Ultrasound Obstet Gynecol 2022; **60**: 570–576
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.24810.

Automatic identification and segmentation of slice of minimal hiatal dimensions in transperineal ultrasound volumes

F. VAN DEN NOORT[1#], C. MANZINI[2#], C. H. VAN DER VAART[2], M. A. J. VAN LIMBEEK[3], C. H. SLUMP[1] and A. T. M. GROB[4]

[1]*Robotics and Mechatronics, Faculty of Electrical Engineering, Mathematics and Computer Science, Technical Medical Centre, University of Twente, Enschede, The Netherlands;* [2]*Department of Obstetrics and Gynecology, University Medical Centre Utrecht, Utrecht, The Netherlands;* [3]*Dynamics of Complex Fluids, Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany;* [4]*Multi-Modality Medical Imaging, Faculty of Science and Technology, Technical Medical Centre, University of Twente, Enschede, The Netherlands*

KEYWORDS: automatic segmentation; deep learning; levator hiatus; pelvic floor; transperineal ultrasound; urogenital hiatus

CONTRIBUTION

What are the novel findings of this work?
Deep learning enabled automatic identification of the slice of minimal hiatal dimensions and segmentation of the urogenital hiatus in transperineal ultrasound (TPUS) volumes of women with symptomatic pelvic organ prolapse. This allowed automatic measurement of the urogenital hiatal area, anteroposterior diameter and coronal diameter.

What are the clinical implications of this work?
Our tool can be implemented in the software of TPUS machines, which should make the analysis of TPUS data less time-consuming and observer-dependent, thus reducing clinical training and analysis time and simplifying the examination of TPUS data for research and clinical purposes.

ABSTRACT

Objective *To develop and validate a tool for automatic selection of the slice of minimal hiatal dimensions (SMHD) and segmentation of the urogenital hiatus (UH) in transperineal ultrasound (TPUS) volumes.*

Methods *Manual selection of the SMHD and segmentation of the UH was performed in TPUS volumes of 116 women with symptomatic pelvic organ prolapse (POP). These data were used to train two deep-learning algorithms. The first algorithm was trained to provide an estimation of the position of the SMHD. Based on this estimation, a slice was selected and fed into the second algorithm, which performed automatic segmentation of the UH. From this segmentation, measurements of the UH area (UHA), anteroposterior diameter (APD) and coronal diameter (CD) were computed automatically. The mean absolute distance between manually and automatically selected SMHD, the overlap (dice similarity index (DSI)) between manual and automatic UH segmentation and the intraclass correlation coefficient (ICC) between manual and automatic UH measurements were assessed on a test set of 30 TPUS volumes.*

Results *The mean absolute distance between manually and automatically selected SMHD was 0.20 cm. All DSI values between manual and automatic UH segmentations were above 0.85. The ICC values between manual and automatic UH measurements were 0.94 (95% CI, 0.87–0.97) for UHA, 0.92 (95% CI, 0.78–0.97) for APD and 0.82 (95% CI, 0.66–0.91) for CD, demonstrating excellent agreement.*

Conclusions *Our deep-learning algorithms allowed reliable automatic selection of the SMHD and UH segmentation in TPUS volumes of women with symptomatic POP. These algorithms can be implemented in the software of TPUS machines, thus reducing clinical analysis time and simplifying the examination of TPUS data for research*

Correspondence to: Dr F. van den Noort, Robotics and Mechatronics, Faculty of Electrical Engineering, Mathematics and Computer Science, Technical Medical Centre, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands (e-mail: friedavanlimbeek@gmail.com)
#F.v.d.N. and C.M. contributed equally to this study.

Accepted: 26 October 2021

INTRODUCTION

Transperineal ultrasound (TPUS) is an imaging technique used to investigate pelvic floor dysfunction[1], which enables assessment of the urogenital hiatus (UH) and levator ani muscle (LAM)[2–4]. The UH, the surface of which is measured as the urogenital hiatal area (UHA) on TPUS, is an opening encircled by the pubic bone and the puborectalis muscle (PRM) and is the largest potential hernial portal in the female body. Pelvic organ prolapse (POP) is the herniation of pelvic organs through the UH and is one of the most common types of pelvic floor dysfunction[5]. An enlarged UHA on TPUS is a sign of impaired pelvic organ support and is associated with POP[2].

The LAM is the largest muscle complex of the pelvic floor. The disconnection of its most medial part from the insertion on the inferior pubic ramus can occur during vaginal delivery and is called LAM avulsion[3]. LAM avulsion is associated with POP and reduced pelvic floor muscle function[3,6,7]. A crucial step for the assessment of both UHA and LAM avulsion on TPUS is the identification of the slice of minimal hiatal dimensions (SMHD). This is performed by locating the shortest line between the pubic symphysis and the anorectal angle in the midsagittal plane[8]. The SMHD is the slice passing through this line, perpendicular to the midsagittal plane. In this slice, UH can be segmented to measure its diameters and UHA, and LAM avulsion can be assessed.

