Baltic J. Modern Computing, Vol. 11 (2023), No. 4, 703-725
https://doi.org/10.22364/bjmc.2023.11.4.09

Usage of Isotropic MRI Images Improves Prostate Cancer Localization Results

Justinas JUCEVIČIUS[1], Povilas TREIGYS[1], Jolita BERNATAVIČIENĖ[1], Mantas TRAKYMAS[2], Ieva NARUŠEVIČIŪTĖ[2]

[1]Institute of Data Science and Digital Technologies, Vilnius University, Akademijos str. 4, Vilnius, LT-08412, Lithuania
[2]National Cancer Institute, Santariskių str. 1, Vilnius, LT-08660, Lithuania

ORCID 0000-0002-5334-657X, ORCID 0000-0002-6608-5508, ORCID 0000-0001-5435-8348, ORCID 0000-0002-3724-0150, ORCID 0009-0007-3137-6149

justinas.jucevicius@mif.vu.lt, povilas.treigys@mif.vu.lt, jolita.bernataviciene@mif.vu.lt, mantas.trakymas@nvi.lt, ieva.naruseviciute@nvi.lt

Abstract. MRI images are often anisotropic which make two dimensional neural networks perform better than three dimensional ones due to the fact they cannot take spacing between adjacent slices into account. The aim of this study was to validate if an earlier proposed technique for converting anisotropic images to isotropic enable three dimensional networks to yield better results than two dimensional networks and increase overall accuracy in the task of segmenting prostate cancer as it was shown to do in the task of segmenting prostate. In order to achieve the aim a combination of previously proposed image conversion technique as well as no-new U-Net (nnU-Net) was used to localize prostate cancer. Majority of published studies on prostate cancer segmentation deal with images acquired at a single institution, while this study deals with the image dataset gathered from 11 different institutions for both model training and validation allowing the assessment of model generalizability. The results of performed experiments confirmed that moving away from anisotropic images and two-dimensional neural network to isotropic images and three-dimensional neural network can improve accuracy while segmenting prostate cancer in MRI images. This study showed that nnU-Net is able to provide similar accuracy to other studies as well as the ability to distinguish clinically significant cancer from clinically insignificant one. This study also showed that a lot of identified prostate zones cannot neither be confirmed nor rejected as being abnormal due to the nature of how ground truth is established thus revealing the need of prospective accuracy evaluation.

Keywords: prostate cancer localization, neural network, isotropic images

1. Introduction

Cancer statistics show that prostate cancer was the most common tumor among males in the United States of America in 2021 (25.62%) (Siegel et al., 2021), in Europe in 2020 (22.2%) (Dyba et al., 2021) and in Lithuania in 2021 (20.49%) (Institute of Hygiene 2022). In addition to that the data from Global Cancer Observatory owned by World

Health Organization/International Agency for Research on Cancer shows that prostate cancer is the second most common among males worldwide (15.14%) following the leading lung cancer (15.37%) (Ferlay et al., 2020). The data is more ruthless when analyzing different age groups and the prevalence among men aged 55 and more increases up to 48.2% in Lithuania (Institute of Hygiene 2022). Even though the mortality rate of prostate cancer is not as high, it still contributes a lot and is the 2nd most common cause of cancer death among men in the United States of America (after lung cancer) (Siegel et al., 2021), the 3rd in Europe (after lung and colorectum) (Dyba et al., 2021), the 2nd in Lithuania (after lung) (Institute of Hygiene 2022) and the 5th worldwide (after lung, liver, colorectum and stomach) (Ferlay et al., 2020).

Guidelines provided by the European Association of Urology strongly recommend performing MRI and localizing cancer prior to performing a prostate biopsy (Mottet et al., 2021). MRI is not only capable of detecting cancer, but also helps in stratifying the risk and reducing the number of biopsies to perform (Fütterer et al., 2015). Moreover, in order to improve sampling precision and diagnostic yield fusion biopsy can be performed by merging MRI information with real time ultrasound (Marks et al., 2013). Despite that, prostate MRI suffers from large inter-reader variability due to its strong relation to readers' expertise (Giannini et al., 2017).

In recent years, the development of artificial intelligence enabled the use of various technologies such as speech and image recognition. One of the image recognition techniques is contour detection which can separate an object of interest from its background (Long et al., 2015). Such technique is already being used in the field of medicine, including prostate cancer detection (Park et al., 2019, Jucevičius et al., 2016, Goldenberg et al., 2019). A study by Bhattacharya et al. (Bhattacharya et al., 2022) reported accuracy as 0.81 of the area under the receiver operating characteristic curve (AUC) and used convolutional neural networks to detect normal tissue, clinically insignificant cancer and clinically significant cancer on prostate MRI training the network on the data of 74 patients who underwent radical prostatectomy at a single institution and ground truth set based on whole-mount histopathology images and 24 patients with no cancer. Another study by Gibbons et al. (Gibbons et al., 2023) reported accuracy as 0.91 AUC and used voxel wise logistic regression models to differentiate prostate cancer from other tissues by creating cancer risk maps. Logistic regression models were trained and cross-validated on the cohort of 73 patients with MRI images acquired at a single institution and ground truth set based on whole-mount histopathology images. A study by Alley et al. (Alley et al., 2022) reported accuracy as 0.85 AUC and used radiomics based classifier to generate voxel wise prostate tumor probability maps, focusing on the pre-processing pipeline. Classifier was trained on the data of 31 patients acquired in a single institution and ground truth set by manual contour delineation by radiation oncologist. One more study by Giannini et al. (Giannini et al., 2015) reported accuracy as 0.91 AUC and used a computer-aided diagnosis system to generate a voxel wise malignancy probability map of peripheral zone of prostate. Computer-aided diagnosis system was trained on the data of 56 patients acquired at a single institution and ground truth set based on histopathology images. Other studies include neural networks trained on radiologist demarcated lesions confirmed by biopsy and report accuracy as 0.80-0.94 AUC and as DSC of 0.34-0.37 (Kwak et al., 2015, Litjens et al., 2014a, Sumathipala et al., 2018, Schelb et al., 2020, Vente et al., 2021). Despite a number of studies performed on prostate cancer detection, most of the previous

Table 1. Overview of existing studies on prostate cancer detection.

Authors	Dataset	Accuracy	Ground truth	Additional information
Bhattacharya et al., 2022	98 patients single institution T2W, ADC	0.81 AUC	WMHI	Accuracy reported on a lesion level
Gibbons et al., 2023	73 patients single institution T2W, ADC, DCE	0.91 AUC	WMHI	Separate models for transitional and peripheral zones, accuracy reported on a lesion level
Alley et al., 2022	31 patients single institution T2W, DWI, ADC, DCE	0.85 AUC	PI-RADS	Only zones with PI-RADS score of 4 or 5 used, accuracy reported on a voxel level
Giannini et al., 2015	56 patients single institution T2W, DWI, DCE	0.91 AUC	WMHI	Only tumors in peripheral zone bigger than 0.5ml in volume used, accuracy reported on a voxel level
Kwak et al., 2015	108 patients single institution T2W, DWI	0.89 AUC	Biopsy	Accuracy reported on a voxel level
Litjens et al., 2014a	347 patients single institution T2W, DWI, ADC, DCE, PDW	0.81 AUC	Biopsy	Accuracy reported on a patient level
Sumathipala et al., 2018	120 patients six institutions T2W, DWI, ADC	0.93 AUC	Biopsy	Accuracy reported on a patient level
Schelb et al., 2020	259 patients single institution T2W, ADC	0.8 AUC 0.99 SE, 0.24 SP, PI-RADS $\geq$ 3 0.83 SE, 0.55 SP, PI-RADS $\geq$ 4 0.34 DSC	Biopsy	Accuracy reported on a voxel level
Vente et al., 2021	112 patients single institution T2W, ADC	0.37 DSC	Biopsy	Segmenting and grading cancer at the same time, accuracy reported for any grade cancer

T2W – T2-weighted, DWI – diffusion weighted imaging, ADC – apparent diffusion coefficient, DCE – dynamic contrast-enhanced, PDW – proton density weighted, AUC – area under receiver operating characteristic curve, WMHI – whole-mount histopathology images from radical prostatectomy, PI-RADS – prostate imaging reporting and data system, DSC – Dice similarity coefficient, SE – sensitivity, SP – specificity.

studies are limited due to data acquired in a single institution and the lack of validation on external datasets, dealing with some areas of the prostate only or ground truth being established based on whole-mount histopathology images, which in turn means focusing only on clinically significant cancers that require immediate treatment and ignore clinically insignificant cancers where active surveillance is required (Arif et al., 2020) (Table 1).

