Automated low-cost ultrasound:

improving antenatal care in resource-limited settings

Thomas van den Heuvel

This book was typeset by the author using LaTeX2$_\varepsilon$.

Cover design by Promotie In Zicht, Arnhem.

ISBN: 978-94-92896-98-8

Printed by Ipskamp Printing, Enschede.

Automated low-cost ultrasound:
improving antenatal care in resource-limited settings

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 9 januari 2019
om 14.30 uur precies

door

Thomas Leon Adrian van den Heuvel

geboren op 15 mei 1990
te Roermond

Promotoren
Prof. dr. ir. B. van Ginneken
Prof. dr. ir. C.L. de Korte

Manuscriptcommissie
Prof. dr. F.P.H.A. Vandenbussche
Prof. dr. W.J. Niessen *(Technische Universiteit Delft)*
Dr. ir. J.G. Bosch *(Erasmus Medisch Centrum Rotterdam)*

The research described in this thesis was carried out at the Diagnostic Image Analysis Group and the Medical Ultrasound Imaging Center, Radboud University Medical Center (Nijmegen, the Netherlands).
The work presented in this thesis was carried out within the Radboud Institute for Health Sciences.
This work was partially funded by the Life Sciences & Health for Development Fund, Project No. LSH14ET04.

TABLE OF CONTENTS

1

General introduction

1.1 Maternal mortality

Each day more than 800 pregnant women die as a consequence of their pregnancy. 99% of these deaths occur in resource-limited countries[1]. Figure 1.1 shows a map of the world with the number of maternal deaths per 100.000 live births in each country. This figure illustrates that the maternal mortality ratio is highest in resource-limited countries. Ultrasound is commonly used to detect maternal risk factors, since it is a real-time, non-invasive imaging method that does not require ionizing radiation. The WHO has published a report with recommendations to decrease maternal mortality in developing countries[2]. Although not all maternal deaths in resource-limited countries can be avoided by introducing ultrasound imaging, the WHO strongly recommends the use of ultrasound imaging. An ultrasound scan is recommended for accurate gestational age estimation, for detection of multiple pregnancies, fetal malpresentation, fetal anomalies, placenta previa, and polyhydramnios, to confirm fetal viability, and for improvement of a woman's pregnancy experience. The accurate estimation of gestational age is critical for the appropriate delivery of time-sensitive interventions in pregnancy, as well as management of pregnancy complications, particularly pre-eclampsia and preterm birth, which are major causes of maternal and perinatal morbidity and mortality in low- and middle-income countries. In addition, estimation of gestational age could be used to detect intra-uterine growth restriction and induction of labor for post-term pregnancy.

The report of the WHO also states that health-care providers sometimes do not feel suitably trained to provide screening and testing procedures for maternal care. This suggests that they might welcome ultrasound scans. The introduction of ultrasound could therefore plausibly increase antenatal care service utilization and reduce morbidity and mortality when accompanied by appropriate referral and management.

Ultrasound imaging requires a trained sonographer to acquire and interpret the ultrasound images. The WHO mentioned that with appropriate training there can be a potential task shift from trained sonographers and doctors to trained nurses, midwives and clinical officers, provided that ongoing training, staff retention, quality improvement activities and supervision are ensured, and mentoring and referral systems are in place.

Unfortunately, there is a severe shortage of well-trained medical personnel in resource-limited countries[3–5]. Training sonographers requires a significant investment of time and resources, which impedes the introduction of ultrasound in these countries. This keeps ultrasound imaging out of reach for most pregnant women in developing countries.

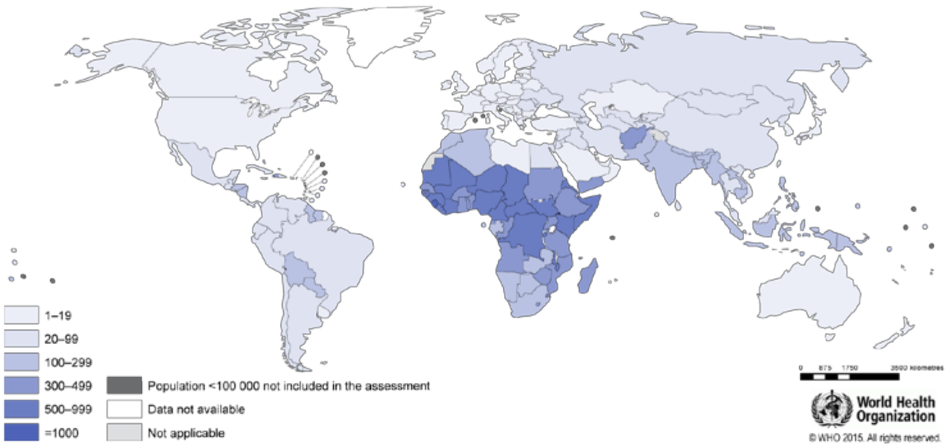

Figure 1.1: Maternal mortality ratio: number of maternal deaths per 100.000 live births (2015).

This thesis describes how low-cost ultrasound in combination with automated image analysis could be used to automatically determine maternal risk factors without the need of a highly trained sonographer. This would decrease the costs and vastly decrease the time that is currently required for training sonographers, which would make the introduction of prenatal ultrasound in resource-limited settings much easier.

1.2 Ultrasound

Medical ultrasound devices make use of ultrasound to create an image of the inside of the human body. The ultrasonic wave produced by the ultrasound transducer cannot be heard with the human ear, because the frequency of the produced sound waves usually ranges from 3 to 12 MHz, while the frequency of the sound that can be picked up with human hearing ranges from 20 Hz to 20 kHz. An ultrasound transducer uses so-called piezo-electric elements to produce and record the ultrasound waves. The piezo-electric element starts oscillating when a voltage is applied. This oscillation generates a sound wave that propagates through the tissue. The sound wave will be partially reflected when it crosses the interface between two tissues with a difference in density. This reflected wave is called an echo and when an echo returns to the transducer, the piezo-electric element will oscillate, generating a voltage over the piezo-electric element. The stronger the echo, the stronger the oscillation, the higher the voltage, and the brighter it will appear on the monitor.

The echo becomes stronger when the density between two tissues is larger. Transitions from soft tissue to bone will cause a strong reflection and will therefore appear bright on an ultrasound image. Most echo signals are caused by inhomogeneities in the tissues that are smaller than the wavelength of the ultrasound. These small reflectors scatter the ultrasound signal in all directions. Tissue contains many scatterers (typically more than 10 per cubic wavelength), the coherent summation of all the scattered signals will create the speckle pattern in the formed ultrasound image.

The speed of ultrasound waves through the human body lies around 1540 meters per second. The echoes that are formed at a larger distance to the transducer take more time to return to the transducer. This time-depth dependency is used to determine from which depth the echo is coming from. For example: if it takes 130 μs for an ultrasound wave to return to the transducer it can be approximated that the wave traveled a total of 20 cm, so the echo was created at a depth of 10 cm.

An ultrasound transducer consists of multiple piezo-electric elements, which are used to acquire a so-called line-by-line acquisition. Figure 1.2 shows a schematic overview of this line-by-line acquisition. A voltage is applied to a subset of piezo-electric elements on the left side of the transducer, which forms an ultrasound wave that travels though the body. All echoes that return to the transducer are recorded to form the first line of the image. After the echoes of the first line of the image are recorded, a voltage is applied to a subset of piezo-electric elements shifted one element to the right. This results in the next line of the image. This process is repeated for all piezoelectric elements to form one 2D image. This line-by-line acquisition is repeated dozens of times per second and displayed on the monitor of the ultrasound device which results in the 'real-time' recording of the imaged area.

There are also ultrasound transducers that can create a 3D volume, but these are not considered in this thesis, because these transducers are too expensive for use in a resource-limited setting. The costs of an ultrasound device are low compared to other medical imaging modalities like computed tomography and magnetic resonance imaging, but most ultrasound devices used in clinical practice today still cost between \$20k and \$200k. Innovations in recent years have made it possible to build ultrasound devices that are more portable and substantially cheaper. There are already devices on the market that can be connected to a laptop, tablet or smartphone. This makes them suitable for areas where there is no steady electrical power supply. The low-cost ultrasound devices that are available on the market today can be purchased between \$2k and \$10k, which makes ultrasound more affordable for use in resource-limited settings. Chapter 2 of this thesis takes the definition of low-cost ultrasound even a step further by describing the development of an ultrasound device with aimed production costs of less than \$100.

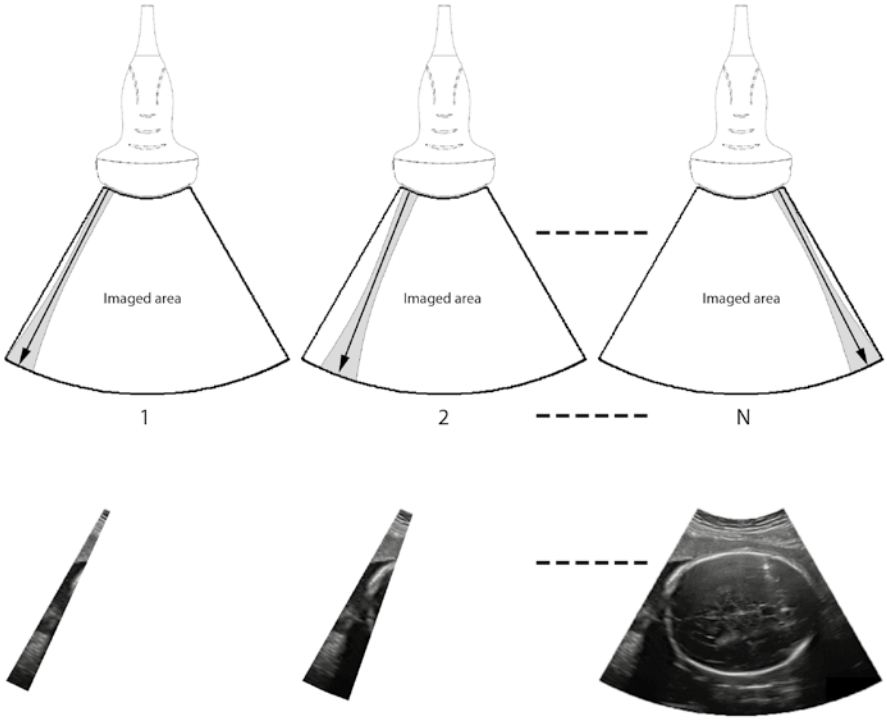

Figure 1.2: Schematic overview of a line-by-line acquisition of a curved ultrasound transducer. N stands for the number of elements, which is typically 192 or 256 for curved ultrasound transducers. The images area in this example includes a cross section of the fetal head.

1.3 Role of the sonographer

A sonographer is trained to acquire and interpret ultrasound images that can be used for the diagnosis of a patient. This requires knowledge of both the ultrasound device and the human anatomy to be able to navigate the ultrasound transducer to the correct location to form an image that contains the information needed to make a diagnosis. Theses correct locations are called "standardized planes" that contains a certain cross-section of the area of interest. For prenatal ultrasound, the sonographer needs to acquire several standardized planes to perform measurements of the fetus. Acquisition of these standard planes requires movement of the transducer over the abdomen of the pregnant woman, while maintaining contact with the skin using ultrasound gel. This gel is required to transfer the sound wave from the transducer into the body. During the examination, sonographers adjust several settings on the ultrasound device to obtain the most optimal image to perform the measurements. Learning the skills that are required to obtain these images takes months or even years. Even after this training, ultrasound imaging still suffers from intra- and inter-observer variability.

As already mentioned, there is a severe shortage of well-trained medical personnel in resource-limited countries, which impedes the introduction of ultrasound in these countries. To obviate the need of a trained sonographer to acquire the images, DeStigter et al.[6] introduced the obstetric sweep protocol (OSP), which consists of six predefined sweeps with the ultrasound transducer over the abdomen of the pregnant woman (Figure 1.3). The OSP can be taught within a day to any health care worker without any prior knowledge of ultrasound. This enables easy to learn acquisition of the ultrasound data, but after acquisition there is still a trained sonographer required to interpret the ultrasound images. Next to this, an internet connection is required to send the acquired ultrasound data to a reader center with trained sonographers. Such an internet connection is not always present in rural areas, which complicates this approach.

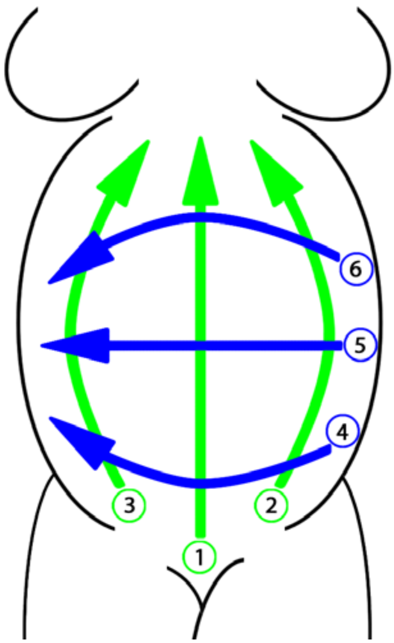

Figure 1.3: Visualization of the obstetric sweep protocol, consisting of six predefined free-hand sweeps with the ultrasound transducer over the abdomen of the pregnant woman.

In this thesis, it was therefore our goal to combine the OSP, obtained with a low-cost ultrasound device, with automated image analysis to automatically detect maternal risk factors. If successful, there would be no need to train a sonographer to both acquire and interpret the data. This would vastly reduce time and costs to train sonographers and implement this approach in resource-limited settings.

1.4 Computer-aided diagnosis

The term computer-aided diagnosis (CAD) pertains to the use of computer algorithms that assist the doctor in analyzing and interpreting medical images. There are several ways to aid the doctor using computer algorithms. The computer algorithm could perform a task that cannot be performed manually in the clinical workflow. This could give the doctor additional information that is useful for making a diagnosis. Another option is the use of computer algorithms to perform a task faster compared to the human observer. This would reduce reading time and therefore reduce costs. An additional benefit of computer algorithms is the fact that they do not suffer from intra- and inter-observer variability. A human observer does not always give the same result when he or she assesses the same image twice (intra-observer variability) and two observers could also produce a different result when the same image is interpreted (inter-observer variability). A computer algorithm always gives the same result when the same image is evaluated twice. This constant performance with a constant error is very important in clinical diagnosis, where each decision could have an influence on the treatment of a patient.

Most CAD systems that are used in clinical practice today only aid the doctor, since the doctor is (legally) responsible for the care of the patient. There are only a few systems used in clinical practice that give an independent diagnosis. An example is the CAD4TB software that autonomously reads X-ray images for the detection of tuberculosis[a]. This software is mostly used in resource-limited settings where there is a shortage of well-trained medical personnel, which therefore impedes the screening of people with tuberculosis. This approach is comparable to the use of automated ultrasound analysis to address the shortage of sonographers presented in this thesis.

The first CAD systems for medical imaging were created in the 1970s[7,8]. These algorithms usually used rule-based approaches to analyze images. In 1990s there was a shift towards feature extraction and statistical models for classification which is referred as classical machine learning in this thesis. This approach is discussed in the next paragraph. In the first year of my PhD, a new trend emerged in medical

[a] https://www.delft.care/cad4tb/

image analysis called deep learning. This approach will be discussed in paragraph 1.4.2.

1.4.1 Classical machine learning

The CAD systems that use classical machine learning usually contain two parts: feature extraction and classification. Features are numerical values derived from the imaging data and are used to discriminate between the classes that you would like to separate. These features are commonly hand-crafted, which means that the human who designs the computer system uses domain knowledge of a specific medical field to determine which values to extract from the image. For example, if you would like to extract all the pixels in an ultrasound image that belong to the fetal skull, you could extract a feature from the image that describes that the intensity of pixels belonging to the fetal skull appear bright on ultrasound images. Unfortunately, not all the bright pixels in an image belong to the fetal skull, so multiple features have to be extracted from the image to be able to discriminate the pixels of the fetal skull from its surroundings.

A classifier is a statistical model that makes a classification given the input features. To be able to decide in which class an observation belongs, the model needs to be trained with examples. In supervised learning, each example has a ground truth value. Based on these training examples the model learns a decision boundary for each combination of features. In the example of the detection of pixels that belong to the fetal skull, the classifier receives features extracted from bright pixels that belong to the fetal skull and from darker pixels that belongs to the background. Next to brightness, other features could describe if the pixels are located on curved, line-like structures (typical for the fetal skull). The classifier combines all the features to optimize the classification task. A separate test is used to evaluate the classifier. In Chapter 4 we make use of a classical machine learning approach in which Haar-like features are used in combination with a Random forest classifier to extract the pixels in an ultrasound image that belong to the fetal skull. The classifier assigns a value between zero and one to each pixel in the image that describes the probability for that pixel to be part of the fetal skull.

1.4.2 Deep learning

The idea of deep learning already existed in the 1970s[9], but application of deep learning in medical imaging became widely used in 2015[7,8], mainly due to improved computing power, better algorithms to train deep networks, and the availability of large datasets for training. The difference with classical machine learning is that deep learning algorithms learn the relevant features from the training data itself[10]. So, there is no need to spend time on manually crafting features. In medical image analysis the most used deep learning algorithms are based on convolutions. These networks consist of multiple layers of convolutions that map the input to a next layer using weighted values that are optimized during training. This approach is loosely inspired on the visual cortex of the human brain, where the convolutions resemble the response of a neuron. A disadvantage of deep learning is that these networks need a lot of data to be able to train relevant features, which is especially difficult in medical imaging where often only limited amounts of data are available.

1.4.3 CAD for ultrasound

Most CAD systems aid the clinician in obtaining a more accurate measurement or acquire the measurement in less time. For ultrasound imaging, a trained sonographer is required to acquire the images. Since the acquisition of the image usually takes more time than obtaining the measurement itself, the time gain for CAD systems in the entire workflow is fairly limited. Research in CAD for ultrasound has therefore not focused on reducing measuring time, but on automated evaluation of 3D ultrasound data to improve upon current clinical practice, since the 3D volume contains more information compared to a 2D cross section. This could lead to more accurate estimation of relevant parameters for diagnosis by decreasing the observer variability, since the sonographer does not have to acquire the exact correct cross section. It is difficult for a human to interpret a complete 3D volume, so these algorithms could also aid the sonographer by construction the correct cross section from a 3D volume. But as mention earlier, the use of 3D ultrasound is not evaluated in this thesis, since this approach is too expensive for use in resource-limited settings.

Another CAD approach is guidance of a user in obtaining the correct standardized plane to extract relevant information for diagnosis. This approach still requires the training of a human to use an ultrasound device and interpret the images to obtain relevant information for diagnosis.

A third approach, which is used in this thesis, is to extract information from predefined free-hand sweeps. The main advantage of this approach is that these predefined sweeps can be taught to any health care worker, without any prior knowledge of ultrasound, within a day. The disadvantage is that the standard plane will mostly likely not be present in the data, which could result in a less accurate measurement. In this thesis we investigate which maternal risk factors could be automatically be detected with the use of predefined sweeps.

1.5 Outline of this thesis

The research in this thesis aims to automatically detect as many maternal risk factors as possible, using the most low-cost ultrasound device as possible. Figure 1.4 gives an overview of the chapters in this thesis. The chapters of this thesis are ordered by increasing level of automation. In the current clinical practice the sonographer acquires the ultrasound images using mid-range devices (ranging from $20.000 until $40.000) or high-end devices (ranging from $100.000 until $200.000). Each chapter makes use of different ultrasound devices.

Chapter 2 discusses the development of a very low-cost ultrasound device that is aimed at a production cost less than $100. This device was evaluated using simulations, phantom experiments and *in vivo* images obtained by a trained sonographer.

Chapter 3 presents a comparison study of low-cost ultrasound devices to manually estimate the gestational age using the OSP, so the acquisition could be performed by any health care worker in resource-limited countries, but a trained sonographer was still required to interpret the images.

Chapter 4 introduces and validates a computer algorithm for the automated measurement of the fetal head circumference using 2D standard plane ultrasound images. This chapter shows the feasibility of automating this measurement, but the acquisition of the 2D standard plane for measuring the fetal head circumference still requires a trained sonographer and all images were obtained using a high-end ultrasound device.

Chapter 5 presents a computer algorithm that is able to fully automatically estimate the gestational age with the use of the OSP. In this chapter the sweeps were obtained with a mid-range ultrasound device at a maternal health clinic in Ethiopia.

Chapter 6 demonstrates that the algorithms can be extended to fully automatically detect twins, estimate gestational age and determine fetal presentation all from data obtained with the use of the OSP. In this study, all data was acquired with a low-cost ultrasound device a maternal health clinic in Ethiopia. With this final chapter we show the potential to detect these maternal risk factors without the need of a trained sonographer in limited-resource countries. The low-cost ultrasound device used in Chapter 6 can be connected to a laptop or tablet, which makes it a portable solution for rural areas.

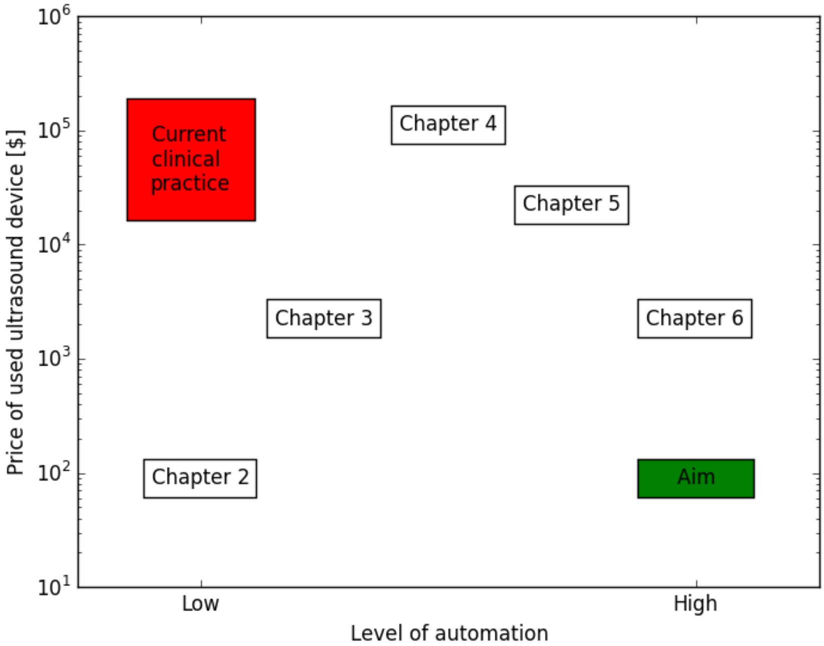

Figure 1.4: Chapter overview of this thesis.

2

Development of a low-cost medical ultrasound scanner

T.L.A. van den Heuvel
D.J. Graham
K.J. Smith
C.L. de Korte
J.A. Neasham

Original title:
Development of a Low-Cost Medical Ultrasound Scanner Using a Monostatic Synthetic Aperture

Published in:
IEEE Transactions on Biomedical Circuits and Systems

Abstract

In this chapter we present the design of low-cost medical ultrasound scanners aimed at the detection of maternal mortality risk factors in developing countries.

Modern ultrasound scanners typically employ a high element count transducer array with multichannel transmit and receive electronics. To minimise hardware costs we employ a single piezoelectric element, mechanically swept across the target scene, and a highly cost-engineered single channel acquisition circuit. Given this constraint, we compare the achievable image quality of a monostatic fixed focus scanner (MFFS) with a monostatic synthetic aperture scanner (MSAS) using post-focusing. Quantitative analysis of image quality was carried out using simulation and phantom experiments, which were used to compare a proof-of-concept MSAS prototype with a MFFS device currently available on the market. Finally, *in vivo* experiments were performed to validate the MSAS prototype in obstetric imaging.

Simulations show that the achievable lateral resolution of the MSAS approach is superior at all ranges compared to the fixed focus approach. Phantom experiments verify the improved resolution of the MSAS prototype but reveal a lower signal to noise ratio. *In vivo* experiments show promising results using the MSAS for clinical diagnostics in prenatal care.

The proposed MSAS achieves superior resolution but lower SNR compared to an MFFS approach, principally due to lower acoustic energy emitted. Significance: The production costs of the proposed MSAS could be an order of magnitude lower than any other ultrasound system on the market today, bringing affordable obstetric imaging a step closer for developing countries.

2.1 Introduction

Worldwide, complications of pregnancy and childbirth lead to approximately 830 deaths every day, of which 99% occur in developing countries. This is mainly caused by the limited access to health services in these areas of the world[1]. With the use of ultrasound imaging it is possible to detect maternal mortality risk factors, but ultrasound devices remain out of reach for healthcare providers in low-resource settings because even the lowest cost ultrasound devices available on the market today are still cost prohibitive. Examples of low-cost devices recently entering the market include the Interson SeeMore probe, the SunBright SUN-806F and the Telemed MicrUs, which can be purchased between $1.5k and $3k. The lowest cost devices emerging from the established ultrasound vendors include Siemens Acuson P10, GE Vscan, Philips Lumify and VISIQ, which are even more expensive options.