The limitation is that SMHD selection and UH measurements are currently performed manually, which makes the analysis time-consuming and requires each observer to complete a learning curve to perform the measurements properly[9]. In addition, even though previous studies showed good inter- and intraobserver variability[10,11], automating this procedure may reduce variability even further.

Several papers have been published recently on the automatic segmentation of the UH[12–14]. However, this automatic segmentation was based on a manually selected SMHD. The selection of the SMHD itself has not been automated to date. Our aim was to use deep learning to develop a tool that would allow both automated selection of the SMHD and UH segmentation, with the purpose of reducing analysis time and observer variability.

METHODS

Data

The data used in this study were collected as part of the GYNecological Imaging using three-dimensional (3D) UltraSound (GYNIUS) project on the assessment of pelvic floor contractility using TPUS, which was performed at a tertiary urogynecological clinic. Women were included in the GYNIUS project between May 2018 and December 2019. The Medical Research Ethics Committee exempted the project from ethical approval (reference number, 18/215) and all women provided written informed consent.

Women underwent TPUS in the supine position after bladder emptying. Women were instructed to perform maximum pelvic floor contraction and maximum Valsalva maneuver according to the method described by Dietz[5]. We used a Philips Epiq 7G machine (Philips, Bothell, WA, USA) with a X6-1 transducer covered with a 2-cm thick gel pad and a glove. The gel pad was used to create more distance between the transducer and the pelvic floor, such that the LAM was fully visible within the opening angle in the coronal plane.

The ultrasound scans used in this study were selected and segmented for analysis in two previous clinical studies[15,16]. The UH was segmented manually by one observer (C.M.) during rest, on maximum pelvic floor contraction and on maximum Valsalva maneuver in the SMHD. This segmentation was performed using in-house developed software[17], which was implemented in MeVisLab 3.0.2. All steps for SMHD selection were saved in the software, including the position of the SMHD in the 3D volume and segmentation of the UH in the SMHD, which enabled automation of the process.

Deep learning

Deep learning is a set of algorithms that try to mimic the learning of the human brain, also known as neural networks. After the ImageNet 2012 Challenge was won by a convolutional neural network (CNN)[18], CNNs quickly became state of the art for (medical) image analysis, often resulting in human-level performance in tasks such as image segmentation and classification[19].

Segmentation CNNs are trained by providing them with manually labeled data. During training, the CNN learns the patterns needed to perform the segmentation task by minimizing a loss function, which quantifies the performance of the network. The more data (preferably from a representative sample of the entire population) that are used, the better a CNN is able to generalize the learned task to the entire population. An independent validation set is used to check during training how well the CNN performs on data that are not part of the training set, which is a measure of the generalization capability of the CNN. The CNN that performs best on the validation set, i.e. for which the loss function has the lowest value on the validation set, is used for further analysis. A test set is then used to analyze the resulting performance of the CNN on new, unseen data (i.e. different from the training set), as well as on data for which the CNN is not optimized (i.e. different from both training and validation sets). Our data were assigned randomly to the training set (104 patients, 381 frames, 337 two-dimensional (2D) slices), validation set (two

patients, 12 frames, six 2D slices) or test set (10 patients, 30 frames, 30 2D slices).

Selection and segmentation pipeline

Two different CNNs that can operate in a pipeline to make the process of SMHD selection and UH segmentation fully automated were trained (Figure 1). The first CNN was a SMHD-selection CNN (SS-CNN), which has the same network architecture as the CNN presented previously to segment automatically the PRM in 3D TPUS volumes[20]. This network processes the data slice by slice, but remembers interslice information, enabling full usage of the 3D context. The same segmentation network was used to estimate the position of the SMHD. This network was trained on sagittal slices because manual selection of the SMHD is done mainly in the midsagittal plane.

The labels of the manual UH segmentation are only one slice thick. However, a shift of a few slices also results in correct visualization of the UH. Such shift may occur between observers or if the same observer performs repeat selection of the SMHD. Therefore, the segmentation mask was enlarged by performing a dilatation operation to cover five slices in order not to restrict SMHD segmentation to the slice of the label. The choice of five slices was arbitrary. The network was trained to maximize the overlap with this enlarged mask by using the dice similarity index (DSI) loss function (DSI-LF)[21]. However, since an overlap (DSI) is not always the perfect indicator of a successful estimation of the position of the SMHD, a loss function that integrated the estimated position of the SMHD with respect to the manual mask was also added (Appendix S1).

