





Article

Stability of Multi-Parametric Prostate MRI Radiomic Features to Variations in Segmentation

Sithin Thulasi Seetha ^{1,2,†,‡} , Enrico Garanzini ^{3,†}, Chiara Tenconi ^{4,5,*}, Cristina Marenghi ⁶ , Barbara Avuzzi ⁷, Mario Catanzaro ⁸, Silvia Stagni ⁸, Sergio Villa ⁷, Barbara Noris Chiorda ⁷, Fabio Badenchini ⁶, Elena Bertocchi ⁶, Sebastian Sanduleanu ², Emanuele Pignoli ⁴, Giuseppe Procopio ⁶, Riccardo Valdagni ^{1,5} , Tiziana Rancati ^{9,*} , Nicola Nicolai ^{8,§} and Antonella Messina ^{3,§}

- ¹ Prostate Cancer Program, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milan, Italy; s.thulasiseetha@maastrichtuniversity.nl (S.T.S.); riccardo.valdagni@istitutotumori.mi.it (R.V.)
- ² Department of Precision Medicine, GROW—School for Oncology and Developmental Biology, Maastricht University, 6211 LK Maastricht, The Netherlands
- ³ Department of Radiology, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milan, Italy; enrico.garanzini@istitutotumori.mi.it (E.G.); antonella.messina@istitutotumori.mi.it (A.M.)
- ⁴ Department of Medical Physics, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milan, Italy; emanuele.pignoli@istitutotumori.mi.it
- ⁵ Department of Oncology and Hematooncology, Università degli Studi di Milano, 20133 Milan, Italy
- ⁶ Unit of Genito-Urinary Medical Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milan, Italy; cristina.marenghi@istitutotumori.mi.it (C.M.); fabio.badenchini@istitutotumori.mi.it (F.B.); elena.bertocchi@istitutotumori.mi.it (E.B.); giuseppe.procopio@istitutotumori.mi.it (G.P.)
- ⁷ Department of Radiation Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milan, Italy; barbara.avuzzi@istitutotumori.mi.it (B.A.); sergio.villa@istitutotumori.mi.it (S.V.); barbara.noris@istitutotumori.mi.it (B.N.C.)
- ⁸ Department of Urology, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milan, Italy; mario.catanzaro@istitutotumori.mi.it (M.C.); silvia.stagni@istitutotumori.mi.it (S.S.); nicola.nicolai@istitutotumori.mi.it (N.N.)
- ⁹ Data Science Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milan, Italy
- * Correspondence: chiara.tenconi@istitutotumori.mi.it (C.T.); tiziana.rancati@istitutotumori.mi.it (T.R.)
- † These authors contributed equally to this work.
- ‡ Current address: Department of Clinical, Surgical, Diagnostic and Pediatric Sciences, University of Pavia, 27100 Pavia, Italy.
- § These authors contributed equally to this work.



Citation: Thulasi Seetha, S.; Garanzini, E.; Tenconi, C.; Marenghi, C.; Avuzzi, B.; Catanzaro, M.; Stagni, S.; Villa, S.; Chiorda, B.N.; Badenchini, F.; et al. Stability of Multi-Parametric Prostate MRI Radiomic Features to Variations in Segmentation. *J. Pers. Med.* **2023**, *13*, 1172. <https://doi.org/10.3390/jpm13071172>

Academic Editor: Serena Monti

Received: 11 May 2023
Revised: 13 July 2023
Accepted: 18 July 2023
Published: 22 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Stability analysis remains a fundamental step in developing a successful imaging biomarker to personalize oncological strategies. This study proposes an in silico contour generation method for simulating segmentation variations to identify stable radiomic features. Ground-truth annotation provided for the whole prostate gland on the multi-parametric MRI sequences (T2w, ADC, and SUB-DCE) were perturbed to mimic segmentation differences observed among human annotators. In total, we generated 15 synthetic contours for a given image-segmentation pair. One thousand two hundred twenty-four unfiltered/filtered radiomic features were extracted applying Pyradiomics, followed by stability assessment using ICC(1,1). Stable features identified in the internal population were then compared with an external population to discover and report robust features. Finally, we also investigated the impact of a wide range of filtering strategies on the stability of features. The percentage of unfiltered (filtered) features that remained robust subjected to segmentation variations were T2w—36% (81%), ADC—36% (94%), and SUB—43% (93%). Our findings suggest that segmentation variations can significantly impact radiomic feature stability but can be mitigated by including pre-filtering strategies as part of the feature extraction pipeline.

Keywords: radiomics; multi-parametric MRI; prostate

1. Introduction

Multi-parametric MRI (mpMRI), including T2-weighted (T2w), diffusion-weighted imaging (DWI), and dynamic contrast-enhanced (DCE) images, has become an essential tool for the detection and characterization of prostate cancer (PCa) [1–3]. Its role has extended beyond tumor staging to encompass cancer detection and monitoring of disease progression during active surveillance (AS) [4–6]. The use of mpMRI in AS seems particularly attractive. Combining morphological and functional images constitutes a non-invasive tool for longitudinal monitoring of patients, interrogating the entire prostate volume, and possibly giving information on the indolence or aggressiveness of the prostate tissue. Presently, mpMRI is widely used in AS to assess an image-based risk stratification score following the guidelines from the prostate imaging reporting and data system (PI-RADS) [7]. However, this semi-quantitative approach relies entirely on standardized acquisition and reporting guidelines.

The increasing use of mpMRI among patients in active surveillance makes radiomics highly attractive. Radiomics is a quantitative approach to medical image analysis that aims to capture information beyond what is visible to the naked eye [8,9]. Only a handful of studies have investigated the utility of mpMRI radiomics features in the premise of AS. Sushentsev, Nikita, et al. [10] examined the complementary value of radiomics features to improve baseline prediction of PCa progression. Another study by Sushentsev et al. [11] compared the performance of delta-radiomics [12] and MRI-derived PRECISE [13] scores in progression prediction. Algohary, Ahmad, et al. [14] evaluated the performance of radiomics in identifying the presence of clinically significant PCa in AS patients. A few other studies [15,16] focused on clinical features and/or a chosen set of shape and first-order features extracted from MRI sequences for progression prediction. Albeit, these studies only included patients with MRI visible lesions wherein lesions served as the ROI for extracting radiomics features. Since around half of the patients in AS population are likely to show MR-visible lesions [17–19], this excludes almost half of the patients enrolled in surveillance studies.

Furthermore, these studies only considered features extracted from bi-parametric (bp) MRI—T2w and ADC (derived from DWI) for predictive modeling. Notably, DCE is acquired as part of routine AS protocol following PIRADS specification. Excluding DCE sequences results in losing readily available diagnostic information [20].

Generally, these image-based signatures must be highly robust to develop reliable models for routine clinical practice. Developing a robust model means setting radiomics in the “big data” analysis framework. Such a model requires extensive training and validation sets from multicentric studies with image data derived from a large patient population for a specific pathology. This introduces the complication of radiomic feature variability due to differences in scanners [21,22], imaging acquisition parameters/protocols [23], reconstruction algorithms [24,25], processing pipelines [26,27], and the annotation of the region of interest (ROI) [28,29]. The variability due to these sources may hide any potential signal from tumor biology, making at least some of the radiomic signatures unreliable and thus hindering the generalization of results.

Delineating the ROI is an essential step before all image-based medical interventions. This is a tedious task, and even in the best scenario (segmentation carried out following standardized and quantitative guidelines), inter- and/or intra-observer variability among trained radiologists is observed. We may attribute these differences to the behavior of radiologists in a clinical setting, where some are more conservative or liberal regarding segmentation. Often, a slight difference in ROIs results in different radiomics feature values [30–32], commonly referred to as feature instability. Developing signatures using such unstable features ultimately leads to lower robustness of signatures. Although many studies have already been carried out to tackle this issue, there are only a few studies in the context of AS. For example, Xu, Lili, et al. [33] and Zhang, Gu-mu-yang, et al. [34] included feature stability assessment to variations in lesion segmentation as a feature filtering strategy earlier in their predictive pipeline. Two radiologists were involved in

these studies, and ICC [35] was used as the statistical metric to measure stability. Conversely, Chen, Tong, et al. [36] merged manual lesion segmentations by two radiologists before feature extraction to reduce the impact of inter-observer variations. Fehr, Duc, et al. [37] employed a segmentation approach where three readers were involved in a consensual delineation of tumor and non-cancerous prostate regions as part of their study.

Generally, it is recommended to include at least three raters in stability assessments [35]. In studies measuring feature stability to segmentation variation, this number usually falls within the range of 2–5 [38–40]. However, obtaining multiple radiologists for segmentation is a challenging task and is often infeasible. One solution could be the application of morphological operations on the ROI to generate perturbed segmentations for feature stability assessment. Sushentsev, Nikita, et al. [10,11] followed this strategy where two versions of ROI were generated by subjecting the original ROI to morphological operations—opening and closing using a spherical structuring element of a 1-pixel radius. However, this method does not approximate clinical inter/intra-observer variations wherein the differences are non-deterministic. An alternative solution was proposed by Haarburger et al. using a probabilistic U-Net model, which was used to generate 25 plausible segmentations [41]. Using this approach, they discovered a set of features stable to variations in segmentation. However, the probabilistic U-Net model suffered from limited segmentation diversity, which can bias the results. A recent extension to this was proposed by the same authors who used PHiseg model to address some of the limitations of their previous work [29]; they even included four radiologists for clinical inter/intra-observer variations analysis. Although using such generative models can scale up such studies, there are a few caveats. The computation cost and resource requirement for training and tuning a generative deep learning (DL) model is quite high. Expertise in DL is also essential to customize and integrate such a model into a radiomics pipeline.

In this work, we try to address these limitations by proposing a simple *in silico* contour generation method inspired by the data augmentation paradigm in DL. We have considered the whole prostate gland as the ROI to endorse the inclusion of AS patients with no MR-visible lesions for future predictive modeling studies. On this account, we also considered the acquired DCE sequence as part of our pipeline routine AS protocol. We intend to simulate various clinical segmentation scenarios using a combination of linear transformations such as rotation, scaling, and shifting that follows a set of predetermined constraints to simulate the behavior of manual annotators. The stable features identified in the internal population will then be compared with an external dataset to report a set of overlapping stable features (i.e., robust features) that could be utilized in future predictive modeling studies.

2. Materials and Methods

2.1. Datasets

2.1.1. Internal Dataset

We included one hundred patients diagnosed with very low-risk prostate cancer and enrolled in active surveillance at the Fondazione IRCCS Istituto Nazionale dei Tumori in Milan. The local Ethics Committee approved the study protocol (INT 113/16, INT 46/07, and INT 95/11), and all patients signed a written informed consent for the study.

MRI acquisitions were performed using an “Ingenia” 1.5 T (Philips Medical System, Best, The Netherlands) equipped with 32-channel phased-array and spine coils in combination with an endorectal receiver coil. Images were acquired using Turbo Spin Echo and Gradient Echo sequences, always including a sequence with axial slicing, according to the PI-RADS v2.1 [7] recommendations. The acquisition protocol was standardized: every set of images included T2w images (TR/TE = 4910/110 ms, slice thickness = 3 mm, pixel spacing = 0.297 mm) and two functional MRI sequences: DWI (b-values of 0, 1500, and 2000 s/mm², TR/TE = 3320/106 ms, slice thickness = 3 mm, pixel spacing = 1.250 mm) and DCE (TR/TE = 4.03/1.88 ms, slice thickness = 3 mm, pixel spacing = 1.136 mm). DCE was

acquired with a high temporal resolution (<10 s) during the administration of the contrast agent in the same position and phase encoding direction as T2w and DWI.

An experienced radiologist (E.G.) segmented the entire prostate gland on the T2w sequence. Interpolation proved sufficient to align the segmentation on the T2w with the other sequences, owed to the restricted motion and sequential acquisition of all the multi-parametric sequences.