In this paper we present an ultrasound device with production costs less than $100. This would make this ultrasound device an order of magnitude cheaper compared to the low-cost ultrasound devices available on the market today. To achieve this goal, a significant reduction in complexity of the hardware is required. The main cost driver of ultrasound systems is the multi-element piezoelectric transducer array which is intricate, expensive and also requires multiple channels of multiplexed transmit and receive electronics. Hardware costs can be vastly reduced by simplifying the transducer array to a single piezoelectric element, as shown in Fig. 2.1, which is mechanically swept across the target scene.

A monostatic design, consisting of a single, mechanically scanned transmit/receive transducer element, poses two main challenges. Firstly, the maximum achievable frame rate will be limited compared to electronic beam-steering, because the single element has to be moved across the target scene. However, as long as this frame rate is sufficient for the diagnostic task in hand, the reduction in production costs outweighs this disadvantage. Second, a monostatic design leads to compromises on lateral resolution compared to an electronically focused transducer array, with resolution constrained by the fixed beam pattern of a naturally focused transducer and the sub-optimal resolution in the near and far field regions. To improve the lateral resolution of the monostatic design, synthetic aperture focusing is explored here.

Besides the lower production costs, the monostatic design may offer some advantages compared to a full array transducer. The response of a monostatic system is perfectly matched at each position and should also be more reliable as individual elements of a transducer array can fail over time and degrade image quality[11]. Given the intended low resource setting, the likelihood of probe damage is increased, hence a cheap and easily replaceable transducer arrangement is preferable. A monostatic

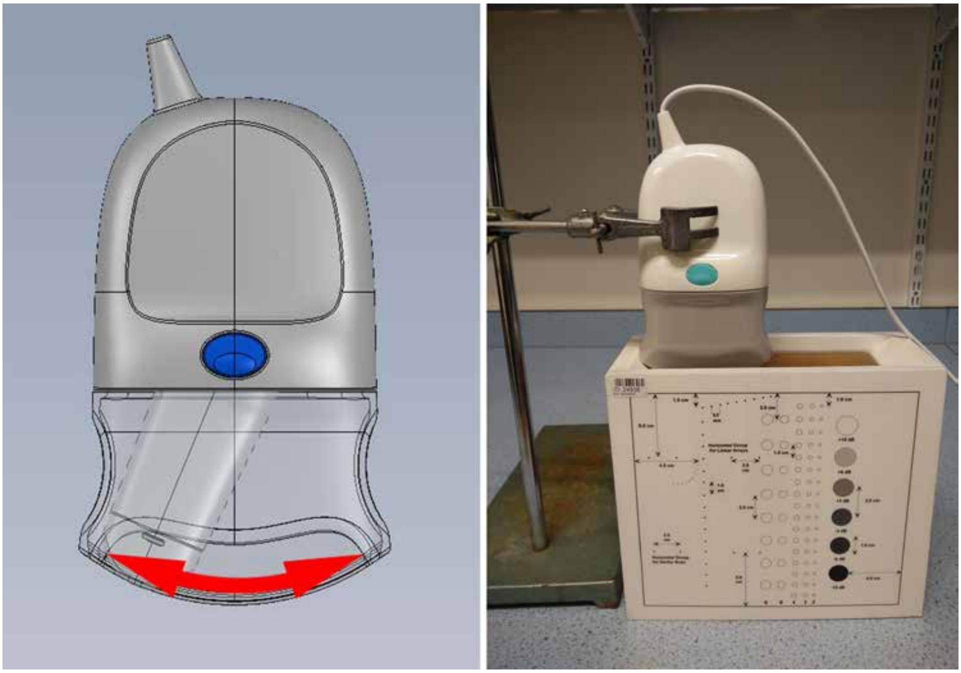

Figure 2.1: Left: A schematic drawing of the MSAS design. The red arrow indicates the path of the single element transducer over the target scene. Right: experimental setup with the MSAS prototype for the phantom experiments.

system may also allow a synthetic array with an element pitch, E_P, smaller than the physical element width, E_W, which is not possible with a physical array, as shown in Fig. 2.2.

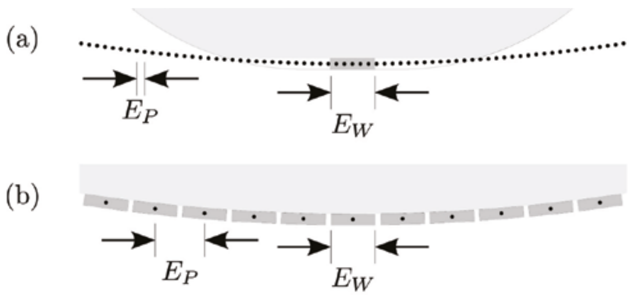

Figure 2.2: Comparison of array geometry using a physical element of width, E_W. Each element position is indicated by a dot, resulting in the effective element pitch distance, E_P. (a) Convex synthetic array formed from a single moving element. (b) Convex physical array.

The monostatic design was evaluated using simulations, phantom experiments and *in vivo* experiments. Simulations were made to compare the fundamental performance of MFFS and MSAS designs. A proof-of-concept MSAS was then produced and compared a MFFS ultrasound device via phantom experiments. *In vivo* experiments were performed to validate the use of the MSAS prototype in prenatal care.

2.2 Literature Overview

Most reported work on synthetic aperture focusing in ultrasound imaging uses a physical transducer array[12–16]. The use of synthetic aperture focusing in medical imaging was introduced by Burckhardt et al.[17] who showed that synthetic aperture with a single element gives a significantly higher lateral resolution compared to a conventional B scan. In 2007, Kortbek et al.[18] used Field II simulations to show that a single rotating mechanically focused concave element, which is used in an anorectal ultrasound transducer, increased the SNR. In 2010 Opretzka et al.[19] used a fixed focus single-element for high frequency ultrasound on a wire phantom and showed a significant reduction of side lobes and of noise compared to delay-and-sum. (They have also a paper in 2012 that shows animal results[20]). In 2011 Andresen et al.[21] used synthetic aperture focusing with a single-element transrectal ultrasound transducer, making a helical motion to obtain 3-D volumes. Simulations and a wire phantom experiment showed a significant improvement in azimuth resolution. Although these papers describe the use of synthetic aperture focusing for medical imaging with a single element, none of this work has shown any *In vivo* results of this technique and this is the first low-cost device using synthetic aperture focusing[22] applied on prenatal care in developing countries.

2.3 Methods

The MSAS prototype and testing methodology is presented in three sections: hardware design, software design and simulation/experiments. A schematic overview of the hardware and software is shown in Fig. 2.3. The hardware section describes the design of the low-cost probe including the transducer, the scanning mechanism and the communication interface. The software section describes the processing of the raw data, acquired by the low-cost probe, to produce the B-mode image, which includes the synthetic aperture focusing method. The simulations and experiments section describes simulations, phantom and *in vivo* experiments that were performed to evaluate the image quality of the MSAS.

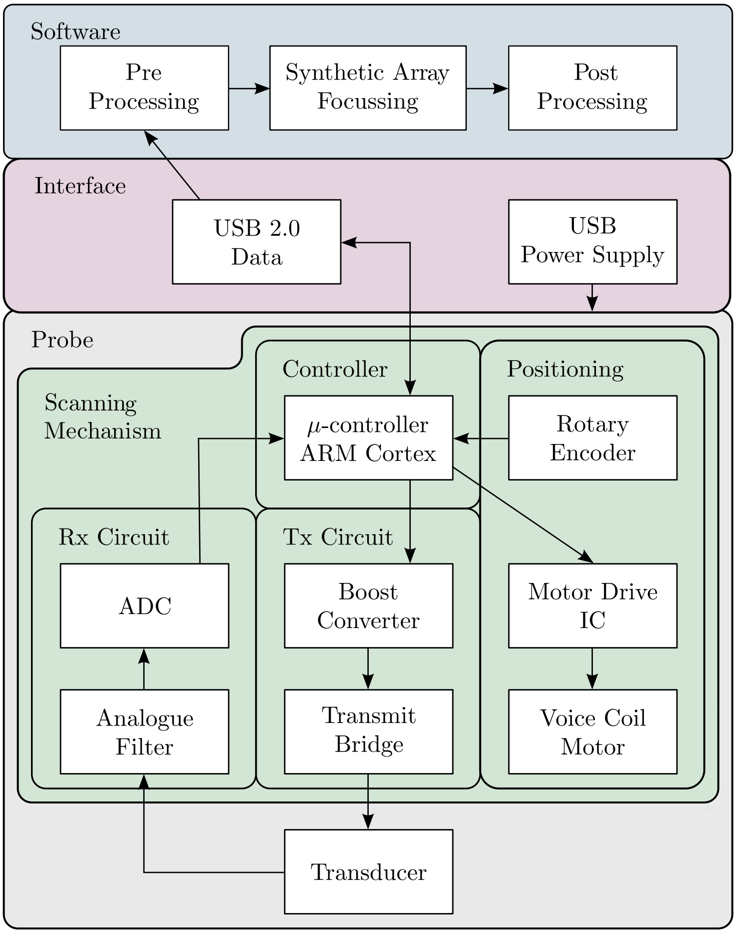

Figure 2.3: A schematic overview of the MSAS prototype design.

2.3.1 Probe Hardware

Transducer

The first major step toward lower cost hardware is to replace the costly construction of a multi-element piezoelectric transducer array and multiple channels of transmit and receive electronics with a single element transducer and single channel of electronics, similar to early imaging systems[17,23]. The monostatic design significantly simplifies the transducer construction to a single piece of piezoelectric material with an electrical connection to each electrode, a quarter-wave matching layer on the active face and an attenuative backing material on the opposite face. The axial resolution is determined by the bandwidth, Δf, whereas the lateral and elevation beam width are determined by the size and geometry of the transducer aperture and the centre frequency, f_c. The physical size of the transducer chosen for the MSAS design is strongly related to the synthetic aperture focusing process which will be explained in Section 2.3.3. Operating frequencies of 2-5MHz are typical for ultrasound abdominal probes as this represents the best trade-off between resolution and penetration depth. In this design the transducer makes direct contact with the skin. With careful design of the transducer housing shape, it is found that coupling to the skin can be maintained via standard coupling gel.

Scanning Mechanism

The single element must be mechanically moved along a predetermined path in order to collect the necessary echo data to build an image. This path was chosen to match that of a standard convex array probe, as shown in Fig. 2.1, which represents the best trade-off between the probe dimensions and the field of view. The transducer of the MSAS prototype makes a sweep of $50°$ on an 8cm radius arc corresponding to a 200 element synthetic array. It takes 0.25 seconds to obtain one sweep, which results in a frame rate of four frames per second (see Table 2.1 for an overview of all parameters).

Positioning system A motor is used to sweep the single transducer across the target scene. The motor design has four design criteria: high torque to overcome the contact friction with the skin, low electrical noise so it does not adversely affect the Signal-to-Noise Ratio (SNR) of echo data, a maximum current of 300mA for operation from a single USB bus power supply and finally appropriate size/weight for the probe to be easily held with one hand. A Voice Coil Motor (VCM) was designed to satisfy these requirements[24]. This motor design uses no brushes and hence is electrically quiet. Furthermore, direct drive means there is no audible noise from a gearbox and there are minimal parts which will suffer mechanical wear. The torque generated by the motor is limited, as described in[24], and if too much pressure is applied by the operator the motor stalls before any discomfort or injury could result. Fluctuations in speed due to variable friction have no inherent effect on image quality as the optical encoder determines when the transducer is excited to ensure that the data is accurately captured. Accurate position registration was achieved using an optical rotation encoder with an angular resolution, θ_{res}, which then initiates each transmit- and receive cycle at $0.25°$ increments. Basic closed loop speed control was implemented, using a pulse-width modulation motor driver together with feedback from the rotary encoder, in order to achieve near uniform sampling of the synthetic array.

Transmit Electronics The transmit circuit is required to produce a short high voltage pulse to excite the transducer at its resonance. The duration of the pulse in cycles should be the upper bound of the Q factor of the transducer (2.1).

$$Q = \frac{f_c}{\Delta f} \tag{2.1}$$

A boost converter was used to produce 48V from the 5V USB supply which was then used to supply a class-E amplifier. The inductor in this amplifier topology was selected to provide a Q-magnification of 2, hence producing a 96V 2-cycle pulse at 4.2MHz to drive the transducer.

Receive Electronics The data acquisition sub-system is the most expensive electronic component of a medical ultrasound device, typically requiring multiple channels of high bandwidth, low noise amplifiers/filters and high specification Analogue to Digital Converters (ADCs). Even when simplifying to a single channel, these components still represent a large proportion of the overall system cost. Careful design and performance trade-offs must be considered to achieve a truly low-cost device. The ADC of the MSAS design was chosen as the optimum balance between price and performance in terms of signal to noise ratio as estimated by the ideal ADC equation (2.2), where b is the bit resolution. Based on these criteria 14-bit ADC was selected capable of sampling at 12MS/s which costs $< \$10$ in large quantities.

$$\mathrm{ADC}_{\mathrm{SNR}} = 1.76 + 6.02b + 10\log_{10}\left(\frac{f_s}{2 \cdot \Delta f}\right)\mathrm{dB} \tag{2.2}$$

Controller The controller is synchronises all positioning, ultrasound transmission and data acquisition operations and transfers echo and position data to the processing software.

2.3.2 Interface

The MSAS probe was designed to interface to a standard USB 2.0 connection which provides sufficient data rate and power supply for this design. This ensures compatibility with many processing platforms, whether new or legacy.

2.3.3 Software

The processing software was developed to perform the signal processing and display the B-mode image on readily available, low-cost platforms such as PCs and laptops. Given the wide availability of PCs and the number of schemes already in place to provide hardware to the developing world, it is believed that this represents the cheapest possible processing and display unit, which lowers the overall system costs. First, the echo data is filtered and down-converted. Second, synthetic aperture focusing is applied to the baseband data. Last, post-processing steps are performed to generate the final grey scale, B-mode image.

Pre-processing

After formatting the echo data, band pass filtering is applied over a bandwidth of 2MHz around the centre frequency, f_c, to remove out of band energy from the data. Then a time varying gain curve is applied to the echo data to compensate for attenuation with depth. Finally the echo data is down-converted to a complex baseband representation.

Table 2.1: MSAS prototype parameters

Name	Symbol	Value
Centre Frequency	f_c	4.2 MHz
Bandwidth	Δf	2 MHz
Range	R	0.15 m
Average speed of sound in tissue	c	1540 ms^{-1}
Radius of Curvature		0.082 m
Lateral Arc Length	θ_L	50°
Angular Resolution	θ_{res}	0.25°
ADC Resolution	$\mathrm{ADC}_{\mathrm{res}}$	14 bit
Sampling Frequency	f_s	12 MS/s
Element Width	E_W	2 mm
Element Length	E_L	7 mm
Element Thickness	E_T	0.5 mm
Near Field Lateral	N_L	2.7×10^{-3} m
Near Field Elevation	N_E	0.034 m
Beam Spread Angle Lateral	α_L	10.8°
Beam Spread Angle Elevation	α_E	3.1°
Frames/s (software limited)		4
Data Throughtput (4FPS)		30 Mb/s

Synthetic aperture focusing

When using a fixed focus approach with a single mechanically swept transducer, the B-mode image is constructed by a simple polar to Cartesian conversion of echo data from each direction viewed. The lateral resolution of such a system is dependent on the beam spread angle, α which favours a large aperture diameter, E_D (2.3). However, the image will also be distorted due to the complex beam shape up to the near field distance, D_{near}, given by (2.4) which is proportional to the square of E_D. Hence any fixed focus system represents a trade-off between the extent of the near field and the beam spread in the far field which suggests that the optimum fixed focus transducer for typical obstetric imaging, at 4MHz and up to 15cm range, consists of a $\approx 15\text{mm}$ diameter ceramic disk with an acoustic lens to bring the natural focus back to $\approx 7\text{cm}$.

$$\frac{\alpha}{2} = \sin^{-1}\left(\frac{0.514 \cdot c}{f_c \cdot E_D}\right) \tag{2.3}$$

$$D_{\text{near}} = \frac{E_D^2 \cdot f_c}{4 \cdot c} \tag{2.4}$$

The MSAS uses synthetic aperture focusing whereby echo data from multiple positions are coherently summed to calculate each image pixel. This favours a transducer with a larger beam spread angle in the lateral direction i.e. the direction of motion, but still requires a narrow beam in the elevation plane. With the use of synthetic aperture focusing each individual pixel_{ij} of an image of size $i \times j$ can be calculated as shown in Fig. 2.4 by combining the received baseband echo data, B_{ij} from N synthetic aperture positions from which that pixel is visible and applying appropriate

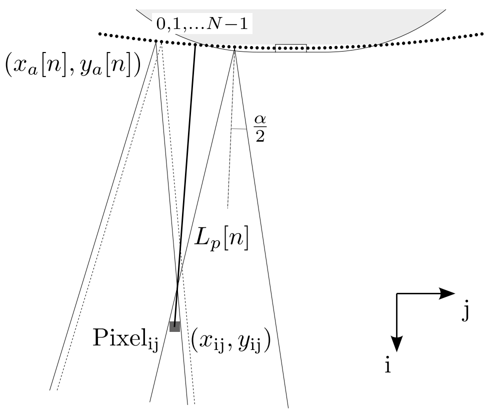

Figure 2.4: Synthetic aperture focusing to calculate Pixel_{ij} from N synthetic aperture elements.

delays, d_n, phase rotations, ϕ, and weightings, W_{ij}, as given by (2.5). The sample delay, d_n, is calculated as in (2.6) with the two-way path length, L_p, the sampling frequency, f_s, and the average speed of sound in tissue c. The phase rotation, ϕ, is calculated as in (2.7) using the centre frequency, f_c. The two way path length, L_p, is calculated using (2.8) where (x_a, y_a) and (x_f, y_f) are the x and y coordinates of the aperture element and the coordinates of focus respectively. The weightings, W_{ij}, are selected from an N length Tukey window function[25] to achieve an acceptable trade-off between main lobe width and side lobe levels.

$$\mathrm{Pixel}_{ij} = \left| \frac{1}{N} \cdot \sum_{n=0}^{N-1} W_{ij}[n] \cdot B_{ij}[n, d_n] \cdot \phi_{ij}[n] \right| \tag{2.5}$$

$$\mathrm{d_n} = \left\lfloor \frac{L_p[n] \cdot f_s}{c} \right\rceil \tag{2.6}$$

$$\phi_{\mathrm{ij}}[n] = \frac{-2 \cdot \pi \cdot L_p[n] \cdot f_c}{c} \tag{2.7}$$

$$L_p[n] = 2 \cdot \sqrt{\left(x_{\mathrm{ij}}^2 - x_a^2[n]\right)^2 + \left(y_{\mathrm{ij}}^2 - y_a^2[n]\right)^2} \tag{2.8}$$

There are two key advantages of this technique. Firstly, the effective transducer aperture increases proportional to the axial range which makes lateral resolution less dependent on range. In the case of a linear scan path, the lateral resolution becomes independent of range but in the case of a convex array some degradation of lateral resolution with respect to range remains, depending on the radius of curvature. Secondly, synthetic focusing of signals gathered from a small physical aperture ensures that the best possible beam pattern is formed at every range and eliminates the distortions seen in the near field of a larger physical aperture.

Post-processing

Two straightforward post-processing steps were performed to improve the visual quality of the B-mode image for clinical use. First, a log compression was performed to adjust the dynamic range. Second, a stick filter[26] was applied to reduce coherent speckle in the final image.

2.3.4 Simulations and experiments

Simulations and experiments were performed to evaluate the MSAS design. First, Field II simulations were used to compare the image quality of MSAS versus MFFS.

Second, phantom experiments were used for quantitative comparison of the MSAS prototype with an MFFS ultrasound device (Interson SeeMore, model 99-5901, centre frequency 3.5 MHz) . The MFFS ultrasound device makes use of a fixed focus single element which moves inside an oil filled housing. Finally, *in vivo* experiments were used to validate the use of the MSAS prototype for common measurements in prenatal care.

Simulations

Field II[27,28] was used to simulate a single element ultrasonic transducer travelling a convex path with an angular resolution, θ_{res}, of $0.25°$ and performing a transmit- and receive cycle at each position. For MSAS a rectangular element of 2mm (lateral) by 7mm (elevation) was simulated whereas for the MFFS simulation a circular aperture of 15mm diameter was used together with a lens, giving similar beam characteristics to the MFFS ultrasound device.

Phantom experiments

The image quality of the MSAS prototype and MFFS ultrasound device were evaluated using Quality Assurance for UltraSound (QA4US) software. This software makes it possible to quantitatively analyse B-mode images and was used to measure the elevation focus, spatial resolution, spatial conformity and contrast sensitivity of each ultrasound device[29].

Elevation focus The elevation direction of the ultrasound probe is perpendicular to the displayed sector and therefore normally not visible. However, special slice thickness phantoms have been developed to visualize the elevation focus using a plane of scatterers which are positioned at an angle of $45°$. This enables the visualization of the slice thickness as a function of the depth of the scatterers in the phantom. The QA4US software defines the elevation focus depth as the depth where the smallest detected slice thickness is located.

Spatial resolution The QA4US software defines the spatial resolution of an ultrasound device as the Full Width Half Maximum (FWHM) of the wire in the phantom that is closest to the elevation focus. The depth of the (in-plane) transmit focus was also set to the elevation focus depth. This will result in the best spatial resolution that can be imaged with the device. The spatial resolution degredates outside the focal point. This degradation was quantified by taking the average of the FWHM of the wires in the phantom two centimetres above and below the focal point and compare this to the best spatial resolution that can be imaged with the device.

Spatial conformity The QA4US software defines the spatial conformity as the percentage difference between two measured wires in the phantom measured on the B-mode image and the actual distance of the wires in the phantom.

Dynamic range and contrast sensitivity The QA4US software defines the dynamic range as the number of dB's within the 0-255 gray level range. It was measured by acquiring multiple images of the contrast objects in a phantom. The contrast sensitivity is defined as the SNR at 3dB as shown in (2.9)

$$\mathrm{SNR_L} = \frac{|\mu_L - \mu_B|}{\sqrt{\sigma^2_{\mu_\mathrm{L}} + \sigma^2_{\mu_\mathrm{B}}}} \tag{2.9}$$

where: μ_L and μ_B are the ensemble mean echo levels of lesion (L) and the surrounding background tissue (B)[29]. This makes the contrast sensitivity dependent on the dynamic range. Therefore the dynamic range needs to be corrected to one reference dynamic range to be able to make a fair comparison of the contrast sensitivity between different ultrasound devices. All pixels in the image are multiplied according to (2.10) to produce a corrected pixel value Pixel'_{ij}, where Pixel_{ij} is the original pixel intensity of a pixel, $\mathrm{DR_{ref}}$ is the reference dynamic range of 2.55 and DR is the measured dynamic range of the device.

$$\mathrm{Pixel}'_{ij} = \mathrm{Pixel}_{ij} \cdot \frac{\mathrm{DR_{ref}}}{\mathrm{DR}} \tag{2.10}$$

In Vivo experiments

In vivo experiments were performed to evaluate the performance of the MSAS prototype in prenatal care. The local ethics committee approved the use of the MSAS prototype on pregnant women. This was achieved by proving that the SESAS provides conformance to the FDA Track 1 standards Fetal Imaging application. Hydrophone measurements showed a Mechanical Index below 0.2, a Derated Spatial-Peak Temporal-Average Intensity below 9 $\mathrm{mWcm^{-1}}$ and a Derated Spatial-Peak Pulse-Average Intensity below 5.2 $\mathrm{Wcm^{-1}}$. The electrical safety of the system was tested according to the NEN-EN-IEC 60601-1 of the NEN 3140. Every pregnant woman in this evaluation study signed a written informed consent. In prenatal care ultrasound can be used to detect maternal mortality risk factors. The performed *in vivo* experiments focussed on biometric measurements that can be used to determine the Gestational Age (GA) and growth of the fetus. In the first trimester the Crown-Rump

Length (CRL) of the fetus is the most reliable measurement to determine the GA. The Head Circumference (HC) and Abdomen Circumference (AC) of the fetus can be used in the second and third trimester to assess the growth of the fetus[30].

2.4 Results

2.4.1 Simulation

Fig. 2.5 shows the results of Field II simulations to generate B-mode images of an arrangement of ideal point targets, similar to that used in phantom experiments. The left image shows the MSAS result and the right image shows the MFFS result. Fig. 2.6 shows a comparison of the lateral resolution versus depth computed from the vertical row of the images in Fig. 2.5.

