Automatic Prostate Volume estimation on Transabdominal Ultrasound using Deep Neural Networks

Liza Maria Kurucz

TU Delft Delft University of Technology

AUTOMATIC PROSTATE VOLUME ESTIMATION ON TRANSABDOMINAL ULTRASOUND USING DEEP NEURAL NETWORKS

Liza Maria Kurucz

Student number : 4535057

27 May 2024

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University Delft University of Technology Erasmus University Rotterdam

Master thesis project (TM30004 ; 35 ECTS)

Dept. of Biomechanical Engineering, TUDELFT

01-12-2023 *TM30004 – 10-06-2024 TM30004*

Supervisors

Prof. Dr. T. Ruers
Dr. B. Dasht Bozorg
Dr. M. Fusaglia
Dr. J. Veenland
Dr. P. Van Leeuwen
MSc. T. Natali

Thesis committee members

Dr. J. Veenland (Chair)
Dr. Theo van Walsum
Dr. P. Van Leeuwen
Dr. B Dasht Bozorg

Automatic Prostate Volume estimation on Transabdominal Ultrasound using deep neural networks

Author
Liza Maria Kurucz

Supervised by
Prof. Dr. T. Ruers
Dr. B. Dasht Bozorg
Dr. M. Fusaglia
Dr. J. Veenland
Dr. P. van Leeuwen
MSc. T. Natali

A Thesis submitted in fulfillment of requirements for the degree of
Master of Science in Technical Medicine: Delft University of Technology, Erasmus Medical Center, Leids Universitair Medisch Centrum

Department of Surgical Oncology - Image Guided Surgery
Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital
May 2024

Abstract

Introduction

Prostate cancer is the most commonly diagnosed cancer in the Netherlands. Accurate assessment of the prostate volume (PV) is a crucial step in prostate cancer (PCa) screening and risk-stratification. In standard clinical care, the PV is obtained by measuring the prostate dimensions with the aid of transrectal ultrasound (TRUS). However, rectal examination is characterised with patient discomfort, for which the feasibility of transabdominal ultrasound (TAUS) is explored, as a more accessible and patient-friendly alternative. However, manual PV measurements are prone to inter-observer variability and require operator training. This study aims to improve the accessibility, complexity, and robustness of PV measurements by developing a framework to automatically estimate the PV based on TAUS acquisitions. The primary components of the framework comprise two deep neural networks, developed for prostate segmentation on axial and sagittal TAUS images, and an algorithm that extracts the prostate's diameters on which the PV is calculated.

Materials and Methods

During this study, a new prostate dataset is developed, comprising sagittal and transverse TAUS image acquisitions of 100 participants, and reference PV measurements based on TRUS and MRI are collected. First, the feasibility of TAUS for manual PV estimation is explored, and the inter-method agreement between TAUS, TRUS and MRI is analysed in Bland Altman diagrams. Additionally, all TAUS acquisitions are assessed on image quality. Secondly, three deep neural networks (using the nnU-Net framework) are developed to segment the prostate on sagittal and/or axial TAUS images. All models, are trained and validated on TAUS image data of 52 participants. Additionally, an algorithm is designed to predict the prostate diameters when prostate segmentations serve as input. To this extend, the PV is estimated according to the widely used Ellipsoid formula. The proposed algorithm is evaluated on input ground-truth segmentations of 42 participants. Essentially, the segmentation models combined with the proposed algorithm result in a framework to automatically estimate the PV on TAUS. Finally, it is tested on unseen TAUS acquisitions of 17 participants, whereby the predicted PV is compared to reference PV measurements on MRI.

Results

Our results show an average volume difference of 3.0 +- 17.6 ml when manual PV estimation on TAUS is compared to MRI. When manual PV estimation on TRUS is compared to MRI, an average volume difference of 12.3 +- 18.8 ml is obtained. The developed segmentations models segment the middle region of the prostate on TAUS with an average DSC = 0.91 +- 0.06 and DSC = 0.83 +- 0.09 for axial and sagittal TAUS images respectively. When the entire prostate region was evaluated, a lower model performance was observed, whereby the prostate was segmented with a DSC of 0.76 +- 0.09 in the axial imaging-plane and DSC of 0.68 +- 0.21 in the sagittal imaging-plane. The algorithm for automatic diameter extraction showed good correspondence with manually assigned prostate diameters on TAUS. When the segmentation models and the algorithm are utilised for

automatic PV estimation, an average volume difference of 2.5 +- 10.2 ml was observed, compared to MRI reference volumes. Ultimately the PV was predicted with a volume difference $< 25\%$ compared to MRI in 14 out of 17 test cases.

Discussion

The results of this study show that it is possible to obtain PV measurements using TAUS that are comparable to those obtained with MRI. Moreover, the variability related to PV measurements using TAUS seem unrelated to TAUS image quality, indicating that manual PV measurements can be performed, even when unfavorable patient characteristics limit the image quality. Still, it is important to note that proper operator training for TAUS examination is essential. The proposed framework for automatic PV estimation on TAUS acquisitions shows good correspondence with MRI reference volumes. Thus expanding the possibilities of PCa risk-stratification, whereby robust, and straightforward PV estimations are desired. In order to adopt the framework for standard clinical care, further research is required on a larger cohort to investigate the generalizability of the framework and ensure reliable results on all future patients.

Acknowledgments

Dear reader,

I would like to take this opportunity to express my gratitude to everyone who has contributed to the completion of this master thesis and supported me during the project.

First, I want to thank my supervisors and colleagues who generously shared their knowledge, time and perspectives. The warm ambiance at the *Group Ruers* research group has motivated me to perform best throughout the entire project. Tiziano, I am very grateful for your guidance and the way you provide feedback. I deeply enjoy our conversations about ultrasound, scientific writing, pythonic life hacks, music and mountains. Behdad, thank you for always providing your honest opinion about my work, and recommendations to improve the project. Matteo, thank you for always finding some time for small discussions, so my thoughts about the project were not a moment alone.

Additionally, I want to thank the Urology Department / Center for Early Diagnostics in the Antoni van Leeuwenhoek Hospital. As a technical medicine student, it is not always ensured that your ideas on innovation are welcomed. Pim, I want to thank you for all your positive feedback throughout the entire project regarding the possibilities of an automatic prostate volume measurement. It was a pleasure to feel so welcomed at the Center of Early Diagnostics every Tuesday and Thursday and I am grateful for all the clinicians who helped me becoming an expert in transabdominal ultrasound examination for prostate evaluation. Hopefully, this manuscript can be seen as a foundation to return the favor to future clinicians.

In my opinion, the high level of scientific expertise in combination with the friendly atmosphere at the NKI-AVL, is a motivation for every (technical medicine) student to pursuit an internship at a hospital like this one.

I would also like to extend my gratitude to my technical supervisor Jifke, who has always been keen to brainstorm about the project. Moreover, the courses you provided during the bachelor and the master, of Technical Medicine have been the key inspiration to take on a project like this.

At last, I am grateful for my family and friends, who have always been there for me over the past years, whether it was about sharing ideas, moral support or having fun. Thank you for your support and encouragement from the first day at the TU Delft.

Enjoy Reading,
Liza

Contents

Abstract i

Acknowledgments ii

List of Acronyms vi

List of Figures vii

List of Tables ix

1 Introduction 1
- 1.1 Introduction . . . 1
- 1.2 Relevant Work . . . 4
  - 1.2.1 Standard clinical care of prostate volume measurements to support PCa risk stratification . . . 4
  - 1.2.2 Transabdominal ultrasound as an alternative imaging modality for prostate volume estimation . . . 4
  - 1.2.3 Deep-learning for automatic prostate segmentation . . . 5
  - 1.2.4 Current techniques for automatic prostate volume estimation . . . 6

2 Materials and Methods 9
- 2.1 Study overview . . . 9
- 2.2 Cohort description and data collection . . . 11
- 2.3 Phase 1: A feasibility study on prostate volume estimation using TAUS . . . 11
  - 2.3.1 Inter-method agreement between TAUS, TRUS, and MRI . . . 11
- 2.4 Phase 2: Deep neural networks for prostate segmentation . . . 12
  - 2.4.1 Image quality assessment . . . 12
  - 2.4.2 Development of prostate segmentation models . . . 13
  - 2.4.3 Model evaluation . . . 14
- 2.5 Proposed Algorithm to extract the prostate diameters . . . 14
- 2.6 Phase 3: Automatic prostate volume estimation using TAUS . . . 15

3 Results 17
- 3.1 Dataset characteristics . . . 17
- 3.2 Phase 1: Prostate volume estimation inter-method agreement: TAUS, TRUS, and MRI . . . 18
  - 3.2.1 PV estimation based on Ellipsoid formula . . . 18
  - 3.2.2 PV estimation based on Axial formula . . . 18
- 3.3 Phase 2: Deep neural networks for prostate segmentation . . . 18
  - 3.3.1 Evaluation Model Sagittal and Model Axial . . . 18
  - 3.3.2 Evaluation Model Sag-Ax . . . 20
  - 3.3.3 Inter-observer variability of manual prostate segmentation on transabdominal ultrasound . . . 21
- 3.4 Evaluation Algorithm to extract the prostate diameters . . . 22
- 3.5 Phase 3: Automatic prostate volume estimation on transabdominal ultrasound . . . 23

4 Discussion **27**
4.1 Prostate volume estimation inter-method agreement: TAUS, TRUS, MRI . . . 27
4.1.1 PV estimation based on the Ellipsoid Formula . . . 27
4.1.2 PV estimation based on Axial formula . . . 29
4.1.3 TAUS vs TRUS vs MRI: Wilcoxon Rank Test . . . 29
4.2 Deep neural networks for prostate segmentation . . . 29
4.2.1 Model Axial, Model Sagittal and Model Sag-Ax . . . 29
4.3 Automatic extraction of the prostate diameters . . . 31
4.4 Automatic prostate volume estimation on transabdominal ultrasound . . . 32
4.4.1 Clinical application . . . 33
4.4.2 Future work . . . 33

Conclusions **35**

A Prostate Imaging-Reporting and Data System **37**

B Quality assessment TAUS acquisitions **39**

C The Application of deep-learning for prostate volume estimation on ultrasound: A Literature Review **41**

Bibliography **43**

List of Acronyms

1. TAUS: transabdominal ultrasoud
2. TRUS: transrectal ultrasoud
3. PCa: prostate cancer
4. PV: prostate volume
5. DRE: digital rectal examination
6. PSA: prostate-specific antigen
7. DL: deep learning
8. PI-RADS: prostate imaging-reporting and data system
9. mpMRI: multi parametric magnetic resonance imaging
10. CT: computed tomography
11. RPCRC: Rotterdam prostate cancer risk calculator
12. AP: anteriorposterior
13. MAVD: mean absolute value difference
14. LOA: limit of agreement
15. SD: standard deviation
16. PPV: presentation of prostate volume
17. VPB: visibility of prostate boundary
18. AIP: image-artefacts interrupting the prostate
19. DICOM: digital imaging and communication in medicine
20. DSC: dice similarity coefficient
21. DSC-TP: average dice similarity coefficient across correctly classified acquisition images
22. DSC mid-plane: dice similarity coefficient on the prostate image, in mid-plane of the acquisition
23. HD mid-plane: maximum haussdorff distance on the prostate imagine, in mid-plane of the acquisition

List of Figures

1.1 Prostate examination using ultrasound: transrectal ultrasound examination and transabdominal ultrasound examination. 2
1.2 The prostate on transabdominal ultrasound: a variety in image quality. 3

2.1 Overview of the project. 10
2.2 The prostate diameters on transabdominal ultrasound. 12
2.3 Three examples of manual prostate segmentation on TAUS 13
2.4 Proposed algorithm to extract the transverse, anterior-posterior and longitudinal prostate diameter. 15

3.1 Bland Altman diagram: the comparison of TAUS, TRUS, and MRI for prostate volume estimation. 19
3.2 The evaluation of *Model Axial* and *Model Sagittal* in a four-fold cross-validation. . 19
3.3 The performance of *Model Sag-Ax*, compared to *Model Axial* and *Model Sagittal*. . 20
3.4 Comparison of prostate segmentations on sagittal and axial TAUS images by *Model Sag-Ax*, *Model Sagittal*, and *Model Axial* 21
3.5 The variance between computed prostate diameters and manually assigned diameters on transabdominal ultrasound. 23
3.6 Bland Altman diagram: the comparison between PV measurements based on computed diameters and the manually assigned diameters 23
3.7 Bland Altman diagram: the comparison between predicted PV measurements and reference volume measurements based on MRI. 24
3.8 A comparison of prostate volume predictions with reference volume measurements on MRI 24
3.9 Examples of prostate segmentations on axial TAUS images. 25
3.10 Examples of prostate segmentations on sagittal TAUS images 26

B.1 The prostate on transabdominal ultrasound images with diverse image quality. . . 40
B.2 The prostate on transabdominal ultrasound: a variety in image quality. 40

List of Tables

3.1 Dataset characteristics 17
3.2 Performance of developed models for prostate segmentation on transabdominal ultrasound 22

1

Introduction

1.1 Introduction

Prostate cancer (PCa) is the most commonly diagnosed cancer in men in the Netherlands [1]. The high incidence of prostate cancer led to the implementation of early detection strategies to promote disease control by identifying PCa in an earlier and less aggressive stage. In the past years, several nomograms have been developed, aiming to stratify the risk of having PCa based on clinically predictive parameters [2]–[5]. In short, these comprise an elevated Prostate-Specific Antigen (PSA) blood level, abnormal digital rectal examination (DRE) outcome, age above 50, the occurrence of PCa in family history, carriers of the BRCA2 mutation, and African race [5]–[8]. Since PSA density (PSA level divided by prostate volume (PV)) significantly improves the accuracy of PCa risk stratification, the PV is routinely obtained in the first stage of diagnostics [9], [10]. Aside from prostate malignancy, the PV is a crucial parameter for the diagnosis of benign prostate diseases such as lower urinary tract symptoms and benign prostatic hyperplasia, affecting one-third of all men [11], [12].

When prostate-related disease is suspected, the PV is determined by measuring the dimensions of the prostate during transrectal ultrasound (TRUS) examination (Figure 1.1a). Even though PV estimation based on TRUS is recognized as standard clinical care, a rectal examination is primarily characterised by patient discomfort, pain, possible tearing of the perianal skin, and infection that particularly occur when the internal or external anal sphincter muscles are tensed [13]. These factors contribute to the complex nature of the procedure and as a result, it is solely performed by expert clinicians. Additionally, cultural and social beliefs play a role in the decreased accessibility of TRUS examination. For instance, several patients cope with fears toward rectal examination, believing that their masculinity and sexuality might be affected by it [14], [15].

An approximation of PV based on transabdominal ultrasound (TAUS) is another option to obtain PSA density, and is often preferred since it is a percutaneous and comfortable approach (Figure 1.1b) [16], [17]. Additionally, it enables the implementation of a PV measurement at an earlier stage of diagnostics. According to a recent study by De Vos *et al.* (2023), a PSA density

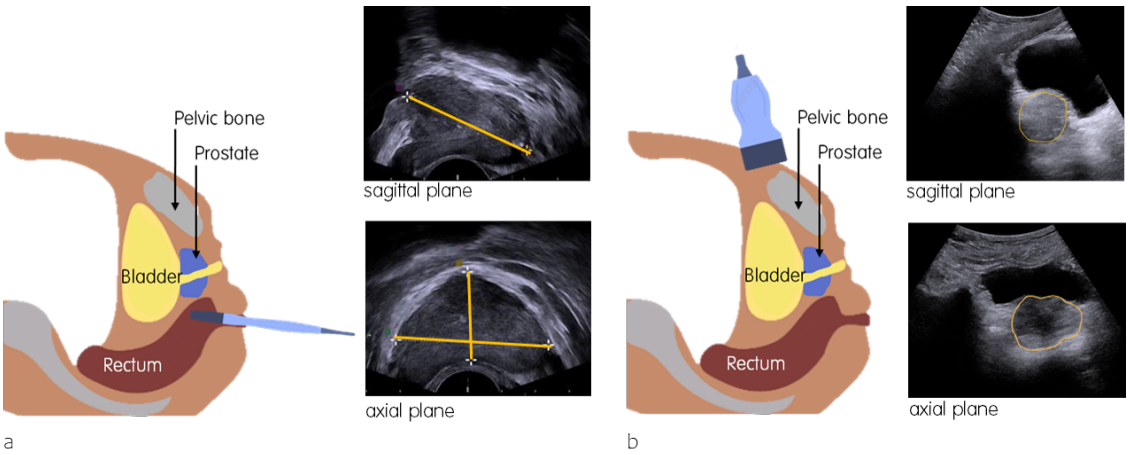


Figure 1.1: Prostate examination using ultrasound. **a:** Transrectal Ultrasound. Diameters of the prostate are assigned with yellow lines. **b:** Transabdominal ultrasound. The prostate is delineated manually with a yellow contour.

measurement based on TAUS at the general practitioner allows for an improved risk-stratification workflow and may reduce unnecessary hospital visits [18].

In practice, it remains challenging to measure the PV manually due to the difficulties associated with the interpretation of ultrasound images, whereby TAUS is considered more prone to inaccuracies compared to TRUS [12]. Manual prostate segmentation, diameter measurements, and annotation of the prostate's outer boundary landmarks are the frequently performed tasks to obtain the PV, and often result in over/under estimations and are susceptible to inter/intra-observer variability [19]. This is primarily the result of poor contrast between the prostate and surrounding tissue, resulting in ambiguous prostate boundaries (Figure B.2c). Additionally, the assessment of ultrasound is complicated by misleading imaging artefacts such as echoic shadows and the presence of calcifications. Notably in TAUS, the prostate can be difficult to observe when shadow artefacts disrupt the prostate boundary as a result of ultrasonic reflection by the pelvic one (Figure B.2d) [20], [21]. Additionally, the ultrasound signal is directly attenuated by abdominal fat which limits the image quality of the prostate in patients with higher (abdominal) fat percentages [22]. The inhomogeneous intensity distribution of the prostate tissue, the presence of other anatomic structures and the large shape variations complicate the assessment and are often considered most challenging in the apex (lower part of the prostate) and base (upper part of the prostate). Figure B.2 provides four TAUS images to illustrate the variety in image quality and occurring artefacts. Finally, prostate annotations are generally conducted during the ultrasound examination. This enhances the complexity of the procedure, since the operator is responsible for the determination of the appropriate imaging-plane and the diameter assessment, while the patient is being examined [23]. These factors limit the implementation of TAUS in a clinical setting and therefore an automatic and straightforward method to obtain the PV is desired.

In recent years, deep-learning (DL) has been utilised for medical imaging analysis to enhance the accuracy, reliability and repeatability of medical image tasks [24]. Specifically, with the development of convolutional neural networks that process the entire image as part of their training, segmentation models enhanced the readability of computed tomography (CT), TRUS, and magnetic resonance imaging (MRI) acquisitions when it comes to defining the prostate borders and volume calculation. [25], [26]. Currently, automatic prostate segmentation techniques have only been studied on CT, TRUS and MRI. Yet, prostate segmentation on TAUS is worth to explore,

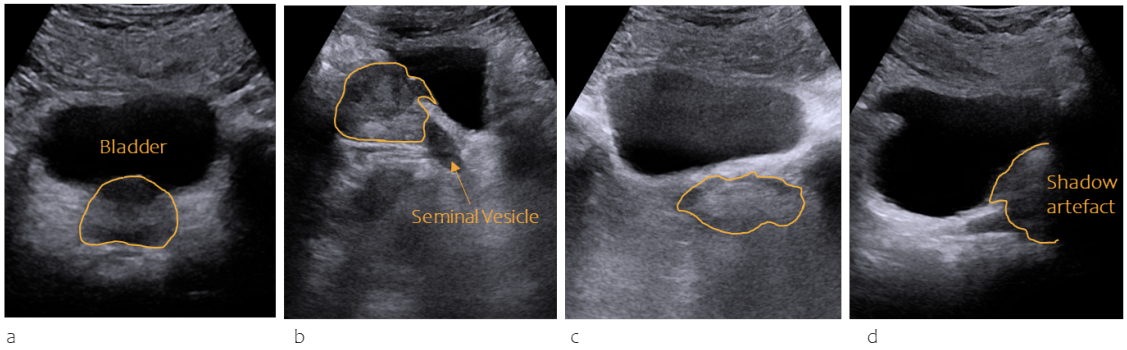


Figure 1.2: Various presentations of the prostate on transabdominal ultrasound. The prostate is delineated manually with a yellow line. **a:** The axial imaging-plane, clear visibility of the prostate. The bladder is annotated. **b:** The sagittal imaging-plane, clear visibility of the prostate with bulging contours that interrupt the bladder. The seminal vesicle is annotated. **c:** The axial imaging-plane, poor visibility of the prostate due to ambiguous boundaries. **d:** The sagittal imaging-plane, part of the prostate boundary is invisible due to interrupting acoustic shadow artefacts.

as a delineation of the prostate can be used to extract the prostate dimensions to enable PV calculation.

The primary aim of this study is to investigate the feasibility of a method to automatically estimate the PV based on TAUS examination and ease the process of PV measurements. With the aid of deep neural networks, the prostate is segmented on TAUS. Then, the predicted segmentation masks serve as input for an algorithm that extracts the prostate diameters to calculate the PV according to the Ellipsoid formula. The feasibility of the framework is explored on a novel dataset that was acquired during this study, including TAUS examinations of 100 participants along with corresponding reference PV measurements based on TRUS and MRI. The overall contribution is summarized as follows:

1. During this study, a new prostate dataset is formed, comprising sagittal and transverse TAUS image acquisitions of 100 participants. For all included participants, the PV is calculated based on TAUS, and corresponding PV measurements based on TRUS and MRI are collected retrospectively. This study analyses the inter-method agreement for PV estimation on all three modalities available in the dataset, namely TAUS, TRUS and MRI.

2. A TAUS image quality assessment is performed to qualify the data regarding prostate visibility and imaging artefacts. One expert manually segmented the prostate in TAUS images of 61 participants.

3. Three deep-learning models are developed to segment the prostate on TAUS acquisitions of 52 participants. The first model (*Model Sagittal*) is trained and validated on sagittal acquisitions, while the second model (*Model Axial*) is trained and validated on axial acquisitions. A third model, *Model Sag-Ax* is developed on all image data to segment the prostate regardless of the imaging-plane. All models are evaluated on: (i) dice similarity coefficient (DSC), (ii) DSC on correctly classified images and (iii) the performance of the models is explored on the mid-prostate region, as this part of the prostate is crucial for the extraction of the prostate diameters.

4. An algorithm is designed to extract the prostate diameters based on input TAUS acquisitions in which the prostate is segmented, enabling PV estimation according to the Ellipsoid

formula. The algorithm is tested on manually delineated prostate segmentations. Then, the computed diameters are compared to manually assigned prostate diameters on TAUS.

5. Finally, unseen test data is utilised to evaluate the entire framework, whereby the predicted PV on TAUS is compared to manual PV estimations based on MRI.

1.2 Relevant Work

1.2.1 Standard clinical care of prostate volume measurements to support PCa risk stratification

According to Dutch health care guidelines, patients with suspected prostate disease are referred to a urologist for further examination when they present an elevated PSA level (> 3.0 ng/ml), unexplainable bone-related complaints, an abnormal DRE outcome or when further screening is requested by the patient [1]. A PSA density exceeding 0.15 ng/ml/cc suggests the need for prostate biopsy, the gold standard procedure for diagnosing PCa [27]. The necessity for prostate biopsy is also be determined based on the outcome of multi-parametric magnetic resonance imaging (mpMRI) scans of the prostate, but the implementation of mpMRI is not standardized in all European clinical workflows. When mpMRI is utilised, targeted prostate biopsy is recommended when a patient scores ≥3 on the Prostate Imaging-Reporting and Data System (PI-RADS) [1], [4], [8]. Appendix A provides a detailed explanation of the PI-RADS. In addition, the Rotterdam Prostate Cancer Risk Calculators (RPCRCs) are utilised to prevent unnecessary prostate biopsies by calculating the risk of having clinically significant PCa. An estimation of the PV is one of the key predictors that is incorporated in such nomogram and therefore required at this stage of diagnostics [10], [28].