Literature overview highlights the difficulty of comparing results of different models given that different methods are used to establish ground truth and different metrics are used to report model accuracy. Most of the studies use database from single institution which does not reveal if models can be generalized well.

The desire to have high-resolution isotropic 3D medical images (defined as uniform voxel size in all three dimensions) in clinical practice together with usually no feasible way to acquire it (Sander et al., 2021) leads to a hypothesis that prostate cancer detection can be improved if dealt with this issue. While there are several other studies investigating this issue (Meyer et al., 2021, Liu et al., 2021), none of them deals with the conversion of anisotropic ground truth labels. While conversion of MRI data before labels are established may work it introduces a lot of burden to experts annotating the images due to introducing many additional slices to be taken care of. The aim of this study therefore was to validate if converting anisotropic MRI image data as well as ground truth labels defined as the ratio between in-plane and out-of-plane spacing being higher than 3 to isotropic can improve prostate cancer segmentation results as it was previously shown to work while segmenting prostate (Jucevičius et al., 2022). In the following section, the dataset as well as the machine learning technique used is presented and two performed experiments are described. Section 3 provides the results for the performance of the experiments executed. Section 4 presents the discussion and provides the highlights.

2. Materials and Methods

2.1. Dataset

In this study MRI images provided by the National Cancer Institute of Lithuania were used. The MRI images themselves were acquired 11 different centers throughout Lithuania (Table 2). Images captured with highest *b*-value for diffusion weighted imaging were used, highest *b*-values varied between different institutions and composed a list of 1000, 1200, 1400, 1500, 1600 and 2000.

Table 2. Imaging tools characteristics at centers where MRI images were acquired.

Center number	Magnetic field strength	Voxel spacing (in-plane/out-of-plane, mm)	Vendor
1	1.5T	0.39 / 2.5	General Electrics
2	3T	0.39-0.78 / 2.5	General Electrics
3	1.5T	0.39-0.65 / 2.5	General Electrics
4	3T	0.70 / 2.5	General Electrics
5	1.5T	0.39-0.70 / 2.5	Philips
6	3T	0.69 / 2.5	Philips
7	1.5T	0.39-0.36 / 2.5	General Electrics
8	1.5T	0.39 / 2.5	General Electrics
9	1.5T	0.39 / 2.5	Siemens
10	3T	0.39-0.70 / 2.5	Siemens
11	1.5T	0.65-0.78 / 2.5	General Electrics

Fusion biopsy with a help of a needle guiding template with a grid of 5mm by 5mm (Figure 1) was performed including both targeted and systematic samples after identifying regions of interest and PI-RADS evaluation at a single center for 146 patients. Data of 120 patients were included in the dataset, excluding data with noise (e.g., due to hip implant) and missing values (missing at least one of axial T2-weighted (T2W), diffusion weighted (DWI) or apparent diffusion coefficient (ADC) MRI sequences).

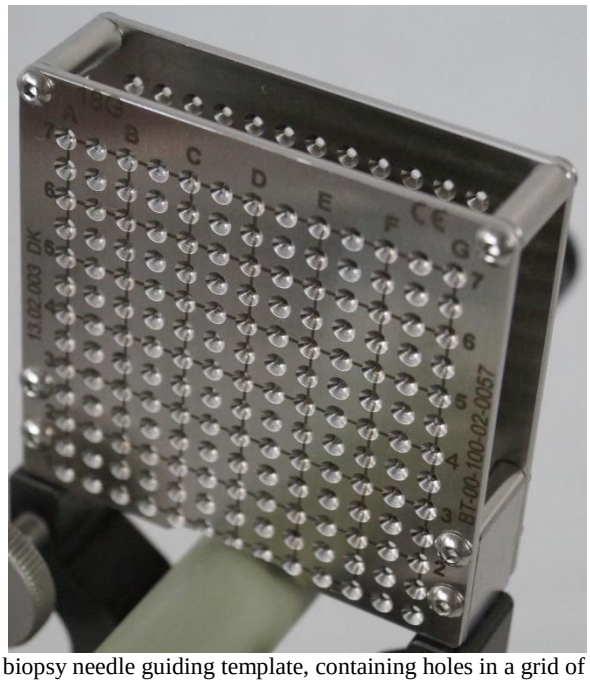

Figure 1. Prostate biopsy needle guiding template, containing holes in a grid of 5mm by 5mm.

All the patients had two additional binary images present: the first one containing prostate mask and the second one containing abnormal zones mask both manually segmented by expert radiologists where zero and one indicated background and prostate/abnormal zone respectively. A total of 216 abnormal zones were identified by radiologists prior to biopsy. All the abnormal zones were classified into 3 categories based on biopsy results: no cancer, clinically insignificant cancer (Gleason score 3+3) and clinically significant cancer (Gleason score 3+4 or 4+3). If there was more than one biopsy sample taken from a single zone, it was classified according to the most malignant one, if there was no biopsy sample taken from a zone, it was classified as no cancer (Figure 2). The total numbers of zones falling into each category were: 45 – clinically significant, 56 – clinically insignificant, 115 – no cancer.

Zone classification showed that 43 out of 120 patients had no cancer at all and the remaining 77 patients had clinically insignificant cancer and/or clinically significant cancer (Table 3).

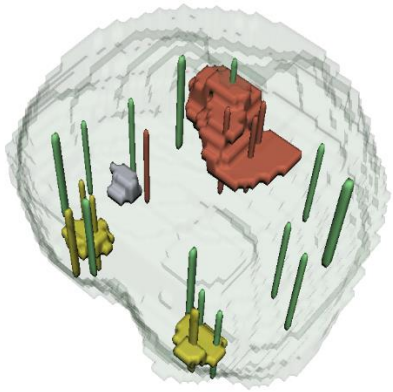

Figure 2. Example of abnormal prostate zone classification. Red, yellow, and green bars indicate clinically significant, clinically insignificant and no cancer biopsy samples respectively. The red zone indicates an area classified as containing clinically significant cancer even though it has biopsy samples with no cancer detecting passing through. Two yellow zones indicate areas classified as clinically insignificant cancers despite having multiple biopsy samples with no cancer detected passing through them as well. Zone colored in gray indicates an area with no biopsy samples taken from and classified as not containing cancer.

Table 3. The distribution of the dataset of 120 patients based on the lesion classification.

No Cancer	At least one lesion with clinically insignificant cancer	At least one lesion with clinically significant cancer
43	43	42

All MRI sequences were resampled to match T2W in terms of size and spacing, as well as registered to have the same position and orientation. Binary masks were segmented on T2W images and were not additionally processed. Image intensity value for each voxel was then set to 0 if the voxel was outside prostate bounds.

2.2. Model Architecture

2.2.1. No-new U-Net

Authors have chosen to use nnU-Net due to it being the best opened source algorithm from the submissions to PROSTATE12 challenge (Litjens et al., 2014b, Jucevičius et al., 2021) as well as having shown its great performance while dealing with 49 different segmentation tasks including, but not limited to prostate segmentation. One more reason for choosing nnU-Net is its key idea of wrapping the standard U-Net by introducing the automated workflow for resolving required hyperparameters and adapting to arbitrary datasets without expensive re-optimization which explains its name – no-new U-Net (Isensee et al., 2020a). The authors of nnU-Net showed that while dealing with anisotropic images, defined as having a ratio greater than 3 between maximum and minimum spacing value of all axes, 2D model generally works better (Isensee et al., 2020a).

