the training data.

Strategies used for model evaluation were limited by the lack of external validation Only 7 studies [6, 9, 14, 30, 33, 36, 37] used both private and public data to evaluate their model. The absence of an external testing dataset is a critical limitation to the clinical applicability of the developed models. Data augmentation and transfer learning were also used to help addressing this issue [6, 14–16, 29, 31, 33, 35–41, 43, 51]. It is important to note that some bias cannot be balanced-out by increasing the sample size by data augmentation or repetition of training. For example, data augmentation of a dataset constituted without prostate cancer patients cannot decrease risk of bias induced by the more homogeneous contours it provides.

Even without data augmentation, MRI images contains wide heterogeneity and most of the times pre-processing steps involving intensity normalization or noise reduction to remove confounding features and improve image quality are necessary [52]. Some authors [6, 13, 15, 31, 33, 35, 51] also reported post-processing. Not reporting some of the pre- or post-processing steps can affect reproducibility and sufficient detail enables readers to determine the quality and generalizability of the work. While several checklists can be used such as those from Enhancing the Quality and Transparency Of health Research (EQUATOR) Network guidelines [53], the use of the recently published Checklist for Artificial Intelligence in Medical Imaging [22] would be helpful to lower risk of bias of ongoing work.

In the future, there is a need for well-sampled databases including large number of representative cases for the anatomical variability of the prostate gland and technical specificities (2D T2 versus 3D T2, slice thickness, FOV, vendors) to account for the anatomical, disease related, acquisition related variabilities, with a multi-readers segmentations and a well-defined delineation guideline of the prostate (as it is already done for example in organs at risk for radiotherapy planning [54]).

Constitution of quality database should be based on latest PI-RADS recommendations, by associating quality criteria such as the consensual quality requirements ESUR/ESUI [55] or Prostate Imaging Quality (PI-QUAL) [56] score to guarantee essential image quality for zonal segmentation and tumor detection.

The main limitation of this review is the absence of details of technical information used; each study making its own contribution for networks with countless hyperparameters, sometimes without enough details to be gathered. This precluded us from comparing models' accuracy without bias.

Some other relevant papers also could be missing because of incongruences between search terms, article keywords, or indexing in the databases, such as for conference proceedings papers. In particular, databases such as ArXiv were not searched as it also provides access to preprints, without peer review.

Conclusion

This review systematically synthesizes published automatic prostate zonal segmentation methods using MRI. We found that no papers in the literature currently have

both sufficiently documented datasets selection and segmentation criteria and enough external validation.

This underlines the critical need for higher quality datasets, a documented reproducible method and terminology for zonal segmentation and sufficient external dataset to develop the best quality methods free from biases: an essential step for future development of automatic detection of prostate cancer.

Abbreviations

AFMS: Anterior fibro-muscular stroma; CG: Central gland; CZ: Central zone; DSC: Dice similarity coefficient; FOV: Field of view; PCa: Prostate cancer; PZ: Peripheral zone; QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies 2; TZ: Transition zone; WG: Whole gland.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13244-022-01340-2.