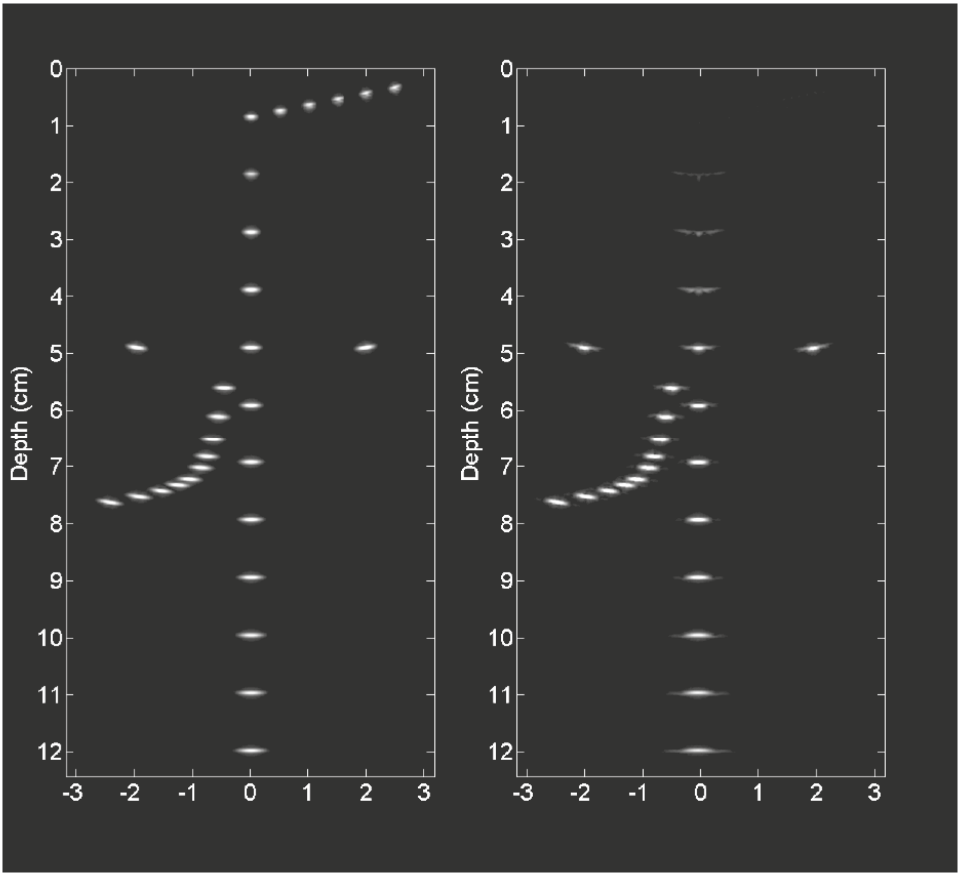

Figure 2.5: Simulated images of point targets. Left: MSAS. Right: MFFS.

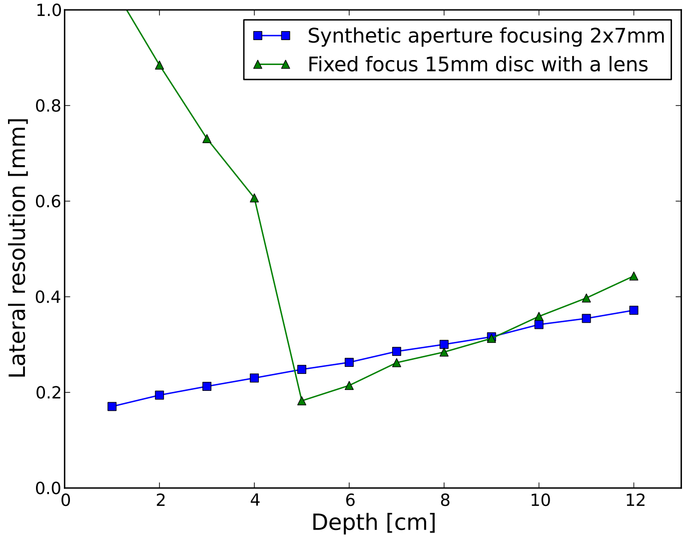

Figure 2.6: Lateral resolution computed from the simulated images of point targets.

2.4.2 Phantom Experiments

Fig. 2.7 shows an example of point target data, at a depth of 10cm, extracted from the resolution phantom experiments with the MSAS prototype. The improvements in lateral resolution and signal to noise ratio resulting from synthetic focusing are clearly shown.

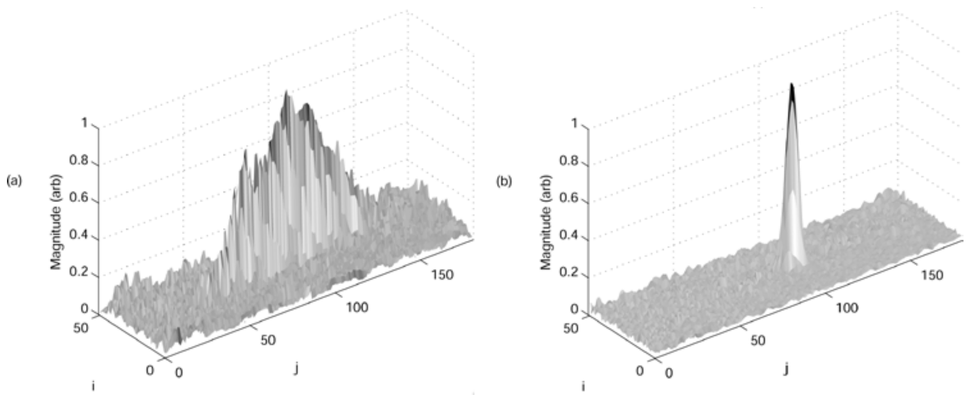

Figure 2.7: Surface plot of a 0.1mm diameter target from a resolution phantom at a depth of 10cm obtained using MSAS prototype. (a) Physical beam pattern (before synthetic focusing). (b) After synthetic focusing.

Fig. 2.8 and Fig. 2.9 show B-mode images obtained in phantom experiments using the MSAS prototype and the MFFS ultrasound device. Quantitative analysis of the image quality from the two ultrasound devices was performed using the QA4US tool. Table 2.2 shows the result of this analysis.

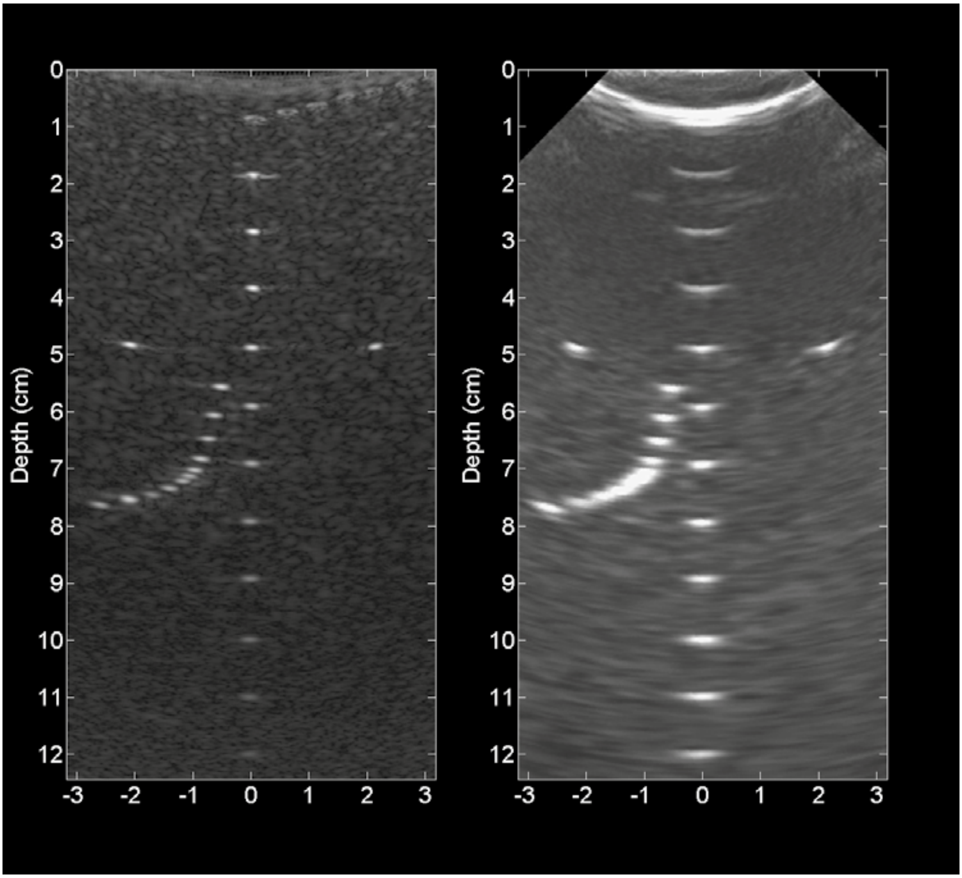

Figure 2.8: Example images of the phantom wires used for calculating spatial resolution and spatial conformity. Left: MSAS prototype. Right: MFFS ultrasound device.

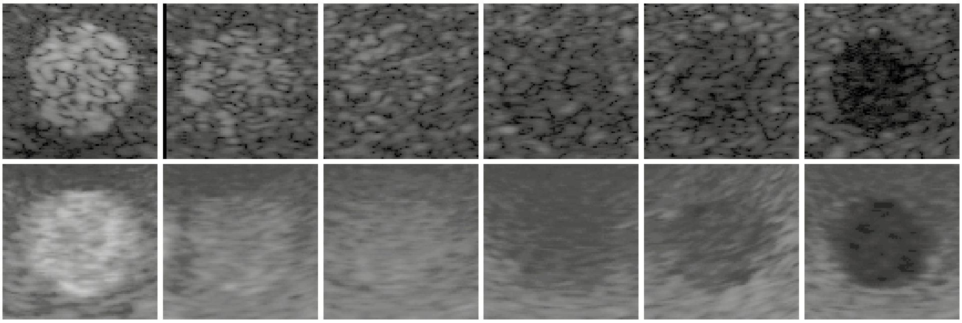

Figure 2.9: Example images of the phantom contrast disks that are used for calculating the dynamic range and contrast sensitivity (left to right: 15dB, 6dB, 3dB, -3dB, -6dB, -15dB). Top: MSAS prototype. Bottom: MFFS ultrasound device.

Table 2.2: QA4US results

Parameter	MSAS	MFFS
Elevation Focus [mm]	30.6	50.5
Slice thickness at elevation focus [mm]	1.08	1.2
Axial Resolution in focus [mm]	0.23 ± 0.03	0.37 ± 0.05
Axial Resolution averaged over depth [mm]	0.28 ± 0.05	0.55 ± 0.13
Lateral Resolution in focus [mm]	0.55 ± 0.02	1.25 ± 0.06
Lateral Resolution averaged over depth [mm]	0.67 ± 0.11	2.71 ± 1.40
Axial spatial conformity [%]	0.2 ± 0.18	2.0 ± 0.06
Dynamic Range [dB]	140	98
Contrast Sensitivity	1.22	2.20
Corrected Contrast Sensitivity	1.46	2.19

2.4.3 In Vivo experiments

Fig. 2.10, shows three example B-mode images of the *in vivo* experiments with the MSAS prototype, one from each of the trimesters of the pregnancy. The top image shows a side view of a fetus in the first trimester, which can be used to measure the crown to rump length (CRL) of a fetus. The middle image shows a cross section of the fetal head in the second trimester, which can be used to measure the head circumference (HC) of the fetus. The bottom image shows a a cross section of the fetal abdomen in the third trimester, which can be used to measure the abdominal circumference (AC) of the fetus.

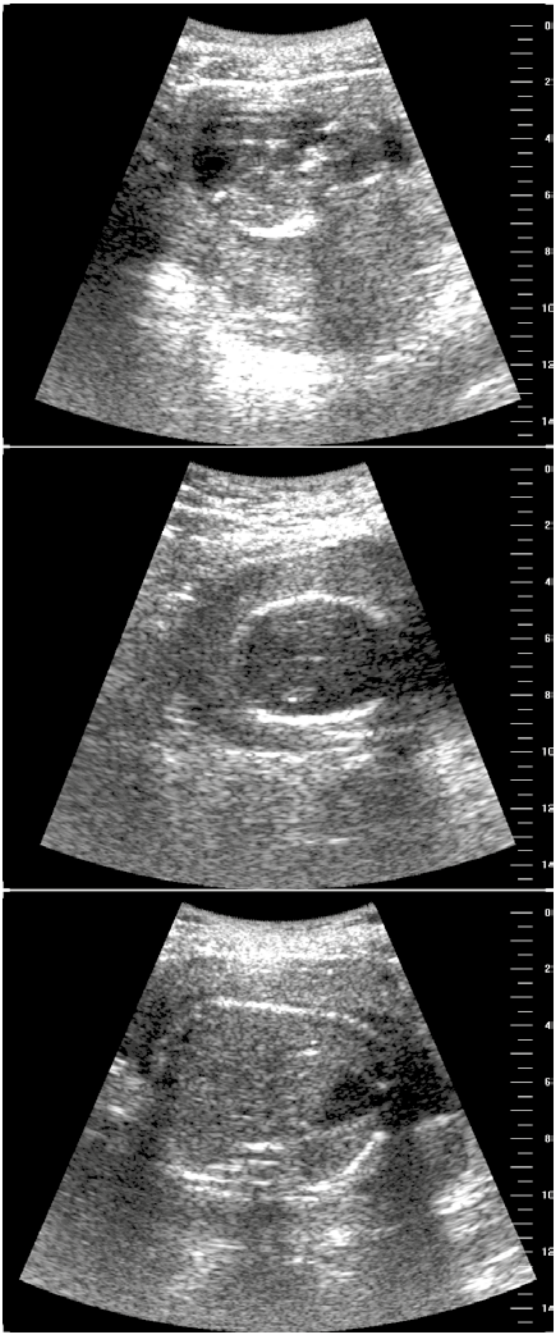

Figure 2.10: From top to bottom: B-mode image of a fetus in the first trimester, which can be used to measure the CRL of the fetus. B-mode image of the head of a fetus in the second trimester, which can be used to measure the HC. B-mode image of the abdomen of a fetus in the third trimester, which can be used to measure the AC.

2.5 Discussion

2.5.1 Simulations

The Field II simulations, shown in Fig. 2.5 and Fig. 2.6, show that the best case lateral resolution of the MSAS and MFFS approaches is similar. However it is clearly demonstrated that the near field distortions seen on the MFFS image, due to the complex beam shape up to the near field distance, are eliminated using the MSAS approach and the resolution at the maximum range is also improved. Overall, the MSAS approach achieves a much more consistent lateral resolution over the full range of the image.

2.5.2 Phantom experiments

The MSAS prototype demonstrates superior axial and lateral resolution compared to the MFFS ultrasound device, which is visible in Fig. 2.8 and was quantified with the QA4US software with results shown in Table 2.2. This superior resolution becomes even more pronounced when the mean resolution over the full depth range is evaluated. The MFFS ultrasound device has a large decay in lateral resolution due to its fixed focus, single element transducer design. The mean lateral resolution of the MFFS ultrasound device is therefore more than twice as poor as the best case resolution. The spatial conformity in axial direction of the MFFS ultrasound device was also found to be worse compared to the MSAS prototype. This may be simply due to the poorer lateral resolution but it may also be affected by the applied frame averaging or smoothing. Unfortunately, post-processing could not be turned off by the software. After correction for dynamic range it can be seen that the contrast sensitivity of the MSAS prototype is significantly lower compared to the MFFS ultrasound device. This is caused by the combination of the small aperture size of the transducer element (2mm x 7mm) and compromises in the analog front end electronics, both of which degrade the achievable SNR even after focusing.

2.5.3 In Vivo experiments

Extensive *in vivo* experiments have shown that it is possible to view the fetus with the MSAS prototype in all trimesters, illustrated by the three examples in Fig. 2.10. The relatively low contrast sensitivity of the MSAS makes it a more challenging task to determine the exact border of the fetus in the first trimester and therefore error bounds on CRL measurements are likely to be higher than with state-of-the-art scanners. The middle image of shows that the MSAS is able to image a sharp edge of the fetal head but limited detail of the internal soft tissue structure, again due to the contrast sensitivity. This makes it more challenging to find internal markers which indicate the ideal cross section of the fetal head for the HC measurement but a reasonable estimate of the head circumference can still be made. In the third trimester it proved possible to view the abdomen of the fetus, but it is more challenging to pick out all deeper structures which are not as clearly visible due to shadowing and the lower contrast sensitivity. Despite the lower frame rate and the lower contrast of the MSAS, it has still proven possible to measure important fetal biometrics such as CRL, HC and AC.

2.5.4 User experience

When designing a low-cost ultrasound devices compromises have to be made to decrease production costs. Like most low-cost ultrasound devices, the MSAS prototype has relatively few parameters that can be adjusted by the user. This may make the use of these devices in developing countries easier but it is not possible to change them if parameters are suboptimal. A significant compromise of the MSAS prototype constructed was found to be the lower frame rate. This diminishes the 'real-time' imaging experience that experienced sonographers are accustomed to. Users could adapt to this lower frame rate but, in combination with the low contrast sensitivity, this made it more difficult and time consuming to locate the appropriate views of a fetus for biometric measurements.

2.5.5 Improvements

The biggest scope for improvement on the MSAS design is in the contrast sensitivity of the device by increasing the received SNR. Three main approaches could be used to obtain a higher contrast sensitivity. Firstly, a larger transducer aperture in the elevation direction in combination with a lens to optimise the focal depth, would increase the transmitted and received ultrasound energy. Second, a higher specification ADC could be used to allow oversampling well above the Nyquist sampling frequency to reduce aliasing and quantisation noise. Thirdly, there is substantial scope to improve to the receiver amplifier to reduce noise floor. User experience would also be improved by a higher frame rate which is possible by optimisation of the processing software. However, careful design will be needed, harnessing ever improving cost/performance of components, to maintain an order of magnitude reduction of system cost. At the time of writing only simple image post-processing steps are performed in the MSAS software. More advanced post-processing steps could be also included to improve the visual quality of the images by optimising dynamic range and reducing speckle.

2.6 Conclusion

In this paper a low component cost, monostatic synthetic aperture scanner (MSAS) was presented with potential applications in low resource countries for the detection of maternal risk factors. This system was designed to reduce the production costs by replacing the multi-element piezoelectric transducer array by a single piezoelectric element which is mechanically swept across the target scene. Since a single piezoelectric element leads to inevitable compromises in image resolution, synthetic aperture focusing was used to achieve consistently high resolution across the B-mode image. Simulations proved that the lateral resolution of the monostatic synthetic aperture focusing approach is superior to that of a monostatic fixed focus scanning (MFFS) approach at almost every range. Phantom experiments, performed with a proof-of-concept MSAS prototype, showed that it has superior axial and lateral resolution compared to another single element 'low-cost' ultrasound device that is available on the market today. However, the phantom experiments also showed that the current MSAS design suffered from a relatively low contrast sensitivity. To validate the performance of the MSAS in prenatal care *in vivo* experiments were performed. The *in vivo* experiments show promising results for clinical diagnostic use of the MSAS. Even with the lower frame rate is was possible to detect the fetus in all three trimesters and image different parts of the fetus that are important for making biometric measurements of the fetus. Therefore it can be concluded that, with further development, the proposed design has the potential to deliver an affordable technology for developing countries to detect maternal risk factors and hopefully reduce the number of maternal deaths in the future.

3

Comparative study of low-cost ultrasound devices

T.L.A. van den Heuvel
D. de Bruijn
Desirée Moens-van de Moesdijk
A. Beverdam
B. van Ginneken
C.L. de Korte

Original title:
Comparison study of low-cost ultrasound devices for estimation of gestational age in resource-limited countries

Published in:
Ultrasound in Medicine and Biology

Abstract

We investigated how accurately low-cost ultrasound devices can estimate gestational age (GA) using both the standard plane and the obstetric sweep protocol (OSP). The OSP can be taught to health care workers without prior knowledge of ultrasound within one day and thus avoid the need to train dedicated sonographers. Three low-cost ultrasound devices were compared with one high-end ultrasound device. GA was estimated with the head circumference (HC), abdominal circumference (AC) and femur length (FL) using both the standard plane and the OSP. The results revealed that the HC, AC and FL can be used to estimate GA using low-cost ultrasound devices in the standard plane within the inter-observer variability presented in the literature. The OSP can be used to estimate the GA by measuring the HC and the AC, but not the FL. This study shows that it is feasible to estimate GA in resource-limited countries with low-cost ultrasound devices using the OSP. This makes it possible to estimate GA and assess fetal growth for pregnant women in rural areas of resource-limited countries.

3.1 Introduction

Worldwide, 99% of all maternal deaths occur in resource-limited countries[1]. Ultrasound can be used to manage obstetric care, but too often remains out of reach for pregnant women in resource-limited countries. There are two main reasons for this. Firstly, ultrasound is too expensive for resource-limited countries. Secondly, a trained sonographer is required to acquire and interpret the ultrasound images. However, there is a severe shortage of well-trained sonographers in resource-limited countries[3–5].

The first problem could be solved with the use of low-cost ultrasound devices. Estimation of gestational age (GA) could be helpful in resource-limited countries[31–38], but it has never been shown how accurate fetal biometrics can be estimated with low-cost ultrasound systems. In this study we therefore compared three low-cost ultrasound devices to measure the head circumference (HC), abdominal circumference (AC) and femur length (FL) by obtaining the standard planes, as described by Verburg et al.[39]. The biparietal diameter was not evaluated in this study, because guidelines state that the HC is more reliable when the head shape is flattened or rounded[40].

The second problem could be solved by using the obstetric sweep protocol (OSP). The OSP was introduced by DeStigter et al.[6] and consists of six predefined free-hand sweeps over the abdomen of the pregnant women with an ultrasound transducer as visualized in Figure 3.1. According to DeStigter et al.[6], the OSP can be taught, within a day, to any health care worker without any prior knowledge of ultrasound, which makes it a suitable approach for resource-limited countries.

We investigated if it is possible to estimate the GA using the OSP. "Correct assessment of GA and fetal growth is essential for optimal obstetric management"[39]. The GA can, for example, be used to estimate due date, to schedule prenatal care and to estimate fetal viability. However, the OSP will most likely not contain the correct standard plane to obtain the fetal biometrics. Therefore, we investigated whether it is possible to accurately estimate the HC, AC and FL by manually selecting the frame within the OSP that best resembles the standard plane. If this is possible, computer-aided detection systems could potentially be used to automatically measure these biometrics. Such a system could make ultrasound more widely available in resource-limited countries, because there would be no need for a trained sonographer to acquire and interpret the image to estimate the GA and monitor growth of the fetus.

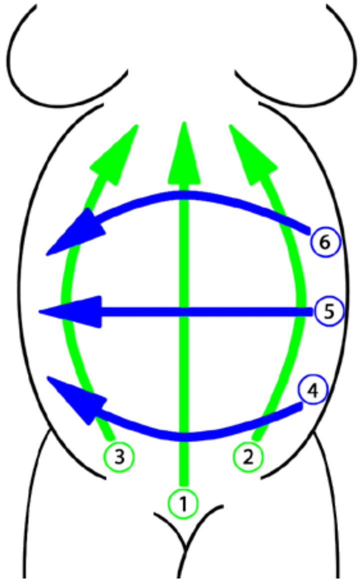

Figure 3.1: Visualization of the obstetric sweep protocol, consisting of six predefined free-hand sweeps with the ultrasound transducer over the abdomen of the pregnant woman.

3.2 Materials and Methods

3.2.1 Data acquisition

Four different ultrasound devices were used to acquire the data for this comparison study: (i) the high-end Voluson E10 in combination with the RM6C transducer (General Electric, Zipf, Austria), which can be purchased around $100,000; (ii) the low-cost MicrUs EXT-1H in combination with the C5-2R60S-3 transducer (TELEMED, Vilnius, Lithuania); (iii) the low-cost SeeMore USB Probe GP 3.5 MHz (Interson Medical Instruments, Pleasanton, USA), (both the MicrUs and SeeMore are approved by the U.S. Food and Drug Administration (FDA) and are commercially available for between $2000 and $3000); (iv) the custom developed very low-cost SESAS (Newcastle University, Newcastle upon Tyne, United Kingdom), which production costs are around $100 and provides conformance to the FDA Track 1 standards—Fetal Imaging application—and is described in detail elsewhere[41]. All three low-cost ultrasound devices were connected to a laptop using USB, thus providing a portable solution for rural areas in resource-limited countries.