We processed DWI and DCE sequences to generate the Apparent Diffusion Coefficient (ADC) and subtraction (SUB) maps. We derived the ADC map by computing the negative gradient associated with a least-square fit (straight line) over the DWI acquisitions with three b-values—0, 1500, and 2000 mm/s². We processed the DCE acquisitions to generate two subtraction (SUB) maps describing the wash-in (SUBwin) and wash-out (SUBwout) phases of the contrast agent. The maps were computed by splitting the DCE acquisitions at a time point close to 90s in the temporal domain, i.e., SUBwin = DCE_{90+ε}-DCE₀ and SUBwout = DCE_{t_n}-DCE_{90+ε}. This was performed to capture the contrast agent inflow (wash-in) and outflow (wash-out) phases which are known to guide radiologists in assessing the malignancy in PCa management [42]. Here, t_n indicates the last DCE acquisition in the temporal domain. ε represents the deviation from the referenced time point. Table 1a presents a simplified summary of the internal dataset properties, and Figure 1, panel (a) highlights mid-gland level axial mpMRI slices of a sample patient from the population.

Table 1. A simplified summary of the specifications for both the internal and external datasets used in this study.

Specifications	(a) Internal Dataset	(b) External Dataset
No. of Patients	100	15
Manufacturer	Ingenia (Philips Medical System, Best, The Netherlands)	GE Signa HDxt platform and GE Discovery MR750w (General Electric Healthcare, Milwaukee, WI) machines.
Magnetic Field Strength	1.5 T	3.0 T
Endorectal Coil	Yes	Yes
PIRADSv2 Compliant	Yes	Yes
Acquisition Protocol	T2w (TR/TE = 4910/110 ms, slice thickness = 3 mm, pixel spacing = 0.297 mm); DWI (b-values = 0, 1500 and 2000 s/mm ² , TR/TE = 3320/106 ms, slice thickness = 3 mm, pixel spacing = 1.250 mm); DCE (TR/TE = 4.03/1.88 ms, slice thickness = 3 mm, pixel spacing = 1.136 mm, acquired with high temporal resolution < 10 s).	T2w (TR/TE = 3350–5109/84–107 ms, slice thickness = 3 mm, pixel spacing = 0.273–0.312 mm); DWI (b-values of 0 and 1400 s/mm ² , TR/TE = 2500–8150/76.7–80.6 ms, slice thickness = 3–4 mm, pixel spacing = 0.625–0.703 mm); DCE (TR/TE = 3.68–4.1/1.3–1.42 ms, slice thickness = 5–6 mm, pixel spacing = 0.547–1.015 mm).
GT Segmentation	Whole prostate gland segmentation on T2w	Whole prostate gland segmentation on T2w, ADC, and SUB

2.1.2. External Dataset

QIN Prostate Repeatability is an open-source [43–45] prostate mpMRI test-retest dataset of 15 men with confirmed or suspected prostate cancer. mpMRI acquisitions were PIRADS v2 compliant and were performed using “GE Signa HDxt platform” and “GE Discovery MR750w” (General Electric Healthcare, Milwaukee, WI) machines. The images were acquired at a magnetic strength of 3.0 T in combination with an endorectal coil. Two scanners were used because of the hardware upgrade during the study. For each patient, the baseline and repeated examinations were taken on the same scanner at a two-week interval. Multi-parametric acquisitions included axial T2w images (TR/TE = 3350–5109/84–107 ms, slice thickness = 3 mm), DWI (b-values of 0 and 1400 s/mm², TR/TE = 2500–8150/76.7–80.6 ms,

slice thickness = 3–4 mm) and DCE (TR/TE = 3.68–4.1/1.3–1.42 ms, slice thickness = 5–6 mm) sequences.

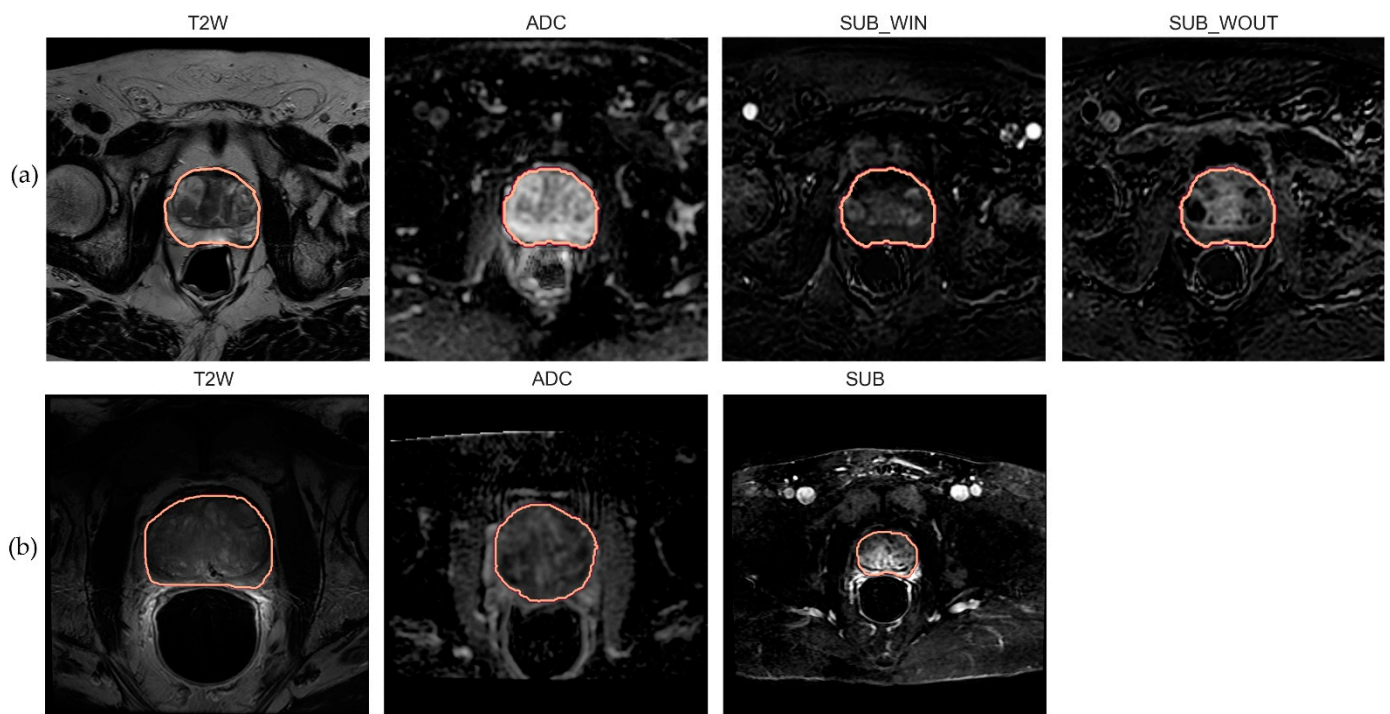


Figure 1. Mid-level axial slice of the prostate gland with ROI annotation. (a) T2w, ADC, SUBwin, and SUBwout images associated with a random patient sampled from the internal population; (b) T2w, ADC, and SUB images associated with a random patient sampled from the external population.

The in-built scanner software generated ADC and DCE SUB maps. The SUB map was computed as the difference between the phases involved in contrast bolus arrival to the baseline. Ultimately manual segmentation of the whole prostate gland (amongst other ROIs) was performed by an experienced radiologist for each sequence and was included in the dataset. Table 1b presents a simplified summary of the external dataset properties, and Figure 1, panel (b), highlights mid-gland level axial mpMRI slices of a sample patient from the population.

2.2. In Silico Contour Generation

To evaluate the impact of segmentation variations on radiomic feature stability, we synthetically generated 15 new prostate ROIs for each patient. We synthesized these contours by subjecting the manual ROI segmentation to bounded perturbations using affine transformations. The transformations include shifting, scaling, and rotation to simulate under-/over-segmentation variations. This approach was inspired by the data augmentation technique commonly used in deep learning [46,47]. TorchIO (v0.18.21) [48], a Python-based library for processing or augmenting 3D medical images, was used to generate synthetic contours dynamically.

By considering bounded (i.e., constrained) combinations of affine transformations, we systematically analyzed three categories of contour augmentations: in-plane, out-plane, and in and out-plane on each mpMRI sequence.

As the name suggests, in-plane augmentation essentially simulates the variability in contouring within the axial plane (i.e., variations within X and Y dimensions associated with a slice). Here the prostate contours are allowed to have variations in their latero-lateral or antero-posterior dimensions by a value randomly sampled from a uniform distribution within the interval $[-2.7 \text{ mm}, +2.7 \text{ mm}]$. In addition to this, the contour is randomly allowed to rotate around the z-axis at a small angle, $\alpha \sim U(-5^\circ, +5^\circ)$ (see Figure 2).

The choice of intervals for contour variability $[-2.7 \text{ mm}, +2.7 \text{ mm}]$ was established by following the results of studies on the inter-observer variability in prostate contouring using MRI [49,50]. These studies report an average standard deviation of 1.1 mm, corresponding to 2.7 mm at a 95% confidence interval.

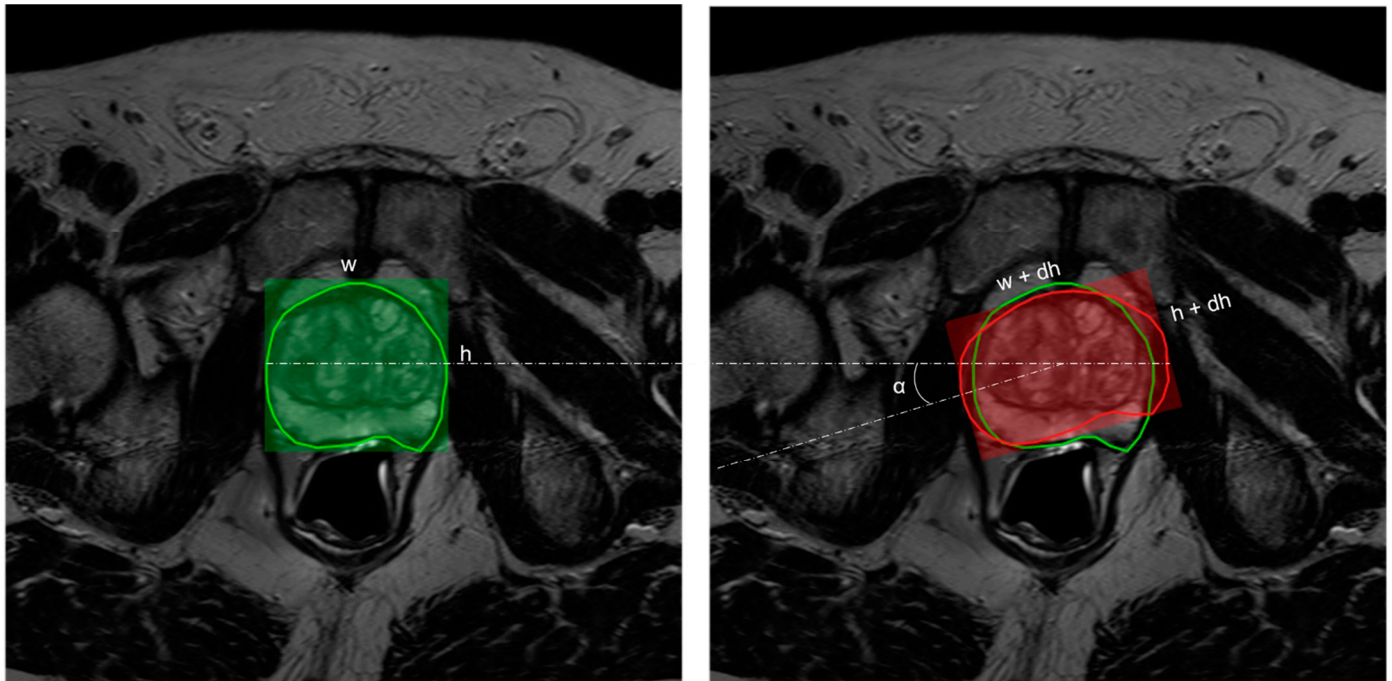


Figure 2. Illustration of in-plane augmentation. The width, w , and height, h , associated with the contour drawn by the radiologist (left in green) in the axial plane are allowed to undergo random perturbation by a delta value— $dw, dh \sim U(-2.7 \text{ mm}, 2.7 \text{ mm})$. This results in a transformed contour (right in red) with width, $w' = w + dw$, and height, $h' = h + dh$. In addition to this, the contour is also allowed to randomly rotate in the z -axis at an angle, $\alpha \sim U(-5^\circ, +5^\circ)$.

Out-plane augmentation essentially represents a scenario where the difference in the segmented ROIs happens due to the difference in the choices of the first and/or last slice in the cranio-caudal direction. In this case, we allowed a maximum shift of one slice on either side related to the choice of the prostate ROI boundary (See Figure 3).