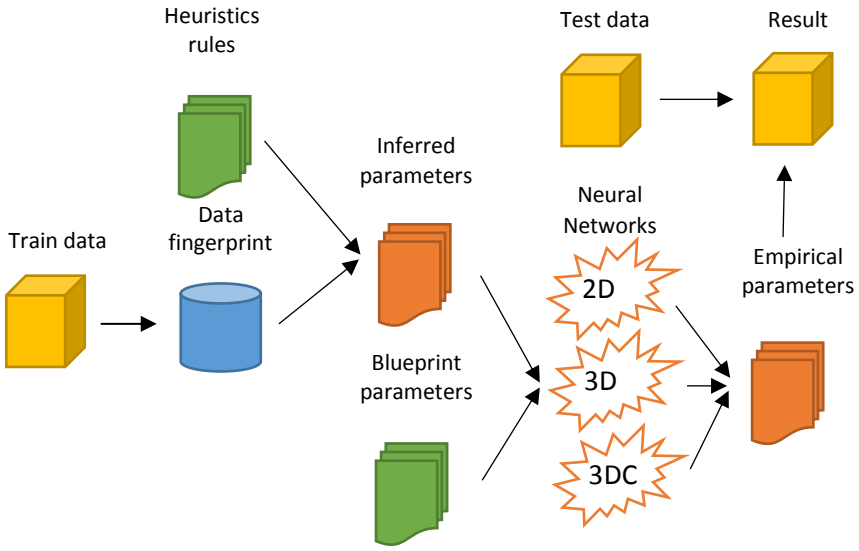

Figure 3. The nnU-Net framework workflow. Framework defines three types of parameters: predefined, derived and collected empirically, combining all type of parameters enables the framework to perform well on unseen datasets. 3DC – 3D network with a cascade.

The nnU-Net workflow (Figure 3) starts by taking in the train data and extracting dataset fingerprint which includes various statistical information on image intensity

values as well as spacing information, image sizes and class ratios. Dataset fingerprint is then used to resolve inferred parameters such as batch and patch sizes, the target image resampling including target voxel spacings as well as image normalization technique based on a set of predefined heuristic rules. Another set of blueprint parameters which do not depend on the data, are predefined and include neural network architecture, loss function, data augmentation and training schedule are then used together with inferred parameters to create two to three different pipelines which are then trained and cross-validated using 5 folds before selecting the best one. Created pipelines include 2D network, 3D network and if applicable 3D network with a cascade which first operates on downsampled images and then refines segmentation result on full resolution.

The nnU-Net is implemented in Python utilizing the PyTorch framework and uses the Batchgenerators library (Isensee et al., 2020b) in its pipeline to augment train data on the fly by using rotation, scaling, Gaussian noise, Gaussian blur as well as other techniques.

2.2.2. Image resampling

Due to convolutional neural networks not taking care of voxel spacing information, the nnU-Net calculates target data spacing and resamples all images to it. The process of resampling includes third order spline interpolation when dealing with MRI data and intensity values, and nearest neighbor interpolation when dealing with binary mask images. Such resampling makes intensity values transition smooth throughout the volume, however the same cannot be said for anisotropic binary mask images, which when visualized in 3D have very rough edges.

A new method for resampling binary mask images has been previously proposed so the masks would transition smoothly throughout the volume as well (Jucevičius et al., 2022). The method calculates a new average mask shape to be placed in the middle between two adjacent slices repeating the process until spacings in all axes are as close to the minimum spacing as possible and repeating the process one more time would make it lower than the minimum spacing. If at least one of the two adjacent slices is empty, then the average mask shape is also empty, otherwise the average mask shape is first calculated on a row-by-row basis by searching for groups of intersecting nonzero values and taking the average of their minimum and maximum coordinates rounded down and up respectively as coordinates for a new group of nonzero values in the intermediate layer (Figure 4).

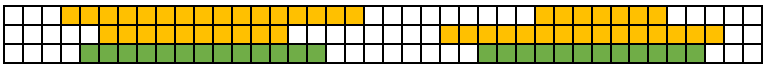

Figure 4. Calculating the boundaries of average shape for one row with multiple groups of intersecting intervals. Yellow – existing rows, green – calculated average row.

If rows contain non-intersecting gaps, they are additionally processed after processing nonzero values by calculating the average between the minimum gap

coordinate and the middle of the gap and the maximum gap coordinate and the middle of the gap and rounding them down and up respectively (Figure 5).

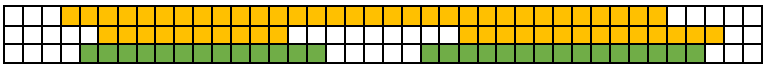

Figure 5. Calculating the boundaries of average shape for one row with a non-intersecting gap. Yellow – existing rows, green – calculated average row.

The process is repeated on a column-by-column basis using the resulting shape and both adjacent slices separately and resulting two shapes are then combined into one where pixel has a value of 1 if it has a value of 1 in at least one of the two shapes. The process is then again repeated with transposed adjacent slices and the transposed resulting shape is combined into one with previous shape where pixel has a final value of 1 if it has a value of 1 in at least one of the two shapes (Algorithm 1).

ALGORITHM 1: BINARY MASK RESAMPLING

```
    Input: Binary mask image
    Output: Resampled binary mask image
 1  spacing_x ← get the spacing of original mask x axis
 2  new_spacing_z ← get the spacing of original mask z axis
 3  target_spacing = new_spacing_z * 1.5
 4  img ← assign input image
 5  while (new_spacing_z > target_spacing) do
 6      n ← get number of slices in img
 7      new_img ← create empty image of same size as img, and increase the
        number of slices by n-1
 8      for (z = 0; z < n - 1; z++) do
 9          slice1 = img[z]
10          slice2 = img[z + 1]
11          new_slice_row ← create a new empty slice of same size as slice1
12          for each line y of slice1 do
13              intervals ← get intersecting intervals of line y in adjacent slices
14              for each pair of intervals do
15                  intervalStart = floor(avg(interval1Start, interval2Start))
16                  intervalEnd = ceil(avg(interval1End, interval2End))
17                  for (x = intervalStart; x <= intervalEnd; x++) do
18                      new_slice_row[x][y] = 1
19                  end
20              end
21              gaps ← get not intersecting gaps of line y in adjacent slices
22              for gap in gaps do
```

```
23          gapStart = floor(avg(gapStart, gapMid))
24          gapEnd = ceil(avg(gapMid, gapEnd))
25          for (x = gapStart; x <= gapEnd; x++) do
26              new_slice_row[x][y] = 0
27          end
28      end
29      end
30      new_slice_col ← repeat rows 11-29 with slice1 and slice2 transposed
        and transpose back the result
31      new_slice_row2 ← repeat rows 11-29 with slice1 and new_slice_row
        transposed and transpose back the result
32      new_slice_row3 ← repeat rows 11-29 with new_slice_row and slice2
        transposed and transpose back the result
33      new_slice_col2 ← repeat rows 11-29 with slice1 and new_slice_col
34      new_slice_col3 ← repeat rows 11-29 with new_slice_col and slice2
35      new_slice = or(new_slice_col2, new_slice_col3, new_slice_row2,
        new_slice_row3)
36      new_img[z * 2] = slice1
37      new_img[z * 2 + 1] = new_slice
38  end
39  new_img[n * 2 - 2] = img[n – 1]
40  img = new_img
41 end
```

The resulting mask images are both isotropic and much smoother than those resampled using nearest neighbor interpolation (Figure 6) giving a more accurate representation of contours of a region of interest to be passed as input to the neural network. Segmentation result comparison when using different segmentation mask resampling techniques is provided further in Section 3.