In clinical practice, TRUS or MRI are utilized to estimate the PV, and the PV extracted from the segmentation masks of the entire prostate gland on MRI is considered the most accurate [29]. Nevertheless, TRUS is frequently employed in clinical settings due to its cost-effectiveness, portability and ability to rapidly estimate the PV with the aid of the Ellipsoid formula [17], [30]. During TRUS examination, an ultrasound probe is inserted in the rectum to view the prostate in the sagittal and axial imaging-plane. (Figure 1.1a) In both image directions, the physician freezes the mid-plane of the prostate whereby the PV is computed by measuring the longitudinal, anterior-posterior (AP), and transverse diameters of the prostate gland and multiplying the product by a coefficient $c = \pi/6$ (Equation 1.1) [13], [19], [23]. In clinical settings that allow for prostate evaluation based on mpMRI, the PV is frequently measured using the same formula due to the rapid nature of this approach [17].

$$PV_{Ellipsoid} = D_{Transverse} \times D_{AnteriorPosterior} \times D_{Longitudinal} \times \pi/6 \quad (1.1)$$

1.2.2 Transabdominal ultrasound as an alternative imaging modality for prostate volume estimation

When the Ellipsoid formula is performed on TAUS images, a high degree of correlation and agreement with PV estimation based on TRUS ($\rho = 0.958$, P≤ 0.01) and MRI ($\rho = 0.914$, P≤ 0.01) is shown [17], [31]. Previous investigations on the inter-method agreement of TAUS, TRUS and MRI have yielded divergent results when the different imaging modalities were compared. For instance

when TAUS was compared to TRUS, Kim *et al.* (2008) reported an average volume difference of 8.4 ±10.5 ml, indicating that PV measurements based on TAUS underestimate PV measurements based on TRUS [32]. In contrast, recent literature reported an average volume difference of -9.9 ±1.99 ml when the inter-method agreement of TAUS and TRUS was investigated [18]. An average volume difference of -4.1 ±11.0 ml and 8.06 ±7.9 ml was reported when MRI was compared to TAUS [17], [33]. Additionally, it has been suggested that the inter-method agreement is influenced differently when smaller or larger PVs are analysed [17]. Currently, there is no exploration on the inter-method agreement of manual PV estimation when TAUS, TRUS and MRI are analysed using the same cohort group.

1.2.3 Deep-learning for automatic prostate segmentation

In recent years, computerized techniques to approximate the PV have been explored, aiming to promote the consistency, efficiency and precision of PV measurements. Predominately, prostate segmentation and boundary extraction methods have been developed to reconstruct the PV on TRUS and MRI. Former research focused on prostate segmentation methods in which deformable models, edge-based segmentation methods and region-based segmentation methods are used and are discussed in several overviews on prostate segmentation [20], [26]. However, all aforementioned methodologies predominantly require the extraction of prostate-related image features such as the relative pixel intensity or prostate shape information. Especially when used in ultrasound, these features are affected by the large contrast diversity and ultrasound artefacts resulting in inaccurate outcomes.

On the contrary, the main advantage of DL is its ability to automatically extract and use multi-level features that consist of abundant semantic as well as detailed image information [34]. As a result, prostate segmentation outcomes on TRUS images have become more accurate and useful over the years. Research in this field mainly focused on the implementation of a standard U-Net and novel ideas have been introduced to advance the performance of deep neural networks on the more challenging ultrasound images.

For instance, the deep neural network designed by Xu *et al.* (2022) outperformed other segmentation techniques by implementing two novel mechanisms to encourage the network to cope with shadow artefacts on image and feature level [35]. More specifically, shadow artefacts were artificially added in training images to enrich the shadow diversity of the utilized dataset. This procedure served as an innovative data augmentation strategy aiming to increase the robustness of the network when shadow artefacts disrupt the prostate boundary. Furthermore, the shadow features were subtracted from the created feature maps. In this way, the model was encouraged to learn the prostate boundary using the remaining shadow-free features.

In the application of Karimi *et al.* (2019), prostate segmentation outcomes on TRUS were enhanced by matching a MRI based statistical shape model with the predicted segmentations, aiming to improve the apex and base region of the prostate segmentations[36].

In the framework proposed by Beitone *et al.* (2022), image input from all imaging-planes is combined to improve segmentation outcomes [37]. The employment of multi-directional images can be of additional value, as the visibility of the prostate varies locally. For instance, the apex and base of the prostate are characterized with ambiguous boundaries on axial US images, while it appears more clearly in the sagittal and coronal imaging-plane. The authors designed a framework in which three 2D U-Nets were trained on axial, sagittal, and coronal TRUS images to produce view-specific segmentation volumes of the prostate. Then, the resulting volumes were used to generate view-specific confident maps upon which the final segmentation was created.

To the best of our knowledge, there is no available literature concerning automatic prostate segmentation on TAUS images. This can be partly explained by the existing variations in TAUS image quality, an aspect that commonly affects the accuracy of prostate segmentation methods [17]. Aside from frequently occurring ultrasound artefacts that challenge the prostate evaluation, TAUS acquisitions may deviate according to the clinician's examination technique. Additionally, patient characteristics influence the image quality. For instance, a high abdominal fat percentage results in poor image resolution and the prostate boundary can be more difficult to observe when the bladder is empty. Additionally, TAUS examination is not embedded in standard clinical care which results in a limited availability of image data, required for the development and validation of automatic segmentation techniques. Nonetheless, automatic prostate segmentation on TAUS could benefice the challenges that occur during prostate examination with TAUS. For instance, medical physicians find difficulty in locating the prostate within the image. Furthermore, automatic prostate segmentation is a crucial step in image registration with other imaging modalities, that may be useful when TAUS is more suitable than TRUS in an operation environment [17]. Finally, automatic prostate segmentation on TAUS can be used to extract the prostate dimensions in the axial and sagittal imaging-planes to estimate the PV in a less complex and robust manner.

1.2.4 Current techniques for automatic prostate volume estimation

At present, solely Albayrak *et al.* (2022) focused on an automatic PV measurement of the prostate using TAUS [33]. The authors automated the workflow of the Ellipsoid formula by establishing a classification method that predicts six outer boundary landmarks to obtain the transverse, AP and longitudinal diameter of the prostate. In their work, two deep neural networks were trained separately for the axial and sagittal imaging-planes. One network focused on 2D axial TAUS images to predict four landmarks related to the transverse and AP diameter, while another network was trained on sagittal TAUS images to identify the remaining two landmarks for the prostate's longitudinal diameter. The plane-specific networks comprised an ensemble of four ResNet-18 CNNs (QDCNN) that were individually trained on image patches that contain the prostate in four image-scales. The aim of both models was to classify the distance and orientation relative to the ground truth landmark given a random point in the training image. Then, the location of each landmark was determined by producing a landmark-specific voting map of all predicted distance and orientation outcomes. Ultimately, the landmarks served as input to utilize the Ellipsoid formula to obtain the PV. The proposed method was validated by comparing the Mean Absolute Value Difference (MAVD) between the predicted PV and a manually estimated PV on TAUS. An average MAVD value of 4.95 ml was obtained, which is smaller than the average inter-expert MAVD value of 5,09 ml for manual PV estimation on TAUS. Additionally, the predicted PV was compared to a PV estimation based on MRI that resulted in an average MAVD value of 6.22 ml. For comparison, the inter-expert average MAVD value between TAUS and MRI was 8.06 ml. (All outcomes were proposed by the research group).

Even though the presented results seem rewarding, there are considerable aspects of their proposed system that warrant discussion. First, the Ellipsoid formula can only be utilized on the mid-axial and mid-sagittal image, where the cross-section of the prostate is at maximum. Currently, the authors only trained the network with TAUS images that satisfy this criterion. As a result, user interaction and expertise is still required when the proposed method would be implemented in a clinical workflow. In our proposed algorithm, the determination of the most suitable prostate slice is implemented to measure the PV automatically as accurate as possible. Secondly, the authors evaluated the system on MAVD of the entire PV, while the outcome relies on

the prediction of six landmarks. Hence, it remains unclear whether the accuracy of the calculated landmarks are contributing equally to the predicted MAVD. During our study, all parameters that contribute to the final prediction of the PV are evaluated to understand which factors possibly result in inaccuracies.

2

Materials and Methods

2.1 Study overview

This research is structured into three phases, each targeting specific aspects of PV estimation to develop and validate a comprehensive framework for automatic PV estimation using TAUS imaging, deep neural networks and algorithms. An overview for the three phases of this study is provided in Figure 2.1.

Phase 1 involved exploring the feasibility of estimating the PV using TAUS. A new dataset was set up, comprising TAUS image acquisitions of 100 participants. For all included participants, the PV was measured manually on TAUS images with the aid of the Ellipsoid formula and corresponding PV measurements based on TRUS and MRI were retrieved from the hospital's electronic database. Then, the inter-method agreement of PV estimation using the different imaging modalities was analysed in Bland Altman diagrams and a Wilcoxon Rank test.

Automatic prostate segmentation and prostate diameters extraction on TAUS images is the main focus of Phase 2. In this phase, multiple experiments were conducted to investigate the application of deep neural networks to segment the prostate on TAUS images. First, all acquisitions were evaluated in a quality assessment on prostate visibility, and resulted in the exclusion of acquisitions with poor image quality. One expert manually segmented the prostate in acquisitions of 61 participants to obtain the ground truth labels required for training, validation and testing of all developed deep neural networks. Subsequently, an algorithm was designed to extract the prostate diameters automatically when input prostate segmentations are provided. The proposed algorithm was evaluated by comparing computed prostate diameters with manually assigned prostate diameters on TAUS on the examinations from 42 participants. The combination of deep neural networks to segment the prostate and an algorithm to extract the prostate diameters results in a novel method for automatic PV estimation.

In Phase 3, an entire framework for automatic prostate volume estimation based on the methods proposed in phase 2, was tested on independent and unseen image data of 17 participants. The predicted PV outcomes were compared to corresponding PV measurements on MRI for each participant.

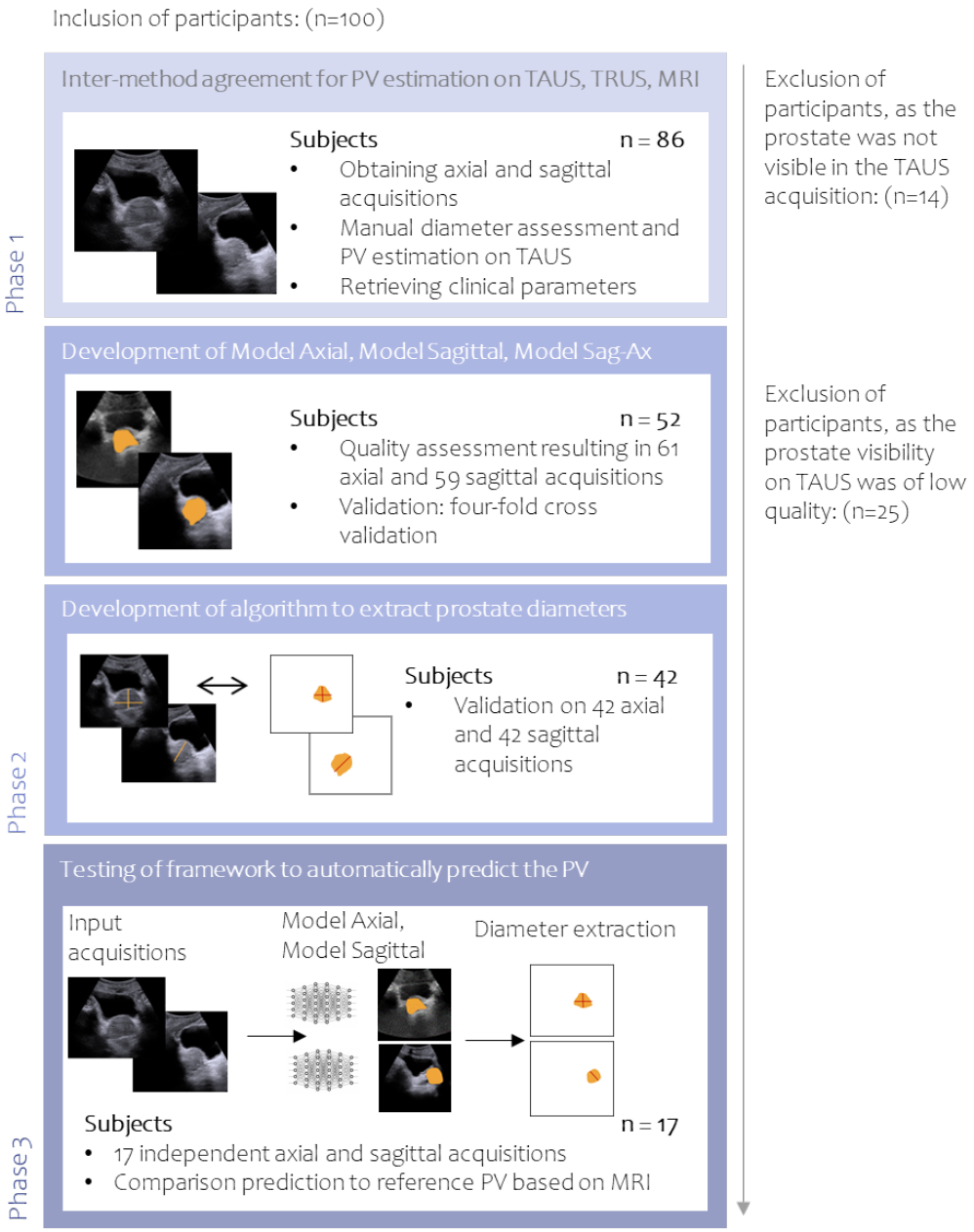


Figure 2.1: Project overview. **Phase 1:** Data of 86 participants is used for analysing the inter-method agreement between manual PV measurement using TAUS, TRUS and MRI. **Phase 2:** Data of 52 participants is used for the development of deep neural networks to segment the prostate on TAUS and data of 42 participants is used to validate the algorithm to extract the prostate diameters. **Phase 3:** The methods developed in Phase 2 result in an framework for automatic prostate volume (PV) estimation on TAUS, and is tested on image-data of 17 participants. Finally, the predicted PV is compared to reference PV measurements based on MRI.

2.2 Cohort description and data collection

To explore the feasibility of TAUS for PV estimation, a cohort study was performed at the Center for Early Cancer Diagnostics, Antoni van Leeuwenhoek Hospital in the Netherlands between August 2023 and February 2024. All men who underwent early diagnostic screening for PCa (including DRE, TRUS examination and mpMRI), were offered to participate in the study and received oral and written information on TAUS examination as well as information on study participation. After the process of shared decision-making, written informed consent was provided by each participant. Subjects who did not speak the Dutch language were excluded from this study. This study was approved by the Institutional Review Board of Antoni van Leeuwenhoek Hospital (IRBd22-319).

A new dataset was formed, comprising TAUS acquisitions of 100 participants in the sagittal and axial imaging-planes. Each acquisition exists of a sequence of 2D TAUS images such that the apex and base of the prostate were covered maximally. For each participant, at least one axial and one sagittal acquisition was conducted and in a few cases, additional acquisitions were carried out to increase the number of images in the dataset. All acquisitions were obtained with a BK Medical 3000 system (BK Medical, Herlev, Denmark) equipped with a 6C2 (9040) Curved Array Transducer[38]. TAUS examination was carried out by one person educated in Technical Medicine (LMK) under the guidance of an experienced urologist (PVL), or by the urologist (PVL).

2.3 Phase 1: A feasibility study on prostate volume estimation using TAUS

For each participant, the AP, transverse and longitudinal diameters of the prostate were manually assigned to enable PV estimation on TAUS according to the Ellipsoid formula (equation 1.1). The maximum transverse diameter was assessed on the mid-axial image and the AP diameter was measured perpendicularly between the anterior and posterior prostate boundary (Figure 2.2a). The longitudinal diameter was assessed on the mid-sagittal image and measured from the bladder to the prostate apex. (Figure 2.2b) When the prostate boundary in the apex region was disrupted due to shadow artefacts, the longitudinal diameter was estimated based on visual inspection. In the case of bulging prostate contours or when the prostate interrupted the bladder, the entire prostate area was included in the diameter assessment. Reference measurements of the PV based on TRUS and MRI were retrieved retrospectively for each participant from the hospital's electronic database. For both imaging modalities, the PV estimations were performed by an expert urologist/radiologist, utilizing the Ellipsoid formula according to clinical standard. The TAUS, TRUS, and MRI examinations were always conducted on the same day. To complete the dataset, clinical information was obtained on age, PSA density based on MRI, and PI-RADS score.

2.3.1 Inter-method agreement between TAUS, TRUS, and MRI

First, the inter-method agreement of manual PV estimation based on TAUS, TRUS, and MRI is analyzed using Bland Altman diagrams and a Wilcoxon Rank Test. Each imaging modality pair is tested at a significance level of $alpha = 0.05$, that was adjusted using the Bonferroni Correction to compensate for multiple comparisons, resulting in $alpha = 0.017$ (0.05/3). Statistical significance was considered when $P < alpha$ was verified. A Bland Altman diagram visualises the difference between PV measurements, obtained by two imaging modalities, on the vertical axis against the mean of the pair on the horizontal axis. The upper and lower limits of agreement (LoA) visualize

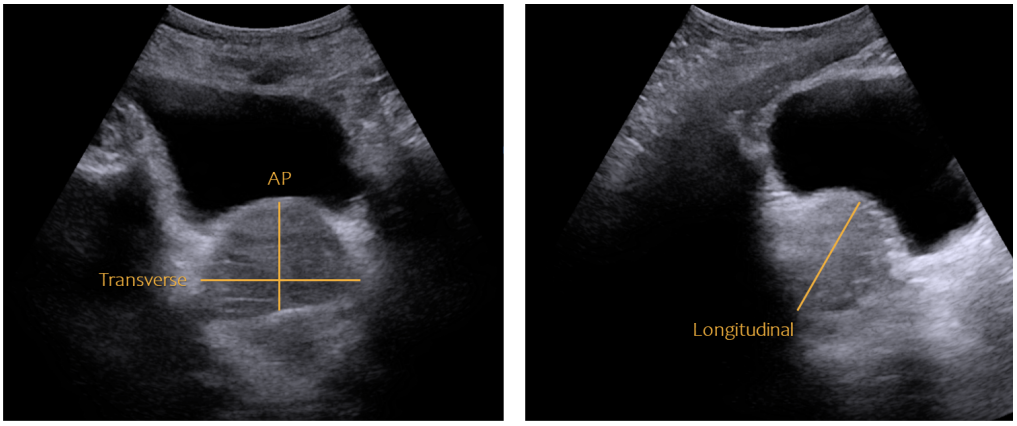


Figure 2.2: The prostate on transabdominal ultrasound (TAUS). **a:** Mid-axial TAUS image depicting the transverse diameter and the anterior-posterior (AP) diameter in yellow. **b:** Mid-sagittal TAUS image depicting the longitudinal diameter in yellow.

the 95% confidence interval of the included measurements [39].

Subsequently, the inter-dependence of prostate diameters in the axial and sagittal imaging-planes was investigated by assessing the correlation between the transverse and the longitudinal diameter using a Spearman Correlation test. The correlation between the AP diameter and the longitudinal diameter was also determined. A strong relation between diameters was considered when correlation coefficient $\rho > 0.75$ and statistically significant when $P < 0.05/2$ [40].

The longitudinal diameter is considered more prone to errors when assessed on the sagittal imaging-plane [32]. Therefore, a novel method to estimate the PV was proposed to investigate whether image information of only the axial imaging-plane is sufficient for a PV estimation on TAUS. In the axial-plane method (PV_{Axial}, Equation2.1), the AP diameter acts as a surrogate for the longitudinal diameter to calculate the volume of the hypothetical prostate shape, in a less complex manner. Hence, for each participant, the PV was calculated according to PV_{Axial}, and compared to PV measurements based on MRI in a Bland-Altman diagram. All statistical analysis was performed using statistical functions present in the SciPy library (version 1.13.0) implemented in Python (version 3.9), or Microsoft Excel (version 2016) **Python**, [41], [42].

$$PV_{Axial} = \quad D_{Transverse} \quad \times \quad D_{AnteriorPosterior} \quad ^2 \times \quad \pi/6 \tag{2.1}$$

2.4 Phase 2: Deep neural networks for prostate segmentation

2.4.1 Image quality assessment

To provide insight into the existing variety of image quality in the dataset, all TAUS acquisitions were assessed and scored by one expert (LMK) on the presentation of the volume (PPV), ranging from 0%. 25%, 50%, 75% and 100%, the visibility of the prostate boundary (VPB) (A, B, C) and artefacts interrupting the prostate (AIP) that contain notes regarding occurring image-artefacts. Elaborate details regarding the quality assessment are covered in Appendix B.

The image data used for the development of all neural networks comprised solely quality A/B acquisitions in which the prostate is visible for at least 75%. In other words, the apex, base

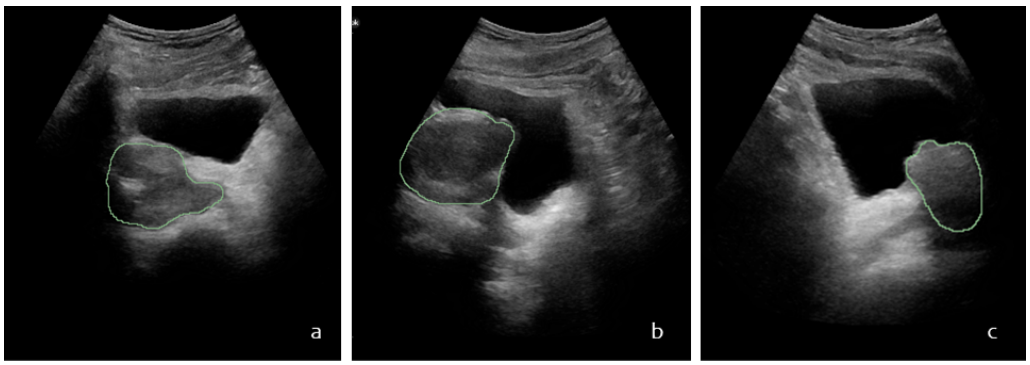


Figure 2.3: Three examples of manual prostate segmentation on sagittal TAUS images. **a:** A manual segmentation where a small shadow region is included as prostate tissue. **b:** A manual segmentation when no shadow artefact is interrupting the prostate boundary. **c:** A manual segmentation where the prostate boundary is partly estimated as shadow artefacts interrupt the prostate boundary.

and middle section of the prostate were visible in the image sequence. After excluding poor-quality acquisitions, 139 acquisitions of 61 participants were regarded as suitable for training, validation and testing. For each acquisition sequence in the training dataset, only one image in every five frames was included in the training set in order to prevent the model from overfitting on similar image data. The dataset used for training and validation contains 61 axial and 59 sagittal acquisitions (1599 sagittal and 1764 axial images of 52 participants). The test dataset contains 17 axial and 17 sagittal acquisitions (1333 sagittal and 1436 axial images of 17 participants). Each image was 1016×936 pixels in size with a pixel size of 0.25×0.25 mm.

First, all acquisitions were converted from Digital Imaging and Communication in Medicine (DICOM) to NIfTI format. Manual prostate segmentation was carried out by a trained observer (LMK) in the open-source software 3D Slicer (version 5.4.0), to obtain the ground truth labels of each acquisition [43]. When the prostate boundary was interrupted by shadow artefacts, the segmentation boundary was estimated based on visual inspection. For instance, adjacent slices were observed to estimate the prostate boundary as accurately as possible (Figure 2.3c).

2.4.2 Development of prostate segmentation models

The state-of-the-art medical segmentation framework nnU-Net is implemented, since it covers the entire segmentation pipeline, including pre-processing, training and post-processing [44]. During this project, two plane-specific models were trained for 1000 epochs in 2D configuration to segment the prostate in the axial and sagittal imaging-planes and are referred to as *Model Axial* and *Model Sagittal*. Additionally, a third model was trained on a combination of the axial and sagittal images to segment the prostate regardless of the imaging-plane (*Model Sag-Ax*).