In and out-plane augmentation combines the in- and out-plane augmentations to generate custom contour variations representing real-world scenarios.

Furthermore, for in-plane augmentation, we considered two possible biases to model intra- and inter-observer variability in contouring: (a) random bias, where the contour associated with each slice can undergo random transformations independently across the axial dimensions, i.e., for each patient, the height and/or width associated with a contour can independently increase/decrease per slice; (b) systematic bias essentially behaves at random but restricts the direction of the variability to remain the same for all the slices associated with a patient. (i.e., the height and width associated with a contour can either increase or decrease for all the slices). Systematic bias mimics the behavior of radiologists in a clinical scenario where some are systematically more “abundant” in their segmentation while others are more “restrictive”.

In summary, we considered five simulated scenarios for each MRI sequence: (1) in-plane augmentation with random bias; (2) in-plane augmentation with systematic bias; (3) out-plane augmentation; (4) in and out-plane augmentation with random bias; (5) in and out-plane augmentation with systematic bias.

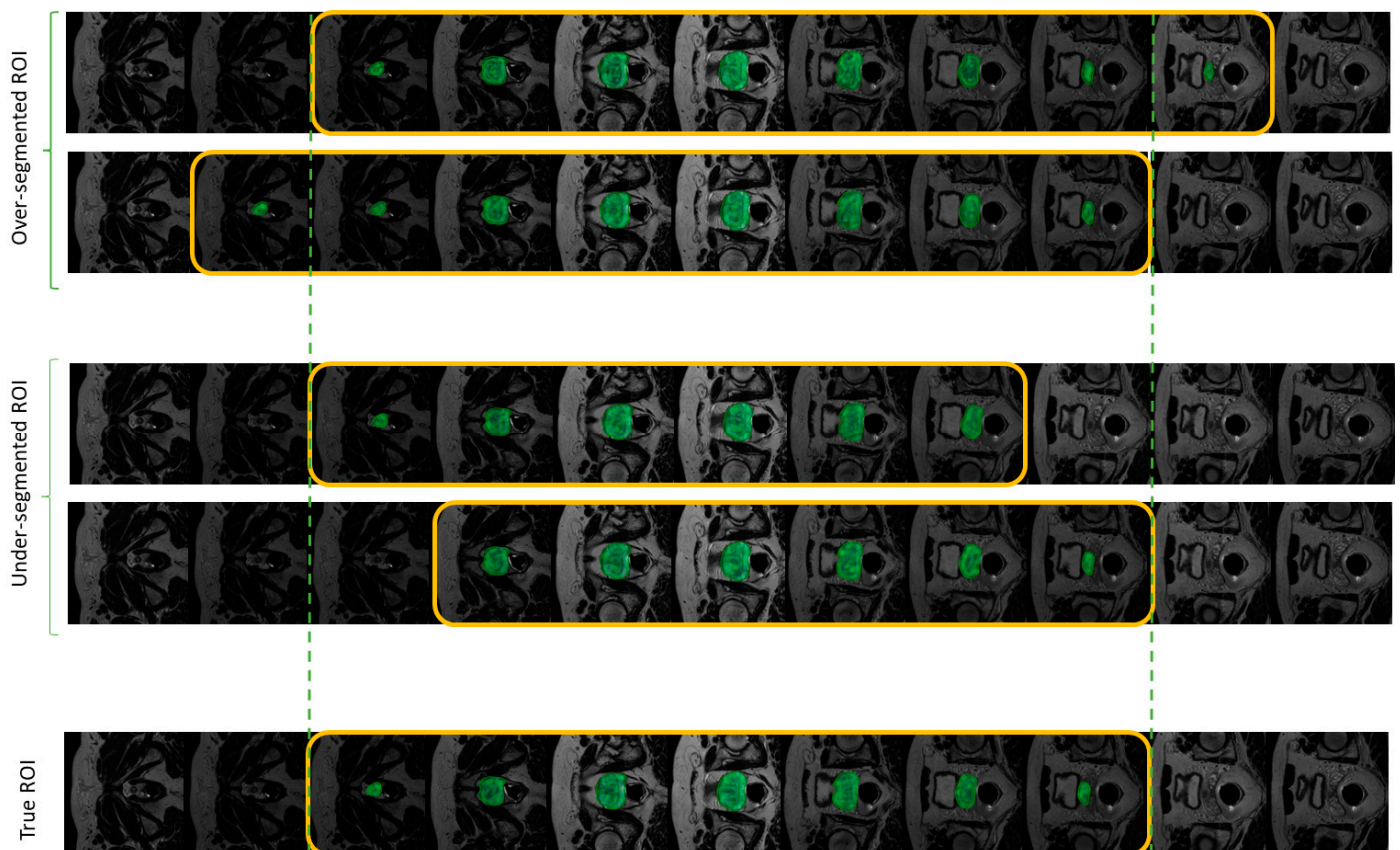


Figure 3. Illustration of various out-plane augmentation scenarios with respect to the true ROI. This augmentation type simulates the variability in the choice of the ROI boundary slice in the craniocaudal direction. The yellow box highlights the slices encompassing the ROI; The vertical dotted green lines indicate the original prostate boundary slices.

2.3. Image Processing Pipeline

In this section, we will summarize some of the preprocessing tasks we have adopted before feeding the image-segmentation pair to the feature extraction pipeline of Pyradiomics [51].

To standardize the voxel size across all the image acquisitions, we resampled the image dimensions to have a common isotropic voxel size of $2 \times 2 \times 2 \text{ mm}^3$. It is important to emphasize that we resampled the in-plane dimensions using third-order B-Spline (cubic) while we resampled the out-plane dimension using nearest neighbor interpolation. We adopted such a strategy to avoid noisy artifacts when upsampling low-resolution images. Subsequently, we used nearest neighbor interpolation to resample all the binary segmentations.

The intensity values in T2w and SUB sequences are relative and are not directly comparable across patients. To this account, we normalized the intensity values using the mean and standard deviation computed on each patient's three-dimensional ROI (i.e., whole prostate). We adopted a similar strategy for ADC. However, since ADC intensities have a global meaning, we computed the mean and standard deviation (σ) across all the patients in the internal dataset rather than normalizing them individually. We then clipped the normalized images at 3σ to further reduce the impact of noises. Finally, we shifted the image mean to a value of 300 with a standard deviation of 100. Assuming a normal distribution, such scaling and shifting ensure that most values lie within the range of 0 to 600, minimizing the influence of negative values on the calculation of radiomic features, which is preferred [26,52,53].

2.4. Radiomics Feature Extraction Pipeline

We used the default settings of Pyradiomics configuration parameters for feature extraction. A notable difference is in the normalization strategy described in the image processing pipeline. The bin-width parameter was also set to 5, such that the number of bins after discretization lies within the range of 30 to 130 (i.e., for the range 0–600, bin-count = $600/5 = 120$), which is shown as having good performance and reproducibility in the literature [54]. Moreover, we believe that a smaller bin width will capture fine-grained information within the whole prostate volume, especially since evidence suggests that only 50% of the patients enrolled in active surveillance have MR visible lesions [17–19].

One thousand two hundred twenty-four features were extracted from the 3D prostate ROI, pertaining to two main feature families and 17 unfiltered/filtered strategies. Table S1 of the Supplementary Materials reports the complete list of all the features considered. For details on their definition, refer to Pyradiomics documentation [51].

The main feature families constitute:

1. First-order statistics (FO, $n = 18$) providing information about the histogram of the grey values inside the prostate ROI; and
2. Texture features, providing information about the spatial distribution of grey values. We used the following textural matrices to compute the textural features: Gray Level Co-occurrence Matrix (GLCM, $n = 22$ features); Gray Level Run Length Matrix (GLRLM, $n = 16$ features); Gray Level Size Zone Matrix (GLSZM, $n = 16$ features).

We utilized all the standard filtering techniques offered by the Pyradiomics package, including LoG (Laplacian of Gaussian filter with kernel sizes, $\sigma = 2, 3, 4, 5$ mm), wavelet (eight decompositions per level based on either applying high (H) or low (L) pass filter along each of the three dimensions—HHH, HHL, HLH, HLL, LHH, LHL, LLH, LLL), squared, square root, logarithm, and exponential filters as part of feature extraction pipeline.

2.5. Stability Analysis

We assessed the stability of radiomic features using the intraclass correlation coefficient form—ICC(1,1) [35] (i.e., model = one-way random effects, type = single rater, and definition = absolute agreement). The model was chosen as one-way random effects since each patient is subjected to random segmentations generated by the augmentation model representing a randomly chosen sample of possible annotators (or raters). The measurement from each rater (i.e., each simulated segmentation) will be the basis of the actual measurement (i.e., the extracted feature); hence the ICC type = single rater. The definition = absolute agreement since we expect the computed feature to remain the same for the same subject across the different annotators. We calculated the ICC(1,1) using the Python library Pingouin (v 0.3.12) [55].

ICC estimate ranges between 0 and 1, with values closer to 1 showing the highest stability. Conventionally to identify stable radiomic features, the ICC estimate is thresholded [35]. In this study, we followed a similar strategy where we categorized a radiomics feature as stable if the lower bound on the 95% confidence interval of the ICC estimate was above 0.90.

We used an external dataset to assess the robustness of stable features identified in the internal population by considering the overlap between stable features across the two datasets. We labeled an overlapped feature as robust if the threshold criterium is satisfied in both datasets, i.e., if the minimum of the lower bound of the ICC estimate in the internal and external datasets is above 0.90. Figure 4 illustrates the overall workflow followed in this study. The Python-based implementation is provided as open-source and is available at https://github.com/sithin-int/stability_study.git (accessed on 19 July 2023) to promote further investigation and reproducibility.

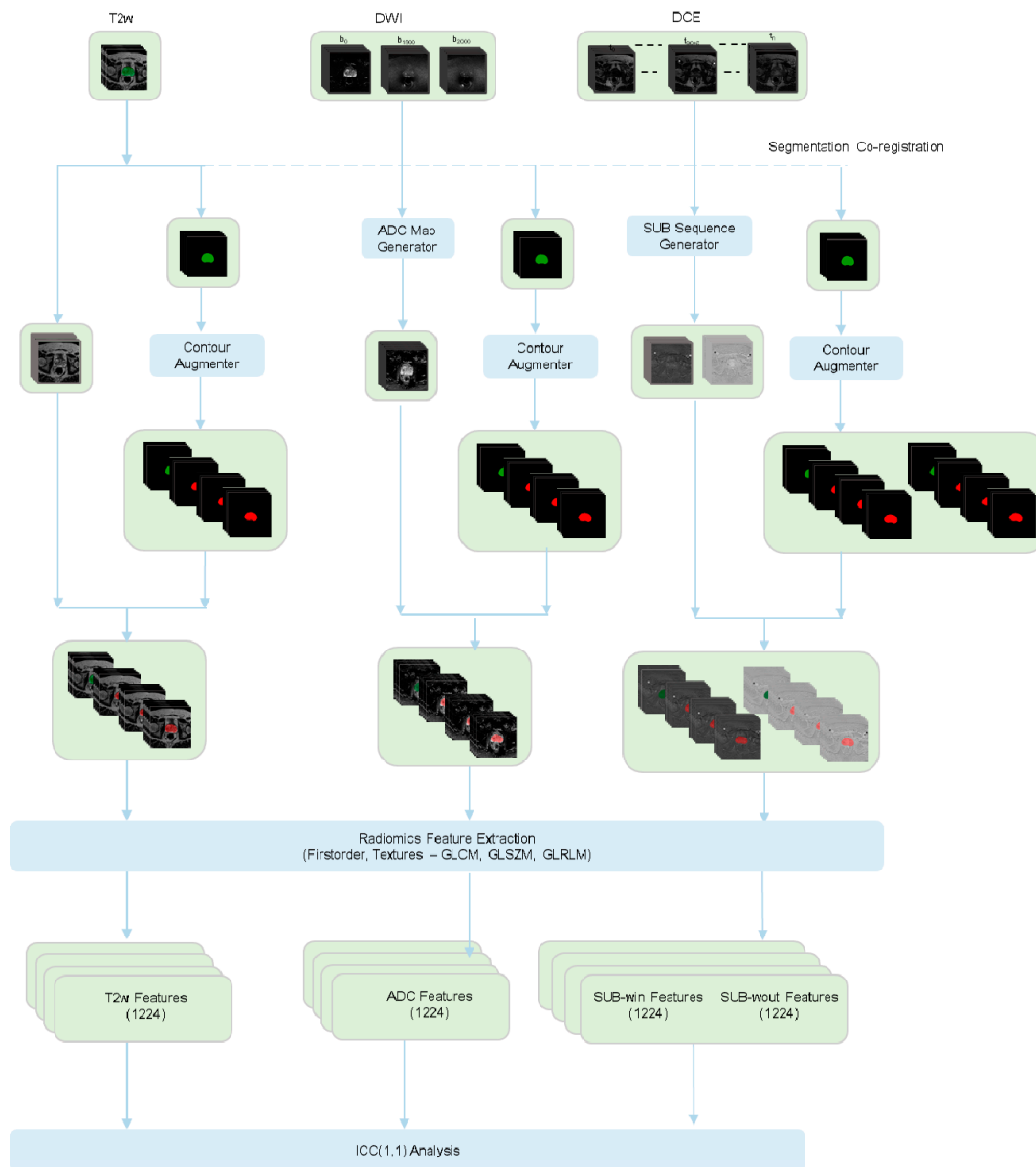


Figure 4. Schematic representation of the workflow involved in the stability study on prostate radiomic features extracted from T2w, DWI, and DCE sequences to variations in segmentation. Manual prostate annotation provided for the T2w sequence was co-registered with the other sequences. The segmentations were then augmented to generate 15 synthetic contours (in the figure as an example, 3 synthetic contours are generated + the original segmentation). A total of 1224 radiomic features were extracted from each of the image-mask pairs. The stability of the features was analyzed using ICC (1,1).

