Semi-Automated Detection of Bladder Neck Funneling and Measurement of Posterior Urethrovesical Angle in Females

Megan Vandermolen

A thesis submitted in partial fulfillment of the requirements for the
Master's degree of Applied Science in Biomedical Engineering

Ottawa-Carleton Institute for Biomedical Engineering
University of Ottawa
Ottawa, Canada
April 2022

Abstract

The pathophysiology of stress urinary incontinence is poorly understood but bladder neck funneling (BNF) and posterior urethrovesical angle (PUVA) enlargement have been implicated. Methods to measure these phenomena are poorly established. The aim of this thesis was to develop and evaluate a semi-automated method to analyze BNF and PUVA from ultrasound images acquired transperineally and test its repeatability and concurrent validity compared to manual segmentation. Agreement between the semi-automated and manual methods was assessed by kappa statistics and intraclass correlation coefficients (ICCs). The repeatability of detection of BNF using the semi-automated approach was almost perfect (κ_C = 1.00 ($p<0.001$)), while the reliability of semi-automated detection of PUVA was good (ICC(3,1) = 0.860 (0.784 – 0.910)). Concurrent validity of BNF classification was almost perfect (κ_L = 1.00 ($p<0.001$)), while PUVA estimation was moderate (ICC(2,1) = 0.610 (0.514 – 0.705)). The method presented here is an acceptable proof of concept; further development is recommended.

Acknowledgements

A thesis is a rollercoaster; it is a process during which you grow as a student, as a researcher, and as a person. One thing is for certain, it is impossible to complete a thesis without being supported, pushed, and carried by the people around you: both those who chose to be part of the process and those who did not.

I want to start off by thanking my thesis supervisor, Dr. Linda McLean, for the opportunity to work in the McLean Function and Measurement Lab (MFM Lab) at the University of Ottawa. She was incredibly patient with me as I learned how to develop a feasible research project from nebulous, and often unconnected, ideas. Linda was unwavering in her confidence that I would finish as planned, despite frequent Covid- and non-Covid-related setbacks. I'd also like to thank Dr. Adrian Chan, my co-supervisor, who began supporting me long before he became my co-supervisor. Adrian's ability to show me the path to submission and his extensive knowledge of image processing truly propelled this project from conception to execution. Linda and Adrian, thank you for your knowledge, guidance, and advice throughout this thesis. Thank-you, also, to the members of my committee, Dr. Duane Hickling and Dr. Leila Mostaço-Guidolin, for donating their time and expertise to aid in the evaluation and completion of this thesis.

I would also like to sincerely thank Marie-Ève Bérubé for sharing the amazing dataset she acquired with me. Without her generosity, I would have no doubt been further delayed by the Covid-19 pandemic. Similarly, I'd like to thank Kisanet Kebedom who dedicated countless hours in contribution to this thesis and kept a positive outlook despite multiple last-minute revisions to her mandate.

Thanks to Ana, Cat, Layla and the rest on my MFM lab mates for their energy, support, and insight. I truly could not have done it without our 8-hour Teams calls to keep me motivated and on track. Lastly, I want to thank my friends and family for their ongoing support. I could not have stayed sane through this without your love and acts of kindness. Thank-you particularly to Viran, Dad, Barbi, Josie, Auntie Jenny, Grandma, Grandad, Auntie Sarah, Auntie Visaka, and Uncle Vishvajit.

Table of Contents

List of Figures

List of Tables

Glossary

AC – Active contour

AER – Area error rate

ANOVA – Analysis of variance

BNF – Bladder neck funneling

CI – Confidence interval

CNN – Convolutional neural network

ICC – Intraclass correlation coefficient

ISD – Intrinsic sphincter deficiency

LUT – Lower urinary tract

MRI – Magnetic resonance imaging

MUCP – Maximum urethral closure pressure

PUVA – Posterior urethrovesical angle

PUVJ – Posterior urethrovesical junction

REB – Research ethics board

SEM – Standard error of the measurement

SUI – Stress urinary incontinence

TAR – Triangle area representation

USI – Ultrasound imaging

VFUD – Videofluorourodynamic

VUD – Videourodynamics

Chapter 1 INTRODUCTION

The two main components of the lower urinary tract are the bladder and urethra [1]. The bladder functions to store liquid waste until it is time to void. Bladder volume is dynamic and can expand to approximately 500 mL [2]. The urethra is a complex tubular structure that functions to maintain urine in the bladder until it is time to void [1]. It has a combination of smooth (involuntary control) and striated (voluntary control) musculature organized both circularly and longitudinally [3], [4]. Bladder filling is accomplished without sensation or rise in pressure until a critical volume is reached, at which time one begins to feel the urge to void; this urge can be delayed for an hour or more [5] until an appropriate time and location are identified. Then there is conscious control of urethral relaxation followed by bladder muscle contraction which allows urine to flow out of the body [5].

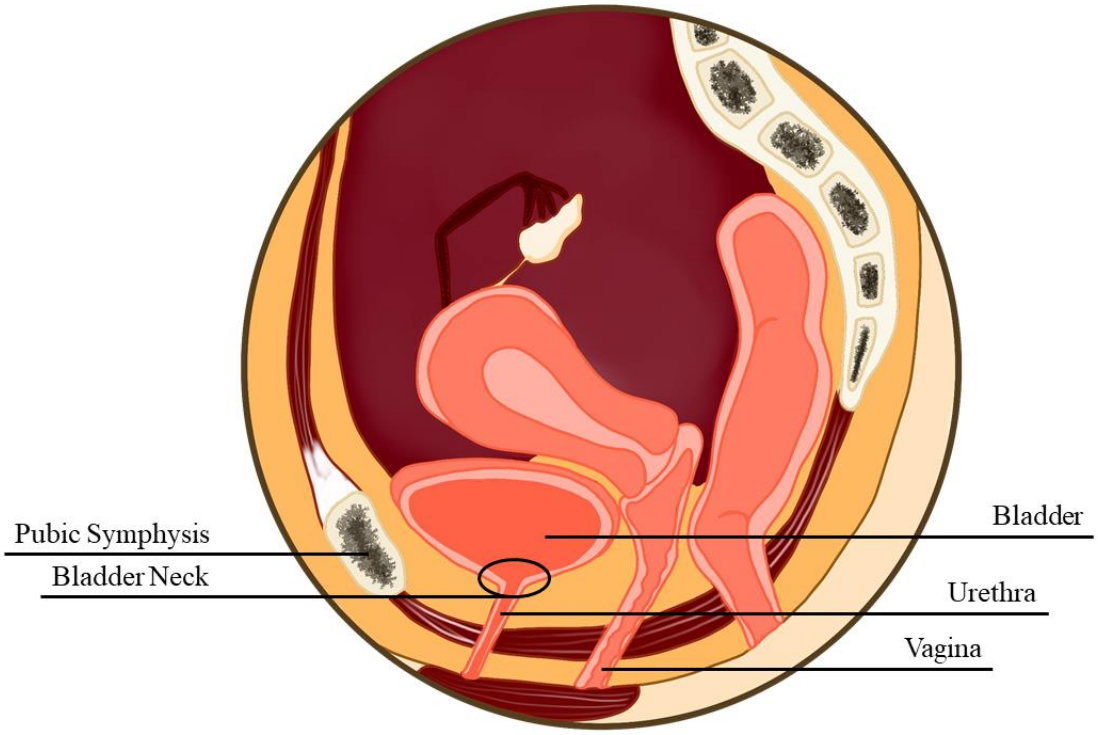


Figure 1.1: Female pelvis [6] *in midsagittal view with annotations*

1.1 NORMAL CONTROL OF MICTURITION

During urine storage, there are numerous mechanisms thought to maintain continence; in females the urethra remains closed through sphincteric action and external compression of the urethra between the anterior vaginal wall and the pubic symphysis (Figure 1.1) [3]–[5],

[7], [8]. Intrinsically, there is a gradual increase in muscle activity in the urethral sphincters, which act to close the lumen and abate further bladder contraction [5]. Extrinsically, if the urethral supports are intact, there is a transmission of caudally-directed forces (from increased intra-abdominal pressure or gravitational loading) which act to compress the urethra against the endopelvic fascia to enhance closure of the urethral lumen [7].

The bladder neck (Figure 1.1) denotes the area, rather than a specific structure, where the urethra meets the bladder base [1]. It is the region where detrusor (bladder) muscle surrounds the trigonal ring and internal urinary meatus [1]. Support for the bladder neck relies on the endopelvic fascia and fascial connections to the arcus tendinous fascia pelvis [4]. This is thought to provide a backstop against abdominal pressure that can compress the bladder neck and urethra to maintain continence [4].

Urinary continence control is learned [9]. Urethral closure pressure must be larger than bladder pressure to maintain continence [4]. This is true both during periods of rest and periods of exertion [4], whereby the continence control mechanisms may be challenged due to rapid rises in intra-abdominal pressure.

When appropriate, voiding occurs through pelvic floor muscle relaxation and/or abdominal straining [10]–[12]. Additionally, during voiding, the bladder neck and the proximal urethra dilate, assuming a funnel-like shape after the initiation of the bladder contraction [5], [13].

1.2 URINARY INCONTINENCE

Urinary incontinence is defined as uncontrolled urine leakage. It can occur for a variety of reasons but generally falls under two main types with different etiologies [14]: 1) stress urinary incontinence (SUI), and 2) urgency urinary incontinence. SUI is defined as the involuntary loss of urine due to an increase in physical exertion [15] and is the most common form of urinary incontinence in females [16]. Urgency incontinence is defined as urine leakage associated with a sudden and strong desire to void that is difficult to resist [14]. These subtypes are not mutually exclusive and often co-occur as mixed urinary incontinence [14]. Recent evidence suggests that SUI and urgency urinary incontinence may have similar etiologies: failure of the urethral closure system [3]. While urinary incontinence

in females is multifactorial, it appears to be related to dysfunction of the bladder, urethra, and pelvic support including fascia and the pelvic floor muscles [4], [17].

Estimates of the prevalence of SUI vary widely [18]–[21], in part because of variations among definitions and populations studied. A recent study from India revealed an observed prevalence of 3% in females aged 30 to 70 [18]. Whereas, Hannestad et al. observed a prevalence of 50% in Norwegian females over the age of 20, with the highest prevalence (65%) occurring among participants aged 45 to 49 [20]. In a more recent review conducted in partnership with the Fourth International Consultation on Incontinence, the prevalence of SUI was observed to range between 4 to 23% in females aged 20-39, 16 and 36% in females aged 40 to 59, and 8 to 23% in females aged 60 to 79 [21]. Additionally, reported prevalence is thought to be underestimated as many females do not report or seek treatment for SUI [14], [22]. Key risk factors for SUI are well documented and include older age, high body mass index (BMI), a history of pregnancy, and a history of vaginal delivery [19], [23]. Urine leakage during physical activity may have a different etiology, as the reported prevalence of urinary incontinence among females who participate in high impact activities is 19.9% to 38.6%, despite most studies examining populations who do not have any of the key risk factors for urinary incontinence [24].

SUI has been shown to have a negative effect on quality of life [18], [19], [25], [26]. While examining American Nurses aged 25 to 42, Minassian et al. determined that 56.6% of nurses with SUI rated their experience as “slightly” bothered while 22.6% rated their experience as “moderately” bothered [25]. In Harris’ examination of the female population of Göteborg, Sweden (aged 65-84 years), females with SUI scored higher (worse) on the Nottingham Health Profile than a representative sample of the population, particularly in terms of lack of energy, emotional disturbance, and mobility problems while also reporting significantly higher social isolation [26].

Although effective interventions for SUI exist, many females are left with continued symptoms [3], [27]. Through meta-analysis, a Cochrane review of 31 trials involving 1817 females from 14 countries found that pelvic floor muscle training in females with SUI resulted in 56% cured cases and 74% cured or improved cases [27]. An assessment of the success of midurethral slings by Richter et al. found that the subjective success of treatment in 597 females with SUI to be 62.2% for retropublic-slings and 55.8% for transobturator-

slings [28]. Where subjective success was defined as the absence of self-reported symptoms of SUI assessed by a validated questionnaire, no leakage recorded in a 3-day bladder diary, and no retreatment for SUI [28]. Intervention success rates may be improved by addressing the specific pathophysiology causing SUI on a case-by-case basis.

However, the pathophysiology of SUI is not fully understood and is thought to be multifactorial [29], involving defects in the urinary sphincters [3], [4], pelvic floor muscles, endopelvic fascia, etc. [4], [7]. A thorough understanding of lower urinary tract anatomy and physiology in symptomatic and asymptomatic females is paramount to understanding the pathophysiology of SUI [30]. Standardized methods to assess the different mechanisms and features involved in continence control may help to identify the underlying causes for SUI in specific females. Further, this may allow clinicians to individualize treatment with new and existing techniques to target their specific deficiencies. In order to achieve this, clinician's must first have reliable and valid methods to measure features associated with SUI.

In particular, two features of interest which appear to be related to SUI are bladder neck funneling (BNF) and posterior urethrovesical angle (PUVA) [29]. Indeed, a recent systematic review and meta-analysis suggests that BNF is five times more prevalent in women with SUI compared to continent women [29]. Similarly, the literature is in agreement that females with SUI have larger PUVAs at rest and on strain than continent females [31]–[37], confirmed through a meta-analysis [29]. To this author's knowledge, BNF and PUVA are analyzed manually. A computer-aided analysis approach could provide time and cost savings.

The aim of this thesis is to develop and evaluate a semi-automated method to classify BNF and estimate PUVA on 2D transperineal ultrasound images in females. The specific objectives of this thesis research were to:

1. Develop a manual analysis approach for BNF classification and PUVA estimation and evaluate its intra- and inter-rater repeatability.
2. Develop an approach for image optimization, semi-automated segmentation, and feature extraction capable of BNF classification and PUVA estimation, and evaluate the repeatability and the concurrent validity of the semi-automated method when compared with manual raters.

Chapter 2 LITERATURE REVIEW

In this review, the utility of imaging techniques for the evaluation of the lower urinary tract are summarized. Then, the relevance of BNF and PUVA in relation to SUI is explored, and definitions of these features as used in the literature are presented and discussed. The effect of measurement conditions (i.e., body position and bladder volume) on observability and magnitude of BNF and PUVA are summarized. The repeatability of different approaches for the classification of BNF and estimation of PUVA are detailed. Lastly, relevant automated and semi-automated approaches to segment the bladder and gallbladder are described forming the basis for the development of the segmentation and estimation approach presented through this thesis.

2.1 IMAGING AS A TOOL IN UROGYNECOLOGY

The use of imaging techniques for the observation of the lower urinary tract (LUT) in patients with urinary incontinence dates back to the 1950s, and although some of these techniques are over 50 years old, their clinical value remains unclear [38]. Indeed, clinical interpretation of images of the LUT in patients with urinary incontinence is limited by our limited understanding of the pathophysiology of urinary incontinence [39]. Insight into the relevance of ultrasound imaging to the management of urinary incontinence requires a thorough understanding of the potential mechanisms involved. The identification of specific morphologic findings that are associated with SUI may lead to a more nuanced and personalized approach to the management of stress urinary incontinence. Imaging of the LUT can be categorized by imaging modality: x-ray, magnetic resonance imaging, and ultrasound imaging [38].

2.1.1 X-Ray Imaging

X-ray imaging is generally used to visualize bony structures in the body and is not ideal for the visualization of soft tissue. However, if a contrast medium is injected, organs such as the bladder and urethra can be visualized [40]. This allows for the analysis of bladder neck mobility, bladder diverticula, etc. [40].

2.1.1.1 Cystourethrography

Cystourethrography is an x-ray imaging technique used to visualize the bladder and urethra after injecting these organs with a contrast medium [41]. It was introduced over 70 years ago [38] and is still in use today. X-ray imaging of the LUT mainly consists of voiding

cystourethrograms for the quantification of bladder neck mobility, which is thought to be key in surgical patient management [39]. While bladder neck displacement is best viewed in the true lateral projection [38], the image quality in this projection is often poor because of the soft tissue and bony structure overlap with the bladder neck area [38]. The advent of digital imaging has started to combat this issue by allowing for digital filtering to remove the influence of soft and bony tissue on the visualization of the bladder and urethra [38]. However, the literature suggests that cystourethrography is not discriminant between continent and incontinent females, and that inter- and intra-rater agreement is not adequate for clinical usefulness [39]. Depending on available medical facilities, however, cystourethrography may be considered for the evaluation of patients with urinary incontinence if the choice of surgical procedure depends on specific supporting tissue deficiencies [38].

2.1.1.2 Videourodynamics

Videourodynamic imaging (VUD) is the synchronous x-ray imaging during measurement of abdominal, bladder, and urethral pressures [40]. VUD has been considered the ‘gold-standard’ of urodynamic evaluation of the LUT, yet it does not play a major role in the primary assessment of female patients presenting with incontinence [38]. VUD may, however, be reasonable in the assessment of more complex cases [38]. In requesting VUD, the benefits must be weighed against the risk associated with radiation exposure as well as the expense and human resources required [40], [42].

2.1.2 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) maps 3-D anatomy and displays it in standard orientations [30]. MRI provides arguably the best images of pelvic and LUT anatomy, including bone, connective tissue, and fat [38] and thus has unparalleled potential as a research tool [30]. In addition to the high resolution of static MRI, MRI using ultra-fast image acquisition in a single plane can now be used to capture morphological changes during functional tasks [30], [38] such as bearing down or contracting the pelvic floor muscles. Indeed, the role of MRI in evaluating the LUT for urethral diverticula, bladder tumors, etc. has been well established [38] and has demonstrated clinical utility to direct surgical approach during the preoperative evaluation of females with multicompartment pelvic organ prolapse [30]. However, MRI is expensive and, to the best of our knowledge, is

not indicated in the evaluation of typical female patients presenting with symptoms of SUI; it is more useful in research for assessing specific anatomical and physiological deficiencies associated with SUI.

Although MRI provides high-resolution images, imaging artifacts are still present [30]. The basic principle of acquisition is to capture a sequence of orthogonal slices [30]. If the patient's position varies, there is a risk of error in linear and angular measurements, called a 'slice angle error' [30]. 'Chemical shift artifact' is due to "spatial misregistration of fat and water molecules" [43] and causes errors in the size of equivalent structures located on opposite sides of an axis [30].

Key limitations of MRI in the evaluation of LUT disorders are: 1) patients can often only be imaged in the supine position, therefore, negating the effects of gravity [44] that are highly relevant to the pathophysiology, and 2) the cost of acquisition is much higher than other imaging techniques, particularly ultrasound imaging. However, the method is considered safe for almost all patients [30] and presents no exposure to ionizing radiation [30], [44].

2.1.3 Ultrasound Imaging

Ultrasound imaging (USI) has been used in urogynecology since the early 1980s but has yet to become a standard tool in this area of clinical practice [45]. Over recent decades, image quality captured by USI has improved immeasurably, which has led to the widespread use of USI as an assessment tool to evaluate the LUT [38]. Indeed, in recent years USI has rapidly replaced x-ray imaging for visualizing the LUT; however, this does not imply any additional clinical utility [39]. Currently, USI is not recommended by the 6th International Consultation on Incontinence in the primary evaluation of patients with urinary incontinence [38].

A key limitation of USI is the need for the imaging probe to be in direct contact with the body [38]. Pressure applied by the probe to the tissue can cause distortion of superficial structures [46]. Additionally, image resolution is dependant on probe frequency, but as frequency is increased, tissue penetration decreases [38]. Another limitation of USI is that the echogenic properties of some tissues change as their orientation with respect to the probe changes [39]. Conversely, USI is often preferred by clinicians because imaging studies can be performed quickly and in their own offices [38]. Similar to MRI, USI presents no

exposure to ionizing radiation [38]. Additionally, USI is the least costly of the imaging methods used to visualize the LUT [38].

Structures within the pelvic cavity can be visualized using a number of different USI approaches, each with their own advantages and disadvantages [45]. Transperineal ultrasound most commonly utilizes a low frequency (3.5 – 6 MHz) curvilinear probe placed on the perineum in the midsagittal plane [45]. This technique allows visualization of the pubic symphysis, urethra, bladder, vagina, rectum, and anal canal [45] and can be performed in either supine or standing. Introital ultrasound uses a front-firing endo-vaginal probe placed at the vaginal introitus; this higher frequency probe provides higher resolution images with reduced penetration of the tissues [45]. Transvaginal ultrasound can be accomplished using endo-vaginal probes and is often used to provide high-resolution images of the anterior and posterior compartments [45]. Imaging can be performed using 2D or 3D probes in all of the above USI approaches, depending on the structures a clinician wishes to visualize; however, 3D imaging options currently remain limited in most clinical settings.

A robust method of segmenting the bladder and proximal urethra from USIs may offer valuable insights into the success or failure of different SUI interventions, while being both less costly and less invasive than other imaging techniques.

2.2 CONTINENCE CONTROL IN RESPONSE TO STRESS: THE RELATIVE IMPORTANCE OF URETHRAL CLOSURE AND URETHRAL SUPPORT

Despite decades of research devoted to the study of SUI, a complete understanding of the mechanisms of SUI has remained elusive [7]. It is well established that in order for continence to be maintained, urethral pressure must be greater than bladder pressure [4]. For many years, the prevailing theory surrounding the mechanism for maintaining this pressure-balance centred around urethral support [3], [7], [8]. Indeed, urethral hypermobility, due to defects in fascial integrity, was theorized to be the primary cause of SUI [7], [47], and surgical intervention in the form of mid-urethral sling insertion, which stabilizes the urethra [3], became the predominant surgical intervention [14]. This intervention was reported to be widely successful, with objective success rates of approximately 80% [28], firmly entrenching urethral hypermobility as the predominant cause of SUI. However, subjective success rates remain low, 62% [28], indicating continued symptoms not captured by office-

based ‘objective’ assessments. Indeed at this success rate, mechanisms other than failure of urethral support may be important drivers of SUI. A recent theory suggests that urethral failure, defined as the reduction in urethral closure pressure, may be a primary mechanism contributing to SUI [3]. This proposed shift highlights a need to better understand urethral function [3], and its role in continence.

2.2.1 The Urethral Support Mechanism

A popular theory regarding the urethral support mechanism is DeLancey’s “Hammock Hypothesis” [7]. In his investigation of female cadavers, he found that the urethra lies on a supportive layer composed of the endopelvic fascia and the anterior vaginal wall (Figure 2.1), and he theorized that it was against this layer that the urethra was compressed in reaction to forces directed caudally [7].

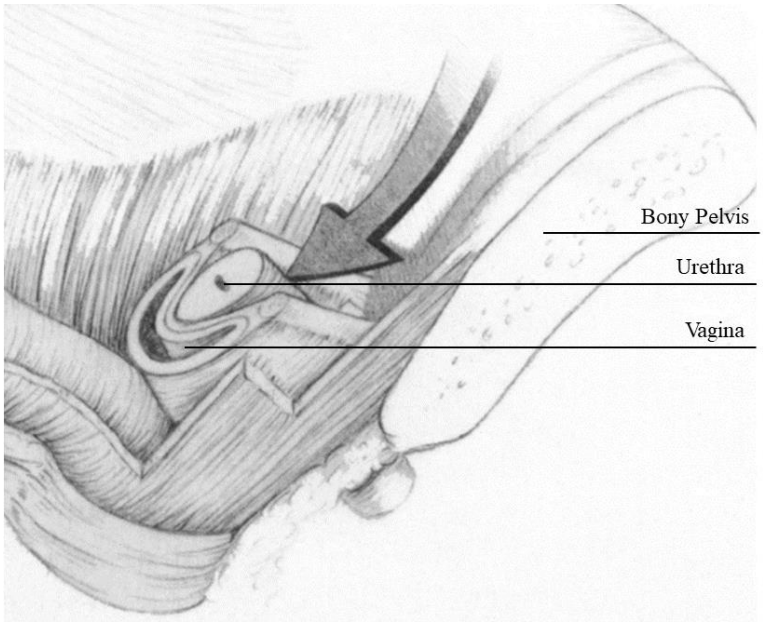


Figure 2.1: Illustration of the urethral support mechanism explored by the “hammock” hypothesis [7] *with added annotations*

If the supporting tissues are stiff, the urethral lumen (internal) diameter should be reduced (possibly to complete closure), which should result in a rise in the pressure required to move urine into the urethra, as explained by the Hagen-Poiseuille law [48]; thus, continence is maintained. DeLancey additionally theorized that the Levator Ani muscles, when contracted, add to the support of the urethra [7].

In Zacharin's investigation of female cadavers, he described a characteristic S-curve of the urethra, created by a series of supporting ligaments [8]. This series of ligaments and the close position of the urethra to the bony contours of the pubic bones was said to limit urethral mobility in continent females [8]. He theorized that the main strength of this support mechanism was the pubo-urethral ligaments, which appeared to contain a large amount of elastic tissue and which connected the proximal urethra to the pubic bone [8]. In contrast with DeLancey, Zacharin suggested that the Levator Ani do not play any direct role in urethral support [8], [49].

DeLancey, in turn, argued that Zacharin's theory of urethral support by fixation of the urethra to the pubic bone is implausible [7]. He highlighted that in a standard standing female, the urethrovesical junction, where the urethra meets the bladder, is well above the location in which the pubo-urethral ligament attaches to the pubic bone [7]. Thus, this proposed mechanism of support could not provide suspensory action to the urethrovesical junction while in the standing position [7]. Additionally, the urethra is not rigidly fixed within the pelvic girdle, but is mobile [7]. Thus, the more modern "hammock" hypothesis appears to be more likely to explain the urethral support mechanism of the female.

2.2.2 Urethral Closure Mechanism

Although pressure transmission to the urethral supporting structures plays a role in maintaining continence, the urethral closure mechanism also has a large role to play [3]. In DeLancey et al. [47], the relative importance of urethral support and urethral closure pressure to SUI was examined through a case-control study. Maximum urethral closure pressure (MUCP), after adjusting for BMI, was found to be the most predictive factor associated with SUI [47]. Indeed, MUCP explained 50% of SUI cases [47]. However, urethral closure pressure is the measured sum of all forces acting on the urethra to cause closure, thus the relative contribution of the different morphological components to urethral function remains unknown [3]. It is thought that urethral striated muscle and the vascular plexus are responsible for the majority of the urethral closure pressure, while the smooth muscle and surrounding connective tissue provide a lower, although still significant, contribution [3].

Anatomically, the urethra can be divided up into fifths. The first fifth of the urethra is most proximal, beginning at the bladder neck, and is surrounded by a ring of smooth muscle

[3], as shown in Figure 2.2. Moving caudally, the next two fifths are encapsulated by a multi-layered tube, whose outer layer is circular striated-muscle [3]. The circular striated-muscle surrounds a thin layer of circular smooth muscle, which surrounds a thicker layer of longitudinal smooth muscle and vascular plexus [3], [4]. In the fourth section, the compressor urethra and urethrovaginal sphincter pass over the urethra [3], illustrated in Figure 2.2. The fifth and most distal portion of the urethra makes up the meatus, which communicates externally [3].

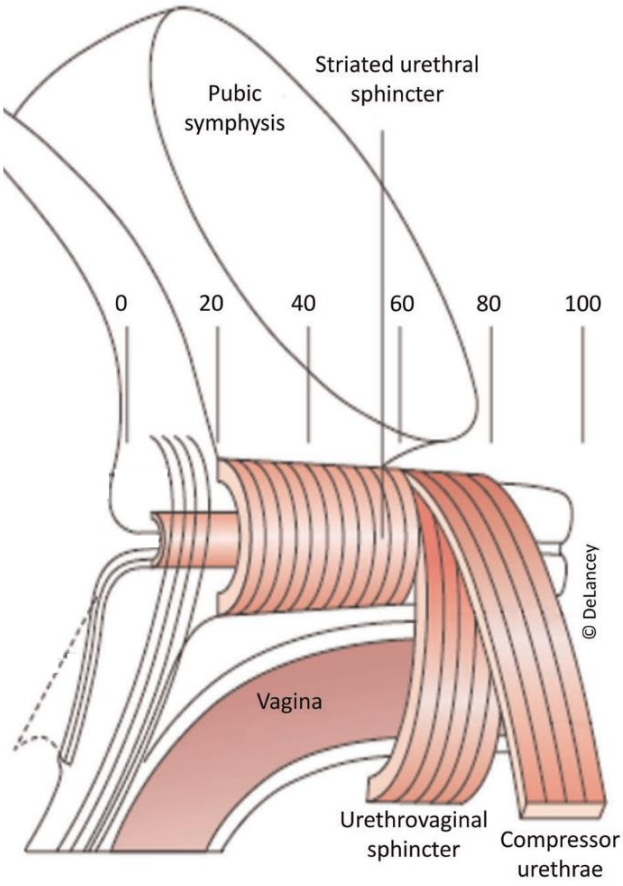


Figure 2.2: Illustration of the urethra in sagittal view [3]

The ability to extract features from the urethra and surrounding structures non-invasively could lead to a clearer understanding of the mechanisms of SUI and to a more individualized approach to the diagnosis and treatment of urinary incontinence. It could additionally lead to novel treatment techniques aimed at correcting deficiencies that were previously not considered to be associated with the disorder.