Model Axial and *Model Sagittal* were trained and validated in a four-fold cross-validation on plane-specific image data, to explore the robustness of the models across the entire dataset. Folds were selected manually, ensuring that no image of the same participant occurred in different folds. Each fold consisted of 8-11 participants and was representative of the acquisitions in the dataset: TAUS images contained multiple cases of shadow artefacts, clear visibility of the seminal vesicles or bladder disruption by the prostate. Secondarily, both models were evaluated on 10 acquisitions from the respective other imaging-plane to explore the performance of each model on image data of controversial image direction.

Model Sag-Ax was trained on 101 sagittal/axial acquisitions of 42 participants and validated on 10 sagittal and 10 axial acquisitions of 10 participants. Then, the performance of *Model Sag-Ax*

was compared to the performance *Model Axial* and *Model Sagittal* on the same evaluation data.

2.4.3 Model evaluation

For all proposed models, the acquisitions in the validation dataset were predicted and compared to ground truth labels. For each image in an acquisition, the Dice similarity coefficient (DSC) and DSC on true positive images (DSC-TP) were computed and the average across all images was obtained for each acquisition. The DSC measures the similarity between the predicted segmentation and the ground truth label for each slice in the acquisitions[45]. The DSC-TP is obtained to explore the model performance when its ability to classify the prostate is optimal. Additionally, the DSC and the Haussdorf distance were calculated on the predicted segmentation that is in mid-plane of the acquisition (DSC mid-plane, HD mid-plane), as the mid-plane prostate region is required for extraction of the prostate diameters. DSC mid-plane and HD mid-plane are calculated on the predicted segmentation, where the ground truth segmentation has the largest cross-section in pixels. HD mid-plane measures the maximum error between the segmentation and its corresponding ground truth label.[45]

For each model, boxplots were generated on DSC, DSC-TP, and DSC mid-plane to visualise the distribution of each evaluation metric outcome across all participants. When multiple acquisitions of the same participant occurred in the validation dataset, the average result of the acquisitions was reported for each participant. For each fold, all evaluation metrics were reported as the average value across all participants and SD values were retrieved. The ultimate performance of *Model Axial* and *Model Sagittal* is reported by taking the average of each evaluation metric outcome in the four validation folds.

The inter-observer variability of manual prostate segmentation on TAUS was evaluated as a reference standard. It was obtained by comparing the segmentations of three observers (BDB, TN, MF) with a ground truth segmentation (LMK) on DSC in 5 sagittal and 5 axial acquisitions. The inter-observer variability for prostate segmentation in the axial and sagittal imaging-plane was reported as the average DSC score across all plane-specific acquisitions that were segmented by the three observers.

2.5 Proposed Algorithm to extract the prostate diameters

An algorithm was developed to extract the prostate diameters based on one sagittal and one axial input acquisition in which the prostate is segmented. The outcome of the algorithm is a computation of the transverse, AP and longitudinal diameter. Similarly to manual diameter assessment on TRUS, the computation of the AP and transverse diameter is based on the axial segmentation whereas the computation of the longitudinal diameter relies on the sagittal segmentation.

Figure 2.4 provides a schematic overview of the algorithm for automatic prostate diameters extraction, and it is designed as follows. For the sequence of 2D prostate segmentations in the sagittal and axial imaging-plane, the mid-axial and mid-sagittal prostate slice is determined by taking the segmentation with the largest cross-section in terms of number of pixels for each imaging-plane. This results in two segmentations: *SegAx* and *SegSag*. Then, the algorithm verifies whether both segmentations consist of one continuous delineation, to ensure an appropriate determination of the centroid of both segmentations. When multiple segments are detected, the largest continuous segment is assigned as *SegAx* and/or *SegSag*. To approximate the maximum transverse and AP diameter, an ellipsoid is fit to the outer boundary of *SegAx*. The major and minor axis lengths of the ellipsoid then simulate the transverse and AP diameters according to the used angle of rotation.

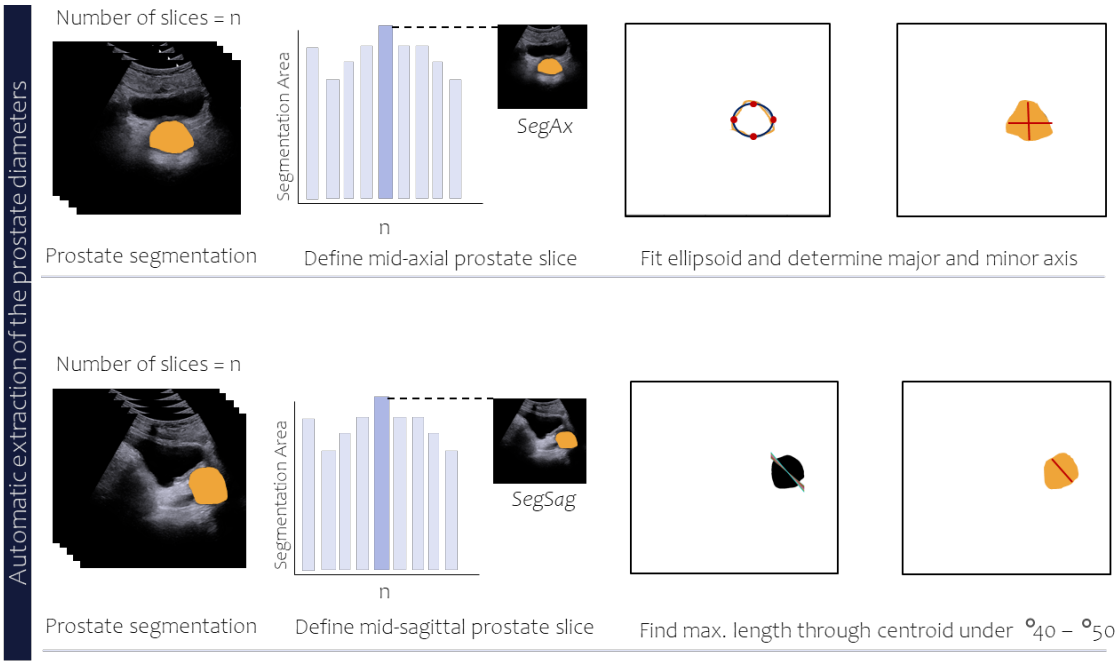


Figure 2.4: Proposed algorithm to extract the transverse, anterior-posterior and longitudinal prostate diameters. The inputs of the algorithm are two sequences of TAUS images (sagittal and axial imaging-plane) in which the prostate is segmented for each slice in the acquisition. The computation of the AP and transverse diameter is based on the axial segmentation whereas the computation of the longitudinal diameter relies on the sagittal segmentation. The algorithm is designed as follows: First the mid-plane prostate segmentation is defined as the segmentation with the largest prediction area, resulting in *SegAx* and *SegSag*. Then an ellipsoid is fit on the outer boundary of *SegAx*, whereby the major and minor axis of the ellipsoid act as the transverse and anterior-posterior diameter. The longitudinal diameter is approximated by defining the maximum line through the centroid of *SegSag* under 40 to 50 degrees.

The major axis corresponds to the transverse diameter when the Ellipsoid's angle of rotation is 0-45, 135-225, and 315-360 degrees, while it corresponds to the AP diameter, when other angles of rotation apply (46-134, 226-314 degrees). The longitudinal diameter is computed by drawing ten lines that go through the centroid of *SegSag* under a range of °40 to°50 degrees. The maximum line L is determined and its intersection coordinates with *SegSag* are obtained. The longitudinal diameter is then calculated according to $L = \Delta x \times tan(\alpha)$.

To evaluate the proposed method, the algorithm is tested on 42 axial/sagittal acquisitions in which the prostate is segmented manually. For each prostate diameter, the computed diameter is compared to the manually assigned diameter on TAUS, visualised in a Bland Altman diagram. Additionally, the PV is predicted according to the $PV_{Ellipsoid}$ and PV_{Axial}, and compared to manual volume measurements based on TAUS to explore the algorithm performance when the input segmentations are correct. Bland Altman diagrams are generated to show the inter-method agreement of the computed PV against the manual measurement.

2.6 Phase 3: Automatic prostate volume estimation using TAUS

Essentially, the combination of the developed segmentation models, and the algorithm that computes prostate diameters results in a framework that enables automatic PV estimation on TAUS.

The entire framework is tested in the last phase of this study by predicting $PV_{Ellipsoid}$ and PV_{Axial} on TAUS acquisitions of 17 participants. To obtain *SegAx* and *SegSag*, the best-performing models of the cross-validated *Model Axial* and *Model Sagittal* were assigned to segment the prostate in both imaging-planes. Then, transverse, AP and longitudinal diameters were computed according to the proposed algorithm. Subsequently, Equation 1.1 is used to predict $PV_{Ellipsoid}$, and Equation 2.1 is used to predict PV_{Axial}. For both predictions, the PV is compared to reference PV measurements based on MRI, whereby Bland Altman diagrams were generated to illustrate the inter-method agreement.

3

Results

3.1 Dataset characteristics

The baseline characteristics of all included participants are shown in Table 3.1. The participants that were included in this study had an average age of 66.5 ±10.6 years. The average prostate volume on MRI was 58.5 ±32.4 ml with a mean PSA density of 0.20 ±0.30 ng/ml/cc. When the acquisitions of each participant were assessed on quality, 61 participants yielded quality A/B images, while the acquisitions of 25 participants contained poor prostate visibility (quality C). The prostate was not present in the acquisitions of 14 participants and were therefore excluded.

Table 3.1: Baseline characteristics of included subjects (n=100

Characteristic	
	Mean ±SD
Age (y)	66.5 ±10.6
PSA-density (ng/mL/cc)	0.2 ±0.3
MRI prostate volume (mL)	58.5 ±32.4
PI-RADS	**Number**
1	3
2	48
3	14
4	10
5	25
TAUS image quality	
incorrect	14
A	26
B	35
C	25

SD: standard deviation, PSA: prostate-specific antigen, MRI: magnetic resonance imaging. TAUS: transabdominal ultrasound, PI-RADS: Prostate Imaging-Reporting and Data System.

3.2 Phase 1: Prostate volume estimation inter-method agreement: TAUS, TRUS, and MRI

3.2.1 PV estimation based on Ellipsoid formula

For 86 participants, the PV was measured on TAUS and MRI using the Ellipsoid formula ($PV_{Ellipsoid}$). TRUS examination was not conducted in nine participants, resulting in 77 PV measurements based on TRUS. When the PV measurements of MRI, TRUS and TAUS were analyzed using the Wilcoxon Rank Test, a significant difference between the two sets of measurements was shown, regardless of the observed imaging modality (TRUS vs TAUS P <0.001, TRUS vs MRI P < 0.001, TAUS vs MRI P = 0.009). In other words, a significant difference was obtained when $PV_{Ellipsoid}$ measurements obtained with MRI were compared to TAUS measurements, when TRUS was compared to MRI, and when TRUS was compared to TAUS.

The average difference between PV measurements on TAUS and MRI was 3.0 ±17.6 ml (n = 86). Figure 3.1a shows the Bland Altman diagram, whereby the positive bias of 3.0 ml indicates that on average, the PV acquired using TAUS tends to underestimate the corresponding measurement obtained with MRI. When PV measurements using TAUS and TRUS were compared, an average volume difference of -8.0 ±17.6 ml was found (n = 77), indicating that TAUS tends to overestimate the PV compared to TRUS. The Bland Altman diagram on TRUS and TAUS shows an increase in variability as the PV enlarges (Figure 3.1d). As a reference, the average difference between PV measurements derived from TRUS and MRI is 12,3 ±18.8 ml (n = 77). Figure 3.1c depicts the variance between TRUS and MRI, illustrating that the difference between volume measurements increases when the prostate volume is enlarged.

3.2.2 PV estimation based on Axial formula

The inter-method agreement between PV_{Axial} and MRI resulted in an average volume difference of 11.2 ±20.3 ml (n = 86), and is shown in Figure 3.1b.

Regarding the inter-dependency of prostate diameters in different imaging-planes, the Spearman's Correlation test shows that the transverse diameter correlates poorly with the longitudinal diameter ($\rho = 0.38$, P < 0.001), while the AP diameter has a strong correlation with the longitudinal diameter in TAUS images ($\rho = 0.77$ P <0.001).

3.3 Phase 2: Deep neural networks for prostate segmentation

3.3.1 Evaluation Model Sagittal and Model Axial

Table 3.2 shows the performance of *Model Axial* and *Model Sagittal*, tested on plane-specific TAUS images in terms of average DSC, DSC-TP, DSC mid-plane, and HD mid-plane. *Model Axial* segments the prostate on axial images with an average DSC of 0.76 ±0.09, while *Model Sagittal* segments the prostate on sagittal images with an average DSC of 0.68 ±0.21. When the mid-plane region of the prostate is evaluated, it shows that *Model Axial* and *Model Sagittal* segment the prostate with an average DSC mid-plane = 0.91 ±0.06 (HD = 0.62 ±0.40 cm), DSC mid-plane = 0.83 ±0.09 (HD = 0.89 ±0.40 cm) respectively. Figure 3.2 presents boxplots on DSC, DSC-TP and DSC- midplane for each fold in the cross-validation, illustrating the performance of both models

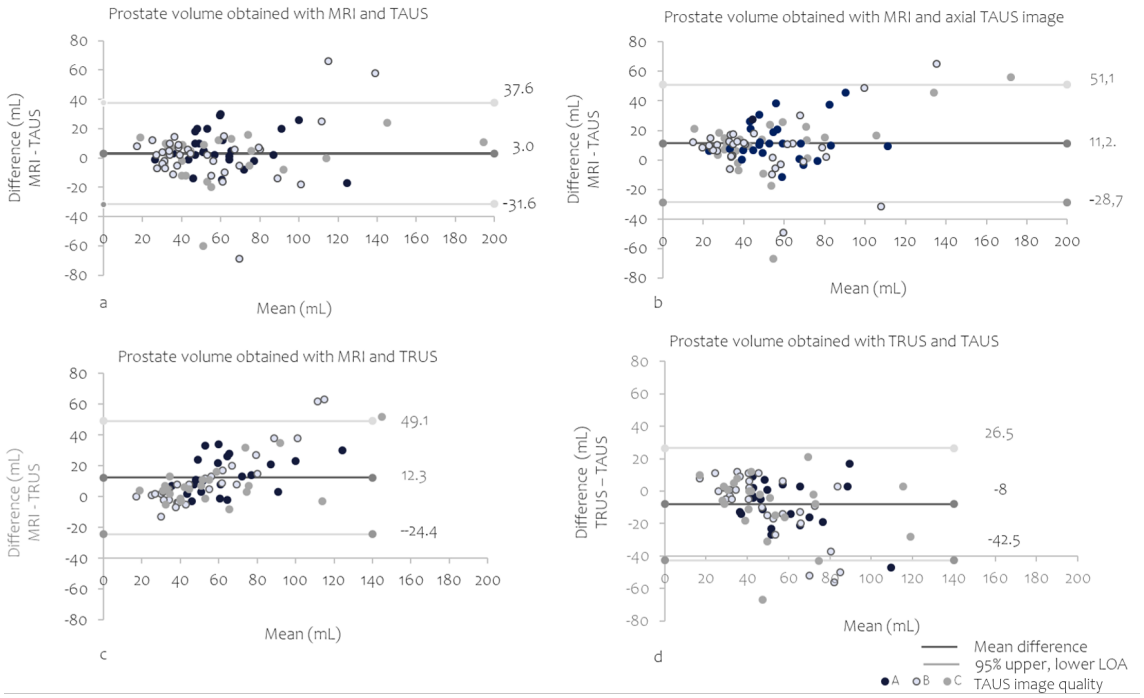


Figure 3.1: The comparison of TAUS, TRUS and MRI for prostate volume estimation, whereby the difference between two sets of PV measurements is visualised on the vertical axis and the mean of the pair is present on the horizontal axis. The upper and lower limits of agreement (LoA) visualise the 95% confidence interval of the included measurements. For each participant, the image quality of the TAUS image is assigned (quality A, B, C). **a:** TAUS compared to MRI ($PV_{Ellipsoid}$). **b:** TAUS compared to MRI (PV_{Axial}). **c:** TRUS compared to MRI. **d:** TRUS compared to TAUS.

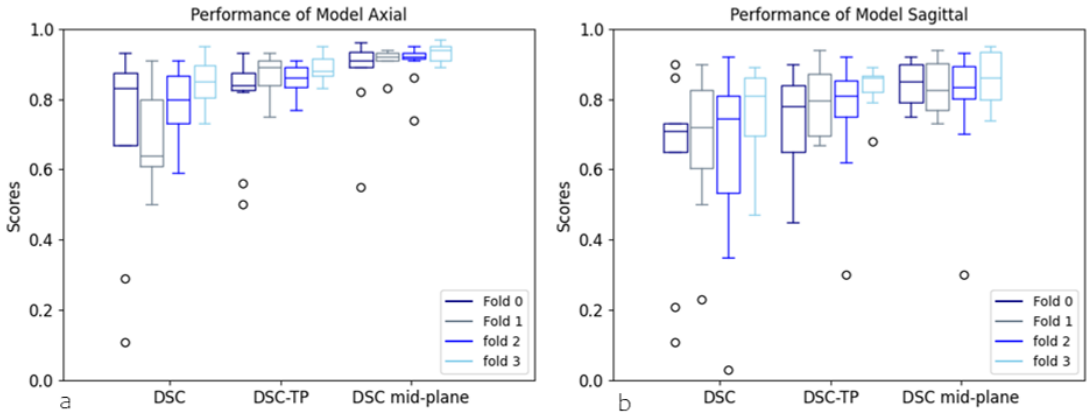


Figure 3.2: Model performance for each fold in the cross-validation, shown in boxplots. For each fold, the dice similarity coefficient (DSC), the DSC on true positive predicted segmentations (DSC-TP) and the DSC in the mid-prostate region were reported. **a:** Model Axial. **b:** Model Sagittal.

across all participants in the dataset. *Model Axial* and *Model Sagittal* achieve the highest DSC score when evaluated on Fold 3.

A qualitative comparison of the performance of *Model Axial* and *Model Sagittal* on different TAUS images is shown in Figure 3.9 and Figure 3.10, where the ground truth segmentation (green) and the prediction (red) are overlaid in the original TAUS image. For both imaging-planes, five cases are presented in which the apex, middle section, and the base of the prostate are displayed. In

line with the quantitative results, the visual inspection shows that in the majority of the cases, the prediction tends to correspond with the ground truth segmentation when the middle region of the prostate is predicted. (Figure 3.9 b,e,h,n, Figure 3.10 b,e,k,n) Additionally, the axial prostate area is segmented correctly when the prostate disrupts the bladder, a sign of prostate enlargement that frequently occurs in middle-aged men (Figure 3.9 c). Deviations in detecting the prostate boundary on axial images mainly comprise overestimations due to present shadow artefacts (Figure 3.9 f) or deviating/missed segmentations in the apex and base prostate region (Figure 3.9 g,o). In a single case, the model falsely classified bladder debris as prostate tissue (Figure 3.9 j).

When the predictions of *Model Sagittal* are examined visually, it is observed that the predictions deviate from the ground truth segmentations more profoundly compared to the predictions of *Model Axial* tested on axial TAUS images. In several cases, the predictions are over- and underestimating the prostate region (Figure 3.10 a,d,g), along with false negative outcomes that mostly occur in the apex and base of the prostate region (Figure 3.10 i,j) When the performance of both models was evaluated on TAUS images of controversy image direction, false negative outcomes were obtained in the majority of the input images. In other words, both models have difficulty in finding the prostate on the imaging-plane that was not included in the training phase of each model.

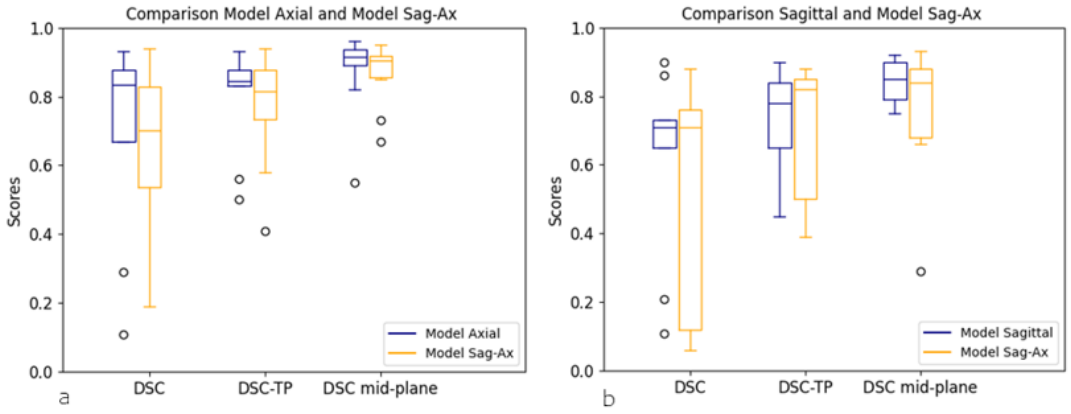


Figure 3.3: A comparison of the performance of *Model Axial, Model Sagittal* and *Model Sag-Ax* on Dice similarity coefficient (DSC), DSC on true positive predicted segmentations (DSC-TP) and DSC in the mid-prostate region (DSC mid-plane). Overall, the plane-specific models outperform *Model Sag-Ax* on all evaluation metrics.**a:** *Model Axial* vs *Model Sag-ax* **b:** *Model Sagittal* vs *Model Sag-ax*

3.3.2 Evaluation Model Sag-Ax

The performance of *Model Sag-Ax* in terms of DSC, DSC-TP, DSC mid-plane, and HD mid-plane is incorporated in Table 3.2. Based on the evaluation on test images, which is a union of axial and sagittal test data, *Model Sag-Ax* segments the prostate with an average DSC = 0.64 ±0.25 and DSC = 0.54 ±0.32 on axial and sagittal TAUS images respectively. When the mid-plane region of the prostate is evaluated, *Model Sag-Ax* segments the prostate with an average DSC mid-plane of 0.87 ±0.09 (HD = 0.65 ±0.66 cm), 0.76 ±0.19 (HD = 0.97 ±0.28 cm) for axial and sagittal TAUS images respectively.

Figure 3.3 displays the comparison of *Model Sag-Ax* to *Model Sagittal* and *Model Axial*, tested on the same image data, on DSC, DSC-TP and DSC mid-plane, visualized in boxplots. When *Model Sag-Ax* is compared to *Model Sagittal*, a lower model performance is shown regardless of the observed evaluation metric. Similar results can be observed when *Model Sag-Ax* is compared

to *Model Axial*. Nevertheless, *Model Sag-Ax* outperforms *Model Axial* in terms of HD mid-plane (0.73 ±0.64 cm vs. 0.65 ±0.66 cm).

When the predictions of *Model Sag-Ax* were inspected visually, it was observed that false positive predictions occurred more frequently compared to the plane-specific models. The false positive segmentations mainly comprised dark shadow artefacts, occurring in the TAUS image. Similar to the plane-specific models, the predictions of *Model Sag-Ax* corresponded closely to the ground truth the mid-prostate region was evaluated. Figure 3.4 shows a comparison between prostate segmentations of *Model Sag-Ax* and *Model Sagittal/ModelAxial* when similar images of the same participant are provided.