All 60 participants in this study received a routine ultrasound examination[42] performed by one of three sonographers (D.d.B., D.M. and A.B.), with 27, 14 and 30 years of experience as a sonographer, respectively. The routine ultrasound examinations were performed between December 2016 and March 2017 at the Department of Obstetrics and Gynecology, Radboud University Medical Center, Nijmegen, the Netherlands. During this examination, the standard planes for obtaining the HC, AC and FL measurements were acquired using the Voluson E10 according to the standards of Verburg et al. [39]. After completion of the examination, the OSP was performed using the Voluson E10. In addition, the three standard planes and the OSP were acquired using one of the three low-cost ultrasound devices. This resulted in three 20-participants groups matched on body mass index of the participant and GA of the fetus. Data were acquired at either 20 or 33 weeks GA, because these are standard time points of routine ultrasound screening for pregnant women in the Netherlands. Only participants with a fetus that did not show any growth abnormalities were included in this study. All ultrasound devices were tested for electrical safety and the SESAS was also tested on acoustic output power to ensure patient safety. All the participants signed an informed consent form approved by the local ethics committee. All data was anonymized according to the tenets of the Declaration of Helsinki.

3

Figure 3.2 shows an example image of the standard plane for obtaining the HC, AC and FL for four different fetuses with a GA around 20 weeks using the four different ultrasound devices.

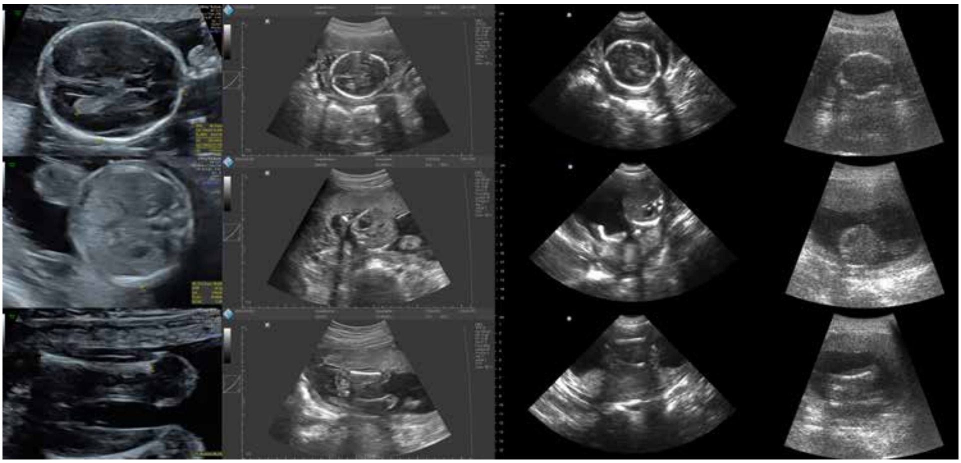

Figure 3.2: Example images of the standard plane. From left to right: Voluson, MicrUs, SeeMore, SESAS. From top to bottom: standard plane to obtain the HC, AC and FL.

Table 3.1: Preset per ultrasound device for the acquisition of the OSP

	Voluson E10	MicrUs	SeeMore	SESAS
Imaging depth	15 cm	15 cm	15 cm	15 cm
Focal depth	8 cm	8.5 cm	7.5 cm	Full depth∗
Imaging Angle	65°	65°	90°	50°
Frame rate	33 Hz	20 Hz	15 Hz	4 Hz
Center frequency	4 MHz	3.5 MHz	3.5 MHz	4.2 MHz

Note: ∗This device uses synthetic aperture focusing[41]

To make the comparison between the different ultrasound devices as fair as possible, a pre-set was defined for each device to minimize the influence of the acquisition protocol on the results. The settings of this pre-set per device can be found in Table 3.1. Note that not all parameters are the same, because some parameters cannot be changed for the low-cost ultrasound devices. The sonographer was asked to acquire around 100 frames per sweep, but since these sweeps were made in free-hand mode, the number of acquired frames per sweep varied. It was not possible to acquire 100 frames with the SESAS, as this device has a frame rate of only 4 Hz. Instead, the sonographer was asked to acquire the sweep with the SESAS within seven to ten seconds, to limit possible motion of the fetus during the acquisition of the OSP. This resulted in 30 to 40 frames per sweep for this device.

3.2.2 Biometric measurements obtained in the standard plane

Measurements of the HC, AC and FL, obtained in the standard planes using the high-end ultrasound device were determined during the routine ultrasound examination and were used as a reference to compare the measurements of the three low-cost ultrasound devices obtained in the standard plane. The HC, AC and FL measurements obtained in the standard planes using the low-cost ultrasound devices were manually determined by one experienced sonographer (D.d.B.). These measurements were obtained at least one week after the routine ultrasound examination, to avoid a bias towards the measurements obtained using the high-end device. During this process, the sonographer was blinded to the measurement obtained using the high-end device.

3.2.3 Biometric measurements obtained utilizing the OSP

The OSP data will most likely not contain the standard plane normally used to measure the HC, AC and FL. Instead the sonographer selected, from the sweep data, the two frames that best resembled the standard planes to obtain the HC and AC measurement. The HC and AC were manually annotated after these frames were selected. The FL was annotated by selecting the ends of the femur over multiple frames. Annotations were made at least one week after the HC, AC and FL were obtained in the standard planes, to avoid a bias towards the measurements obtained in the standard planes. During the annotation process, the sonographer was blinded to the measurements obtained in the standard planes obtained using both the high-end and low-cost ultrasound devices.

3.2.4 Estimation of the gestational age

The HC, AC and FL can be used to estimate the GA. The curve of Verburg et al.[39] was used to estimate the GA from each HC, AC and FL measurement. The crown-rump length (CRL), obtained between 8^{+4} weeks and 12^{+6} weeks, was used as the reference GA. Only fetuses with a reference GA below 23 weeks were used to compare the GA, because the 95% confidence interval for GA prediction using biometric parameters becomes more than one week after 23 weeks[43].

3.2.5 Comparison of the results

The biometric measurement obtained using both the standard plane and the OSP were compared to the inter-observer variability presented in literature to determine whether it is possible to obtain a measurement with an ultrasound device..

The 95% limits of agreement (LoA) for the GAs estimated from the HC, AC and FL were compared to the LoA obtained from the curve of Verburg et al.[39]. When the LoA for the GA fell within the LoA of the curve of Verburg et al.[39], we concluded that it was possible to estimate the GA with an ultrasound device. The LoA for the GAs were calculated using the formula of Hayes and Krippendorff[44]. The LoA of the curves of Verburg et al.[39] were caused by differences in fetal growth during the pregnancy and inter-observer variability of the sonographers. The standard deviation (SD) reported by Verburg et al.[39] was used to determine the LoA for the HC, AC and FL. The SD reported by Verburg et al.[39] was dependent on the GA, so the GA determined by the CRL was used to determine the SD for the participants scanned with each ultrasound device.

3.2.6 Statistical analysis

Paired statistical tests were performed to test if the measures (HC, AC and FL) obtained in either the standard plane or utilizing the OSP were significantly different

($p<0.05$) from the measurement obtained in the standard plane using the Voluson E10. A paired t-test was used when the data was normally distributed according to the Shapiro-Wilk test. When this was not the case, the Wilcoxon Signed Rank test was used. The same paired statistical tests were performed to test if the GA estimated from the HC, AC or FL obtained in either the standard plane or utilizing the OSP significantly differed ($p<0.05$) from the GA estimated from the CRL. The paired statistical tests were also used to test if the results between the low-cost ultrasound devices and the Voluson E10 were significantly different. Unpaired statistical tests were performed to test if the results between the low-cost ultrasound devices significantly differed ($p<0.05$). An independent t-test was used when the data were normally distributed according to the Shapiro-Wilk test. When this was not the case, the Mann-Whitney U-Test was used.

3.3 Results

A total of 60 participants were included in this study. Table 3.2 lists maternal age and body mass index for the participants and the GA of the fetus. There are no significant differences between the groups. A total of 348 biometric measurements were obtained in the standard planes. The sonographer could measure the HC, AC and FL in the standard plane for all participants using the Voluson, MicrUs and SeeMore. With the use of the SESAS, the sonographer could measure the HC, AC and FL in 19, 17 and 12 participants, respectively. The AC of one fetus was difficult to measure using the Voluson, due to the position of the fetus (GA of 32^{+3} weeks). The AC of this fetus measured using the MircUs resulted in an outlier which was excluded from the results. A total of 339 measurements were obtained utilizing the OSP. The sonographer could measure the HC in all participants using the MicrUs, SeeMore and SESAS. One HC could not be measured using the Voluson, because the fetus was low-lying and the OSP was acquired too high on the abdomen. The sonographer could measure the AC in all participants using the Voluson and the MicrUs. One AC could not be measured using the SeeMore, because it did not fall completely within the FOV of any of the six sweeps due to the small footprint of the transducer. Three ACs could not be measured using the SESAS, because the number of frames per sweep in combination with the lower signal to noise ratio made it impossible to detect the borders of the fetal abdomen. The sonographer could measure the FL in all participants using the SeeMore and Voluson. One FL could not be measured using the MicrUs, because the femur was not visible due to the position of the fetus (GA of 32^{+5} weeks). Fifteen FLs could not be measured using the SESAS, because the number of frames per sweep was too low to be able to detect the femur. A total of 45 participants (15 per low-cost device) had a GA below 23 weeks according to the CRL measurement in the first trimester and were therefore included in the comparison of the GA.

Table 3.2: Maternal age and body mass index of participants and GA of the fetus

	All (N=60)	MicrUs (N=20)	SeeMore (N=20)	SESAS (N=20)
Maternal age (years)	31.1±0.1	31.7±0.2	29.9±0.2	31.7±0.2
Body mass index	23.2±2.5	23.0±2.4	22.8±2.5	23.9±2.6
Gestational age (weeks)	23.3±5.3	23.3±5.4	23.3±5.3	23.5±5.4

3.3.1 Biometric measurements obtained in the standard plane

Figure 3.3 shows a scatterplot for the HC, AC and FL measurements obtained in the standard plane. The x-axis shows the reference measurement obtained using Voluson E10 in the standard plane. The y-axis shows the measurement obtained using the three low-cost ultrasound devices in the standard plane. The legend shows how many measurements were obtained using each ultrasound device.

Table 3.3 lists the difference (mean ± SD) between the measurements obtained in the standard plane using the Voluson E10 and the measurements obtained in the standard plane using the three low-cost devices. The differences were computed in millimeter and in percentage. All three low-cost devices significantly overestimated the HC. The SeeMore also significantly overestimated the AC. The SeeMore and MicrUs significantly underestimated the FL.

Table 3.3: Mean and SD difference of the HC, AC and FL in millimeter and percentage between the Voluson obtained in the standard plane and the three low-cost ultrasound devices obtained in the standard plane

	Mean±SD (mm)			Mean±SD (%)		
	HC	AC	FL	HC	AC	FL
SESAS	8.5±10.3∗†	4.3±12.3	-1.2±3.4	4.0±4.7∗†	3.0±7.3	-2.6±7.0
SeeMore	10.5±3.7∗†	8.4±6.8∗†	-1.6±1.9∗	5.3±2.2∗†	4.7±3.3∗†	-3.7±4.4∗
MicrUs	2.4±4.0∗	0.6±7.4	-2.2±2.0∗	1.1±1.8∗	0.6±4.0	-5.5±5.3∗

Note: ∗significantly different compared to the Voluson in the standard plane, †significantly different compared to the MicrUs in the standard plane

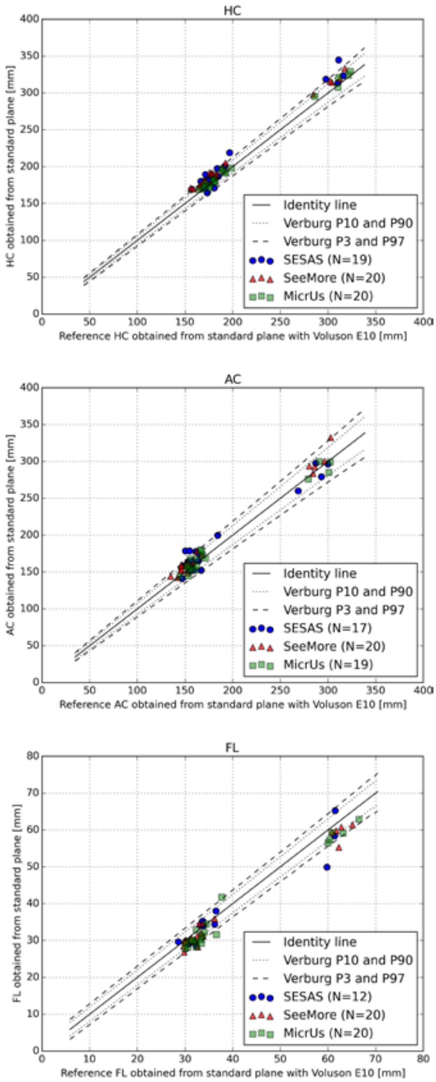

Figure 3.3: Scatterplot of the measurement obtained using the Voluson in the standard plane compared to the measurement obtained using the low-cost ultrasound devices in the standard plane. From top to bottom: measurements HC, AC and FL.

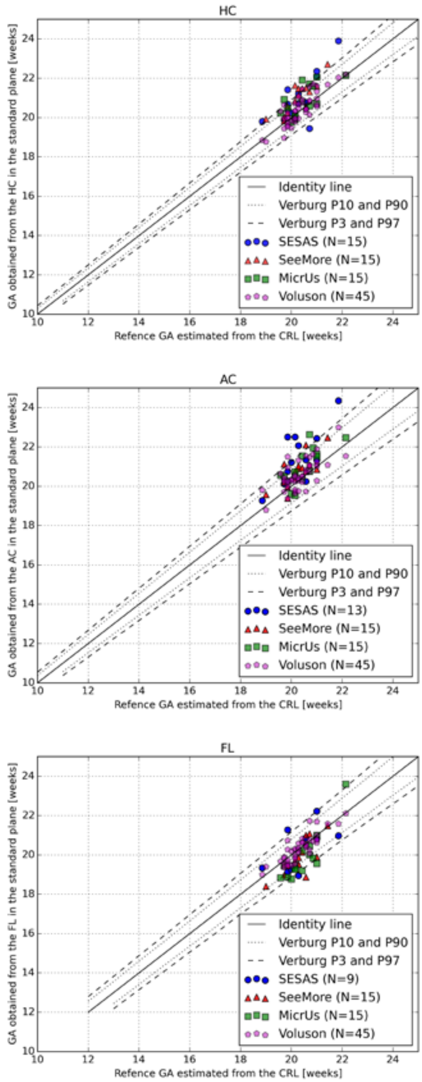

Figure 3.4: Scatterplot of the GA estimated from the CRL compared to the measurement obtained using the four ultrasound devices in the standard plane. From top to bottom: GA estimated from HC, AC and FL.

Figure 3.4 shows a scatterplot of GA estimates based on the HC, AC and FL measurements obtained in the standard plane. The x-axis shows the reference GA estimated from the CRL measurement in the first trimester. The y-axis shows the GA estimated from the measurement obtained using all four ultrasound devices in the standard plane. The legend shows how many measurements were obtained using each ultrasound device.

On the left side of Table 3.4 are the LoA for the GA estimated from the measurements obtained in the standard plane using all four ultrasound devices. On the right side of Table 3.4 are the LoA for the GA estimated with the Verburg curve for the same participant groups. The GA estimated from the HC using the low-cost ultrasound devices is significantly higher than the GA estimated from the CRL. The GA estimated from the AC is significantly higher for all four ultrasound devices than the GA estimated from the CRL. The GA estimated from the FL using the SeeMore and MicrUs are significantly lower compared to the GA estimated from the FL using the Voluson E10 and the CRL.

Table 3.4: Left side: limits of agreement (LoA) for the GA estimated from the HC, AC and FL using the ultrasound devices obtained in the standard plane compared to the GA estimated from the CRL measurement. Right side: LoA for the curve of Verburg et al.[39] for each device

	LoA in the standard plane (days)			LoA of Verburg curve (days)		
	HC	**AC**	**FL**	**HC**	**AC**	**FL**
SESAS	-7.4 to 15.8∗§	-5.9 to 21.8∗†	-15.1 to 14.4	-6.6 to 6.7	-9.3 to 9.4	-8.4 to 8.7
SeeMore	-1.0 to 11.4∗§	-4.1 to 11.3∗§	-10.9 to 4.5∗§	-6.6 to 6.7	-9.3 to 9.4	-8.4 to 8.6
MicrUs	-5.9 to 11.5∗	-5.7 to 12.8∗	-14.4 to 6.0∗§	-6.7 to 6.8	-9.4 to 9.5	-8.5 to 8.7
Voluson	-5.4 to 6.7	-6.1 to 9.5∗	-5.0 to 4.8	-6.6 to 6.7	-9.4 to 9.4	-8.4 to 8.7

Note: ∗significantly different from GA estimated from CRL, †significantly different from GA estimated using the MicrUs in the standard plane, ‡significantly different from GA estimated using the SeeMore in the standard plane, §significantly different from GA estimated using the Voluson in the standard plane

3.3.2 Biometric measurements obtained utilizing the OSP

Figure 3.5 shows a scatterplot for the HC, AC and FL measurement measurements obtained utilizing the OSP. The x-axis shows the reference measurement obtained using Voluson in the standard plane. The y-axis shows the measurement obtained using all four ultrasound devices utilizing the OSP. The legend shows how many measurements were obtained using each ultrasound device.

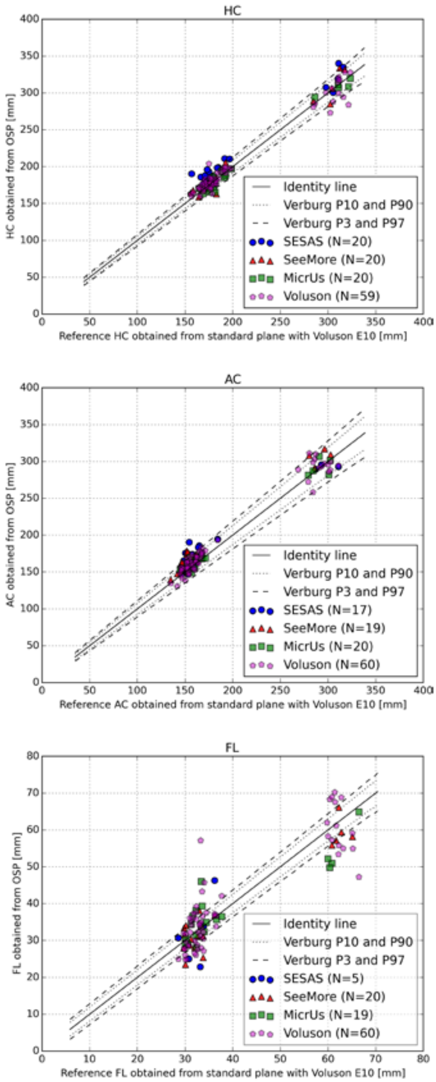

Figure 3.5: Scatterplot of the measurement obtained using the Voluson in the standard plane compared to the measurement obtained using the four ultrasound devices utilizing the OSP. From top to bottom: measurements HC, AC and FL.

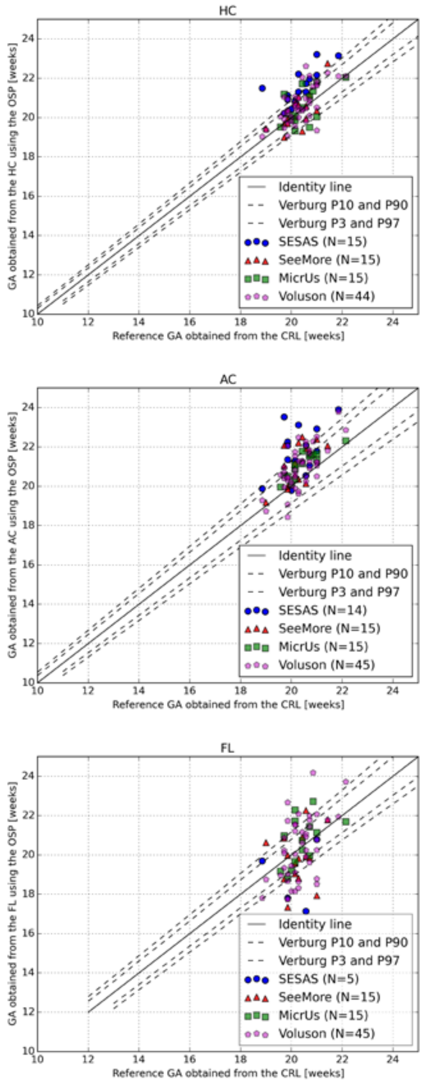

Figure 3.6: Scatterplot of the GA estimated from the CRL compared to the measurement obtained using the four ultrasound devices utilizing the OSP. From top to bottom: GA estimated from HC, AC and FL.

Table 3.5 lists the differences (mean ± SD) between the measurements obtained in the standard plane using the Voluson E10 and the measurements obtained utilizing the OSP using all four ultrasound devices. The differences were computed in millimeter and in percentage. The SESAS significantly overestimated the HC compared to the HC obtained in the standard plane using the Voluson E10. The mean difference in HC using the SESAS is significantly higher compared to the mean difference in HC using the other three ultrasound devices. The SESAS, SeeMore and Voluson significantly overestimated the AC compared to the AC obtained in the standard plane using the Voluson E10. The mean difference in AC using the SeeMore is significantly higher compared to the mean difference in AC using the MicrUs and the Voluson E10.

Table 3.5: Mean and SD difference of the HC, AC and FL in millimeter and percentage between the Voluson obtained in the standard plane and the three low-cost ultrasound devices obtained utilizing the OSP

	Mean±SD (mm)			**Mean±SD (%)**		
	HC	**AC**	**FL**	**HC**	**AC**	**FL**
SESAS	13.7±8.7∗†‡§	8.7±12.6∗	-0.8±7.0	7.1±4.9∗†‡§	5.7±7.5∗	-3.0±20.5
SeeMore	2.2±9.5	12.7±8.7∗†§	-1.5±3.8	1.0±4.4	7.4±5.2∗†§	-3.4±10.3
MicrUs	-2.3±6.6	2.4±7.7	-0.4±5.4	-1.1±3.4	1.6±3.9	1.0±13.3
Voluson	0.4±9.9	3.3±10.1∗	0.1±6.3	0.6±4.3	2.0±4.9∗	0.8±15.8

Note: ∗significantly different compared to the Voluson in the standard plane, †significantly different compared to the MicrUs utilizing the OSP, ‡significantly different compared to the SeeMore utilizing the OSP, §significantly different compared to the Voluson utilizing the OSP

Figure 3.6 shows a scatterplot for GA estimation obtained from the HC, AC and FL measurements obtained utilizing the OSP. The x-axis shows the reference GA estimated from the CRL measurement in the first trimester. The y-axis shows the GA estimated from the measurement obtained using all four ultrasound devices utilizing the OSP. The legend shows how many measurements were obtained using each ultrasound device.

On the left side of Table 3.6 are the LoA for the GA estimated from the measurements obtained utilizing the OSP using all four ultrasound devices. On the right side of Table 3.6 are the LoA for the GA estimated with the Verburg curve for the same groups. The GA estimated from the HC utilizing the OSP using the SESAS is significantly worse compared to the GA estimated from the CRL and using the other three ultrasound device utilizing the OSP. The GA estimated from the CRL is significantly different from the GA estimated from the AC utilizing the OSP for all four ultrasound devices.

Table 3.6: Left side: limits of agreement (LoA) for the GA estimated from the HC, AC and FL using the ultrasound devices obtained utilizing the OSP compared to the GA estimated from the CRL measurement. Right side: LoA for the curve of Verburg et al.[39] for each device

	LoA in the standard plane (days)			LoA of Verburg curve (days)		
	HC	**AC**	**FL**	**HC**	**AC**	**FL**
SESAS	-0.3 to 17.5∗†‡§	-5.7 to 26.9∗†§	-48.5 to 44.8	-6.6 to 6.7	-9.3 to 9.4	-8.5 to 8.7
SeeMore	-9.1 to 9.7	-5.9 to 18.5∗§	-22.8 to 16.9	-6.6 to 6.7	-9.3 to 9.4	-8.4 to 8.6
MicrUs	-10.4 to 11.6	-0.5 to 9.4∗	-17.2 to 24.8	-6.7 to 6.8	-9.4 to 9.5	-8.5 to 8.7
Voluson	-7.5 to 11.4∗†	-7.1 to 15.2∗	-27.7 to 31.7∗	-6.6 to 6.7	-9.4 to 9.4	-8.4 to 8.7

Note: ∗significantly different from GA estimated from CRL †significantly different from GA estimated using MicrUs utilizing the OSP ‡significantly different from GA estimated using SeeMore utilizing the OSP §significantly different from GA estimated using Voluson utilizing the OSP

3.3.3 Comparison to literature

Table 3.7 shows a literature overview of the inter-observer variability for the HC, AC and FL measurements. Some cells are empty, because some papers present the results in millimeters and some in percentages. In addition, not all papers show results of all three biometric measurements.