2.3 BLADDER NECK FUNNELING

2.3.1 Definition and Relevance

For the purposes of this thesis, BNF is defined as the opening of the proximal urethra when functionally inappropriate. BNF is thought, in this context, to represent failure of the proximal structure of the urethra. There is discourse in the literature about the relationship of BNF to SUI [29], [50]. Remember that, a recent meta-analysis suggests that BNF is more prevalent in females with SUI compared to continent females [29]. Others argue that BNF is not indicative of a functional issue because it is also observed in continent females [50], [51]. Despite this discourse, measurement or categorization of BNF is included in the studies of urinary incontinence pathophysiology, but seldom as a primary outcome [33], [34], [60]–[69], [52]–[59].

2.3.2 Imaging Approaches and Measures

There is a conspicuous gap in the literature in terms of what constitutes BNF. Several methods of measuring BNF have been reported, the most common of which is binary classification (i.e., present/absent) [34], [52], [63]–[67], [69], [53]–[55], [58]–[62]. Occasionally, the presence of BNF is further classified as “extensive”; however, exact definitions for “extensive” funneling are not provided [68]. Varied definitions have been reported to describe the presence of BNF [52]–[55], [58], [59], [62], [64]–[66]. These definitions can be grouped into 5 categories, as shown in Table 2.1, which are similar in spirit (i.e., all reflecting continuity between the bladder and urethra) but differ in the details of their precise criteria.

In a recent systematic review [29], a meta-analysis of studies in which BNF was used as an outcome to compare women with and without SUI suggested that BNF was strongly associated with SUI [29]. However, the authors noted wide variation in the definition of BNF [29]; a standardized definition of BNF remains elusive. Of note, however, is that among studies that included only women with SUI, fewer than half of women with SUI demonstrate observable BNF [58], [59], [64], [65]. These findings highlight a clear gap in the literature around what constitutes BNF, how to measure it, and its potential role in SUI pathology.

The variability in the methods used to evaluate BNF may, at least in part, explain seemingly contradictory findings. To better understand the pathophysiology of SUI,

researchers must be able to reliably measure features that are relevant to the pathophysiology. Therefore, selecting a standardized and intuitive definition of BNF is paramount if we are to understand its role in SUI.

Table 2.1:Frequency of definitions of BNF provided in the included literature

Definition	Number of Papers	Imaging Method		Imaging Plane	
The widening / opening / separating of the bladder neck/proximal urethra/internal meatus [53]–[55], [58], [60]–[62]	7	Magnetic Resonance Imaging	4	Sagittal, Coronal, and Transverse	3
				Sagittal and Transverse	1
		Ultrasound Cystourethrography	1	Midsagittal	1
		Ultrasound Imaging	2	Midsagittal	2
No definition given [34], [63], [67]–[69]	5	Magnetic Resonance Imaging	1	Sagittal, Coronal and Axial	1
		Ultrasound Cystourethrography	1	Midsagittal	1
		Ultrasound Imaging	3	Midsagittal	3
Opening the proximal urethra as two diverging lines/edges [59], [65]	2	Ultrasound Imaging	1	Midsagittal	1
		Video Urodynamics	1	Coronal	1
Appearance of radiographic contrast material in the proximal urethra in the absence of a detrusor contraction or an increase in intra-abdominal pressure [66]	1	Video Urodynamics	1	Sagittal and Oblique	1
The funneling of the proximal urethra and opening of the entire urethra during leaking [52]	1	Ultrasound Imaging	1	Midsagittal	1
Urethral internal orifice was opened during a bearing down manoeuvre [64]	1	Ultrasound Imaging	1	Midsagittal	1

In addition to binary classification, BNF has been reported as a continuous variable by some research groups, measured by its length (anterior curve length, midline curve length, or midline straight-line length [33], [56], [57], [64]), and/or its width [56], [57], [64]. Studies in which aspects of BNF morphology have been measured are described below.

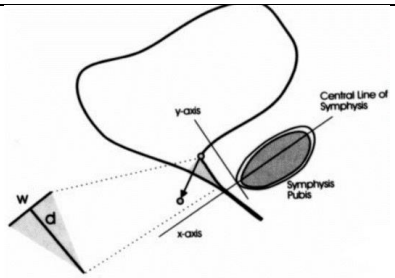
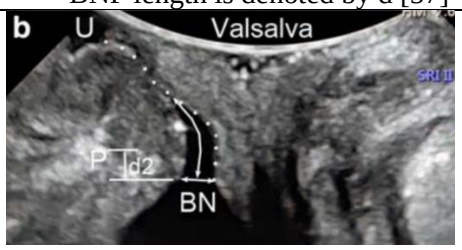
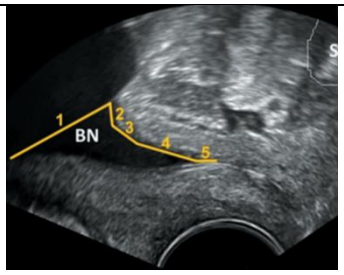
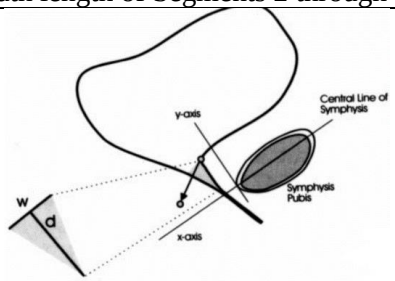
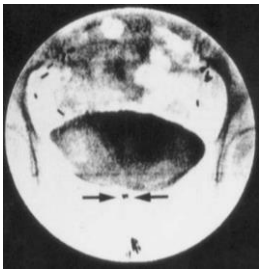
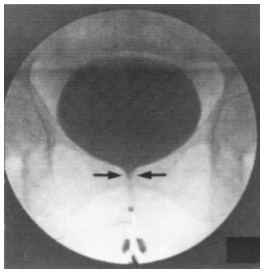
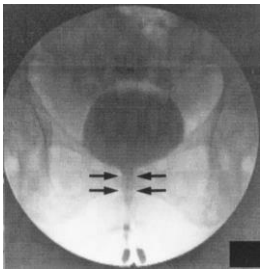
Using ultrasound imaging in the midsagittal plane, Wlaźlak et al. retrospectively analysed images of the proximal urethra captured from women presenting with and without symptoms of SUI (including a positive cough test). They found that 100% of participants presenting with symptoms of SUI also demonstrated BNF while only 16.3% of participants presenting without SUI demonstrated BNF [56]. Wlaźlak et al. additionally measured the anterior curve length of the BNF and reported it to be significantly larger in participants with SUI compared to controls [56]. They reported no concurrent significant difference between groups in terms of BNF width; BNF width was not explicitly defined, but it can be inferred to be the distance between the anterior and posterior aspects of the urethrovestical junction [56]. Both Zhang et al. [33] and Schaer et al. [57] also reported on the length of the BNF. While neither group explicitly defined their measurement, the diagrams presented in their publications illustrate BNF length as the length of a straight line drawn from the midpoint of the open bladder neck and the midpoint of the apex of the BNF within the urethra [33], shown in Table 2.2 as BNF midline straight-line length. Consistent with Wlaźlak et al., Zhang et al. concluded that BNF midline straight-line length was significantly larger, during both rest and straining, in females with SUI compared to those without SUI [33]. Schaer et al. also found that BNF midline straight-line length was greater, during both strain and cough, in females with SUI compared to those without, but this observation did not reach statistical significance [57]. Contrary to Wlaźlak et al., Schaer et al. also reported that BNF width, drawn as the distance between the anterior and posterior urethrovesical junction, was larger in females with SUI than continent controls [57].

Wen et al. retrospectively evaluated ultrasound images from female patients with SUI [64]. BNF was quantified by both midline curve length and width [64]. Midline curve length was defined explicitly as “the length of the opened urethra”; a curve was drawn which passed through the midpoints between the anterior and posterior aspects of the urethra from the most cranial aspect to the most caudal aspect of the visible funnel [64]. Width was explicitly defined as the anteroposterior diameter of the orifice [64], presumably the urethrovesical orifice. They reported that funneling midline curve length and width did not change significantly between supine and standing positions [64].

The final method of characterizing BNF is through its shape as observed in the frontal plane on videourodynamics [70]. A retrospective analysis by Ghoniem et al. included

100 consecutive female patients who had both a clinical and a videofluorourodynamic (VFUD) diagnosis of SUI with intrinsic sphincter deficiency (ISD) defined as proximal urethral closure pressure <10cmH2O [70]. BNF shape was categorized as: subtle/urodynamic (ISD-A), beak-shaped (ISD-B), and pipe stem (ISD-C) [70]. ISD-A was described as an open bladder neck at stress, ISD-B was characterized by a beak-shaped bladder neck regardless of its position within the pelvis and ISD-C was characterized by an open, fixed urethra with a high positioned bladder neck [70]. Ghoniem et al. found that 32% of females with SUI presented with ISD-A, 45% presented with ISD-B, and 14% presented with ISD-C [70]. While all women in the study had very low proximal urethral closure pressure (< 10 cmH_2O), proximal urethral closure pressure was significantly lower in participants with ISD-B or ISD-C compared to those with ISD-A [70]. The authors concluded that these qualitative differences among characteristic shapes indicated different pathologies that required different treatment approaches [70].

Table 2.2: Bladder neck funneling measurement techniques reported in the literature.

Classification Type	Picture
Midline Straight Length	 BNF length is denoted by d [57]
Midline Curve Length	 [64]
Anterior Side Curve Length	 The path length of Segments 2 through 5 [56]
Width	 BNF width is denoted by w [57]
Characteristic Shape	 ISD-A  ISD-B  ISD-C [70]

2.3.3 What Does BNF Represent in Terms of Pathophysiology

This plethora of definitions for BNF suggests that this characteristic has become separated from the meaning of bladder neck closure failure. BNF length could be an attempt to identify which aspects of the urethral closure mechanism are likely deficient, yet such identification is rarely discussed in the literature [33], [56], [64]. Indeed, variations in BNF length definition (anterior curve length vs. midline curve length vs. midline straight-line length) do not seem to point to any anatomical feature or physiological relationship [33], [56], [64]. Schaer et al. attempted to describe the relationship between their measurement of BNF and urethral morphology, yet did little more than suggest that "[BNF] without incontinence is an abnormal finding" [57].

On the other hand, the differentiation between short and long funneling may point to urethral sphincter function, as discussed by Wlaźlak et al. [56]. Yet when comparing this to the description of the urethral closure mechanism described by Pipitone et al., there is no distinct morphological structure at the 50% point of the urethra [3], the threshold for long funneling [56]. Therefore, this definition seems to lack specificity in regards to its relationship with urethral morphology. However, the pelvic floor and Levator Ani muscles do act on the urethra at approximately its midpoint [71], [72], and thus properly functioning extrinsic compression may be relevant to funneling length.

Considering the measurement of BNF width, it seems intuitive that this could be interpreted as a measure related to laxity in the trigonal ring in the most proximal region of the urethra. Yet again, the actual mechanism generating the opening of the proximal urethra is not discussed in the investigations involving measurement of BNF width [56], [57], [64]. Additionally, if care is not taken when measuring this feature, it is easy to measure the width of a funnel-like shape occurring at the trigone instead of funneling at the proximal urethra itself. In that case, the measurement would not only be erroneous but also seek to describe a relationship to an entirely different morphological mechanism than that of urethral closure.

Lastly, while again not discussed by Ghoniem et al. [70], the shape of the funnel may be related to different pathophysiological processes. It could be postulated that ISD-A is related to trigonal ring and circular smooth muscle failure at the proximal urethra, ISD-B is associated with ISD-A as well as failure of the smooth and striated sphincters, while ISD-C

encompasses ISD-B as well as failure of the extrinsic (compressor urethra and urethrovaginal sphincter) closure.

2.3.3 Ultrasound Imaging Approach for the Visualization of BNF

The literature suggests that the observation of BNF is dependent on bladder volume and, although inconsistent, patient position. It has been suggested that in order to optimize visualization of BNF on USI, participants should be standing with a full bladder [2], [64], [73], [74].

In a study of 109 female participants with SUI (urine leakage observed during a cough or bearing down), bladder volume was found to significantly affect transperineal USI parameters, including BNF [73]. BNF was more frequently and easily observed when participants were measured with a self-defined full bladder (61.5% with BNF) compared to an emptier bladder (50 mL) (53.2% with BNF) [73]. Another examination of females with SUI (n=30) confirmed that funneling is more frequently observed when participants have larger bladder volumes (33.3% of participants demonstrated BNF at 300 mL) versus smaller bladder volumes (10% of participants demonstrated BNF at 100 mL) [2].

This same study also suggested that BNF was more observable when females (n=80) were asked to bear down while in the standing position (27.5% with BNF) compared to the supine position (3.8% with BNF) [2]. Another study of 132 female participants presenting with lower urinary tract symptoms examined BNF using transperineal USI while women were bearing down and also concluded that BNF was more observable in standing (71.2% with BNF) compared to supine (56.8% with BNF) [74]. However, another study of transperineal ultrasound images acquired from 81 females with SUI concluded that funneling midline curve length and width did not change significantly between standing and supine positions [64].

In terms of task, many investigations have examined BNF on Valsalva or straining [33], [34], [52], [53], [56], [57], [64], [70], but an analysis of task's effect on the observability of BNF was not found. Effort on Valsalva or during straining can not be standardized [15] and this may bias findings – women who are more inclined to leak urine may not perform the task as strongly as women with no history of SUI. Although the literature suggests that position and bladder volume affect the observation of BNF, many investigations that have examined the relationship between SUI and BNF using USI do not

follow these suggestions [33], [55], [68]. Variability in patient position and bladder volume may indeed explain some of the conflicting findings in the literature.

2.3.4 Reliability Data on BNF Classification and Estimation

Of the published methods for the classification and/or quantification of BNF, only binary, width, and midline curve length have published data regarding their repeatability [53], [55], [64], [65]. Indeed, the classification of BNF as present or absent appears to have excellent repeatability. Cassadó et al., found agreement on the classification of BNF using USI, categorized using the binary method, was 100% both between and within raters [53] when evaluated in the supine position during Valsalva or cough with a bladder volume of 75 to 275 mL. Pontbriand-Drolet et al. reported on the intra-rater reliability of BNF width and length (no measurement parameters described) on MRI both at rest and on straining using the Intraclass Correlation Coefficient (ICC) [55]. Both BNF length [rest: ICC = 0.98 (SEM: 0.46) , on straining: ICC =0.76 (SEM: 0.56)] and BNF width [rest: ICC = 0.87 (SEM: 0.32), on straining: ICC = 0.92 (SEM: 0.76)] were reported to have good to excellent reliability [55]. Another investigation found 100% agreement between raters when classifying BNF from USI using the binary method while participants were in supine with a bladder volume of 300 mL at both rest and during Valsalva [65]. Wen et al. performed a test-retest series on 20 patients with two different raters and reported excellent reliability [ICC = 0.95 (95% CI: 0.83–0.98] for all BNF measurements (binary, width, midline curve length) measured from ultrasound images [64]. Although these results suggest that BNF can be reliably classified, the vague definitions/terms surrounding many of the classification measures contributes to the uncertainty surrounding the implications of BNF in terms of SUI.

Reflecting on the various BNF definitions and their association to lower urinary tract anatomy, none of the uncovered definitions seem to examine the relationship between the trigone, proximal urethra, and SUI. Indeed, BNF definitions seem to relate most closely to morphological aspects associated with urethral closure. A feature thought to be associated with the urethral support mechanism, such as PUVA, should also be included when analyzing the lower urinary tract in relation to SUI.

2.4 POSTERIOR URETHROVESICAL ANGLE

2.4.1 Definitions and Relevance

For the purposes of this thesis, posterior urethrovesical angle (PUVA) is defined as the angle formed between the posterior aspect of the urethra and the inferior-posterior bladder wall, as shown in Figure 2.3. Remember that the literature is in agreement that females with SUI have larger PUVAs at rest and on straining than continent females [29], [31]–[37].

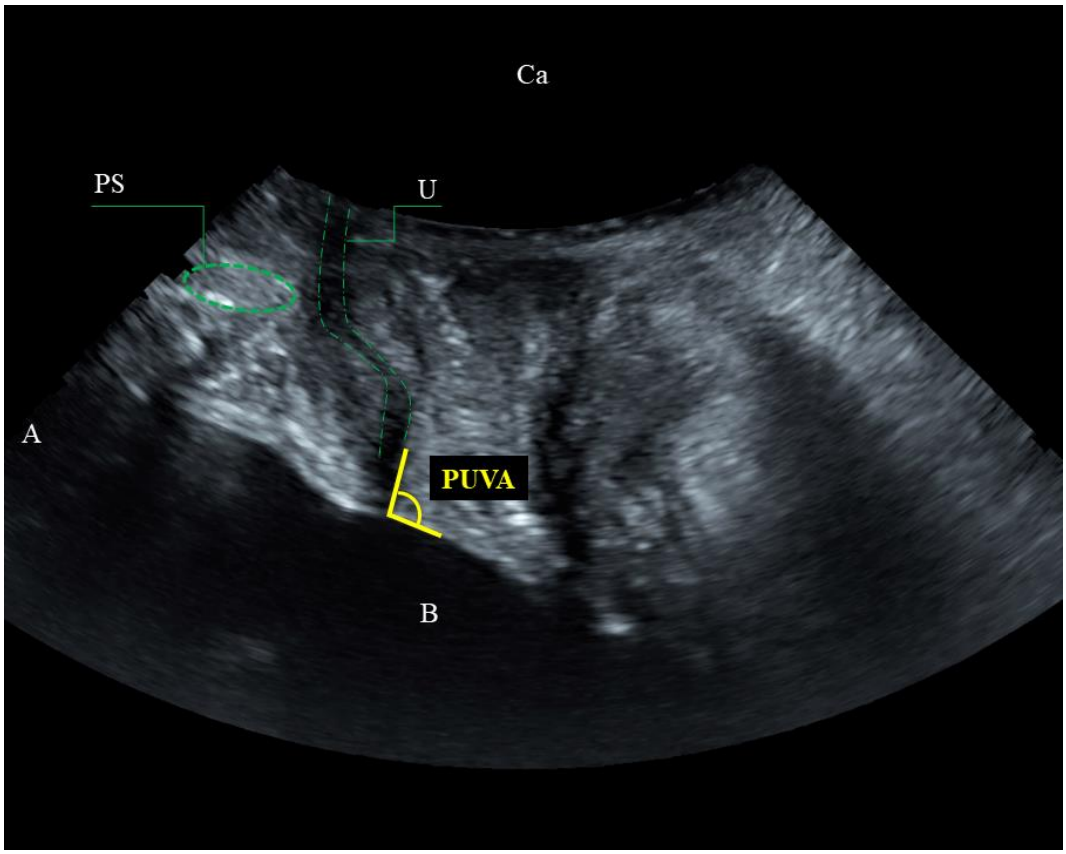


Figure 2.3: Example of PUVA as defined in this thesis. A: Anterior, B: Bladder, Ca: Caudal, PS: Pubic Symphysis, U: Urethra

2.4.2 Imaging Approaches and Measures

There are several definitions of PUVA in the literature that are similar in nature but distinct in terms of the methods of measurement [31], [32], [35], [55], [75]. These definitions are displayed in Table 2.3. The majority of investigations in the literature fail to include a definition of PUVA [33], [34], [36], [37], [64], [76]. However, because existing definitions are so similar in nature, this may not be a cause for concern.

Table 2.3: Definitions of posterior urethrovesical angle

First Author	Year	Imaging Parameters	Definition of Posterior Urethrovesical Angle
Al-Saadi, Wasan Ismail [35]	2016	Transperineal Ultrasound, Midsagittal plane	• "[T]he angle between the proximal urethra and the posterior vesical wall."
Fielding, J. [32]	1998	Magnetic Resonance Imaging, Midsagittal plane	• "The posterior urethrovesical angle was measured in the same fashion as that measured for voiding cystourethrography."
Kim, Jeong Kon [75]	2003	Magnetic Resonance Imaging, Sagittal plane parallel to the long axis of the urethra	• "Two lines were drawn, one through the long axis of the urethra and one parallel to the bladder base. The intersection of these lines determined the vesicourethral angle."
Pontbriand-Drolet, Stéphanie [55]	2016	Magnetic Resonance Imaging, Sagittal plane centred at the pubic symphysis	• "Measured at the intersection of the lines drawn along the urethral lumen and bladder floor."
Sendag, Fatih [31]	2003	Transperineal Ultrasound, Midsagittal plane	• "The angle between the urethral axis and the floor of the bladder axis one-third closer to the urethra."

2.4.3 Ultrasound Imaging Approach for the Visualization of PUVA

Literature surrounding USI parameters that affect PUVA suggests that at rest, PUVA is significantly larger in standing than in supine [2], [74], presumably due to the gravitational loading of the pelvic organs. However, on a straining task, the literature is inconclusive about the impact of body position on PUVA, with one paper concluding PUVA is significantly larger in standing than in supine [2] and another concluding that posture has no significant effect on PUVA measured during straining [74]. This discrepancy could be due to the non-standardized nature of the straining task.

The impact of bladder volume on PUVA is also unknown. Dietz et al. observed in participants that were positioned in supine at rest that, PUVA was significantly larger when the bladder was full than when it was empty [73]. Bladder volume had no impact on PUVA when participants were observed in supine during straining [73]. The findings of Schaer et al. were consistent with those of Dietz et al. for the straining task but not at rest [2]. While Dietz et al. investigated a larger sample than Shaer et al. (n = 109 vs. n = 30) [2], [73], Dietz et al. evaluated PUVA at only two bladder volumes [50 mL and maximum capacity (mean

355 mL)] [73], whereas Shaer et al. evaluated PUVA, among other measures, at 100 mL increments of bladder filling up to 500 mL and found that PUVA was not impacted by bladder volume at rest [2]. More research is required to determine whether positioning and bladder volume affect PUVA in females.

2.4.4 Reliability of PUVA Estimation

Similar to measurements of BNF, few investigations have reported on the reliability of their measurement approach for PUVA [55], [64]. In Wen et al. [64], inter-rater reliability of PUVA on USI in with unspecified position or bladder filling (n = 20) was excellent with a [ICC = 0.91 (95% CI: 0.83-0.97)] [64] while in Pontbriand-Drolet et al., intra-rater reliability for PUVA measured on MRI in supine (n = 66) was good both at rest [ICC = 0.81 (SEM: 8.48)], and on straining [ICC = 0.87 (SEM: 6.54)] [55]. Although these results are promising, there is limited evidence for intra- and inter-rater reliability in evaluating PUVA. The reliability of PUVA evaluated using USI should be investigated further.

2.5 AUTOMATIC AND SEMI-AUTOMATIC SEGMENTATION OF FLUID-FILLED BODIES ON ULTRASOUND IMAGING

To our knowledge, there are no existing methods described for the automated or semi-automated segmentation of the proximal urethra, and no methods have been described for the feature extraction of BNF or PUVA from USI of the female pelvis. To inform our approach to this novel problem, the literature concerning automated and semi-automated segmentation of images of other fluid-filled bodies (the bladder and gallbladder) acquired using USI was explored.

2.5.1 Automatic Segmentation of the Bladder and Bladder Walls

Three methods of automatic bladder segmentation were found in the literature: two concerned measurements of the bladder wall thickness [77], [78], and one aimed to measure bladder diameter in order to estimate bladder volume [79].

Akkus et al. evaluated 80 B-mode ultrasound (Alpinion E-CUBE 12R, curved-linear Alpinion C1-6 probe with centre frequency: 2.5 MHz) images acquired from 11 healthy volunteers [77]. No pre-processing was applied to the image set [77]. The bladder was segmented using two approaches: the first utilized dynamic programming, and the second implemented an encoder-decoder type fully convolutional neural network (CNN). The CNN

used the rectified linear unit activation function and U-Net architecture. A Glorot uniform distribution was used to initialize the weights of the model, and a softmax function was used to produce a 2-class output. Five-fold cross-validation was used to reduce the impact of the limited size of the training set [77]. Manual segmentation of the bladder was performed by 3 independent raters and inter-rater reliability was evaluated using the Dice similarity index. The performance of both segmentation algorithms was evaluated against manual segmentation, and the segmentations were also compared to each other using the Dice similarity index. Mean and median Dice indices for all comparisons between segmentation methods and raters were similar in magnitude, and all were above 0.9 [77]. Akkus et al. concluded that although both segmentation methods were successful, the CNN segmentation method was superior because it represented the shortest computation time [77].

In Padmapriya et al. [78], an edge-based segmentation technique was used to detect and segment the bladder wall. An unspecified number of ultrasound B-mode images (256x256x8 greyscale) were captured from an unspecified number of participants [78]. A 7x7 mean filter was used to smooth the input images and to attempt to remove noise while preserving detail. Automatic adaptive thresholding was then used to enhance contrast through binarization. The bladder wall was segmented using the Sobel edge detection algorithm, and no post-processing was applied. Padmapriya et al. concluded that the segmentation method was highly reliable, but they did not present quantitative measures or methods to support this claim [78].

Matsumoto et al. developed a method to automatically detect bladder diameter [79]. They used an unspecified number of B-mode ultrasound images captured from an unspecified number of participants using the Sonosite iViz system. Firstly, all images were resized to 512x384 pixels. Additionally, brightness change, contrast change, and horizontal flip were applied randomly to the image set. No preprocessing was applied to the image set. A CNN was utilized for bladder segmentation, a VGG16-like CNN was used as the encoder and the decoder was constructed from FCN-8s. The learning rate was initialized to $1e^{-4}$ and was reduced by 10% every 50 epochs. The training set was composed of 81 pairs of transverse and longitudinal ultrasound images. After training, a set of 23 image pairs from healthy volunteers with a voided volume of <150 mL were collected for the validation set using the same sonographer and imaging system [79]. A set of 54 image pairs from 27

participants was used to evaluate the automatic segmentation method [79]. Evaluation of manual bladder volume estimation resulted in an interclass correlation coefficient of 0.99. Matsumoto et al. concluded their method performed comparably to manual segmentation [79].

All of the three methods found were fully automated and two utilized machine learning as their segmentation algorithm [77]–[79]. The choice to use a CNN versus another, simpler approach to segment the bladder is perhaps superfluous. Segmentation of the bladder should be a simple task, whether approaching the problem from an automated or semi-automated lens because the bladder is a hypoechoic body surrounded by a relatively thick hyperechoic wall. Applying such an algorithm to dynamically segment the bladder during acquisition could be supported, as this method has the power to provide fast and accurate segmentation when provided with unreliable landmarks and excessive noise. However, neither method was attempting this more difficult challenge [77], [79]. Indeed, Padmapriya et al. presented the only straightforward algorithm for segmenting the bladder but did not provide objective evidence to support the claim that the method is highly reliable [78]. For this reason, none of the segmentation algorithms presented provide strong arguments for augmentation and application to the current problem. However, pre- and post-processing methods will be considered when developing the method presented in this thesis. Based on the limited literature for segmentation of the bladder, it was deemed imperative to look elsewhere in the biological image segmentation literature to determine viable methods of segmentation on hypoechoic bodies on ultrasound.