After training, the network provided estimates of the position of the SMHD in the 3D volume, which were not in a perfectly straight plane. Therefore, a plane (least-square error) was fitted through the datapoints of the estimation, and 3D data in this plane were interpolated to obtain a 2D slice (the SMHD).

The second network was the 2D-segmentation CNN (2DS-CNN), which has been presented in our previous work on 2D-SMHD segmentation of the PRM and UH[14].

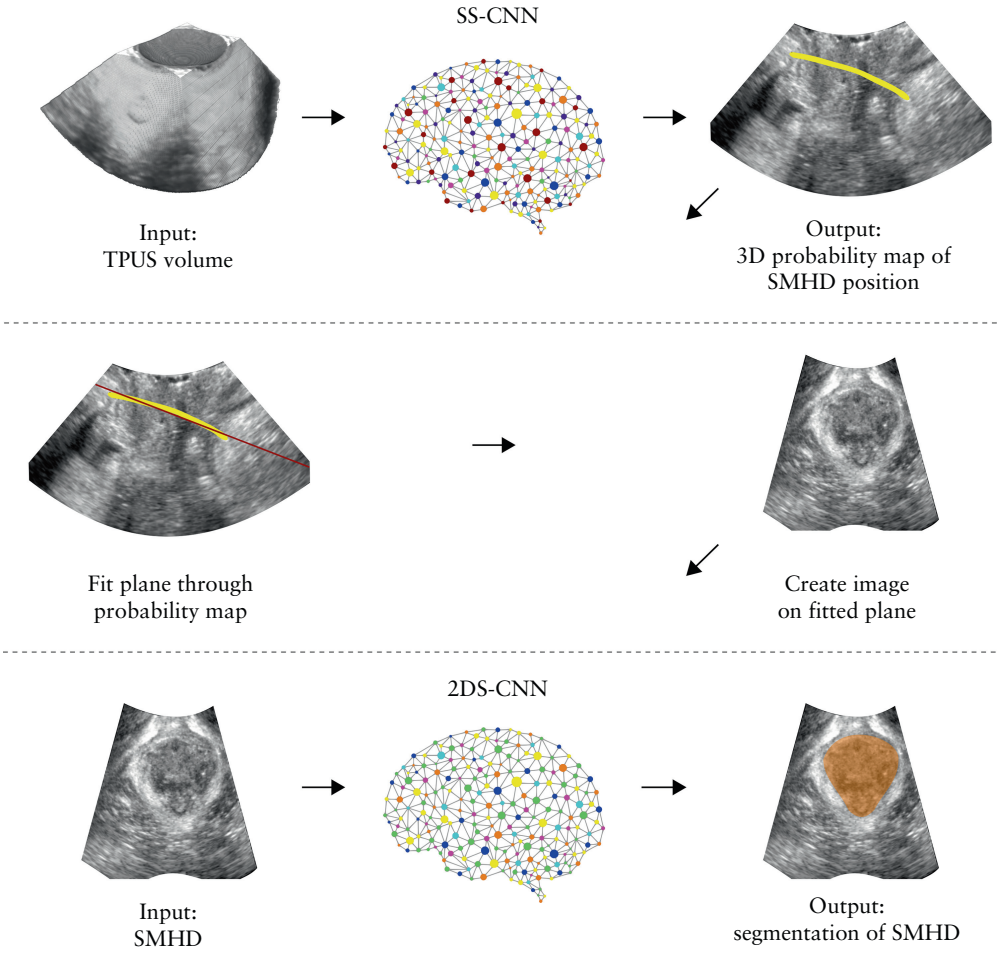

Figure 1 Pipeline for selection of the slice of minimal hiatal dimensions (SMHD) and segmentation of the urogenital hiatus in a transperineal ultrasound (TPUS) volume. A TPUS volume is fed into the trained SMHD-selection-convolutional-neural-network (SS-CNN). The SS-CNN provides a three-dimensional (3D) probability map (yellow line) of the position of the urogenital hiatus in the SMHD in the TPUS volume. A plane (red) is fitted through the probability map, which is used to create a two-dimensional (2D) image of the SMHD, provided that the SS-CNN has identified the urogenital hiatus correctly. The SMHD is fed into the trained 2D-segmentation CNN (2DS-CNN), which performs segmentation of the urogenital hiatus (orange area) in the SMHD. From this segmentation, the area, anteroposterior diameter and coronal diameter of the urogenital hiatus are calculated.

For the current study, the network was trained (with DSI-LF) on the manually selected 2D-SMHD to perform UH segmentation. Based on the output of this CNN, relevant parameters, including UHA and anteroposterior (APD) and coronal diameter (CD) of the UH, were measured automatically.