3. Results

In this study, we investigated the impact of variations in segmentation on the stability of radiomic features using an *in silico* contour variability simulator covering three augmentation scenarios—in-plane vs. out-plane vs. in and out-plane; and two segmentation biases—random vs. systematic. Table 2 summarizes the distribution of pairwise dice scores between the ground truth (manual segmentation by the experienced radiologist) and generated contour across all the augmentation-bias configurations. Table 3 presents a simplified summary of the percentage of stable and robust features across all these configurations. Among them, in and out-plane systematic variations significantly impacted

radiomic features’ stability, while out-plane variations seem to affect the least. We observed that the variability margin also depends on the sequence and the filtering strategy.

Table 2. The mean and standard deviation of the dice distribution for each sequence subjected to five different augmentation scenarios.

(a) Internal											
aug config	T2w		ADC		SUB _{win}		SUB _{wout}				
	mean	std	mean	std	mean	std	mean	std			
InP-R	0.95	0.01	0.95	0.01	0.95	0.01	0.95	0.01	0.95	0.01	
InP-S	0.95	0.02	0.95	0.02	0.95	0.02	0.95	0.02	0.95	0.02	
OutP	0.99	0.01	0.99	0.01	0.99	0.01	0.99	0.01	0.99	0.01	
In&OutP-R	0.95	0.01	0.95	0.01	0.95	0.02	0.95	0.02	0.95	0.01	
In&OutP-S	0.94	0.03	0.95	0.03	0.94	0.02	0.94	0.02	0.94	0.03	
(b) External											
aug config	T2w		ADC		SUB						
	mean	std	mean	std	mean	std					
InP-R	0.95	0.01	0.96	0.01	0.95	0.02					
InP-S	0.95	0.03	0.95	0.02	0.95	0.03					
OutP	0.99	0.01	0.99	0.01	0.99	0.01					
In&OutP-R	0.94	0.02	0.95	0.01	0.94	0.02					
In&OutP-S	0.94	0.03	0.95	0.03	0.94	0.03					

aug config—augmentation configuration; InP-R/S—InPlane Random/Systematic; OutP—OutPlane; In&OutP-R/S—In and Out Plane Random/Systematic.

The sheer amount of output data generated by our analysis makes it challenging to discuss each configuration in detail. To simplify, the remainder of the paper will focus exclusively on the most clinically relevant configuration, i.e., in and out plane augmentation with systematic bias. For all the other scenarios, we recommend referring to the supplementary materials. Another overhead may be attributed to the 16 distinct filtering strategies investigated in our study. To streamline our analysis, we only considered filter(s) that showed stability for a feature in the internal population, referred to as best-filter(s), to be compared with the external population for robustness evaluation. The terms “stability” and “robustness” used in this section need to be carefully interpreted. Stable features refer to radiomic features that are stable to variations in segmentation exclusively based on their behavior on the internal population. Robust features, on the other hand, are the overlapped stable features in both the internal and external populations.

Among unfiltered/original radiomic features subjected to in and out-plane systematic variations, T2w (stability = 69%) and SUBwin (65%) sequences showed high stability, followed by SUBwout (53%) and ADC (47%). On the contrary, during robustness evaluation, the fraction of stable features dropped by a significant margin for T2w (drop margin~30%) and SUBwin (~20) while it remained within the range of 10% for SUBwout, and ADC. Consequently, the robustness of features proceeds in the order of SUBwin (robust = 46%), SUBwout (43%), ADC (36%), and T2w (36%). Although T2w sequence features exhibited high stability, they showed the least robustness among all the sequences.

Filtering, on the other hand, improved radiomic features’ stability considerably compared to the unfiltered counterpart. All the filtered sequences had a mean improvement margin of 38%. ADC and T2w features showed a stability of 97%, followed by SUBwin (96%) and SUBwout (94%). The robustness assessment also indicated a high overlap among stable features between the internal and external datasets. The T2w sequence showed the least robustness yet had almost 81% of robust filtered features. This is an improvement of nearly 50% compared to its behavior in the unfiltered configuration. ADC (robustness = 94%) exhibited a similar trend with almost 60% of improvement margin. For both SUBwin and SUBwout sequences, 93% of all the stable features were robust.

Table 3. A summary of the fraction of stable and robust features associated with the mpMRI sequences considered in the study. A feature is considered as stable if the lower bound of the 95% CI of the ICC estimate > 0.90. A stable feature will be categorized as robust if the feature remains stable in both the internal and external dataset.

(a) T2w																				
aug config	firstorder				glcm				glrlm				glszm				Overall			
	O		BF		O		BF		O		BF		O		BF		O		BF	
	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R
InP-R	0.72	0.72	1	1	0.95	0.86	1	1	0.94	0.75	1	1	0.81	0.62	1	0.88	0.86	0.75	1	0.97
InP-S	0.44	0.22	1	0.94	0.73	0.41	1	0.86	0.94	0.44	1	0.62	0.75	0.44	1	0.81	0.71	0.38	1	0.82
OutP	0.83	0.83	1	1	1	1	1	1	1	1	1	1	0.94	0.81	1	1	0.94	0.92	1	1
In&OutP-R	0.72	0.61	1	1	0.95	0.82	1	1	0.94	0.56	1	1	0.81	0.56	1	0.88	0.86	0.65	1	0.97
In&OutP-S	0.44	0.22	1	0.94	0.73	0.41	1	0.86	0.94	0.44	1	0.62	0.69	0.38	0.88	0.75	0.69	0.36	0.97	0.81
(b) ADC																				
aug config	firstorder				glcm				glrlm				glszm				Overall			
	O		BF		O		BF		O		BF		O		BF		O		BF	
	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R
InP-R	0.94	0.89	1	1	1	0.91	1	1	1	0.94	1	1	0.81	0.69	0.94	0.94	0.94	0.86	0.99	0.99
InP-S	0.5	0.5	1	1	0.55	0.41	1	1	0.69	0.5	1	1	0.44	0.31	0.88	0.88	0.54	0.43	0.97	0.97
OutP	1	0.89	1	1	1	0.95	1	1	0.94	0.94	1	1	0.81	0.81	1	1	0.94	0.9	1	1
In&OutP-R	0.94	0.89	1	0.94	1	0.91	1	1	0.94	0.88	1	1	0.75	0.56	0.94	0.94	0.92	0.82	0.99	0.97
In&OutP-S	0.56	0.56	1	0.94	0.5	0.36	1	1	0.5	0.31	1	1	0.31	0.19	0.88	0.81	0.47	0.36	0.97	0.94
(c) SUB_{win}																				
aug config	firstorder				glcm				glrlm				glszm				Overall			
	O		BF		O		BF		O		BF		O		BF		O		BF	
	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R
InP-R	0.72	0.67	1	1	0.91	0.91	1	1	0.81	0.62	1	1	0.81	0.56	1	0.88	0.82	0.71	1	0.97
InP-S	0.5	0.33	1	1	0.82	0.5	1	1	0.56	0.56	0.94	0.81	0.75	0.56	0.88	0.88	0.67	0.49	0.96	0.93
OutP	0.83	0.78	1	1	1	1	1	1	1	1	1	1	1	0.88	1	1	0.96	0.92	1	1
In&OutP-R	0.72	0.61	1	1	0.91	0.82	1	1	0.81	0.62	1	0.88	0.81	0.5	1	0.88	0.82	0.65	1	0.94
In&OutP-S	0.5	0.33	1	1	0.82	0.5	1	1	0.56	0.56	0.94	0.81	0.69	0.44	0.88	0.88	0.65	0.46	0.96	0.93
(d) SUB_{wout}																				

Table 3. Cont.

aug config	firstorder				glcm				glrlm				glszm				Overall			
	O		BF		O		BF		O		BF		O		BF		O		BF	
	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R	S	R
InP-R	0.78	0.67	1	1	0.95	0.91	1	1	0.94	0.62	1	1	0.88	0.62	0.94	0.88	0.89	0.72	0.99	0.97
InP-S	0.33	0.33	1	1	0.68	0.41	1	1	0.56	0.56	1	0.81	0.56	0.5	0.94	0.88	0.54	0.44	0.99	0.93
OutP	0.89	0.78	1	1	1	1	1	1	0.94	0.94	1	1	1	0.88	1	1	0.96	0.9	1	1
In&OutP-R	0.72	0.61	1	1	0.91	0.82	1	1	0.81	0.62	1	0.88	0.81	0.56	0.88	0.88	0.82	0.67	0.97	0.94
In&OutP-S	0.33	0.33	1	1	0.64	0.41	1	1	0.56	0.56	0.88	0.81	0.56	0.44	0.88	0.88	0.53	0.43	0.94	0.93

aug config—augmentation configuration; InP-R/S—InPlane Random/Systematic; OutP—OutPlane; In&OutP-R/S—In and Out Plane Random/Systematic; O—Original/Raw/Unfiltered; BF—Best Filtered; S—Stable; R—Robust.

In summary, ADC-filtered features demonstrated the highest degree of robustness, followed by SUB and T2w. Figures 5–8, in their panels (a), highlight the impact of the unfiltered v/s filtering strategy on the stability and, consequently, on the robustness of radiomic features as a heatmap. Figures 5–8, in their panels (b), portray the overlap among the ICC estimates between the internal and external population for unfiltered and best-filtered feature configurations.