2.5.2 Automatic and Semi-Automatic Segmentation of the Gallbladder

Because the gallbladder presents a similar segmentation problem, in that it is also a hypoechoic body surrounded by reasonably hyperechoic tissue, literature was reviewed in the hope of informing our approach. Four methods of automatic or semi-automatic segmentation of the gallbladder were found in the literature [80]–[83]

In a recent publication, Muneeswaran and Rajasekaran developed an automatic segmentation method inspired by spider-web construction for the segmentation of the gallbladder on ultrasound images [80]. An unspecified number of ultrasound images from 8 participants were used to develop their method [80]. Histogram equalization and a difference-of-offset Gaussian filter were applied to improve image quality. The bioinspired

spider-web algorithm was then applied to segment the desired structures. Firstly, the ‘spider’ was placed at a random pixel in the image, and then the algorithm selected pixels in the ‘spiders’ eight neighborhood with equal intensity to the spider pixel. When a neighboring pixel is iso-intense with the ‘spider’ pixel, the ‘spider’ moved from its original position to the specified neighbor. If a neighboring pixel was of different intensity, then the ‘spider’ returned to its original position. When the ‘spider’ moved between two pixels, these pixels were said to be silked. The segmented region is represented by the set of silked pixels. An unspecified number of ultrasound images from 60 patients were used to evaluate the algorithm [80]. This image set was also examined using other state-of-the-art methods, including a snake balloon, a parameter-adaptive pulse-coupled neural network, a modified edge-based model and a modified morphological model in order to comparatively evaluate the efficacy of the spider-web based model [80]. The image set was additionally manually segmented for comparative analysis [80]. Algorithms were evaluated using the Dice similarity index, overlap function, and overlap value which evaluated the similarity between the automatic segmentation and manual segmentation as well as position error which evaluated the error between these two regions [80]. The spider-based algorithm outperformed all other methods on all metrics [80].

In Ciecholewski’s investigation into the development of an active contour method for the automatic segmentation of the gallbladder, a set of 220 images was captured from an unspecified number of participants, including cases without lesions and with lesions of varying types [83]. In this method, no preprocessing was applied to the image set [83]. A motion equation active contour (AC) model was used to segment the gallbladder from the background [83]. Area error rate (AER) was used to evaluate the difference between the automatic segmentation and manual segmentation [83]. AER was lowest in gallbladders without lesions and highest in gallbladders with folds or turns [83]. The author concluded the method was reasonably precise [83].

Ciecholewski built upon his earlier results using a set of 220 images which were captured from an unspecified number of participants, including cases without lesions and with lesions of varying types [81]. First, the image histograms were normalized to improve image contrast [81]. Then a motion equation AC model, centre-point AC model, and balloon AC model were applied to the images, and the results were summed to determine the

segmentation of the gallbladder [81]. AER was used to evaluate the difference between the automatic segmentation and manual segmentation [81]. AER was lowest in gallbladders without lesions and highest in gallbladders with polyps [81]. The author concluded that the algorithm yielded reasonably precise results but further development was needed to reduce the AER in gallbladders containing lesions [81]. It should also be noted that the AER of this method was lower than the AER of Ciecholewski's first method [81], [83].

In Ciecholewski and Chochołowicz's method to segment the gallbladder, a set of 800 grey-level ultrasound images was generated from an unspecified number of participants using a GE Healthcare Logiq C3 USI system [82]. Captured images had a size of 512 x 512 pixels, with 8-bit resolution [82]. Histogram normalization and a Gaussian filter were applied to enhance contrast and smooth the image, respectively [82]. Edge-based and region-based AC models were applied to segment the gallbladder from the background, and lastly, the convex hull algorithm was applied to the contours to eliminate concavities [82]. If necessary, missing fragments of the gallbladder edge could be manually approximated [82]. Manual segmentation was performed by two radiologists [82]. The method was evaluated using the Dice similarity index, the overlap fraction, the overlap value, the extra fraction and position error [82]. On all measures, both semi-automatic segmentation methods (edge-based and region-based) performed better than fully automatic segmentation methods, with no approximation of missing fragments of the gallbladder edge [82]. There was no significant difference in the average areas determined by the active contour methods and those segmented manually [82]. Edge-based models performed slightly better than region-based models, although the difference in performance was minor [82]. The authors concluded that the edge-based method significantly improved upon previous methods [82].

Ciecholewski's extensive development of methods to segment the gallbladder suggests that an active contour method could provide good results on other hypoechoic bodies surrounded by hyperechoic tissues. Additionally, unlike CNN methods, active contour methods require low computational power and no training, thus allowing for application on a smaller dataset without immediate concerns of overfitting. Conversely, Muneeswaran and Rajasekaran present a method that is less established than active contour methods but is still simple compared to CNNs. Both methods are strong candidates for application to the identification of BNF and estimation of PUVA.

2.6 CONCLUSION

There is a need to better understand the pathophysiology of SUI [39] and to use this knowledge to develop a more personalized approach to the management of SUI. We know that both BNF and PUVA are strongly associated with SUI [29], yet their implications are poorly understood. Before we can begin to understand the links among BNF, PUVA, and SUI, standardized definitions and valid and reliable methods of measurement are needed. However, manual analysis is time-consuming and expensive. A fully automated approach is the ideal solution, but can be onerous and complex to implement. In particular, automated approaches, by their very nature, are more rigid thus they respond poorly to unforeseen image variations or artifacts. A semi-automated approach can be developed relatively efficiently and if this preliminary approach demonstrates good potential, additional future efforts towards a full-automated approach can be pursued. This semi-automated method will borrow from bladder and gallbladder segmentation literature with the intention of creating a simple method to analyze BNF and PUVA.

Chapter 3 METHODS

3.1 IMAGE SET

A set of 227 B-mode US videos captured from 39 female participants was collated from a concurrent research study (REB #: H-06-18-759) examining urogenital morphology and its relationship to urinary incontinence in female runners. Two-dimensional, B-mode US videos were captured transperineally in the midsagittal plane using a GE Voluson S6 USi system coupled with a RAB6-RS 3D curvilinear probe (2-5 MHz, GE). Participant inclusion and exclusion criteria are detailed below in Table 3.1.

Table 3.1: Inclusion and exclusion criteria for experimental group (REB #: H-06-18-759)

Inclusion criteria	Exclusion criteria
• Cis-female runners older than 18 years old • Run at least 10 km/week for at least one year • For women who leak urine during running: 4 episodes of leakage or more per month (or 1 episode of leakage or more per week)	• Risk factors for physical activity according to Par-Q results • History of urogenital surgery • Neurological impairments involving the central nervous system or the sacral nerves • Known connective tissue disorders • Pelvic organ prolapse that moves caudally past the hymen on Valsalva maneuver standing up • Pregnant • Have delivered a baby during the last year • Dyspareunia or pelvic pain • Symptoms of female athlete triad as reported by the LEAF questionnaire • In situ devices that would not be suitable for ultrasound imaging, intra-vaginal dynamometry and intra-vaginal pressure sensor

Participants were imaged in the standing position after their estimated bladder volume was between 150-300 mL as determined using trans-abdominal ultrasound and the volume function on the ultrasound machine. Participants stood with their feet shoulder-width apart and used an examination table for stability to limit postural sway. Among other tasks beyond the scope of this study, US videos were captured during quiet standing.

Because effort during Valsalva or straining cannot be standardized [15], frames acquired during quiet standing were selected for analysis.

Three image clips were acquired before and after a standardized running protocol as the purpose of the original study was to evaluate changes in pelvic morphology and function induced by running. The running protocol had three stages. In the first stage, participants ran for two minutes at each of the following speeds: 7 km/hr, 10 km/hr, 15 km/hr. The second stage was a 30-minute run at a self-selected speed, > 7 km/hr. The third stage involved running for 30 seconds at the following speeds: 10 km/hr and 7km/hr. The view angle was maximized to either 85 or 90 degrees for all US acquisitions, while gain and depth were optimized for each individual in order to visualize the pubic symphysis, the anorectal angle, and the urethra within the image frame throughout all the study tasks. The data were not acquired for the purposes of optimizing the image quality at the bladder neck. As such, a subset of data, consisting of 74 frames acquired from 30 participants, was created based on the following inclusion criteria:

1. Clear visualization of the posterior bladder wall
2. Clear visualization of the hypoechoic representation of the urethra
3. Clear visualization of the pubic symphysis and anal rectal angle
4. Minimal motion blur

Frames were extracted from the US videos using ImageJ (Version 1.8.0). Extracted frames were 1136x852 pixels (approximately 135 mm x 101 mm) in size and saved locally to construct the final dataset. Frames larger than this size were cropped and visually inspected to ensure that none of the landmarks were removed.

3.2 MANUAL ANALYSIS OF BNF AND PUVA

Manual classification of BNF and manual estimation of PUVA were each performed twice by two independent raters who received two two-hour group training sessions with a urologist. BNF was defined as present when the proximal urethra was observed to be open as indicated by [58], [60]. BNF was classified using visual inspection in ImageJ. BNF classification was paired with a binary confidence measure, where raters declared that they were either sure of their classification or unsure of their classification. The binary confidence measure was used to inform the validation of the semi-automatic analysis

software against manual classification, as classification disagreements between manual and semi-automated classification were deemed to be more acceptable when the rater was uncertain about their manual classification. This approach was used to provide insight into differences between the raters and to evaluate agreement within the context of the validity of manual segmentation.

PUVA was defined as the angle formed between the most posterior aspect of the hypoechoic representation of the urethra, and the inferior-posterior bladder wall (Figure 2.3). Raters determined the PUVA using the ImageJ angle tool. This tool allows the rater to select three points, then to calculate the angle between them (Figure 3.1). The rater is additionally aided by the visualization of the angle made by 2 straight lines allowing them to fine-tune their point selection to ensure the best estimation.

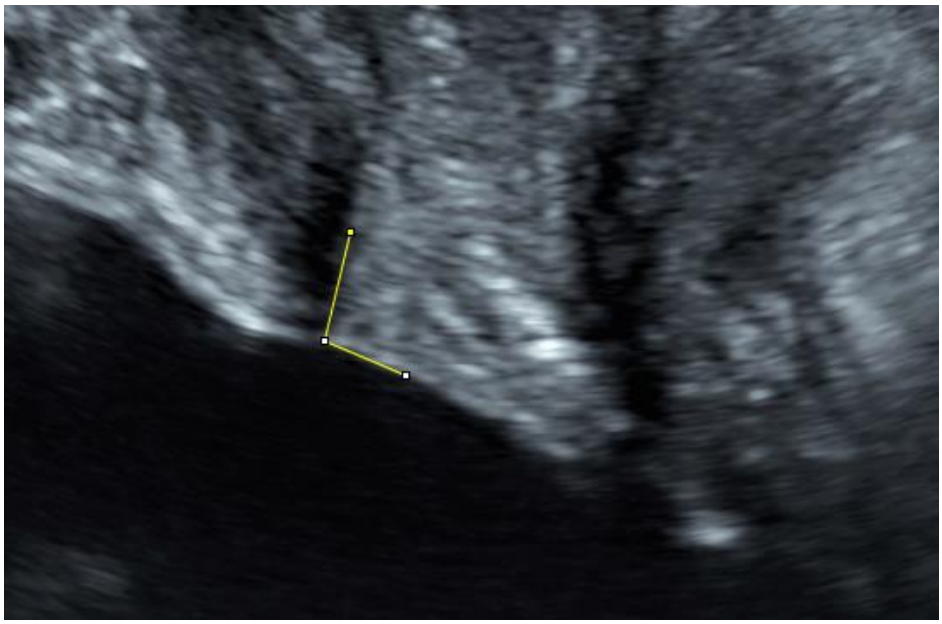


Figure 3.1: ImageJ angle tool

All statistics were calculated using SPSS version 28 (SPP Inc, Chicago, IL). Intra- and inter-rater agreement for BNF (binary: present or absent) was evaluated using Cohen's Kappa statistic and Light's Kappa statistic, respectively. The inter-rater agreement of BNF classification was assessed over 2 datasets. The first was the original dataset and the second was the subset of data over which both raters were confident in their classification of BNF on both trials, called the 'Confidently Classified BNF Dataset' (n=17). Intra-rater reliability for PUVA was evaluated using intra-class correlation (ICC) coefficients and their 95% confidence intervals based on single-rating, absolute agreement using a 2-way mixed-effects

analysis of variance (ANOVA) model [84]. Inter-rater reliability for PUVA was evaluated using ICC coefficients and their 95% confidence intervals based on single-rating, absolute agreement, using a 2-way random-effects ANOVA model [84]. A Bland-Altman analysis was also conducted to analyze the agreement in PUVA estimation between Rater 1 and Rater 2.

3.3 SEMI-AUTOMATED SEGMENTATION

A semi-automated method was developed to classify BNF and estimate PUVA using MATLAB 2020a. Segmentation of bladder and proximal urethra was performed through three main stages (Figure 3.2): 1) pre-processing, to remove or reduce image noise and artifacts, 2) segmentation of the regions of interest, and 3) post-processing, to refine the segmentation. After segmentation, features were extracted to classify BNF and estimate PUVA.

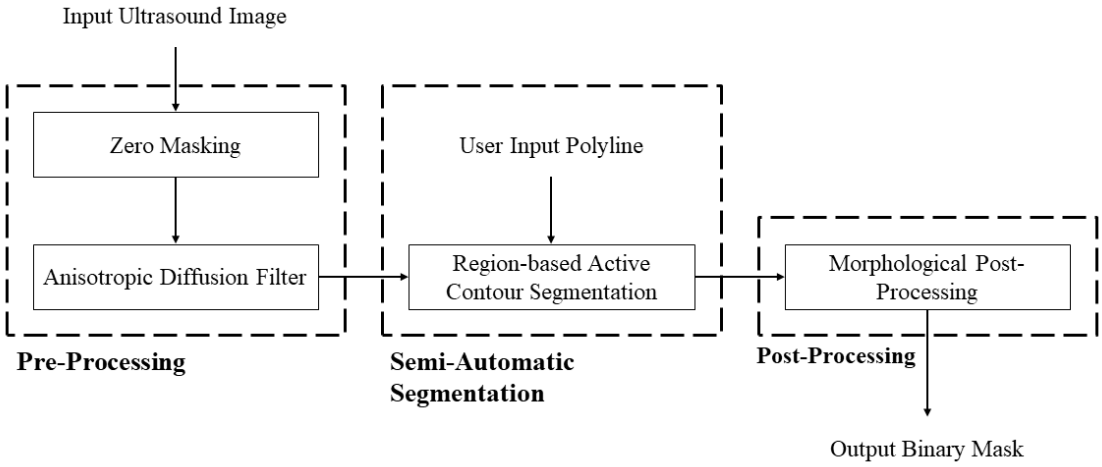


Figure 3.2: Semi-Automated segmentation of bladder and proximal urethra flowchart

3.3.1 Pre-Processing

First, a mask was applied to remove text and other irrelevant features surrounding the ultrasound image. Then, the image was converted to grayscale using a weighted-sum method [85]. Anisotropic diffusion filtering [86] was used to reduce the effect of speckle noise in the ultrasound image. Anisotropic diffusion filtering was chosen over other smoothing filters because it successfully reduced noise while maintaining the integrity of the edges based on qualitative assessment on a subset of images. Anisotropic diffusion filter

parameters (number of iterations and gradient threshold) were estimated for each individual image in the data set using the MATLAB function ‘imdiffuseest’. Estimated parameters were applied only on the image for which they were estimated on.

3.3.2 Segmentation Algorithm

A region-based active contour method (Chan-Vese Algorithm [87]) was utilized to roughly segment the bladder, bladder neck, and proximal urethra. This algorithm uses a mask provided by the user (Figure 3.3) as a starting point and performs a maximum of 275 iterations. The algorithm will stop evolution before the maximum number of iterations if the solution is found to be stationary [87]. An active contour method was chosen because similar methods performed well in Ciecholewski [81], [83] and Ciecholewski, et al. [82] for segmenting the gallbladder [81]–[83].

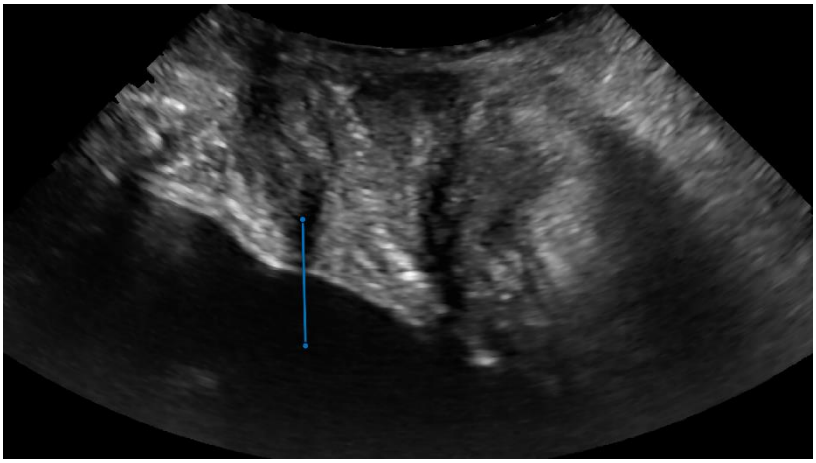


Figure 3.3: Example of user-provided mask, shown in blue, for semi-automated segmentation

3.3.3 Post-Processing

A morphological closing operation was performed with a disk structuring element (radius = 3), and morphological reconstruction was used to fill any holes in segmentation [88]. A radius of 3 was chosen by trial and error on a subset of images as the smallest diameter of structure element that could reliability close small holes and the largest diameter capable of reducing edge noise without erroneously joining distinct (verified by visual inspection) blobs.

a.  b. 

c. 

Figure 3.4: a. Rough segmentation output from active contour algorithm, b. Segmentation after morphological closing and hole filling, c. Segmentation after morphological opening.

Then a morphological opening operation [88] with a disk structuring element (radius = 2) was used to remove any tiny features which may erroneously connect blobs. A small radius of 2 was chosen to affect small features. Segmented blobs having areas less than 150 pixels were also removed to reduce the incidence of false negatives during BNF feature extraction. After initial testing, a moving median filter (k=10) was also applied to the spatial data to smooth the segmentation edges and prevent the influence of noise on feature extraction. In this case, k was chosen to be small compared to the number of entries in the total segmentation array and was fine-tuned on a subset of images.

3.4 AUTOMATED BNF CLASSIFICATION AND PUVA ESTIMATION

3.4.1 Automated BNF Classification

BNF classification was performed by analysing the number of segmented blobs. If there was only a single blob, BNF was classified as present, as a single connected blob

indicates no closure at the proximal urethra (Figure 3.5a,c). If there were two or more blobs, BNF was classified as absent (Figure 3.5b,d). Occasionally, the urethral segmentation is represented as multiple blobs because it appears to be broken up by noise and artifact.

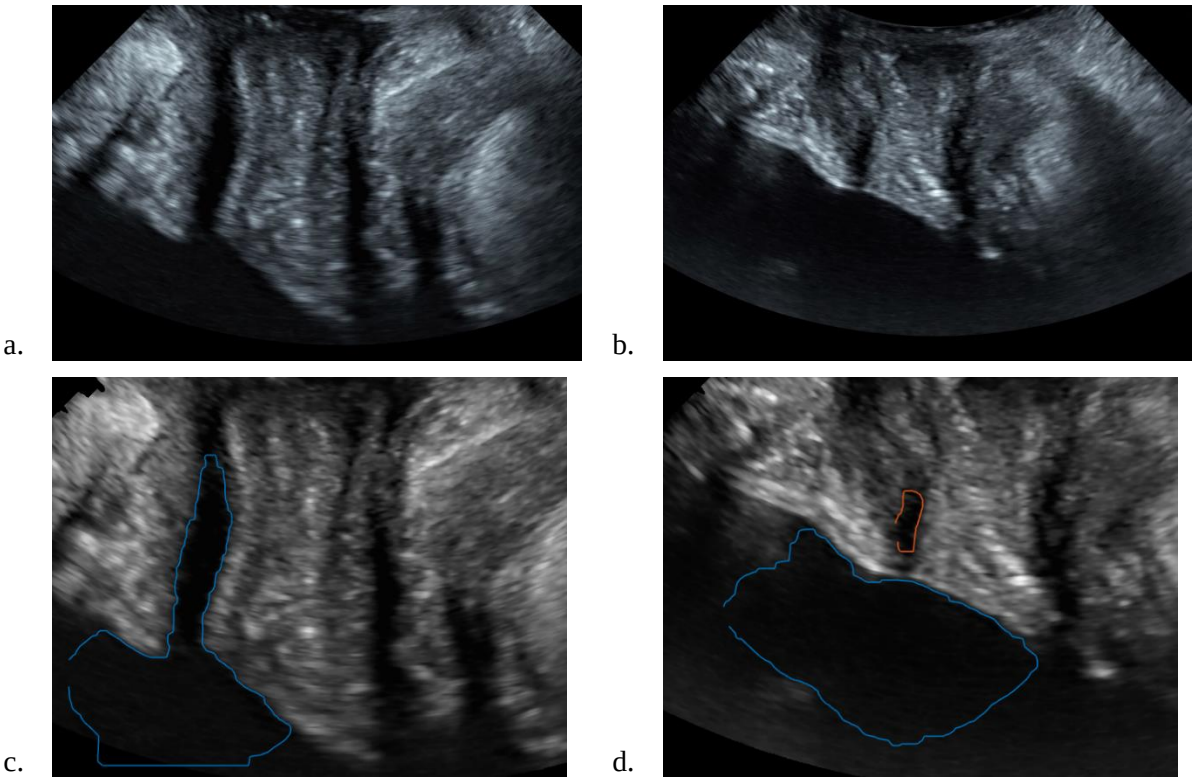


Figure 3.5: a. Example of a frame with BNF present, b. Example of a frame with BNF absent, c. Example of segmentation of a frame with BNF present, d. Example of segmentation of a frame with BNF absent

3.4.2 Automated PUVA Estimation

To estimate PUVA, the posterior urethrovesical junction (PUVJ) must be determined. In this thesis, PUVJ is defined as the point at which the posterior urethral wall and posterior bladder wall meet. The number of identified blobs was used to direct the method for determining the PUVJ; specifically, if only one blob or more than one blob (equivalently BNF classification) was identified. Next polylines were constructed, from the PUVJ, to describe the posterior bladder wall and the proximal posterior urethral wall. Then the polylines were fit with first-order polynomials, with the angle between the polynomials being the PUVA estimate.

3.4.2.1 Determining PUVJ with BNF absent

With BNF classified as absent, there was more than one blob segmented. Firstly, the bladder and urethral segmentation were separated using area filtering, with the largest blob

labeled as the bladder. The remaining blob or blobs were subjected to a morphological closing operation using a disk structuring element (radius = 9) to join smaller blobs that were in close proximity (Figure 3.6). A radius of 9 was chosen by trial and error on a subset of images with the intent of creating a urethral segmentation that approximated an ellipse with its major diameter generally oriented parallel to the cranial-caudal axis. After these morphological operations the largest urethral blob was labelled as the urethra. The largest blob was almost always the blob that was most proximal in the urethra.

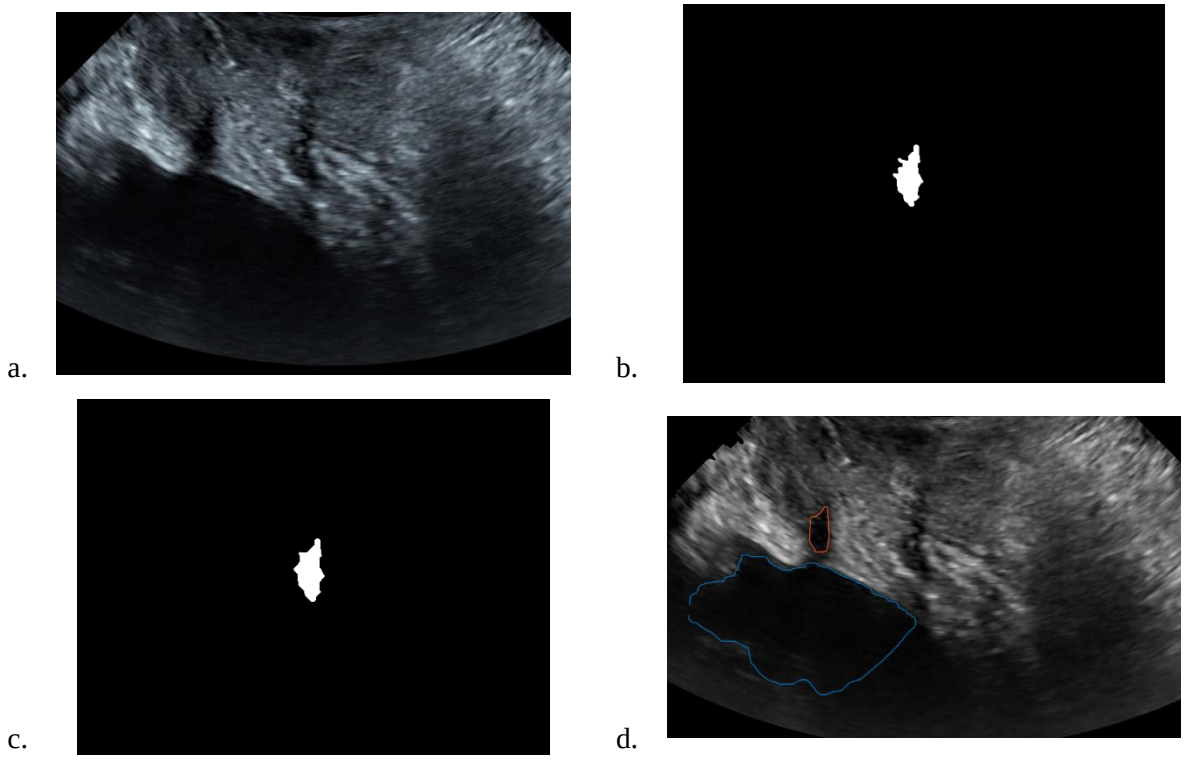


Figure 3.6: a. Original image, b. Urethral segmentation, c. Urethral segmentation after morphological closing applied, d. Final segmentation with smoothing.

The PUVJ was then determined from two candidates. To determine Candidate 1 first a urethral characteristic line had to be determined. If the urethral segmentation involved only one blob, the characteristic line was defined as the axis of least inertia (Figure 3.7). However, if the urethral segmentation involved more than one blob, the characteristic line was defined as the line of best fit through the blobs' centroids (Figure 3.8).

After a urethral characteristic line was determined, the slope signature of the bladder mask was determined using segments one-fiftieth of the bladder segmentation perimeter in length (Figure 3.9b). This threshold was determined by first selecting a starting point fraction of the segmentation perimeter that was thought would sufficiently smooth the noise

in segmentation to reveal the gross shape of the segmentation. The slope signature of the segment closest to the intersection point of the characteristic line and bladder segmentation, as defined above, and the slope signatures of the two segments before that in the posterior direction were analyzed. The beginning point of the segment with the largest magnitude slope signature, which was simultaneously less than magnitude one, was chosen as Candidate 1.

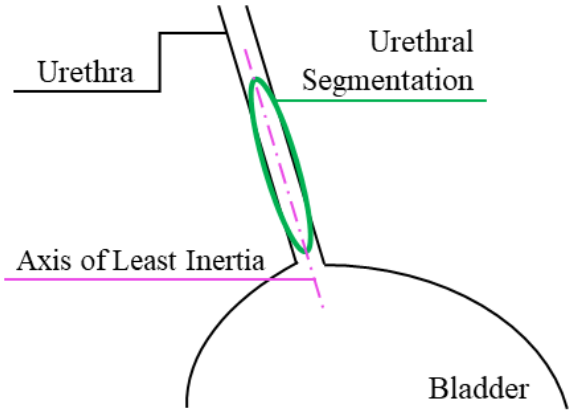


Figure 3.7: Example of when the axis of least inertia is used as the characteristic line to determine the PUVJ. Here it is seen that the axis of least inertia align parallel with the urethra.