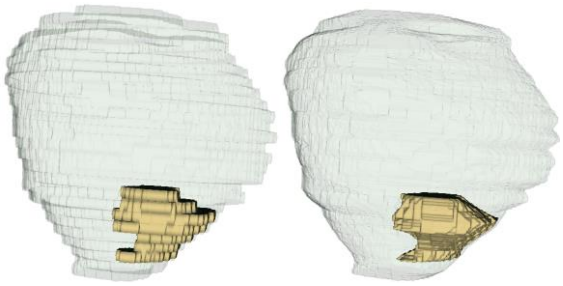

Figure 6. Example of resampling using nearest neighbor interpolation (on the left) and the proposed method (on the right). Transparent green – whole prostate, yellow – abnormal zone. Even though the number of slices is the same in both cases, segmentation masks on the right are much smoother and provide more accurate boundaries to be passed as input to the neural network.

2.2.3. Performance Measures

The performance metrics are crucial in verifying the robustness of a model. Dice similarity coefficient (DSC) (Klein et al., 2008) was selected to be used in the context of experiments performed due to it being part of the loss function, AUC values were calculated on voxel level, patient level, and lesion level to be able to compare the result to other studies even though it's not a suitable metric for unbalanced datasets. While calculating AUC on a lesion level true positive was set if there was any overlay between ground truth and identified lesion, true negative was set only when the patient had no ground truth lesions and no lesions were identified. Dice similarity coefficient measures the overlap between the ground truth and resulting segmentations. The value of Dice similarity coefficient varies from one to zero, where 1 depicts full overlap and 0 represents no intersection. Dice similarity coefficient is expressed by the formula:

$$DSC = 2 \times TP / (2 \times TP + FP + FN), \quad (1)$$

where TP represents the number of voxels correctly identified as abnormal, FN represents the number of voxels incorrectly identified as background and FP represents the number of voxels incorrectly identified as abnormal.

To validate generalization and performance on unseen data authors have used a k fold cross validation (Ravi et al., 2017) with k=5. The k-fold cross validation consists of partitioning the dataset into k-fold and performing training on all but one fold and then testing on the one that has been left out. This procedure is repeated until each fold has been left out once.

In addition to Dice similarity coefficient, authors have also evaluated the confusion matrices, classifying segmented zone as true positive if there's an overlap of at least 1 voxel, false positive if segmented zone does not overlap any of the ground truth zones and false negative if ground truth zone is not overlapped by any of the segmented zones. There were no zones to be classified as true negative, therefore authors have counted the number of patients who had no cancerous zones as a reference and no zones had been identified by the system as false negative. Authors have calculated two confusion matrices for each of the experiment by comparing identified zones against reference zones and biopsy samples.

2.3. Experiment Setup

2.3.1. Isotropic and Anisotropic Data

The first experiment was aimed at confirming whether 3D neural network model gives better results in segmenting prostate cancer when converting anisotropic data to isotropic than its 3D counterpart trained on anisotropic images and both 2D neural network models trained with anisotropic and isotropic data. Such performance improvement was already identified in the previous study on prostate segmentation (Jucevičius et al., 2022). Data is converted from anisotropic to isotropic using the method described in Section 2.2.2. As part of preprocessing built-in in the nnU-Net framework all images were normalized and clipped to a largest non-zero area. MRI image modalities of T2W, DWI and ADC were used as separate input image channels. In this part of experiment a

total of 4 models were trained on the same data, using the same data splits into 5 different folds: 2D model trained on anisotropic images, 2D model trained on isotropic images, 3D model trained on anisotropic images and 3D model trained on isotropic images.

In order to compare the results of models using different data, segmentation results of the models that used isotropic data were resampled back to their original size by using the nearest neighbor interpolation.

The first part of experiment was carried out by setting only 45 zones classified as clinically significant cancer and distributed among 42 patients out of 120 as ground truth.

In addition to using the default probability threshold of 0.5 defined in the nnU-Net framework when binarizing the resulting softmax probability map, authors have also tried to find the optimal threshold for this task. Since abnormal regions are relatively small compared to the whole prostate the resulting voxel set is imbalanced, therefore in order to find the optimal threshold, authors have used precision recall curve (Davis et al., 2006) and the F-Score, which is a harmonic mean of precision and recall (Sofaer et al., 2019). To find the optimal threshold all unique probability values from resulting probability maps calculated during cross validation were used as possible threshold values, calculating the F-Score for each of them and looking for the maximum value. F-Score is calculated from the precision and recall, where the precision is the number of true positive results divided by the number of all positive results and the recall is the number of true positive results divided by the number of all samples that should have been identified as positive. F-Score is expressed by the formula:

$$\text{F-Score} = 2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall}). \quad (2)$$

2.3.2. Clinically Significant and Clinically Insignificant Cancer

The second experiment was aimed at evaluating whether the system is able to segment both clinically significant and clinically insignificant cancers as well as distinguish one from the other. For this part of the experiment one additional 3D model was trained using isotropic data and the same pipeline. To train the new model a total of 101 zones distributed among 77 out of 120 patients were used to set ground truth and included both clinically insignificant and clinically significant zones in the same class. For this experiment additional model was trained using both clinically significant and clinically insignificant zones as cancerous zones passed for training. Both conversions to isotropic data as well as finding the optimal threshold were applied for the second experiment as well.

2.3.3. Case Analysis

The third part of the study was aimed at manually analyzing segmentation results in an attempt to try and evaluate whether discrepancies between ground truth segmentations and those provided by the system are due to errors in the system or due to original misclassification.

3. Results

3.1. Isotropic and Anisotropic Data

As expected, the first experiment confirmed that 3D neural network model trained on isotropic data gives the best segmentation results measured by Dice similarity coefficient (0.2510 – 3D isotropic vs. 0.1221 – 2D isotropic vs. 0.0243 – 3D anisotropic vs. 0.1558 – 2D anisotropic). All different values from probability maps constructed during cross validation of 3D isotropic model were used as thresholds to plot precision-recall curve and find optimal threshold to be used for classifying voxels (Figure 7).

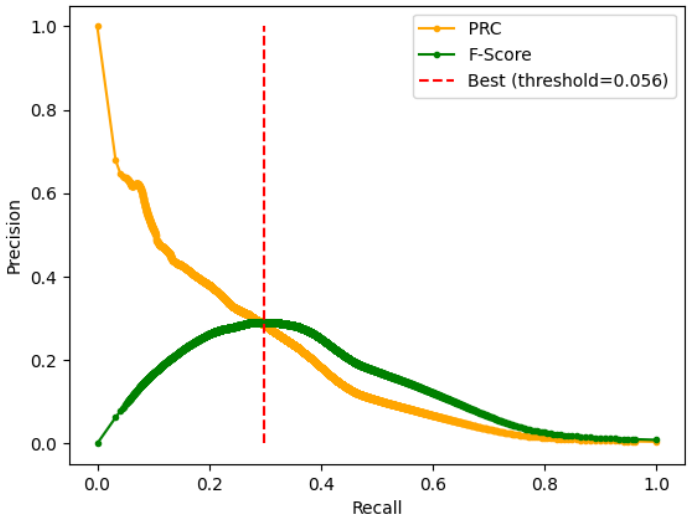

Figure 7. Precision-recall curve plotted by using all unique values from probability maps constructed during cross validation while training isotropic 3D neural network to locate clinically significant prostate cancer as voxel classification thresholds. Precision, recall and F-Score were calculated at each threshold, optimal threshold selected based on maximum F-Score value. PRC – Precision-Recall curve.

Using the optimal threshold of 0.0569 instead of the default one of 0.5 increased the score from 0.2510 to 0.3037 (Table 4).