Validation

The different steps of the pipeline were validated separately to ensure their proper functioning. The functioning of the SS-CNN and plane fitting was validated by measuring the mean absolute distance (MAD) and the Hausdorff distance (HDD), i.e. the maximum absolute distance, between the manually and automatically selected SMHD of the test set. In addition, a non-quantitative, visual inspection of the manually and automatically selected SMHD was performed.

The 2DS-CNN was validated by applying it to the SMHD of the test set. The overlap between automatic and manual UH segmentation was quantified using the DSI, according to the following formula: $DSI = 2(X \cap Y)/(X + Y)$, where $X \cap Y$ is the number of overlapping pixels of the two segmentations, and X and Y are the number of pixels of the two segmentations, respectively. A DSI of 1 represents maximum segmentation overlap, while a DSI of 0 indicates no segmentation overlap.

The results of the complete pipeline were investigated by comparing the manual and automatic measurements of UHA, APD and CD. The intraclass correlation coefficients (ICCs) with 95% CI were calculated for each parameter and evaluated according to the subgroup definitions of Landis and Koch[22]. Box plots were created to compare the distribution of the manual and automatic measurements, and the mean difference and limits of agreement were investigated using Bland–Altman analysis[23].

RESULTS

Table 1 shows the demographic and clinical characteristics of the included patients ($n = 116$). Mean ± SD age was 59.5 ± 11.8 years and mean ± SD body mass index was 24.7 ± 3.6 kg/m^2. The majority (98.3%) of women were vaginally parous and 45 (38.8%) had complete LAM avulsion. Two (1.7%) women had Stage-I POP, 67 (57.8%) had Stage-II POP and 47 (40.5%) had Stage-III POP.

In some cases, data from the same patient were acquired and analyzed more than once, resulting in 423 frames on which the 3D position of the SMHD was saved successfully (150 at rest, 137 on maximum pelvic floor contraction, 136 on maximum Valsalva maneuver). The 2D segmentations were not always saved successfully, resulting in a dataset of 112 women and 373 training images.

Figure 2 shows all manually and automatically selected SMHD in the test set for visual comparison. All DSI values between manual and automatic UH segmentations were above 0.85, and the mean and median MAD between the manually and automatically selected SMHD was 0.20 and 0.14 cm, respectively. The mean and median HDD was 0.50 and 0.39 cm, respectively (Figure 3).

Table 1 Demographic and clinical characteristics of 116 included patients with symptomatic pelvic organ prolapse (POP)

Parameter	*Value*
Age (years)	59.5 ± 11.8
BMI (kg/m^2)	24.7 ± 3.6
Postmenopausal	88 (75.9)
Vaginally parous	114 (98.3)
Hysterectomy	16 (13.8)
POP surgery*	12 (10.3)
Incontinence surgery	3 (2.6)
Type of POP	
Anterior	66 (56.9)
Apical	6 (5.2)
Posterior	11 (9.5)
Anterior and apical	4 (3.4)
Anterior and posterior	21 (18.1)
Apical and posterior	3 (2.6)
Anterior, apical and posterior	5 (4.3)
POP stage	
I	2 (1.7)
II	67 (57.8)
III	47 (40.5)
LAM avulsion	45 (38.8)

Data are presented as mean ± SD or *n* (%). *Hysterectomy excluded. BMI, body mass index; LAM, levator ani muscle.

To validate the performance of the complete pipeline, the manual and automatic measurements of UHA, APD and CD were compared. All parameters were visualized using boxplots (Figure 4), and the agreement between manual and automatic UHA measurements was also evaluated using the Bland–Altman method (Figure 5). The ICC values between manual and automatic UH measurements were 0.94 (95% CI, 0.87–0.97) for UHA, 0.92 (95% CI, 0.78–0.97) for APD and 0.82 (95% CI, 0.66–0.91) for CD (Table 2).

DISCUSSION

In this study, we present a pipeline for fully automated SMHD selection and UH segmentation, which allows automatic measurement of UHA, APD and CD. An automated selection of the SMHD has not been reported previously in the literature. We have presented the performance results of each CNN and the accuracy of the measurements by the complete pipeline. The validation analysis comparing automatic and manual measurements showed excellent agreement, demonstrating high reliability of the automatic pipeline.