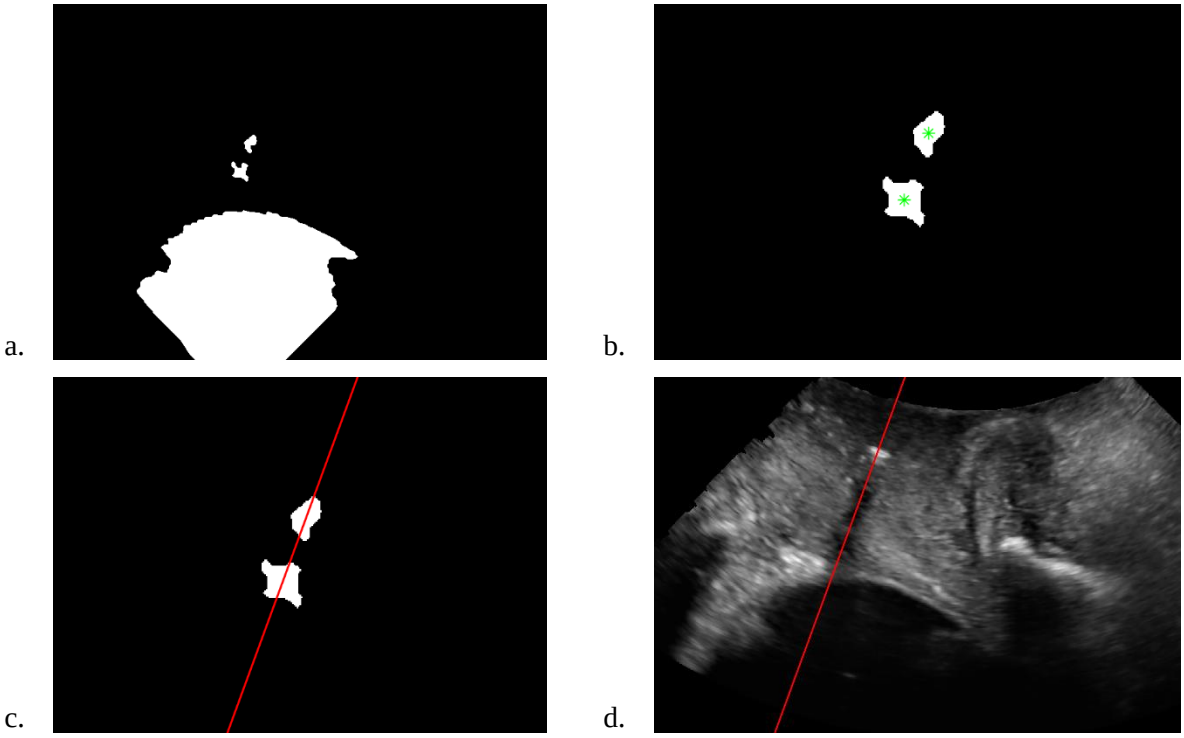
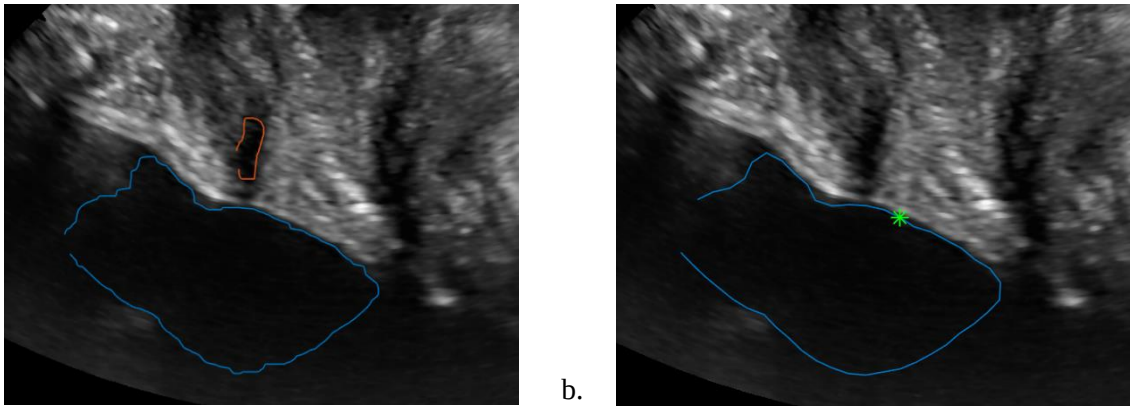


Figure 3.8: a. Bladder and urethral segmentation mask, b. Urethral segmentation mask with centroids in green, c. Urethral segmentation mask with characteristic line as per method 1b. shown in red, d.

Characteristic line shown on USI to accurately replicate the gross orientation of the proximal urethra long axis.

a. b.



c.

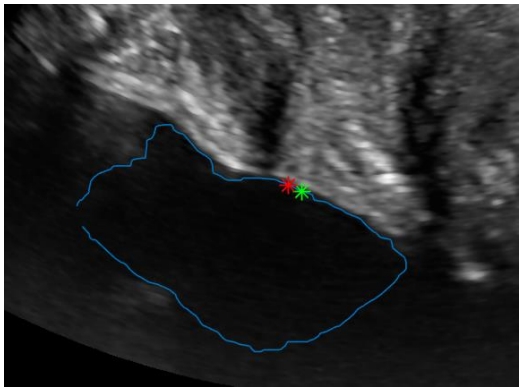


Figure 3.9: a. Segmentation of the bladder and proximal urethra, b. Polygon used to determine slope signature (blue) with Candidate 1 highlighted (green), c. Bladder segmentation with Candidate 1 highlighted in green and Candidate 2 highlighted in red. Candidate 2 was selected in this case.

To determine Candidate 2, the bladder mask was first rotated such that the lateral axis of the bladder (constructed of its most anterior and most posterior point) was horizontal. Candidate 2 was defined as the most caudal point of the bladder segmentation within 20 pixels in the lateral direction of Candidate 1 in this orientation.

If Candidate 2 was found to be sufficiently caudal to (>3 pixels) and reasonably near (<40 pixels) to Candidate 1, it was chosen as the PUVJ; otherwise, Candidate 1 was chosen. Thresholds were chosen to ensure that Candidate 2 would be sufficiently caudal to Candidate 1, to avoid errors attributed to noise, and reasonably near to Candidate 1 so as not to erroneously confuse the shadowing behind the pubic symphysis as the bladder neck and proximal urethra (Figure 3.9c).

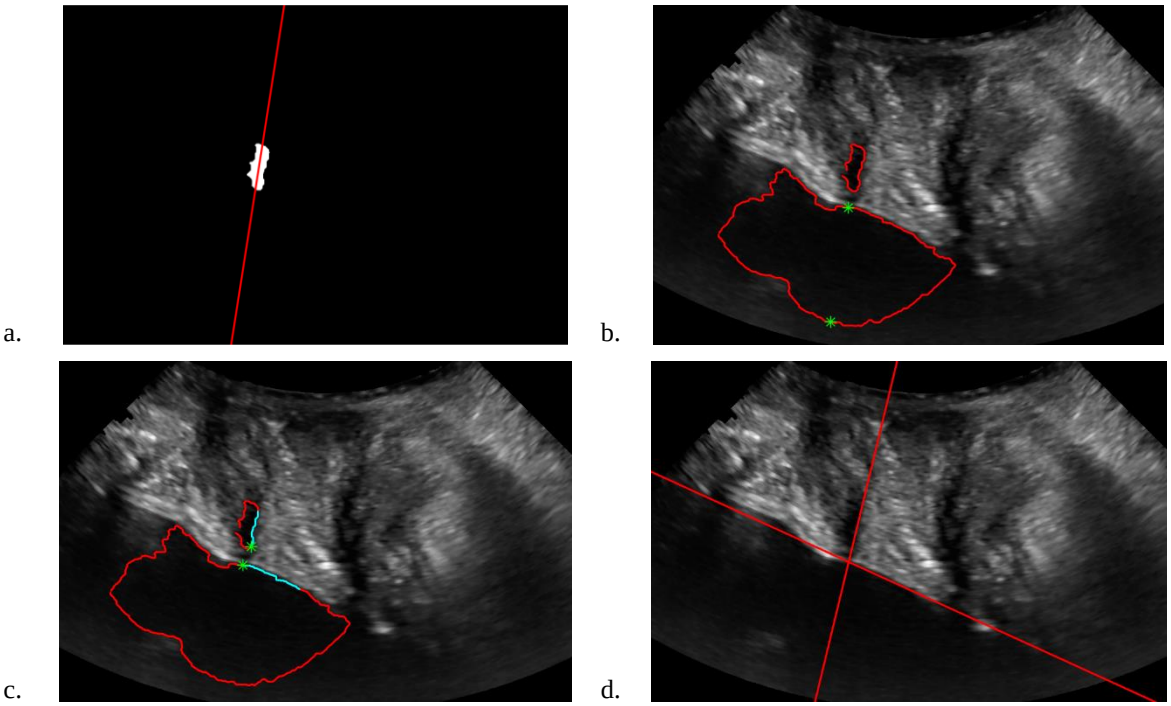


Figure 3.10: a. Urethra binary mask and respective axis of least inertia, b. Urethra and bladder segmentation lines (red) with axis of least inertia intersection points (green), c. Urethra and bladder segmentations with polyline segments highlighted in cyan and starting points highlighted in green, d. First-order polynomials which define the PUVA

3.4.2.2 Determining PUVJ with BNF Present

When BNF is present, the bladder and urethral segmentation cannot be separated, resulting in one segmented blob. Three triangle area representations (TARs) [89] of the segmented blob were calculated using polygons constructed with sides equal to one thirty-fifth, one-fiftieth, and one sixty-fifth of the segmentation perimeter (Figure 3.11b). These thresholds were determined by first selecting a starting point fraction of the segmentation perimeter that was thought would sufficiently smooth the noise in segmentation to reveal the gross shape, then selecting thresholds above and below this level, such that they produced resulting polygons which were clearly different. These thresholds were fine-tuned by trial and error on a subset of images. On each polygon, the index of the most caudal vertex was determined. Using this index, the vertex with the most concave TAR score (the most concave vertex) for each polygon was isolated from a subset of vertices containing the most caudal point of the segmentation and the 7, 9, or 11 vertices closest to it in the posterior-cranial direction, respectively. These ranges were determined by selecting a starting range that was thought would capture the correct most concave vertex while excluding the noise

that was occasionally present further up the posterior bladder wall. The median result from the three polygons was deemed to represent the PUVJ (Figure 3.11c).

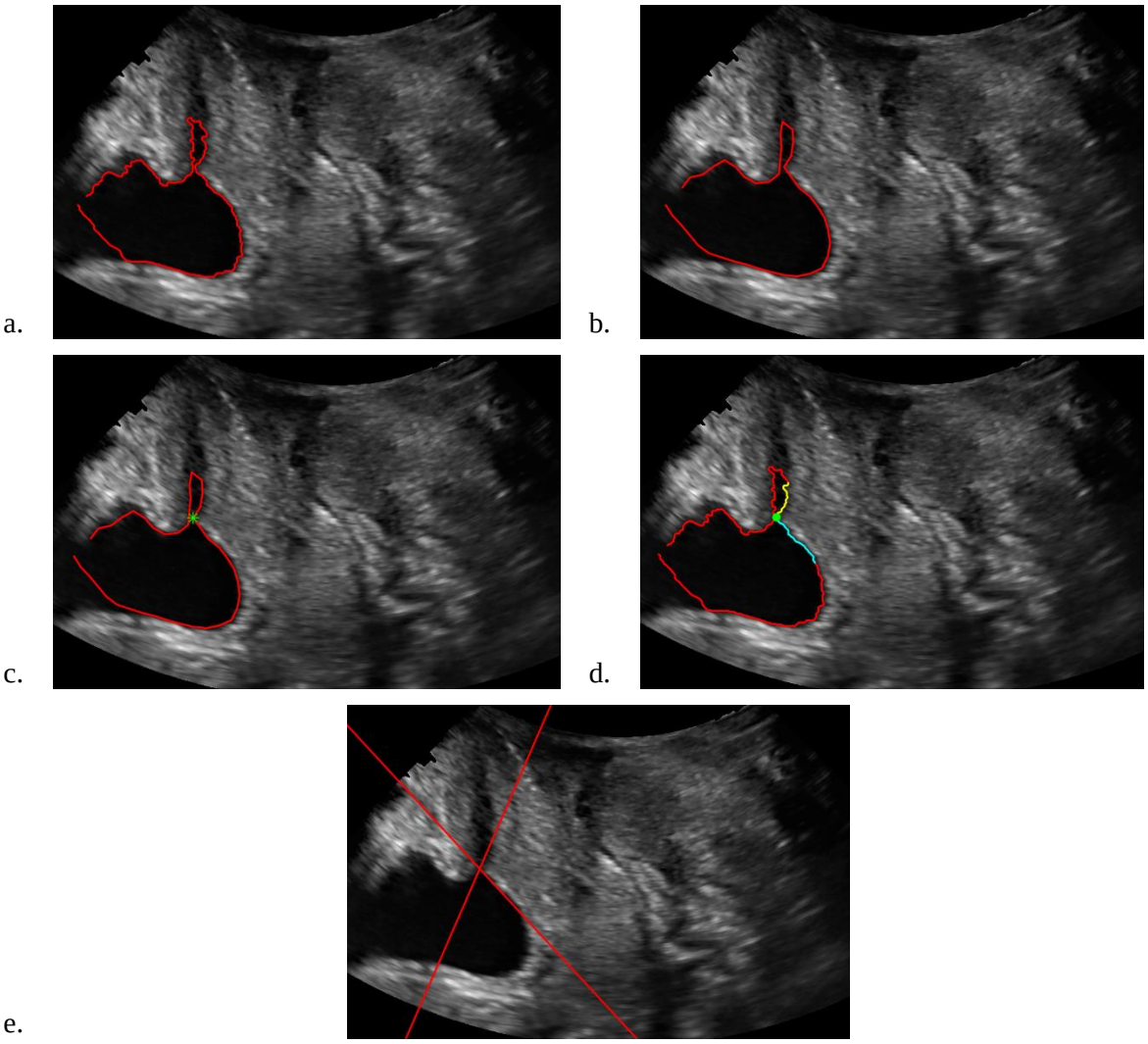


Figure 3.11: a. Urethra and bladder segmentation, b. Polygon used to determine TAR representation (segment length = segmentation perimeter/50), c. Polygon (segment length = segmentation perimeter/50) used to determine TAR representation with most concave vertex highlighted in green, d. Urethra and bladder segmentation with urethra polyline highlighted in yellow and bladder polyline highlighted in blue, e. First-order polynomials which define the PUVA

3.4.3 PUVA Estimation

A polyline was constructed to describe the posterior bladder wall, which began at the PUVJ and extended 75 pixels in length. The polyline length was chosen to reflect a similar length to that which the manual raters used when determining PUVA. Then a second polyline was used to define the posterior urethra. This polyline either began at the most posterior and most cranial aspect of the urethral segmentation, in the case of multiple blobs

after segmentation (BNF absent), or at the PUVJ, in the case of a single blob after segmentation (BNF present). The urethral polyline extended two-thirds the length of the segmented urethra; this generally represented approximately one third of the total urethral length. Two special cases were determined to improve PUVA estimations from small urethral segmentations:

1. In the case of multiple blobs after segmentation, if the urethral segmentation was less than 1500 pixels in area, the axis of least inertia was used as the urethra polyline. This threshold was determined by analyzing urethral segmentation area versus visually determined PUVA success on a subset of images.
2. In the case of a single blob after segmentation, if two thirds the length of the urethra was less than 70 pixels long (threshold determined by trial and error on a subset of data after a starting point was chosen, which was thought to represent a small urethral segmentation length) the portion of the TAR polygon which began at the PUVJ and extended half the length of the segmented urethra was used as the urethra polyline.

The urethra and bladder polylines were fit with first-order polynomials, and the angle between them was extracted as the PUVA.

3.3.4 Evaluation of the Performance of Semi-Automated Classification of BNF and Estimation of PUVA

All statistical analyses were performed using SPSS statistical package version 28 (SPP Inc, Chicago, IL). Bland-Altman analyses were performed using MATLAB 2020a. The semi-automated method was applied twice over the dataset; variations in the results are anticipated from the manually supplied mask used to start the segmentation. The intra-rater reliability of the estimated PUVA was evaluated using ICC coefficients and their 95% confidence intervals based on both single and average ratings, and absolute agreement based on a 2-way mixed-effects analysis of variance (ANOVA) model [84]. The intra-rater agreement for BNF classification was evaluated using Cohen's Kappa [90]. In the absence of a true gold standard, the concurrent validity of PUVA estimation was evaluated by comparing the semi-automated estimation to that of the manual measurements of each of the raters using ICC coefficients and their 95% confidence intervals based on both single-rating and average-ratings, and absolute agreement, this time based on a 2-way random-effects

ANOVA model [84]. The concurrent validity of BNF classification was evaluated using Light's Kappa using the subset of the data over which the manual raters were in agreement on the presence/absence of BNF and over which they reported that they were confident in their rating on both trials of manual analysis. Interpretation of kappa statistics followed McHugh's guideline [90] and interpretation of ICC statistics followed Koo et al.'s guideline [84]. Analyses of differences were conducted using the Bland-Altman method performed on the average of semi-automated estimations of PUVA over the two assessments and the average of the individual raters' estimations of PUVA over both assessments. To complement this analysis, linear regression analyses were used to evaluate systemic biases between PUVA estimated by the manual raters and the semi-automated method. Data sets were tested for normality before statistical methods were applied.

PUVA analysis was preformed over 3 datasets. The 'Full Dataset' (n=70), included the original dataset, excluding four images in which the PUVA measurement could not be determined. The 'Filtered Dataset' (n=52), excluded images from the Full Dataset that had poor posterior bladder wall or urethral segmentation caused chiefly by poor clarity (Appendix A). Excluded images were identified by the author, based on visual identification of urethral and bladder segmentation that contained obvious errors (Figure 3.12).

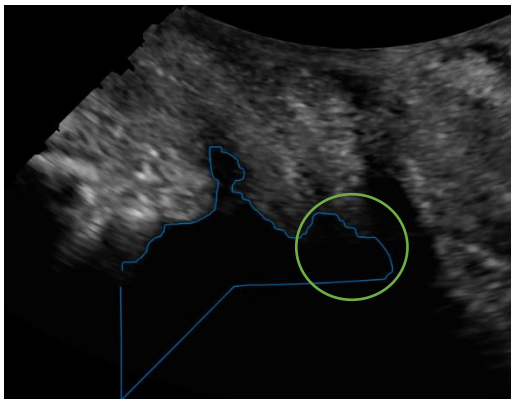
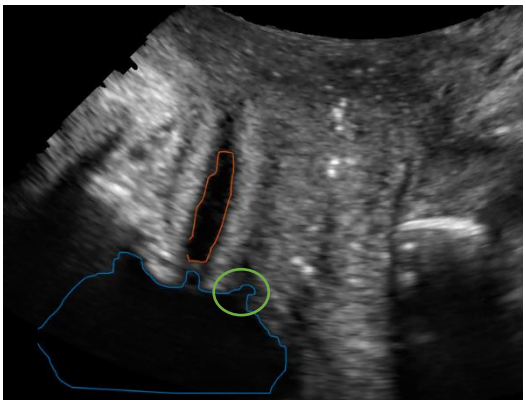
a.  b. 

Figure 3.12: Examples of obvious segmentation errors highlighted in green

The third dataset, called the 'Manually Corrected Dataset' (n=74), included all original images, but images with poor clarity were manually corrected in ImageJ using the paintbrush tool (all manually corrected images are displayed in Appendix B). Manual correction entailed painting over noise, artifact, or shadowing based on the author's best judgement. When manual correction was applied, care was taken to select a correction colour that closely matched that of segmented pixels.

The concurrent validity of the BNF classification was assessed on 2 datasets. The ‘Confidently Classified, Total Agreement BNF Dataset’ (n=15), was defined as the set over which the manual raters were in agreement on the presence/absence of BNF and over which they reported that they were confident in their rating on both manual assessments. The second dataset was constructed of trials where at least 3 of the 4 manual ratings were in agreement, regardless of confidence, called the ‘Moderate Agreement BNF Dataset’ (n=60).

Chapter 4 RESULTS

4.1 DATASET DEMOGRAPHIC INFORMATION

Key demographic information from the sample of participants is presented in Table 4.1 below. Forty-three frames were retained from images acquired before the run while 31 were retained from those acquired after the run.

Table 4.1: Participant demographic information of participant subset from REB #: H-06-18-759

	Total (n=30)
Age (years); mean (std)	38.4 (9.0)
Parity (children); median (range)	0 (0-3)
BMI (kg/m^2); mean(std)	22.60 (3.44)
Waist-to-Hip Ratio; mean(std)	0.78 (0.06)
Symptoms of SUI when Running (N)	13

Rater 1 identified BNF as present in 32 of 74 (43%) frames on both assessments. On the first assessment, they were confident in their classification on 48 out of 74 (64.9%) frames (Table C.1) and 47 of 74 (63.5%) frames on the second assessment (Table C.2). Rater 2 classified BNF as present in 38 of 74 (51.4%) frames on initial assessment and 48 of 74 (64.9%) frames in the second assessment (Table 4.2). Rater 2 reported they were confident in their classification on 54 of 74 (73.0%) frames at both assessments (Table C.3, Table C.4).

The mean (standard deviation) of the estimated PUVA for Rater 1 was 122.9 (17.7) degrees on the first assessment and 123.1 (14.4) degrees on the second assessment. The mean (standard deviation) of estimated PUVA for Rater 2 on the first assessment was 120.3 (21.8) degrees and 117.5 (22.7) degrees on the second assessment.

4.2 RELIABILITY OF MANUAL ANALYSIS OF BNF AND PUVA

The intra-rater agreement of BNF classification was moderate to almost perfect [90] (Table 4.2), while the inter-rater agreement of BNF classification was weak (κ_L= 0.498 (95% CI: 0.312 – 0.683)). Inter-rater agreement of BNF classification was significantly improved when examining the Confidently Classified BNF Dataset (n=17), (κ_L= 0.823 (95% CI: 0.556 – 1.000)). For the subset of data where both raters were confident in at least one of their

assessments (called the 'Moderately Confident BNF Dataset' (n=57)) inter-rater agreement was weak (κ_L= 0.596 (95% CI: 0.409 – 0.782)). Rating and confidence matrices for each rater over each assessment are available in Appendix C.

Intra-rater reliability of PUVA estimation was good (Table 4.2) while inter-rater reliability was moderate (ICC(2,1) = 0.725 (95% CI: 0.637 – 0.802)).

Table 4.2: Intra-rater reliability of manual classification of BNF and estimation of PUVA (n=74)

Bladder Neck Funneling Classification				
	Number of Individuals rated as BNF present		Cohen's Kappa (95% CI)	p-value
Rater 1	Trial 1	32	1.00 (1.00 – 1.00)	< 0.001
	Trial 2	32		
Rater 2	Trial 1	38	0.728 (0.577 – 0.879)	< 0.001
	Trial 2	48		
Posterior Urethrovesical Angle Estimation				
	Mean PUVA (std) (degrees)		ICC(3,1) (95% CI)	p-value
Rater 1	122.981 (16.132)		0.823 (0.733 – 0.885)	< 0.001
Rater 2	118.918 (22.252)		0.840 (0.756 – 0.897)	< 0.001

4.3 AGREEMENT OF SEMI-AUTOMATED CLASSIFICATION OF BNF

The intra-rater agreement of BNF classified using the semi-automated method was almost perfect.

Table 4.3: Intra-rater reliability of semi-automated classification of BNF

	Number of Individuals rated as BNF present		**Cohen's Kappa (95% CI)**	**p-value**
Semi-Automated Method (n=74)	Trial 1	38	1.00 (1.00 – 1.00)	< 0.001
	Trial 2	38		

4.4 RELIABILITY OF SEMI-AUTOMATED ESTIMATION OF PUVA

A PUVA measurement could not be determined on four images because an intersection between the characteristic line and the bladder segmentation did not exist. This was generally caused by a small urethral segmentation caused by noise, or by noise and shadowing in the bladder.

Over the Full Dataset, the intra-rater reliability of PUVA estimation was good when the model used single ratings and excellent when the model used average ratings. Evaluated over the Filtered Dataset and the Manually Corrected Dataset, the intra-rater reliability of PUVA estimation was excellent for both single- and average-ratings.

Table 4.4: Intra-rater reliability of semi-automated estimation of PUVA

	Mean PUVA (std) (degrees)	**ICC(3,1) (95% CI)**	**p-value**
Full Dataset* (n=70)	121.282 (21.777)	0.860 (0.784 – 0.910)	< 0.001
Manually Corrected Dataset↓ (n=74)	122.441 (19.769)	0.942 (0.909 – 0.963)	< 0.001
Filtered Dataset** (n=52)	120.934 (19.491)	0.946 (0.908 – 0.969)	< 0.001
	Mean PUVA (std) (degrees)	**ICC(3,2) (95% CI)**	**p-value**
Full Dataset* (n=70)	121.282 (21.777)	0.925 (0.879 – 0.953)	< 0.001
Manually Corrected Dataset↓ (n=74)	122.441 (19.769)	0.970 (0.953 – 0.981)	< 0.001
Filtered Dataset** (n=52)	120.934 (19.491)	0.972 (0.952 – 0.984)	< 0.001

** Full Dataset included the original dataset, excluding four images in which the PUVA measurement could not be determined.*

*** Filtered Dataset excluded images from the Full Dataset that had poor segmentation caused chiefly by poor clarity*

↓ Manually Corrected Dataset included all original images, but images with poor clarity were manually corrected in ImageJ

4.5 CONCURRENT VALIDITY OF SEMI-AUTOMATED CLASSIFICATION OF BNF AND ESTIMATION OF PUVA

Agreement between BNF classification by the semi-automated method and the manual analysis performed by each of the raters on the Confidently Classified, Total Agreement BNF Dataset (n=15) was almost perfect, with Light's Kappa Statistic 1.00 (95% CI: 1.00 – 1.00). Concurrent validity of BNF classification was strong (κ_L= 0.811 (95% CI: 0.667 – 0.955)) when assessing the Moderate Agreement BNF Dataset (n=60).

Concurrent validity of PUVA estimation over the Full Dataset between the semi-automated method and each of the two raters was moderate for single measures and excellent on average measures (Table 4.5). Evaluated over the Filtered Dataset and Manually Corrected Dataset described in Section 3.3.4, the concurrent validity of PUVA estimation was moderate for single measures and excellent for average measures, with higher ICC values than observed for the Full Dataset.

Table 4.5: Concurrent validity of PUVA estimation as assessed by ICC against the manual raters

Single Ratings	**ICC(2,1) (95% CI)**	**p-value**
Semi-Automated Method with Full Dataset v. Manual Raters (n=70)	0.610 (0.514 – 0.705)	< 0.001
Semi-Automated Method with Manually Corrected Dataset v. Manual Raters (n=74)	0.719 (0.640 – 0.792)	< 0.001
Semi-Automated Method with Filtered Dataset v. Manual Raters (n=52)	0.691 (0.591 – 0.784)	< 0.001
Average Ratings	**ICC(2,2) (95% CI)**	**p-value**
Semi-Automated Method with Full Dataset v. Manual Raters (n=70)	0.904 (0.864 – 0.935)	< 0.001
Semi-Automated Method with Manually Corrected Dataset v. Manual Raters (n=74)	0.939 (0.914 – 0.958)	< 0.001
Semi-Automated Method with Filtered Dataset v. Manual Raters (n=52)	0.931 (0.897 – 0.956)	< 0.001

4.6 ANALYSIS OF DIFFERENCES IN PUVA ESTIMATION BETWEEN MANUAL AND SEMI-AUTOMATED METHODS

Before performing statistical analyses, all dataset were tested and confirmed to be normally distributed using the Kolmogorov-Smirnov test (α=0.05). The Bland-Altman analysis of the average of semi-automated estimations of PUVA over the two assessments compared to the average of the individual raters' estimations of PUVA was conducted over the Full Dataset. This analysis, shown in Figure 4.1 below, resulted in a non-significant bias of 0.16 degrees. Three trials lay outside of the limits of agreement (-36 to 36 degrees), all of which were found to be the result of poor segmentation resulting from shadowing, noise, or artifact. The limits of agreement for this dataset cover a large range of values, suggesting poor confidence in the measure. The linear regression shows that only 30% of the variance in the model is explained by the line of best fit. Additionally, a large bias (38.5 degrees) was found.