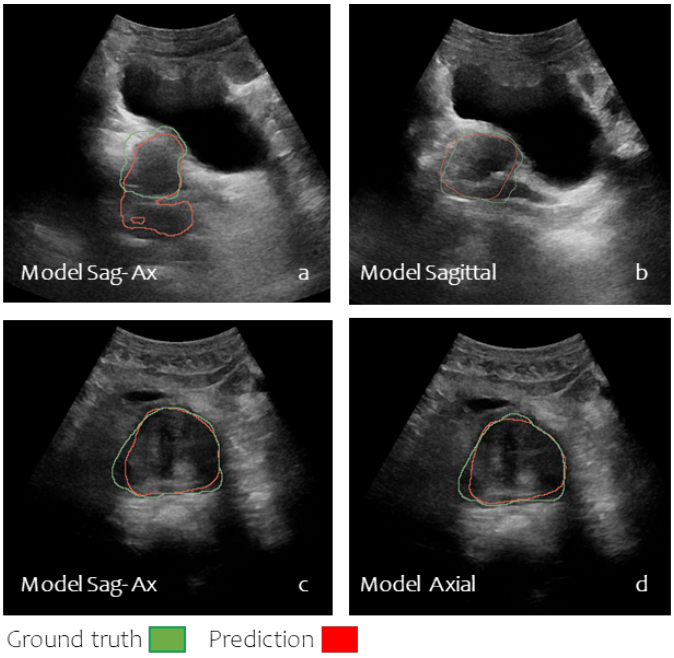


Figure 3.4: A comparison of the predictions of *Model Sag-Ax*, *Model Sagittal*/*Model Axial* when similar images of the same participant are provided. The examples show increased false positive predictions of *Model Sag-Ax* compared to *Model Sagittal*. At the same time, the prediction of *model Sag-Ax* corresponds closely to the prediction of *Model Axial* and the ground truth segmentation **a:** Model Sag-Ax **b:** Model Sagittal **c:** Model Sag-Ax **d:** Model axial

3.3.3 Inter-observer variability of manual prostate segmentation on transabdominal ultrasound

The analysed inter-observer variability of manual prostate segmentation on TAUS showed a DSC = 0.66 ±0.17 for axial TAUS images. For prostate segmentation on sagittal TAUS images, a DSC = 0.62 ±0.13 was reported.

Table 3.2: Performance of Model Axial, Model Sagittal and Model Sag-Ax to segment the prostate on transabdominal ultrasound

Model	DSC	DSC-TP	DSC mid-plane	HD mid-plane [cm]
Model Axial				
Fold 0	0.70 ±0.25	0.80 ±0.13	0.89 ±0.11	0.73 ±0.64
Fold 1	0.69 ±0.14	0.87 ±0.05	0.91 ±0.04	0.65 ±0.37
Fold 2	0.78 ±0.11	0.86 ±0.04	0.93 ±0.03	0.47 ±0.15
Fold 3	0.85 ±0.07	0.89 ±0.03	0.91 ±0.05	0.61 ±0.41
mean	**0.76 ±0.09**	**0.86 ±0.06**	**0.91 ±0.06**	**0.62 ±0.40**
Model Sagittal				
Fold 0	0.62 ±0.26	0.74 ±0.13	0.84 ±0.06	0.78 ±0.26
Fold 1	0.67 ±0.21	0.79 ±0.10	0.83 ±0.07	0.98 ±0.46
Fold 2	0.65 ±0.25	0.76 ±0.16	0.80 ±0.16	1,20 ±0.58
Fold 3	0.76 ±0.12	0.83 ±0.06	0.86 ±0.07	0,60 ±0.37
mean	**0.68 ±0.21**	**0.78 ±0.12**	**0.83 ±0.09**	**0.89 ±0.40**
Model Sag-Ax				
Axial data	0.64 ±0.25	0.77 ±0.15	0.87 ±0.09	0.65 ±0.66
Sagittal data	0.54 ±0.32	0.70 ±0.18	0.76 ±0.19	0.97 ±0.28

The Dice similarity coefficient (DSC), DSC on true positive segmentations (DSC-TP), DSC score in the mid-plane image of the acquisition (DSC mid-plane), Haussdorf Distance in the mid-plane image of the acquisition (HD mid-plane) Results are reported as average outcome across all acquisitions ±standard deviation. *Model Axial* was trained and validated on axial TAUS images, while *Model Sagittal* was trained and validated on sagittal TAUS images. *Model Sag-Ax* is trained on the combination of sagittal and axial TAUS images and tested individually on images of the axial and sagittal imaging-plane.

3.4 Evaluation Algorithm to extract the prostate diameters

Figure 3.5 illustrates the expected variability related to the diameter computation when the segmentation performance is optimal. Compared to manually assigned diameters, it is shown that the proposed algorithm derives the longitudinal diameter with an average difference of 0.18 ±0.8 cm. In other words, the algorithm tends to underestimate the longitudinal diameter whereby a maximum underestimation of 1.74 cm can be expected (Figure 3.5a). On the contrary, when the computed transverse and AP diameters were analysed, an average difference of -0.05 ±0.4 and -0.3 ±0.3 cm were observed respectively (Figure 3.5b, c).

When the computed diameters are utilized for the calculation of $PV_{Ellipsoid}$, an average volume difference of -1.5 ±10.1 ml compared to manual PV estimation on TAUS is observed (Figure 3.6a). Similar results are obtained for PV_{Axial}, and Figure 3.6b shows that the computed PV_{Axial} results in an average volume difference of 1.3 ±9.4 ml when compared to manual PV_{Axial} measurements on TAUS.

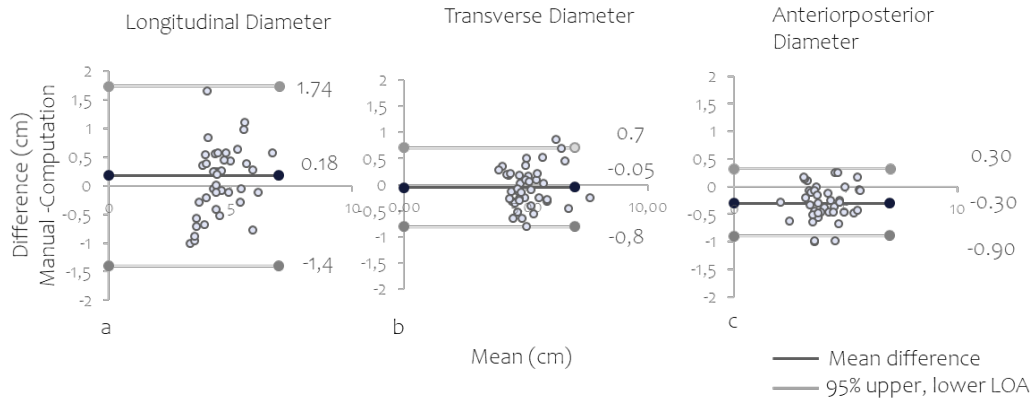


Figure 3.5: The difference between computed prostate diameters and manually assigned diameters visualised in Bland Altman diagrams. **a:** The longitudinal diameter **b:** The transverse diameter. **c:** The anteriorposterior diameter.

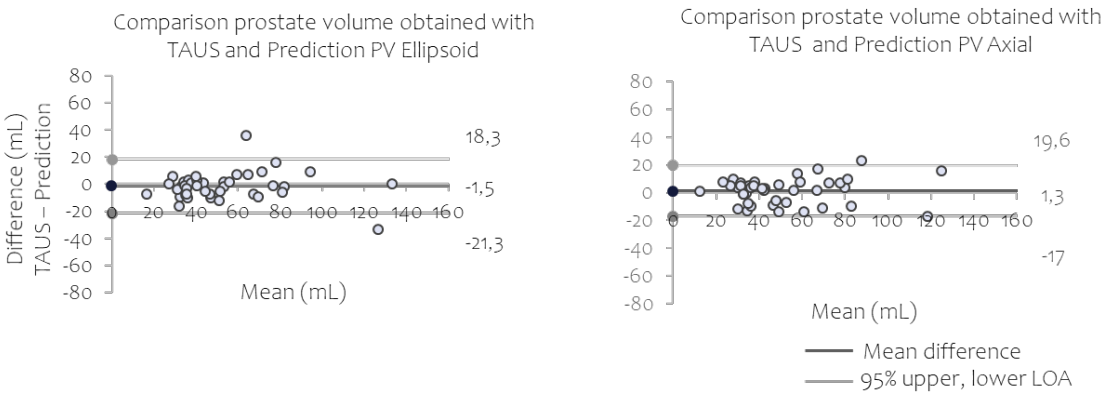


Figure 3.6: A comparison between PV measurements based on computed diameters and the manually assigned diameters when ground truth segmentations are provided. **a:** Comparison manual $PV_{Ellipsoid}$ and prediction **B:** Comparison manual PV_{Axial} and prediction

3.5 Phase 3: Automatic prostate volume estimation on transabdominal ultrasound

This section covers the final evaluation of the framework (automatic segmentation and diameter extraction), tested on TAUS acquisitions of 17 participants. Figure 3.8 shows the prediction of $PV_{Ellipsoid}$, the prediction of PV_{Axial} and reference PV measurements on MRI for each participant in the test dataset. Additionally, the relative deviation of the predicted $PV_{Ellipsoid}$ compared to MRI is reported. For 14 out of 17 cases, the prediction of $PV_{Ellipsoid}$ deviates with less than 25% compared to the MRI volume. A maximum overestimation of 47% is reported for a single case in the test dataset. In one case, the prediction of $PV_{Ellipsoid}$ correspond with 100% to the PV measurements based on MRI.

In figure 3.7 the inter-method agreement between PV predictions and PV estimations based on MRI is illustrated in Bland Altman diagrams. Figure 3.7a shows that the prediction of $PV_{Ellipsoid}$ deviates on average with 2.5 $\pm$10.2 ml compared to MRI. When the predictions of PV_{Axial} are analysed, an average volume difference of -1.2 $\pm$10.8 ml is observed (Figure 3.7b).

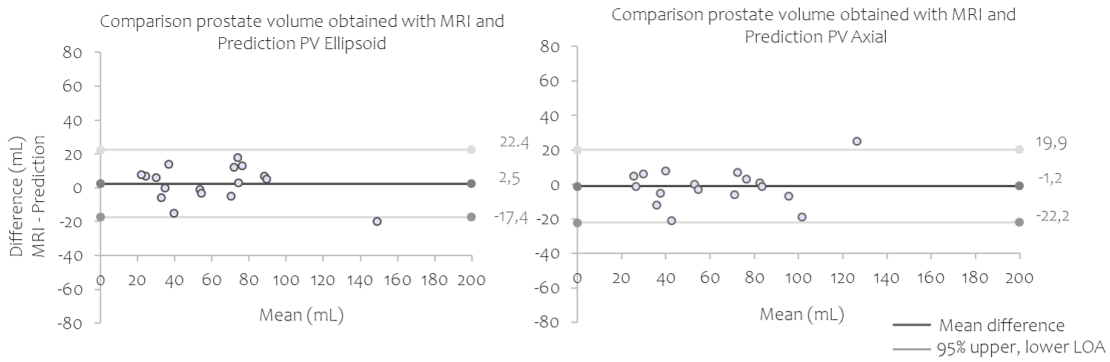


Figure 3.7: A comparison between predicted prostate volume (PV) measurements and reference volume measurements based on MRI, whereby the mean difference and the limit of agreements (LoA) compared to MRI are visualised.**a:** MRI compared to the predicted PV according to $PV_{Ellipsoid}$ **b:** MRI compared to the predicted PV according to PV_{Axial}.

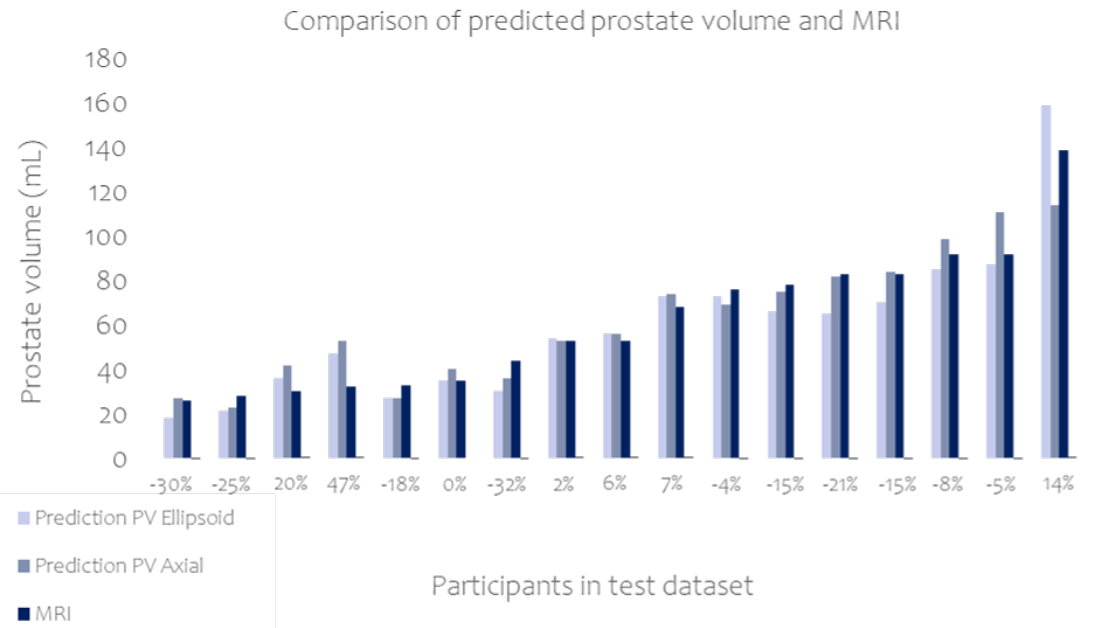


Figure 3.8: A comparison of prostate volume predictions with reference volume measurements on MRI. The barplot shows the predicted PV according to PV_{Axial} and $PV_{Ellipsoid}$ and the reference MRI volume for each participant in the test dataset. The relative difference between Predicted $PV_{Ellipsoid}$ and the reference MRI volume is reported. In 14/17 test cases, the relative difference compared to MRI is $\leq$25%.

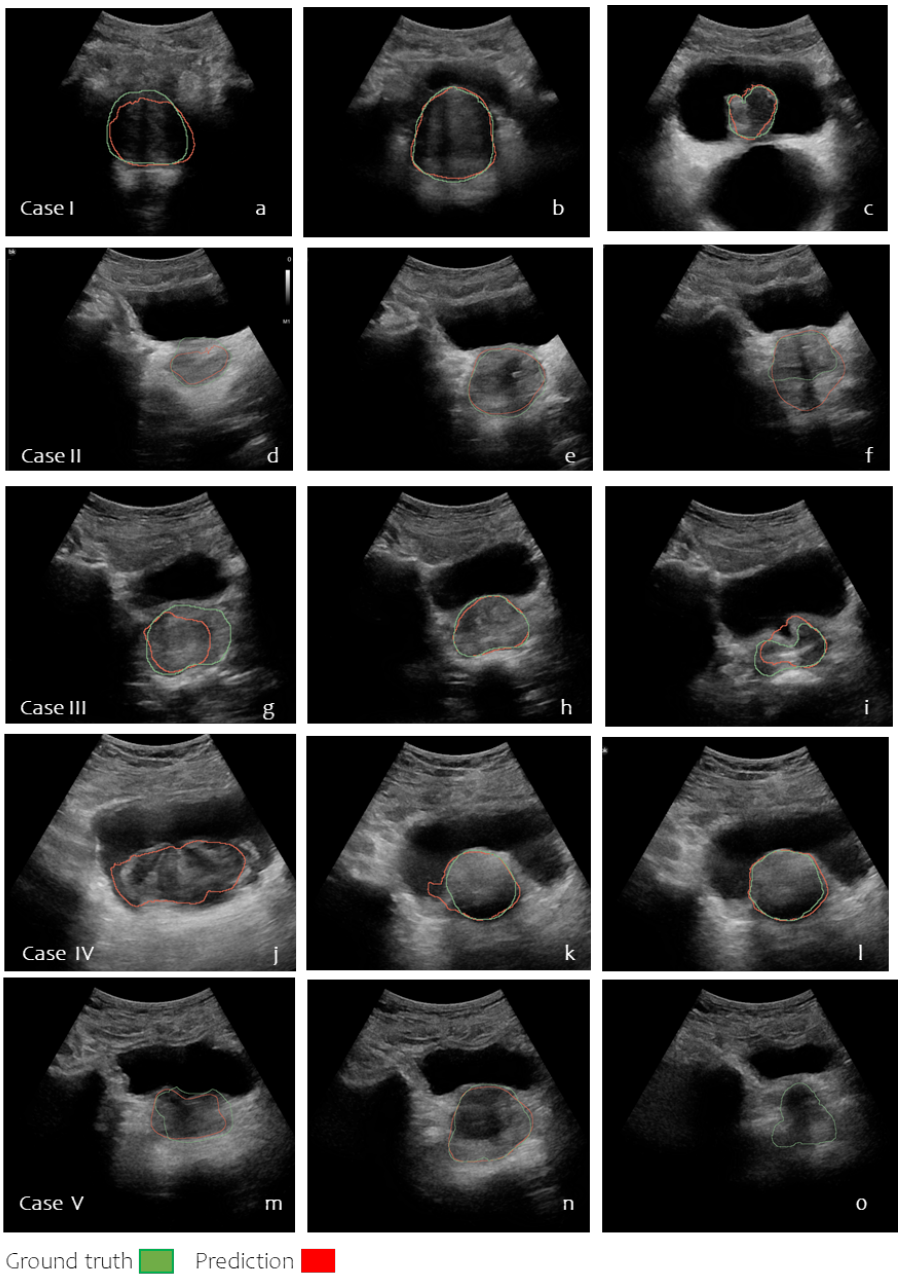


Figure 3.9: A qualitative comparison of the performance of Model Axial where the ground truth segmentation (green) and the prediction (red) are overlaid in the TAUS image. For each case, an image from the apex, middle, and base prostate region is provided. **a, b, c:** Predicted segmentations that correspond closely to the ground truth segmentation. **f:** Misclassification of shadow artefact. **j:** False-positive segmentation of bladder debris occurring in the image. **o:** False-negative segmentation in the base region of the prostate.

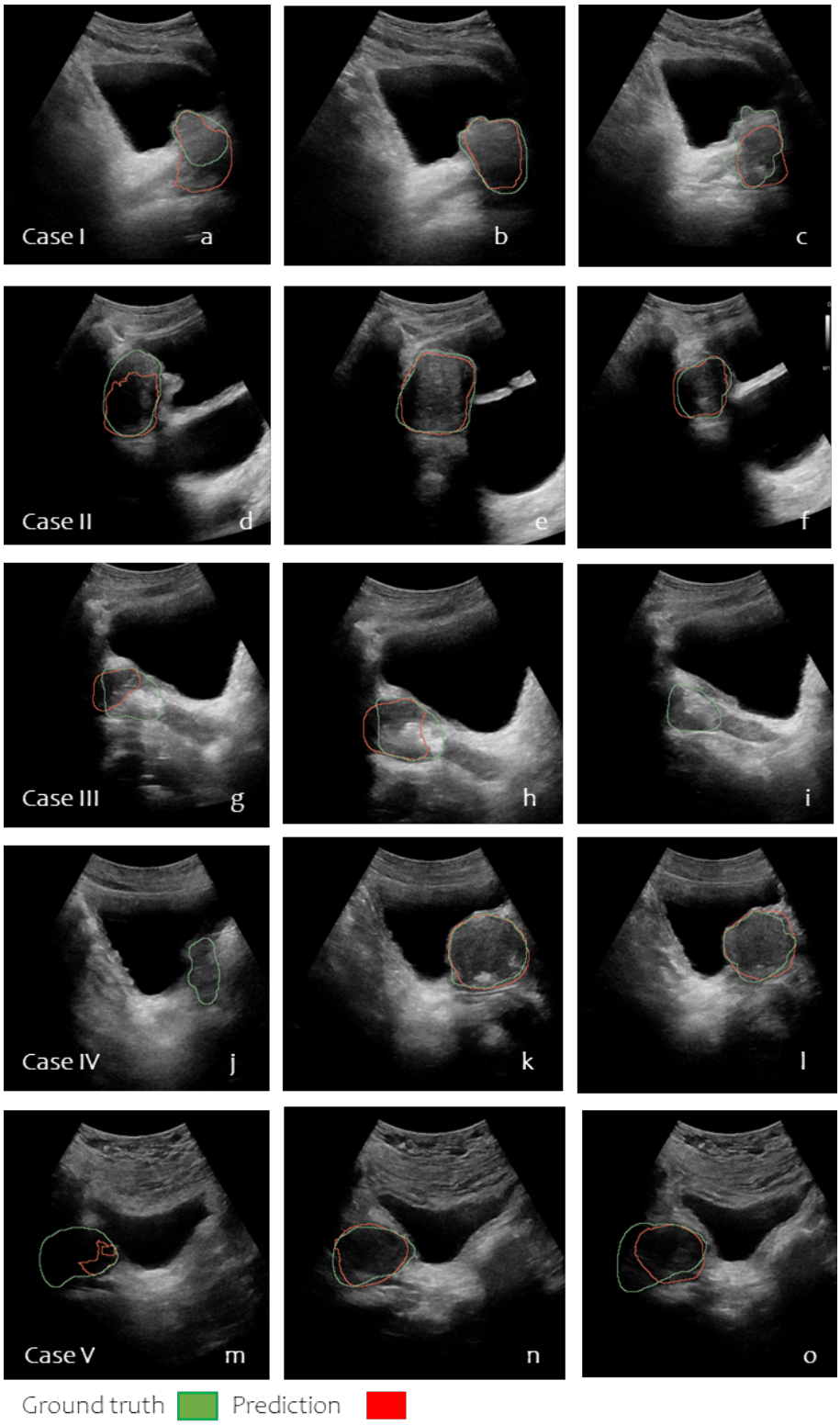


Figure 3.10: A qualitative comparison of the performance of Model Sagittal where the ground truth segmentation (green) and the prediction (red) are overlaid in the TAUS image. For each case, an image from the apex, middle, and base prostate region is provided. **a, b, c:** Predicted segmentations showing more deviation in the apex and base prostate region compared to the middle region.**b, e, k, n:** Predictions of the mid-prostate region that correspond closely to the ground truth segmentation. **d, g:** The predicted segmentation missed the prostate region partly. **i, j:** False negative segmentation in the apex/base region of the prostate.

4

Discussion

Prostate cancer (PCa) risk-stratification relies on accurate determination of the prostate volume (PV) whereby transrectal ultrasound (TRUS) is predominantly utilized in clinical practice. This study proposes a framework for automatic PV estimation based on transabdominal ultrasound (TAUS) as a patient-friendly, and robust alternative to infer the PV. Therefore, the feasibility of TAUS for manual PV estimation is explored and compared to PV estimation based on TRUS and magnetic resonance imaging (MRI). Secondly, three deep neural networks are developed to segment the prostate on TAUS images and serve as input to a proposed algorithm that computes prostate diameters. Ultimately, the framework enables automatic PV estimation when input TAUS acquisitions are provided.

4.1 Prostate volume estimation inter-method agreement: TAUS, TRUS, MRI

4.1.1 PV estimation based on the Ellipsoid Formula

TAUS is a non-invasive imaging modality, well-tolerated by patients, and its general availability makes it feasible for routine use in PCa risk stratification workflows. Former research on the topic show that PV estimation based on TRUS, TAUS, and MRI results in correlated outcomes whereby the Ellipsoid formula ($PV_{Ellipsoid}$) can be employed as a rapid technique with reasonable accuracy and repeatability [17], [32].

When $PV_{Ellipsoid}$ measurements based on TAUS are compared to MRI, an average volume difference of 3.0 ±17,6 ml was observed. Figure 3.1a suggests that the variance of PV measurements is independent of the PV in a range of 20 to 100 ml. Since our dataset only comprised six cases above 100 ml, it remains uncertain whether the increased SD in these cases can be attributed to prostate enlargement. Furthermore, the variance in PV measurements seems unrelated to the TAUS image quality, indicating that manual diameter assessment can be performed even when unfavorable patient/TAUS characteristics limit the image quality. In a previous study conducted by Guo *et al.*(2023), a comparison between $PV_{Ellipsoid}$ utilizing TAUS and MRI showed an average difference of -4.1 ±11.0 ml [17]. In contrast to our findings, their proposed Bland Altman diagram suggests a decreased inter-method agreement when the PV increases. Additionally, TAUS overestimated the PV measurements based on MRI whereas our results show an average underestimation of

3.0 ml. The discrepancy in results may be attributed to the inter-observer variability related to the Ellipsoid Formula and the variation is used datasets. [33]. Additionally, in their work the AP diameter was assessed on the sagittal imaging-plane, while the AP diameter in our work was derived from the axial TAUS image. Currently, there is no standard method of obtaining the prostate's diameters on TAUS, which may be of interest in future explorations on PV estimation using prostate diameters on TAUS.