The selection of the SMHD is difficult because it is not aligned with any principal anatomical plane (i.e. coronal, sagittal or axial plane). In the literature, there have been several studies of slice detection methods for other imaging tasks, such as L3 slice detection in computed tomography data[24,25] and slice detection in fetal ultrasound data[26–29]. Among these studies, only Li *et al.*[28] described the method

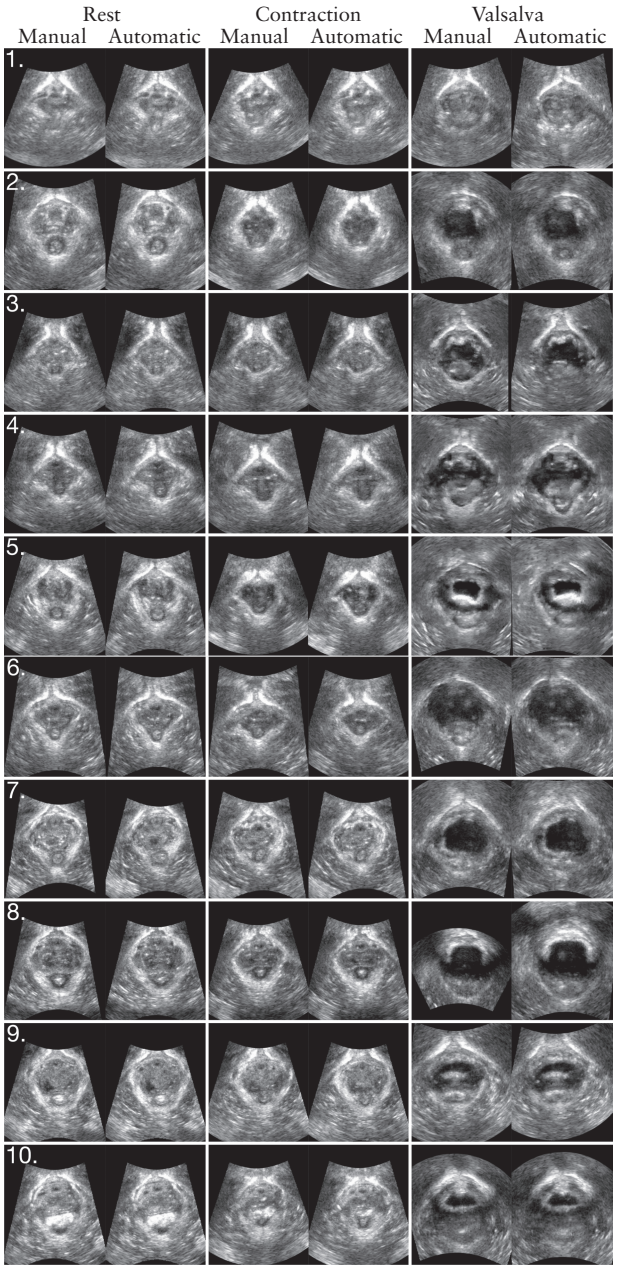

Figure 2 Manually and automatically selected slices of minimal hiatal dimensions, obtained during rest, on maximum pelvic floor contraction and on maximum Valsalva maneuver from all patients (Cases 1–10) in the test set.

for detecting a slice that is not aligned with any principal anatomical plane. Although their approach was different from the one presented in this work, their results in terms of distance between manually and automatically selected slices were comparable to ours, with a slightly better agreement in this study (average MAD of 0.20 cm in this study *vs* 0.34 cm and 0.35 cm in the study by Li *et al.*[28]). However, we restricted the distance calculation to the area of the manual UH segmentation, which is our region of interest, instead of calculating the average distance between manual and automatic SMHD. This restriction may have had a slight positive influence on our results.

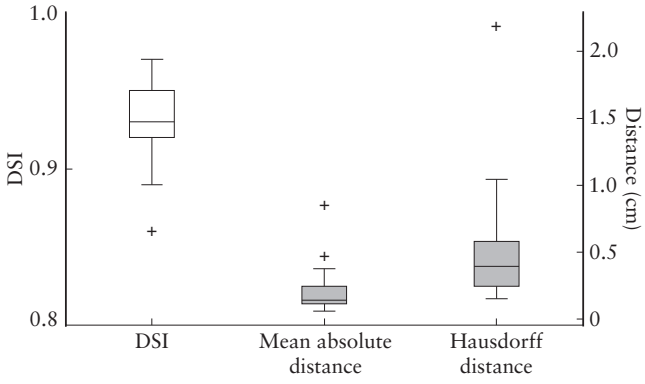

Figure 3 Overlap between manual and automatic segmentation of the urogenital hiatus, expressed as dice similarity index (DSI), and mean absolute distance and Hausdorff distance between the manually and automatically selected slices of minimal hiatal dimensions. Boxes are median and interquartile range, and whiskers are range after excluding outliers (+) lying outside 1.5 × interquartile range. Shading denotes distance measurements.