Table 3.7: Literature overview of the inter-observer variability for the HC, AC and FL

		Mean±SD (mm)			Mean±SD (%)		
	N	**HC**	**AC**	**FL**	**HC**	**AC**	**FL**
Sarmandal et al.[45]	22	-0.1±8.9	-0.6±7.7	-1.3±2.3			
Perni et al.[46]	122	0.1±5.6	1.0±11.6	0.4±1.9			
Rijken et al.[47]	90	-1.6±4.8	-0.6±5.7	-0.4±1.4			
Lima et al.[48]	102		0.0±13.0	0.0±1.1			
Chang et al.[49]	40	-1.6±5.8	-1.9±7.5	-0.1±1.4	-0.5±1.9	-0.5±2.4	-0.2±2.2
Sarris et al.[50]	175	0.9±6.1	0.9±10.7	0.0±2.2	0.5±2.5	1.2±2.9	0.0±5.7
Verburg et al.[51]	20				1.3±5.4	0.3±5.6	-1.4±5.1
Napolitano et al.[52]	100				-0.8±2.5		

3.4 Discussion

We have shown the feasibility of measuring the HC, AC and FL with low-cost ultrasound devices using both standard planes and OSP. The results indicate that the HC, AC and FL measurements obtained in the standard planes with the low-cost ultrasound devices are similar to the inter-observer variability presented in the literature. The results also indicate that it is possible to measure the HC and AC utilizing the OSP to estimate GA.

3.4.1 Biometric measurements obtained in the standard plane

The HC was overestimated using all three low-cost devices in the standard plane. This could be caused by the lower image quality, which made it more difficult to determine the correct standard plane and accurately delineate the HC. The difference between the MicrUs and the reference was only 2.4 mm (1.1%), which is significantly better compared to the HC measured with the SESAS and SeeMore and falls within inter-observer variability presented in Table 3.7. Due to this overestimation, the upper limit of the LoA for GA estimated using the low-cost devices was increased, while the LoA interval remained close to the 13.3 days of the Verburg curve.

The AC measured using the MicrUs and SeeMore show similar results compared to the inter-observer variability presented in the literature, which indicates that it is possible to measure the AC with these low-cost devices. The LoA interval of the GA was 15.4 and 18.5 days for the SeeMore and MicrUs, which is smaller compared to the LoA interval of the Verburg curve, but the estimated GA from the AC was significantly higher compared to the GA estimated from the CRL for all four devices. Since the AC was also overestimated with the Voluson E10, we conclude that the average AC of the fetuses in this study population was larger compared to the population average. The AC could not be measured in three of the twenty participants using the SESAS and the LoA interval for the estimated GA for the remaining participants was 27.7 days, which is larger compared to the LoA interval of the Verburg curve.

It was not possible to measure the FL in the standard plane using the SESAS in eight of the twenty participants. This was caused by low frame rate of the SESAS, which made it very difficult to image the femur of a moving fetus. The GA estimated from the FL using the SeeMore and MicrUs was significantly lower compared to the GA estimated from the FL using the Voluson E10. This indicates that these two low-cost devices underestimate the FL and therefore underestimates the GA. The results show that the LoA interval therefore increases, but this increase is only three days compared to the LoA interval of the Verburg curve. Therefore, we conclude that the FL can be measured using the SeeMore and MicrUs.

3.4.2 Biometric measurements obtained utilizing the OSP

It is possible to measure the HC using the SeeMore, MicrUs and Voluson E10 utilizing the OSP, because the difference between the HC measured utilizing the OSP and the HC measured using the Voluson E10 in the standard plane is close to the inter-observer variability presented in literature. The LoA interval for the HC obtained utilizing the OSP was 22.0 days, which is nine days larger compared to the 13.3 days of the Verburg curve.

The AC obtained utilizing the OSP is significantly higher compared to the AC measured using the Voluson E10 in the standard plane. The OSP will most likely not contain the standard plane and will therefore result in an obliquely section of the abdomen. This results in a larger AC compared to the standard plane, but the AC measured with the Voluson E10 and MicrUs utilizing the OSP still fall within the inter-observer variability. The LoA interval for the AC obtained utilizing the OSP was 24.4 days, which is six days larger compared to the 18.8 days of the Verburg curve.

The SESAS could not be used to accurately measure the HC and AC utilizing the OSP. This was caused by the limited number of frames within the OSP data of the SESAS in combination with the lower contrast sensitivity, which made it more difficult to select the correct frame to obtain an accurate measurement.

The results indicate that it was not possible to accurately measure the FL utilizing the OSP. The OSP will most likely not contain the standard plane to measure the FL. A random cross section through the femur bone will differ substantially from the FL measured in the standard plane and will therefore not give an accurate estimation of the FL.

3.4.3 Study limitations

The GA of the acquired data ranged from 18^{+6} to 33^{+0} weeks, so the feasibility for measuring the HC and AC utilizing the OSP in the first trimester could not be investigated. Data from the first trimester would be required for this, but it should be noted that most women in resource-limited countries will not receive an ultrasound examination in the first trimester of their pregnancy.

3.4.4 Clinical implications

The results show that the OSP can be used to measure the HC and AC for estimation of GA with the use of low-cost ultrasound devices. In this work, a well-trained sonographer was still required to interpret the OSP data and manually obtain the biometric measurements. In the future, computer aided detection systems could be used to automatically measure these biometrics. This would obviate the need of a well-trained sonographer to both obtain and interpret the data for estimation of GA and monitoring fetal growth.

3.5 Conclusions

We show that it is possible to accurately estimate GA with low-cost ultrasound devices using both the standard plane and the OSP. The results indicate that a trained sonographer was able to determine the standard plane and measure the HC, AC and FL to estimate GA using the SeeMore and MicrUs within the inter-observer variability presented in literature. The SESAS can be used to measure the HC and AC to estimate GA, but showed a larger standard deviation. This study also shows that the OSP can be used to accurately estimate GA by measuring the HC and AC, but not the FL. Since the OSP can be taught to health care workers without prior knowledge of ultrasound within one day, it is feasible to estimate GA and assess fetal growth with low-cost ultrasound devices without training dedicated sonographers. In the future, computer-aided detection systems could be used to automatically measure these biometrics. This would obviate the need of a well-trained sonographer to both obtain and interpret the data for estimation of GA and monitor fetal growth.

4

Automated measurement of fetal head circumference

T. L. A. van den Heuvel
D. de Bruijn
C.L. de Korte
B. van Ginneken

Original title:
Automated Measurement of Fetal Head Circumference using 2D Ultrasound Images

Published in:
PLoS One

Abstract

In this chapter we present a computer aided detection (CAD) system for automated measurement of the fetal head circumference (HC) in 2D ultrasound images for all trimesters of the pregnancy. The HC can be used to estimate the gestational age and monitor growth of the fetus. Automated HC assessment could be valuable in developing countries, where there is a severe shortage of trained sonographers. The CAD system consists of two steps: First, Haar-like features were computed from the ultrasound images to train a random forest classifier to locate the fetal skull. Secondly, the HC was extracted using Hough transform, dynamic programming and an ellipse fit. The CAD system was trained on 999 images and validated on an independent test set of 335 images from all trimesters. The test set was manually annotated by an experienced sonographer and a medical researcher. The reference gestational age (GA) was estimated using the crown-rump length measurement (CRL). The mean difference between the reference GA and the GA estimated by the experienced sonographer was 0.8 ± 2.6, -0.0 ± 4.6 and 1.9 ± 11.0 days for the first, second and third trimester, respectively. The mean difference between the reference GA and the GA estimated by the medical researcher was 1.6 ± 2.7, 2.0 ± 4.8 and 3.9 ± 13.7 days. The mean difference between the reference GA and the GA estimated by the CAD system was 0.6 ± 4.3, 0.4 ± 4.7 and 2.5 ± 12.4 days. The results show that the CAD system performs comparable to an experienced sonographer. The presented system shows similar or superior results compared to systems published in literature. This is the first automated system for HC assessment evaluated on a large test set which contained data of all trimesters of the pregnancy.

4.1 Introduction

Ultrasound imaging is widely used for screening and monitoring of pregnant women, since it is a low-cost, real-time and non-invasive imaging method. However, acquisition of ultrasound images is operator-dependent and the images are characterized by attenuation and speckle and may contain artifacts such as shadows and reverberations, making their interpretation complex. During the ultrasound screening examination, biometric measurements of the fetus such as the crown-rump length (CRL) and the head circumference (HC) are often computed to determine the gestational age (GA) and to monitor growth of the fetus. The CRL is the most accurate measurement for estimating the GA of the fetus between 8 weeks and 4 days (commonly noted as: 8^{+4} weeks) and 12^{+6} weeks. After 13 weeks, the HC is used the most accurate measurement to determine the GA, because it is not possible to accurately measure the CRL anymore. The guidelines state that HC should be measured in a transverse section of the head with a central midline echo, interrupted in the anterior third by the cavity of the septum pellucidum with the anterior and posterior horns of the lateral ventricles in view[39]. The biometric measurements are obtained manually, which leads to inter- and intra-observer variability. An accurate automated system could reduce measuring time and variability, because it does not suffer from intra-observer variability. Worldwide, 99% of all maternal deaths occur in developing countries. Skilled care before, during and after childbirth can save the lives of women and newborn babies[1]. Unfortunately, there is still a severe shortage of well-trained sonographers in low resource settings. This keeps ultrasound screening out of reach for most pregnant women in these countries[3]. An automated system could assist inexperienced human observers in obtaining an accurate measurement. In this work, we focus on measuring the HC because this measurement can be used to determine the GA and monitor growth of the fetus. In addition, the fetal head is more easily detectable compared to the fetal abdomen.

Systems for automatic HC measurement have been presented using randomized Hough transform[53,54], Haar-Like features[55–58], multilevel thresholding[59], circular shortest paths[60], boundary fragment models[61], semi-supervised patch based graphs[62], active contouring[63,64], intensity based features[65] and texton based features[66]. Although these methods show promising results, they were evaluated on a relatively small amount of data (10 to 175 test images). Furthermore, none of these papers used images of fetuses from all trimesters of pregnancy. We present a system that was developed using 999 ultrasound images and evaluated on a large independent test set of 335 ultrasound images from all trimesters. The presented quantification system was designed to be as fast and robust as possible and the results were compared to the methods presented in literature. A complete overview of the comparison between our method and previous publications is presented in Section 4.4.7.

4.2 Materials and Methods

4.2.1 Data

A total of 1334 two-dimensional (2D) ultrasound images of the HC were collected from the database of the Department of Obstetrics of the Radboud University Medical Center, Nijmegen, the Netherlands. The ultrasound images were acquired from 551 pregnant women who received a routine ultrasound screening exam between May 2014 and May 2015. Only fetuses that did not exhibit any growth abnormalities were included in this study. Images were acquired by experienced sonographers using either the Voluson E8 or the Voluson 730 ultrasound device (General Electric, Austria). The local ethics committee (CMO Arnhem-Nijmegen) approved the collection and use of this data for this study. Due to the retrospective data collection, informed consent was waived. All data was anonymized according to the tenets of the Declaration of Helsinki.

The size of each 2D ultrasound image was 800 by 540 pixels with a pixel size ranging from 0.052 to 0.326 mm. This large variation in pixel size is a result of ad-

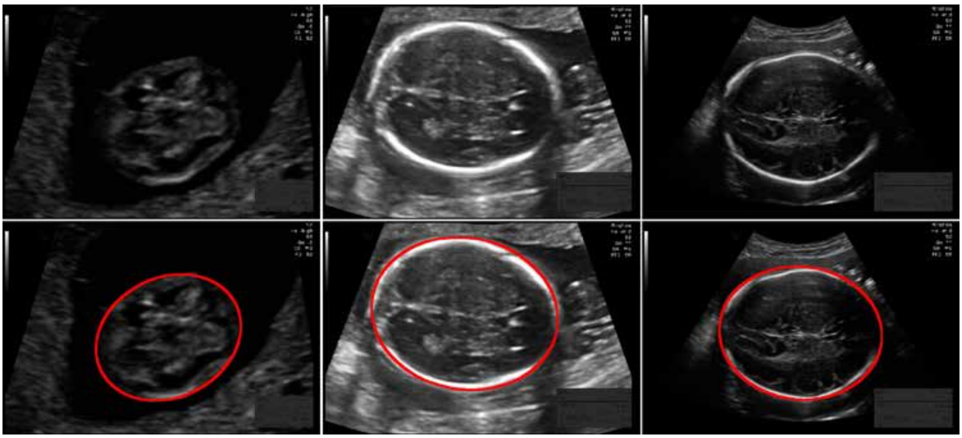

Figure 4.1: Example ultrasound images. From top to bottom: without annotation and with annotation in red. From left to right: first trimester with an HC of 65.1 mm (pixel size of 0.06 mm), second trimester with an HC of 167.9 mm (pixel size of 0.12 mm) and third trimester with an HC of 278.4 mm (pixel size of 0.24 mm). Note that the skull is not yet visible as a bright structure in the first trimester.

justments in the ultrasound settings by the sonographer (depth settings and amount of zoom are routinely varied during the examination) to account for the different sizes of the fetuses. Fig 4.1 shows example ultrasound images from each trimester. The distribution of the GA in this study is shown in Fig 4.2. Most data were acquired after 12 and 20 weeks of pregnancy, since these are standard time points of routine ultrasound screening for pregnant women in the Netherlands. During each exam, the sonographer manually annotated the HC. This was done by drawing an ellipse that best fits the circumference of the head. Fig 4.2 also shows the comparison between the distribution of the HC and the growth curve of Verburg et al.[39]. The reference GA was determined with a CRL measurement between 20 mm (8^{+4} weeks) and 68 mm (12^{+6} weeks). All the HCs that fell outside the 3-97 percent confidence interval of the curve of Verburg et al.[39] were individually checked to ensure no mistakes were made during data collection.

The data was randomly divided into a training set and a test set of 75 percent and 25 percent, respectively. The GAs were proportionally balanced between the data sets as shown in Table 4.1. All images that were made during one echographic examination were assigned to either the training or the test set. An independent data set of HC annotations of the images in the test set was created by TLAvdH, a medical researcher who has a technical background in ultrasound imaging and received training by an experienced sonographer in measuring the HC.

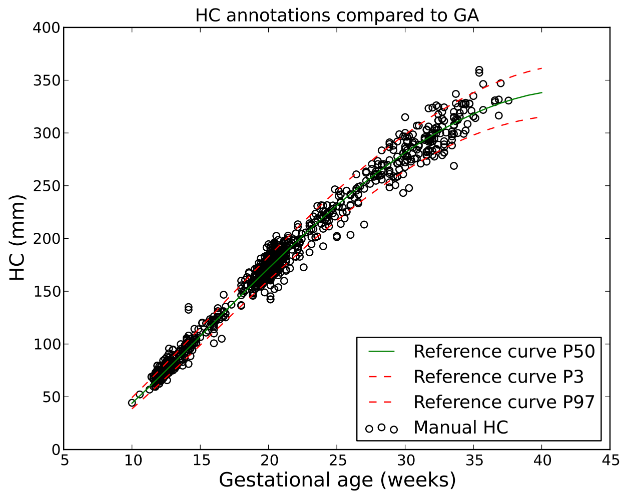

Figure 4.2: Distribution of HC and GA for the study data. The x-axis represents the GA that was estimated using the CRL. The y-axis represents the HC measured by the experienced sonographer.

Table 4.1: Number of images in the training and the test set

Trimester	**Training set**	**Test set**
First	165	55
Second	693	233
Third	141	47
Total	**999**	**335**

4.2.2 Quantification system

In this study, three variations of the quantification system, indicated as system A, B, or C, were optimized and evaluated to investigate the influence of the changing appearance of the fetal head during pregnancy on the performance of the system. An overview of the three systems is shown in Fig 4.3. All three systems contain the same two steps: First, Haar-like features were computed from the ultrasound images to train a random forest classifier (RFC) to locate the fetal skull. Next, the HC was extracted using Hough transform, dynamic programming and an ellipse fit. Both steps are described in detail in the following subsections. System A uses one pipeline that was optimized on training data from all trimesters. It can be seen in Fig 4.1 that the fetal skull is not clearly visible in the first trimester. To deal with this different appearance, system B uses two pipelines to measure the HC: one pipeline was optimized on training data from the first trimester and the other pipeline was optimized on training data from the second and third trimesters. System C uses three pipelines, which were optimized on training data from the first, second and third trimester separately. In a low-resource setting the trimester of the fetus is commonly unknown. For systems with multiple pipelines, a selection method was used to automatically select the best fitted ellipse. This allows the system to automatically measure the HC without requiring the trimester to be known in advance.

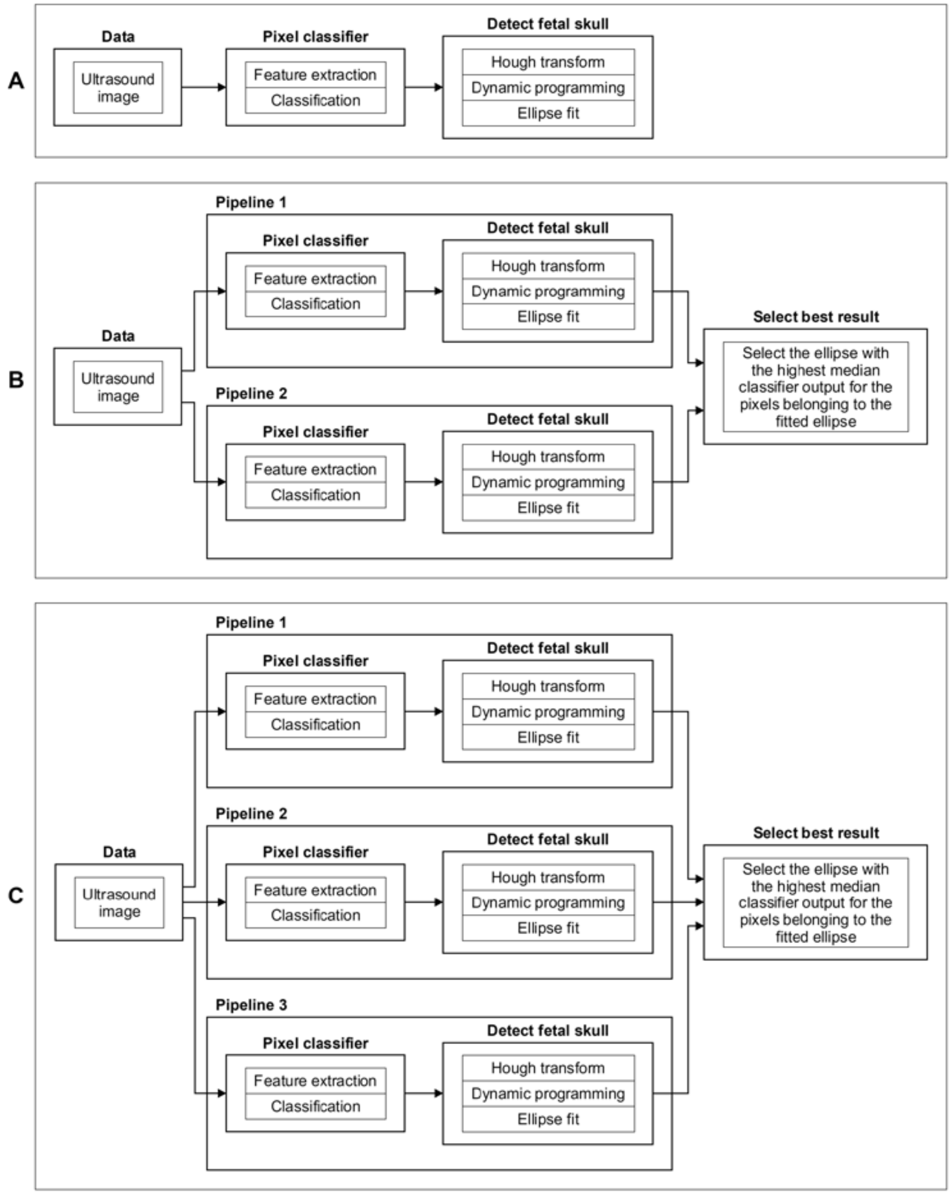

Figure 4.3: Overview of the three evaluated quantification systems A, B, and C. System A was optimized on training data from all trimesters. System B has two pipelines: pipeline 1 was optimized on training data from trimester one and pipeline 2 was optimized on training data from trimester two and three. System C uses three pipelines: pipeline 1, 2 and 3 were optimized on training data from trimester one, two and three, respectively. All pipelines of a quantification system are computed when the HC is measured in a test ultrasound image.

4

Pixel classifier

The first step of the three quantification systems consists of a pixel classifier that emphasizes the fetal skull and reduces artifacts in the ultrasound image, by computing the likelihood that each pixel in the image has of being part of the fetal skull. This makes the detection of the fetal skull in the second step more robust.

Feature extraction: Haar-like features [67] were used to be able to discriminate between background pixels and pixels that belong to the fetal skull. Viola and Jones [68] have shown that using an integral image enables the rapid computation of these features. Fig 4.4 shows the twelve different Haar-like features that were used for the pixel classification. The Haar-like features in rotated direction have a larger kernel width and height compared to the upright direction, but they capture the same relationship between the neighboring pixels. The Haar-like features were computed in different kernel sizes. To make these kernels invariant to the pixels size of the ultrasound image, all features were computed in millimeters. The pixel size of each Haar-like feature was chosen as close to the millimeter scale as possible. As a consequence, the kernel size of the Haar-like features increases when the pixel size of an ultrasound image decreases. A larger kernel size will result in a higher kernel response. To make the response of the feature independent from its kernel size, the Haar-like features were normalized. Normalization was performed by dividing the positive and negative coefficients of the kernel by their respective areas.

Classification: An OpenCV implementation of the RFC [69] was used for pixel classification. Positive samples were obtained from pixels annotated by the sonographers as the HC. The same number of negative samples were obtained from pixels randomly taken from the background with a minimal distance d_{min} from the annotation. When negative samples were obtained too close to the annotation they resemble positive samples, since the manually drawn ellipse will never fit the outer edge of the skull perfectly. This problem was solved by increasing d_{min}, which was optimized within the training set. Data augmentation was applied by flipping the ultrasound image horizontally, which resembles an acquisition with a flipped ultrasound transducer. The pixel classifier produces a likelihood map with a per pixel estimate of being part of the fetal skull. This likelihood map was visualized with a color map ranging from green to red, where a high likelihood was shown in red.

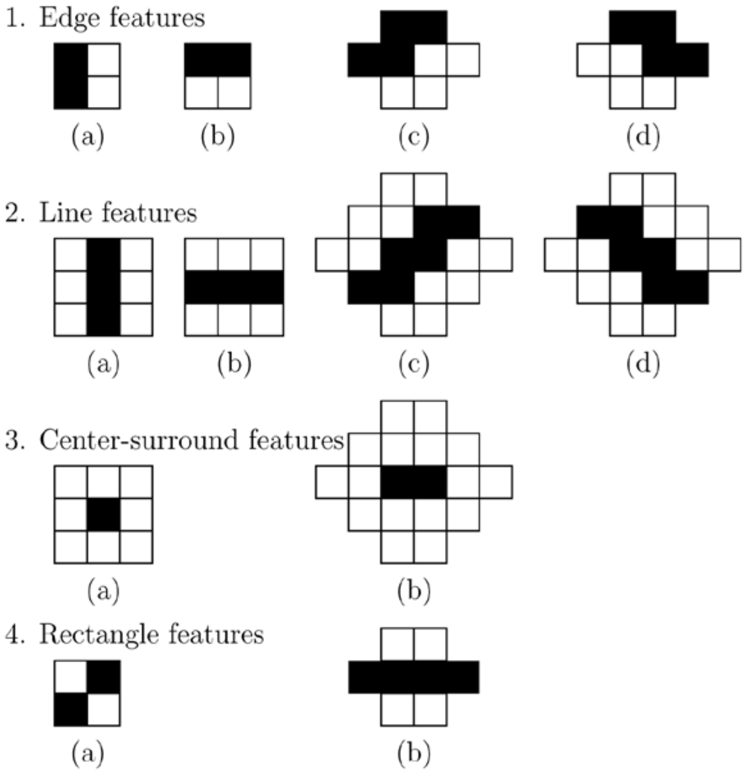

Figure 4.4: Overview of the twelve Haar-like features utilized in the quantification system. From top to bottom: 1. Edge features in horizontal and vertical direction (kernel size of two by two pixels). 2. Line features in horizontal en vertical direction (kernel size of three by three pixels). 3. Center-surround features (kernel size of three by three pixels). 4. Rectangle features (kernel size of two by two pixels). The left side of each row represents the features in upright direction. The right side of each row represents the features in rotated direction. The height and width of the features in rotated direction are larger compared to the upright direction, but they capture the same relationship between the neighboring pixels.