Table 4. Comparison of clinically significant cancer localization results between 2D and 3D models trained on anisotropic and isotropic images.

	2D anisotropic	**2D isotropic**	**3D anisotropic**	**3D isotropic**
DSC with a default threshold of 0.5	0.1558	0.1221	0.0243	**0.2510**
DSC with an optimal threshold of 0.0569	0.1834	0.2388	0.1230	**0.3037**
AUC on voxel level	0.6912	0.6243	0.7021	**0.8543**
AUC on patient level	0.7013	0.6564	0.7102	**0.7265**
AUC on lesion level	0.4904	0.3409	**0.5468**	0.5420

3D neural network model identified a total of 70 zones classified as containing clinically significant cancer by segmenting each left out fold during cross validation. A confusion matrix (Table 5) was composed by comparing identified zones against reference zones, where true negative was defined as patient not having neither identified nor referenced any clinically significant cancer zones. This matrix showed that only half of the clinically significant cancers could be identified and that almost two thirds of identified zones were false positive.

Table 5. Confusion matrix composed by comparing identified zones against reference zones, using 3D model trained on isotropic data including only clinically significant cancer.

	Predicted positive	**Predicted negative**
Actual positive	25 (53.19%)	22 (46.81%)
Actual negative	45 (47.87%)	49 (52.13%)

3.2. Clinically Significant and Clinically Insignificant Cancer

The second experiment used only 3D models and was aimed at comparing the ability to detect all cancer cases opposing to detecting only clinically significant cancer at the same time evaluating the ability to distinguish them one from the other. Segmentation results of the second experiment reached DSC of 0.3590 when using default threshold of 0.5. An optimal threshold of 0.0008 (Figure 8) was identified for this model which increased DSC to 0.4536. Segmentation results from AUC perspective yielded 0.8558 on a voxel level, 0.7992 on a patient level and 0.5140 on a lesion level.

3D neural network trained to detect all prostate cancer cases including both clinically significant and clinically insignificant cancer identified a total of 182 zones when segmenting left out folds during cross validation. Corresponding confusion matrix (Table 6) shows that 40.66% out of all identified zones were actually positive, however 29.52% of the reference zones were not identified as cancerous.

Confusion matrix composed by comparing identified zones against biopsy samples (Table 7) showed similar results, however it’s worth noting, that only 29 out of 104 false

positive zones were actually negative, the remaining 75 zones had no overlap with any of the biopsy samples and the assigned class cannot be taken for granted.

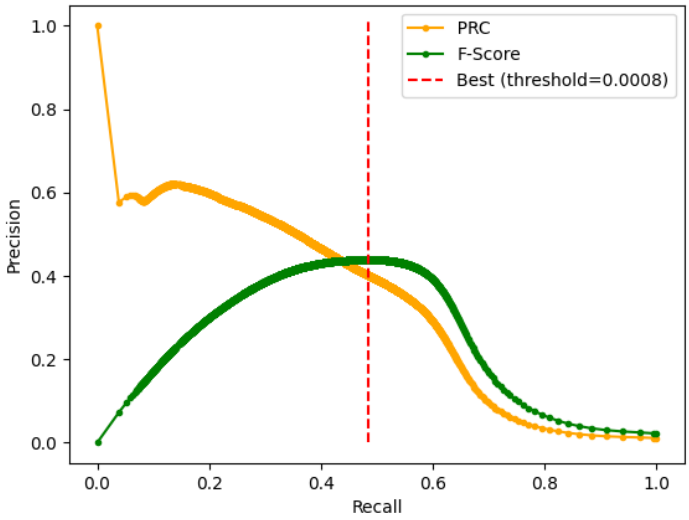

Figure 8. Precision-recall curve plotted by using all unique values from probability maps constructed during cross validation while training isotropic 3D neural network to locate clinically significant and clinically insignificant prostate cancer as voxel classification thresholds. Precision, recall and F-Score were calculated at each threshold, optimal threshold selected based on maximum F-Score value. PRC – Precision-Recall curve.

Table 6. Confusion matrix composed by comparing identified zones against reference zones, using 3D model trained on isotropic data including both clinically significant and clinically insignificant cancers.

	Predicted positive	**Predicted negative**
Actual positive	74 (70.48%)	31 (29.52%)
Actual negative	108 (89.26%)	13 (10.74%)

Table 7. Confusion matrix composed by comparing identified zones against biopsy samples, using 3D model trained on isotropic data including both clinically significant and clinically insignificant cancers.

	Predicted positive	**Predicted negative**
Actual positive	78 (71.56%)	31 (28.44%)
Actual negative	29 + 75* (88.89%)	13 (11.11%)

* 75 zones had no overlap with any of the biopsy samples and were classified as negative.

Since the aim of the system is to aid in determining the location to perform biopsy on and the biopsy itself is often performed by using a needle guiding template with a grid of

5mm by 5mm (Figure 1), authors tried ruling out identified zones which if placed in a bounding box would have both dimensions lower than 5mm looking from a plane perspective.

The results of applying such a filter were interesting: 87 out of previously identified 182 zones using the model trained to detect all cancer cases and the identified optimal threshold had to be ruled out. The corresponding confusion matrix (Table 8) showed that true positive rate dropped a bit, but the rate of false positive decreased significantly, even though still more than half of false positive cases cannot be taken for granted due to not overlapping with any biopsy sample and could potentially be reclassified as true positive.

Table 8. Confusion matrix composed by comparing identified zones with at least one dimension greater than 5mm in plane axes against biopsy samples, using 3D model trained on isotropic data including both clinically significant and clinically insignificant cancers.

	Predicted positive	**Predicted negative**
Actual positive	69 (66.35%)	35 (33.65%)
Actual negative	24 + 31* (74.32%)	19 (25.68%)

* 31 zone had no overlap with any of the biopsy samples and were classified as negative.

3.3. Case Analysis

While manually analyzing the differences between ground truth segmentations and those provided by the model described in Section 2.3.2, authors have noticed that there were quite a few cases when identified zones were close to but did not intersect the reference zones. In addition to that, the same biopsy samples from both of those identified and reference zones and looking from a biopsy perspective such identified zones would be classified as true positive. There were also some cases where zones identified by the model were far away from reference segmentation but considered true positive based on biopsy results (Figure 9).

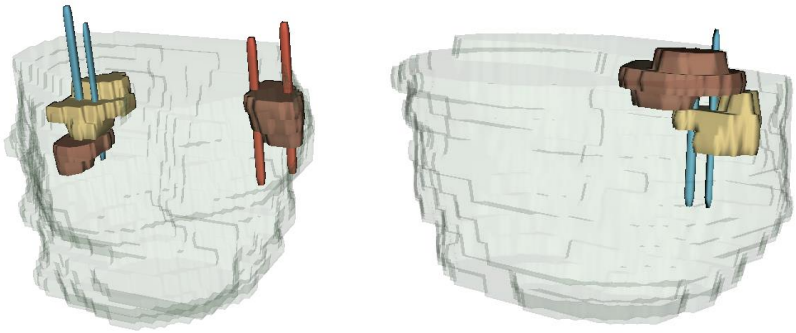

Figure 9. Examples of segmentation results. Brown zones depict zones identified by the model and yellow zones indicate reference segmentation. Blue and red sticks represent clinically insignificant and clinically significant biopsy samples respectively.

Radiologists have retrospectively analyzed some of such cases in order to evaluate zones identified by the model trained to detect both clinically significant and clinically insignificant prostate cancer (Figure 10) and reported that zones identified by the model could indeed contain clinically significant cancer as indicated by the biopsy sample, but it's hard to confirm that due to the uncertainty in which part of the biopsy sample cancer was detected. In addition to that, radiologists also informed that such zones might have been missed in the initial segmentation due to difficult image interpretation in the central zone of the prostate.