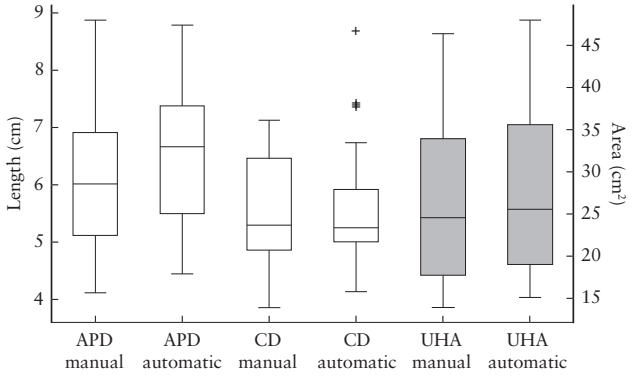

Figure 4 Manual and automatic measurements of anteroposterior diameter (APD), coronal diameter (CD) and area (UHA) of the urogenital hiatus in the test set. Boxes are median and interquartile range, and whiskers are range after excluding outliers (+) lying outside 1.5 × interquartile range. Shading denotes area measurements.

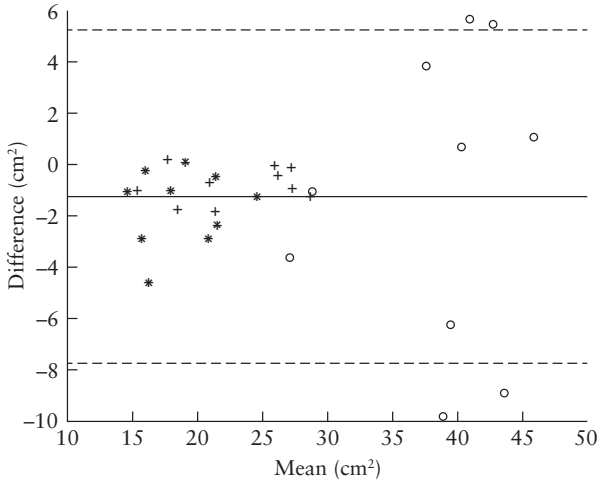

Figure 5 Bland–Altman plot of urogenital hiatal area measurements in the test set obtained during rest (+), on maximum pelvic floor contraction (∗) or on maximum Valsalva maneuver (○). Mean difference (——) and limits of agreement (– – -) are also shown.

Table 2 Comparison of manual and automatic measurements of urogenital hiatal area, anteroposterior diameter and coronal diameter

Measurement	*Automatic pipeline*	*Manual*	*Mean difference (LOA)*	*ICC (95% CI)*
Urogenital hiatal area (cm^2)	27.4 ± 9.8	26.1 ± 10.1	−1.3 ± 3.3 (−7.7 to 5.2)	0.94 (0.87–0.97)
Anteroposterior diameter (cm)	6.5 ± 1.2	6.2 ± 1.3	−0.3 ± 0.4 (−1.1 to 0.6)	0.92 (0.78–0.97)
Coronal diameter (cm)	5.6 ± 0.9	5.5 ± 1.0	−0.1 ± 0.6 (−1.3 to 1.1)	0.82 (0.66–0.91)

Data are given as mean ± SD, unless stated otherwise. ICC, intraclass correlation coefficient; LOA, limits of agreement.

Only one automatically selected SMHD was more than 0.5 cm away from the manually selected SMHD (Figure 3), which corresponds to the image of Case 8 on maximum Valsalva maneuver in Figure 2. However, this difference had little influence on the UH measurements. The clear outlier among CD measurements (Figure 4) corresponds to the CD measurement on maximum Valsalva maneuver in Case 10 (Figure 2) and may be explained by the unclear borders of the automatically selected SMHD, which hindered proper automated segmentation.

The UHA measurements obtained on maximum Valsalva maneuver had the largest differences between manual and automatic assessments (Figure 5). However, the average UHA measurements on Valsalva maneuver of most women in the test set were almost twice as large as those obtained on maximum pelvic floor contraction and during rest. Therefore, the error appears to be proportional to the size of the measured area. There was an overall trend of slight overestimation of UHA (average of 1.3 cm^2) by the automatic pipeline. This error is similar to the interobserver variability for manually performed UHA measurements[30].

The overlap results of the 2DS-CNN are comparable to those of other automated and semiautomated methods presented in the literature[12–14], which report average DSI values of 0.92–0.94 (0.93 in this study). The ICC for UHA, CD and APD measurements between manual analysis and the fully automated pipeline were excellent, supporting our previous results on automatic segmentation of UHA, APD and CD[14]. This indicates that automation of slice selection (SS-CNN) does not have a negative impact on the segmentation results. The interobserver variability for the manually performed measurements differs substantially across studies[9,10,30,31]. Our ICC values are higher than those reported in the literature, proving the success of the presented automated SMHD selection and segmentation pipeline.

The presented fully automated pipeline requires a few seconds to process a single volume. Therefore, it can be implemented in the software of TPUS machines, which should make the analysis of TPUS data less time-consuming and observer-dependent, thus reducing clinical training and analysis time and simplifying the examination of TPUS data for research and clinical purposes. This will lower the barriers that clinicians may experience when using TPUS in their clinical practice.