Detect fetal skull

The likelihood map of the pixel classifier was used to detect the fetal skull in three steps. First, a Hough transform was applied to detect the center of the fetal skull. Secondly, dynamic programming was used to detect the outside of the fetal skull. Finally, an ellipse was fitted on the result of the dynamic programming algorithm to measure the HC.

Hough transform: An itk implementation of the Hough transform algorithm[70] was used to detect the center of the fetal skull from the likelihood map of the pixel classifier. Every classification pipeline has a GA ranging from the minimum GA, GA_{min}, to the maximum GA, GA_{max}. The minimum radius, r_{min}, of each classification pipeline was set to the half of the biparietal diameter (BPD) of the GA_{min} on the P3 curve of Verburg et al.[39]. The maximum radius, r_{max}, of each classification pipeline was computed using Eq. (4.1) in which the HC and BPD are taken from the GA_{max} of the P97 curve of Verburg et al.[39]. The Hough transform was not used to measure the HC because the fitted circle will not give a good estimation of the elliptical shape of the fetal skull. Instead, the detected center was used for initialization of the dynamic programming algorithm (as explained in the next step), which is computational more efficient than fitting an ellipse using Hough transform.

$$r_{max} = \left\lceil \frac{HC}{\pi} - \frac{BPD}{2} \right\rceil \tag{4.1}$$

Dynamic programming: Dynamic programming was used to extract the pixels belonging to the outside of the fetal skull[71]. Dynamic programming was used, because it can be computed very efficiently compared to other methods like active contouring. Fig 4.5 shows a schematic example of the dynamic programming algorithm. Dynamic programming was used in a polar transform of the pixel classifier likelihood map to find the shortest path from the left to the right side of Fig 4.5.B. The polar transform uses a preset number of angles, N_{angles}, around the center point that was detected with the Hough transform algorithm. The sampling distance, S_{dis}, in radial direction was increased to make the algorithm less sensitive to noise and spurious responses in the likelihood map and to a decrease computation time. When S_{dis} becomes too large, the resolution of the polar transform decreases and eventually the dynamic programming algorithm will fail to detect the fetal skull. An optimal value for S_{dis} was determined on the training set. To make the dynamic programming algorithm less sensitive to small circular structures in the likelihood map, a radial offset of 5 mm and 10 mm was taken for the second and third trimester, respectively. According to the annotation protocol for HC measurements, the HC must be detected at the outside edge of the fetal skull[39]. Although the RFC was trained with annotations that describe the outside of the fetal skull, the Haar-like features were not able to distinguish between inside and outside of the fetal skull. Therefore, the RFC detected all pixels belonging to the fetal skull instead of only those that belong to the outside of the fetal skull. For this reason, the dynamic programming algorithm detected the midline of the skull. To solve this problem, a second dynamic programming algorithm was computed in the polar transform of the ultrasound image. This algorithm uses the same center and number of angles, N_{angles}, as the first dynamic programming algorithm, but without any downsampling in radial direction to maintain detailed information about the edge of the skull. To detect the outside of the fetal skull, the derivative of the ultrasound image in radial direction was computed. Pilot experiments showed that the fetal skull is only a few millimeters thick. To restrict the second dynamic programming algorithm to the area that is likely to contain the fetal skull, the second dynamic programming algorithm was only computed on the area within a distance of 2 mm from the first dynamic programming result. It is not advisable to directly apply dynamic programming to the derivative of the ultrasound image in radial direction because this would be overly sensitive to noise in ultrasound image. The result of the second dynamic programming algorithm, computed on the derivative of the ultrasound image, was taken as the final result for the ellipse fit in the next step.

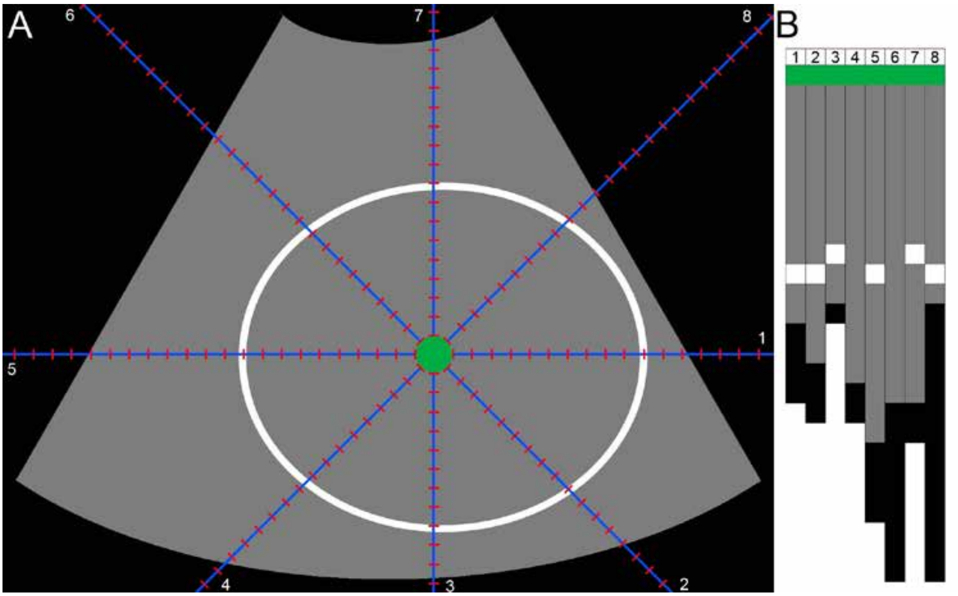

Figure 4.5: A: Perfect pixel classifier likelihood map where only the fetal skull has a high probability (depicted in white) and the background a low probability (depicted in gray). The pixels outside of the FOV are depicted in black. The center detected by the hough transform is depicted in purple and the radial offset is depicted in green. This schematic example uses eight angles (N_{angles}) for the polar transform (depicted in blue). The sampling distance (S_{dis}) is depicted in red. B: The output of the polar transform. The dynamic programming algorithm is used to extract the shortest path from left to right.

Ellipse fitting: A direct least square fitting of ellipses[72] was used to determine the HC from the extracted pixels of the dynamic programming algorithm. Only the pixels detected by the dynamic programming algorithm within the highest fifth percentile of the likelihood map of the pixel classifier were used to fit the ellipse, because these pixels have a high likelihood for being part of the fetal skull. The fitted ellipse was required to have a circumference of at least 38.6 mm. This is the smallest reported HC on the curve of Verburg et al.[39] and will therefore prevent the quantification system from detecting small circular structures or noise in the image.

Select best result

All pipelines of a quantification system were computed when the HC was measured in a test ultrasound image. In a low-resource setting the trimester of the fetus is commonly unknown, so quantification systems B and C will produce two and three fitted ellipses, respectively. To allow the system to fully automatically measure the HC, the ellipse with the highest median value of the first dynamic programming algorithm on the pixel classifier likelihood map was selected as the final result.

4.3 Experiments

Four experiments were performed to evaluate the performance of the three quantification systems and compare them to the manual annotations of the experienced sonographer (observer 1) and the medical researcher (observer 2). First, the parameters of the pipelines were optimized for each system. Secondly, the HC measured by observer 1 was used as a reference to compare the HC measured by the three systems and the HC measured by observer 2. Thirdly, the measured HCs were used to estimate the GAs which were compared to the GAs that were estimated using the CRL (measured in the first trimester of the pregnancy). Finally, we checked for indications of overfitting.

4.3.1 System parameter optimization

All parameters in the three quantification systems were optimized within the training set using a three-fold cross-validation. Optimization of five parameters was performed to improve the system performance (the parameter settings can be found in Table 4.2). First, the number of trees in the RFC was increased until the performance of the classifier was stable. Increasing the number of trees increases the computation time, so the lowest number of trees which showed a stable performance was used during optimization of the other parameters. Secondly, the scales of the Haar-like

Table 4.2: Parameter sets for optimizing systems A, B, and C.

Parameter	Set
Number of trees in RFC	$N_{trees} \in \{1, 2, 5, 10, 20, 50, 100\}$
Haar-like feature scales (mm)	$F_{scales} \in \{0.1, 0.2, 0.5, ..., 40, 45, 50\}$
Background sampling (mm)	$d_{min} \in \{0, 0.1, 0.2, 0.3\}$
Polar transform (mm)	$S_{dis} \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$
Polar transform	$N_{angles} \in \{360, 270, 180\}$

features were optimized. Starting with the optimum single scale, additional scales were only included when they improved the result. Thirdly, both the minimal distance, d_{min} and S_{dis} were increased until the performance did not improve anymore. Finally, the number of angles, N_{angles}, used for the polar transform was decreased as long as the performance of the system did not decrease, to speed up computation time.

4.3.2 HC comparison

4

The HC annotations of observer 1 were used as a reference to compare the performance of quantification system A, B, or C, as well as the observer 2 using the difference (DF), the absolute difference (ADF), the Hausdorff distance (HD)[73] and the Dice similarity coefficient (DSC)[74].

DF was defined as:

$$DF \quad = HC_S - HC_R, \tag{4.2}$$

where HC_R is the HC measured by observer 1 and HC_S is the HC measured by observer 2 or quantification system A, B or C.

ADF was defined as:

$$ADF \quad = |HC_S - HC_R| \tag{4.3}$$

HD was defined as:

$$H(S,R) \quad = max(h(S,R), h(S,R)), \tag{4.4}$$

where $R = \{r_1, \dots, r_q\}$ are the pixels from observer 1 and $S = \{s_1, \dots, s_p\}$ are the pixels from observer 2 or quantification system A, B or C, given:

$$h(S,R) \quad = \max_{s \epsilon S} \max_{r \epsilon R} ||s - r||. \tag{4.5}$$

DSC was defined as:

$$DSC \quad = \frac{2 \cdot |Area_S \cap Area_R|}{|Area_S| + |Area_R|}, \tag{4.6}$$

where $Area_R$ is the area of the annotation of observer 1 and $Area_S$ is the area of the annotation of observer 2 or the quantification system A, B or C.

Statistical analysis was performed to determine whether the difference was significant ($p < 0.05$). When the tested data was normally distributed according to the Shapiro-Wilk test, a paired T-Test was performed using SPSS (version 20.0). Otherwise, a Wilcoxon Signed Rank Test was performed. Although not all distributions were normally distributed, the tables in the Results Section show the mean and standard deviation, because this makes a comparison with values provided in previous literature possible.

4.3.3 GA comparison

The GA from the HC of the quantification systems and the observers was estimated using the P50 curve from Verburg et al.[39]. The reference GA was determined with a CRL measurement between 20 mm (8^{+4} weeks) and 68 mm (12^{+6} weeks). The differences between the estimated GA and the reference GA were computed for evaluation of the results. The same statistical tests as explained in the previous Section were used to determine whether the difference in GA was significant.

4.3.4 Overfitting

The best performing quantification system was evaluated on the training data to investigate whether overfitting of the system parameters had occurred.

4.4 Results

4.4.1 System parameter optimization

Table 4.3 shows the final parameter settings of the three quantification systems, as determined by running the optimization procedure on the training set explained in Section 4.3. The Haar-like feature scales are sorted by importance. Note that both the most important Haar-like feature scale, F_{scales}, and the downsampling of the dynamic programming, S_{dis}, increases with the trimester.

Table 4.3: Final parameter settings of quantification systems A, B, and C after parameter optimization.

	System A	**System B**		**System C**		
Computed on trimester(s)	1, 2 and 3	1	2 and 3	1	2	3
Number of trees RFC, N_{trees}	10	10	10	10	10	10
Haar-Like feature scales, F_{scales}(mm)	6, 20	2.5, 0.5, 11	7,11	2.5, 0.5, 11	7	9,12
Background sampling, d_{min} (mm)	0	0.2	0.2	0.2	0.1	0.1
Hough transform, r_{min} (mm)	5	5	12	5	12	34
Hough transform, r_{max} (mm)	61	18	61	18	50	61
Polar transform, S_{dis} (mm)	0.4	0.2	0.4	0.2	0.3	0.5
Polar transform, N_{angles}	270	270	270	270	270	270

4.4.2 Visualization of computation steps of quantification system C

Fig 4.6 shows the output of each step in quantification system C for an ultrasound image in the test set of a fetus with a GA of 20^{+0} weeks. All three pipelines of system C are computed on the input image from the test set. The second row shows the output of the pixel classifiers for each pipeline. It can be seen that the pixel classifier of the first pipeline, which is optimized on the training data of the first trimester, does not give a high response on this image. The third row shows the polar transform of the pixel classifier, where it can be seen that the radial dimension of the image decreases as the trimester increases due to the increase in sampling distance S_{dis}. The middle image of the fourth row shows that the second dynamic programming result (green) is re-positioned towards the outside of the fetal skull compared to the first dynamic programming result (red). Row six shows the final three fitted ellipses. In this example, the pipeline that was optimized on the training data of the second trimester gave the highest median pixel classifier response on the edge of the fitted ellipse. This ellipse was therefore selected as the final result.

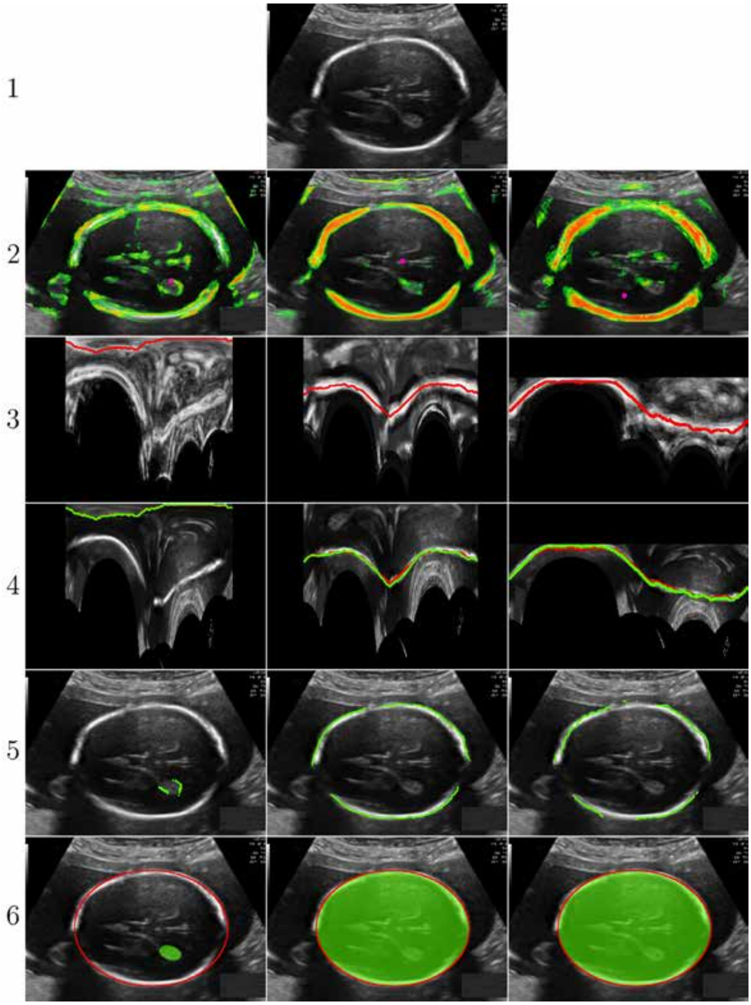

Figure 4.6: Steps of quantification system C. From left to right: pipeline 1, 2, and 3, respectively. From top to bottom: (1) Input image. (2) Input ultrasound image with overlay of pixel classifier likelihood ranging from green to red and Hough transform result in pink. (3) Polar transformed pixel classifier likelihood with overlay of dynamic programming in red. (4) Polar transformed ultrasound image with overlay of dynamic programming in red and repositioned dynamic programming result in green. (5) Ultrasound image with overlay of the highest five percentile repositioned dynamic programming pixels. (6) Ultrasound image with fitted ellipse in green and annotation of the experienced sonographer in red. In this example image, the pipeline that was optimized for the second trimester is automatically selected as the best result, since the edge of this fitted ellipse has the highest median pixel classifier output.

4.4.3 HC comparison

Table 4.4 shows the DF, ADF, HD and DSC of the measured HC from observer 1 compared to quantification systems A, B and C and observer 2. The ADF, HD and DSC of system A are significantly worse in the first trimester than systems B and C and observer 2. In the second trimester, system A fails on one image because the system fits an ellipse smaller than 38.6 mm. Therefore, the values for system A in the second trimester only consist of 232 values. The DF and ADF of observer 2 are significantly worse in the second trimester than systems A, B and C. There are no significant differences in HC between the systems and observer 2 in the third trimester.

Table 4.4: Results of the experienced sonographer (observer 1) compared to the classifier A, B and C and the medical researcher (observer 2) on the test set.

		Trimester 1	**Trimester 2**	**Trimester 3**
DF(mm)	Observer 2	1.4±1.9•°	3.4±2.8∗•°	1.4±7.0
	System A	3.2±20.5	1.1±3.0•°	0.7±6.3
	System B	-0.3±6.0	0.9±3.8	1.3±6.7
	System C	-0.3±6.1	0.8±3.3	0.6±5.9
ADF(mm)	Observer 2	1.8±1.5	3.7±2.5∗•°	5.4±4.6
	System A	11.3±17.3‡•°	2.3±2.2	5.1±3.7
	System B	3.1±5.1	2.5±3.0	5.4±4.0
	System C	3.1±5.2	2.4±2.4	4.8±3.4
HD (mm)	Observer 2	0.9±0.5	1.8±0.9	3.3±1.6
	System A	5.0±5.7‡•°	1.8±1.1	3.5±1.6
	System B	1.7±2.3‡	1.8±1.4	3.9±2.3
	System C	1.7±2.3‡	1.8±1.3	3.3±1.6
DSC(%)	Observer 2	96.8±1.7	97.5±1.0∗•	97.4±1.0
	System A	84.1±15.2‡•°	97.6±1.3	97.3±1.1
	System B	94.4±5.4‡	97.6±1.5	96.9±1.5
	System C	94.4±5.5‡	97.6±1.4	97.2±1.2

Note: ‡significantly different from observer 2, ∗significantly different from system A, •significantly different from system B, °significantly different from system C

4

4.4.4 GA comparison

The difference between the reference GA (estimated from the CRL) and the GA computed from the HC is shown in Table 4.5 and visualized in Fig 4.7. The difference between the reference and observer 2 in the first trimester is significantly worse than the difference between the reference and observer 1, system B and system C. The difference between the reference and observer 2 in the second trimester is significantly worse compared to observer 1 and systems A, B and C. Fig 4.7 shows that observer 2 tended to manually annotate the HC a few millimeters larger compared to observer 1, which resulted in a larger estimated GA. Fig 4.7 shows that system A has a large interquartile range and four outliers with a difference of more than 20 days. System B is significantly worse than that of system C in the third trimester. This is caused by two outliers with a difference of more than 30 days, which are shown in Fig 4.7.

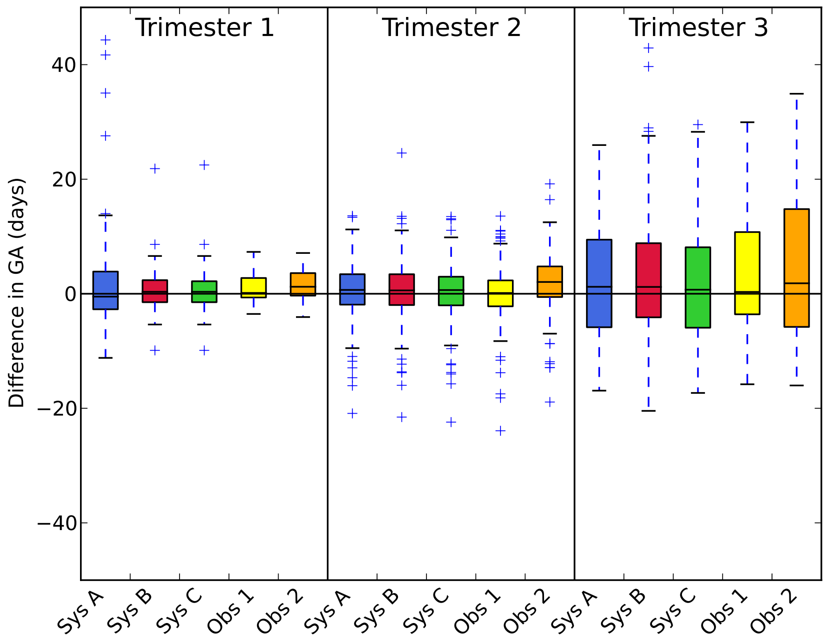

Figure 4.7: The difference with the reference GA (days) that was estimated using the CRL in the first trimester.

Table 4.5: Mean difference with the reference GA (days) that was estimated using the CRL in the first trimester

	Trimester 1	**Trimester 2**	**Trimester 3**
Observer 1	0.8±2.6	-0.0±4.6	1.9±11.0
Observer 2	1.6±2.7†•°	2.0±4.8∗•°	3.9±13.7
System A	2.5±11.2	0.6±4.7†•°	2.9±12.5
System B	0.6±4.3	0.6±4.9†	3.8±14.4°
System C	0.6±4.3	0.4±4.7†	2.5±12.4

Note: †significantly different from observer 1, ∗significantly different from system A, •significantly different from system B, °significantly different from system C

4.4.5 Visual results of quantification system C

To get an idea how the median ADF of system C looks like, the result closest to the median ADF of system C is visualized in Fig 4.8. The images of the first, second and third trimester have an ADF of 1.8 mm, 1.6 mm and 4.2 mm, which results in a difference in GA of -1.0 days, -0.9 days and -4.3 days, respectively. The median ADF in the first trimester is a lot smaller compared to the mean ADF of 3.1 mm (shown in Table 4.4), due to one outlier. This outlier is shown in the right column of Fig 4.8 and has a ADF of 36.8 mm, which results in a difference in GA of 22.5 days with the GA estimated from the CRL.

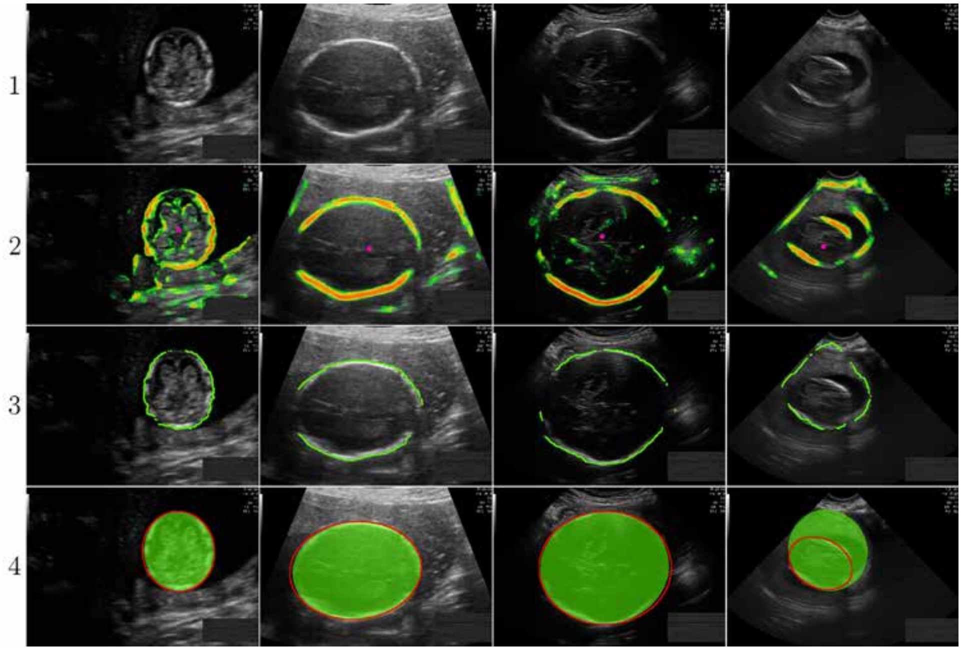

Figure 4.8: Results of quantification system C closest to the median ADF of system C. From left to right: first trimester with a ADF of 1.8 mm, result second trimester with a ADF of 1.6 mm, result third trimester with an ADF of 4.2 mm and worst result first trimester with an ADF of 36.8 mm. From top to bottom: (1) The ultrasound image. (2) The ultrasound image with overlay of the pixel classifier likelihood ranging from green to red and the Hough transform result in pink. (3) The ultrasound image with an overlay of the highest fifth percentile repositioned dynamic programming pixels. (4) The ultrasound image with the fitted ellipse in green and the annotation of observer 1 in red.