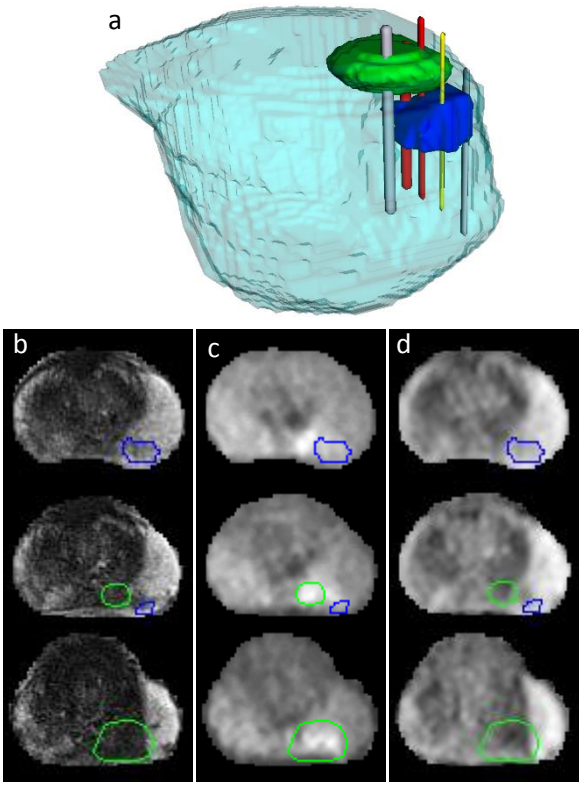

Figure 10. Example of retrospectively analyzed case. Zone identified by the model trained to detect any prostate cancer is represented in green and reference segmentation zone is represented in blue. Red, yellow and gray sticks represent clinically significant, clinically insignificant and no cancer biopsy samples respectively. a – contains 3D model of the whole prostate, identified and reference zones and biopsy samples, b, c, d, contain T2W, DWI and ADC slices respectively with identified and reference zones overlayed on top.

4. Discussion and Conclusions

In this study, authors validate whether converting data to isotropic gives better results while localizing prostate cancer using nnU-Net as it was previously shown to work while

segmenting prostate itself. The Dice similarity coefficient while using 3D neural network with isotropic data to detect clinically significant cancer was 0.2510, which is about 60% better than 2D network with anisotropic data (0.1558). This study also showed that the DSC can be further improved up to 0.3037, by using a non-default segmentation threshold.

Furthermore, including clinically insignificant cancers to the training data yields the DSC of 0.4536. Having several models including both clinically insignificant and clinically significant cancers is of value since this way the system is enabled to not only detect prostate cancer but classify detected lesions as well.

Despite that, however, it is interesting to note that resulting optimal thresholds for binarizing probability maps are really low, but it can be explained by the fact that the dataset is unbalanced as there is a relatively small number of voxels containing cancerous tissue opposing to the number of non-cancerous voxels. Another reason for that is the way ground truth is established, which as showed in Figure 2 assigns a single class to a whole region of connecting voxels when multiple biopsy samples with different results were taken from the area. It is also worth mentioning that such low thresholds do not introduce a lot of new possible abnormal prostate zones, but in most cases only increases the volume of the same zones identified while using a default threshold of 0.5 (Figure 11).

Our study has some important strengths. First, our data comes from 11 different institutions having different MRI equipment and using different protocols which confirms the generalizability of our findings, while other datasets are limited to 1 or 2 scanners (Giannini et al., 2015, Giannini et al., 2021, Pellicer-Valero et al., 2022).

Secondly, our dataset contains cases with no confirmed cancer in addition to both clinically significant and clinically insignificant cancers as well as cancer located throughout the whole prostate including both peripheral and transitional zones even though tumors in both zones have different texture characteristics (Turkbey et al., 2019). Whereas other studies either use only cases with confirmed cancer (Giannini et al., 2015) or use cancers of a single zone only (Niaf et al., 2012, Niaf et al., 2014, Vos et al., 2010).

Our study also has some limitations. First, cancers in both transitional and peripheral zones were not separated and have been segmented in the same way, which might have added to relatively low Dice similarity coefficients. Second, the reference standard was established by using radiologist's annotations which can miss up to 40% clinically significant cancer (Le et al., 2015) and were not reannotated after biopsy, resulting in lesions having clinically significant and clinically insignificant cancer and even no cancer according to biopsy results still classified as clinically significant as shown in Figure 1, leading to a lot larger zone set as a ground truth than it is in reality. Finally, due to lesions often being small with ill-defined margins and a very high inter-observer variability (Steenbergen et al., 2015), the relatively low DSC in our study as well as other studies that reported 0.34 (Schelb et al., 2020) and 0.37 (Vente et al., 2021) must be interpreted with caution and some other metrics should be chosen instead to provide a more objective look on the actual performance of the model.

In this study, authors have demonstrated that the use of a novel way in converting binary masks to isotropic which are then used for model training yields better performance results. Authors have also shown that unmodified nnU-Net can give similar prostate localization accuracy as other studies even though they are not high enough to be used in clinical practice. Authors have developed a model that can detect up to 70%

of actual positive prostate cancers. Future developments will include the separation of transitional and peripheral zones as well as the prospective evaluation of the model, which is extremely interesting considering the large number of identified zones that there neither confirmed nor rejected by existing biopsy samples and additional retrospective analysis of such cases showed promising results of being able to confirm them.