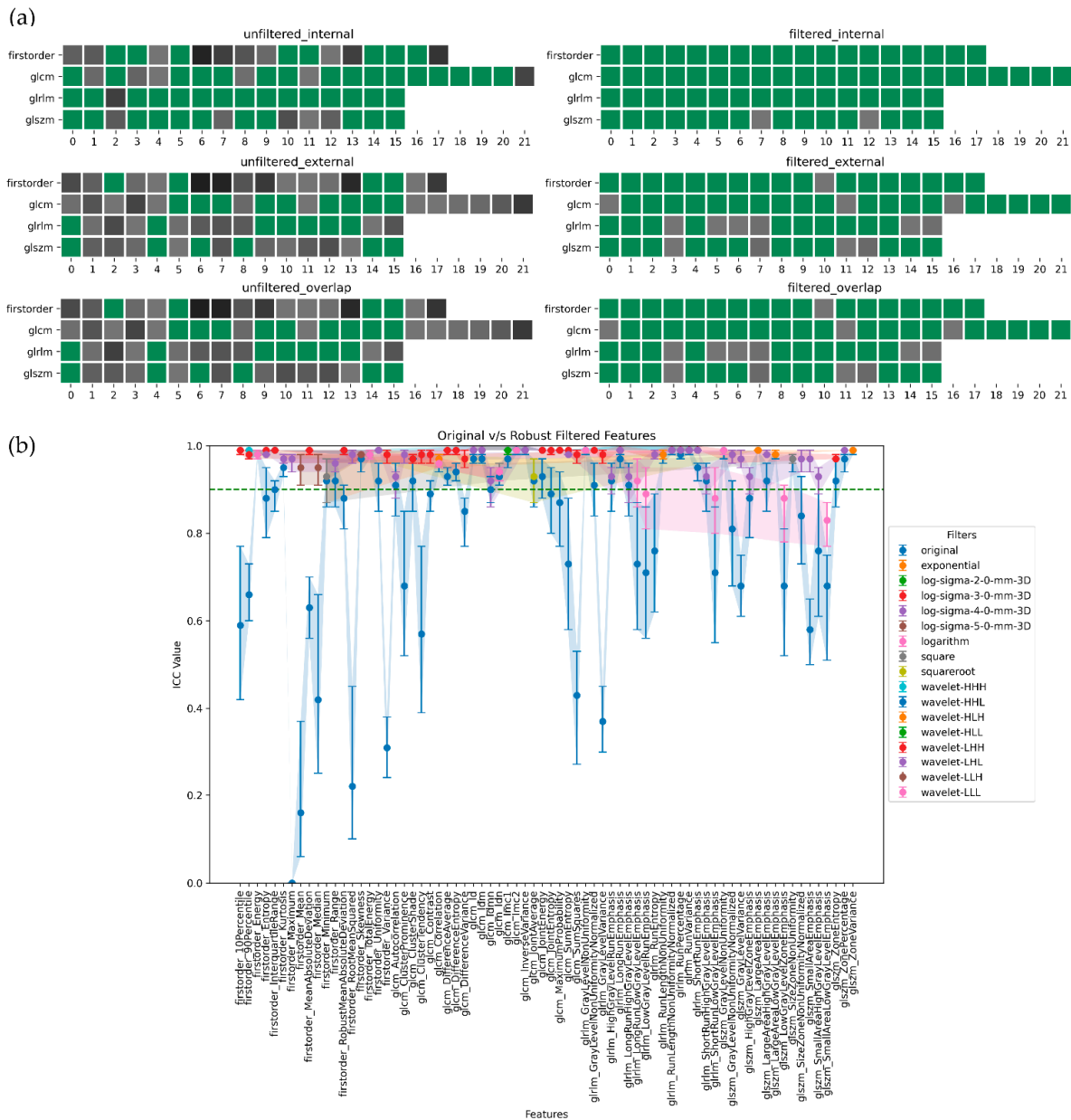


Figure 5. Stability and robustness of filtered/unfiltered T2w radiomics features subjected to in and out-plane-systematic segmentation variations. A feature is considered stable if the lower bound of the 95% CI of the ICC estimate > 0.90. (a) Stable v/s unstable feature heatmap—grey cells indicate unstable features, with darker shades of grey indicating lower ICC bounds. All the green cells represent stable features. (b) ICC plot portrays the overlap computed as the minimum stability value of a feature in the internal and external dataset grouped by both unfiltered and best-filtered configurations. For simplicity, we are only displaying the best filter(s) that yielded the highest ICC lower bound after overlap. The dotted green line indicates the stability threshold.

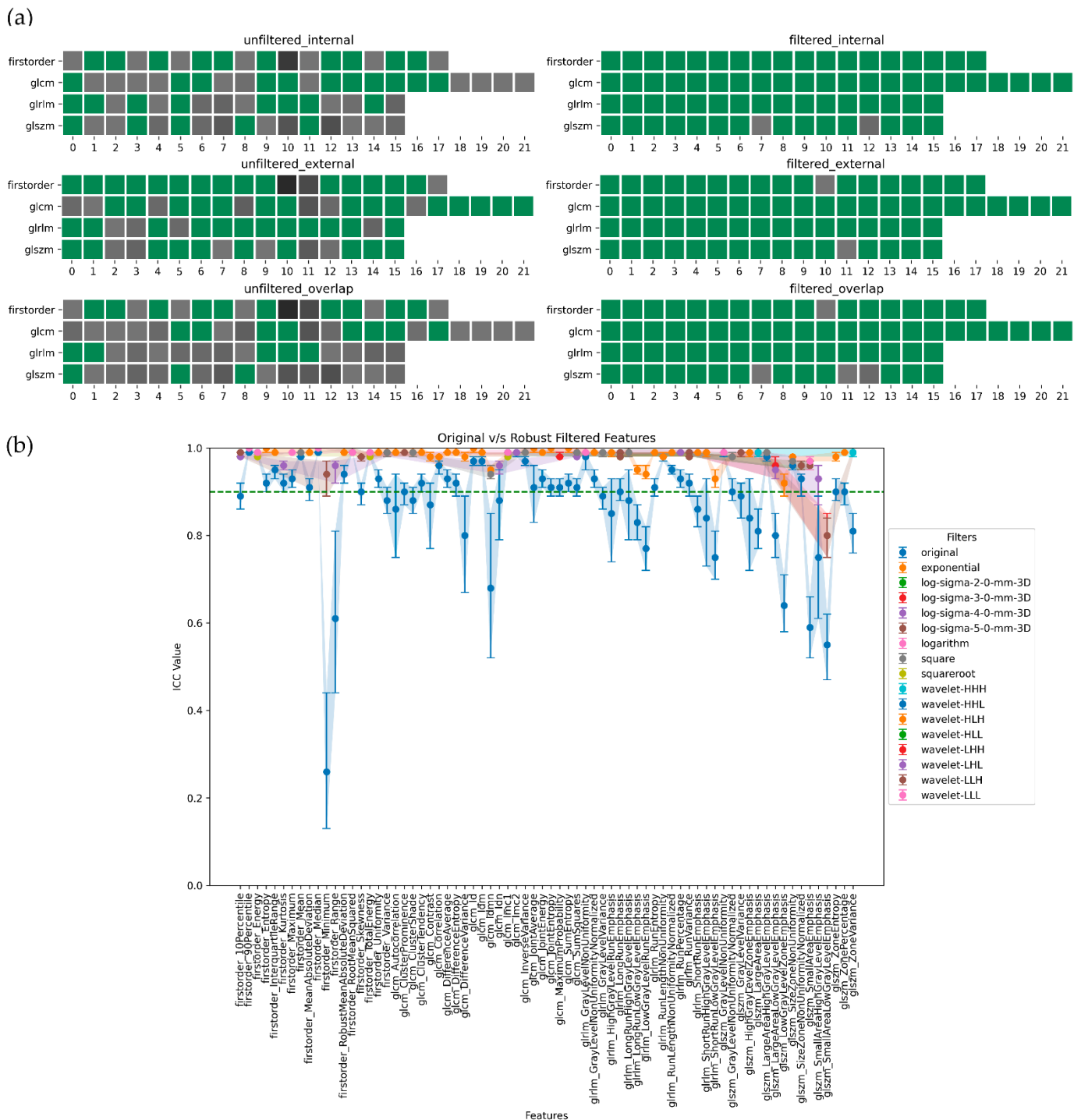


Figure 6. Stability and robustness of filtered/unfiltered ADC radiomics features subjected to in and out-plane-systematic segmentation variations. A feature is considered stable if the lower bound of the 95% CI of the ICC estimate > 0.90. **(a)** Stable v/s unstable feature heatmap—grey cells indicate unstable features, with darker shades of grey indicating lower ICC bounds. All the green cells represent stable features. **(b)** ICC plot portrays the overlap computed as the minimum stability value of a feature in the internal and external dataset grouped by both unfiltered and best-filtered configurations. For simplicity, we are only displaying the best filter(s) that yielded the highest ICC lower bound after overlap. The dotted green line indicates the stability threshold.

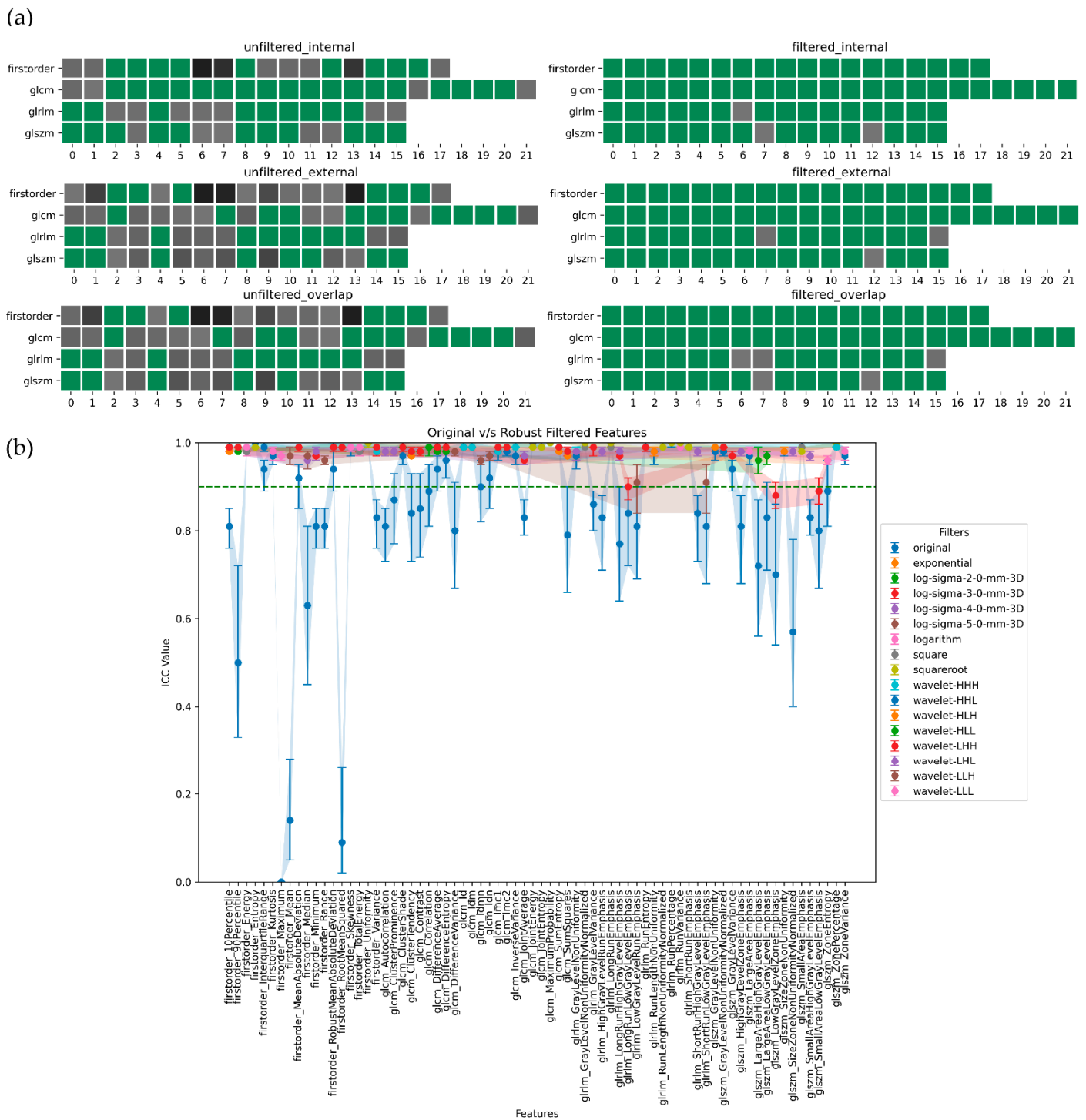


Figure 7. Stability and robustness of filtered/unfiltered SUBwin radiomics features subjected to in and out-plane-systematic segmentation variations. A feature is considered stable if the lower bound of the 95% CI of the ICC estimate > 0.90. **(a)** Stable v/s unstable feature heatmap—grey cells indicate unstable features, with darker shades of grey indicating lower ICC bounds. All the green cells represent stable features. **(b)** ICC plot portrays the overlap computed as the minimum stability value of a feature in the internal and external dataset grouped by both unfiltered and best-filtered configurations. For simplicity, we are only displaying the best filter(s) that yielded the highest ICC lower bound after overlap. The dotted green line indicates the stability threshold.