Additional file 1: Figure S1. Schematic of the 17 types of protocol of zonal segmentation. Type A1:Detailed segmentation of WG, PZ and CG (including CZ + TZ + AFMS) [23, 43]. Type A2:Detailed segmentation of WG, CG (including CZ + TZ + AFMS) [25]. Type A3:Detailed segmentation of PZ, CG (including CZ + TZ + AFMS) [36]. Type A4:Segmentation of PZ and urethra only [32]. Type B1a: Segmentation of PZ, CG (including TZ and CZ), without detail for AFMS [31]. Type B1b: Detailed segmentation of PZ, «TZ» (including TZ and CZ), AFMS not segmented [37]. Type B2a: Segmentation of WG, PZ, CG (including TZ + CZ) without detail for AFMS (does not seem segmented) [13]. Type B2b: Segmentation of WG, PZ, «TZ» (including TZ + CZ), without detail for AFMS [8]. Type B2c: Segmentation of WG, PZ, CG (including TZ + CZ) without detail for AFMS (seems segmented with PZ) [9, 26]. Type B3: Detailed segmentation of PZ, CG (including TZ et CZ) and AFMS [34]. Type C1a: Segmentation of PZ, CG without detail for CZ or AFMS [7, 15, 17, 28, 33]. Type C1b: Segmentation of TZ et PZ, with WG = CG + PZ, without detail for CZ or AFMS [10, 16, 29, 41]. Type C2: Segmentation of WG, PZ et CG, without detail for CZ or AFMS [18, 27, 39]. Type C3: Detailed segmentation of WG, «TZ» (including TZ + AFMS). PZ and CZ were not segmented [40]. Type C4:Segmentation of WG and PZ, with WG − PZ = CG, without detail for CZ or AFMS [6, 12]. Type D1a:Segmentation of WG, TZ without detail for CZ or AFMS [14, 35, 38]. Type D1b:Segmentation of WG, CG, without detail for CZ or AFMS [24, 30]. Type D2:Segmentation of WG, TZ, PZ, without detail for CZ or AFMS [42]. Segmentation protocols details not reported in full text were extrapolated from figures of those articles. CZ: central zone. TZ: transition zone. AFMS: anterior fibro-muscular stroma. PZ: peripheral zone. CG: central gland. WG: whole gland. **Table S1.** Overview of most used public databases characteristics. **Table S2.** Detailed quality assessment for risk of bias and applicability of included studies.

Acknowledgements

The authors thank Thomas Desmousseaux for his help with prostate drawings.

Author contributions

CW (corresponding author) has made substantial contributions to the conception, design of the work, acquisition, analysis, interpretation of data, drafted the work and substantively revised it. SM has made substantial contributions to the conception, analysis, interpretation of data, revised the work. DH has made substantial contributions to the analysis and interpretation of data, and revised the work. NA has made substantial contributions to revise the work. HD has made substantial contributions to revise the work. RRP has made substantial contributions to the conception and design of the work, interpretation of data, has drafted and substantially revised the work. All authors read and approved the final manuscript.

Funding

The authors state that this work has not received any funding.

Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

Declarations

Ethics approval and consent to participate

Institutional Review Board approval was not required because it is a systematic review of literature, which does not involve research on human subjects.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Sorbonne Université, Paris, France. [2]Academic Department of Radiology, Hôpital Tenon, Assistance Publique des Hôpitaux de Paris, 4 Rue de La Chine, 75020 Paris, France. [3]Academic Department of Radiology, Hôpital Pitié-Salpétrière, Assistance Publique des Hôpitaux de Paris, Paris, France. [4]GRC N° 5, Oncotype-Uro, Sorbonne Université, Paris, France. [5]Inria, Epione Team, Sophia Antipolis, Université Côte d'Azur, Nice, France.

Received: 11 August 2022 Accepted: 27 November 2022
Published online: 21 December 2022