4.4.6 Overfitting

Table 4.6 shows the results of quantification system C on the training and the test set. Overfitting occurs when the results on the training set are much better than the results on the test set.

4.4.7 Comparison to literature

Table 4.7 shows a comparison of system C with the reported results published in literature.

Table 4.6: Results of quantification system C for training and test set compared to observer 1

	Trimester 1		Trimester 2		Trimester 3	
	Train set	**Test set**	**Train set**	**Test set**	**Train set**	**Test set**
DF (mm)	-0.7±6.4	-0.3±6.1	0.7±2.1	0.8±3.3	1.0±6.1	0.6±5.9
ADF (mm)	3.4±5.5	3.1±5.2	2.3±2.1	2.4±2.4	4.6±4.1	4.8±3.4
HD (mm)	2.1±2.7	1.7±2.3	1.7±0.9	1.8±1.3	3.4±2.0	3.1±1.9
DSC (%)	93.2±7.5∗	94.4±5.5	97.6±1.8	97.6±1.4	97.2±1.5	97.3±1.5

Note: $*p<0.05$

Table 4.7: Comparison of system C against the reported results published in literature.

Method	No.	GA (weeks)	DF (mm)	ADF (mm)	HD (mm)	DSC (%)
Our method	335	11-37	0.6±4.3	2.8±3.3	2.0±1.6	97.0±2.8
Zhang et al.[66]	10	-	-0.22±9.53	-	3.30±1.09	-
Anto et al.[65]	50	-	-	-	-	75±-
Perez-Glez. et al.[64]	10	-	-2.73±2.04	-	2.64±0.57	97.19±0.97
Jatmiko et al.[58]	100	-	-	8.21±-	-	-
Satwika et al.[54]	72	Trim 1&2	-	14.6±-	-	-
Foi et al.[75]	90	21,28,33	-2.01±3.29	-	2.16±1.44	97.80±1.04
Ciurte et al.[75]	90	21,28,33	11.93±5.32	-	4.6±1.64	94.45±1.57
Stebbing et al.[75]	90	21,28,33	-3.46±4.06	-	2.59±1.14	97.23±0.77
Sun et al.[75]	90	21,28,33	3.83±5.66	-	3.02±1.55	96.97±1.07
Ponomarev et al.[75]	90	21,28,33	16.39±24.88	-	6.87±9.82	92.53±10.22
Ni et al.[57]	175	17-38	-	5.58±1.74%	-	-
Zalud et al.[56]	80	-	-	5.1±5.4	-	-
Carneiro et al.[55]	20	-	-	2.76±1.40	4.15±2.05	-
Lu and Tan[53]	11	13-34	-	3.41±1.74%	-	-

4.5 Discussion

We presented three variations of a quantification system, indicated as system A, B or C, that measures the fetal HC in all trimesters of the pregnancy. The systems were evaluated on a large test set of 335 ultrasound images. The best system, system C, performs comparable to an experienced sonographer (observer 1) and significantly better than a medical researcher (observer 2) in the first and second trimester. The presented system shows similar or superior results compared to other systems published in literature. This is the first system in literature that was evaluated on a very larger test set of 335 ultrasound images which contained data of all trimesters of the pregnancy (Table 4.7). In the next Section, we discuss various aspects of the system and the experimental results.

4.5.1 HC comparison

Table 4.4 shows that system A performs significantly worse in the first trimester compared to systems B and C. This results from the fact that the appearance of the fetal skull in the first trimester differs from that in the second and third trimester. Since the fetal skull is relatively soft in the first trimester, it does not always appear brighter than the inside of the fetal head. Therefore, it is sometimes very difficult to detect the edge of the fetal head, especially when it lies close to the wall of the uterus. Thus, it is more difficult to automatically measure the HC in the first trimester. For this reason, it is important to show the performance of an automated system for each trimester separately. It is also known that the standard deviation in HC increases as the size of the fetus increases. The primary reason for this results from the fact that the natural variation in fetal size increases with GA. Since the fetal head becomes larger with GA, the pixel size of the ultrasound image also increases. This can also be noticed in the curve of Verburg et al.[39], where the P3-P97 interval gets wider with increasing GA. Separation of the different trimesters is hereby essential when evaluating the results. It also underlines the clinical importance to estimate the GA of the fetus in the first or second trimester to obtain a reliable estimate of the fetal GA.

4.5.2 GA comparison

Table 4.5 shows that system B performs significantly worse than system C in the third trimester, so it is beneficial to train a separate classifier for the third trimester as well. Together with the results from the previous Section, it can be concluded that system C performs superior to systems A and B and was therefore chosen as the final system.

Table 4.5 and Fig 4.7 show that the standard deviation of the GA in the first trimester of quantification system C is larger than the that of the two observers. This is mainly caused by one outlier. When this outlier was removed, the standard deviation decreases from 4.3 to 3.1 days, which is similar to the standard deviation of observers 1 and 2.

The mean GA estimation of system C is significantly better than observer 2 in the first and second trimesters, compared to the reference GA estimated from the CRL. The underlying reason for this is that observer 2 systematically annotated the HC a few millimeters larger compared to observer 1. This indicates that the system may aid inexperienced human observers in measuring the HC. Furthermore, the standard deviation of the HC in the second trimester is similar for both observers and system C. The performance of observer 1 is significantly better in the second trimester compared to system C, but the mean difference of 0.4 days is not clinically relevant.

4.5.3 Visual results of quantification system C

Fig 4.8 shows the result of system C with the median ADF for each trimester. It can be seen that the median result of system C is very similar to the manual annotations of observer 1. The increase in ADF for later trimesters is mainly caused by the increase in pixel size. The right column in Fig 4.8 shows the outlier of system C in the first trimester. In this image, the right and left side of the fetal skull are hardly visible. In addition, a large shadow appears next to the dark amniotic fluid at the right side of the fetal skull. While the Hough transform still detects the center of the fetal head, the dynamic programming algorithm is not able to follow the fetal skull. Instead, it follows the border between the amniotic fluid and the shadow, resulting in a HC that is completely off. This results in a difference in GA of 22.5 days with the reference GA.

4.5.4 Overfitting

Table 4.6 shows the results of system C on the training and the test sets. Note that no overfitting occurs because the results from the training and test sets did not differ significantly. The DSC in the first trimester was even significantly worse in the training set compared to the test set.

4.5.5 Comparison to literature

Table 4.7 shows an overview of previously reported results in literature. Ideally, these methods were evaluated on the same test set to make a direct comparison possible. Unfortunately, such a dataset was not available and implementation of other methods is a difficult task due to the lack of implementation details. Even though a direct comparison of the results is not possible, Table 4.7 highlights three strengths of our method. First, four methods[53,55,64,66] were only evaluated on a dataset of 10, 11 or 20 images. Our method was evaluated on a large independent test set of 335 images, which shows not only the feasibility but also the robustness of the method. Secondly, it was shown that the first trimester is the most challenging trimester to measure the HC, but almost all other methods either did not mention the GA of the test set, or only evaluated their system only on data of the second and third trimester. We therefore recommend that future research will report the GA and evaluate the results for each trimester separately. This would make a comparison with previous work easier. Thirdly, only Satwika et al.[54] have evaluated their system on a relatively large test set of 72 images which included data of the first trimester. They have reported a mean ADF of 14.6 mm, which is much larger compared to the ADF of 2.8±3.3 mm of our proposed method. Even though these systems were not evaluated on the same test set, it illustrates the potential of our proposed method.

4.5.6 Study limitations

The data for this study was acquired in only one hospital using two different ultrasound devices from the same vendor. Future work should include multi-center data from different vendors to be able to further evaluate the performance of the proposed method. The results show that the system performs significantly better than a medical researcher in the first and second trimester, but it is still required to obtain the 2D standard plane. Other work in literature focuses on aiding less skilled sonographers in obtaining the 2D standard plane, or reconstructing the 2D standard plane from a 3D volume[76–81]. Combining these methods with our proposed system could further improve inter-observer variability, but this is out of the scope of this work.

4.6 Conclusions

We presented an automated system for the detection of fetal HC in 2D ultrasound images. This is the first system presented in literature that was evaluated on a large independent test set of 335 ultrasound images that included data of all trimesters. It was shown that it is important to separate the results for each trimester, because the uncertainty of the estimated GA increases with GA due fact that the natural variation in fetal size increases with GA. This is the first system that evaluated results for each trimester separately. The GA can be estimated more accurately in the first trimester, but the fetal skull is not clearly visible in the first trimester, which makes automated detection of the HC a more challenging task. The performance of the presented system was comparable to an experienced sonographer.

5

Automated fetal head detection for estimation of the head circumference

T.L.A. van den Heuvel
H. Petros
S. Santini
C.L. de Korte
B. van Ginneken

Original title:
Automated fetal head detection and circumference estimation from free-hand ultrasound sweeps using deep learning in resource-limited countries

Accepted for publication in:
Ultrasound in Medicine and Biology

Abstract

Ultrasound imaging remains out of reach for most pregnant women in developing countries, because it requires a trained sonographer to acquire and interpret the images. We address this problem by presenting a system that can automatically estimate the fetal head circumference (HC) from data obtained with the use of the obstetric sweep protocol (OSP). The OSP consists of multiple predefined sweeps with the ultrasound transducer over the abdomen of the pregnant woman. The OSP can be taught within a day to any health care worker without prior knowledge of ultrasound.

An experienced sonographer acquired both the standard plane—to obtain the reference HC—together with the OSP from 183 pregnant women in St. Luke's Hospital, Wolisso, Ethiopia. The OSP data—which will most likely not contain the standard plane—was used to automatically estimate the HC using two fully convolutional neural networks. First, a VGG-Net inspired network was trained to automatically detect the frames which contained the fetal head. Second, a U-net inspired network was trained to automatically measure the HC for all frames in which the first network detected a fetal head. The HC was estimated from these frame measurements and the curve of Hadlock was used to determine the gestational age (GA).

The results show that most automatically estimated GAs fell within the P2.5-P97.5 interval of the Hadlock curve compared to the GAs obtained from the reference HC, so it is possible to automatically estimate the GA from the OSP data. Our method has therefore potential application for providing maternal care in resource-constrained countries.

5.1 Introduction

Worldwide, 99% of all maternal deaths occur in developing countries. In absolute numbers, this corresponds to approximately 820 deaths per day[1]. Ultrasound is widely used to detect maternal risk factors during pregnancy, because it is a low-cost, real-time and non-invasive imaging method. However, images suffer from noise, shadows, and reverberations, making it hard to interpret them. More importantly, a trained sonographer is required to acquire and interpret the images. In first world countries, sonographers are extensively trained to obtain precisely defined standard imaging planes in which to perform biometric measurements of the fetus[42,82–85]. The fetal head circumference (HC) is one of the most important measurements. The HC can be used to determine the gestational age (GA) and monitor growth of the fetus. The guidelines describe that the standard plane for obtaining the HC should be measured at the level of the thalami, where the cavum septi pellucidi interrupts the anterior one-third of the falx. Ideally, the falx is positioned horizontally on the screen. The cerebellum should not be visible in this scanning plane[42,82–85]. Unfortunately, there is a severe shortage of trained sonographers in developing countries[3], which keeps ultrasound imaging out of reach for most pregnant women in these countries. In this paper, we present a system that automatically estimates the HC with the use of the obstetric sweep protocol (OSP). The OSP consists of multiple free-hand sweeps with the ultrasound transducer over the abdomen of the pregnant woman. The OSP can be taught to any health care worker without any prior knowledge of ultrasound within a day, obviating the need for a trained sonographer to obtain the ultrasound images. By combining the OSP with a system that can automatically estimate the HC from the sweep data, there would also be no need for a trained sonographer to interpret the images for this task.

The literature describes several methods to automatically measure the HC when the standard plane is acquired[56–58,75,76,86]. However, acquisition of this standard plane requires a trained sonographer which is the problem that we would like to solve in this study. There are three different approaches presented in literature to aid less experienced sonographers in obtaining information about the fetus using ultrasound data.

In the first approach, 3D ultrasound is used to automatically extract the standard plane for the fetal brain[76–81], abdomen[80,87], heart[88,89], nuchal translucency[90], and face[91,92]. Unfortunately, 3D ultrasound is more expensive compared to 3D ultrasound[93] and is therefore considered unsuitable for developing countries. Additionally, it is unknown whether this approach is suited for the third trimester when the fetus does not completely fit within the field of view (FOV) of the 3D probe.

In the second approach, a video of a free-hand 2D ultrasound probe is used to detect the standard plane of the fetal abdomen[94,95], heart[96] and face[97], and to automatically obtain multiple standard planes[98,99]. Systems analyzing such video streams can be used to aid less experienced sonographers in obtaining the correct standard plane. Even though this approach uses 2D ultrasound, it still requires training of the sonographer to use an ultrasound device and interpret the images to obtain the biometric measurement of the fetus.

In the third approach, which was used in this study, predefined free-hand sweeps are acquired to obtain information about the fetus. The main advantage of this approach is that these sweeps can be taught to any health care worker without any knowledge of ultrasound within a day. The disadvantage is that the sweep data will most likely not contain the standard plane that is usually used to perform the biometric measurement. Kwitt et al.[100], used a free-hand sweep on phantom data to automatically detect structures of interest. To the authors knowledge there is only one paper in literature that developed a system that automatically detects the fetal position and heartbeat with the use of a single predefined free-hand sweep[101]. But in this paper the authors mention that this single sweep did not contain either the head or abdomen in 31% of the 129 test cases. In this paper we use the obstetric sweep protocol (OSP) introduced by DeStigter et al.[6]. The OSP consists of multiple predefined free-hand sweeps. The three transversal sweeps used in this study are shown in Figure 5.1. The use of three sweeps increases the chance that the fetal head is visible in at least one of the sweeps.

This is the first study to propose an automated system which estimates the HC from predefined sweeps without the acquisition or reconstruction of the standard plane. In previous work, we have shown that it is possible to manually select an optimal frame from the sweep data to estimate the HC[102].The aim of this study was to develop and validate a method for fully automatic estimation of the HC and the GA with the use of the OSP, obviating the need for a trained sonographer to provide point-of-care obstetric ultrasound. The data for this study was acquired in Ethiopia, and therefore represents data from the target population for this application. We specifically aimed to reduce the computational complexity of the proposed method as much as possible, to be able to deploy the system on a low-cost laptop or tablet. This would facilitate widespread application of this system in developing countries.

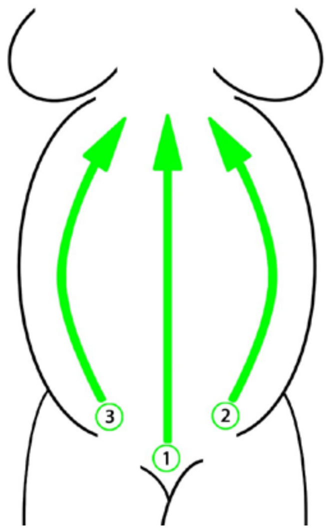

Figure 5.1: Visualization of the three free-hand sweeps of the obstetric sweep protocol. The three sweeps are obtained by moving the ultrasound transducer from the pubic bone to the breast bone, as indicated by the green arrows.

5.2 Methods

5.2.1 Data

For this study, an experienced sonographer acquired both the standard plane—for measuring the reference HC—together with the OSP from 183 pregnant women using the SonoAce R3 (Samsung Medison, Korea). The data was acquired in St. Luke's Catholic Hospital and College of Nursing and Midwifery in Wolisso, Ethiopia. Patient identifiers were removed and the data were saved anonymously. This study was approved by the local ethics committee (Ref. No. BEFO/AHBTHQO/4004/1-20). Every pregnant woman in this study signed a written informed consent. All data was anonymized according to the tenets of the Declaration of Helsinki. The image of the standard plane was used to measure the reference HC for each patient. An example image of the standard plane is shown in Figure 5.2. The reference HCs in the data varied between 116 mm and 361 mm (Figure 5.3); most pregnant women visited the hospital in the third trimester of their pregnancy, which is typical for maternal care in developing countries. Note that Figure 5.3 shows only 181 HCs; data from two anencephalic fetuses was excluded.

5

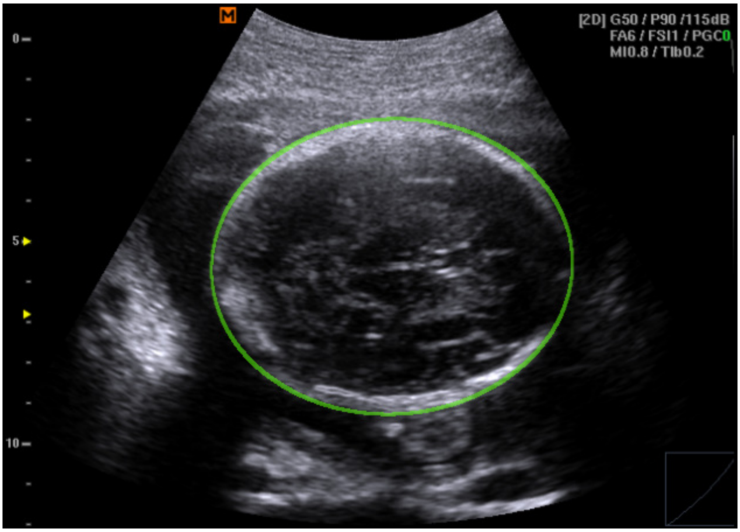

Figure 5.2: Example ultrasound image of the standard plane that was used to obtain the fetal head circumference (green).

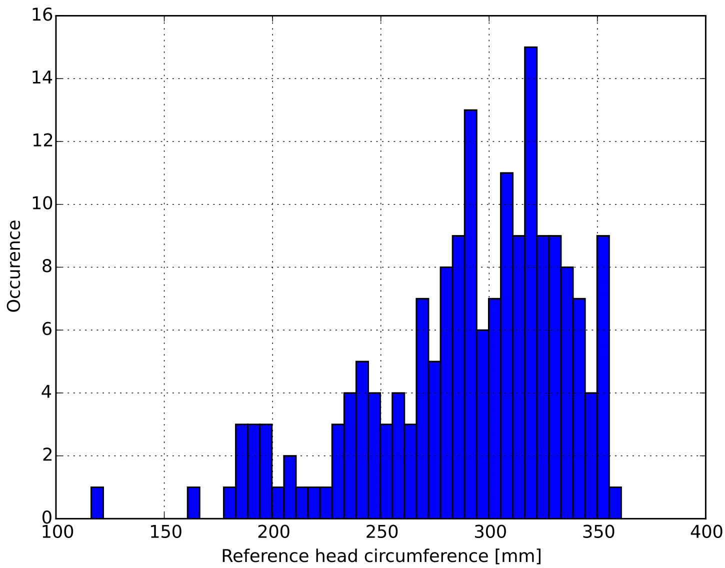

Figure 5.3: Overview of the distribution of head circumferences (N=181) measured in the standard plane.

The OSP, introduced by DeStigter et al.[6], recommends six free-hand sweeps over the abdomen of the pregnant women. Saving one sweep with the SonoAce R3 took on average one minute. To prevent a delay in the clinical work flow, it was decided to only obtain the three transverse sweeps of the OSP. The transverse sweeps were obtained from the pubic bone to the breast bone (Figure 5.1). The first sweep was obtained at the midline, the second sweep was obtained at the left side of the patient and the third sweep was obtained at the right side of the patient. The sonographer was asked to acquire around one hundred frames per sweep, but since these sweeps are made in free-hand mode, the number of frames per sweep was variable. The imaging depth was set to 12 cm and each frame in the sweep had a size of 630×450 pixels.

An overview of the number of patients, sweeps and frames can be found in Table 5.1. The data was divided in a training set, validation set and test set of 60%, 20% and 20%, respectively. The training set was used to train the systems, the validation set was used to optimize hyperparameters of the systems and select the stopping criterion, the test set was used to evaluate performance of the systems. The number of frames was matched as closely to the target ratio as possible, while making sure that the data of one patient was always part of either the training, validation set, or test set.

All frames within the OSP data were manually labeled for the presence of the fetal head (Table 5.1). Four different labels were used: present, partially present, not present and possibly present. Present meant that the fetal head falls within the FOV of the frame. Partially present meant that the fetal head falls partially outside of the FOV of the frame, which would make a circumference measurement inaccurate. Not present meant that the fetal head was not present in the frame. Possibly present meant that either the frame contains a fraction of the fetal head but not enough to

Table 5.1: Overview of the data

	Total	**Training**	**Validation**	**Test**
Targeted ratio (%)	100	60	20	20
No. of patients	183	109	35	39
No. of sweeps	621	369	128	124
No. of frames	49.269	29.181	10.197	9.891
Head present	3.199	2.097	478	624
Head partially present	1.238	772	266	200
Head not present	35.496	20.966	7.394	7.136
Head possibly present	9.336	5.346	2.059	1.931

recognize a fetal head circumference, or it was impossible to be sure whether the fetal head was present in the frame, due to artifacts or poor image quality. The frames labeled as possibly present were not used for the training. Figure 5.4 shows an example of a labeled sweep data set. Despite instructions to the sonographer, not all three sweeps were acquired for all patients, and not all sweeps contained more than one hundred frames.

For evaluation of the estimated HC in the test set we excluded ten twins, four fetuses for which not all three sweeps were recorded, two fetuses in shoulder presentation, two low-lying fetuses for which a depth setting of 12 cm was not sufficient to detect the HC, one anencephalic fetus and one patient with polyhydramnios. This resulted in 23 patients in the validation set and 31 patients in the test set.

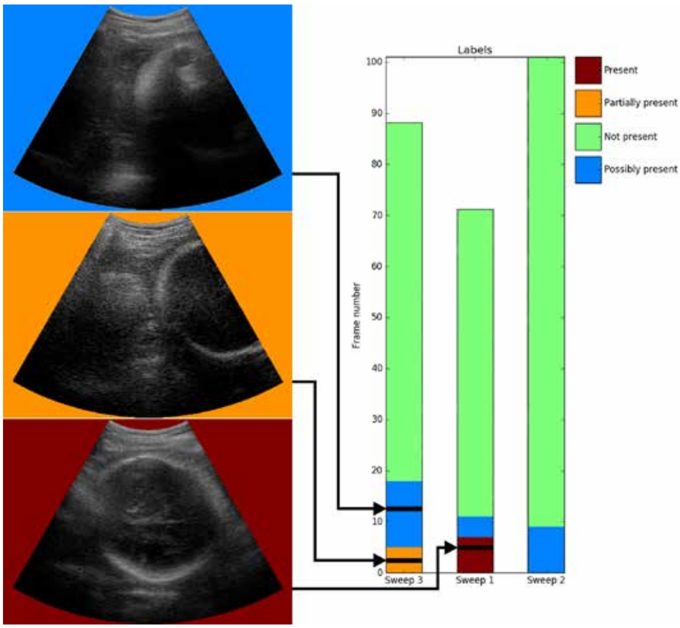

Figure 5.4: Labels of an example data set. A schematic drawing of the three sweeps is shown on the right. Sweep 1 contains 71 frames, sweep 2 contains 101 frames, and sweep 3 contains 88 frames. Each frame was manually labeled if the frame contains a fetal head. The color-bar on the right indicates the colors of the four classes. Three example frames are shown on the left side. One frame from sweep 1 in which the fetal head is present is shown in red. One frame from sweep 3 in which the fetal head is partially present is shown in orange. One frame from sweep 3 in which the fetal head is possibly present is shown in blue; this frame shows an eye of the fetus, but there is no indication of the fetal head circumference visible in this frame.