The data of this study were acquired from women with symptomatic POP, which are more complex to analyze compared with data of women with intact LAM and/or without pelvic floor dysfunction[9]. Due to its reproducibility, this pipeline is an excellent tool to standardize the UH measurements in complex patient populations, making them less observer-dependent. However, it should be noted that the training data of the pipeline were based on the manual analysis carried out by a single experienced observer. Therefore, it can be assumed that the pipeline is biased towards the way this specific observer analyzed the data. Before implementation into clinical practice, we recommend to add training data from multiple observers, which should eliminate personal bias in the network.

For a more reliable analysis of minimal hiatal dimensions on Valsalva maneuver, a rendered volume[32] and OmniView combined with volume contrast imaging[33] have been suggested in the literature. However, those methods require the (approximate) position of the SMHD to compute their interpolated 2D images. Since these types of images look very similar to a single slice SMHD, we expect that the 2D segmentation results using these images would be similar to the results of this study.

Even when the pipeline generates some errors, these can be identified easily in clinical practice by a quick visual examination of the selected and segmented SMHD. In these cases, manual analysis is recommended, which can also be used to update the pipeline, making it more robust over time. The only step that still needs automation is the selection of the correct frame (i.e. rest, pelvic floor contraction or Valsalva maneuver). However, this is the least time-consuming step in the manual examination and, based on the literature[26,27,29], it can be expected that its automation is feasible.

In conclusion, we have presented a pipeline that selects and segments the SMHD reliably and thus automates the analysis of the SMHD on TPUS in women with symptomatic POP. Implementation of this pipeline in the software of TPUS machines should make the analysis of TPUS data less time-consuming and observer-dependent. This should reduce clinical training and analysis time and simplify the examination of TPUS data for research and clinical purposes.

REFERENCES

1. Dietz HP. Pelvic Floor Ultrasound: A Review. *Clin Obstet Gynecol* 2017; **60**: 58–81.
2. Dietz HP, Shek C, De Leon J, Steensma AB. Ballooning of the levator hiatus. *Ultrasound Obstet Gynecol* 2008; **31**: 676–680.
3. Dietz HP, Simpson JM. Levator trauma is associated with pelvic organ prolapse. *BJOG* 2008; **115**: 979–984.
4. Grob ATM, Veen AAC, Schweitzer KJ, Withagen MIJ, van Veelen GA, van der Vaart CH. Measuring echogenicity and area of the puborectalis muscle: Method and reliability. *Ultrasound Obstet Gynecol* 2014; **44**: 481–485.
5. Dietz HP. Ultrasound in the assessment of pelvic organ prolapse. *Best Prac Res Clin Obstet Gynaecol* 2019; **54**: 12–30.
6. Abdool Z, Shek KL, Dietz HP. The effect of levator avulsion on hiatal dimension and function. *Am J Obstet Gynecol* 2009; **201**: 89.e1–5.