References

1. Mottet N, van den Bergh RCN, Briers E et al (2021) EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer—2020 update. Part 1: screening, diagnosis, and local treatment with curative intent. Eur Urol 79:243–262. https://doi.org/10.1016/j.eururo.2020.09.042
2. Rozet F, Hennequin C, Beauval J-B et al (2018) Recommandations françaises du Comité de Cancérologie de l'AFU—Actualisation 2018–2020:cancer de la prostate. Prog Urol 28:R81–R132. https://doi.org/10.1016/j.purol.2019.01.007
3. Turkbey B, Rosenkrantz AB, Haider MA et al (2019) Prostate imaging reporting and data system version 2.1:2019 update of prostate imaging reporting and data system version 2. Eur Urol 76:340–351. https://doi.org/10.1016/j.eururo.2019.02.033
4. Benson MC, Seong WI, Pantuck A et al (1992) Prostate specific antigen density: a means of distinguishing benign prostatic hypertrophy and prostate cancer. J Urol 147:815–816. https://doi.org/10.1016/S0022-5347(17)37393-7
5. Korsager AS, Fortunati V, van der Lijn F et al (2015) The use of atlas registration and graph cuts for prostate segmentation in magnetic resonance images. Med Phys 42:1614–1624. https://doi.org/10.1118/1.4914379
6. Zavala-Romero O, Breto AL, Xu IR et al (2020) Segmentation of prostate and prostate zones using deep learning: a multi-MRI vendor analysis. Strahlenther Onkol 196:932–942. https://doi.org/10.1007/s00066-020-01607-x
7. Litjens G, Debats O, van de Ven W et al (2012) A pattern recognition approach to zonal segmentation of the prostate on MRI. In: Ayache N, Delingette H, Golland P, Mori K (eds) Medical image computing and computer-assisted intervention—MICCAI 2012. Springer, Berlin, pp 413–420
8. Padgett KR, Swallen A, Pirozzi S et al (2019) Towards a universal MRI atlas of the prostate and prostate zones: comparison of MRI vendor and image acquisition parameters. Strahlenther Onkol 195:121–130. https://doi.org/10.1007/s00066-018-1348-5
9. Chilali O, Puech P, Lakroum S et al (2016) Gland and zonal segmentation of prostate on T2W MR images. J Digit Imaging 29:730–736. https://doi.org/10.1007/s10278-016-9890-0