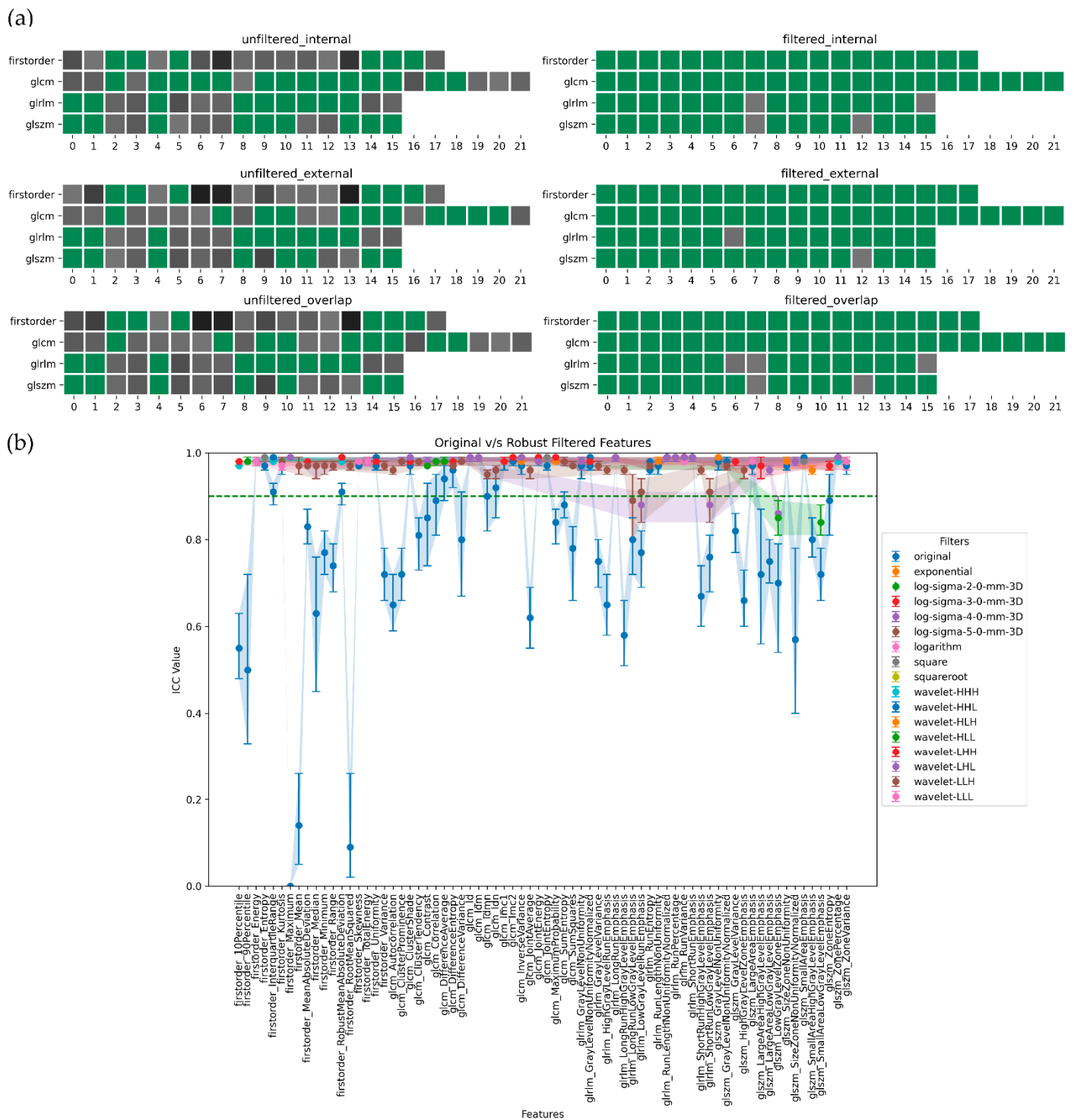


Figure 8. Stability and robustness of filtered/unfiltered SUBwout radiomics features subjected to in and out-plane-systematic segmentation variations. A feature is considered stable if the lower bound of the 95% CI of the ICC estimate > 0.90. (a) Stable v/s unstable feature heatmap—grey cells indicate unstable features, with darker shades of grey indicating lower ICC bounds. All the green cells represent stable features. (b) ICC plot portrays the overlap computed as the minimum stability value of a feature in the internal and external dataset grouped by both unfiltered and best-filtered configurations. For simplicity, we are only displaying the best filter(s) that yielded the highest ICC lower bound after overlap. The dotted green line indicates the stability threshold.

4. Discussion

Reliability (or stability) is essential to use quantitative image-based features as potential biomarkers for clinical applications. While numerous factors could influence radiomics feature repeatability, this study focused exclusively on the stability of features to variations in segmentation. This was accomplished by designing an *in silico* contour generator that simulates variations commonly observed among manual annotators. This study investigated five distinct configurations covering three augmentation scenarios—in-plane, out-plane, and in and out-plane—and two segmentation biases—random and systematic. The generator's design was inspired by the data augmentation paradigm in DL and utilized bounded affine transformations.

In the premise of prostate mpMRI analysis, only a few studies have investigated the stability of features to variations in segmentation. For example, Xu, Lili, et al. [33] obtained lesion annotations from two radiologists to assess the feature stability to develop a robust radiomics model for predicting extraprostatic extension. Another study undertaken by Sushentsev, Nikita, et al. [10] used morphological perturbations such as opening and/or closing on the lesion ROI to simulate contour variations without involving manual annotators. The robust features they identified were then used to predict PCa progression. It is important to note that, in both studies, the lesions were used as the ROIs. This may not be ideal in active surveillance, especially since only half the patients will likely show MR-visible lesions [17–19]. To this account, we recommend using the whole prostate gland to analyze MRI images from very low-risk PCa patients on active surveillance.

Segmentation of the prostate gland is challenging due to the lack of a clear visual boundary and significant variations in its size and shape among patients. These differences lead to intra-/inter-observer segmentation variations among human annotators. Our results highlight that these variations can notably impact stability, particularly among unfiltered radiomic features. However, it can be mitigated by incorporating filtering strategies. While the Image Biomarker Standardisation Initiative (ISBI) does not address pre-filtering strategies, filters such as Wavelet and LoG have been shown to yield highly predictive signatures [9,56]. Our results suggest that the use of Wavelet and LoG filters could also lead to considerable improvements in terms of stability (see Figure 9).

Radiomics mpMRI studies rarely include DCE sequences but rely primarily on bi-parametric (bp) or uni-parametric MR sequences. This may be attributed to two major reasons: (i) despite the loss in diagnostic information [20], bpMRI is recommended for biopsy naïve PCa patients and is more suitable for large studies [57] as it eliminates the risk of adverse effects due to contrast agent and speeds up image acquisition; (ii) the processing pipeline associated with DCE is complicated as we need to consider the temporal domain. Conventionally pharmacokinetic maps are extracted from DCE images and are used for radiomics analysis [58].

PCa patients on active surveillance do not fall into the category of biopsy naïve patients; hence, we chose to include DCE sequence in our analysis. Instead of computing pharmacokinetic maps, we derived subtraction maps to encapsulate the contrast agent's wash-in and wash-out phases. Schwier, Michael, et al. [26] conducted a test-retest feature repeatability study on prostate mpMRI sequences, examining SUB maps. They reported that none of the radiomic features extracted from the whole-prostate gland was stable. Our findings suggest the contrary, where SUB sequence features showed high robustness among unfiltered and filtered configurations. However, it is essential to note that these results are not directly comparable. Their study focused on the stability of SUB features in a test-retest configuration, while we investigated feature stability subjected to variations in segmentation. Nonetheless, we would like to emphasize this point to promote further investigation of SUB maps.

A conventional approach to evaluating the stability of radiomic features involves using a test-retest paradigm [53,59,60]. It is often the case that test-retest acquisitions may be difficult to obtain or not readily available. Zwanenburg et al. [61] proposed a solution to this approach by leveraging the data augmentation paradigm in DL. They synthesized

different image acquisitions by considering linear/non-linear perturbations of the original image. They used a combination of transformations such as translation, rotation, volume growth/shrinkage, super-voxel-based contour randomization, and noise addition. The *in silico* contour segmentation tool proposed in this study also follows a similar pipeline. Yet, instead of augmenting the images, we aim to induce variations in the segmentation mask. Moreover, we restricted augmentations to replicate real-world variations commonly seen among manual annotators. The potential application of this tool lies in promoting future stability studies to segmentation as it can be easily integrated into any radiomics pipeline.

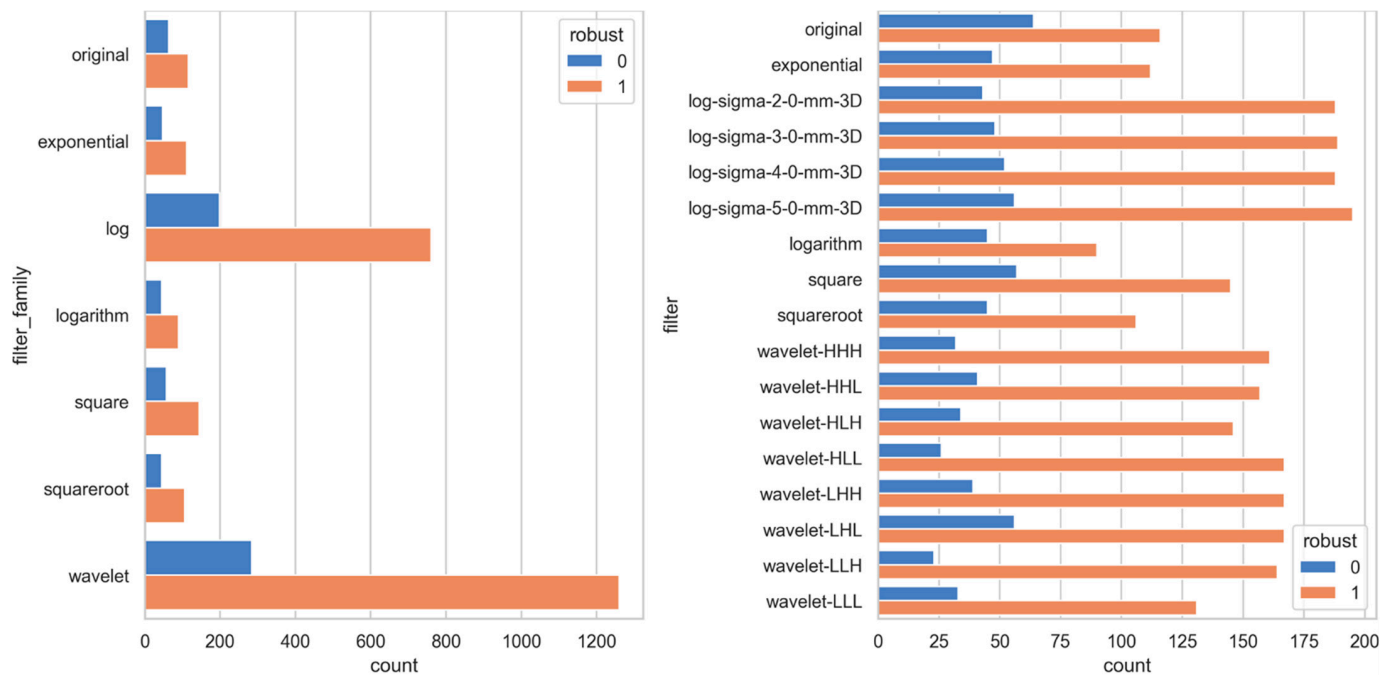


Figure 9. The histogram plot presents a summary of the frequency at which a robust feature is associated with a particular filter/filter-family subjected to systematic-in and out-plane augmentation across all image sequences.

The limitations of this study highlight possible future research directions. We assumed we could simulate the segmentation variations resulting from inter/intra-observer variability by subjecting ground truth annotation to bounded affine transformations. Nevertheless, it is uncertain how much this approximation reflects the variations observed in the real world. Therefore, further investigation is warranted to validate the proposed method by comparing it with the conventional clinical inter/intra-observer variations study involving manual annotators. This increases the overhead of involving multiple radiologists, which was not feasible in this study. Although we considered the overlap in terms of stability between the internal and external datasets, it is essential to emphasize that these populations may not be directly comparable. This is particularly true for SUB maps. For the external population, SUB maps were scanner derived considering the early post-contrast and pre-contrast. On the other hand, SUB maps were derived manually for the internal population by considering the contrast agent’s wash-in and wash-out phases. Yet another limitation may be attributed to the scope of this study, where the analysis is limited to the stability of prostate radiomic features subjected to variations in segmentation. In reality, numerous other sources of variations affect feature stability, such as the heterogeneity of study protocols, scan acquisition parameters, reconstruction settings, and feature extraction pipeline, which need to be considered to improve the overall generalizability of radiomic signatures.