When $PV_{Ellipsoid}$ measurements using TAUS and TRUS were compared, an average volume difference of -8.0 ±17.6 was observed, whereby the variance of PV measurements increased as the PV increases (Figure 3.1d). In line with this, Gerald *et al.* (1996) also observed a connection between PV and variability when TRUS is utilised, demonstrating that TRUS underestimates PV measurements more frequently when the PV exceeds 30 ml [46]. Moreover, De Vos et al. (2023) explored the feasibility of PV measurements with the aid of TAUS in a primary care setting and reported an average volume difference of -9.9 ml ±2.0 ml when TAUS was compared to TRUS [18]. Similarly, the authors showed that the variability of PV measurements between TRUS and TAUS increased as the PV enlarged. One possible explanation for this phenomenon is that enlarged prostates often extend into the bladder area, which is difficult to visualize in sagittal TRUS images. As a result, the assessment of the longitudinal diameter becomes more difficult especially when larger PVs are encountered. On the contrary, MRI and TAUS acquisitions have in common that the bladder region is more visible, resulting in better visualization of larger prostates even when growing into the bladder. This may explain the larger variance of TRUS compared to TAUS when they are analyzed against MRI volumes (Figure 3.1c). Yet, it must be acknowledged that deviations in TRUS measurements may be the result of study limitations. In several cases, TRUS examination was conducted by clinicians who were still in training and the diameter assessment was always performed during patient examination. In contrast, the diameter assessment based on TAUS and MRI were always performed after patient examination. More importantly, during this study, all clinicians knew beforehand that an MRI-based PV estimation was conducted for each participant and further decisions were based on this outcome.

The results from this study suggests that it is possible to achieve PV estimations through TAUS images that are similar to MRI. Still, it is important to notice that the determination of the apex prostate region on the sagittal imaging-plane influences the accuracy of manual PV measurements based on TAUS. When TAUS examination is conducted improperly, shadow artefacts induced by the reflection of the pelvic bone, limit the prostate's visibility. As a result, the clinician is forced to guess the longitudinal diameter, leading to undesirable variability. The variability in the estimation of diameters, due to improper TAUS examination has been reported to decrease in more expert users, whereby the learning curve associated with TAUS examination indicates that more experienced clinicians learn how to perform better acquisitions [18], [32]. The same learning curve was also accounted by the operator (LMK) when TAUS examination was performed during this study. During the initial phase of data collection, several acquisitions were carried out improperly (n=14), and therefore excluded, whilst the prostate was clearly visible in the acquisitions that were acquired later on. One possible solution to minimize the (initial) complexity of TAUS examination, is a real time prostate detection system that aids inexpert operators in finding the prostate during TAUS examination. Recently, Natali *et al.*(2024) investigated an automatic prostate detection system, to promote appropriate prostate visualisation when TAUS acquisitions are conducted.[47]

4.1.2 PV estimation based on Axial formula

To exclude the variability related to determination of the longitudinal diameter on the sagittal imaging-plane, this study introduces a new method to estimate the PV in which the (axial) AP diameter substitutes the (sagittal) longitudinal diameter for their observed inter dependency (ρ=0.77, P <0.001). More importantly, PV_{Axial} enables a PV approximation that necessitates only one axial TAUS acquisition, thereby reducing the patient examination duration. Figure 3.1b shows the comparison of PV_{Axial} and MRI, where an average volume difference of 11.2 ±20.3 ml was observed. Compared to $PV_{Ellipsoid}$, the SD increased by 2.7 ml, showing that the approximation of PV_{Axial} results in a less accurate estimation. The increased variability of PV_{Axial} can be explained as the interdependence of the AP diameter and the longitudinal diameter is an assumption that not always agrees. Yet, in several cases PV_{Axial} corresponded closely to MRI and further research is required to explore its usability. For instance, it is worth to explore in which PV range, more accurate PV estimations are conducted when PV_{Axial} is utilised.

4.1.3 TAUS vs TRUS vs MRI: Wilcoxon Rank Test

Regardless of the used imaging modality (TAUS vs. MRI, TRUS vs. MRI, TAUS vs.TRUS) the difference between PV measurements was statistically significant. Similar results were observed in a previous study that compared the employment of the Ellipsoid formula on TRUS and MRI [48]. Therefore, consistency in the choice of imaging modality is important to ensure the generation of robust volume estimations.

In conclusion, the analysed inter-method agreement of TAUS, TRUS and MRI for PV estimation indicates that TAUS and TRUS measurements vary to a similar extent when compared to MRI reference volumes. In contrast to TRUS, TAUS seems less prone to inaccuracies when larger prostate volumes are encountered. However, both TRUS and TAUS are characterised with a limited prostate visibility in the sagittal image-plane, which can be regarded as the main contribution regarding differences with MRI. The employment of PV_{Axial} excludes the necessity of a sagittal TAUS acquisition, however, further research is required when more accurate PV estimations are desired. Finally, statistically different PV outcomes are obtained when TAUS, TRUS and MRI are compared. This highlights the importance of a consistent use of one imaging modality when PV-based protocols are developed.

4.2 Deep neural networks for prostate segmentation

4.2.1 Model Axial, Model Sagittal and Model Sag-Ax

This is the first attempt to develop deep neural networks for prostate segmentation on axial and sagittal TAUS images. It enables an automatic delineation of the prostate on which the prostate's diameters can be derived in both imaging-planes. Since the main objective is to estimate the transverse, AP and longitudinal diameter in the mid-prostate region, the segmentation performance of all proposed models was evaluated on the mid-plane TAUS images where the ground truth prostate segmentation is the largest. The results of Table 3.2 show that *Model Axial* and *Model Sagittal* segment the mid-plane prostate region with an average DSC score of 0.91 ±0.06 and 0.83 ±0.09. Even though, *Model Sag-Ax* can segment the prostate regardless of the input image-plane, a lower performance was observed when tested on axial and sagittal TAUS images (DSC = 0.87

±0.09 on axial images, DSC = 0.76 ±0.19 on sagittal images).

The proposed models are also evaluated on all images in the TAUS acquisition to quantify the model's ability to segment the apex, middle and base prostate region. When *Model Axial*, *Model Sagittal*, and *Model Sag-Ax* are evaluated on the entire prostate region, a decreased model performance is found compared to the evaluation on the mid-prostate region, indicating that the apex and base region of the prostate are segmented with less accuracy (Table 3.2). The lower DSC score is primarily the result of false-negative predictions, occurring in apex and base prostate region (Figure 3.9o, 3.10i,j).

Table 3.2 shows that the performance of each model evaluated on the mid-prostate region (DSC mid-plane) outperformed the evaluation on the entire prostate region (DSC-TP). In other words, even when the models classify the prostate tissue correctly, the predictions of the apex and base prostate region tend to deviate more from the ground truth segmentations compared to predictions in the mid-prostate region. These outcomes are expected when it is taken into account that the prostate's upper/lower parts are visualized more ambiguously on TAUS images and the same phenomenon is reported in several studies on prostate segmentation on TRUS [26], [37]. On the other hand, it was shown that the base prostate region was segmented correctly when the prostate was interrupting the bladder region. This may be the result of the better contrast of the prostate tissue against the anechoic bladder in the TAUS image (Figure 3.9 c). Furthermore, prostate-bladder interruption was frequently seen in the training images of *Model Axial*, resulting in an improved segmentation performance on unseen cases of bladder intrusion.

In one specific case, the prostate was falsely segmented when bladder debris occurred in the TAUS image (3.9 j). False positive segmentations must be limited to maximize the reliability of *Model Axial*, which can be overcome by increasing the representation of specific image/patient characteristics in the training dataset.

Overall, *Model Axial* outperforms *Model Sagittal* on all used evaluation metrics, suggesting that the segmentation task in the sagittal image-plane is more challenging. This is most likely related to the presence of shadow artefacts that limit the prostate visibility in the image (Figure 3.10 a). When a higher model performance is desired for prostate segmentation in sagittal TAUS images, additional training data may be required due to the complex nature of these images. Another interesting approach to enhance the model performance would be to investigate the shadow augmentation strategy proposed by Xu *et al.* (2022) to encourage the model to segment the prostate while the shadow region is ignored [35].

During the qualitative and quantitative evaluation, *Model Axial* and *Model Sagittal* show a better performance compared to *Model Sag-Ax.* This indicates that the increased amount of training images used for the development of *Model Sag-Ax*, does not contribute to better segmentation performance. Conversely, the misclassification of shadow artefacts was observed more often in the predictions of *Model Sag-Ax* compared to the models that were trained on less but task-specific data (Figure 3.4 a). These results suggest that the improved performance of *Model Sagittal* and *Model Axial* partly rely on the localization of the prostate in the image whereby the prostate is usually positioned in the middle region of the axial TAUS image, or on the side of the sagittal TAUS image. As a consequence, the usability of *Model Axial* and *Model Sagittal* may be operator-specific, which can be regarded as a limiting factor when not taken into account when utilised. Further explorations are required to investigate the model performances when the prostate location varies in the images, and additional training may be required on TAUS acquisitions whereby the prostate is positioned in uncommon image regions.

The investigation on automatic prostate segmentation performed in Section 2.4 presents several

other limitations. First, all models are developed on higher-quality prostate images in which the prostate boundary is visible (Quality A/B images, Appendix B). The exclusion of poor-quality TAUS images, likely creates bias and can potentially affect future predictions on images with lower quality that are not included in the training set of this study. Given that TAUS acquisitions are operator-dependent, and the quality relies on the learning curve of the clinician and patient characteristics, a model trained solely on high-quality images may prevent the generalizability of the model. Further studies are required to analyze if the performance is affected when poor-quality images are provided.

Moreover, all models are developed on images acquired from 52 men. To ensure the generalizability of the models and study their application for clinical application, further investigation on a larger dataset, including a wider patient population and acquisitions from different US manufacturers, is desirable.

Additionally, the reference segmentations were provided by one expert. While this avoids inter-expert variability resulting in more robust outcomes, it may introduce bias. For instance, in several cases the ground truth prostate segmentation was guessed partly based on visual inspection when the prostate boundary was interrupted by the presence of artefacts. A potential improvement to avoid this bias is to create an enlarged dataset in which the prostate is segmented by multiple experts. Then, the final ground truth segmentation can be defined by the prostate pixels occurring in the segmentations of all experts.

This study provides the first insight on automatic prostate segmentation in TAUS. Our results show the ability to segment the middle prostate region with a DSC > 0.85. Compared to the variability related to manual prostate segmentation in TAUS, the developed models show a higher performance.This is a first step towards an automatic prostate delineation, from which the PV can be derived.

4.3 Automatic extraction of the prostate diameters

In this study, an algorithm is presented for the computation of the transverse, AP and longitudinal diameter, whereby a sequence of prostate segmentations in the axial and sagittal imaging-plane serve as input. Ultimately, the algorithm enables automatic PV estimation on TAUS.

To validate the algorithm design, ground truth prostate segmentations on TAUS are provided and the computed diameters are compared to manually assigned prostate diameters on TAUS. Figure 3.5 shows that the transverse and AP diameters are computed with an average difference of -0.05 $\pm$0.4 cm and 0.3 $\pm$0.3 cm respectively. For both computations, the random scatter in the Bland-Altman diagrams suggests that the variability is independent of the observed PV range. The computation of the longitudinal diameter presents an increased variability, whereby an average difference of 0.18 $\pm$1.74 cm was observed.

The variability related to the longitudinal diameter may be explained by the algorithm design, as the position of the computed diameter is derived from the maximum diameter that goes through the segmentation under an angle of 40 to 50 degrees. The provided angle range results in sufficient diameter positioning, however, the visual inspection of the extracted diameters exposed that in several acquisitions the actual prostate position deviates due to the free-hand nature of TAUS. Consequently, the longitudinal diameter must be estimated according to another angle to measure the prostate length. This makes the current design prone to inter-operator variability, regarding the TAUS examination in sagittal direction. One simple solution to address this constraint might

be to provide a larger range of angles, that would allow more deviation in prostate position. In a different manner, the longitudinal diameter can be approximated by resolving the relative position between the axial TAUS images, allowing for a reconstruction of 2D TAUS images [49].

Notwithstanding the observed variability in the computation of the prostate diameters, little variation between the computed $PV_{Axial}/PV_{Ellipsoid}$ compared to manual PV estimation on TAUS can be observed from Figure 3.6 (average difference < 1.5 ml). Hence a computation the PV according to the proposed algorithm confirms a strong level of agreement with manual PV measurements on TAUS when the input segmentations are optimal.

4.4 Automatic prostate volume estimation on transabdominal ultrasound

The combination of *Model Axial* and *Model Sagittal* and the proposed algorithm demonstrate similar capability of extracting volume measurements from TAUS as compared to manual PV measurements on MRI. This study provides insights for automatic PV estimation using TAUS as an efficient and patient-friendly alternative to current methods, thus expanding the possibilities regarding PV approximation during PCa risk -stratification workflows.

Although previous research has highlighted the value of automatic PV estimation, the majority of the studies focus on prostate segmentation models based on TRUS and MRI, where a 3D volume is generated [20], [26]. PV estimation with the aid of TAUS has not received sufficient attention, primarily due to the required examination training and for the existing variability in image quality, limiting the use of deep neural networks. Additionally, 3D volume reconstruction of prostate segmentations is not (yet) possible in free-hand ultrasound.

Previously, one study on PV estimation using TAUS, was able to approximate the PV with an average volume difference of 6.22 ml $\pm$7.2 ml compared to reference PV measurements based on MRI (Section 1.2.4).Nonetheless, the proposed method neglected the determination of the mid-plane prostate image, and requires therefore operator input. In our study this limitation was addressed, by developing *Model Axial* and *Model Sagittal* to segment the prostate when a sequence of prostate TAUS images are provided in the respective image-plane. The sequence of prostate segmentations in both directions serve as input for our algorithm to extract the prostate diameters whereby determination of the mid-plane prostate segmentation is incorporated, enabling a fully automatic framework.

The performance of the framework was evaluated across 17 test participants, and in the majority of the cases (14/17), the prediction of $PV_{Ellipsoid}$ deviates with less than $\leq$25% compared to MRI reference volumes. On average, a difference of 2.5 $\pm$10.2 ml was reported, shown in Figure 3.7a and the following section enlightens two cases in which the largest over and underestimation was observed.

In the first case, an underestimation of 14 ml was observed (relative deviation compared to MRI: -32%). Further investigation showed accurate predictions of *SegAx* and *SegSag*, whereby the mid-plane was defined properly. In other words, the segmentation step of the framework narrowly contributed to discrepancy in $PV_{Ellipsoid}$ outcomes. When the diameter computation was compared to the manual diameter assessment, a difference of 0.025, 0.02 and 0.18 cm was observed for the transverse, AP and longitudinal diameter. The small deviation for both steps of the framework show that the difference in PV can be attributed to the employment of the Ellipsoid Formula on TAUS. Notably, the predicted PV differed from the manual PV estimation on TAUS with 1 ml. This confirms the ability of accurately calculating the mid-prostate images and the corresponding

diameters. Yet, considering the existing variability of PV estimation when ultrasound is used, the proposed framework appears less reliable.

Additionally, an overestimation of 15 ml was reported, through which a PV prediction of 47 ml was obtained (relative deviation compared to MRI: +47%). In this case, the manual PV estimation on TAUS was 76 ml. The discrepancy in the manual PV estimation on TAUS, the prediction and the reference PV measurement on MRI was mostly the result of shadow artefacts limiting the prostate boundary. As a consequence, the longitudinal diameter was overestimated by the operator, as part of the shadow region was false positively included in the diameter assessment. In a lesser extent, a similar result was observed in the predicted $PV_{Ellipsoid}$, whereby the longitudinal diameter was overestimated by the algorithm. Limited prostate visibility in sagittal TAUS images remain a limitation of $PV_{Ellipsoid}$, and this case shows how this phenomenon is translated in the proposed framework. To address this challenge, the feasibility of using solely axial prostate diameters is explored. PV_{Axial} offers several advantages, as a single segmentation model is required. This results in a reduced patient examination duration and less experience required for the clinicians to get acquisition of acceptable quality. The prediction outcomes show that PV Axial is obtained with an increased SD of 0.6 ml compared to a prediction based on $PV_{Ellipsoid}$. This suggest the possibility of using the prostate in one imaging-plane for the approximation of the PV depending on the required LoA.

4.4.1 Clinical application

Currently, it has been shown that a PV classification of PV $<$ 30 ml, PV = 30 - 50 ml and PV $>$ 50 ml hardly affects the predictive capability of PCa risk-stratification [50]. Based on our final evaluation, for both the predicted $PV_{Ellipsoid}$, and PV_{Axial} , 16/17 participants are classified correctly when this LoA is persisted. Additionally, an accuracy, deviating with $\leq$25% from the ground truth is thought to be sufficient in the first stage of PCa risk-stratification [18]. To this extent, the proposed framework may benefice prostate cancer screening when it is centralized in a regional care pathway. This has been established in Rotterdam by the Urology Department of the Erasmus University Medical Center in collaboration with the primary care laboratory STAR-SHL, where screening consultations are performed with multi-variable risk stratification to reduce redundant hospital referrals [51]. In this setting, an approximation of the PV with the aid of a handheld ultrasound system may be convenient. Even so, it is recommended to investigate the performance of the framework on TAUS acquisitions of the respective manufacturer. The employment of TAUS in combination with an automatic PV measurement, must be a straightforward approach and easy to use for every operator. Therefore, it is suggested to expand the framework, with a real time prostate detection system, to ensure appropriate input acquisitions for optimal PV approximation.

4.4.2 Future work

In order to adopt the framework as standard of care, further research is required. Therefore, a larger cohort needs to be analysed to investigate the performance of the framework on different target groups. Additionally, the framework must be tested on poor quality acquisitions and the intended use of the product must be clearly explained: essentially, the framework serves as an aid to diagnostics, whereby the PV can be approximated in an accessible and simple manner. Additionally, it is recommended to explore the performance of the developed segmentation models on TAUS images of different imaging manufacturers. More specifically, the framework must be tested when a handheld ultrasound device is utilised. Thereafter the contribution of the proposed

framework is suggested to be explored to understand if redundant hospital referrals are reduced when it is utilised in an earlier stage of diagnostics.

Conclusions

Accurate determination of the PV is essential for PCa risk-stratification and therefore routinely obtained in during an early stage of diagnostics. This study explores the feasibility of TAUS to approximate the PV as a more patient-friendly, accessible and inexpensive approach, compared to TRUS and MRI. The results in this study indicate that it is possible to obtain PV measurements based on TAUS examination, that are similar to PV measurements on MRI. The existing variability of PV measurements between TAUS and MRI seems unrelated to the TAUS image quality, suggesting that manual PV assessment can be performed, even when unfavorable patient characteristics limit the image quality (e.g. patients with high abdominal fat percentages). In contrast to TRUS, TAUS seems less prone to inaccuracies when larger PVs are encountered. Yet, when TAUS examination is conducted improperly, shadow artefacts induced by the reflection of the pelvic bone, may limit the prostate's visibility, resulting unreliable PV measurements. Consequently, proper operator training of TAUS examination and image interpretation is essential.

To minimize the initial complexity of PV estimation based on TAUS and to ensure robust PV measurements, this study proposed a framework for automatic PV estimation based on input TAUS acquisitions. The main components of the framework comprise: *Model Axial* and *Model Sagittal*, two segmentation models that predict the prostate area in plane-specific TAUS images, and an algorithm that extracts the prostate's diameters based on the predicted prostate segmentations. All components were evaluated individually and show good correspondence to manual prostate delineations, especially when the mid-prostate region is assessed, and diameter assessment on TAUS, ensuring a reliable automated framework.

Final tests on TAUS image data of unseen participants showed that is possible to achieve automated PV measurements based on TAUS, comparable to manual PV measurements based on MRI. In the majority of the test participants the predicted PV deviated with less than 25% compared to the ground truth. Moreover, in the first stage of PCa diagnostics, a PV classification of PV $<$ 30 ml, PV = 30-50 ml and PV $>$ 50 ml ensures a reasonable predictive capability for PCa risk-stratification. In our final evaluation, 16/17 test participants were classified correctly when this LoA is persisted. Thereby, the proposed framework may benefice PCa risk stratification workflows, and expands the possibilities regarding PV estimation using TAUS.

In order to adapt the framework for clinical settings, further research on larger cohorts is required to ensure the generalizability of the method. Additionally it is recommended to evaluate its performance on poor quality TAUS images and images from different US manufacturers.

Ultimately, the framework must serve as an aid to diagnostics whereby the PV is approximated in an accessible and simple manner. For this reason, it is recommended to expand it with a real time prostate detection system, to ensure appropriate input TAUS images when utilised by inexpert TAUS operators.

A

Prostate Imaging-Reporting and Data System

- **PI-RADS 1:** Very low. Clinically significant cancer is very unlikely to be present.
- **PI-RADS 2:** Low. Clinically significant cancer is unlikely to be present.
- **PI-RADS 3:** Intermediate. The presence of clinically significant cancer is equivocal.
- **PI-RADS 4:** High. Clinically significant cancer is likely to be present.
- **PI-RADS 5**: Very High. Clinically significant cancer very likely to be present.[52]

B

Quality assessment TAUS acquisitions

Each acquisition is assessed on image quality in which the following parameters are taken into account:

Present Prostate Volume (PPV): To what extent is the prostate volume visible in the acquisition. The acquisition can be scored on a PPV of 0%, 25%,50%, 75%, 100%.

Visibility Prostate Boundary (VPB): To what extent is the prostate visible in the acquisition. **A:** The majority of the boarders are clear in the acquisition. **B:** More or less clear boarders throughout the acquisition. **C:** The majority of boarders are hard to interpret in the acquisition.

Artefacts Interrupting the Prostate (AIP): To what extent are imaging artefacts present in the acquisition leading to a reduced PPV / VPB. For instance, shadow artefacts that interrupt the prostate boundary may lead to a decrease in PPV. The VPB may decrease in patients with a high abdominal fat percentage. Additionally, the PPV can be reduced due to existing acoustic shadow artefacts that occur when the pelvic bone reflects the ultrasound signal.

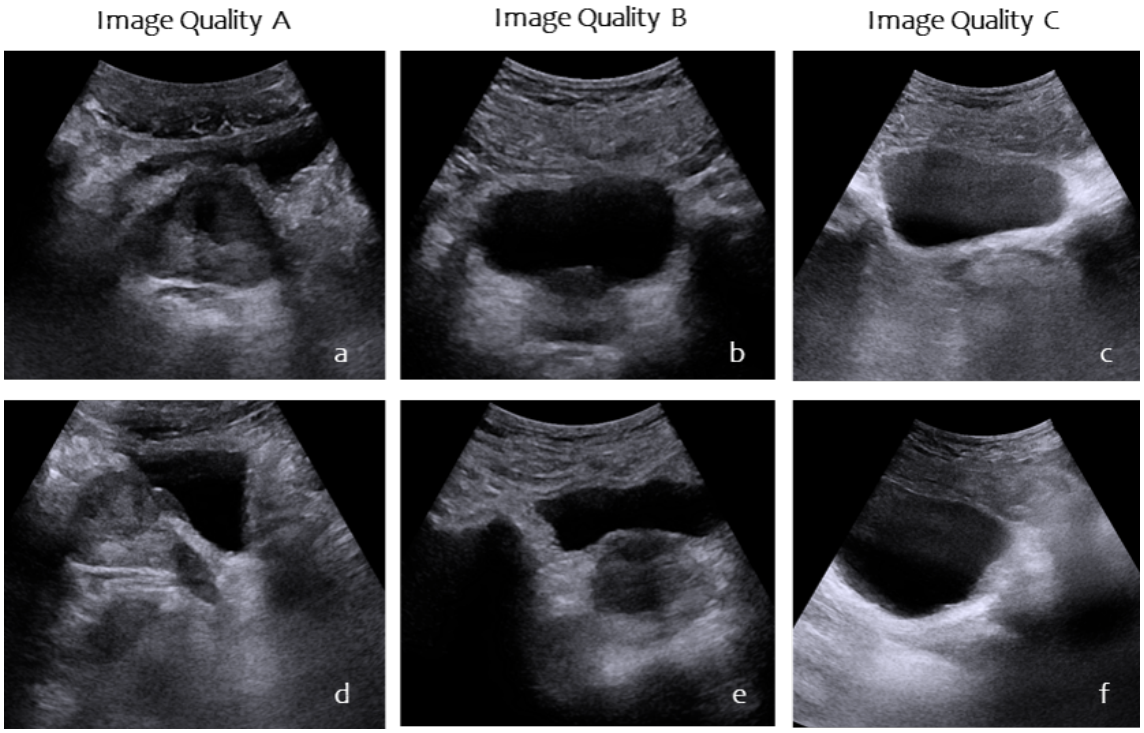


Figure B.1: The prostate on transabdominal ultrasound images with diverse image quality. **a:** Axial quality A image. **b:** Axial quality B image. **c:** Axial quality C image. **d:** Sagittal quality A image. **e:** Sagittal quality B image. **f:** Sagittal quality C image.