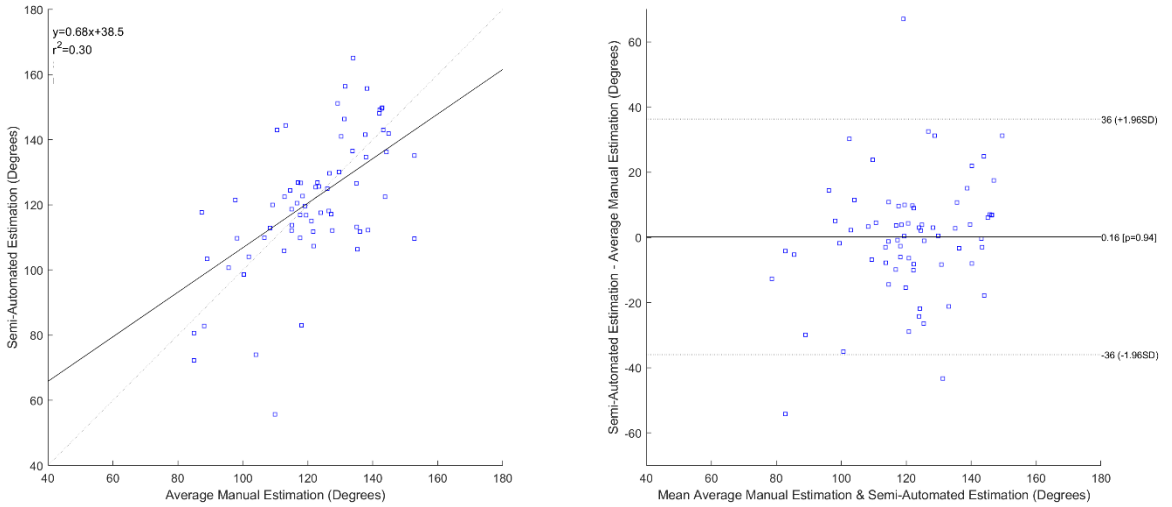


Figure 4.1: Bland-Altman plot of manual estimation vs semi-automated estimation of PUVA over the Full Dataset (where SD stands for standard deviation)

The Bland-Altman plot for the Manually Corrected Dataset is shown in Figure 4.2. This analysis resulted in a non-significant bias of 1.5 degrees, with the semi-automated estimation resulting in slightly larger angles than the manual estimation. Six trials lay outside of the limits of agreement. Table 4.6 details the errors thought to have led to poor agreement for these six trials. Note that the three trials previously outside the limits of agreement due to poor segmentation from shadowing or noise were now within the limits of agreement after manual correction. The limits of agreement were much smaller for the Manually Corrected Dataset (-23 to 26 degrees). The linear regression explained 61% of the variance in the data with a 16.4 degree bias, which is comparable to the results of the linear regression comparing PUVA estimated by Rater 1 to PUVA estimated by Rater 2 (Appendix C).

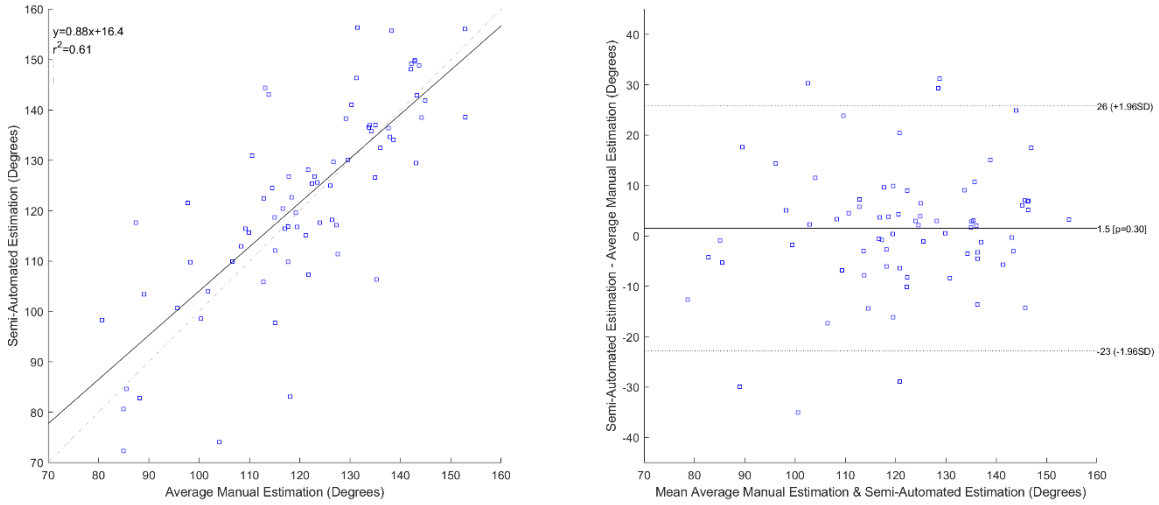


Figure 4.2: Bland-Altman plot of PUVA estimated through manual vs. semi-automated method over the Manually Corrected Dataset (where SD stands for standard deviation)

Table 4.6: Errors thought to lead to poor agreement between PUVA estimated through manual and semi-automated method over the Manually Corrected Dataset.

Reason	Frequency
Incorrect PUVJ identified.	3
Cause unknown	2
“Loose” bladder neck segmentation	1

The Bland-Altman plot for PUVA estimated using the Manually Corrected Dataset with manual rater disagreements that were greater than +/- 10 degrees removed (n = 50) is shown in Figure 4.3. This analysis resulted in a non-significant bias of 1.1 degrees. Note that limits of agreement (-18 to 20 degrees) were smaller after manual rater disagreements were removed. Three trials lay outside the limits of agreement, Table 4.7 details the errors thought to have led to poor agreement for the three trials that lay outside of the limits of agreement. For this dataset, the linear regression explained 75% of the variance in the data, and had a 4.37 degree bias which is superior to the results of the linear regression comparing Rater 1 to Rater 2 (Appendix C).

Table 4.7: Errors thought to lead to poor agreement between manual estimation method and semi-automated estimation method over Manually Corrected Dataset with rater disagreements removed.

Reason	Frequency
Incorrect PUVJ identified.	2
Cause unknown	1

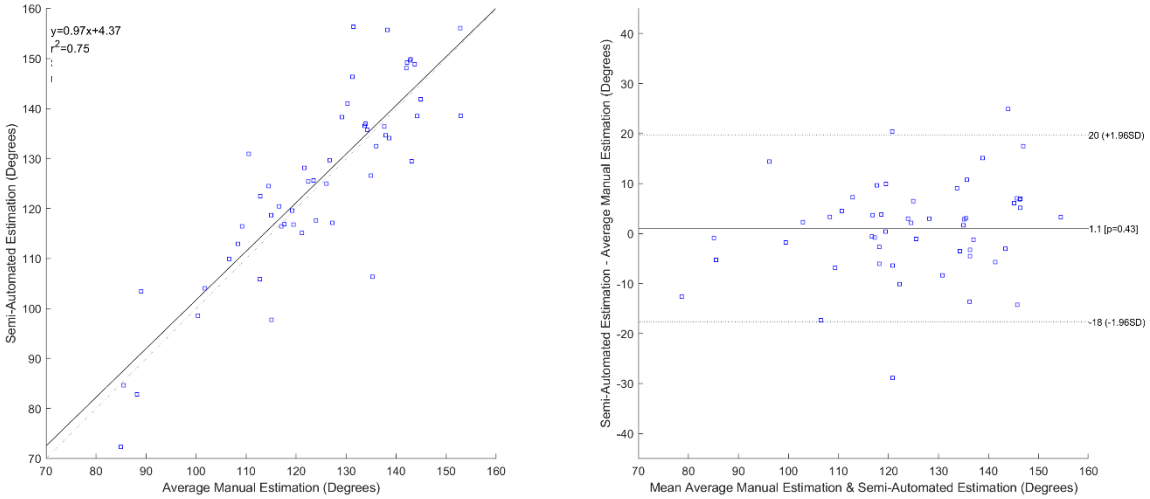


Figure 4.3: Bland-Altman plot of manual estimation vs. semi-automated estimation of PUVA over the Manually Corrected Dataset when manual rater disagreements are removed (where SD stands for standard deviation)

Lastly, the Bland-Altman plot for the Filtered Dataset is shown in Figure 4.4. This analysis resulted in a non-significant bias of 1.5 degrees, suggesting that the semi-automated estimation produces slightly larger PUVAs than manual estimation. Five values lay outside of the limits of agreement. Table 4.8 details the errors thought to have led to this poor agreement. Note that one value previously outside the limits of agreement due to incorrect identification of the PUVJ was removed when filtering the dataset. Interestingly, the limits of agreement were slightly larger when the data were filtered (-24 to 27 degrees) when they were compared to manually corrected (-23 to 26 degrees). The linear regression of the filtered data explained 57% of the variance in the data, with a smaller bias than both the Full Dataset and Manually Corrected Dataset.

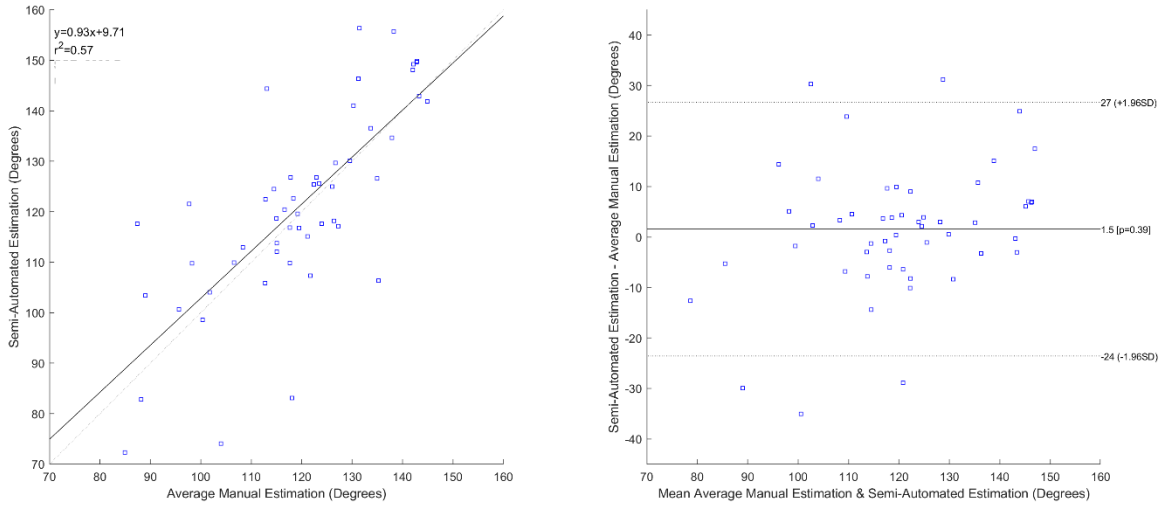


Figure 4.4: Bland-Altman plot of manual estimation vs. semi-automated estimation of PUVA over the Filtered Dataset (where SD stands for standard deviation)

Table 4.8: Errors thought to lead to poor agreement between PUVA estimated using manual method and semi-automated method over the Filtered Dataset.

Reason	Frequency
Incorrect PUVJ identified.	2
Cause unknown	2
"Loose" bladder neck segmentation	1

In the figures below, examples are given of excellent (Figure 4.5), moderate (Figure 4.6), and poor (Figure 4.7) agreement cases. The poor agreement case shown is a case where the PUVJ is incorrectly identified because the urethra was not segmented.

a. 

b. 

Figure 4.5: Example of excellent agreement: a. Rater 2 estimation of PUVA, b. Semi-automated method of estimating PUVA

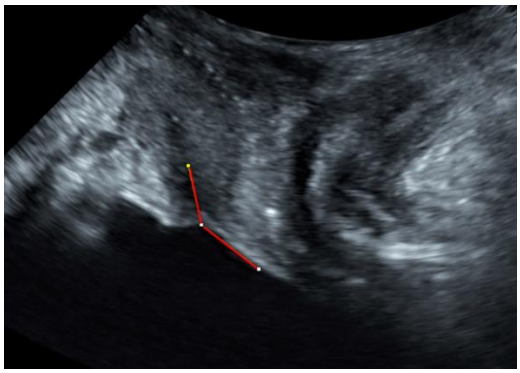
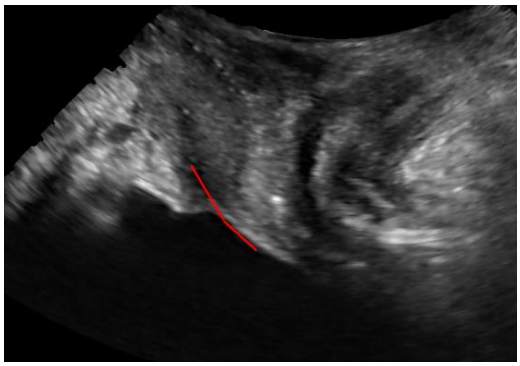
a.  b. 

Figure 4.6: Example of moderate agreement: a. Rater 2 estimation of PUVA, b. Semi-automated method of estimating PUVA

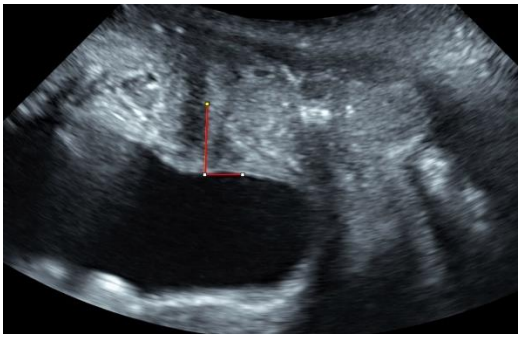
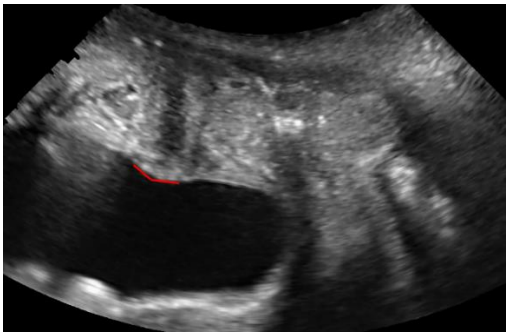
a.  b. 

Figure 4.7: Example of poor agreement: a. Rater 2 estimation of PUVA, b. Semi-automated method of estimating PUVA (Note that the segmentation missed the urethra)

Chapter 5 DISCUSSION

The objectives of this thesis were met - manual and semi-automated methods of classifying BNF and estimating PUVA were developed and evaluated. The intra-rater agreement of manual BNF classification and intra-rater reliability PUVA estimation were consistent with the existing literature [53], [55]. The inter-rater agreement of the manual BNF classification and inter-rater reliability PUVA estimation were lower than reports in the literature [53], [64], [65]. This may be due to the low resolution in the image set around the bladder neck and proximal urethra, since imaging was not optimized for high resolution in this region. The intra-rater agreement BNF classification and intra-rater reliability PUVA estimation of the semi-automated method was consistent with or exceeded those of the manual raters. The concurrent validity of semi-automated BNF classification against the manual classification was better than inter-rater agreement observed between the manual raters. The concurrent validity of semi-automated PUVA estimation against manual estimation was lower than the inter-rater reliability between manual raters for the Full Dataset and Filtered Dataset. However, the concurrent validity of semi-automated PUVA estimation against manual estimation for the Manually Corrected Dataset was consistent with inter-rater reliability between manual raters. The analysis revealed that segmentation errors and incorrect identification of the PUVJ are two areas that require improvement.

5.1 MANUAL ANALYSIS

5.1.1 Reliability of BNF Classification

The level of intra-rater agreement for BNF classification found in our investigation is lower (κ_C = 0.728 – 1.00) but comparable to the agreement reported by Cassadó et al. [53] (100% agreement). Indeed, our estimation of intra-rater agreement may be more conservative as Cassadó et al. chose not to evaluate agreement using Cohen's Kappa [53], which seeks to account for agreement due to chance. The definition for BNF used by Cassadó et al. ("widening of the bladder neck" [53]) differs from the definition used in this thesis ("opening of the proximal urethra"). It is possible that bladder neck widening is more easily detected than proximal urethra opening, which requires visible isoechogeneity between the bladder and the proximal urethra. Cassadó et al. also evaluated BNF at the peak of straining [53] vs at rest. In our experience, during straining the hypoechoic representation

of the urethra is often lost, and therefore it is difficult to determine whether the proximal urethra is indeed open or whether the bladder neck has merely descended and become funnel-shaped thus appearing that the proximal urethra is opened.

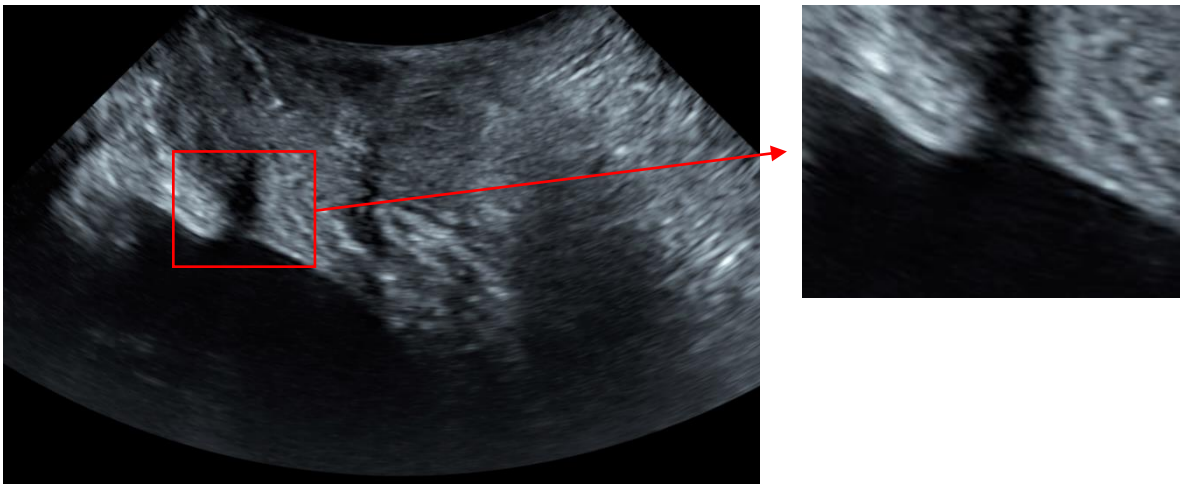


Figure 5.1:Example of a frame which rater 1 and rater 2 have conflicting ratings for BNF classification

The inter-rater agreement of BNF classification was considerably lower (κ_L= 0.498 (95% CI: 0.312 – 0.683)) than expected based on reports in the literature [53], [65]. Schaer et al. used the same definition for BNF as used in this research (i.e., the opening of the proximal urethra) and evaluated agreement on both rest and during straining [65]. Thus, the agreement reported by Schaer's is more relevant than that of Cassadó et al. in terms of comparisons with our findings. When evaluating inter-rater agreement in BNF classification over the Moderately Confident Dataset (κ_L= 0.596 (95% CI: 0.409 – 0.782)), levels of agreement comparable to the literature were not observed [53], [65]. However, when examining the Confidently Classified Dataset (κ_L= 0.823 (95% CI: 0.556 – 1.000)) the results are comparable to those of Schaer, though our result is still notably poorer [53], [65]. This leads us to question why the inter-rater agreement was so low in terms of BNF classification?

Rater 2 was more confident in their ratings (approximately 73% of their ratings) than Rater 1 (approximately 64% of their ratings). Additionally, Rater 1 was more likely classify BNF as absent, whereas Rater 2 was more likely classify BNF as present. This discrepancy may be due to differences in interpretation of hyperechoic pixels around the bladder neck; Rater 1 may have interpreted this as legitimate tissue that was poorly visualized, while Rater 2 may have interpreted this as noise (Figure 5.1). For this reason, it is theorized that the

collection of a dataset where the bladder neck is more clearly visualized would improve the intra- and inter-rater agreement. Additionally, a dataset with clearer visualization around the bladder neck may increase the confidence of the raters in their classification. Bearing in mind that the images used for this analysis were not acquired with specific attention to resolution at the bladder neck, a newly acquired data set with such a focus may yield better repeatability in the classification of BNF.

5.1.2 Reliability of Manual PUVA Estimation

Intra-rater reliability of PUVA estimation was good, as expected (ICC(3,1): 0.823 – 0.840), with results similar to reports from Pontbriand-Drolet et al. (ICC: 0.81 (SEM: 8.48)) [55], and the PUVA definition used by Pontbriand-Drolet et al. is sufficiently similar to the definition of PUVA used in this analysis [55]. That said, inter-rater reliability of PUVA estimation was lower than anticipated (ICC(2,1) = 0.725 (95% CI: 0.637 – 0.802). In Wen et al. the inter-rater reliability of PUVA estimates was reported as excellent (ICC: 0.91 (95% CI: 0.83-0.97)) [64]. Although Wen et al. did not explicitly define PUVA, their definition can be inferred through their figures and appears to be sufficiently similar to ours [64].

However, Wen et al. did not provide any details of the ICC model [64], which creates challenges when comparing results. Firstly, describing the analysis as both a test-retest and inter-rater reliability evaluation is problematic. Normally, a test-retest evaluation would utilize a 2-way **mixed**-effects ANOVA model, whereas an inter-rater reliability assessment would utilize a 2-way **random**-effects ANOVA model, as suggested in Koo et al.'s ICC guideline [84]. Additionally, Wen et al. do not specify whether they evaluated reliability using single measures or average measures, nor whether they evaluated absolute agreement or consistency [64]. Indeed, if Wen et al. used a 2-way random effects ANOVA model to evaluate **average**-rating, absolute agreement, our assessment of inter-rater reliability (ICC(2,2): 0.913, 95% CI: 0.875-0.942) would be consistent with that of Wen et al. [64]. Similar to what is described above for BNF classification, the intra- and inter-rater reliability of PUVA estimation would likely be improved by acquiring a new dataset of images with a focal point for high resolution at and around the bladder neck.

5.2 RELIABILITY OF THE SEMI-AUTOMATED METHOD

The intra-rater agreement of BNF classification, as assessed by Light's Kappa, was almost perfect (κ_L = 1.00). The results are consistent with those from the manual raters using the same dataset (κ_C = 0.728 – 1.00) and those from manual raters reported in the literature (100% agreement) [53]. While there are no semi-automated or automated methods for BNF classification in the literature with which to provide a direct comparison, this result suggests that the method developed here has excellent repeatability.

The intra-rater reliability of PUVA estimation for the Full Dataset, as assessed by ICC using single ratings, was good (ICC(3,1): 0.860 (95% CI: 0.784-0.910)). In fact, it was slightly more reliable than the manual raters when assessed over the entire dataset. Additionally, the method was much more reliable than the manual raters (ICC(3,1): 0.823 – 0.840) on the Manually Corrected Dataset (ICC(3,1): 0.942 (95% CI: 0.909-0.963)) and on the Filtered Dataset (ICC(3,1): 0.946 (95% CI: 0.908-0.969)). The method also performed better than the reliability reported by Pontbriand-Drolet et al. using manual rating based on an MRI dataset recorded at rest [ICC = 0.81 (SEM: 8.48)], and on straining [ICC = 0.87 (SEM: 6.54)] [55]. Similar to BNF classification, there are no semi-automated or automated methods of PUVA estimation described in the literature against which we can compare our method, however the method appears to be adequately reliable to be useful in both research and clinical evaluation.

As expected, the intra-rater reliability of PUVA estimation, was better when using average ratings than single ratings. This suggests that estimating PUVA multiple times on the same frame by slightly varying the user-provided mask or evaluating multiple frames in succession using the same mask could result in a more reliable estimation of PUVA. These results could also imply that to improve reliability, a more robust segmentation approach should be implemented, one that is either fully automated, or is more invariant to small variations in the user-provided mask.

The intra-rater reliability results demonstrated in this investigation highlight the efficacy of our method at removing inconsistencies related to manual segmentation, particularly around the estimation of PUVA. In addition to having better reliability than manual segmentation, the nature of the method provides explicit definitions for BNF and PUVA, which can be explained in a series of steps with set thresholds, allowing for exact

replication. Conversely, the definitions provided to inform manual rating rely somewhat on judgement, thus they allow room for interpretation which may lead to discrepancies in results between investigations. Additionally, the semi-automated method provides time and effort savings to the user.

5.3 CONCURRENT VALIDITY OF SEMI-AUTOMATED METHOD

Concurrent validity of the semi-automated method of BNF classification exceeded expectations, resulting in 100% agreement with the manual raters over the Confidently Classified, Total Agreement BNF Dataset. However, the sample size of this subset was small (n=15). Thus, although the results are promising, their generalizability is poor. The results over the Moderate Agreement Dataset (n=60) showed strong concurrent validity (κ_L= 0.811 (95% CI: 0.667 – 0.955)) and presented a more generalizable estimation of how the method would perform on an optimized dataset. The results on the Moderate Agreement Dataset are lower than inter-rater agreement levels reported in the literature (100% agreement); this is likely influenced by poor agreement between the manual raters [53], [65].

The concurrent validity of the semi-automated method of PUVA estimation, as assessed by ICC using single ratings, was lower than expected over the Full Dataset (ICC(2,1): 0.610 (95% CI: 0.514 – 0.705)) and Filtered Dataset (ICC(2,1): 0.691 (95% CI: 0.591 – 0.784)). However, it performed similarly to the inter-rater reliability of the manual raters (ICC(2,1): 0.725 (95% CI: 0.637 – 0.802)) on the Manually Corrected Dataset (ICC(2,1): 0.719 (95% CI: 0.640 – 0.792)). This suggests that segmentation error was a large contributor to low validity. When comparing the concurrent validity of the Filtered Dataset to the Manually Corrected Dataset, the reliability decreases. This points to errors in the feature extraction process which become more apparent at a smaller sample size (n=74 vs n=52) once the frames with poor segmentation results are removed.

Concurrent validity may be improved by consolidating the way in which PUVA estimates are calculated between the manual and semi-automated methods. The semi-automated method calculates the PUVA from two lines. The manual method also calculates PUVA from two lines; however, the manual method uses three points to construct these two lines, which encourages the manual rater to place the middle dot at the PUVJ (Figure 3.1). The semi-automated method does not impose this constraint but instead defines the line of

best fit of the posterior bladder wall and posterior urethral wall without considering where these lines intersect. Often it is near the PUVJ, but it is possible that this intersection point is distinct from the PUVJ as shown in Figure 4.6. Additionally, it is possible the urethra line of best fit, although a good approximation of the gross direction of the urethra, could be a poor approximation of the proximal section of the urethra. These phenomena may be corrected by first acquiring a dataset that provides more clear visualization of the proximal urethra. Then by weighing the co-ordinates of the polyline used to create the polynomial such that points closer to the identified PUVJ are more influential in the polynomial's creation. Additionally, more accurate identification of the PUVJ could reduce the effect of these differing constraints by resulting in lines of best fit that more accurately follow the paths of the posterior bladder and posterior urethral walls.

The ICC values for average ratings are far superior (ICC(2,2): 0.904 – 0.931 [depending on dataset]) than those of the single ratings when evaluating the concurrent validity of PUVA estimation. As such, as discussed above, the method may be improved by estimating PUVA multiple times either on the same frame or using multiple frames in succession, then using the average result as the PUVA estimation. Conversely, these results could also suggest that a more robust segmentation approach is needed, one that is either fully automated, or is more invariant to small variations in the user-provided mask. Moreover, reliable identification of the PUVJ would likely improve the estimation of PUVA using the approach developed for this thesis.

Results from the Bland-Altman analysis for PUVA estimation over the Full Dataset were worse than expected. Although the bias on the Bland-Altman plot (Figure 4.1) was low (0.16 degrees), the limits of agreement were extremely large (-36<LOI<36 degrees) and the linear regression explained little of the variance in the data ($r^2 = 0.30$). All of the data points which fell outside the limits of agreement could be explained by segmentation errors caused by shadowing, noise, or artifact in the images. When comparing the Bland-Altman results on the Full Dataset to those of the manual raters (Figure C.1), although the bias over the Full Dataset is smaller (d = - 4.1 degrees vs. 0.16 degrees), the limits of agreement were wider (-30<LOI<22 vs. -36<LOI<36 degrees) and the regression explained less of the variance in the data (r^2 =0.62 vs 0.30); this concurrent validity is unacceptable. Segmentation error appears to be a large source of disagreement between the manual and semi-automated raters.

Evaluation of the Filtered Dataset leads to a similar conclusion (Figure 4.4). Although the bias (1.5 degrees) of the Bland-Altman plot and the limits of agreement (-24 to 27 degrees) are comparable to the agreement between raters, the relationship is still weak (r^2 = 0.57). Examining the results from images that fell outside the limits of agreement for the Filtered Dataset highlighted that the PUVJ is sometimes incorrectly identified. Indeed, during the development of the method, the identification of the correct PUVJ was a challenging task. We expect that generating a dataset with better visualization of the bladder neck will allow for fine-tuning of the thresholds in the PUVJ identification method, leading to higher performance. Additionally, on two images that produced PUVA estimates outside the limits of agreement, it appears that the semi-automated method was 'more' representative of the true PUVA than the results of the manual raters (Figure 5.2). Thus improved rater performance may result in improved concurrent validity without requiring further method development.