10. Makni N, Betrouni N, Colot O (2014) Introducing spatial neighbourhood in Evidential C-Means for segmentation of multi-source images: application to prostate multi-parametric MRI. Inf Fusion 19:61–72. https://doi.org/10.1016/j.inffus.2012.04.002
11. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. arXiv:1505.04597 [cs] (2015).
12. Zhu Y, Wei R, Gao G et al (2019) Fully automatic segmentation on prostate MR images based on cascaded fully convolution network. J Magn Reson Imaging 49:1149–1156. https://doi.org/10.1002/jmri.26337
13. Zabihollahy F, Schieda N, Krishna Jeyaraj S, Ukwatta E (2019) Automated segmentation of prostate zonal anatomy on T2-weighted (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets. Med Phys 46:3078–3090. https://doi.org/10.1002/mp.13550
14. Clark T, Zhang J, Baig S et al (2017) Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks. J Med Imag 4:1. https://doi.org/10.1117/1.JMI.4.4.041307
15. Rundo L, Han C, Nagano Y, et al. USE-Net: incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. arXiv:1904.08254 [cs] (2019).
16. Liu Y, Sung K, Yang G et al (2019) Automatic prostate zonal segmentation using fully convolutional network with feature pyramid attention. IEEE Access 7:163626–163632. https://doi.org/10.1109/ACCESS.2019.2952534
17. Khan Z, Yahya N, Alsaih K, Meriaudeau F. Zonal segmentation of prostate T2W-MRI using atrous convolutional neural network. In: 2019 IEEE student conference on research and development (SCOReD). Bandar Seri Iskandar: IEEE; 2019. p. 95–99.
18. Nai Y-H, Teo BW, Tan NL et al (2020) Evaluation of multimodal algorithms for the segmentation of multiparametric MRI prostate images. Comput Math Methods Med 2020:8861035. https://doi.org/10.1155/2020/8861035
19. Page MJ, McKenzie JE, Bossuyt PM et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. https://doi.org/10.1136/bmj.n71
20. Booth A, Clarke M, Dooley G, et al. An international registry of systematic-review protocols. Lancet. 2012. https://doi.org/10.1016/S0140-6736(10)60903-8.
21. Bristol U of QUADAS-2. https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/. Accessed 2 Jul 2021.
22. Mongan J, Moy L, Kahn CE (2020) Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. Radiol Artif Intell 2:e200029. https://doi.org/10.1148/ryai.2020200029
23. Makni N, Iancu A, Colot O et al (2011) Zonal segmentation of prostate using multispectral magnetic resonance images: zonal segmentation of prostate using multispectral MR images. Med Phys 38:6093–6105. https://doi.org/10.1118/1.3651610
24. Yin Y, Fotin SV, Periaswamy S, et al. Fully automated 3D prostate central gland segmentation in MR images: a LOGISMOS based approach. In: Haynor DR, Ourselin S, editors. San Diego, California, USA; 2012. p. 83143B.
25. Moschidis E, Graham J. Automatic differential segmentation of the prostate in 3-D MRI using Random Forest classification and graph-cuts optimization. In: 2012 9th IEEE international symposium on biomedical imaging (ISBI). 2012. p. 1727–1730.
26. Toth R, Ribault J, Gentile J et al (2013) Simultaneous segmentation of prostatic zones using active appearance models with multiple coupled levelsets. Comput Vis Image Underst 117:1051–1060. https://doi.org/10.1016/j.cviu.2012.11.013
27. Chi Y, Ho H, Law YM, et al. A compact method for prostate zonal segmentation on multiparametric MRIs. In: Medical imaging 2014: image-guided procedures, robotic interventions, and modeling. International Society for Optics and Photonics. 2014. p. 90360N.
28. Can YB, Chaitanya K, Mustafa B et al (2018) Learning to segment medical images with scribble-supervision alone. In: Stoyanov D, Taylor Z, Carneiro G et al (eds) Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer International Publishing, Cham, pp 236–244
29. Mooij G, Bagulho I, Huisman H. Automatic segmentation of prostate zones. arXiv:1806.07146 [cs] (2018).
30. Cheng R, Lay N, Roth HR et al (2019) Fully automated prostate whole gland and central gland segmentation on MRI using holistically nested networks with short connections. J Med Imaging (Bellingham) 6:024007. https://doi.org/10.1117/1.JMI.6.2.024007
31. Jensen C, Sørensen KS, Jørgensen CK et al (2019) Prostate zonal segmentation in 1.5T and 3T T2W MRI using a convolutional neural network. J Med Imaging 6:1. https://doi.org/10.1117/1.JMI.6.1.014501
32. Hambarde P, Talbar SN, Sable N et al (2019) Radiomics for peripheral zone and intra-prostatic urethra segmentation in MR imaging. Biomed Signal Process Control 51:19–29. https://doi.org/10.1016/j.bspc.2019.01.024
33. Rundo L, Han C, Zhang J, et al. CNN-based prostate zonal segmentation on T2-weighted MR images: a cross-dataset study. arXiv:1903.12571 [cs] (2019).
34. Meyer A, Rakr M, Schindele D, et al. Towards patient-individual PI-Rads v2 sector map: Cnn for automatic segmentation of prostatic zones from T2-weighted MRI. In: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). Venice: IEEE; 2019. p. 696–700.
35. Motamed S, Gujrathi I, Deniffel D, et al. A transfer learning approach for automated segmentation of prostate whole gland and transition zone in diffusion weighted MRI. arXiv:1909.09541 [cs, eess, q-bio] (2020)
36. Qin X, Zhu Y, Wang W et al (2020) 3D multi-scale discriminative network with multi-directional edge loss for prostate zonal segmentation in bi-parametric MR images. Neurocomputing 418:148–161. https://doi.org/10.1016/j.neucom.2020.07.116
37. Liu Y, Yang G, Hosseiny M et al (2020) Exploring uncertainty measures in Bayesian deep attentive neural networks for prostate zonal segmentation. IEEE Access 8:151817–151828. https://doi.org/10.1109/ACCESS.2020.3017168
38. Lee DK, Sung DJ, Kim C-S et al (2020) Three-dimensional convolutional neural network for prostate MRI segmentation and comparison of prostate volume measurements by use of artificial neural network and ellipsoid formula. AJR Am J Roentgenol 214:1229–1238. https://doi.org/10.2214/AJR.19.22254
39. Aldoj N, Biavati F, Michallek F et al (2020) Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net. Sci Rep 10:14315. https://doi.org/10.1038/s41598-020-71080-0
40. Sanford TH, Zhang L, Harmon SA et al (2020) Data augmentation and transfer learning to improve generalizability of an automated prostate segmentation model. AJR Am J Roentgenol 215:1403–1410. https://doi.org/10.2214/AJR.19.22347
41. Lai C-C, Wang H-K, Wang F-N et al (2021) Autosegmentation of prostate zones and cancer regions from biparametric magnetic resonance images by using deep-learning-based neural networks. Sensors 21:2709. https://doi.org/10.3390/s21082709
42. Bardis M, Houshyar R, Chantaduly C et al (2021) Segmentation of the prostate transition zone and peripheral zone on MR images with deep learning. Radiol Imaging Cancer 3:e200024. https://doi.org/10.1148/rycan.2021200024
43. Cuocolo R, Comelli A, Stefano A et al (2021) Deep learning whole-gland and zonal prostate segmentation on a public MRI dataset. J Magn Reson Imaging. https://doi.org/10.1002/jmri.27585
44. Armato SG, Huisman H, Drukker K et al (2018) PROSTATEx challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images. J Med Imag 5:1. https://doi.org/10.1117/1.JMI.5.4.044501
45. Nicholas Bloch AM. NCI-ISBI 2013 challenge: automated segmentation of prostate structures. The Cancer Imaging Archive; 2015.
46. Litjens G, Toth R, van de Ven W et al (2014) Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. Med Image Anal 18:359–373. https://doi.org/10.1016/j.media.2013.12.002
47. Taha AA, Hanbury A (2015) Metrics for evaluating 3D medical image segmentation:analysis, selection, and tool. BMC Med Imaging 15:29. https://doi.org/10.1186/s12880-015-0068-x
48. Montagne S, Hamzaoui D, Allera A et al (2021) Challenge of prostate MRI segmentation on T2-weighted images:inter-observer variability and impact of prostate morphology. Insights Imaging 12:71. https://doi.org/10.1186/s13244-021-01010-9
49. Becker AS, Chaitanya K, Schawkat K et al (2019) Variability of manual segmentation of the prostate in axial T2-weighted MRI: a multi-reader study. Eur J Radiol 121:108716. https://doi.org/10.1016/j.ejrad.2019.108716