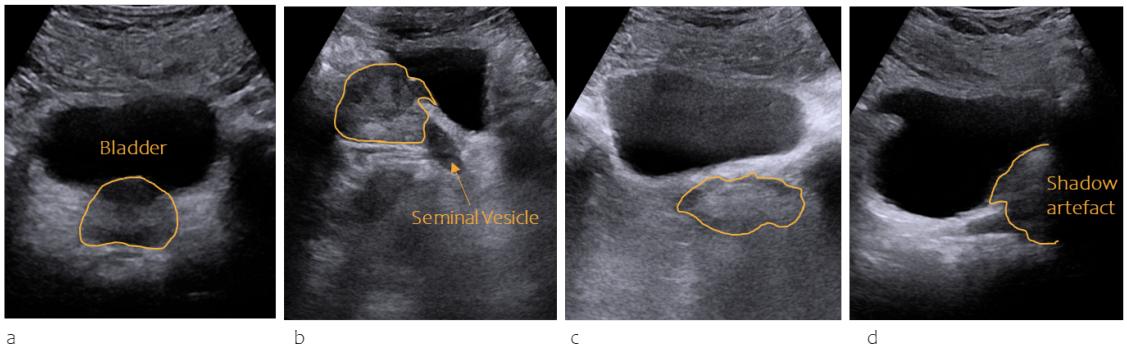


Figure B.2: Various artefacts interrupting the prostate vision on TAUS. The prostate is delineated manually with a yellow line. **a:** The axial imaging-plane, clear visibility of the prostate. The bladder is annotated. **b:** The sagittal imaging-plane, clear visibility of the prostate with bulging contours that interrupt the bladder. The seminal vesicle is annotated. **c:** The axial imaging-plane, poor visibility of the prostate due to ambiguous boundaries. **d:** The sagittal imaging-plane, part of the prostate boundary is invisible due to interrupting acoustic shadow artefacts.

Bibliography

[1] M. Blanker, *Richtlijnen database: Prostaatkanker*, Mar. 2024. [Online]. Available: https://richtlijnen.nhg.org/standaarden/prostaatkanker.

[2] M. J. Roobol, E. W. Steyerberg, R. Kranse, *et al.*, "A risk-based strategy improves prostate-specific antigen–driven detection of prostate cancer," *European urology*, vol. 57, no. 1, pp. 79–85, 2010.

[3] F. K.-H. Chun, T. Steuber, A. Erbersdobler, *et al.*, "Development and internal validation of a nomogram predicting the probability of prostate cancer gleason sum upgrading between biopsy and radical prostatectomy pathology," *European urology*, vol. 49, no. 5, pp. 820–826, 2006.

[4] P. J. van Leeuwen, A. Hayen, J. E. Thompson, *et al.*, "A multiparametric magnetic resonance imaging-based risk model to determine the risk of significant prostate cancer prior to biopsy," *BJU international*, vol. 120, no. 6, pp. 774–781, 2017.

[5] A. Morlacco, D. Modonutti, G. Motterle, F. Martino, F. D. Moro, and G. Novara, *Nomograms in urologic oncology: Lights and shadows*, 2021. DOI: 10.3390/jcm10050980.

[6] D. W. Bruner, D. Moore, A. Parlanti, J. Dorgan, and P. Engstrom, "Relative risk of prostate cancer for men with affected relatives: Systematic review and meta-analysis.," *International journal of cancer*, vol. 107, pp. 797–803, 5 Dec. 2003, ISSN: 0020-7136. DOI: 10.1002/ijc.11466.

[7] R. Marima, M. Mbeje, R. Hull, D. Demetriou, N. Mtshali, and Z. Dlamini, "Prostate cancer disparities and management in southern africa: Insights into practices, norms and values," *Cancer Management and Research*, pp. 3567–3579, 2022.

[8] H. Van Poppel, R. Hogenhout, P. Albers, R. C. van den Bergh, J. O. Barentsz, and M. J. Roobol, "A european model for an organised risk-stratified early detection programme for prostate cancer," *European urology oncology*, vol. 4, no. 5, pp. 731–739, 2021.

[9] I. Yusim, M. Krenawi, E. Mazor, V. Novack, and N. J. Mabjeesh, "The use of prostate specific antigen density to predict clinically significant prostate cancer," *Scientific Reports*, vol. 10, 1 2020, ISSN: 20452322. DOI: 10.1038/s41598-020-76786-9.

[10] M. J. Roobol, H. A. van Vugt, S. Loeb, *et al.*, "Prediction of prostate cancer risk: The role of prostate volume and digital rectal examination in the erspc risk calculators," *European urology*, vol. 61, no. 3, pp. 577–583, 2012.

[11] K. T. McVary and S. A. Kaplan, *A tower of babel in today's urology: Disagreement in concepts and definitions of lower urinary tract symptoms/benign prostatic hyperplasia re-treatment*, 2020. DOI: 10.1097/JU.0000000000001073.

[12] J. F. Tyloch and A. P. Wieczorek, "The standards of an ultrasound examination of the prostate gland. part 2," *Journal of ultrasonography*, vol. 17, no. 68, pp. 43–58, 2017.

[13] J. A. Villanueva Herrero, A. Abdussalam, and A. Kasi, *Rectal exam*, 2023. [Online]. Available: http://europepmc.org/books/NBK537356.

[14] F. R. Romero, K. R. P. S. Romero, T. Brenny Filho, R. Pilati, D. Kulysz, and F. C. de Oliveira Júnior, "Reasons why patients reject digital rectal examination when screening for prostate cancer," *Archivos Españoles de Urología*, vol. 61, no. 6, pp. 759–765, 2008.

[15] S. Walker *et al.*, *Prostate cancer support groups: An evaluation*. Department of Psychology, Swinburne University of Technology, 2005.

[16] W. R. Pate, N. Garg, L. B. Wang, S. E. Wason, and P. V. Barbosa, "Comparison of transabdominal and transrectal ultrasound for sizing of the prostate," *Urology*, vol. 141, 2020, ISSN: 15279995. DOI: 10.1016/j.urology.2020.04.054.

[17] S. Guo, J. Zhang, J. Jiao, *et al.*, "Comparison of prostate volume measured by transabdominal ultrasound and mri with the radical prostatectomy specimen volume: A retrospective observational study," *BMC Urology*, vol. 23, 1 2023, ISSN: 14712490. DOI: 10.1186/s12894-023-01234-5.

[18] I. I. de Vos, F.-J. H. Drost, L. P. Bokhorst, *et al.*, "Prostate cancer risk assessment by the primary care physician and urologist: Transabdominal-versus transrectal ultrasound prostate volume-based use of the rotterdam prostate cancer risk calculator," *Translational Andrology and Urology*, vol. 12, no. 2, p. 241, 2023.

[19] J. S. Lee and B. H. Chung, "Transrectal ultrasound versus magnetic resonance imaging in the estimation of prostate volume as compared with radical prostatectomy specimens," *Urologia Internationalis*, vol. 78, 4 2007, ISSN: 00421138. DOI: 10.1159/000100836.

[20] S. Ghose, A. Oliver, R. Martí, *et al.*, "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images," *Computer Methods and Programs in Biomedicine*, vol. 108, 1 2012, ISSN: 01692607. DOI: 10.1016/j.cmpb.2012.04.006.

[21] N. Betrouni, M. Vermandel, D. Pasquier, S. Maouche, and J. Rousseau, "Segmentation of abdominal ultrasound images of the prostate using a priori information and an adapted noise filter," *Computerized Medical Imaging and Graphics*, vol. 29, 1 2005, ISSN: 08956111. DOI: 10.1016/j.compmedimag.2004.07.007.

[22] R. N. Uppot, "Technical challenges of imaging & image-guided interventions in obese patients," *The British journal of radiology*, vol. 91, no. 1089, p. 20 170 931, 2018.

[23] J. F. Tyloch and A. P. Wieczorek, "The standards of an ultrasound examination of the prostate gland. part 1," *Journal of Ultrasonography*, vol. 16, no. 67, pp. 378–390, 2016.

[24] B. Wiestler and B. Menze, "Deep learning for medical image analysis: A brief introduction," *Neuro-oncology advances*, vol. 2, no. Supplement_4, pp. iv35–iv41, 2020.

[25] Y. Wang, H. Dou, X. Hu, *et al.*, "Deep attentive features for prostate segmentation in 3d transrectal ultrasound," *IEEE transactions on medical imaging*, vol. 38, no. 12, pp. 2768–2778, 2019.

[26] J. Jiang, Y. Guo, Z. Bi, Z. Huang, G. Yu, and J. Wang, "Segmentation of prostate ultrasound images: The state of the art and the future directions of segmentation algorithms," *Artificial Intelligence Review*, vol. 56, no. 1, pp. 615–651, 2023.

[27] F. Pellegrino, A. L. Tin, A. Martini, *et al.*, "Prostate-specific antigen density cutoff of 0.15 ng/ml/cc to propose prostate biopsies to patients with negative magnetic resonance imaging: Efficient threshold or legacy of the past?" *European urology focus*, vol. 9, no. 2, pp. 291–297, 2023.

[28] S. Remmers, D. Nieboer, L. L. Rijstenberg, T. Hansum, G. J. van Leenders, and M. J. Roobol, "Updating the rotterdam prostate cancer risk calculator with invasive cribriform and/or intraductal carcinoma for men with a prior negative biopsy," *European Urology Open Science*, vol. 36, pp. 19–22, 2022.

[29] S. Y. Youn, M. H. Choi, Y. J. Lee, *et al.*, "Prostate gland volume estimation: Anteroposterior diameters measured on axial versus sagittal ultrasonography and magnetic resonance images," *Ultrasonography*, vol. 42, 1 2023, ISSN: 22885943. DOI: 10.14366/usg.22104.

[30] S. Aprikian, M. Luz, F. Brimo, *et al.*, "Improving ultrasound-based prostate volume estimation," *BMC urology*, vol. 19, 1 2019, ISSN: 14712490. DOI: 10.1186/s12894-019-0492-2.

[31] I. Ajayi, A. Aremu, A. Olajide, T. Bello, F. Olajide, and V. Adetiloye, "Correlation of transrectal and transabodominal ultrasound measurement of transition zone volume with postoperative enucleated adenoma volume in benign prostatic hypertrophy," *The Pan African Medical Journal*, vol. 16, 2013.

[32] S. H. Kim and S. H. Kim, "Correlations between the various methods of estimating prostate volume: Transabdominal, transrectal, and three-dimensional us," *Korean Journal of Radiology*, vol. 9, no. 2, p. 134, 2008.

[33] N. B. Albayrak and Y. S. Akgul, "Estimation of the prostate volume from abdominal ultrasound images by image-patch voting," *Applied Sciences*, vol. 12, p. 1390, 3 Jan. 2022, ISSN: 2076-3417. DOI: 10.3390/app12031390.

[34] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, Springer, 2015, pp. 234–241.

[35] X. Xu, T. Sanford, B. Turkbey, S. Xu, B. J. Wood, and P. Yan, "Shadow-consistent semi-supervised learning for prostate ultrasound segmentation," *IEEE Transactions on Medical Imaging*, vol. 41, pp. 1331–1345, 2022.

[36] D. Karimi, Q. Zeng, P. Mathur, *et al.*, "Accurate and robust deep learning-based segmentation of the prostate clinical target volume in ultrasound images," *Medical image analysis*, vol. 57, pp. 186–196, 2019.

[37] C. Beitone and J. Troccaz, "Multi-expert fusion: An ensemble learning framework to segment 3d trus prostate images," *Medical Physics*, vol. 49, no. 8, pp. 5138–5148, 2022.

[38] C. B. medical, *Bk medical real time active imaging for decision support*, 2024. [Online]. Available: https://www.bkmedical.com.

[39] P. Watson and A. Petrie, "Method agreement analysis: A review of correct methodology," *Theriogenology*, vol. 73, no. 9, pp. 1167–1179, 2010.

[40] H. Akoglu, "User's guide to correlation coefficients," *Turkish journal of emergency medicine*, vol. 18, no. 3, pp. 91–93, 2018.

[41] C. M. Corporation, *Microsoft excel*, 2024. [Online]. Available: https://www.microsoft.com/.

[42] T. S. Community, *Https://docs.scipy.org/doc/scipy/reference/stats.html*, 2008-2024.

[43] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, *et al.*, "3d slicer as an image computing platform for the quantitative imaging network," *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.

[44] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "Nnu-net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[45] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, vol. 14, no. 1, p. 6086, 2024.

[46] G. J. Matthews, J. Motta, and J. Fracchia, "The accuracy of transrectal ultrasound prostate volume estimation: Clinical correlations," *Journal of clinical ultrasound*, vol. 24, no. 9, pp. 501–505, 1996.

[47] T. Natali, M. Wijkhuizen, L. Kurucz, *et al.*, "Automatic real-time prostate detection in transabdominal ultrasound images," in *Medical Imaging with Deep Learning*, 2024. [Online]. Available: `https://openreview.net/forum?id=jh1KZ4cyWA`.

[48] S. Y. Youn, M. H. Choi, Y. J. Lee, *et al.*, "Prostate gland volume estimation: Anteroposterior diameters measured on axial versus sagittal ultrasonography and magnetic resonance images," *Ultrasonography*, vol. 42, no. 1, p. 154, 2023.

[49] H. Guo, H. Chao, S. Xu, B. J. Wood, J. Wang, and P. Yan, "Ultrasound volume reconstruction from freehand scans without tracking," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 3, pp. 970–979, 2022.

[50] N. Pereira-Azevedo, I. Braga, J. F. Verbeek, *et al.*, "Prospective evaluation on the effect of interobserver variability of digital rectal examination on the performance of the rotterdam prostate cancer risk calculator," *International Journal of Urology*, vol. 24, no. 12, pp. 826–832, 2017.

[51] D. F. Osses, A. R. Alberts, G. C. Bausch, and M. J. Roobol, "Multivariable risk-based patient selection for prostate biopsy in a primary health care setting: Referral rate and biopsy results from a urology outpatient clinic," *Translational Andrology and Urology*, vol. 7, no. 1, p. 27, 2018.

[52] A. B. Rosenkrantz, S. Kim, R. P. Lim, *et al.*, "Prostate cancer localization using multiparametric mr imaging: Comparison of prostate imaging reporting and data system (pi-rads) and likert scales," *Radiology*, vol. 269, no. 2, pp. 482–492, 2013.

C

The Application of deep-learning for prostate volume estimation on ultrasound: A Literature Review

Applied Deep-Learning for Prostate Volume Estimation on Ultrasound: A Literature Review

L.M. Kurucz

Image-Guided Surgery, Netherlands Cancer Institute – Antoni van Leeuwenhoek

Abstract. The high incidence of prostate cancer led to the implementation of screening strategies to demote disease burden. Prostate volume (PV) plays a crucial role in managing prostate-related diseases, particularly in the risk-stratification of prostate cancer. In practice, diagnostic ultrasound (US) is commonly utilized to obtain the PV. To alleviate the clinician's workload and to reduce patient examination duration, automated methods for PV estimation are required. Hence, this review focusses on the application of deep-learning (DL) for PV estimation, by means of prostate segmentation on US or the employment of the Prolate Ellipsoid Formula (PEF). Currently, proposed DL architectures are able to segment the prostate with a Dice Similarity Score (DSC) that ranges between 0.87 and 94%. Strong implementations in this field include the employment of multi-directional image data, the implementation of attention mechanisms and novel pre-processing techniques to combat frequently seen US image artefacts. Unfortunately, most proposed methods still suffer from inconsistent results that prevent clinical implementation. Limitations of the proposed explorations frequently comprise the use of homogeneous datasets, no implementation of ablation studies and the lack of statistical significance due to the limited availability of US data. Finally, the absence of public datasets complicate the comparison of DL architectures in this field and are therefore required.

1. Introduction

Prostate cancer is the second most commonly diagnosed cancer in men and accounts for the third cause of cancer death. (1) Due to the high prevalence of prostate cancer, the implementation of early detection and screening has been widespread since the late 1980s waiming to demote disease burden. (2) In multiple stages of prostate disease management, the prostate is evaluated carefully with routine measurement of the prostate volume (PV). (3) PV is an important parameter as it is used to obtain prostate-specific antigen density to determine the risk of prostate cancer. (4) When a more invasive intervention like brachytherapy is required, accurate determination of the prostate's clinical target volume supports the development of an effective therapeutic plan and PV is a critical prerequisite to calculate dose distribution. (5) Additionally, PV is acquired for the diagnosis of benign prostate diseases such as lower urinary tract symptoms or benign prostatic hyperplasia. (6)

In clinical practice, physicians rely on the aid of computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound (US) to evaluate the prostate on malignancy or volume. However, the choice of the most suitable imaging option depends on the clinical purpose as each modality offers distinct characteristics. PV estimation based on MRI-segmentation is considered as gold standard owing to the better signal-to-noise ratio. Nonetheless, transrectal ultrasound (TRUS) is commonly utilized in clinical settings due to its cost-effectiveness, portability, and ability to rapidly estimate the PV. (7) The evaluation of the prostate based on transabdominal ultrasound (TAUS) is another feasible option, often preferred since it is a percutaneous and more comfortable approach. (8) In regard to stratifying the risk of prostate cancer, estimating PV with the aid of TAUS can be of additional value, given that the examination can take place at an earlier stage of diagnosis (e.g. at the General Practitioner).

Notwithstanding the favorable features of US, accurate PV estimation remains challenging due to difficulties associated with the interpretation of US images. Frequently performed tasks to obtain the PV, like prostate segmentation, annotation of the prostate's outer boundary landmarks or diameter measurements often result in over/under estimations and are prone to inter/intra-observer variability. (9) This is primarily because of ambiguous boundaries resulting from poor contrast between the prostate and surrounding tissue. Additionally, the assessment is complicated by misleading imaging artefacts such as echoic shadows and the presence of calcifications. Notably in TAUS, the prostate can be difficult to observe when shadow artefacts disrupt the prostate boundary as a result of ultrasonic reflection by the pelvic bone. (10, 11) The inhomogeneous intensity distribution of the prostate tissue and the large shape variations also complicate the assessment and are often considered most demanding in the apex and base of the prostate. Finally, prostate annotations are generally conducted during US examination, adding complexity to the procedure.

Hence, computer-aided methods have been introduced aiming to alleviate the workload on clinicians, decrease examination duration and improve the accuracy of segmentation and annotations. Figure 1 depicts examples of the prostate image obtained by TAUS, TRUS and MRI, to highlight the differences obtained by all three imaging modalities.

There has been a notable increase in the demand for automated prostate segmentation methods, which allow for the derivation of the PV. Initially, deformable models, edge-based segmentation methods, region-based segmentation methods have been introduced and are discussed in multiple overviews on prostate segmentation. (10, 12) However, all aforementioned segmentation methods predominantly require the extraction of prostate related image features such as the relative pixel intensity or prostate shape information. Especially when used in US, these features affected by the large contrast diversity and US artefacts. . Consequently, the generation of reliable segmentations on US remains challenging and this paper aims to cover recent advances in this field. In the past years, deep-learning (DL) architectures that process the entire image as part of their training, have been introduced as an alternative way to assess the prostate automatically. Compared to other machine learning methods, the main advantage of DL is its ability to automatically extract and use multi-level features that consist of abundant semantic as well as detailed image information. In spite of the promising aspects of DL, existing methods still suffer from inaccuracy in prostate boundary extraction. (13) Therefore experts on the topic explore the implementation of DL architectures to improve prostate segmentation. For clarity, in this review, DL architecture refers to the entire segmentation pipeline that includes the used neural network and its additional implementations. With this pursuit in mind, the purpose of this review is to provide insights in academic literature of the field, to better understand the advancements of DL architectures for the accurate determination of PV in diagnostic US imaging. In summary, this paper presents:

1. A systematic search to include developed DL architectures for prostate segmentation and PV estimation on US images.

2. An overview of proposed methods, including a quantitative comparison of results.

3. Multiple designs are enlightened and evaluated with respect to context.

4. Suggestions for further research are provided.

The paper is organized as follows. The search methodology is explained in Section 2. In Section 3, the results are provided, subdivided in the application of DL for prostate segmentation on US and the application of DL to estimate the PV, utilizing the Prolate Ellipsoid Formula (PEF). In regard to segmentation, explicit decisions aiming to improve DL architectures are explained. The following subdivision is maintained: US pre-processing methods, the employment of multi-directional image data, the implementation of additional shape information, introducing attention mechanisms, feature map refinement, and the consistency and robustness of DL architectures. Thereafter, a quantitative comparison of results is provided between DL architectures that focus on prostate segmentation. Finally, the review ends with a discussion in Section 4.

Figure 1 Cross-section of the prostate on different imaging modalities, where the prostate is manually delineated with a yellow color line. a Transabdominal Ultrasound. b Transrectal Ultrasound. (42) c Magnetic Resonance Imaging (gold-standard for prostate evaluation).

2. Search methodology

The core search concepts of this review were DL, PV and US. A search was conducted in October 2023 on three scientific platforms: PubMed (14), IEEE Xplore (15) and Scopus (16) using the search queries included in Appendix A. The inclusion and exclusion criteria were defined a priori. During title and abstract screening, articles were selected if they met the following inclusion criteria: 1) Studies reporting on PV estimation or prostate segmentation using DL in US. In this manner, acquisitions based on TRUS and TAUS were both covered. 2) Studies reported on registration of US images were included as segmentation is frequently required in registration pipelines. 3) Studies written in the English language and published from 2016. In regard to exclusion criteria, articles were excluded when the papers focused on other organs than the prostate or when they were unrelated to US. Finally, invalid

records, reviews and duplicates were excluded as well.

During full-text screening, articles were included when the proposed segmentation architectures were evaluated on performance. Finally, articles were excluded when a proposed registration pipeline dit not comprise an automatic segmentation step, or when the segmentation was based on other imaging modalities than US.

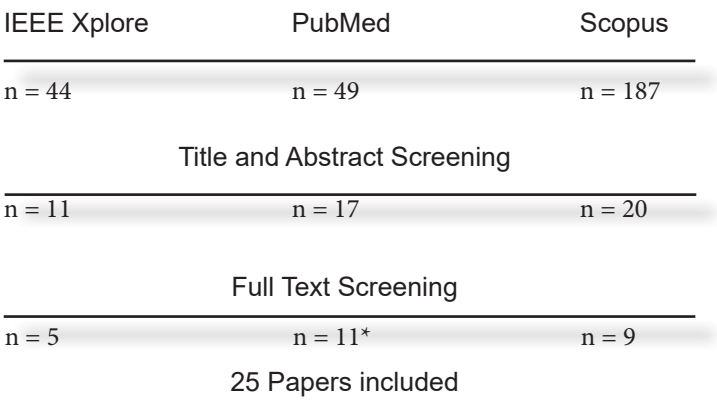


Figure 2 Flow diagram of the search process. Ultimately, 25 papers were included. *The continuation of (17), (48) was also included, although it was not available in explored scientific platforms.