7. DeLancey JOL, Morgan DM, Fenner DE, Kearney R, Guire K, Miller JM, Hussain H, Umek W, Hsu Y, Ashton-Miller JA. Comparison of levator ani muscle defects and function in women with and without pelvic organ prolapse. *Obstet Gynecol* 2007; **109**: 295–302.
8. Dietz HP. Pelvic Floor Ultrasound. In: *Female Urology*, Raz S, Rodríguez LV (eds). Elsevier: Philadelphia, PA, 2008; 100–124.
9. van Veelen GA, Schweitzer KJ, van der Vaart CH. Reliability of pelvic floor measurements on three- and four-dimensional ultrasound during and after first pregnancy: Implications for training. *Ultrasound Obstet Gynecol* 2013; **42**: 590–595.
10. Brækken IH, Majida M, Ellstrøm-Engh M, Dietz HP, Umek W, Bø K. Test-Retest and intra-observer repeatability of two-, three- and four- dimensional perineal ultrasound of pelvic floor muscle anatomy and function. *Int Urogynecol J Pelvic Floor Dysfunct* 2008; **19**: 227–235.
11. Tan L, Shek KL, Atan IK, Rojas RG, Dietz HP. The repeatability of sonographic measures of functional pelvic floor anatomy. *Int Urogynecol J* 2015; **26**: 1667–1672.
12. Sindhwani N, Barbosa D, Alessandrini M, Heyde B, Dietz HP, D'Hooge J, Deprest J. Semi-automatic outlining of levator hiatus. *Ultrasound Obstet Gynecol* 2016; **48**: 98–105.
13. Bonmati E, Hu Y, Sindhwani N, Dietz HP, D'hooge J, Barratt D, Deprest J, Vercauteren T. Automatic segmentation method of pelvic floor levator hiatus in ultrasound using a self-normalizing neural network. *J Med Imaging (Bellingham)* 2018; **5**: 021206.
14. van den Noort F, van der Vaart CH, Grob ATM, van de Waarsenburg MK, Slump CH, van Stralen M. Deep learning enables automatic quantitative assessment of puborectalis muscle and urogenital hiatus in plane of minimal hiatal dimensions. *Ultrasound Obstet Gynecol* 2019; **54**: 270–275.
15. Manzini C, van den Noort F, Grob ATM, Withagen MIJ, van der Vaart CH. The effect of pessary treatment on puborectalis muscle function. *Int Urogynecol J* 2021; **32**: 1409–1417.
16. Manzini C, Withagen MIJ, van den Noort F, Grob ATM, van der Vaart CH. Transperineal ultrasound to estimate the appropriate ring pessary size for women with pelvic organ prolapse. *Int Urogynecol J* 2022; **33**: 1981–1987.
17. Ritter F, Boskamp T, Homeyer A, Laue H, Schwier M, Link F, Peitgen HO. Medical image analysis. *IEEE Pulse* 2011; **2**: 60–70.
18. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 2012; **1**: 1097–1105.
19. Litjens G, Kooi T, Bejnordi BE, Arindra A, Setio A, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Sanchez CI. A Survey on Deep Learning in Medical Image Analysis. *Med Image Anal* 2017; **42**: 60–88.
20. van den Noort F, Sirmacek B, Slump CH. Recurrent U-net for automatic pelvic floor muscle segmentation on 3D ultrasound. 2021. DOI: 10.48550/arXiv.2107.13833.
21. Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of the 4th International Conference on 3D Vision, IEEE, Stanford, CA, USA, 2016; 565–571.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
23. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurements. *Lancet* 1986; **327**: 307–310.
24. Belharbi S, Chatelain C, Hérault R, Adam S, Thureau S, Chastan M, Modzelewski R. Spotting L3 slice in CT scans using deep convolutional network and transfer learning. *Comput Biol Med* 2017; **87**: 95–103.
25. Kanavati K, Islam S, Aboagye EO, Rockall A. Automatic L3 slice detection in 3D CT images using fully-convolutional networks. 2018. DOI: 10.48550/arXiv.1811.09244.
26. Chen H, Dou Q, Ni D, Cheng JZ, Qin J, Li S, Heng PA. Automatic fetal ultrasound standard plane detection using knowledge transferred recurrent neural networks. In: *Medical Image Computing and Computer-Assisted Intervention*, Navab N, Hornegger J, Wells W, Frangi A (eds). MICCAI 2015, Munich, Germany; 507–514.
27. Baumgartner CF, Kamnitsas K, Matthew J, Smith S, Kainz B, Rueckert D. Real-time standard scan plane detection and localisation in fetal ultrasound using fully convolutional neural networks. In: *Medical Image Computing and Computer-Assisted Intervention*, Ourselin S, Joskowicz L, Sabuncu M, Unal G, Wells W (eds). MICCAI 2016, Athens, Greece; 203–211.
28. Li Y, Khanal B, Hou B, Alansary A, Cerrolaza JJ, Sinclair M, Matthew J, Gupta C, Knight C, Kainz B, Rueckert D. Standard plane detection in 3D Fetal Ultrasound using an iterative transformation network. In: *Medical Image Computing and Computer-Assisted Intervention*, Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G (eds). MICCAI 2018, Grenada, Spain; 392–400.
29. Qu R, Xu G, Ding C, Jia W, Sun M. Standard plane identification in fetal brain ultrasound scans using a differential convolutional neural network. *IEEE Access* 2020; **8**: 83821–83830.
30. Dietz HP, Shek C, Clarke B. Biometry of the pubovisceral muscle and levator hiatus by three-dimensional pelvic floor ultrasound. *Ultrasound Obstet Gynecol* 2005; **25**: 580–585.
31. Weinstein MM, Jung SA, Pretorius DH, Nager CW, den Boer DJ, Mittal RK. The reliability of puborectalis muscle measurements with 3-dimensional ultrasound imaging. *Am J Obstet Gynecol* 2007; **197**: 68.e1–6.
32. Dietz HP, Wong V, Shek KL. A simplified method for determining hiatal biometry. *Aus N Z J Obstet Gynaecol* 2011; **51**: 540–543.
33. Youssef A, Montaguti E, Sanlorenzo O, Cariello L, Elbadawy Awad E, Pacella G, Ghi T, Pilu G, Rizzo N. A new simple technique of 3-dimensional sonographic assessment of the pelvic floor muscles. *J Ultrasound Med* 2015; **34**: 65–72.

SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:

Appendix S1 The loss function used for the convolutional neural network for selection of the slice of minimal hiatal dimensions (SS-CNN)