5. Conclusions

This study presents a method to evaluate the stability of radiomic features to variations in segmentation. The technique was then employed to examine the stability of mpMRI radiomic features extracted from the whole prostate gland among PCa patients in active surveillance. Our findings highlight that unfiltered radiomic features are susceptible to variations in segmentation. However, by incorporating pre-filtering strategies, the feature stability improved. We also recommend using external datasets to validate the robustness of stable features identified on the internal dataset.

The contour augmentation method proposed in this study also has the potential to enhance the robustness of PI-RADS [7] determination, i.e., by evaluating PI-RADS using multiple augmented contours, a distribution of scores can be obtained mirroring the uncertainties associated with the single definition of the regions of interest.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jpm13071172/s1>. Table S1: The complete list of radiomics features and filtering strategies considered in the study. Folder B.1 ICC Plots: The folder named “plots” contains all the figures associated with each sequence and augmentation scenario for both internal and external population. Eg. To visualize the ICC plot associated with internal T2w sequence subjected to in-plane-random augmentation; Open plots->t2w->in_plane_random_internal.png. Folder B.2 Overlap Plots: The folder named “overlap_plots” contains all the figures associated with each sequence and augmentation scenario. The overlap here indicates the overlap (or minimum) of ICC estimates between both the internal and external population. Eg. To visualize the overlap plot associated with T2w sequence subjected to in-plane-random augmentation; Open overlap_plots->t2w->in_plane_random.png. Folder B.3 Heatmaps: The folder named “heatmaps” contains all the heatmaps associated with each sequence and augmentation scenario. This is another representation of overlapping plots highlighting robust features. Robust features are highlighted green while unstable features are highlighted gray. Eg. To visualize the heatmap associated with T2w sequence subjected to in-plane-random augmentation; Open heatmaps->t2w->in_plan_random.png. Folder B.4 Robust Filter Histograms: The folder named “histplots” contains all the figures associated with each augmentation scenario. This plot counts the number of times a filter/filter family was found among robust (1) non-robust (0) features. Eg: To visualize the histogram associated with in-plane-random augmentation scenario; Open histplots->in_plane_random.png. Folder Robust Features: All the robust features associated with each augmentation configuration are contained in folder “robust_features”. The filtering technique available in MS excel can guide the user to extract information necessary for a specific sequence.

Author Contributions: Conceptualization, T.R., N.N. and A.M.; Data curation, S.T.S., E.G., C.T. and F.B.; Formal analysis, S.T.S.; Funding acquisition, R.V. and N.N.; Methodology, S.T.S., E.G., C.T., S.S. (Sebastian Sanduleanu) and T.R.; Project administration, E.B.; Resources, C.M., B.A., M.C., S.S. (Silvia Stagni), S.V., B.N.C., E.P., G.P. and R.V.; Software, S.T.S.; Supervision, T.R. and A.M.; Validation, S.T.S.; Writing—original draft, S.T.S.; Writing—review and editing, E.G., C.T., C.M., B.A., M.C., S.S. (Silvia Stagni), S.V., B.N.C., F.B., E.B., S.S. (Sebastian Sanduleanu), E.P., G.P., R.V., T.R. and A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by “5 per 1000” funds (Italian Ministry of Health 2016—financial support for healthcare research), title of the project “Microstructural evaluation of prostate cancer by multiparametric magnetic resonance” to “Nicola Nicolai” and by Fondazione Italo Monzino. The APC by “5 per 1000” funds (Italian Ministry of Health 2016—financial support for healthcare research).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Fondazione IRCCS Istituto Nazionale dei Tumori (INT 113/16, INT 46/07, and INT 95/11).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Raw data were generated at Fondazione IRCCS Istituto Nazionale dei Tumori. Derived data supporting the findings of this study are available from the corresponding authors [T.R. and C.T.] on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Johnson, L.M.; Turkbey, B.; Figg, W.D.; Choyke, P.L. Multi-parametric MRI in Prostate Cancer Management. *Nat. Rev. Clin. Oncol.* **2014**, *11*, 346–353. [[CrossRef](#)]
2. Barrett, T.; Haider, M.A. The Emerging Role of MRI in Prostate Cancer Active Surveillance and Ongoing Challenges. *AJR Am. J. Roentgenol.* **2017**, *208*, 131–139. [[CrossRef](#)]
3. Thurtle, D.; Barrett, T.; Thankappan-Nair, V.; Koo, B.; Warren, A.; Kastner, C.; Saeb-Parsy, K.; Kimberley-Duffell, J.; Gnanaprasam, V.J. Progression and Treatment Rates Using an Active Surveillance Protocol Incorporating Image-Guided Baseline Biopsies and Multi-parametric Magnetic Resonance Imaging Monitoring for Men with Favourable-Risk Prostate Cancer. *BJU Int.* **2018**, *122*, 59–65. [[CrossRef](#)]
4. Schoots, I.G.; Petrides, N.; Giganti, F.; Bokhorst, L.P.; Rannikko, A.; Klotz, L.; Villers, A.; Hugosson, J.; Moore, C.M. Magnetic Resonance Imaging in Active Surveillance of Prostate Cancer: A Systematic Review. *Eur. Urol.* **2015**, *67*, 627–636. [[CrossRef](#)]
5. Schoots, I.G.; Nieboer, D.; Giganti, F.; Moore, C.M.; Bangma, C.H.; Roobol, M.J. Is Magnetic Resonance Imaging-Targeted Biopsy a Useful Addition to Systematic Confirmatory Biopsy in Men on Active Surveillance for Low-Risk Prostate Cancer? A Systematic Review and Meta-Analysis. *BJU Int.* **2018**, *122*, 946–958. [[CrossRef](#)] [[PubMed](#)]
6. Ghavimi, S.; Abdi, H.; Waterhouse, J.; Savdie, R.; Chang, S.; Harris, A.; Machan, L.; Gleave, M.; So, A.I.; Goldenberg, L.; et al. Natural History of Prostatic Lesions on Serial Multi-parametric Magnetic Resonance Imaging. *Can. Urol. Assoc. J.* **2018**, *12*, 270–275. [[CrossRef](#)] [[PubMed](#)]
7. Turkbey, B.; Rosenkrantz, A.B.; Haider, M.A.; Padhani, A.R.; Villeirs, G.; Macura, K.J.; Tempany, C.M.; Choyke, P.L.; Cornud, F.; Margolis, D.J.; et al. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. *Eur. Urol.* **2019**, *76*, 340–351. [[CrossRef](#)]
8. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; van Stiphout, R.G.P.M.; Granton, P.; Zegers, C.M.L.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis. *Eur. J. Cancer* **2012**, *48*, 441–446. [[CrossRef](#)]
9. Aerts, H.J.W.L.; Velazquez, E.R.; Leijenaar, R.T.H.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding Tumour Phenotype by Non-invasive Imaging Using a Quantitative Radiomics Approach. *Nat. Commun.* **2014**, *5*, 4006. [[CrossRef](#)] [[PubMed](#)]
10. Sushentsev, N.; Rundo, L.; Blyuss, O.; Gnanaprasam, V.J.; Sala, E.; Barrett, T. MRI-Derived Radiomics Model for Baseline Prediction of Prostate Cancer Progression on Active Surveillance. *Sci. Rep.* **2021**, *11*, 12917. [[CrossRef](#)]
11. Sushentsev, N.; Rundo, L.; Blyuss, O.; Nazarenko, T.; Suvorov, A.; Gnanaprasam, V.J.; Sala, E.; Barrett, T. Comparative Performance of MRI-Derived PRECISE Scores and Delta-Radiomics Models for the Prediction of Prostate Cancer Progression in Patients on Active Surveillance. *Eur. Radiol.* **2022**, *32*, 680–689. [[CrossRef](#)] [[PubMed](#)]
12. Fave, X.; Zhang, L.; Yang, J.; Mackin, D.; Balter, P.; Gomez, D.; Followill, D.; Jones, A.K.; Stingo, F.; Liao, Z.; et al. Delta-Radiomics Features for the Prediction of Patient Outcomes in Non-Small Cell Lung Cancer. *Sci. Rep.* **2017**, *7*, 588. [[CrossRef](#)] [[PubMed](#)]
13. Moore, C.M.; Giganti, F.; Albertsen, P.; Allen, C.; Bangma, C.; Briganti, A.; Carroll, P.; Haider, M.; Kasivisvanathan, V.; Kirkham, A.; et al. Reporting Magnetic Resonance Imaging in Men on Active Surveillance for Prostate Cancer: The PRECISE Recommendations—A Report of a European School of Oncology Task Force. *Eur. Urol.* **2017**, *71*, 648–655. [[CrossRef](#)] [[PubMed](#)]
14. Algohary, A.; Viswanath, S.; Shiradkar, R.; Ghose, S.; Pahwa, S.; Moses, D.; Jambor, I.; Shnier, R.; Böhm, M.; Haynes, A.-M.; et al. Radiomic Features on MRI Enable Risk Categorization of Prostate Cancer Patients on Active Surveillance: Preliminary Findings. *J. Magn. Reson. Imaging* **2018**, *48*, 818–828. [[CrossRef](#)] [[PubMed](#)]
15. Sushentsev, N.; Caglic, I.; Rundo, L.; Kozlov, V.; Sala, E.; Gnanaprasam, V.J.; Barrett, T. Serial Changes in Tumour Measurements and Apparent Diffusion Coefficients in Prostate Cancer Patients on Active Surveillance with and without Histopathological Progression. *Br. J. Radiol.* **2022**, *95*, 20210842. [[CrossRef](#)] [[PubMed](#)]
16. Nayan, M.; Salari, K.; Bozzo, A.; Ganglberger, W.; Lu, G.; Carvalho, F.; Gusev, A.; Schneider, A.; Westover, B.M.; Feldman, A.S. A Machine Learning Approach to Predict Progression on Active Surveillance for Prostate Cancer. *Urol. Oncol.* **2022**, *40*, 161.e1–161.e7. [[CrossRef](#)]
17. Radtke, J.P.; Kuru, T.H.; Boxler, S.; Alt, C.D.; Popeneciu, I.V.; Huettenbrink, C.; Klein, T.; Steinemann, S.; Bergstraesser, C.; Roethke, M.; et al. Comparative Analysis of Transperineal Template Saturation Prostate Biopsy versus Magnetic Resonance Imaging Targeted Biopsy with Magnetic Resonance Imaging-Ultrasound Fusion Guidance. *J. Urol.* **2015**, *193*, 87–94. [[CrossRef](#)]
18. Filson, C.P.; Natarajan, S.; Margolis, D.J.A.; Huang, J.; Lieu, P.; Dorey, F.J.; Reiter, R.E.; Marks, L.S. Prostate Cancer Detection with Magnetic Resonance-Ultrasound Fusion Biopsy: The Role of Systematic and Targeted Biopsies. *Cancer* **2016**, *122*, 884–892. [[CrossRef](#)]