50. Weinreb JC, Barentsz JO, Choyke PL et al (2016) PI-RADS prostate imaging—reporting and data system: 2015, Version 2. Eur Urol 69:16–40. https://doi.org/10.1016/j.eururo.2015.08.052
51. Meyer A, Chlebus G, Rak M et al (2021) Anisotropic 3D multi-stream CNN for accurate prostate segmentation from multi-planar MRI. Comput Methods Programs Biomed 200:105821. https://doi.org/10.1016/j.cmpb.2020.105821
52. Gupta J, Saini SK, Juneja M (2020) Survey of denoising and segmentation techniques for MRI images of prostate for improving diagnostic tools in medical applications. Mater Today Proc 28:1667–1672. https://doi.org/10.1016/j.matpr.2020.05.023
53. The EQUATOR Network|Enhancing the QUAlity and Transparency of Health Research. https://www.equator-network.org/. Accessed 5 Aug 2021.
54. Vrtovec T, Močnik D, Strojan P et al (2020) Auto-segmentation of organs at risk for head and neck radiotherapy planning: From atlas-based to deep learning methods. Med Phys 47:e929–e950. https://doi.org/10.1002/mp.14320
55. 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
56. Giganti F, Kirkham A, Kasivisvanathan V et al (2021) Understanding PI-QUAL for prostate MRI quality: a practical primer for radiologists. Insights Imaging 12:59 https://doi.org/10.1186/s13244-021-00996-6

Publisher's Note

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

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com