To evaluate DL architectures that were covered in this review, several architecture parameters were retrieved to illustrate the context in which the DL architectures are proposed: 1) The purpose of the DL-architecture with respect to 2D or 3D prostate segmentation. 2) The amount of US data used for training and testing. 3) The amount of patients that comprise the dataset 4) The presented results concerning the used evaluation metrics.

Finally, innovations of the DL architecture that primarily played a role in an increased performance were explained and compared. All stages of screening were done by an independent reviewer (LMK) and senior authors (TN, MF, BDB, and JV) were addressed when necessary.

3. Results

3.1 Study selection

Based on the search in PubMed, IEEE Xplore and Scopus, a total of 280 studies were retrieved. After title and abstract screening, 61 studies satisfied the inclusion and exclusion criteria. When the following studies were screened thoroughly on full text, 24 studies were included. Aside from the systematic search process, one paper was included as well as it was thought to enhance the breadth of this review. This is a continuation of (17) that was not available in Scopus, PubMed or IEEE Xplore. Hence, 25 papers on advances in DL for PV estimation and prostate segmentation on US were included. Figure 2 illustrates the flow diagram depicting the number of papers identified at each step of the search process.

3.2 The application of deep-learning for prostate segmentation on ultrasound

Currently, automated PV estimation primarily relies on prostate segmentation methods based on TRUS. Based on utilized search strategy, there is no available literature concerning prostate segmentation based on TAUS. In general, DL architectures in this review use a traditional U-Net (18) as the back-bone for the proposed DL architecture. Several authors modified the traditional U-Net by implementing residual connections or substituting the encoding path with another model. However, the majority of architectural designs incorporate additional modules aimed at improved segmentation results and are discussed in Section 3.2.1-3.2.7. Table 1 provides a brief summary of explorations that focus on prostate segmentation on TRUS.

3.2.1 Ultrasound pre-processing methodologies

Pre-processing techniques aimed at improving the image readability have been explored since many imaging artefacts inherent of US result in inaccurate prostate segmentations. (19,20) For instance, dark image areas, which are referred to as acoustic shadows, may arise as structures that have a high echogenicity (e.g. skeleton, calcifications) reflect the US wave signal.

To alleviate this problem in prostate segmentation, Xu et al. (2022) designed two novel mechanisms to encourage a network to cope with shadow regions on image and feature level. (19) More specifically, shadow artefacts were artificially added in training images to enrich the shadow diversity of the data set. This procedure served as an innovative data augmentation strategy aiming to increase the robustness of the network on prostate images that suffer from shadow artefacts. The authors simulated artefacts by extracting the shadow region of other images using soft thresholding. Then, additional training data was created by fusing the shadow artefacts of one image to another image. (Figure 3). Furthermore, the shadow features were subtracted from the created feature maps. (Figure 4) This results in a feature space that ignores shadow regions and can be seen as an approach to conduct a systematic drop-out layer. In this way, the model is encouraged to learn the prostate boundary using the remaining shadow-free features.

When the effectiveness of both methods was evaluated as an addition to a traditional U-Net architecture their results showed that both methods increased the Dice Similarity Coefficient (DSC) with 1.29% and reduced de Hausdorff Distance (HD) with 0.33mm.

Another applied method to enhance the quality of training data is to pre-process US images with handcrafted filters. Lei et al. (2019) explored the use of 3D Mean, Median and Gaussian filters to enhance the

feature extraction of the prostate boundary in their DL architecture. (20) However, this pre-processing step had no significant effect when DSC, Precision and Recall were evaluated. Notably, the pre-processed data, decreased the model performance significantly in terms of HD. (HDoriginal = 3.911 mm, HDpre-process = 3.996 mm) This indicates that the use of these particular filters in this design is not applicable to improve segmentation results.

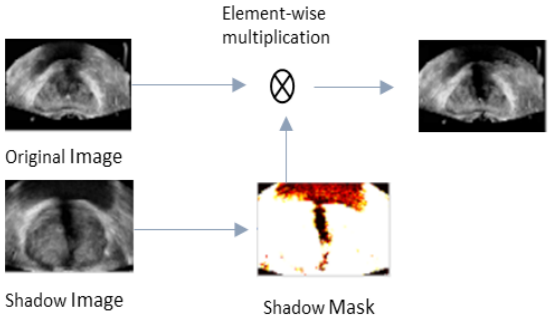
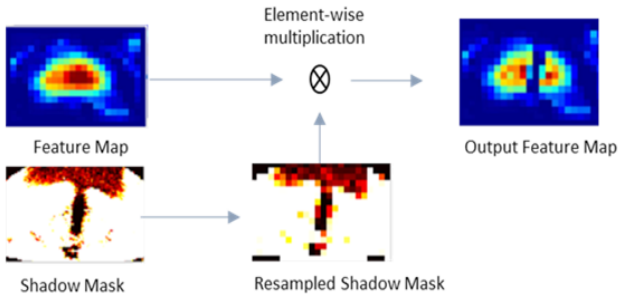


Figure 3 The implementation of shadow artefacts of one image to another image to conduct Shadow Augmentation designed by (19).



b

Figure 4 The subtraction of shadow features to enhance the attention to the prostate region. (19).

from all directions to improve segmentation results. (21) The authors designed a framework in which three 2D U-Nets were trained on axial, sagittal and coronal TRUS images to produce view-specific segmentation volumes of the prostate. The resulting volumes were intended to serve as input for a subsequent fusion network. In this network, the view-specific segmentations and the original volume were used to generate view-specific confident maps upon which the final segmentation was created.

Similarly, Lei et al. (2019) trained three view-specific V-Nets and introduced a Multi-Directional Contour Refinement (MDCR) method to post-process the view-specific volumes to estimate the PV. (20) The core concept of this post-processing method is to define the final boundary contour by taking the mean of boundary intersection coordinates of axial, sagittal and coronal directions. In this way, information from all directions is embedded in the final segmentation. When transverse segmentations of the apex and base are lacking in accuracy, the contour refinement is implemented differently by taking the average of the sagittal and coronal segmentation only. After evaluating the addition of MDCR to the DL architecture, it was shown that the DSC-score increased from 0.912 to 0.919 ($P < 0.001$).

Orlando et al. (2020) proposed a novel 3D segmentation method by utilizing a modified U-Net to segment twelve 2D radial slices, which were subsequently used for a 3D volume reconstruction. (22) The decision to obtain radial prostate segmentations, opposed to transverse segmentations, was motivated by the experience that the prostate presents more clearly when the center of the gland is in plane. Figure 5 illustrates the difference between radial, sagittal and axial image slices. During their exploration, a modified U-Net was trained on 6773 2D images that were obtained randomly at axial, coronal, sagittal and radial image planes. Thereafter, the resulting 12 segmentations were placed in the initial 3D space, and the 3D prostate boundary was reconnected following the reconstruction method proposed by (23).

3.2.2 The employment of multi-directional image data

The employment of multi-directional images for prostate segmentation can be useful as the visibility of the prostate varies locally. For instance, the apex and base of the prostate are characterized with ambiguous boundaries on transverse US images, while it appears more clearly in the sagittal and coronal direction. This is the rationale behind the Multi-eXpert Fusion (MXF) framework proposed by Beitone et al. (2022) that uses image input

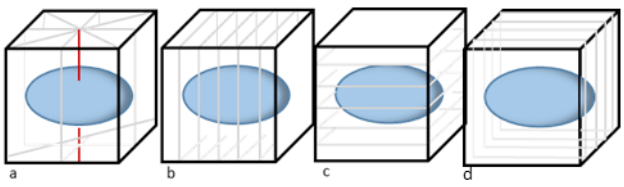


Figure 5 Illustration of different 2D slice directions (grey). **a** The prostate (blue) radially sliced from a 3D TRUS acquisition around its axis of rotation (red). **b** Sagittal prostate slices. **c** Transverse prostate slices, **d** Coronal prostate slices.

Table 1 A brief summary of proposed deep learning architectures to segment the prostate on transrectalultrasound.* 3D reconstruction of twelve 12 2D slices.

Paper	Segmentation	Description
Anas et al. (2018)(36)	2D	ResU-Net with recurrent connections to exploit temporal information.
Anas et al. (2017)(39)	2D	ResU-Net with boundary attended loss function
Beitone et al. (2022)(21)	3D	Three view-specific U-Nets are trained to produce view-specific segmentation that are the input of a fusion network that produces a confidence map on which the final segmentation is created.
Bi et al. (2022)(25)	2D	In a traditional U-Net, the loss function is evaluated at different stages of the network with the aid of a prostate boundary map
Feng et al. (2023) (33)	2D	A multi-stage network that enhances the feature space to improve segmentation results. A traditional U-Net is modified by replacing the encoding branch with a pre-trained VGG16 model. Then the output feature maps are then the input of a similar second network. This process is repeated three times and the final segmentation is conducted on the refined feature map and the segmentations that were created after each network stage.
Ghavami et al.(2018) (38,47)	3D	Intergration of spatial information in standart U-Net by taking 2-6 adjacent slices as training input.
Ghavami et al.(2018) (38, 47)	2D	Intergration of spatial information in standart U-Net by taking 2-6 adjacent slices as training input.
Girum et al. (2020) (24)	3D	A prostate shape mask reconstruction is generated based on prostate boundary coordinates that are predicted from a U-Net back-bone layer. Then, the final segmentation is obtained by merging the shape mask reconstrution with the output of the U-Net decoder.
Karimi et al. (2019) (26)	2D	The final segmentation is created based on the average of 5 U-Nets and adjusted with the aid of MRI based statistical shape model
Lei et al. (2021) (27)	3D	Modified U-Net that learns based on ROI specific feature maps.
Lei et al. (2019) (20)	3D	Thee view-specific V-Nets that estimate prostate volume with the aid of Multi Directional Contour refinement. The view specific V-Nets are optimized with the aid of deep supervision.
Liu et al. (2022) (43)	3D	A 3D resU-net with channel attention implemented.
Liu et al. (2023) (32)	3D	Modified U-Net in which spatial attention and a detail compenstation module is intergrated.
Orlando et al. (2020)(22)	3D*	Modified U-net to predict radial slices that are utilized for 3D volume reconstruction.
Palladino et al. (2022)(44)	2D	Modified U-Net that learns based on ROI specific feature maps.
Peng et al. (2024) (41, 46)	2D	Architecture that inherits the ability of deep-learning models to locate the ROI with incorporation of mathematical models to smooth the contour of ROI.
van Sloun et al. (2021) (34, 45)	2D	U-Net trained by SDa to improve generalization.
Vesal et al. (2022) (30)	3D	Coordination Dilated-ResU-Net that was trained by SDa to improve generalization.
Wang et al.(2019) (29)	3D	Modified U-Net in which spatial attention is intergrated.
Wu et al. (2020) (40)	2D	Traditional U-Net.
Xu et al. (2022) (19)	3D	Modified U-Net that is trained by shadow enhanced images and a modified feature space.

The resulting 3D segmentation after reconstruction was compared to a traditional V-Net and, an improved segmentation performance was presented. (DSC = 0.94, vs 0.91) Although, the quantitative comparison of results must be regarded carefully as the authors mentioned that the V-Net was not rigorously optimized. However, it remains a reasonable comparison to explore if segmentation based on computationally expensive 3D information allows for better results compared to less spatially-dense image information.

In clinical practice, the adjacent slices in a 2D acquisition are often considered during the process of manual segmentation to distinguish the prostate boundary from surrounding tissue. Therefore, Ghavami et al. (2018) explored the integration of neighboring slices as input to a traditional U-Net to improve 2D TRUS segmentation. Based on the evaluation, DSC. increased significantly with 1% when an addition of two adjacent slices on each side of the input image was implemented. Importantly, when three adjacent slices were integrated, no significant improvement was shown indicating that the addition of solely two adjacent slices is sufficient to increase the segmentation performance.

3.2.3 The implementation of additional shape information

To accurately segment ambiguous prostate regions, additional shape information can be incorporated in DL architectures to improve the final outcome. For instance, in the work of Girum et al. (2020), the use of boundary landmarks is explored to improve thesegmentation in 2D TRUS images. (24) The authors employed automatically extracted boundary landmarks (Figure 6) to form a mask reconstruction that was fused with the output of the decoder. The aim was to make the final segmentation less prone to low-contrast and artefacts across the full organ's boundary. In their design, the encoder part of a modified U-Net architecture, extracts anatomic landmarks from the U-Net's bottle neck when the prostate is present in the image. When the contribution of mask reconstruction was evaluated, it was shown that the HD decreased averagely by 2.04mm.

Furthermore, Bi et al. (2022) aimed to improve the training phase of a traditional U-Net by combining boundary information with the aid of an active shape model. (25) In their work, a boundary map was generated by processing the ground truth segmentation with morphological operations and key point initialization to eliminate the disturbance of boundary irregularities. Thereafter, the boundary map was utilized to evaluate the intermediate feature maps / segmentations as a deep supervision strategy. In the end, the DL architecture was trained with a loss function that considered the discrepancy between the boundary map and the intermediate segmentation outcomes and increased the DSC from 0.90 to 0.93.

In the application of Karimi et al. (2019), the implementation of a MRI based statistical shape model (SSM) is explored to improve the segmentation results of a U-Net ensemble that was trained on TRUS data. (26) By matching a SSM that contained MRI information of the expected prostate shape, the HD was decreased by ± 1 mm.

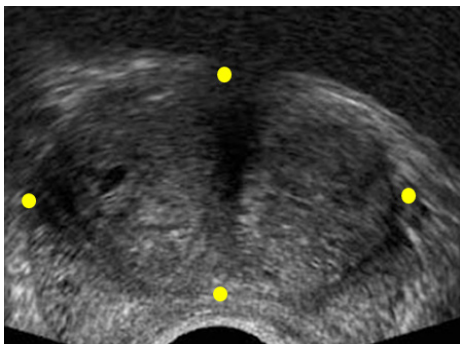


Figure 6 The prostate on transrectal ultrasound where four outer landmarks are marked in yellow.

3.2.4 The implementation of attention mechanisms

Since US images of the prostate are commonly marked by artefacts, methods for diverting attention to the most relevant region in an image are desired. Briefly, this can be achieved on image-level by first defining a Region of interest (ROI) that directs the relevant feature space. This is also referred to as hard attention whereby the 'useless' areas are not considered during segmentation.

In the work of Lei et al. (2021), the ROIs of the prostate is inherently obtained by a probability map of the organ's center of mass. (27) Thereafter the probability map is used to form a bounding box that corresponds to the ROI of the prostate. In contrast, in PROST-Net, the proposed system by Palladino et al (2022), the ROI is defined using the Region Proposal Network designed by (49). (44) When both architectures were evaluated, PROST-NET was able to segment 2D TRUS images with a DSC of 0.87, whereas Lei et al. (2021) achieved a DSC of 0.93.

Alternatively, attention can be obtained on feature-level by introducing channel / spatial attention modules to the DL architecture, also referred to as soft attention. In this way, the weights of each channel or region in the feature space are adaptively recalibrated to determine what to pay attention to. (28)

A pioneer in introducing soft attention in the field of prostate segmentation on TRUS is Wang et al. (2019). (29) The authors integrated an attention module that leverages spatial information, aiming to select more discriminative features. Although, the proposed DL-architecture was able to segment the prostate with as DSC of 0.90, the contribution of the attention module was not evaluated quantitatively, nor was the network tested on separate test data. Following this work, Vesal et al. (2022) retrained their architecture based on 802 training samples and achieved even better results, (DSC = 0.92) proving the robustness of their network on unseen data.

Similarly, in the proposed architecture of Vesal et al. (2022) Coordinate Attention Blocks (CAM) were incorporated to increase the segmentation performance. (30) The primary aim of CAM is to perform channel attention while positional information is preserved. (31) (Figure 7a) Ultimately, the proposed method presented similar DSC outcomes compared to Wang et al. (2019) when their method was evaluated on equal test data.

In the proposed ADC-Net of Liu et al. (2023), a spatial-based attention module was introduced to enhance the network's ability to learn more useful features and improved the DSC and JC metric with 3% and 4% respectively. (32) Figure 7b illustrates the proposed attention mechanism.

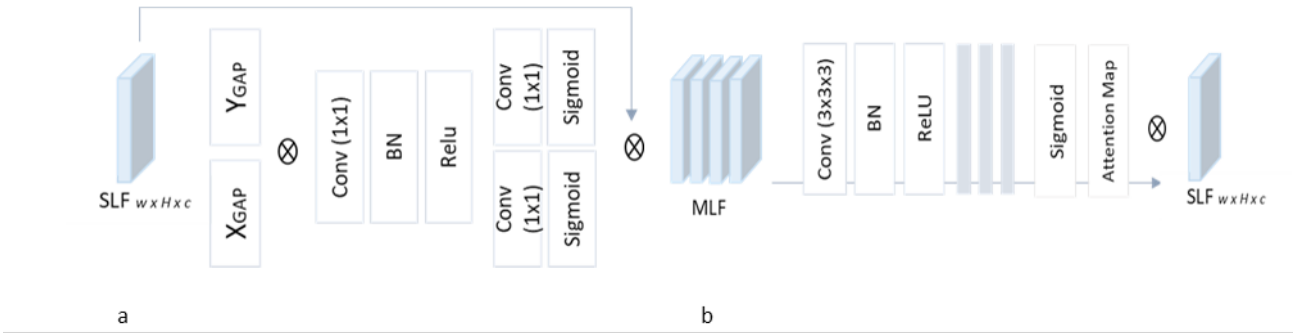


Figure 7a Illustration of coordinate attention mechanism (CAM). The input is a single-layer feature map (SLF) with height, width and channels C. First each channel is encoded in x and y direction by means of Global Average Pooling (GAP). Then, a shaded 1x1 convolutional transformation is applied to concatenate the output weights of the two GAP layers. Then the resulting tensor is split to yield attention vectors with the same number of channels in x and y direction. This is multiplied by the input SLF to produce a channel attentive feature map.
bIllustration of spatial-information attention mechanism. The fused multi-layer feature map (MLF) that contain semantic information are passed through multiple convolution blocks to obtain the weight scores. Normally, each block consists of a convolution layer, a batch normalization (BN) and a non-linear ReLU operation. Thereafter, the weights are constrained between 0 and 1 by means of a sigmoid function. Ultimately the weights are multiplied by the single layer features to produce attentive feature maps.

3.2.5 Feature map refinement

Feng et al. (2023) explored the enhancement of feature maps by the employment of a Multi-Stage U-Net. (33) Their proposed architecture consists of three U-Nets in which the encoding paths are replaced by a VGG16-model. Precisely, the first network is trained on 2D TRUS data to obtain a premature segmentation. Then the feature maps that are encoded serve as input for a second network upon which a second segmentation is created. Finally, this step is repeated and a third segmentation is obtained based on the second feature maps in a third network. At last, the three interim segmentations are fused by element-wise multiplication to generate the final segmentation outcome. Compared to the performance of a single U-net with VGG16 integration, the DSC scores improved with 3.7%. Interestingly, the model performance decreased when interim segmentation outcomes were not included in the formation of the final segmentation. This indicates that certain aspects of the prostate boundary are learned by the different networks, as the pixels in the interim segmentation must agree with the other interim segmentations to be included in the final segmentation.

In the proposed ADC-Net of Liu et al. (2023), the feature space is refined by the introduction of a Detail Compensation Module (DCM). The primary aim of this study was to estimate the PV with the aid of 3D U-Net. (32) To reduce the complexity that arises in 3D convolution, the authors decided to decrease the feature channel dimensions to 16, 32, 64 and 128. Even though this procedure reduces the network's complexity, it is also results in the loss of detailed image information. Hence, a DCM was introduced to enrich the residuary feature maps. In essence, DCM embeds a pre-trained ResNet-34 model that encodes detailed feature maps that are merged to the feature maps of the 3D U-Net, aiming to enhance the feature space. During an ablation study the effect of the incorporated DCM was evaluated and achieved an improvement of 3%, 4%, 3% and 3% on DSC, JC, Precision and Recall respectively.

3.2.6 Consistency and robustness of deep-learning models

A common difficulty in the implementation of DL architectures is the achievement of robust and consistent results across data from multiple institutions owing to the diversity in US images. This is mainly due to the existing variety of imaging manufacturers and the difference in patient characteristics. For instance, abdominal fat limits the US signal to be processed, resulting in a poorer image resolution. Furthermore, US acquisitions may deviate according to the clinician's technique of US examination.

Usually, to improve generalizability, large amount of heterogeneous data is required to optimize DL architectures. Nevertheless, many studies suffer from scarcity of available data. Therefore, van Sloun et al. (2021) and Vesal et al. (2022) aimed to improve the generalization of segmentation models on new data with the aid of supervised Domain Adaption (sDA). (30, 34) To this extent, the proposed networks are pre-trained on one TRUS dataset in a supervised manner. Thereafter the trained networks are retrained on a small number of TRUS images of another data set to improve the network on unseen image data.

In the work of Van Sloun et al. (2021), sDA was implemented in a traditional U-Net architecture that was fine-tuned using three different datasets. (34) Nevertheless, the model performance did not improve significantly by the implementation of sDA. The lacking generalizability of their proposed network is most likely a result of the homogeneous data that was used for training

and fine-tuning, as the datasets contained multiple acquisitions per patient. (436 acquisitions of 181 patients)

With similar ambition, Vesal et al. (2021) explored sDA by using a multi-class loss function. (30) Aside from using a segmentation loss to train a model on different datasets, a Knowlegde Distillation loss was also incorporated. The primary aim of this regularization technique is to reduce the effect of the changes in weights of the encoded features when a well-trained model is updated on new data. When the implementation of knowledge distillation was quantitatively compared on two other datasets, the authors showed an increased model performance on the transferred datasets indicating that their proposed architecture performs well on multi-institutional data. (DSC = 0.94, HD = 2.29 on dataset 1, 0.91, 3.69 on dataset 2, 0.82, 7.13 on dataset 3) Given that the segmentation performance surpassed the performance of a network that was solely trained on the transferred datasets, knowledge distillation, in addition to a DSC loss, appears to be a promising strategy for handling the extensive diversity in US acquisitions

3.2.7 A quantitative evaluation

Various evaluation metrics exist in order to assess the segmentation performance of DL architectures. Most commonly DSC and Jaccard index (JC) are used to quantify volume similarity between two contours whereas HD and Average surface distance (ASD) are used to quantify boundary similarity. In general, desirable segmentation results are associated with lower HD and MSD and higher DSC and JC scores. (37) Table 2 provides an overview of proposed DL architectures and the corresponding evaluation scores.

At present, DL architectures are able to segment the prostate on TRUS with a DSC similarity that ranges between 0.87 - 0.94, and a HD distance of 1.12 – 8.37 mm. For comparison, manual 2D segmentations of different experts are characterized with a DSC of 0.93. (38) As most proposed architectures consist of a modified U-Net, it is worth to mention that a DSC score of 0.90 is obtained when a traditional U-Net is trained. (39, 40) Furthermore, Anas et al. (2017) showed that a ResU-Net was able to segment the prostate with a DSC of 0.93.

Table 2 A comparison of DL architectures on Dice Similarity Coefficient (DSC), Jaccard Index (JC), Average Surface Distance (ASD) and Haussdorf Distance (HD). *evaluation on HD95. ** 3D reconstruction of twelve 2D segmentations.