19. Johnson, D.C.; Raman, S.S.; Mirak, S.A.; Kwan, L.; Bajgirani, A.M.; Hsu, W.; Maehara, C.K.; Ahuja, P.; Faiena, I.; Pooli, A.; et al. Detection of Individual Prostate Cancer Foci via Multi-parametric Magnetic Resonance Imaging. *Eur. Urol.* **2019**, *75*, 712–720. [[CrossRef](#)]
20. de Rooij, M.; Israël, B.; Bomers, J.G.R.; Schoots, I.G.; Barentsz, J.O. Can Biparametric Prostate Magnetic Resonance Imaging Fulfill Its PROMIS? *Eur. Urol.* **2020**, *78*, 512–514. [[CrossRef](#)]
21. Peerlings, J.; Woodruff, H.C.; Winfield, J.M.; Ibrahim, A.; Van Beers, B.E.; Heerschap, A.; Jackson, A.; Wildberger, J.E.; Mottaghy, F.M.; DeSouza, N.M.; et al. Stability of Radiomics Features in Apparent Diffusion Coefficient Maps from a Multi-Centre Test-Retest Trial. *Sci. Rep.* **2019**, *9*, 4800. [[CrossRef](#)] [[PubMed](#)]
22. Rai, R.; Holloway, L.C.; Brink, C.; Field, M.; Christiansen, R.L.; Sun, Y.; Barton, M.B.; Liney, G.P. Multicenter Evaluation of MRI-Based Radiomic Features: A Phantom Study. *Med. Phys.* **2020**, *47*, 3054–3063. [[CrossRef](#)] [[PubMed](#)]
23. Bologna, M.; Corino, V.; Mainardi, L. Technical Note: Virtual Phantom Analyses for Preprocessing Evaluation and Detection of a Robust Feature Set for MRI-Radiomics of the Brain. *Med. Phys.* **2019**, *46*, 5116–5123. [[CrossRef](#)] [[PubMed](#)]
24. Kim, H.; Park, C.M.; Lee, M.; Park, S.J.; Song, Y.S.; Lee, J.H.; Hwang, E.J.; Goo, J.M. Impact of Reconstruction Algorithms on CT Radiomic Features of Pulmonary Tumors: Analysis of Intra- and Inter-Reader Variability and Inter-Reconstruction Algorithm Variability. *PLoS ONE* **2016**, *11*, e0164924. [[CrossRef](#)]
25. Choe, J.; Lee, S.M.; Do, K.-H.; Lee, G.; Lee, J.-G.; Lee, S.M.; Seo, J.B. Deep Learning-Based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses. *Radiology* **2019**, *292*, 365–373. [[CrossRef](#)]
26. Schwier, M.; van Griethuysen, J.; Vangel, M.G.; Pieper, S.; Peled, S.; Tempny, C.; Aerts, H.J.W.; Kikinis, R.; Fennessy, F.M.; Fedorov, A. Repeatability of Multi-parametric Prostate MRI Radiomics Features. *Sci. Rep.* **2019**, *9*, 9441. [[CrossRef](#)]
27. Scalco, E.; Belfatto, A.; Mastropietro, A.; Rancati, T.; Avuzzi, B.; Messina, A.; Valdagni, R.; Rizzo, G. T2w-MRI Signal Normalization Affects Radiomics Features Reproducibility. *Med. Phys.* **2020**, *47*, 1680–1691. [[CrossRef](#)]
28. Saha, A.; Harowicz, M.R.; Mazurowski, M.A. Breast Cancer MRI Radiomics: An Overview of Algorithmic Features and Impact of Inter-reader Variability in Annotating Tumors. *Med. Phys.* **2018**, *45*, 3076–3085. [[CrossRef](#)]
29. Haarbuerger, C.; Müller-Franzes, G.; Weninger, L.; Kuhl, C.; Truhn, D.; Merhof, D. Radiomics Feature Reproducibility under Inter-Rater Variability in Segmentations of CT Images. *Sci. Rep.* **2020**, *10*, 12688. [[CrossRef](#)]
30. Baeßler, B.; Weiss, K.; Pinto Dos Santos, D. Robustness and Reproducibility of Radiomics in Magnetic Resonance Imaging: A Phantom Study. *Investig. Radiol.* **2019**, *54*, 221–228. [[CrossRef](#)]
31. van Timmeren, J.E.; Cester, D.; Tanadini-Lang, S.; Alkadhi, H.; Baessler, B. Radiomics in Medical Imaging—“how-to” Guide and Critical Reflection. *Insights Imaging* **2020**, *11*, 91. [[CrossRef](#)] [[PubMed](#)]
32. Galavis, P.E. Reproducibility and Standardization in Radiomics: Are We There Yet? In Proceedings of the XVI Mexican Symposium On Medical Physics, Mexico City, Mexico, 26–30 October 2020; AIP Publishing: Melville, NY, USA, 2021.
33. Xu, L.; Zhang, G.; Zhao, L.; Mao, L.; Li, X.; Yan, W.; Xiao, Y.; Lei, J.; Sun, H.; Jin, Z. Radiomics Based on Multi-parametric Magnetic Resonance Imaging to Predict Extraprostatic Extension of Prostate Cancer. *Front. Oncol.* **2020**, *10*, 940. [[CrossRef](#)] [[PubMed](#)]
34. Zhang, G.-M.-Y.; Han, Y.-Q.; Wei, J.-W.; Qi, Y.-F.; Gu, D.-S.; Lei, J.; Yan, W.-G.; Xiao, Y.; Xue, H.-D.; Feng, F.; et al. Radiomics Based on MRI as a Biomarker to Guide Therapy by Predicting Upgrading of Prostate Cancer from Biopsy to Radical Prostatectomy. *J. Magn. Reson. Imaging* **2020**, *52*, 1239–1248. [[CrossRef](#)]
35. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [[CrossRef](#)] [[PubMed](#)]
36. Chen, T.; Li, M.; Gu, Y.; Zhang, Y.; Yang, S.; Wei, C.; Wu, J.; Li, X.; Zhao, W.; Shen, J. Prostate Cancer Differentiation and Aggressiveness: Assessment with a Radiomic-Based Model vs. PI-RADS V2. *J. Magn. Reson. Imaging* **2019**, *49*, 875–884. [[CrossRef](#)] [[PubMed](#)]
37. Fehr, D.; Veeraraghavan, H.; Wibmer, A.; Gondo, T.; Matsumoto, K.; Vargas, H.A.; Sala, E.; Hricak, H.; Deasy, J.O. Automatic Classification of Prostate Cancer Gleason Scores from Multi-parametric Magnetic Resonance Images. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6265–E6273. [[CrossRef](#)]
38. Khoo, E.L.H.; Schick, K.; Plank, A.W.; Poulsen, M.; Wong, W.W.G.; Middleton, M.; Martin, J.M. Prostate Contouring Variation: Can It Be Fixed? *Int. J. Radiat. Oncol. Biol. Phys.* **2012**, *82*, 1923–1929. [[CrossRef](#)]
39. Balagurunathan, Y.; Gu, Y.; Wang, H.; Kumar, V.; Grove, O.; Hawkins, S.; Kim, J.; Goldgof, D.B.; Hall, L.O.; Gatenby, R.A.; et al. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. *Transl. Oncol.* **2014**, *7*, 72–87. [[CrossRef](#)]
40. Kalpathy-Cramer, J.; Mamomov, A.; Zhao, B.; Lu, L.; Cherezov, D.; Napel, S.; Echegaray, S.; Rubin, D.; McNitt-Gray, M.; Lo, P.; et al. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. *Tomography* **2016**, *2*, 430–437. [[CrossRef](#)]
41. Haarbuerger, C.; Schock, J.; Truhn, D.; Weitz, P.; Mueller-Franzes, G.; Weninger, L.; Merhof, D. Radiomic Feature Stability Analysis Based on Probabilistic Segmentations. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020.
42. Berman, R.M.; Brown, A.M.; Chang, S.D.; Sankineni, S.; Kadakia, M.; Wood, B.J.; Pinto, P.A.; Choyke, P.L.; Turkbey, B. DCE MRI of prostate cancer. *Abdom. Radiol.* **2016**, *41*, 844–853. [[CrossRef](#)]
43. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [[CrossRef](#)]

44. Fedorov, A.; Vangel, M.G.; Tempny, C.M.; Fennessy, F.M. Multi-parametric Magnetic Resonance Imaging of the Prostate: Repeatability of Volume and Apparent Diffusion Coefficient Quantification. *Investig. Radiol.* **2017**, *52*, 538–546. [[CrossRef](#)]
45. Fedorov, A.; Schwier, M.; Clunie, D.; Herz, C.; Pieper, S.; Kikinis, R.; Tempny, C.; Fennessy, F. An Annotated Test-Retest Collection of Prostate Multi-parametric MRI. *Sci. Data* **2018**, *5*, 180281. [[CrossRef](#)] [[PubMed](#)]
46. Perez, L.; Wang, J. The Effectiveness of Data Augmentation in Image Classification Using Deep Learning. *arXiv* **2017**, arXiv:1712.04621.
47. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
48. Pérez-García, F.; Sparks, R.; Ourselin, S. TorchIO: A Python Library for Efficient Loading, Preprocessing, Augmentation and Patch-Based Sampling of Medical Images in Deep Learning. *Comput. Methods Programs Biomed.* **2021**, *208*, 106236. [[CrossRef](#)]
49. Rasch, C.; Barillot, I.; Remeijer, P.; Touw, A.; van Herk, M.; Lebesque, J.V. Definition of the Prostate in CT and MRI: A Multi-Observer Study. *Int. J. Radiat. Oncol. Biol. Phys.* **1999**, *43*, 57–66. [[CrossRef](#)]
50. Smith, W.L.; Lewis, C.; Bauman, G.; Rodrigues, G.; D'Souza, D.; Ash, R.; Ho, D.; Venkatesan, V.; Downey, D.; Fenster, A. Prostate Volume Contouring: A 3D Analysis of Segmentation Using 3DTRUS, CT, and MR. *Int. J. Radiat. Oncol.*Biol.*Phys.* **2007**, *67*, 1238–1247. [[CrossRef](#)] [[PubMed](#)]
51. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J.W. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [[CrossRef](#)] [[PubMed](#)]
52. Huang, W.; Chen, Y.; Fedorov, A.; Li, X.; Jajamovich, G.H.; Malyarenko, D.I.; Aryal, M.P.; LaViolette, P.S.; Oborski, M.J.; O'Sullivan, F.; et al. The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge. *Tomography* **2016**, *2*, 56–66. [[CrossRef](#)] [[PubMed](#)]
53. Bianchini, L.; Botta, F.; Origgi, D.; Rizzo, S.; Mariani, M.; Summers, P.; García-Polo, P.; Cremonesi, M.; Lascialfari, A. PETER PHAN: An MRI Phantom for the Optimisation of Radiomic Studies of the Female Pelvis. *Phys. Med.* **2020**, *71*, 71–81. [[CrossRef](#)]
54. Tixier, F.; Rest, L.; Hatt, C.C.; Albarghach, M. Intratumor Heterogeneity Characterized by Textural Features on Baseline 18F-FDG PET Images Predicts Response to Concomitant Radiochemotherapy in Esophageal. *J. Nucl.* **2011**, *52*, 369–378. [[CrossRef](#)]
55. Vallat, R. Pingouin: Statistics in Python. *J. Open Source Softw.* **2018**, *3*, 1026. [[CrossRef](#)]
56. Velazquez, R. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res* **2017**, *77*, 3922–3930. [[CrossRef](#)]
57. Eklund, M.; Jäderling, F.; Discacciati, A.; Bergman, M.; Annerstedt, M.; Aly, M.; Glaessgen, A.; Carlsson, S.; Grönberg, H.; Nordström, T.; et al. MRI-Targeted or Standard Biopsy in Prostate Cancer Screening. *N. Engl. J. Med.* **2021**, *385*, 908–920. [[CrossRef](#)]
58. Zhou, X.; Gao, F.; Duan, S.; Zhang, L.; Liu, Y.; Zhou, J.; Bai, G.; Tao, W. Radiomic Features of Pk-DCE MRI Parameters Based on the Extensive Tofts Model in Application of Breast Cancer. *Phys. Eng. Sci. Med.* **2020**, *43*, 517–524. [[CrossRef](#)] [[PubMed](#)]
59. Jha, A.K.; Mithun, S.; Jaiswar, V.; Sherkhane, U.B.; Purandare, N.C.; Prabhash, K.; Rangarajan, V.; Dekker, A.; Wee, L.; Traverso, A. Repeatability and Reproducibility Study of Radiomic Features on a Phantom and Human Cohort. *Sci. Rep.* **2021**, *11*, 2055. [[CrossRef](#)] [[PubMed](#)]
60. Ibrahim, A.; Refae, T.; Leijenaar, R.T.H.; Primakov, S.; Hustinx, R.; Mottaghy, F.M.; Woodruff, H.C.; Maidment, A.D.A.; Lambin, P. The Application of a Workflow Integrating the Variable Reproducibility and Harmonizability of Radiomic Features on a Phantom Dataset. *PLoS ONE* **2021**, *16*, e0251147. [[CrossRef](#)]
61. Zwanenburg, A.; Leger, S.; Agolli, L.; Pilz, K.; Troost, E.G.C.; Richter, C.; Löck, S. Assessing Robustness of Radiomic Features by Image Perturbation. *Sci. Rep.* **2019**, *9*, 614. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.