Lastly, the bias on the Bland-Altman plot (1.5 degrees), and the limits of agreement (-23<LOI<26 degrees) were both lower for the Manually Corrected Dataset (Figure 4.2) than what was observed between the manual raters, while the regressions explained similar amounts of the variance in the data (r^2 = 0.61). However, there were more datapoints outside the limits of agreement for the Manually Corrected Dataset than there were for the agreement between the manual raters. Most outliers in the Manually Corrected Dataset appear to be caused by incorrect PUVJ identification, or are cases where the semi-automated estimation appears to be accurate, and the manual raters appear to be incorrect (Figure 5.2). As stated above, this implies that improved manual rater performance could improve concurrent validity.

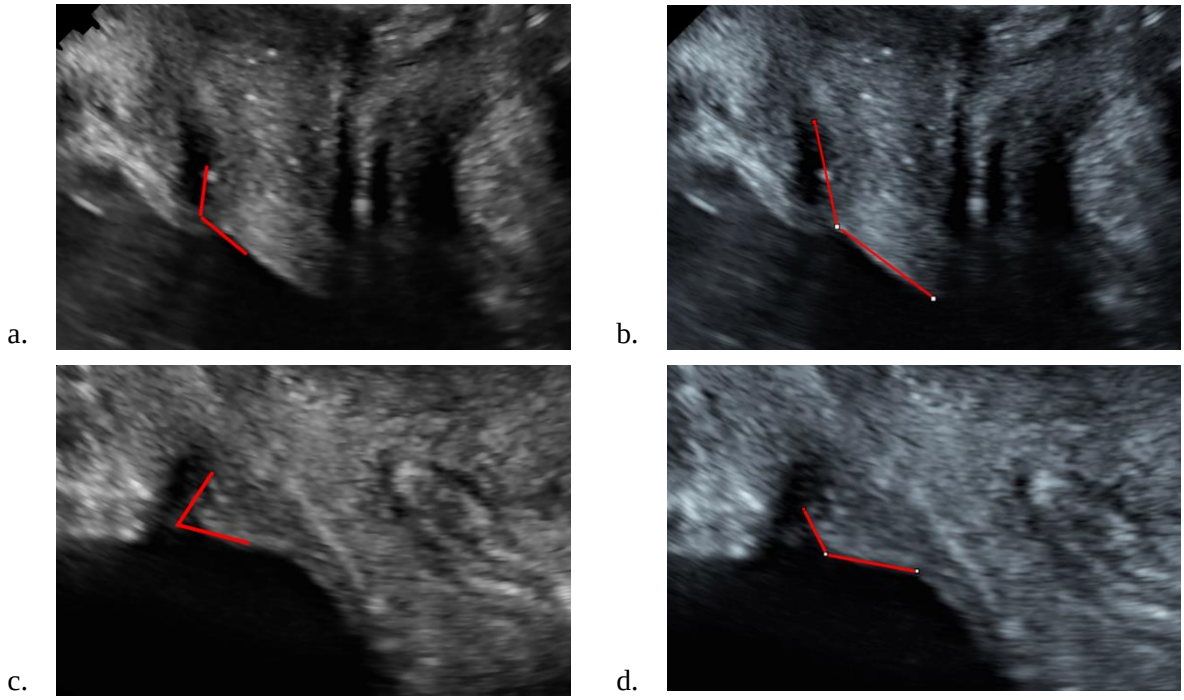


Figure 5.2: a. Semi-automated PUVA for RUN 20 Pre-run 1, b. Rater 1 PUVA for RUN 20 Pre-run 1, c. Semi-automated PUVA for RUN 43 Post-run 1, d. Rater 1 PUVA for RUN 43 Post-run 1.

With the exception of the errors mentioned above, it would be challenging to improve the performance of the semi-automated method on this dataset as there is low agreement between the manual raters. When rater disagreements were removed, the results of the Bland-Altman and linear regression improved drastically (bias: 1.1 degrees, limits of agreement: -18<LOI<20 degrees, r^2: 0.75). The agreement between manual raters appears to have been limited by poor visualization at the bladder neck and proximal urethra. Similarly, the semi-automated method appears to be limited by this same issue. The analysis of the Manually Corrected Dataset with rater disagreements removed provides an approximation of method performance on a dataset with better visualization at the bladder neck and demonstrated that improved agreement between the manual raters could result in improved concurrent validity without further development of the semi-automated method.

In a recent meta-analysis Falah-Hassani et al. reported that PUVA is 19.2 degrees (13.3 – 25.2 degrees) larger in females with SUI than in continent controls [29]. All limits of agreement for PUVA (i.e., between manual raters, and between the raters and the semi-automated method) on this dataset were greater than this difference. This suggests that neither method would be able to reliably differentiate between females with and without SUI using this image set.

5.5 LIMITATIONS AND RECOMMENDATIONS

5.5.1 Limitations

There are several limitations associated with this investigation. Firstly, the dataset was small (n=74). For example, most machine learning approaches benefit from a large training set (e.g., n=1000), and additional new data for testing and validation of the method. It may be the case that using a first-order polynomial to represent the posterior urethral wall limits the accuracy of the method when estimating the PUVA. The urethra, even at its most proximal aspect, is curved [8]. However, when visually examining PUVA estimations from the semi-automated method compared to those estimated by the manual raters, the curve of the proximal urethra did not appear to be a large source of error, even in cases where the manual and semi-automated estimations differed greatly. This is likely because the manual estimation of PUVA must also use a first-order polynomial to represent the most proximal aspect of the urethra.

As noted above, the images were collected with a focus on visualizing the pubic symphysis and the anorectal angle, not the bladder neck. While the bladder neck is close to these structures, it is generally cranial to them, thus the images were not optimized for high resolution in this region. This resulted in a data set wherein the proximal urethra was often poorly visualized, and the posterior bladder wall was often lost in the shadow of unidentified structures. Image quality could have been improved if parameters were set with the intention to visualize the bladder neck and proximal urethra. Indeed, both the manual and semi-automated analysis results were greatly impacted by poor image quality at the bladder neck. Additionally, imaging parameters such as gain, depth, and view angle varied with each participant, resulting in challenges during analysis. Although imaging participants in standing is recommended, and likely resulted in increased observability of BNF and magnitude of PUVA, the motion blur induced by postural sway appeared to result in the exclusion of many participants and made analyzing multiple frames per video too onerous as many frames were too blurry.

5.5.2 Recommendations for Future Work

In future studies, participants should be imaged at rest with a comfortably full bladder as recommended in the literature [2], [73]. While bladder volume is not thought to

impact PUVA magnitude [2], [73], bladder volume should be estimated and recorded. Similar to the methods described by Schaer et al. [65], a curvilinear probe with a large, standardized view angle should be placed at the external urethral meatus. Care should be taken to ensure the hypoechoic representation of the proximal urethra and the posterior bladder wall are clearly visualized. Technicians should take care to minimize probe motion during acquisition. If the intention is to pursue an automated approach without machine learning, gain and depth should be standardized. Then one of two approaches is recommended.

1. 3D static ultrasound images should be collected in braced standing as recommended in the literature [2], [74], while participants pause their breathing. Straining should be avoided to limit motion blur. A total of 3 volumetric images should be taken at each time point (i.e,. before the run and after the run).
2. 2D ultrasound videos should be collected in standing similarly to the recommendation above; participants should pause their breathing and should not strain. The collection of 2D ultrasound videos will allow for the capture of multiple frames in quick succession from which measures can be averaged. Videos need not be more than 5 seconds in length. A total of 3 videos should be taken at each time point (i.e., before the run and after the run).

In both cases, files should be saved with minimal compression at each step of analysis.

Manual estimation of PUVA should be altered so as not to require the middle point of the three-point method to sit at the PUVJ. This will consolidate the manual measurement definition with the semi-automated method. Augmentations could be made to allow for multiple frames from each trial to be analyzed with the same mask, where PUVA estimation could be the average result from the frames, and BNF classification could be based on a voting scheme across the frames. These methods should provide more reliable results, given the ICC evaluations using average-ratings demonstrated better performance. If incorrect PUVJ identification causes continued errors on the optimized dataset, combining TAR with slope signatures may result in better identification of PUVJ in the one blob case, especially in cases where the PUVA is large. If the above augmentations do not result in the desired performance, pursuing a method based on machine learning may be required.

Additional analysis is recommended to evaluate between-frame and between-day reliability of the semi-automated method. Augmentations to the method should be made to pursue a between-frame reliability similar to that of the manual raters on the optimized dataset. Also, evaluation of reliability and concurrent validity of semi-automated PUVA estimation after manually correcting the PUVJ is recommended.

The findings of this study indicate that there is good potential for continued development of a semi-automated or automated approach for the detection of BNF and PUVA for use in future research on the pathophysiology of SUI. Collection of a new image set, optimized to provide high resolution in the region of the bladder neck, is imperative as a first step to test the reliability and validity of the method developed through this thesis.

Chapter 6 Conclusion

The aim of this thesis, to develop and evaluate a semi-automated method to classify BNF and estimate PUVA on 2D transperineal ultrasound images in females, was satisfied. A method of manual analysis was developed and its intra- and inter-rater reliability were evaluated, with almost perfect intra-rater agreement for BNF and good intra-rater reliability for PUVA estimation. The results of this analysis are consistent with reports in the literature [53], [55]. The inter-rater agreement of BNF classification and inter-rater reliability PUVA estimation was found moderate in both cases, which was lower than expected based on the available literature [53], [64], [65]. Limitations in repeatability are thought to be due to poor visualization of the proximal urethra, bladder neck and posterior bladder wall in the dataset.

Semi-automated methods were developed to classify BNF and estimate PUVA from ultrasound images. The intra-rater repeatability of these methods were found to match or exceed intra-rater repeatability of manual analysis on the same dataset and in the literature [53], [55]. The method appears to remove some human error from PUVA estimation.

The concurrent validity of BNF classification using the semi-automated method was excellent given the uncertainty between raters when using visual classification. The concurrent validity of PUVA estimation using the semi-automated approach was lower than anticipated. These results suggest that segmentation and feature extraction errors must be addressed during further development. That said, the concurrent validity of the Manually Corrected Dataset using single measures was comparable to the inter-rater reliability over the same dataset. It is thought that the concurrent validity of PUVA using single ratings could be improved through the acquisition of a new imaging dataset optimized to provide high resolution at the bladder neck and proximal urethra, and by consolidating the method by which PUVA is calculated in the manual and semi-automated methods. Concurrent validity of PUVA using average-ratings was excellent, suggesting that using an average rating from a single mask over multiple frames could improve method performance. Bland-Altman plots suggested that neither the manual nor the semi-automated approach to PUVA estimation is reliable enough to distinguish between females with and females without SUI. This analysis additionally showed that method performance was limited by poor agreement between manual raters which may be corrected by optimizing imaging parameters in a new dataset.

The results presented in this thesis show the potential efficacy of a semi-automated or automated method to analyze BNF and PUVA from transperineal ultrasound images. However, both the dataset and method should be improved before being applied to clinical research. It is recommended to image participants in standing with a full bladder while they are at rest, using an approach similar to that detail by Wlaźlak et al. [56]. The technician should take care to ensure that the hypoechoic representation of the urethra and the posterior bladder wall are clearly visualized. Participants should hold their breath during image acquisition, and probe motion should be minimized during acquisition through supporting the technician's forearm. A dataset collected in this manner, in addition to frame-averaging augmentations, should result in improved method performance.

Remember that BNF and PUVA are features which have been shown, in a recent meta-analysis, to be strongly correlated with SUI in female patients [29] and appear to be related to mechanisms of SUI. Thus, continued development of a semi-automated method to analyze BNF and PUVA on ultrasound images of female patients could be a crucial tool to improve pathophysiological understanding of SUI. Indeed, the method developed through this thesis is an important stepping-stone towards development of methods for SUI diagnosis and treatment-planning that may allow clinicians to isolate specific deficiencies in the continence system. The ability to isolate and treat deficiencies in the continence system will in turn likely result in improved outcomes for patients and reduced spending across healthcare infrastructure.

REFERENCES

[1] J. O. L. Delancey, “Anatomy,” in *Textbook of Female Urology and Urogynecology - Two-Volume Set*, 4th ed., L. Cardozo and D. Staskin, Eds. Boca Raton: CRC Press, 2016, pp. 198–214.

[2] G. N. Schaer, O. R. Koechli, B. Schuessler, and U. Haller, “Perineal ultrasound: Determination of reliable examination procedures,” *Ultrasound Obstet. Gynecol.*, vol. 7, no. 5, pp. 347–352, 1996, doi: 10.1046/j.1469-0705.1996.07050347.x.

[3] F. Pipitone, Z. Sadeghi, and J. O. L. DeLancey, “Urethral function and failure: A review of current knowledge of urethral closure mechanisms, how they vary, and how they are affected by life events,” *Neurourol. Urodyn.*, pp. 1–11, 2021, doi: 10.1002/nau.24760.

[4] J. A. Ashton-Miller and J. O. L. DeLancey, “Functional anatomy of the female pelvic floor,” *Ann. N. Y. Acad. Sci.*, vol. 1101, pp. 266–296, 2007, doi: 10.1196/annals.1389.034.

[5] J. Blaivas, M. B. Chancellor, J. Weiss, and M. Verhaaren, “Normal Micturition,” in *Atlas of Urodynamics*, J. Blaivas, M. B. Chancellor, J. Weiss, and M. Verhaaren, Eds. 2007, pp. 11–21.

[6] K. Brooks, “Female Pelvis Illustration.” Ottawa, 2020.

[7] J. O. L. DeLancey, “Structural support of the urethra as it relates to stress urinary incontinence: The hammock hypothesis,” *Am. J. Obstet. Gynecol.*, vol. 170, no. 5, pp. 1713–1723, 1994, doi: 10.1016/s0002-9378(12)91840-2.

[8] R. F. ZACHARIN, “The Suspensory Mechanism of the Female Urethra.,” *J. Anat.*, vol. 97, pp. 423–7, 1963, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/14047361%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1244203.

[9] S. Tyagi, P. Tyagi, N. Yoshimura, and M. B. Chancellor, “Physiology of Micturition,” in *Textbook of Female Urology and Urogynecology - Two-Volume Set*, 4th ed., L. Cardozo and D. Staskin, Eds. Boca Raton: CRC Press, 2016, pp. 232–246.

[10] R. Patel and V. Nitti, “Bladder outlet obstruction in women: prevalence, recognition, and management.,” *Curr. Urol. Rep.*, vol. 2, no. 5, pp. 379–387, 2001, doi: 10.1007/s11934-996-0024-y.

[11] S. Yande and M. Joshi, “Bladder outlet obstruction in women,” *J. Midlife. Health*, vol. 2, no. 1, pp. 11–17, 2011, doi: 10.4103/0976-7800.83257.

[12] J. O. DeLancey, “The pathophysiology of stress urinary incontinence in women and its implications for surgical treatment,” *World J. Urol.*, vol. 15, no. 5, pp. 268–274, Oct. 1997, doi: 10.1007/BF02202011.

[13] T. R. Coblentz and M. Gray, “Bladder neck obstruction in the female.,” *Urol. Nurs. Off. J. Am. Urol. Assoc. Allied*, vol. 21, no. 4, pp. 265–268, 2001.

[14] Y. Aoki, H. W. Brown, L. Brubaker, J. N. Cornu, J. O. Daly, and R. Cartwright, “Urinary incontinence in women.,” *Nat. Rev. Dis. Prim.*, vol. 3, 2017, doi: 10.1038/nrdp.2017.42.Urinary.

[15] B. T. Haylen *et al.*, “An International Urogynecological Association (IUGA)/International Continence Society (ICS) joint report on the terminology for female pelvic floor dysfunction,” *Neurourol. Urodyn.*, vol. 29, no. 4, pp. 4–26, 2010.

[16] V. Minassian, H. Drutz, and A. Al-Badr, “Urinary incontinence as a worldwide problem,” *Int. J. Gynaecol. Obstet.*, vol. 82, pp. 327–338, 2003, doi: 10.1016/S00207292(03)00220-0.

[17] K. J. Macura and R. R. Genadry, “Female urinary incontinence: Pathophysiology, methods of evaluation and role of MR imaging,” *Abdom. Imaging*, vol. 33, no. 3, pp. 371–380, 2008, doi: 10.1007/s00261-007-9257-6.

[18] B. K. Agarwal and N. Agarwal, “Urinary incontinence: prevalence, risk factors, impact on quality of life and treatment seeking behaviour among middle aged women,” *Int. Surg. J.*, vol. 4, no. 6, pp. 1953–1958, May 2017, doi: 10.18203/2349-2902.isj20172131.

[19] A. H. MacLennan, A. W. Taylor, D. H. Wilson, and D. Wilson, “The prevalence of pelvic floor disorders and their relationship to gender, age, parity, and mode of

delivery," *Br. J. Obstet. Gynaecol.*, vol. 107, pp. 1460–1470, 2000.

[20] Y. S. Hannestad, G. Rortveit, H. Sandvik, and S. Hunskaar, "A community-based epidemiological survey of female urinary incontinence: the Norwegian EPINCONT study. Epidemiology of Incontinence in the County of Nord-Trondelag.," *J. Clin. Epidemiol.*, vol. 53, no. 11, pp. 1150–1157, 2000.

[21] B. S. Buckley and M. C. M. Lapitan, "Prevalence of urinary incontinence in men, women, and children-current evidence: Findings of the fourth international consultation on incontinence," *Urology*, vol. 76, no. 2, pp. 265–270, 2010, doi: 10.1016/j.urology.2009.11.078.

[22] J. Perera, D. S. Kirthinanda, S. Wijeratne, and T. K. Wickramarachchi, "Descriptive cross sectional study on prevalence, perceptions, predisposing factors and health seeking behaviour of women with stress urinary incontinence," *BMC Womens. Health*, vol. 14, no. 1, pp. 1–7, 2014, doi: 10.1186/1472-6874-14-78.

[23] D. H. Thom, S. K. Van Den Eeden, and J. S. Brown, "Evaluation of parturition and other reproductive variables as risk factors for urinary incontinence in later life," *Obstetrics and Gynecology*, vol. 90, no. 6. pp. 983–989, 1997, doi: 10.1016/S0029-7844(97)00537-1.

[24] S. Almousa and A. Bandin van Loon, "The prevalence of urinary incontinence in nulliparous adolescent and middle-aged women and the associated risk factors: A systematic review," *Maturitas*, vol. 107, no. August 2017, pp. 78–83, 2018, doi: 10.1016/j.maturitas.2017.10.003.

[25] V. A. Minassian, E. Devore, K. Hagan, and F. Grodstein, "Severity of urinary incontinence and effect on quality of life in women, by incontinence type," *Obstet. Gynecol.*, vol. 121, no. 5, p. 1083, 2013.

[26] A. Harris, "Impact of urinary incontinence on the quality of life of women.," *Br. J. Nurs.*, vol. 8, no. 6, pp. 375–380, 1999, doi: 10.12968/bjon.1999.8.6.6664.

[27] C. Dumoulin, L. P. Cacciari, and E. J. C. Hay-Smith, "Pelvic floor muscle training versus no treatment, or inactive control treatments, for urinary incontinence in

women," *Cochrane Database Syst. Rev.*, no. 10, p. Art. No.: CD005654, 2018, doi: 10.1002/14651858.CD005654.pub4.

[28] H. E. Richter *et al.*, "Retropubic versus Transobturator Midurethral Slings for Stress Incontinence," *N. Engl. J. Med.*, vol. 362, no. 22, pp. 2066–2076, 2010, doi: 10.1056/nejmoa0912658.

[29] K. Falah-Hassani, J. Reeves, R. Shiri, D. Hickling, and L. McLean, "The pathophysiology of stress urinary incontinence: a systematic review and meta-analysis," *Int. Urogynecol. J.*, vol. 32, no. 3, pp. 501–552, 2021, doi: 10.1007/s00192-020-04622-9.

[30] E. Costantini and M. J. Drake, "Magnetic Resonance Imaging in Urogynecology," in *Textbook of Female Urology and Urogynecology*, 4th ed., L. Cardozo and S. David, Eds. Boca Raton: CRC Press, 2016, pp. 414–421.

[31] F. Sendag *et al.*, "Role of perineal sonography in the evaluation of patients with stress urinary incontinence," *Aust. New Zeal. J. Obstet. Gynaecol.*, vol. 43, no. 1, pp. 54–57, 2003, doi: 10.1046/j.0004-8666.2003.00012.x.

[32] J. R. Fielding, D. J. Griffiths, E. Versi, R. V. Mulkern, M. L. T. Lee, and F. A. Jolesz, "MR imaging of pelvic floor continence mechanisms in the supine and sitting positions," *Am. J. Roentgenol.*, vol. 171, no. 6, pp. 1607–1610, 1998, doi: 10.2214/ajr.171.6.9843296.

[33] X. Zhang *et al.*, "Application of perineal ultrasound measurement and urodynamic study in the diagnosis and typing of stress urinary incontinence ultrasound and urodynamic study," *Urologia*, vol. 80, no. 3, pp. 233–238, 2013, doi: 10.5301/urologia.5000027.

[34] Y. Q. Li, J. Geng, C. Tan, J. Tang, and X. Yang, "Diagnosis and classification of female stress urinary incontinence by transperineal two-dimensional ultrasound," *Technol. Heal. Care*, vol. 25, no. 5, pp. 859–866, 2017, doi: 10.3233/THC-160786.

[35] W. I. Al-Saadi, "Transperineal ultrasonography in stress urinary incontinence: The significance of urethral rotation angles," *Arab J. Urol.*, vol. 14, no. 1, pp. 66–71,

2016, doi: 10.1016/j.aju.2015.11.003.

[36] B. Zhao, L. Wen, W. Chen, Z. Qing, D. Liu, and M. Liu, “A Preliminary Study on Quantitative Quality Measurements of the Urethral Rhabdosphincter Muscle by Supersonic Shear Wave Imaging in Women With Stress Urinary Incontinence,” *J. Ultrasound Med.*, vol. 39, no. 8, pp. 1615–1621, 2020, doi: 10.1002/jum.15255.

[37] H. Chang, S. Chang, H. C. Kuo, and T.-C. Tsai, “Transrectal sonographic cystourethrography: Studies in stress urinary incontinence,” *Urology*, vol. 36, no. 6, pp. 488–492, 1990, doi: 10.1016/0090-4295(90)80183-N.

[38] V. Khullar *et al.*, “Imaging, neurophysiological testing and other tests,” in *6th International Consultation on Incontinence*, pp. 671–804.

[39] A. Tubaro, K. Kluivers, F. Puccini, and A. Carbone, “Imaging of the Upper and Lower Urinary Tract (Radiology and Ultrasound),” in *Textbook of Female Urology and Urogynecology*, 4th ed., L. Cardozo and S. David, Eds. Boca Raton: CRC Press, 2016, pp. 392–413.

[40] I. Giarenis and M. J. Drake, “Special Urodynamic Tests : Videourodynamics,” in *Textbook of Female Urology and Urogynecology*, 4th ed., L. Cardozo and S. David, Eds. Boca Raton: CRC Press, 2016, pp. 355–362.

[41] Merriam-Webster, “Cystourethrogram,” *Merriam-Webster.com Medical Dictionary*. https://www.merriam-webster.com/medical/cystourethrogram.

[42] G. Singh, M. Lucas, L. Dolan, S. Knight, C. Ramage, and P. T. Hobson, “Minimum Standards for Urodynamic Practice in the UK,” *Neurourol. Urodyn.*, vol. 29, pp. 1365–1372, 2010, doi: 10.1002/nau.20883.

[43] F. Gaillard and A. Haouimi, “Chemical shift artifact.,” *Radiopaedia.org*, 2021. .

[44] P. Rahmanou, C. Chaliha, and V. Khullar, “Role of imaging in urogynaecology,” *BJOG An Int. J. Obstet. Gynaecol.*, vol. 111, no. SUPPL. 1, pp. 24–32, 2004, doi: 10.1111/j.1471-0528.2004.00462.x.

[45] M. Bogusiewicz, “Ultrasound imaging in urogynecology - State of the art 2016,” *Prz.*

Menopauzalny, vol. 15, no. 3, pp. 123–132, 2016, doi: 10.5114/pm.2016.63060.

[46] B. G. WISE, G. BURTON, A. CUTNER, and L. D. CARDOZO, "Effect of Vaginal Ultrasound Probe on Lower Urinary Tract Function," *Br. J. Urol.*, vol. 72, no. 2, pp. 12–16, 1993, doi: 10.1111/j.1464-410X.1992.tb16135.x.

[47] J. O. L. DeLancey *et al.*, "Stress Urinary Incontinence: Relative Importance of Urethral Support and Urethral Closure Pressure," *J. Urol.*, vol. 179, no. 6, pp. 2286–2290, Jun. 2008, doi: 10.1016/j.juro.2008.01.098.

[48] H. Kolumban and Y. Wang, "Pipe Flows," in *Fluid and Thermodynamics: Volume 1: Basic Fluid Mechanics.*, Springer International Publishing, 2016, p. 582.

[49] R. F. ZACHARIN, "The Anatomic Supports of the Female Urethra," *Obstet. Gynecol.*, vol. 32, no. 6, pp. 754–759, 1968.

[50] E. Versi, L. D. Cardozo, J. W. W. Studd, M. Brincat, T. M. O'dowd, and D. J. Cooper, "Internal urinary sphincter in maintenance of female continence," *Br. Med. J. (Clin. Res. Ed).*, vol. 292, no. 6514, pp. 166–167, 1986, doi: 10.1136/bmj.292.6514.166.

[51] E. Versi, L. Cardozo, J. Studd, E. McGuire, E. Versi, and L. D. Cardozo, "Distal urethral compensatory mechanisms in women with an incompetent bladder neck who remain continent, and the effect of the menopause," *Neurourol. Urodyn.*, vol. 9, no. 6, pp. 579–590, 1990, doi: 10.1002/nau.1930090603.

[52] R. Tunn, K. Goldammer, A. Gauruder-Burmester, B. Wildt, and D. Beyersdorff, "Pathogenesis of urethral funneling in women with stress urinary incontinence assessed by introital ultrasound," *Ultrasound Obstet. Gynecol.*, vol. 26, no. 3, pp. 287–292, 2005, doi: 10.1002/uog.1977.

[53] J. Cassadó *et al.*, "Introital ultrasonography: A comparison of women with stress incontinence due to urethral hypermobility and continent women," *BJU Int.*, vol. 98, no. 4, pp. 822–828, 2006, doi: 10.1111/j.1464-410X.2006.06404.x.

[54] K. J. Macura, "Magnetic resonance imaging in assessment of stress urinary incontinence in women: Parameters differentiating urethral hypermobility and

intrinsic sphincter deficiency," *World J. Radiol.*, vol. 7, no. 11, p. 394, 2015, doi: 10.4329/wjr.v7.i11.394.

[55] S. Pontbriand-Drolet *et al.*, "Differences in Pelvic Floor Morphology Between Continent Stress Urinary Incontient, and Mixed Urinary Incontinent Elderly Women: An MRI Study," *Neurourol. Urodyn.*, vol. 35, pp. 515–521, 2016, doi: 10.1002/nau.22743.

[56] E. Wlaźlak, T. Kluz, G. Surkont, and J. Kociszewski, "Urethral funneling visualized during pelvic floor sonography - analysis of occurrence among urogynecological patients," *Ginekol. Pol.*, vol. 89, no. 2, pp. 55–61, 2018, doi: 10.5603/GP.a2018.00010.

[57] G. N. Schaer, D. Perucchini, E. Munz, U. Peschers, O. R. Koechli, and J. O. L. DeLancey, "Sonographic evaluation of the bladder neck in continent and stress-incontinent women," *Obstet. Gynecol.*, vol. 93, no. 3, pp. 412–416, 1999, doi: 10.1016/S0029-7844(98)00420-7.

[58] W. C. Huang and J. M. Yang, "Bladder neck funneling on ultrasound cystourethrography in primary stress urinary incontinence: a sign associated with urethral hypermobility and intrinsic sphincter deficiency," *Urology*, vol. 61, no. 5, pp. 936–941, 2003, doi: 10.1016/S0090-4295(02)02558-X.