Paper	Segmentation	Patient	Dataset	DSC	JC	ASD(mm)	HD(mm)
Anas et al. (2018)	2D	18	2875/1017	0.93 ± 0.03		1.12 ± 0.79	2.97± 1.96
Anas et al. (2017)	2D	598	4284/1081	0.93 ± 0.04		1.13 ± 0.81	3.41 ± 2.18
Beitone et al. (2022)	3D	100	80/20	0.93 ± 0.04	0.86 ± 0.06	0.83 ± 0.41	5.48 ± 2.66
Bi et al. (2022)	2D	-	90/10	0.93	0.93		
Feng et al. (2023)	2D	364	1638/182	0.95	0.89		
Ghavami et al.(2018)	3D	109	99/10	0.89 ± 0.05			1.79 ± 2.05
Ghavami et al.(2018)	2D	109	3017/1017	0.89 ± 0.01			1.12 ± 0.79
Girum et al. (2020)	3D	145	125/20	0.88 ± 0.02		0.10 ± 0.06	2.01 ± 0.54*
Karimi et al. (2019)	2D	675	5400/1350	0.94 ± 0.03			2.50 ± 1.7
Lei et al. (2021)	3D	50	40/10	0.93 ± 0.03		0.57 ± 0.20	2.28 ± 0.64*
Lei et al. (2019)	3D	46	35/11	0.92 ± 0.03		0.59 ± 0.26	3.938 ± 1.6
Liu et al. (2022)	3D	50	50/?	0.91 ± 0.02		1.1 ± 0.18	4.38 ± 1.13*
Liu et al. (2023)	3D	46	46/?	0.91 ± 0.06	0.84 ± 0.9		
Orlando et al. (2020)	3D**	246	5418/1355	0.94 ± 0.02		0.89 ± 0.15	2.89 ± 1.45
Palladino et al. (2022)	2D	22	-	0.87			
Peng et al. (2024)	2D	266	741/204	0.94 ± 0.04	0.93 ± 0.05		
van Sloun et al. (2021)	2D	78	158/44	0.93 ± 0.01			3.0 ± 5.7
Vesal et al. (2022)	3D	954	802/190	0.94 ± 0.03			2.29 ± 1.45*
Wang et al.(2019)	3D	40	30/10	0.90 ± 0.03	0.82 ± 0.04	3.32 ± 1.15	8.37 ± 2.52*
Wu et al. (2020)	2D	-	490/106	0.90	0.82		
Xu et al. (2022)	3D	1150	1064/ 679	0.92 ± 0.02	0.93 ± 0.29		5.89 ± 1.93

3.3 Prostate volume estimation with the aid of the Prolate Ellipsoid Formula

Mathematically, PV can be directly derived from prostate segmentation or the estimation is achieved with the aid of the PEF. (35) In this way, the total volume is computed by measuring the length (L), height (H) and width (W) of the prostate gland and multiplying the product by a coefficient (c) (Formula 1). (9, 35) In clinical practice, the PEF is frequently used for PV estimation as it is a faster approach compared to manual prostate segmentation.

$$\text{Prolate Ellipsoid Formula: } W \times H \times L \times \pi/6. \quad (1)$$

Albayrak et al. (2022) aimed to automate the workflow of PEF by establishing a method to estimate the six outer boundary landmarks that are required to calculate the height, width and length of the prostate. (17, 48) In their work, two DL architectures were trained separately for different image orientations. One network focused on 2D transverse TAUS images to predict four landmarks related to height and width, while another network was trained on sagittal TAUS images to identify the remaining two landmarks to determine the prostate's length. The view-specific DL architectures comprised an ensemble of four ResNet-18 CNNs (QDCNN) that were designed to classify the distance and orientation relative to the landmarks given a random point in the training image. Then, the location of each landmark was determined by producing a landmark-specific voting map of all predicted distance and orientation outcomes. Ultimately, the landmarks served as input to utilize Formula 1 to obtain the PV.

Notably, QDCNN was trained on a variety of image patches of different scales around a random point in the training image. The rationale behind this approach is that patches of different sizes with matching centers are able to extract image details from different scales while reducing the attention to surrounding structures that occur in TAUS acquisitions.

The proposed method was validated by comparing the Mean Absolute Value Difference (MAVD) between the predicted PV and a manually estimated PV on TAUS. An average MAVD value of 4.95 cm3 was obtained, which is smaller than the average inter-expert MAVD value of 5.09 cm3 for manual PV estimation on TAUS. Additionally, the predicted PV was compared to a PV estimation based on MRI that resulted in an average MAVD value of 6.22 cm3. For comparison, the inter-expert average MAVD value between TAUS and MRI was 8.06 cm3.

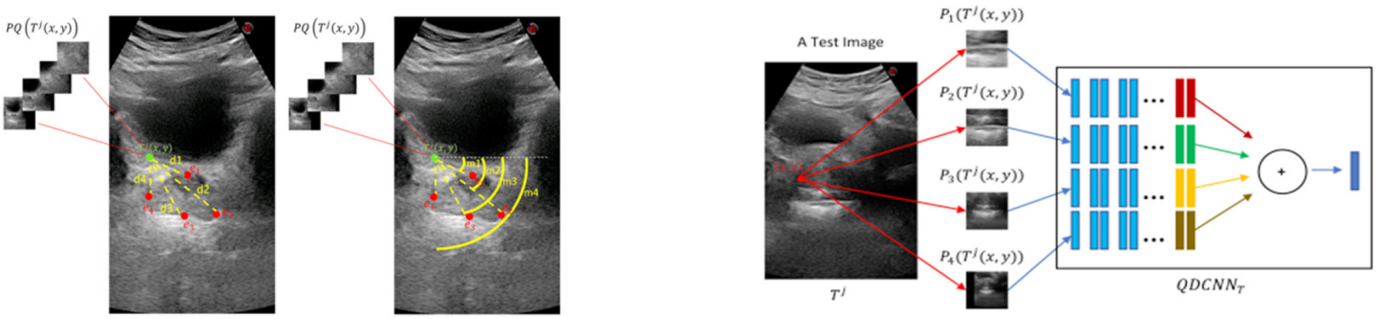


Figure 8 Illustration of the patch-center (green) and its corresponding scale-specific patches to predict the distance and orientation to a respective landmark (red). To achieve this, an ensemble four scale-specific DCNNs are proposed.

4. Discussion

To effectively manage prostate diseases, it is essential to determine the prostate volume (PV), whereby ultrasound (US) is commonly employed as it allows for a cost-effective and rapid assessment. Mathematically, PV can be obtained by means of US image segmentation or with the aid of the Prolate Ellipsoid Formula (PEF). However, in clinical practice, it remains challenging for physicians to obtain an accurate estimation of the PV since US images, are frequently marked by imaging artefacts that contribute to an ambiguous prostate appearance. Hence, computerized methods have been introduced to improve the determination of PV and

to reduce patient examination duration. This paper aims to review recent explorations that focus on the application of deep-learning to estimate the PV on US.

4.1 The application of deep-learning for prostate segmentation on ultrasound

Based on the evaluation of DL architectures for prostate segmentation, the presented methods demonstrate comparability in regard to model performance. Nevertheless, a comparison merely on evaluation metrics remains short-sighted as various aspects influence the result. Briefly, the segmentation outcome is influenced by the architecture design and dataset characteristics in which dataset homogeneity, US image quality and quantity play a role. Experts should be aware of the used dataset-quality and whether the presented performance can be translated to clinical practice, where poor-quality US acquisitions occur frequently. This marks the necessity for the availability of public datasets to allow a fair comparison of DL architectures in the field of prostate segmentation on US. Moreover, standard deviation values need to be acknowledged to provide transparency concerning the consistency of results.

In the majority of explorations, the performance of the DL architecture is enlightened by the presentation of better- and worse segmentation results. Inaccurate predictions typically involve the apex and base region, presumably due to the poor prostate visibility in this area. As a result, the implementation of multi-directional image information has been explored, covering an ensemble of view-specific architectures whereby the final segmentation outcome is produced on prostate information in axial, coronal and sagittal direction. However, in future experiments, it is recommended to explore how much multi-directional image input contributes to the final segmentation performance as its implementation increases the training workload. For instance, when a 3D reconstruction is established on a certain number of 2D segmentations, it is recommended to explore whether the use of more slices would significantly increase the model performance to illustrate the trade-off between input size and segmentation accuracy. Additionally, when post-processing strategies like MDCR (20) are introduced, it should not be neglected that MDCR is dependent on a set distance parameter that may be prone to inconsistent results when the diversity in image data is increased. Even though the incorporation of multi-directional information appears to increase the segmentation performance, it remains unclear if the design of Beitone et al. (2022) can be translated to clinical practice, as the authors manually cropped each acquisition to a Region Of Interest (ROI): The prostate with 40 voxels around it. Owing to this pre-processing step, image artefacts that usually degrade a model performance may have been excluded in the evaluation.

Multiple authors explored the employment of additional shape information to enhance segmentation outcomes in which the use of statistical shape models and boundary reconstructions were explored. (24, 25, 26) Overall, all additional shape-implementations contributed to improved segmentation outcomes.

When it comes to the integration of attention mechanisms, the proposed soft attention strategies that were proposed by Wang et al. (2019), Feng et al. (2023), Vesal et al. (2022) and Liu et al. (2022) all seemed to improve segmentation results. The proposed experiments present a DSC score above 0.92 which is presumably also the result of the high number of patient inclusion. Usually, a low number of included patients results in poor segmentation generalizability as the network is unable to capture the diversity of the prostate across different patients. This aspect is possibly seen when the performance of Palladino et al. (2022) is evaluated. However, the performance of Anas et al. (2018) contradicts this hypothesis by achieving a DSC score of 0.93 on an architecture that is trained on 18 patients.

The experiments of Vesal et al. (2022) and van Sloun et al. (2021) emphasize the importance of patient heterogeneity in used datasets. In this light, employment of supervised Domain Adaption is proven to be of additional value, to improve the generalizability of DL architectures that are trained on a low number of patients.

Furthermore, inaccurate segmentations as a result of present shadow artefacts may be reduced with the aid of the pre-processing strategies proposed by Xu et al. (2020). Their method is worth to explore more elaborately, particularly in the case of transabdominal ultrasound images, where shadow artefacts frequently disrupt the prostate boundary. In contrast, pre-processing training data with 3D Mean, Median and Gaussian filters appear to be less useful for improving segmentation outcomes. (20)

At last, it is recommended to conduct ablation studies when additional models are implemented or when an ensemble of networks is proposed. For instance, Lei et al. (2021) deigned a DL architecture that consists of multiple systems to enhance segmentation outcomes, however the contribution of each implementation was not explored. In contrast, in other experiments it is clearly reported which components of the proposed DL architectures improve segmentation outcomes. () Specifically when taken into account that efficient DL models are usually required to optimize the opportunity for real time employment, it is reasonable to understand how an addition in complexity contributes to model performance.

The employment of DL for prostate segmentation is worth to explore further as it allows for a volume estimation based on the whole prostate region. This is particularly beneficial in patients suspected of prostate

cancer, as the prostate shape variates across patients. Furthermore, prostate segmentation architectures can be utilized in the field of image-registration.

4.1 The application of deep-learning for prostate volume estimation with the aid of the Prolate ellipsoid formula.

In this review, solely one research group concentrates on PV estimation with the aid of PEF. Strikingly, Albayrak et al. (2022) focused on TAUS to obtain the PV, whereas prostate segmentation methods are most commonly applied on TRUS. Even though the presented results seem rewarding with an average volume difference of 4.95 cm3, there are considerable aspects of their proposed system that warrant discussion. First, PEF can only be utilized when all diameters are present at maximum in de 2D TAUS image. Therefore, in both directions, the prostate cross-section must be at maximum when given to the system. Currently, the authors only trained the network with TAUS images that satisfy this criteria. As a result, human knowledge is still required when the proposed method would be implemented in a clinical workflow. Hence, the proposed design requires an implementation that determines the most suitable prostate slice in an acquisition. The methods proposed by (21) and (palladino) may serve as a solution when they are modified to classify the prostate region first on its cross section, before the computation of the landmarks start.

Secondly, the system is evaluated on MAVD of the entire PV. Since the PV relies on the prediction of six landmarks, it remains unclear whether the accuracy of the calculated landmarks are contributing equally to the presented MAVD. When the predictions are evaluated separately, the performance is less black-boxed and insight is provided on which diameter, and therefore image direction, inaccuracies depend.

The application of DL for the employment of PEF is unique and can be advantageous on TAUS image data since the acquisitions consist of a 2D coordinate space that complicate the acquirement of a 3D volume reconstruction. Utilizing TAUS for PV estimation is promising as the examination is characterized by more patient comfort and a decreased work-load on clinical practitioners. Nevertheless, PEF assumes the prostate to be of an ellipsoid shape which may lead to over and under estimations of PV in clinical practice.

.

5. Bibliography

1.Wei JT, Barocas D, Carlsson S, Coakley F, Eggener S, Etzioni R, et al. Early Detection of Prostate Cancer: AUA/SUO Guideline Part I: Prostate Cancer Screening. Journal of Urology. 2023;210(1).

2.Ito K, Oki R, Sekine Y, Arai S, Miyazawa Y, Shibata Y, et al. Screening for prostate cancer: History, evidence, controversies and future perspectives toward individualized screening. Vol. 26, International Journal of Urology. 2019.

3.Tewari A, Indudhara R, Shinohara K, Schalow E, Woods M, Lee R, et al. Comparison of transrectal ultrasound prostatic volume estimation with magnetic resonance imaging volume estimation and surgical specimen weight in patients with benign prostatic hyperplasia. Journal of Clinical Ultrasound. 1996;24(4).

4.Jue JS, Barboza MP, Prakash NS, Venkatramani V, Sinha VR, Pavan N, et al. Re-examining Prostate-specific Antigen (PSA) Density: Defining the Optimal PSA Range and Patients for Using PSA Density to Predict Prostate Cancer Using Extended Template Biopsy. Urology. 2017;105.

5.Albert JM, Swanson DA, Pugh TJ, Zhang M, Bruno TL, Kudchadker RJ, et al. Magnetic resonance imaging-based treatment planning for prostate brachytherapy. Brachytherapy. 2013 Jan 1;12(1):30–7.

6.Nickel JC. NEW PERSPECTIVES ON BPH Benign Prostatic Hyperplasia: Does Prostate Size Matter? Rev Urol. 2003;5.

7.Youn SY, Choi MH, Lee YJ, Grimm R, von Busch H, Han D, et al. Prostate gland volume estimation: anteroposterior diameters measured on axial versus sagittal ultrasonography and magnetic resonance images. Ultrasonography. 2023;42(1).

8.Pate WR, Garg N, Wang LB, Wason SE, Barbosa P V. Comparison of Transabdominal and Transrectal Ultrasound for Sizing of the Prostate. Urology. 2020;141.

9.Lee JS, Chung BH. Transrectal ultrasound versus magnetic resonance imaging in the estimation of prostate volume as compared with radical prostatectomy specimens. Urol Int. 2007;78(4).

10.Ghose S, Oliver A, Martí R, Lladó X, Vilanova JC, Freixenet J, et al. A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images. Comput Methods Programs Biomed. 2012;108(1).

11.Betrouni N, Vermandel M, Pasquier D, Maouche S, Rousseau J. Segmentation of abdominal ultrasound images of the prostate using a priori information and an adapted noise filter. Computerized Medical Imaging and Graphics. 2005;29(1).

12.Jiang J, Guo Y, Bi Z, Huang Z, Yu G, Wang J. Segmentation of prostate ultrasound images: the state of the art and the future directions of segmentation algorithms. Artif Intell Rev. 2023;56(1).

13.Peng T, Wang C, Tang C, Gu Y, Zhao J, Li Q, et al. A multi-center study of ultrasound images using a fully automated segmentation architecture. Pattern Recognit. 2024;145.

14."PubMed", PubMed. https://pubmed.ncbi.nlm.nih.gov/.

15."IEEE Xplore". https://ieeexplore.ieee.org/Xplore/home.jsp . \

16."Scopus Preview - Scopus - Welcome to Scopus". https://www.scopus.com/home.uri.

17.Albayrak NB, Yildirim E, Akgul YS. Prostate size inference from abdominal ultrasound images with patch based prior information. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2017.

18.Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2015.

19.Xu X, Sanford T, Turkbey B, Xu S, Wood BJ, Yan P. Shadow-Consistent Semi-Supervised Learning for Prostate Ultrasound Segmentation. IEEE Trans Med Imaging. 2022 Jun 1;41(6):1331–45.

20.Lei Y, Tian S, He X, Wang T, Wang B, Patel P, et al. Ultrasound prostate segmentation based on multidirectional deeply supervised V-Net. Med Phys. 2019;46(7).

21.Beitone C, Troccaz J. Multi-eXpert fusion: An ensemble learning framework to segment 3D TRUS prostate images. Med Phys. 2022;49(8).

22.Orlando N, Gillies DJ, Gyacskov I, Romagnoli C, D'Souza D, Fenster A. Automatic prostate segmentation using deep learning on clinically diverse 3D transrectal ultrasound images. Med Phys. 2020;47(6).

23.Qiu W, Yuan J, Ukwatta E, Fenster A. Rotationally resliced 3D prostate TRUS segmentation using convex optimization with shape priors. Med Phys. 2015;42(2).

24.Girum KB, Lalande A, Hussain R, Créhange G. A deep learning method for real-time intraoperative US image segmentation in prostate brachytherapy. Int J Comput Assist Radiol Surg. 2020;15(9).

25.Bi H, Sun J, Jiang Y, Ni X, Shu H. Structure boundary-preserving U-Net for prostate ultrasound image segmentation. Front Oncol. 2022;12.

26. Karimi D, Zeng Q, Mathur P, Avinash A, Mahdavi S, Spadinger I, et al. Accurate and robust deep learning-based segmentation of the prostate clinical target volume in ultrasound images. Med Image Anal. 2019;57.

27.Lei Y, Wang T, Roper J, Jani AB, Patel SA, Curran WJ, et al. Male pelvic multi-organ segmentation on transrectal ultrasound using anchor-free mask CNN. Med Phys. 2021;48(6).

28.Guo MH, Xu TX, Liu JJ, Liu ZN, Jiang PT, Mu TJ, et al. Attention mechanisms in computer vision: A survey. Vol. 8, Computational Visual Media. 2022.

29.Wang Y, Ni D, Dou H, Hu X, Zhu L, Yang X, et al. Deep Attentive Features for Prostate Segmentation in 3D Transrectal Ultrasound. IEEE Trans Med Imaging. 2019;38(12).

30.Vesal S, Gayo I, Bhattacharya I, Natarajan S, Marks LS, Barratt DC, et al. Domain generalization for prostate segmentation in transrectal ultrasound images: A multi-center study. Med Image Anal. 2022;82.

31.Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2021.

32.Liu M, Wu K, Jiang L. ADC-Net: adaptive detail compensation network for prostate segmentation in 3D transrectal ultrasound images. In 2023.

33.Feng Y, Atabansi CC, Nie J, Liu H, Zhou H, Zhao H, et al. Multi-stage fully convolutional network for precise prostate segmentation

in ultrasound images. Biocybern Biomed Eng. 2023;43(3).

34.van Sloun RJG, Wildeboer RR, Mannaerts CK, Postema AW, Gayet M, Beerlage HP, et al. Deep Learning for Real-time, Automatic, and Scanner-adapted Prostate (Zone) Segmentation of Transrectal Ultrasound, for Example, Magnetic Resonance Imaging–transrectal Ultrasound Fusion Prostate Biopsy. Eur Urol Focus. 2021;7(1).

35.Aprikian S, Luz M, Brimo F, Scarlata E, Hamel L, Cury FL, et al. Improving ultrasound-based prostate volume estimation. BMC Urol. 2019;19(1).

36.E. M. A. Anas, P. Mousavi, and P. Abolmaesumi, "A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy," Med Image Anal, vol. 48, 2018, doi: 10.1016/j.media.2018.05.010.

37.Yeghiazaryan V, Voiculescu I. Family of boundary overlap metrics for the evaluation of medical image segmentation. Journal of Medical Imaging. 2018;5(01).

38.Ghavami N, Hu Y, Bonmati E, Rodell R, Gibson E, Moore C, et al. Integration of spatial information in convolutional neural networks for automatic segmentation of intraoperative transrectal ultrasound images. Journal of Medical Imaging. 2018;6(01).

39.Anas EMA, Nouranian S, Mahdavi SS, Spadinger I, Morris WJ, Salcudean SE, et al. Clinical target-volume delineation in prostate brachytherapy using residual neural networks. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2017.

40.Wu H, Chen Y, Huang B, Zhang S, Zhu Y, Chen Y. Contour-based medical image fusion for biopsy. In: 12th International Conference on Advanced Computational Intelligence, ICACI 2020. 2020.

41.Peng T, Zhao J, Gu Y, Wang C, Wu Y, Cheng X, et al. H-ProMed: Ultrasound image segmentation based on the evolutionary neural network and an improved principal curve. Pattern Recognit. 2022;131.

42.Trus Prostate by Dr Rohit Malik [online video]. 9 januari 2019. Available at: TMT: TRUS Prostate by Dr Rohit Malik - YouTube (accessed 28-11-2023) .

43. D. Liu, L. Wang, Y. Du, M. Cong, and Y. Li, "3-D Prostate MR and TRUS Images Detection and Segmentation for Puncture Biopsy," IEEE Trans Instrum Meas, vol. 71, 2022, doi: 10.1109/TIM.2022.3192292

44. L. Palladino, B. Maris, A. Antonelli, and P. Fiorini, "PROST-Net: A Deep Learning Approach to Support Real-Time Fusion in Prostate Biopsy," IEEE Trans Med Robot Bionics, vol. 4, no. 2, pp. 323–326, May 2022, doi: 10.1109/TMRB.2022.3145667.

45. R. J. G. Van Sloun et al., "Zonal segmentation in transrectal ultrasound images of the prostate through deep learning."

46. T. Peng et al., "H-ProMed: Ultrasound image segmentation based on the evolutionary neural network and an improved principal curve," Pattern Recognit, vol. 131, 2022, doi: 10.1016/j.patcog.2022.108890.

47.N. Ghavami et al., "Automatic slice segmentation of intraoperative transrectal ultrasound images using convolutional neural networks," 2018. doi: 10.1117/12.2293300.

48. Albayrak, Nur Banu, and Yusuf Sinan Akgul. 2022. "Estimation of the Prostate Volume from Abdominal Ultrasound Images by Image-Patch Voting" Applied Sciences 12, no. 3: 1390.

49. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," IEEE Trans Pattern Anal Mach Intell, vol. 42, no. 2, 2020, doi: 10.1109/TPAMI.2018.2844175.

Appendix A

Search Query Pubmed: EXPORT DATE: 11 October 2023

(“deep learning”[MeSH Terms] OR (“deep”[All Fields] AND “learning”[All Fields]) OR “deep learning”[All Fields] OR (“deep”[All Fields] AND “neural”[All Fields] AND “network”[All Fields]) OR (“neural”[All Fields] AND “network”[All Fields]) OR “U-Net”[All Fields] (“deep learning” [All Fields] AND “segmentation” [All Fields])) AND (“prostate”[MeSH Terms] OR “prostate”[All Fields] OR “prostates”[All Fields] OR “prostatic”[All Fields] OR “prostatism”[MeSH Terms] OR “prostatitis”[All Fields] OR (“prostate”[All Fields] AND “volume”[All Fields]) OR (“prostate”[All Fields] AND “segmentation”[All Fields]) OR (“prostate”[All Fields] AND “image”[All Fields] AND “segmentation”[All Fields])) AND (“ultrasonography”[All Fields] OR “ultrasonography”[MeSH Terms] OR “ultrasonographies”[All Fields] OR “ultrasound”[All Fields] OR “TRUS”[All Fields] OR “TAUS”[All Fields] OR “US”[All Fields] OR (“transrectal”[All Fields] AND “ultrasound”[All Fields]) OR (“transabdominal”[All Fields] AND “ultrasound”[All Fields]) OR (“abdominal”[All Fields] AND “ultrasound”[All Fields]))

Search Query IEEE Xplore: EXPORT DATE: 16 October 2023

(“Abstract”:”deep learning” OR “Abstract”:”neural network” OR “Abstract”:”Net”) AND (“Abstract”:”prostate”) AND (“Abstract”:”ultrasound” OR “Abstract” OR “Abstract”:”US” OR “Abstract”:”TRUS” OR “Abstract”:”TAUS”)
Filter: 2016-2023

Search Query Scopus: EXPORT DATE: 25 October 2023

(TITLE-ABS-KEY (deep AND learning) AND TITLE-ABS-KEY (prostate) AND TITLE-ABS-KEY (ultrasound)) AND PUBYEAR > 2016 AND PUBYEAR < 2025