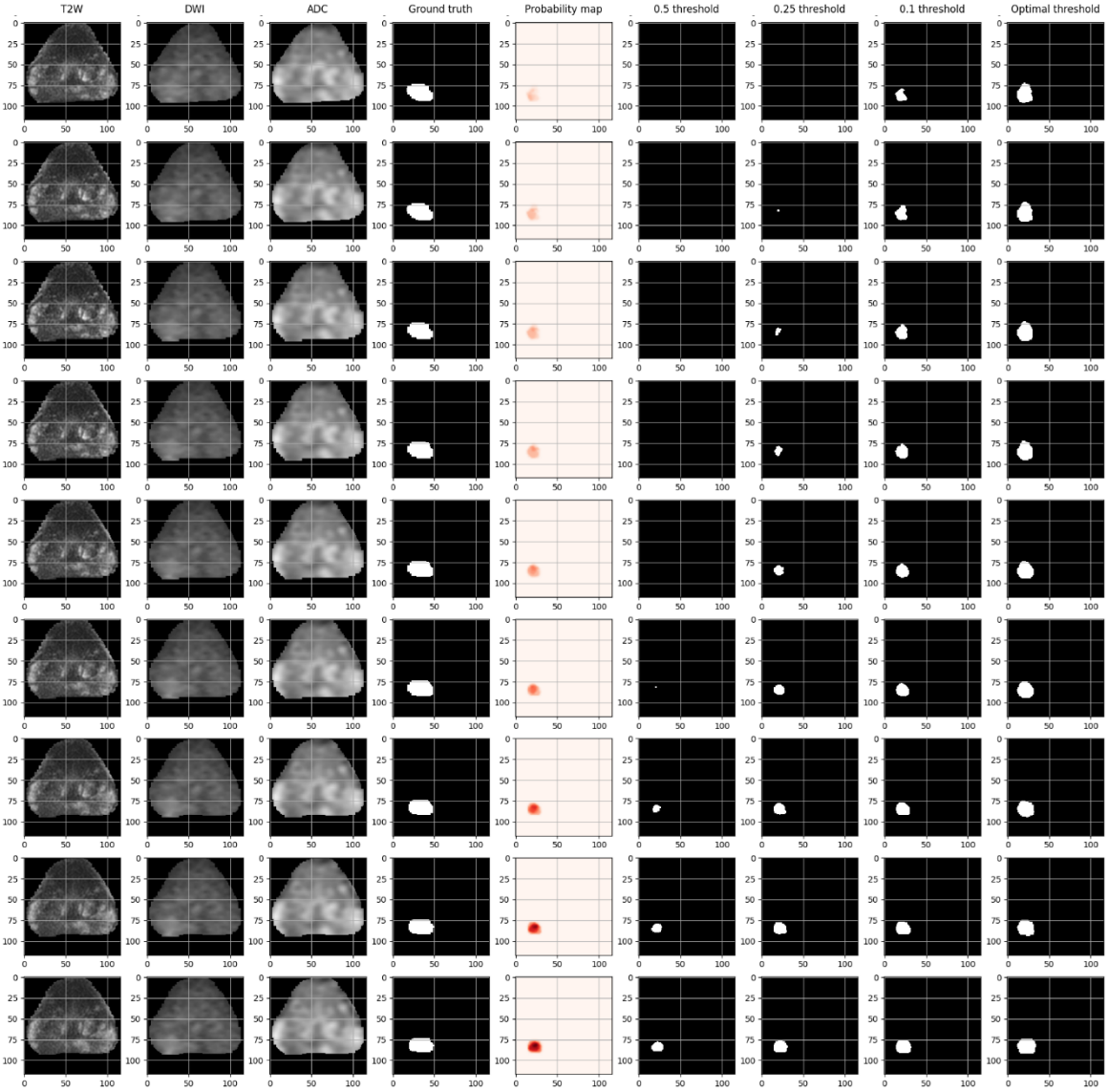

Figure 11. Example of how segmented zones changes while reducing the threshold used for binarizing resulting probability map.

Acknowledgements

The authors are thankful for the high-performance computing resources provided by the Information Technology Open Access Center at the Faculty of Mathematics and Informatics of Vilnius University Information Technology Research Center.

References

Alley, S., Jackson, E., Olivié, D., Van Der Heide, U. A., Ménard, C., Kadoury, S. (2022). Effect of magnetic resonance imaging pre-processing on the performance of model-based prostate tumor probability mapping. Physics in Medicine and Biology, 67(24), 245018. https://doi.org/10.1088/1361-6560/ac99b4

Arif, M., Schoots, I. G., Castillo Tovar, J., Bangma, C. H., Krestin, G. P., Roobol, M. J., Niessen, W., Veenland, J. F. (2020). Clinically significant prostate cancer detection and segmentation in low-risk patients using a convolutional neural network on multi-parametric MRI. European Radiology, 30(12), 6582–6592. https://doi.org/10.1007/s00330-020-07008-z

Bhattacharya, I., Seetharaman, A., Kunder, C., Shao, W., Chen, L. C., Soerensen, S. J., Wang, J. B., Teslovich, N. C., Fan, R. E., Ghanouni, P., Brooks, J. D., Sonn, G. A., Rusu, M. (2022). Selective identification and localization of indolent and aggressive prostate cancers via CorrSigNIA: an MRI-pathology correlation and deep learning framework. Medical Image Analysis, 75, 102288. https://doi.org/10.1016/j.media.2021.102288

Davis, J., Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. Proceedings of the 23rd International Conference on Machine Learning - ICML '06. https://doi.org/10.1145/1143844.1143874

Dyba, T., Randi, G., Bray, F., Martos, C., Giusti, F., Nicholson, N., Gavin, A., Flego, M., Neamtiu, L., Dimitrova, N., Negrão Carvalho, R., Ferlay, J., Bettio, M. (2021). The European cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers. European Journal of Cancer, 157, 308–347. https://doi.org/10.1016/j.ejca.2021.07.039

Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2020). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.fr/today, accessed August 19, 2022.

Fütterer, J. J., Briganti, A., De Visschere, P., Emberton, M., Giannarini, G., Kirkham, A., Taneja, S. S., Thoeny, H., Villeirs, G., Villers, A. (2015). Can Clinically Significant Prostate Cancer Be Detected with Multiparametric Magnetic Resonance Imaging? A Systematic Review of the Literature. European Urology, 68(6), 1045–1053. https://doi.org/10.1016/j.eururo.2015.01.013

Giannini, V., Mazzetti, S., Armando, E., Carabalona, S., Russo, F., Giacobbe, A., Muto, G., Regge, D. (2017). Multiparametric magnetic resonance imaging of the prostate with computer-aided detection: experienced observer performance study. European Radiology, 27(10), 4200–4208. https://doi.org/10.1007/s00330-017-4805-0

Giannini, V., Mazzetti, S., Defeudis, A., Stranieri, G., Calandri, M., Bollito, E., Bosco, M., Porpiglia, F., Manfredi, M., De Pascale, A., Veltri, A., Russo, F., Regge, D. (2021). A Fully Automatic Artificial Intelligence System Able to Detect and Characterize Prostate Cancer Using Multiparametric MRI: Multicenter and Multi-Scanner Validation. Frontiers in Oncology, 11. https://doi.org/10.3389/fonc.2021.718155

Giannini, V., Mazzetti, S., Vignati, A., Russo, F., Bollito, E., Porpiglia, F., Stasi, M., Regge, D. (2015). A fully automatic computer aided diagnosis system for peripheral zone prostate cancer detection using multi-parametric magnetic resonance imaging. Computerized Medical Imaging and Graphics, 46, 219–226. https://doi.org/10.1016/j.compmedimag.2015.09.001

Gibbons, M., Simko, J. P., Carroll, P. R., Noworolski, S. M. (2023). Prostate cancer lesion detection, volume quantification and high-grade cancer differentiation using cancer risk maps derived from multiparametric MRI with histopathology as the reference standard. Magnetic Resonance Imaging. https://doi.org/10.1016/j.mri.2023.01.006

Goldenberg, S. L., Nir, G., Salcudean, S. E. (2019). A new era: artificial intelligence and machine learning in prostate cancer. Nature Reviews Urology, 16(7), 391–403. https://doi.org/10.1038/s41585-019-0193-3

Institute of Hygiene (2022). Health Statistics. Vilnius, Lithuania. Available from: https://stat.hi.lt/default.aspx, accessed August 19, 2022.

Isensee, F., Jaeger, P. F., Kohl, S. a. A., Petersen, J., Maier-Hein, K. H. (2020). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z

Isensee, F., Jäger, P., Wasserthal, J., Zimmerer, D., Petersen, J., Kohl, S., Schock, J., Klein, A., Roß, T., Wirkert, S., Neher, P., Dinkelacker, S., Köhler, G., Maier-Hein, K. (2020). batchgenerators - a python framework for data augmentation (0.19.6). Zenodo. https://doi.org/10.5281/zenodo.3632567

Jucevičius, J., Treigys, P., Bernatavičienė, J., Briedienė, R., Naruševičiūtė, I., Dzemyda, G., Medvedev, V. (2016). Automated 2D Segmentation of Prostate in T2-weighted MRI Scans. International Journal of Computers Communications and Control, 12(1), 53-60.