[59] A. C. Santiago, L. H. Quiroz, and S. A. Shobeiri, "Decreased Urethral Volume Is Comparable to Funneling as a Predictor of Intrinsic Sphincter Deficiency," *Female Pelvic Med. Reconstr. Surg.*, vol. 26, no. 5, pp. 287–298, 2020, doi: 10.1097/SPV.0000000000000386.

[60] M. Li, B. Wang, X. Liu, P. Qiao, W. Jiao, and T. Jiang, "MR defecography in the assessment of anatomic and functional abnormalities in stress urinary incontinence before and after pelvic reconstruction," *Eur. J. Radiol.*, vol. 126, no. January, p. 108935, 2020, doi: 10.1016/j.ejrad.2020.108935.

[61] N. Li, C. Cui, Y. Cheng, Y. Wu, J. Yin, and W. Shen, "Association between magnetic resonance imaging findings of the pelvic floor and de novo stress urinary

incontinence after vaginal delivery," *Korean J. Radiol.*, vol. 19, no. 4, pp. 715–723, 2018, doi: 10.3348/kjr.2018.19.4.715.

[62] J. M. Yang and W. C. Huang, "Discrimination of bladder disorders in female lower urinary tract symptoms on ultrasonographic cystourethrography," *J. Ultrasound Med.*, vol. 21, no. 11, pp. 1249–1255, 2002, doi: 10.7863/jum.2002.21.11.1249.

[63] K. M. Rinne, S. Kainulainen, S. Aukee, S. Heinonen, and C. G. Nilsson, "Dynamic magnetic resonance imaging of the behavior of the mid-urethra in healthy and stress incontinent women," *Acta Obstet. Gynecol. Scand.*, vol. 89, no. 3, pp. 373–379, 2010, doi: 10.3109/00016340903555982.

[64] L. Wen, B. Zhao, W. Chen, Z. Qing, and M. Liu, "Real-time assessment of the behaviour of the bladder neck and proximal urethra during urine leaking in the cough stress test (CST) in supine and standing positions using transperineal ultrasound," *Int. Urogynecol. J.*, 2020, doi: 10.1007/s00192-020-04273-w.

[65] G. N. Schaer, O. R. Koechli, B. Schuessler, and U. HALLER, "Perineal Ultrasound for Evaluating the Bladder Neck in Urinary Stress Incontinence," *Obstet. Gynecol.*, vol. 85, no. 2, pp. 220–224, 1995, doi: 10.1016/0029-7844(94)00369-O.

[66] S. F. English, C. L. Amundsen, and E. J. McGuire, "Bladder neck competency at rest in women with incontinence," *J. Urol.*, vol. 161, no. 2, pp. 578–580, 1999, doi: 10.1016/S0022-5347(01)61956-6.

[67] H. Y. Chen, Y. L. Huang, Y. C. Hung, and W. C. Chen, "Evaluation of stress urinary incontinence by computer-aided vector-based perineal ultrasound," *Acta Obstet. Gynecol. Scand.*, vol. 85, no. 10, pp. 1259–1264, 2006, doi: 10.1080/00016340600880951.

[68] H. P. Dietz and B. Clarke, "The urethral pressure profile and ultrasound imaging of the lower urinary tract," *Int. Urogynecol. J. Pelvic Floor Dysfunct.*, vol. 12, no. 1, pp. 38–41, 2001, doi: 10.1007/s001920170092.

[69] S. Hajebrahimi, A. Azaripour, and H. Sadeghi-Bazargani, "Clinical and Transperineal Ultrasound Findings in Females with Stress Urinary Incontinence Versus Normal

Controls," *Pakistan J. Biol. Sci.*, vol. 12, no. 21, pp. 1434–1437, 2009, doi: 10.3923/pjbs.2009.1434.1437.

[70] G. M. Ghoniem, A. N. Elgamasy, R. Elsergany, and D. S. Kapoor, "Grades of intrinsic sphincteric deficiency (ISD) associated with female stress urinary incontinence," *Int. Urogynecol. J.*, vol. 13, no. 2, pp. 99–105, 2002, doi: 10.1007/s001920200023.

[71] A. H. Kegel, "Early genital relaxation: New technic of diagnosis and nonsurgical treatment," *Obstetrics and Gynecology*, vol. 8, no. 5. pp. 545–550, 1956.

[72] S. A. Jung *et al.*, "Vaginal high-pressure zone assessed by dynamic 3-dimensional ultrasound images of the pelvic floor," *Am. J. Obstet. Gynecol.*, vol. 197, no. 1, pp. 52.e1-52.e7, 2007, doi: 10.1016/j.ajog.2007.04.026.

[73] H. P. Dietz and P. D. Wilson, "The influence of bladder volume on the position and mobility of the urethrovesical junction," *Int. Urogynecol. J.*, vol. 10, no. 1, pp. 3–6, 1999, doi: 10.1007/PL00004011.

[74] H. P. Dietz and B. Clarke, "The influence of posture on perineal ultrasound imaging parameters," *Int. Urogynecol. J.*, vol. 12, no. 2, pp. 104–106, 2001, doi: 10.1007/PL00004030.

[75] J. K. Kim, Y. J. Kim, M. S. Choo, and K. S. Cho, "The urethra and its supporting structures in women with stress urinary incontinence: MR imaging using an endovaginal coil," *Am. J. Roentgenol.*, vol. 180, no. 4, pp. 1037–1044, 2003, doi: 10.2214/ajr.180.4.1801037.

[76] H. S. Virtanen and P. Kiilholma, "Urogynecologic ultrasound is a useful aid in the assessment of female stress urinary incontinence - A prospective study with TVT procedure," *Int. Urogynecol. J.*, vol. 13, no. 4, pp. 218–223, 2002, doi: 10.1007/s001920200048.

[77] Z. Akkus, B. H. Kim, R. Nayak, A. Gregory, A. Alizad, and M. Fatemi, "Fully automated segmentation of bladder sac and measurement of detrusor wall thickness from transabdominal ultrasound images," *Sensors (Switzerland)*, vol. 20, no. 15, pp.

1–11, 2020, doi: 10.3390/s20154175.

[78] B. Padmapriya, T. Kesavamurthi, and H. W. Ferose, “Edge based image segmentation technique for detection and estimation of the bladder wall thickness,” *Procedia Eng.*, vol. 30, no. 2011, pp. 828–835, 2012, doi: 10.1016/j.proeng.2012.01.934.

[79] M. Matsumoto *et al.*, “Development and evaluation of automated ultrasonographic detection of bladder diameter for estimation of bladder urine volume,” *PLoS One*, vol. 14, no. 9, pp. 1–9, 2019, doi: 10.1371/journal.pone.0219916.

[80] V. Muneeswaran and M. P. Rajasekaran, “Automatic segmentation of gallbladder using bio-inspired algorithm based on a spider web construction model,” *J. Supercomput.*, vol. 75, no. 6, pp. 3158–3183, 2019, doi: 10.1007/s11227-017-2230-4.

[81] M. Ciecholewski, “Gallbladder Boundary Segmentation in 2-D Ultrasound Images Using Deformable Contour Methods,” in *Modeling Decisions for Artificial Intelligence: 7th International Conference*, 2010, pp. 163–174.

[82] M. Ciecholewski and J. Chochołowicz, “Gallbladder shape extraction from ultrasound images using active contour models,” *Comput. Biol. Med.*, vol. 43, no. 12, pp. 2238–2255, 2013, doi: 10.1016/j.compbiomed.2013.10.009.

[83] M. Ciecholewski, “Gallbladder Boundary Segmentation from Ultrasound Images Using Active Contour Model,” in *Intelligent Data Engineering and Automated Learning – IDEAL 2010: 11th International Conference*, 2010, pp. 63–69.

[84] T. K. Koo and M. Y. Li, “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research,” *J. Chiropr. Med.*, vol. 15, no. 2, pp. 155–163, 2016, doi: 10.1016/j.jcm.2016.02.012.

[85] The Mathworks Inc., “rgb2gray,” *The Mathworks, Inc.*, 2021. https://www.mathworks.com/help/matlab/ref/rgb2gray.html#buiz8mj-9.

[86] P. Perona and J. Malik, “Scale-Space and Edge Detection Using Anisotropic Diffusion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, 1990, doi: 10.1109/34.56205.

[87] T. F. Chan and L. A. Vese, “Active contours without edges,” *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, 2001, doi: 10.1109/83.902291.

[88] P. Soille, *Morphological Image Analysis: Principles and Applications*, vol. 148. Springer-Verlag, 1999.

[89] N. Alajlan, I. El Rube, M. S. Kamel, and G. Freeman, “Shape retrieval using triangle-area representation and dynamic space warping,” *Pattern Recognit.*, vol. 40, no. 7, pp. 1911–1920, 2007, doi: 10.1016/j.patcog.2006.12.005.

[90] M. L. McHugh, “Interrater reliability: the kappa statistic,” *Biochem. Medica*, vol. 22, no. 3, pp. 276–282, 2012.

Appendix A Frames Removed in Filtered Dataset Due to Poor Segmentation

Figure A.1: Participant 2 pre-run trial 3

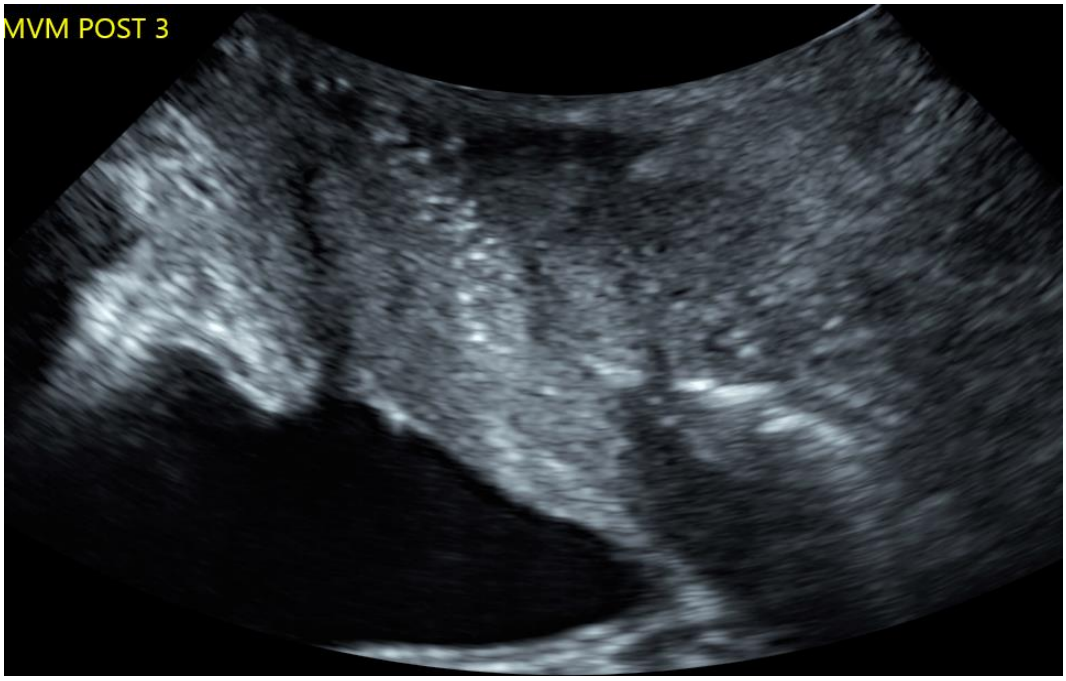


Figure A.2: Participant 3 post-run trial 3

Figure A.3: Participant 4 post-run trial 3

Figure A.4: Participant 4 pre-run trial 1

Figure A.5: Participant 4 pre-run trial 2

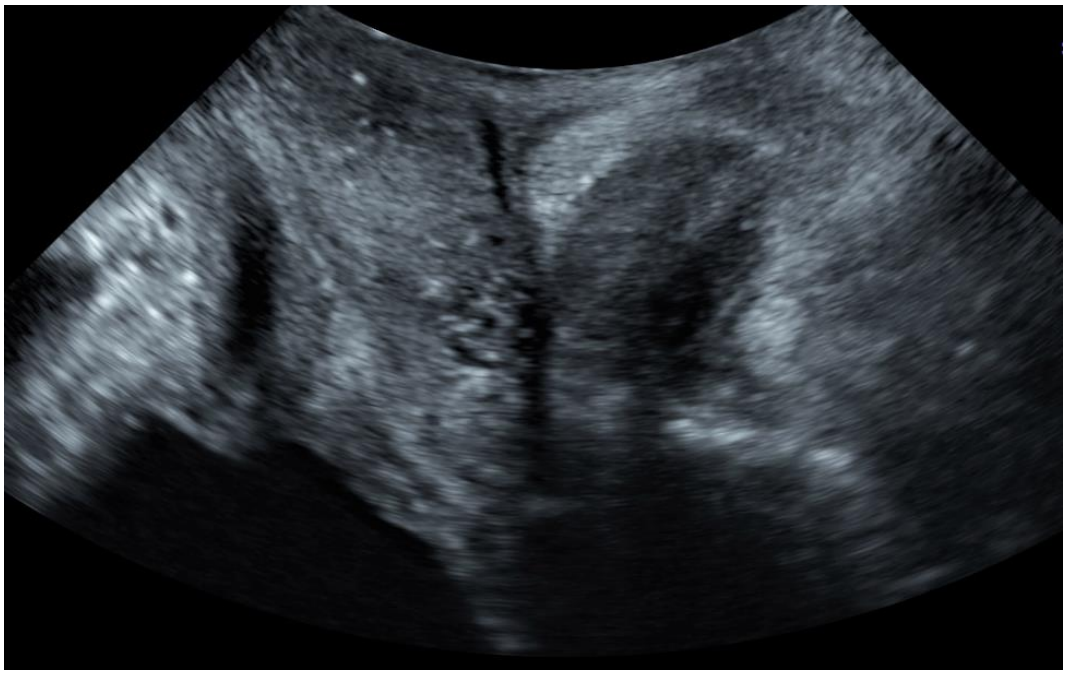


Figure A.6: Participant 11 pre-run trial 1

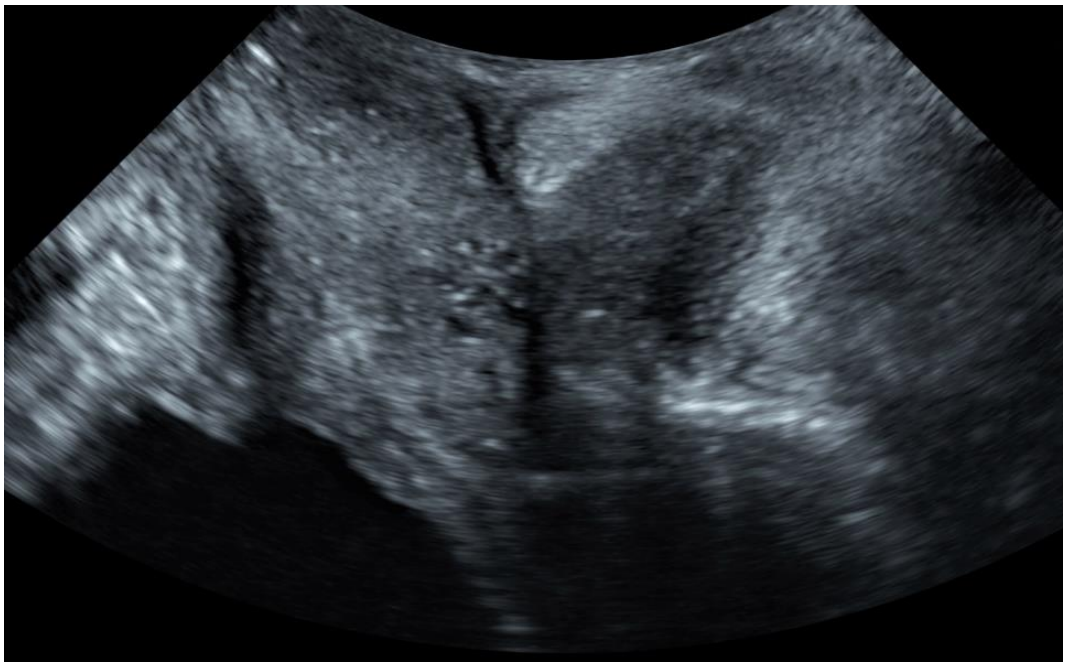


Figure A.7: Participant 11 pre-run trial 2

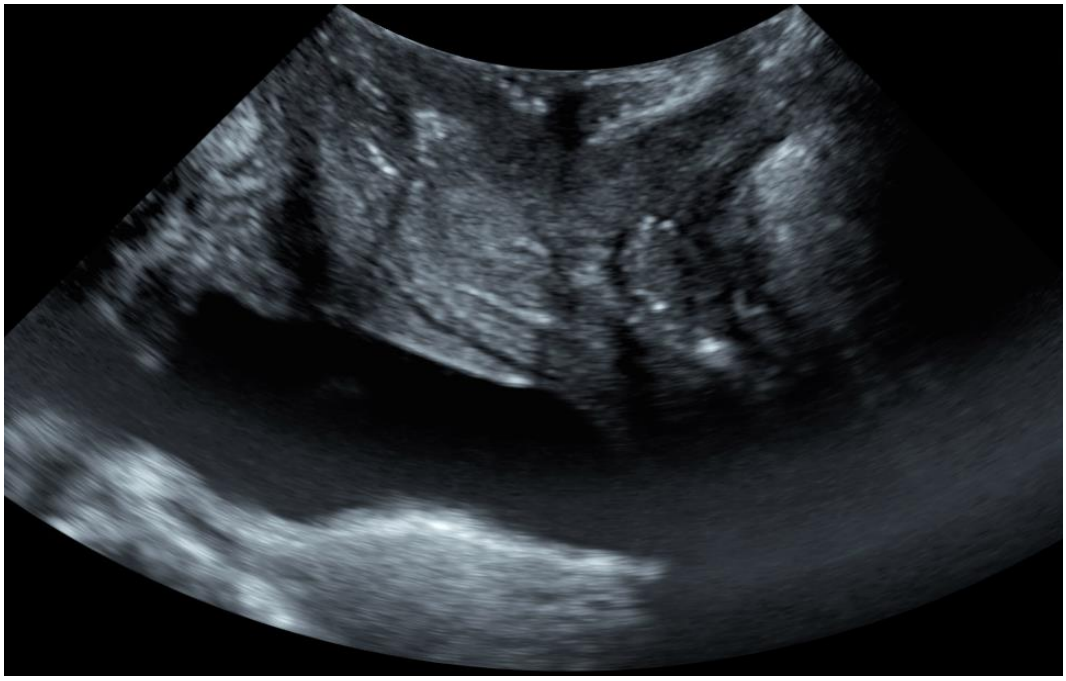


Figure A.8:Participant 13 pre-run trial 1

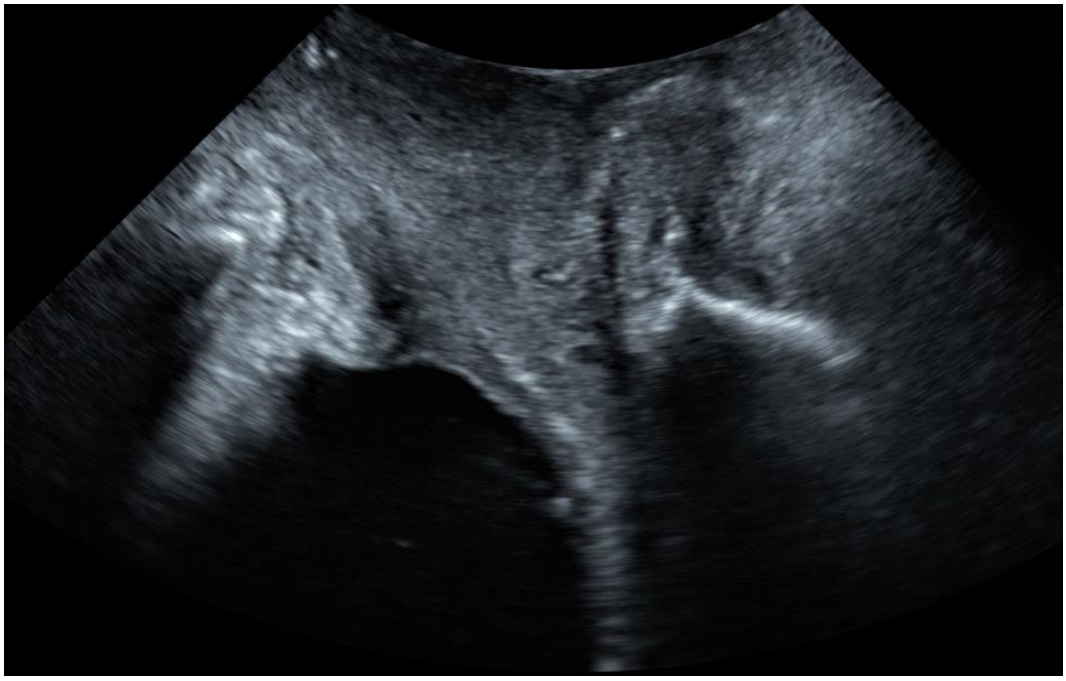


Figure A.9: Participant 16 pre-run trial 1

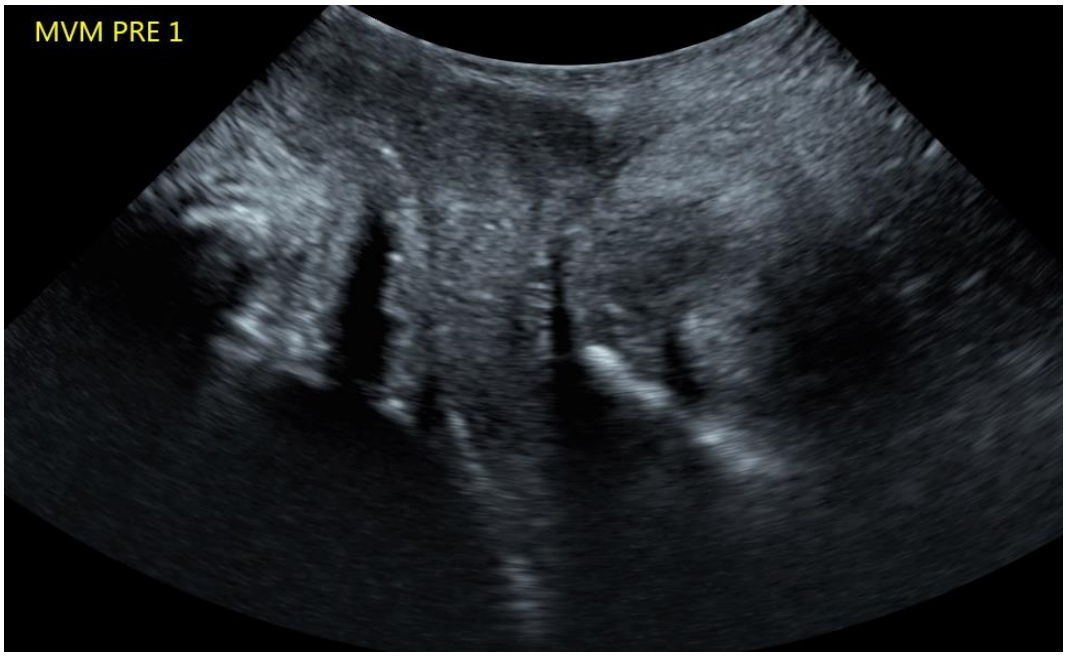


Figure A.10: Participant 19 pre-run trial 1

Figure A.11: Participant 20 post-run trial 1

Figure A.12: Participant 31 pre-run trial 2

Figure A.13: Participant 33 pre-run trial 1

Figure A.14: Participant 34 post-run trial 3

Figure A.15: Participant 35 post-run trial 2

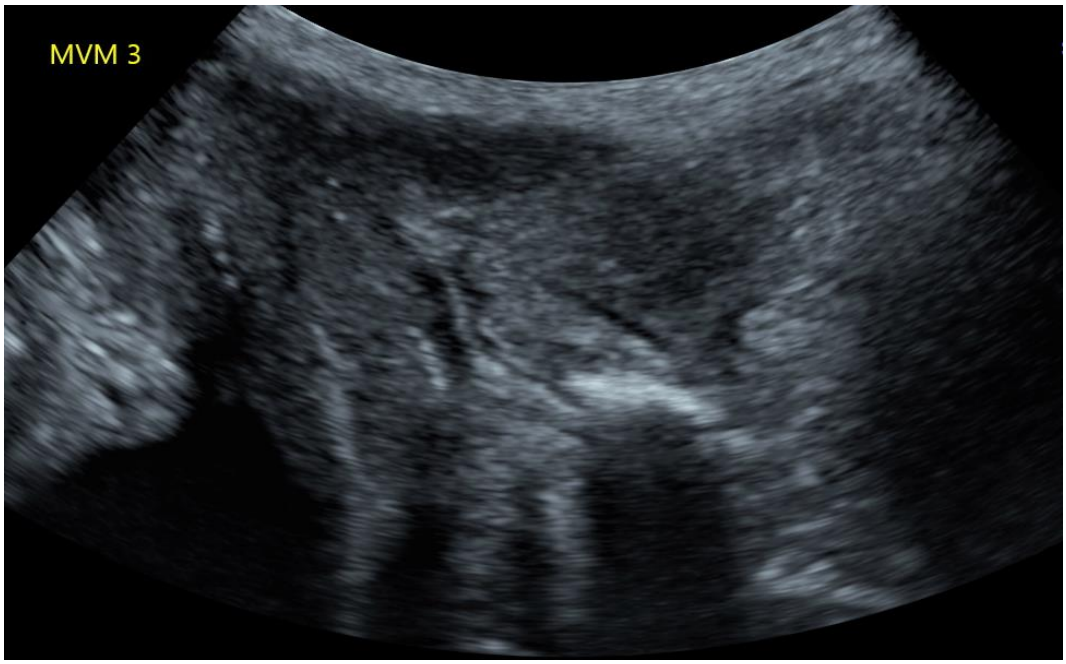


Figure A.16: Participant 37 post-run trial 3

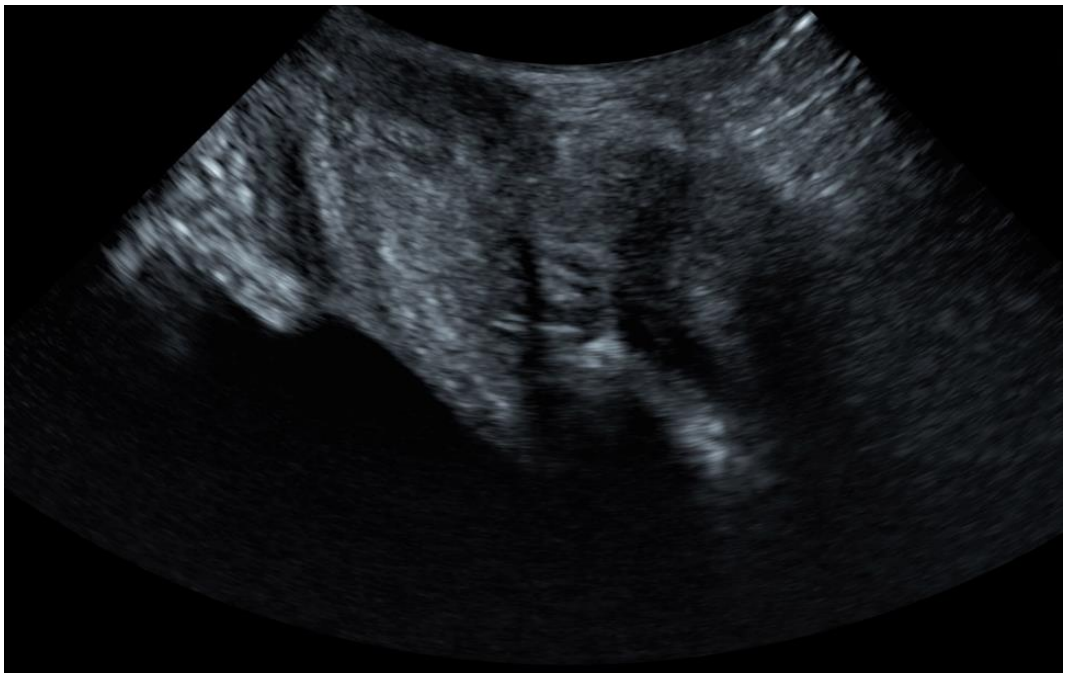


Figure A.17: Participant 42 pre-run trial 1

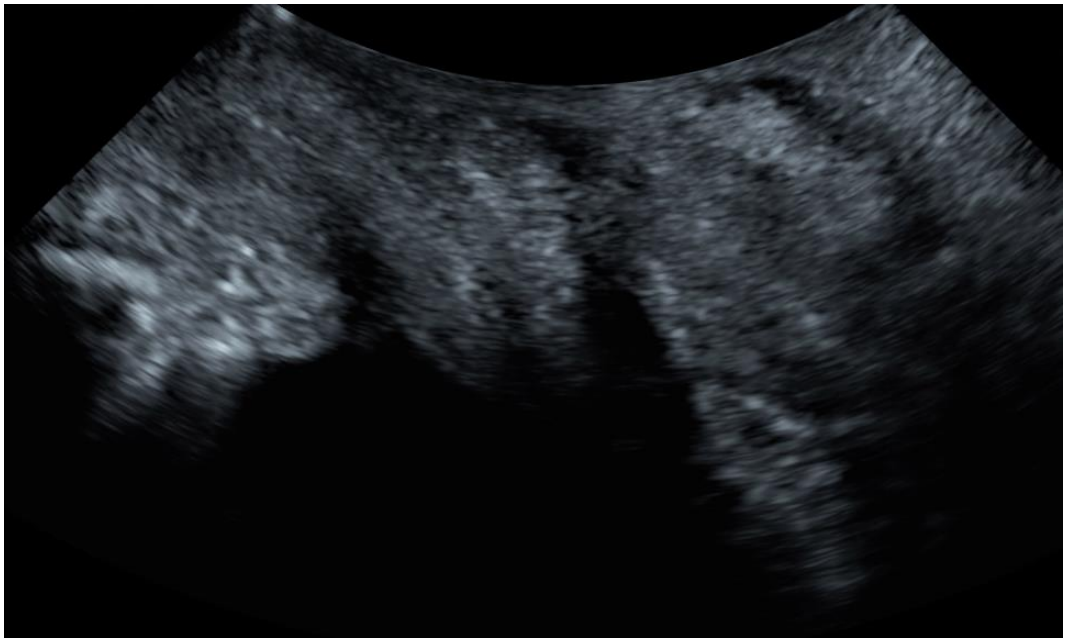


Figure A.18: Participant 46 pre-run trial 2

Figure A.19: Participant 49 pre-run trial 1

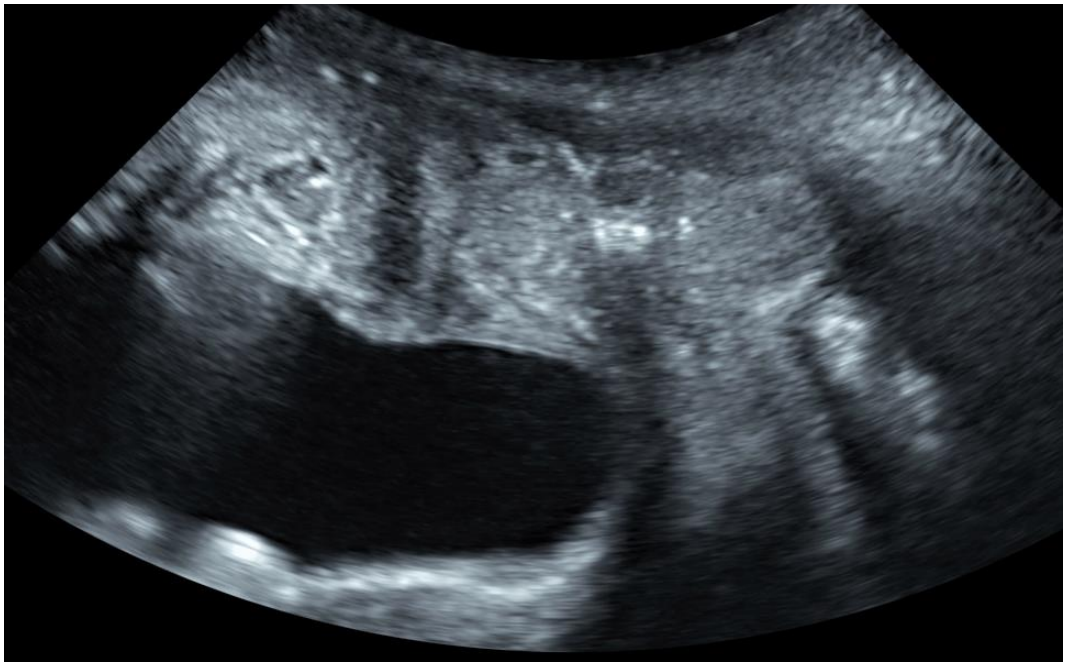


Figure A.20: Participant 51 post-run trial 1

Figure A.21: Participant 56 pre-run trial 3

Figure A.22: Participant 59 pre-run trial 2

Appendix B Manual Correction of Poor Segmentations

Figure B.1: Manually corrected participant 2 pre-run trial 3 (corrected in dark grey)

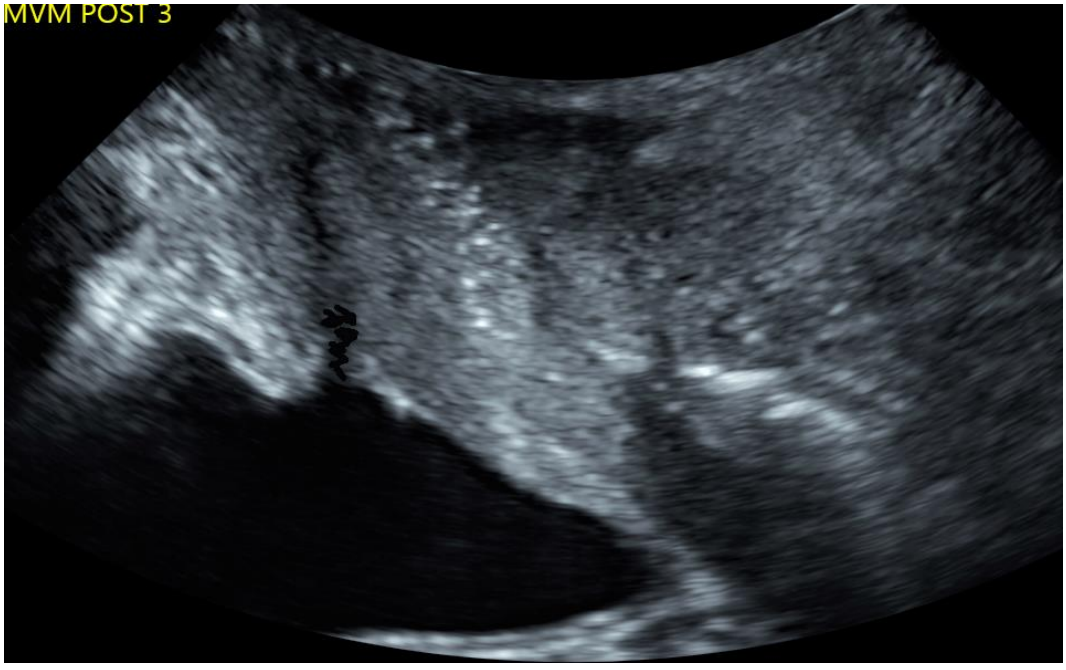


Figure B.2: Manually corrected participant 3 post-run trial 3 (corrected in dark grey)

Figure B.3: Manually corrected participant 4 post-run trial 3 (corrected in grey)

Figure B.4: Manually corrected participant 4 pre-run trial 1 (corrected in grey)

Figure B.5: Manually corrected participant 4 pre-run trial 2 (corrected in grey)

Figure B.6: Manually corrected participant 11 pre-run trial 1 (corrected in dark grey)

Figure B.7: Manually corrected participant 11 pre-run trial 2 (corrected in grey)

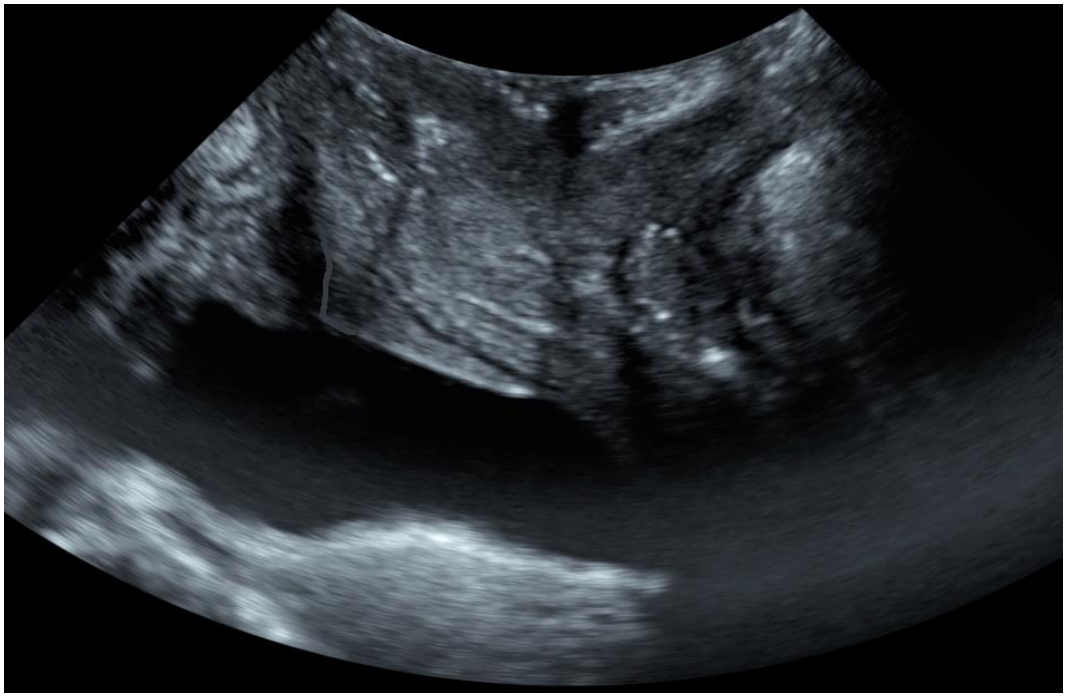


Figure B.8: Manually corrected participant 13 pre-run trial 1 (corrected in grey)

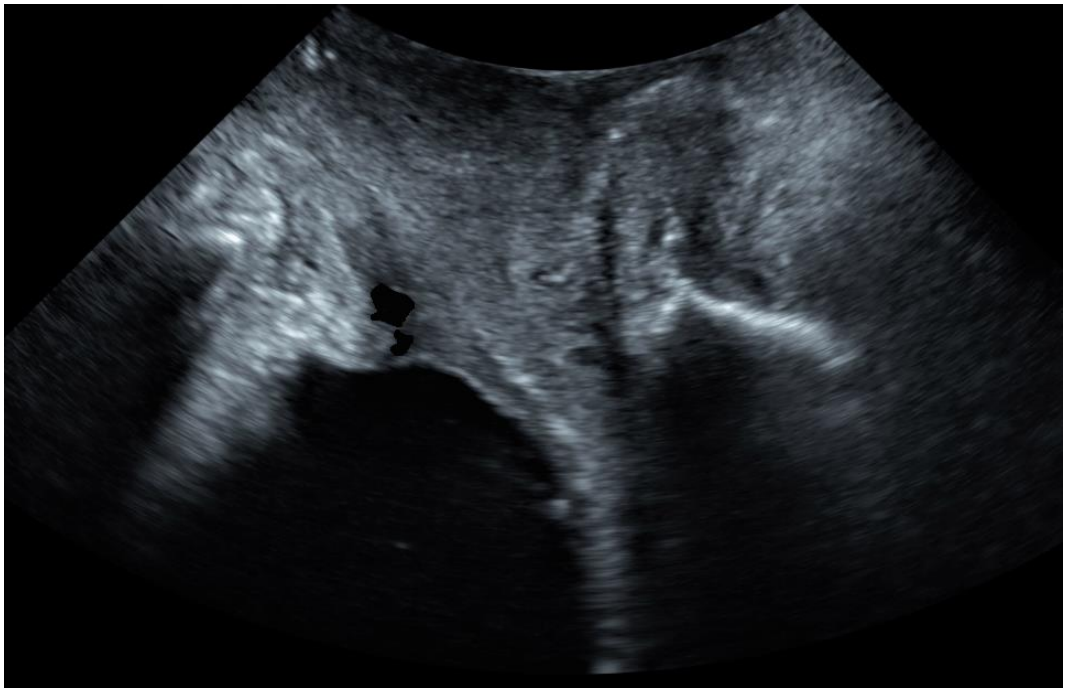


Figure B.9: Manually corrected participant 16 pre-run trial 1 (corrected in dark grey)

Figure B.10: Manually corrected participant 19 pre-run trial 1 (corrected in grey)

Figure B.11: Manually corrected participant 20 post-run trial 1 (corrected in dark grey)

Figure B.12: Manually corrected participant 31 pre-run trial 2 (corrected in grey)

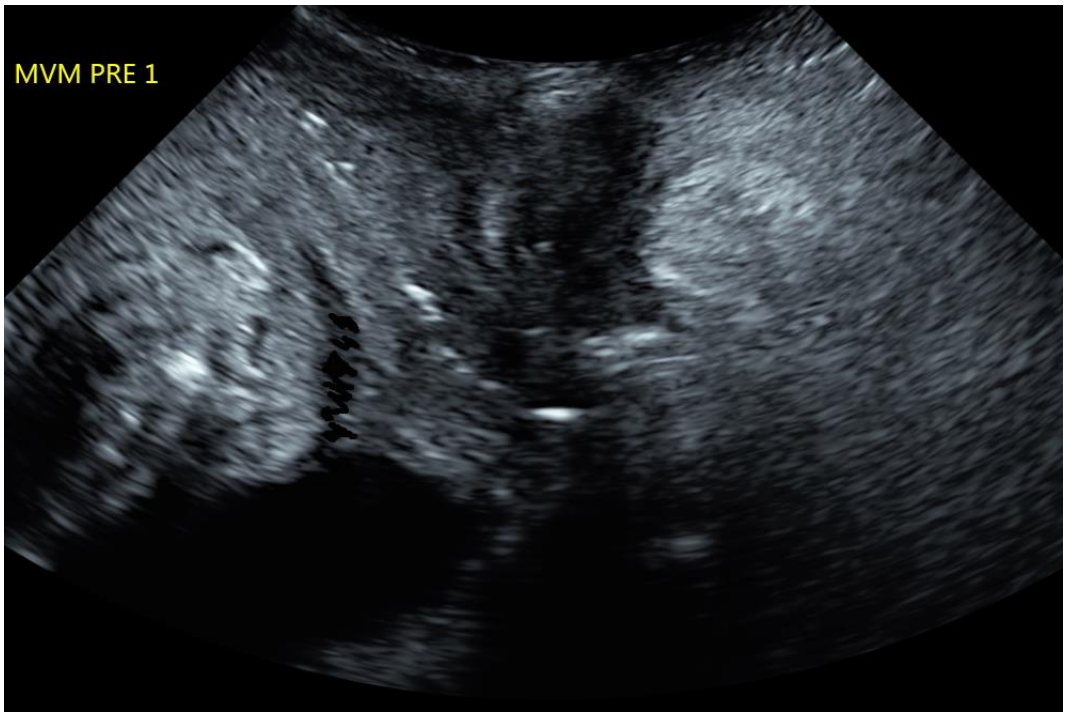


Figure B.13: Manually corrected participant 33 pre-run trial 1 (corrected in dark grey)

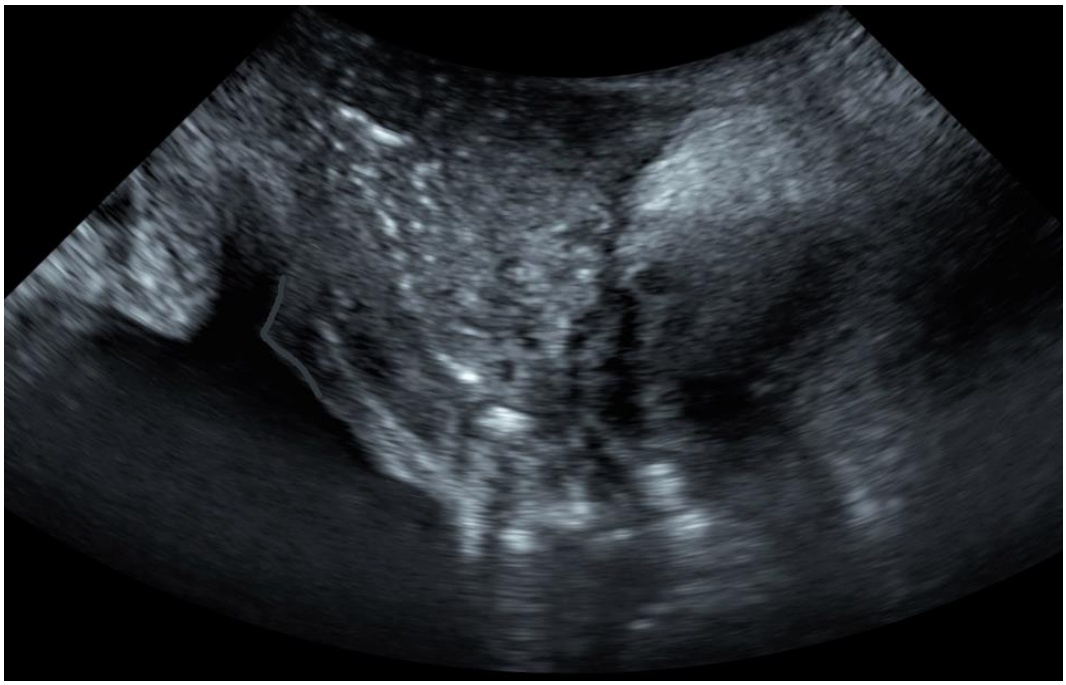


Figure B.14: Manually corrected participant 34 post-run trial 3 (corrected in grey)

Figure B.15: Manually corrected participant 35 post-run trial 2 (corrected in grey)

Figure B.16: Manually corrected participant 37 post-run trial 3 (corrected in grey)

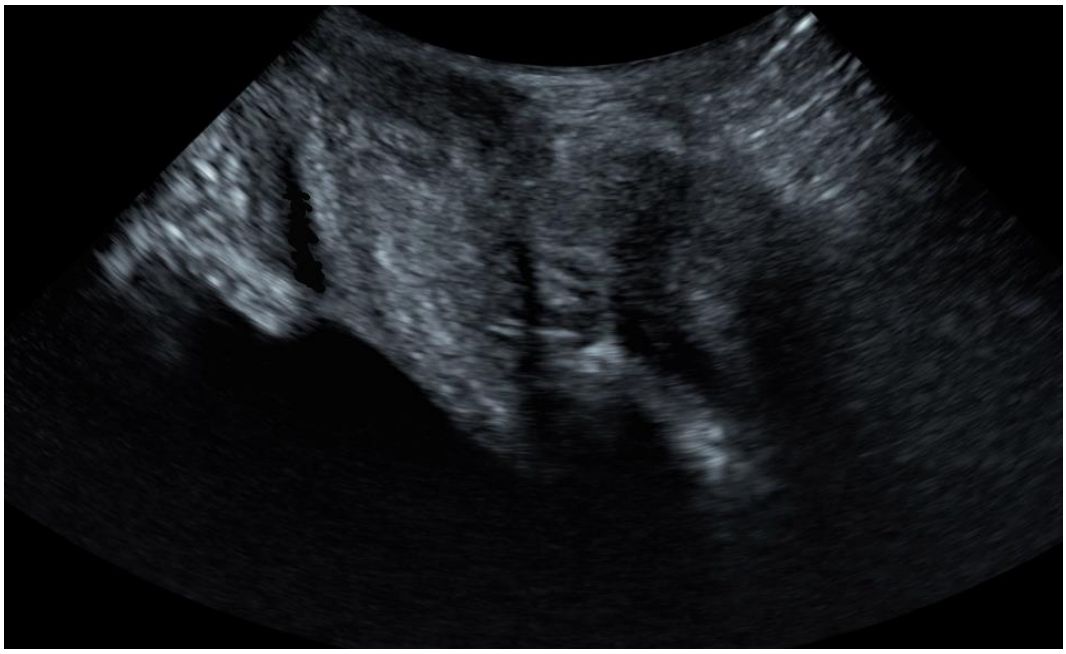


Figure B.17: Manually corrected participant 42 pre-run trial 1 (corrected in dark grey)

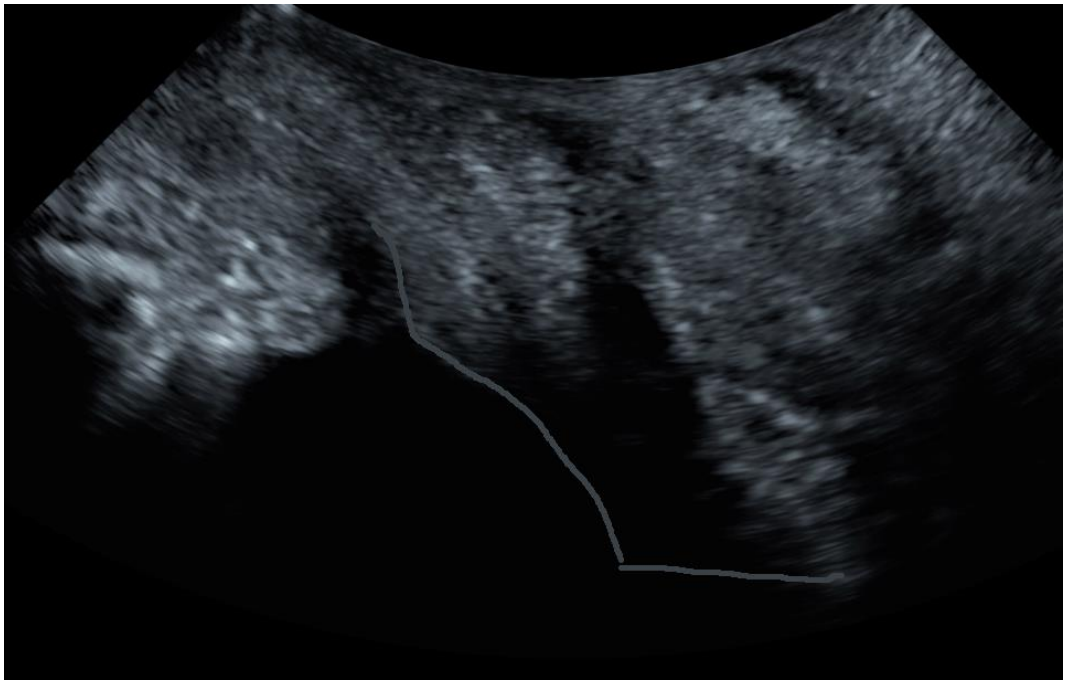


Figure B.18: Manually corrected participant 46 pre-run trial 2 (corrected in grey)

Figure B.19: Manually corrected participant 49 pre-run trial 1 (corrected in grey)

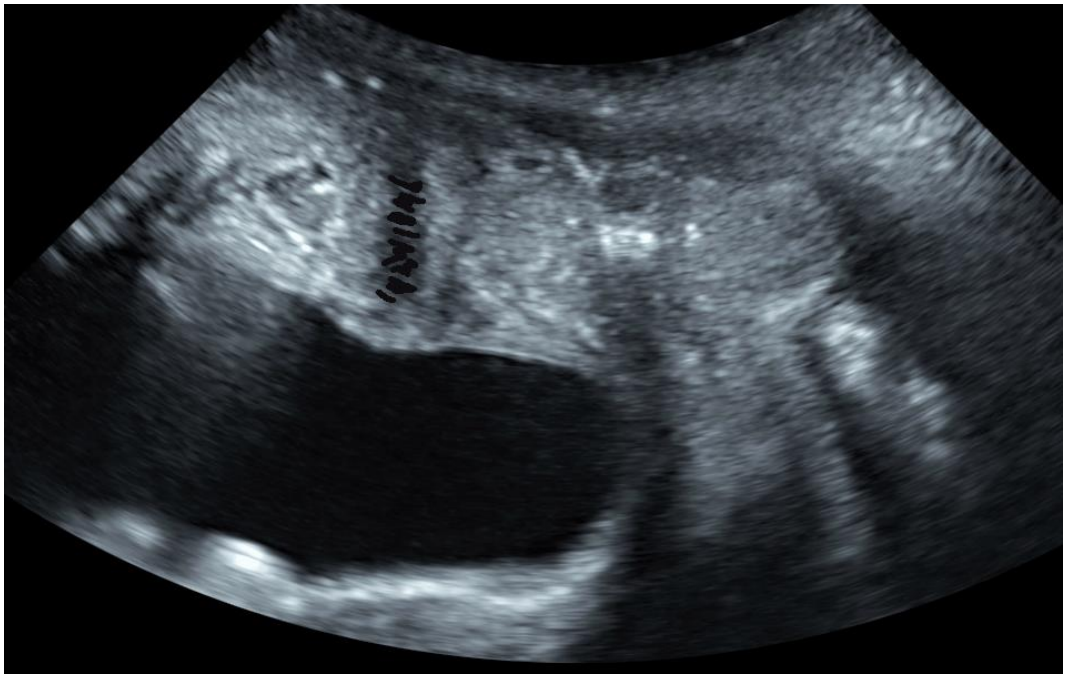


Figure B.20: Manually corrected participant 51 post-run trial 1 (corrected in dark grey)

Figure B.21: Manually corrected participant 56 pre-run trial 3 (corrected in grey)

Figure B.22: Manually correct participant 59 pre-run trial 2 (corrected in dark grey)

Appendix C BNF Classification Rating and Confidence Matrices for Individual Raters

Table C.1: Rater 1 trial 1 rating and confidence matrix

	Positive	Negative
Confident	19	29
Unconfident	13	13

Table C.2: Rater 1 trial 2 rating and confidence matrix

	Positive	Negative
Confident	20	27
Unconfident	12	15

Table C.3: Rater 2 trial 1 rating and confidence matrix

	Positive	Negative
Confident	25	29
Unconfident	13	7

Table C.4: Rater 2 trial 2 rating and confidence matrix

	Positive	Negative
Confident	39	15
Unconfident	9	11

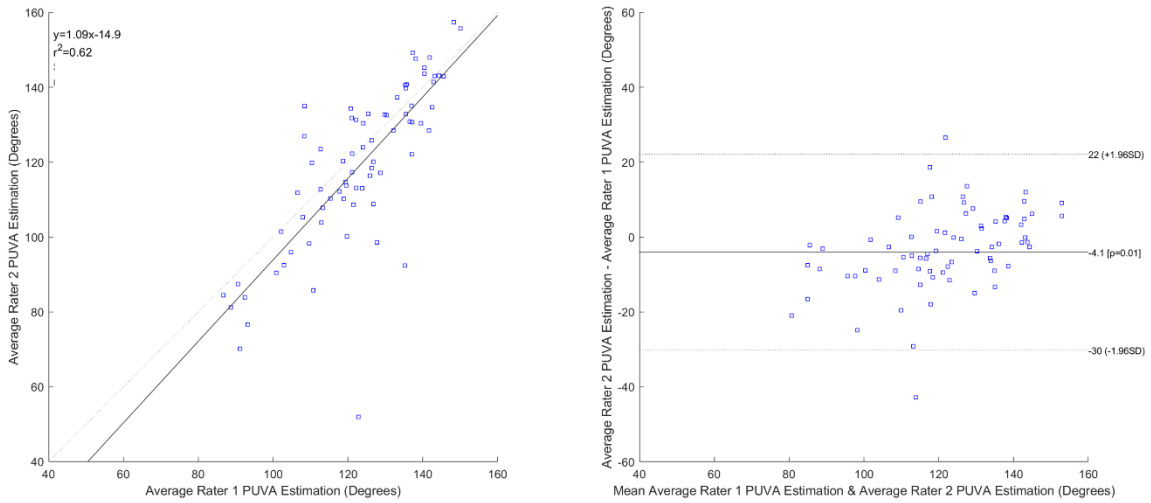


Figure C.1: Bland-Altman Plot of Rater 1 vs Rater 2 Estimation of PUVA (where SD stands for standard deviation)

*Note: Difference between Rater 1 and Rater 2 PUVA estimation was not normally distributed