Jucevičius, J., Treigys, P., Bernatavičienė, J., Briedienė, R., Naruševičiūtė, I., Trakymas, M. (2021). Investigation of MRI Prostate Localization using Different MRI Modality Scans. 2020 IEEE 8th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE). https://doi.org/10.1109/aieee51419.2021.9435797

Jucevičius, J., Treigys, P., Bernatavičienė, J., Trakymas, M., Naruševičiūtė, I., Briedienė, R. (2022). Segmentation Mask Resampling for MRI Prostate Localization Improvement. 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET). https://doi.org/10.1109/icecet55527.2022.9872602

Klein, S., Van Der Heide, U. A., Lips, I. M., Van Vulpen, M., Staring, M., Pluim, J. P. W. (2008). Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Medical Physics, 35(4), 1407–1417. https://doi.org/10.1118/1.2842076

Kwak, J. T., Xu, S., Wood, B. J., Turkbey, B., Choyke, P. L., Pinto, P. A., Wang, S., Summers, R. M. (2015). Automated prostate cancer detection using T2-weighted and high-b-value diffusion-weighted magnetic resonance imaging. Medical Physics, 42(5), 2368–2378. https://doi.org/10.1118/1.4918318

Le, J. D., Tan, N., Shkolyar, E., Lu, D. Y., Kwan, L., Marks, L. S., Huang, J., Margolis, D. J., Raman, S. S., Reiter, R. E. (2015). Multifocality and Prostate Cancer Detection by Multiparametric Magnetic Resonance Imaging: Correlation with Whole-mount Histopathology. European Urology, 67(3), 569–576. https://doi.org/10.1016/j.eururo.2014.08.079

Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., Huisman, H. (2014). Computer-Aided Detection of Prostate Cancer in MRI. IEEE Transactions on Medical Imaging, 33(5), 1083–1092. https://doi.org/10.1109/tmi.2014.2303821

Litjens, G., Toth, R., Van De Ven, W., Hoeks, C., Kerkstra, S., Van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., Strand, R., Malmberg, F., Ou, Y., Davatzikos, C., Kirschner, M., Jung, F., Yuan, J., Qiu, W., Gao, Q., Edwards, P. E., Maan, B., Heijden, F., Ghose, S., Mitra, J., Dowling, J., Barratt, D., Huisman, H., Madabhushi, A. (2014). Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. Medical Image Analysis, 18(2), 359–373. https://doi.org/10.1016/j.media.2013.12.002

Liu, Y., Liu, Y., Vanguri, R., Litwiller, D., Liu, M., Hsu, H. Y., Ha, R., Shaish, H., Jambawalikar, S. (2021). 3D Isotropic Super-resolution Prostate MRI Using Generative Adversarial Networks and Unpaired Multiplane Slices. Journal of Digital Imaging, 34(5), 1199–1208. https://doi.org/10.1007/s10278-021-00510-w

Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2015.7298965

Marks, L., Young, S., Natarajan, S. (2013). MRI–ultrasound fusion for guidance of targeted prostate biopsy. Current Opinion in Urology, 23(1), 43–50. https://doi.org/10.1097/mou.0b013e32835ad3ee

Meyer, A., Chlebus, G., Rak, M., Schindele, D., Schostak, M., Van Ginneken, B., Schenk, A., Meine, H., Hahn, H. K., Schreiber, A., Hansen, C. (2021). Anisotropic 3D Multi-Stream CNN for Accurate Prostate Segmentation from Multi-Planar MRI. Computer Methods and Programs in Biomedicine, 200, 105821. https://doi.org/10.1016/j.cmpb.2020.105821

Mottet, N., Van Den Bergh, R. C., Briers, E., Van Den Broeck, T., Cumberbatch, M. G., De Santis, M., Fanti, S., Fossati, N., Gandaglia, G., Gillessen, S., Grivas, N., Grummet, J., Henry, A. M., Van Der Kwast, T. H., Lam, T. B., Lardas, M., Liew, M., Mason, M. D., Moris, L., Oprela-Lager, D. E., Poel, H. G., Rouvière, O., Schoots, I. G., Tilki, D., Wiegel, T., Willemse, P. P. M., Cornford, P. (2021). EAU-EANM-ESTRO-ESUR-SIOG Guidelines on Prostate Cancer—2020 Update. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. European Urology, 79(2), 243–262. https://doi.org/10.1016/j.eururo.2020.09.042

Niaf, E., Lartizien, C., Bratan, F., Roche, L., Rabilloud, M., Mège-Lechevallier, F., Rouvière, O. (2014). Prostate Focal Peripheral Zone Lesions: Characterization at Multiparametric MR Imaging—Influence of a Computer-aided Diagnosis System. Radiology, 271(3), 761–769. https://doi.org/10.1148/radiol.14130448

Niaf, E., Rouvière, O., Mège-Lechevallier, F., Bratan, F., Lartizien, C. (2012). Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. Physics in Medicine and Biology, 57(12), 3833–3851. https://doi.org/10.1088/0031-9155/57/12/3833

Park, J., Yun, J., Kim, N., Park, B., Cho, Y., Park, H. J., Song, M., Lee, M., Seo, J. B. (2019). Fully Automated Lung Lobe Segmentation in Volumetric Chest CT with 3D U-Net: Validation with Intra- and Extra-Datasets. Journal of Digital Imaging, 33(1), 221–230. https://doi.org/10.1007/s10278-019-00223-1

Pellicer-Valero, O. J., Marenco Jiménez, J. L., Gonzalez-Perez, V., Casanova Ramón-Borja, J. L., Martín García, I., Barrios Benito, M., Pelechano Gómez, P., Rubio-Briones, J., Rupérez, M. J., Martín-Guerrero, J. D. (2022). Deep learning for fully automatic detection, segmentation, and Gleason grade estimation of prostate cancer in multiparametric magnetic resonance images. Scientific Reports, 12(1). https://doi.org/10.1038/s41598-022-06730-6

Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., Yang, G. Z. (2017). Deep Learning for Health Informatics. IEEE Journal of Biomedical and Health Informatics, 21(1), 4–21. https://doi.org/10.1109/jbhi.2016.2636665

Sander, J., De Vos, B. D., Išgum, I. (2021). Unsupervised super-resolution: creating high-resolution medical images from low-resolution anisotropic examples. Medical Imaging 2021: Image Processing. https://doi.org/10.1117/12.2580412

Schelb, P., Wang, X., Radtke, J. P., Wiesenfarth, M., Kickingereder, P., Stenzinger, A., Hohenfellner, M., Schlemmer, H. P., Maier-Hein, K. H., Bonekamp, D. (2020). Simulated clinical deployment of fully automatic deep learning for clinical prostate MRI assessment. European Radiology, 31(1), 302–313. https://doi.org/10.1007/s00330-020-07086-z

Siegel, R. L., Miller, K. D., Fuchs, H. E., Jemal, A. (2021). Cancer Statistics, 2021. CA: A Cancer Journal for Clinicians, 71(1), 7–33. https://doi.org/10.3322/caac.21654

Sofaer, H. R., Hoeting, J. A., Jarnevich, C. S. (2019). The area under the precision-recall curve as a performance metric for rare binary events. Methods in Ecology and Evolution, 10(4), 565–577. https://doi.org/10.1111/2041-210x.13140

Steenbergen, P., Haustermans, K., Lerut, E., Oyen, R., De Wever, L., Van Den Bergh, L., Kerkmeijer, L. G., Pameijer, F. A., Veldhuis, W. B., Van Der Voort Van Zyp, J. R., Pos, F. J., Heijmink, S. W., Kalisvaart, R., Teertstra, H. J., Dinh, C. V., Ghobadi, G., Van Der Heide, U. A. (2015). Prostate tumor delineation using multiparametric magnetic resonance imaging: Inter-observer variability and pathology validation. Radiotherapy and Oncology, 115(2), 186–190. https://doi.org/10.1016/j.radonc.2015.04.012

Sumathipala, Y., Lay, N., Turkbey, B. (2018). Prostate cancer detection from multi-institution multiparametric MRIs using deep convolutional neural networks. Journal of Medical Imaging, 5(04), 1. https://doi.org/10.1117/1.jmi.5.4.044507

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., Thoeny, H. C., Verma, S., Barentsz, J., Weinreb, J. C. (2019). Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. European Urology, 76(3), 340–351. https://doi.org/10.1016/j.eururo.2019.02.033

Vente, C. D., Vos, P., Hosseinzadeh, M., Pluim, J., Veta, M. (2021). Deep Learning Regression for Prostate Cancer Detection and Grading in Bi-Parametric MRI. IEEE Transactions on Biomedical Engineering, 68(2), 374–383. https://doi.org/10.1109/tbme.2020.2993528

Vos, P. C., Hambrock, T., Barenstz, J. O., Huisman, H. J. (2010). Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI. Physics in Medicine and Biology, 55(6), 1719–1734. https://doi.org/10.1088/0031-9155/55/6/012

Received August 14, 2023, revised December 8, 2023, accepted December 11, 2023