5.2.2 Pre-processing

Three pre-processing steps were performed on the data. Firstly, the pixel values were rescaled from 0-255 to floating point values from 0-1. Secondly, all frames were automatically masked to remove surrounding markers like the lookup table, the ruler and the ultrasound acquisition settings. Finally, since this system is intended to be used in developing countries, it should be possible to run it on a low-cost laptop or tablet. To achieve this goal, the frames were downsampled with ten different downsampling factors—ranging from two to twenty—to decrease the dimensions of the input layer. Experiments were performed at each downsampling factor to investigate how much the input image could be downsampled without decreasing the performance of the system. Figure 5.5 shows four example images with different downsampling factors. A frame without downsampling had a size of 630×450 pixels and a frame with a downsampling factor of twenty had a size of 32×23 pixels

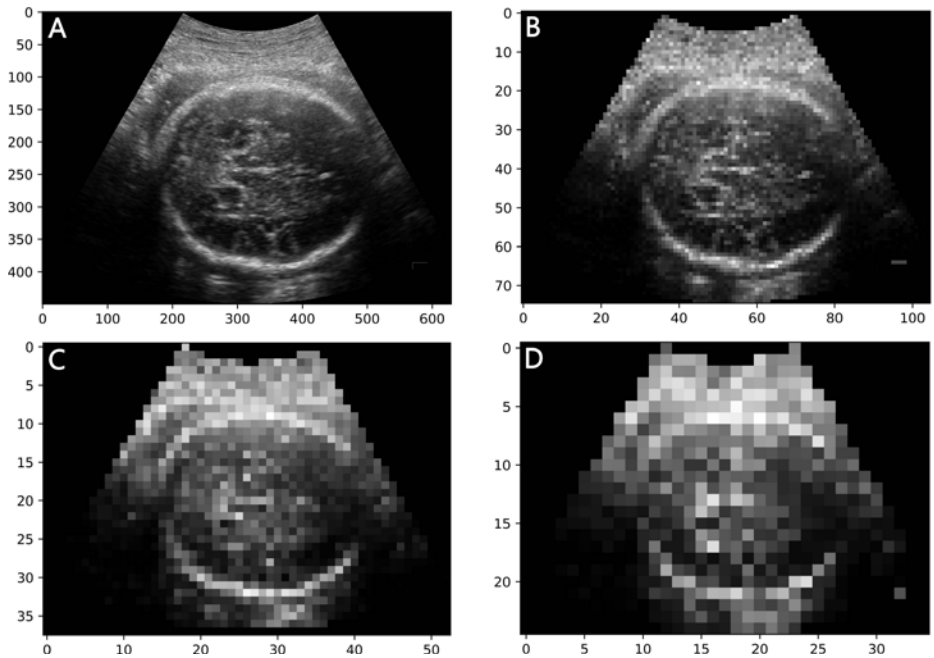

Figure 5.5: Visualization of four different downsampling factors of a frame in which the fetal head is present. From A to D: no downsampling, downsampling factor of 6, downsampling factor of 12 and downsampling factor of 18. This results in a frame size of 630×450, 105×75, 53×38 and 35×25 pixels, respectively.

5

5.2.3 Fetal head detection

Network architecture

Table 5.2 shows an overview of the five network architectures that were evaluated for the fetal head detection, indicated by the letters A, B, C, D and E. The network architectures were inspired by the VGG-Net of Simonyan and Zisserman[103]. The downsampled frame from the OSP data was used as the input layer of the network. Network architecture A, B, C, D and E contained one, two, three, four and five stack(s) of two convolution layers and one max-pooling layer, respectively. The number of filters in the convolution layer was doubled after each max-pooling layer. Three fully connected (FC) layers follow after the last max-pooling layer. The last FC layer contained three neurons followed by a soft-max to a posterior over the three classes. These three classes were: fetal head present, partially present and not present. All hidden layers (convolutional and FC) were equipped with a rectified linear activation function. The number of network parameters needed to be as low as possible to make deployment on a low-end laptop or tablet feasible in the future. This decrease in network parameters was achieved by decreasing the number of filters in the convolution layers and decreasing the number of neurons in the FC layers compared to the VGG-Net of Simonyan and Zisserman. An overview of the number of network parameters for each network architecture per downsampling factor is shown in Table 5.3. Note that not all network architectures can be computed for each downsampling factor because the input image could become too small.

Network parameters

The weight parameters in the network were initialized using He weight initialization[104]. The network was implemented using Theano[105] and Lasagne[106]. Training was performed on a Nvidia GeForce GTX 1080 graphics card. During training, a batch size of 102 frames was used for all experiments, since this was the maximum number of frames that fitted in the GPU memory when a downsampling factor of two was used. The size of the batch size will change the rate of convergence[107], but this influence was not evaluated since an exhaustive search of all hyper parameters is not feasible due to the required computation power. The Adam update method[108] was used with an initial learning rate of 0.001 and L2 regularization of 0.0001. Dropout was used in all FC layers (p=0.5)[109].

Table 5.2: Network architectures A, B, C, D and E. conv3-N is a 3×3 convolution with N filters. FC-M is a fully connected layer with M neurons.

A	**B**	**C**	**D**	**E**
		input-layer		
conv3-16	conv3-16	conv3-16	conv3-16	conv3-16
conv3-16	conv3-16	conv3-16	conv3-16	conv3-16
maxpool	maxpool	maxpool	maxpool	maxpool
	conv3-32	conv3-32	conv3-32	conv3-32
	conv3-32	conv3-32	conv3-32	conv3-32
	maxpool	maxpool	maxpool	maxpool
		conv3-64	conv3-64	conv3-64
		conv3-64	conv3-64	conv3-64
		maxpool	maxpool	maxpool
			conv3-128	conv3-128
			conv3-128	conv3-128
			maxpool	maxpool
				conv3-256
				conv3-256
				maxpool
		FC-256		
		FC-128		
		FC-3		
		soft-max		

Table 5.3: Number of network parameters per downsampling factor (in thousands)

Downsampling factor	**A**	**B**	**C**	**D**	**E**
2	69,873	32,613	13,868	5,242	2,195
4	17,067	7,423	2,727	917	-
6	7,204	2,876	843	-	-
8	3,976	1,492	400	-	-
10	2,412	836	236	-	-
12	1,707	541	154	-	-
14	1,183	378	-	-	-
16	872	247	-	-	-
18	650	173	-	-	-
20	552	132	-	-	-

5

Training of the network

During training, one epoch was defined as one pass over all frames in the training set. The training set included 20,966 frames labeled as fetal head not present, 2,097 frames labeled as fetal head present and 772 frames labeled as fetal head partially present. To balance the class distribution in each mini-batch, we oversampled the data in the present and partially present classes. Each mini-batch contained 102 frames (34 for each class). Augmentation was performed by horizontally flipping a random selection of 50% of the frames within a batch. To avoid a bias towards the negative samples during training, we defined a validation set which contained all 266 frames labeled as fetal head partially present and a random selection of 266 frames labeled as fetal head present and 266 frames labeled as fetal head not present. Every ten iterations, the linear weighted kappa was computed on this balanced validation set. The stopping criteria for training was reached when the weighted kappa on this balanced validation did not increase during ten epochs.

Network result

The network with the highest weighted kappa on the balanced validation set during training was selected as the final network. The accuracy on the full validation set was computed to compare the final result for each network architecture and downsampling factor.

5.2.4 Head circumference estimation

The frames in the OSP data will most likely not contain the standard plane which is normally used to measure the fetal HC, since the sweeps were predefined. The frames will therefore contain a random cross section of the fetal head. This section explains how the HC was estimated from the OSP data without obtaining or reconstructing the standard plane.

Network architecture

A network architecture inspired by the U-net of Ronneberger et al.[110] was used to determine which pixels in a frame belong to the outer edge of fetal head. The U-net proposed by Ronneberger et al.[110] contains over 31 million parameters. The number of parameters was reduced to 1.9 million by decreasing the number of channels in each feature map by a factor of four. The output segmentation map of the network contains with two classes: pixels that belong to the outer edge of fetal head and background pixels. The network was evaluated on ten different downsampling factors, ranging from two to twenty.

Network parameters

During training, a batch size of five 2D standard planes was used for all experiments. The Adam update method was used with an initial learning rate of 0.0001. Dropout was used in the two bottom layers (p=0.5), similar to the U-net proposed by Ronneberger et al.[110]

Training of the network

The 109 patients in the training set only included 90 2D standard plane images, because the 2D standard plane was not acquired for 19 patients. The pixels annotated as part of the HC were labeled positive and the other pixels were labeled as background, which resulted in a highly unbalanced class distribution. To balance the class distribution a weight map was introduced. The foreground weight was set to one and the background weight was set to the number of positive divided by the number of negative pixels per 2D standard plane image. Augmentation was performed by horizontally flipping a random selection of 50% of the 2D images. The system was evaluated on the OSP frames which were classified as fetal head present by the best performing fetal head detection network described in Section 5.2.3.

Network result

Connected component analysis was used to extract the largest component for each frame. Smaller components were included when the component had at least half the size of the largest component. A least square ellipse fit[72] was used to extract the HC for each frame. Since the HC measured in the standard plane is one of the largest HCs one can measure from a fetal head, the frame with the highest non-outlier HC was selected as the final HC. Outliers were removed to make the estimated HC robust to segmentation errors. An outlier was defined as a HC larger than $Q_3 + 1.5 \times (Q_3 - Q_1)$[111]. The network with the lowest mean absolute difference (MAD) between the reference HC, measured in the standard plane, and the automated HC, measured using the OSP, on the validation set was selected as the final network.

5.2.5 Gestational age estimation

The curve of Hadlock et al. [112] was used to determine the GA from the HC. The reference GA, determined from the HC obtained using the standard plane, was compared to the automatically estimated GA, determined from the automatically estimated HC using the OSP data. It was not possible to determine the GA when the HC fell outside the curve of Hadlock, therefore these fetuses were excluded in the GA comparison.

5.3 Evaluation and Results

All results were evaluated for ten different downsampling factors—ranging from two to twenty—to investigate to which extent the input image could be downsampled without decreasing the performance of the deep learning systems. Section 5.3.1 shows the performance for the head detection, Section 5.3.2 and show the results for the HC estimation and Section 5.3.3 show the results for the GA estimation.

5.3.1 Fetal head detection

Figure 5.6.A shows the accuracy for detection of the fetal head with the five network architectures on the full validation set for ten different downsampling factors. Network C is the best performing network for downsampling factors two until eight, network B is the best performing network for downsampling factors 10, 12, 14 and 18 and network A is the best performing network for downsampling factor 16 and 20. Network E could only be computed on a downsampling factor of two, but it did not converge and therefore the accuracy was 0.909. This accuracy is reached when all validation frames were classified as fetal head not present. The best performing network architecture per downsampling factor was also computed on the test set. Figure 5.6.B shows the accuracy of the best performing network architecture for both the validation set and test set. The frame based accuracy on the validation set was highest at downsampling factor six; network architecture C performed best at this downsampling factors and was therefore selected as the final network.

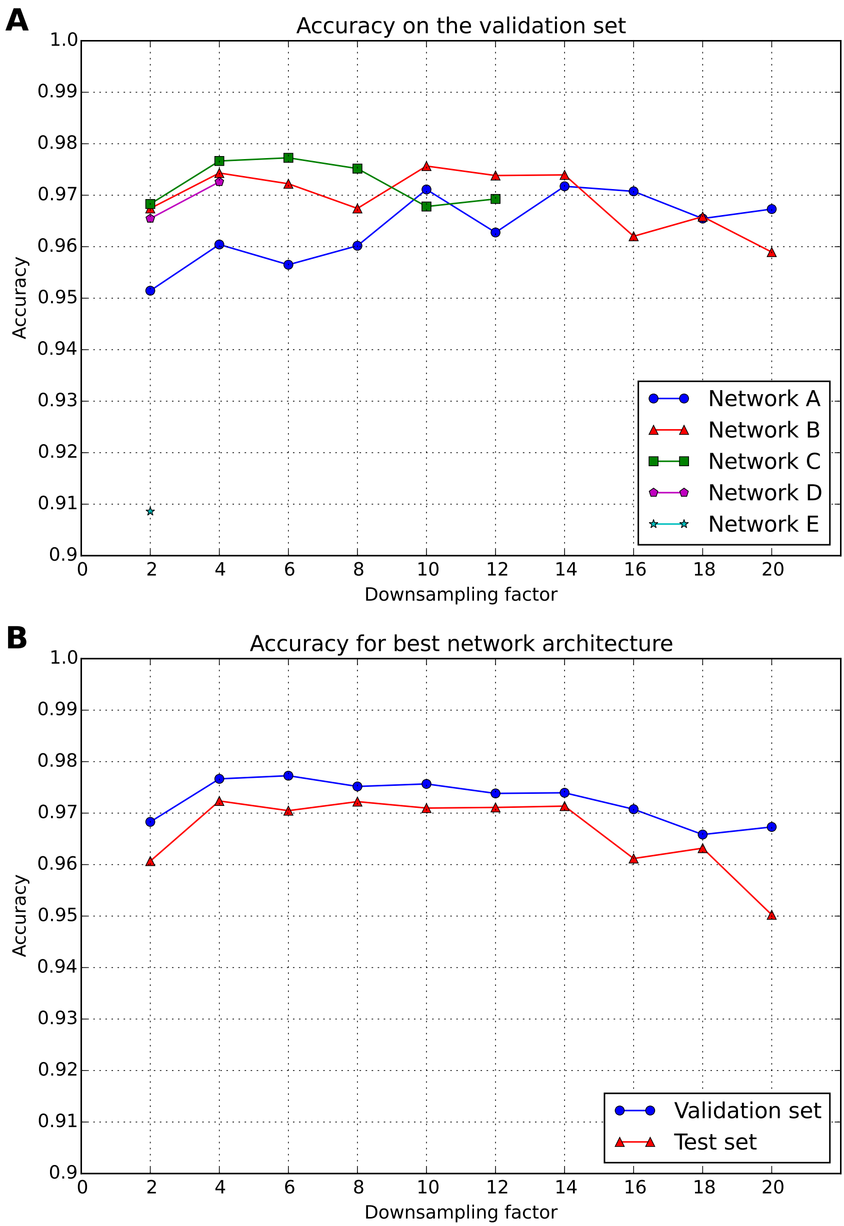

Figure 5.6: Accuracy at ten different downsampling factors. A: Accuracy on the full validation set for the five different network architectures. B: Based on graph A, we selected the best network architecture per downsampling factor and show the accuracy on both the validation set and test set.

5

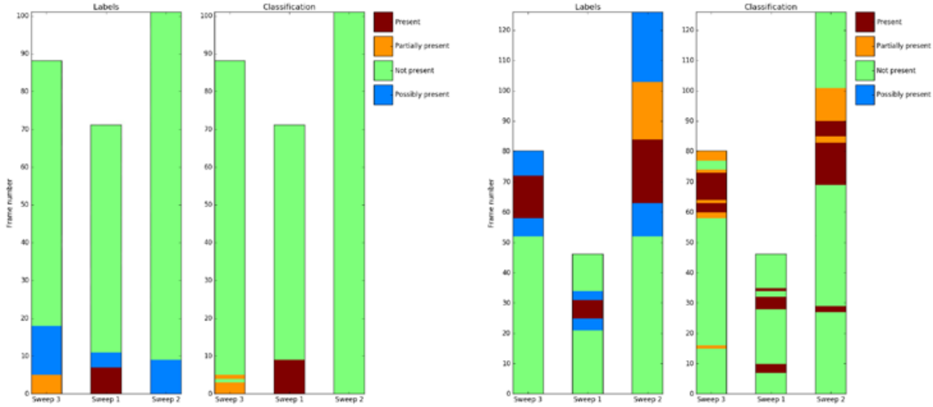

Figure 5.7: Labels and classification of two patients from the test set. The colors present the four different classes as indicated by the color-bar. Left: Labels and classification of example patients shown in Figure 5.4. Sweeps 1, 2 and 3 have an accuracy of 1.000, 1.000 and 0.987, respectively. Right: Labels and classification of the test patient with the lowest accuracy (0.861). Sweeps 1, 2 and 3 have an accuracy of 0.821, 0.826 and 0.939, respectively.

Figure 5.7 shows the manual labels and the classification of network architecture C with a downsampling factor six for two patients in the test set. The patient on the right has the lowest accuracy (of 0.861) of all patients in the test set.

There was one fetus in the test set in which the network did not detect any frames as fetal head present. This fetus had a HC of 116 mm, which was the smallest HC in the complete dataset (Figure 5.3).

5.3.2 Head circumference estimation

Figure 5.8 shows the MAD between the reference HC and the automatically estimated HC at ten different downsampling factors for both the validation set and the test set. The MAD on the validation set was lowest at a downsampling factor of four. The MAD on the test set at this downsampling factor was 10.3 mm. To make a comparison with literature possible, the 95% confidence interval was estimated using 1.96 times the standard deviation (SD) of the mean difference (MD). The MD±SD was -3.0 ± 13.3 mm and $-1.3 \pm 4.6\%$, resulting in a 95% confidence interval of 26.1 mm and 9.1%.

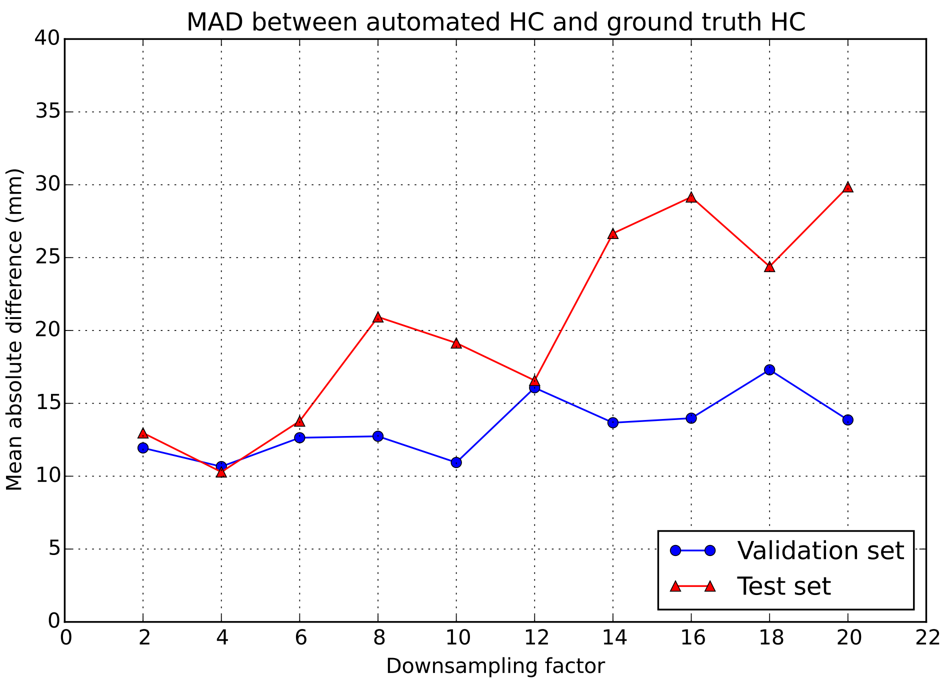

Figure 5.8: Mean absolute difference between the reference HC and the automatically estimated HC at eleven different downsampling factors for both the validation set and the test set.

5.3.3 Gestational age estimation

Figure 5.9 shows a scatter plot with the GA obtained from the reference HC using the standard plane and the GA obtained from the automatically estimated HC using the OSP data evaluated on the test set. The figure shows 29 patients, because the HC of one fetus was larger than the largest reported HC of Hadlock et al.[112], so for this fetus the GA could not be determined. The figure shows that most points fall within the P2.5-P97.5 interval of the Hadlock curve. The MD±SD between the reference GA and the automatically estimated GA was -3.6 ± 9.8 days.

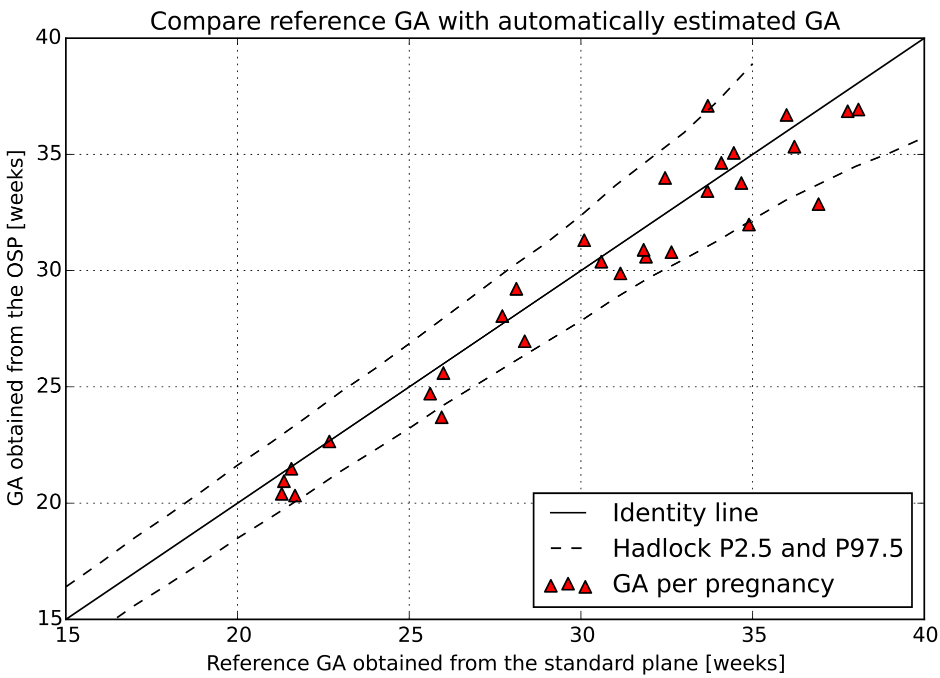

Figure 5.9: Comparison between reference GA and the GA obtained from the OSP on the test set. The red triangles represent the GA per pregnancy. The black dashed lines represent the ± 1.96 SD in GA, with an SD of 1 cm as reported by Hadlock et al.[112]

5.4 Discussion

Until now, ultrasound imaging has relied on a human sonographer to determine the standard plane to obtain biometric measurements of the fetus. In this study, we show that it is possible to measure the HC with the use of the OSP. A deep learning network was used to detect the frames in the OSP data which present the fetal head. Using a simple network architecture, a frame based accuracy of 0.97 was achieved on an independent test set. The frames in which the network detects the fetal head were used to automatically estimate the HC. The MAD between the automatically measured HC—estimated using the OSP—and the reference HC—obtained by an experienced sonographer in the standard plane—was 10.3 mm. This corresponded to a MAD in GA of 8.1 days, for which most estimated GAs fall within the P2.5-P97.5 interval of the curve of Hadlock et al.[112]

5.4.1 Fetal head detection

Figure 5.6 shows that the highest accuracy for detection of the fetal head using the VGG-Net inspired network on the validation set was achieved at downsampling factor six. Network architecture C achieved the highest frame based accuracy of 0.98 on the validation set, which results in an accuracy of 0.97 on the test set. It is interesting to see that accuracy of the deep learning network remains above 0.97 for both the validation and the test until a downsampling factor of fourteen. At a downsampling factor of fourteen, the image size is only 45×32 pixels. The performance decreases when the downsampling factor is larger than fourteen, which shows that the image quality became insufficient to detect the fetal head. There is also a performance drop when the downsampling factor is smaller than four, which could be caused by a decrease in the receptive field of the network. Figure 5.7 shows the labels and classification of network architecture C of two patients from the test set. The lowest accuracy within the test set of 0.861 was obtained from the patient shown on the right. In this case, each sweep contained frames of the fetal abdomen which were misclassified as the fetal head. Nevertheless, HC estimation was still possible as most of the frames that contained the fetal head were still detected. There was one fetus in the test set in which the deep learning network did not detect any frames with the fetal head present. This fetus had a reference HC of 116 mm, which was the smallest HC in the dataset (Figure 5.3). The deep learning network was not trained to classify frames containing a small fetal head as such. If more training data from the first and second trimester were available, this problem could possibly be solved.

5.4.2 Head circumference estimation

Figure 5.8 shows that the lowest MAD for measuring the HC using the U-net inspired network on the validation set was achieved at downsampling factor four, which resulted in an MAD of 10.3 mm on the test set. The literature reports different inter-observer variabilities for the HC measurement. Napolitano et al.[52] reported a

95% limits of agreement of 4.9%, Sarris et al.[50] reported a 1.96 SD of 12.1 mm, and Perni et al.[46] reported a 1.96 SD of 11.0 mm. The 1.96 SD of our system is 26.1 mm and 9.1%. This is more than double compared to most inter-observer variabilities reported in literature, which is caused by the fact that the OSP does not contain the correct cross section to measure the HC. The next paragraph will discuss how this increase influences the estimation of the GA.

5.4.3 Gestational age estimation

The result shows that the MAD on the test set is 8.1 days. Figure 5.9 shows that most estimated GA from the OSP fall within the P2.5-P97.5 interval of the curve of Hadlock et al.[112]. Most of the data for this study was acquired from pregnant women in the third trimester of pregnancy, since most pregnant women visit the hospital in Ethiopia around this time. Using the HC obtained in the third trimester is less reliable for estimating the GA, due to for example fetal growth restrictions[50]. Hadlock et al.[112] reported that the reliability (2 $\times$ the SD) for the GA can reach a maximum of 20.9 days. So even though the 1.96 SD of the estimated HC was twice as high compared to inter-observer variability reported in literature, it is still possible to estimate the GA within the P2.5-P97.5 confidence interval of the Hadlock curve. We therefore conclude that it is feasible to automatically estimate the GA by measuring the HC utilizing the OSP. The GA can be estimated more accurately by measuring the HC in the first and second trimesters of pregnancy, but this data was not available for this study. Future research should investigate how accurate this system can estimate the GA in the first and second trimester.

5.4.4 Deployment on low-cost hardware

Table 5.3 shows that best performing head detection architecture only contains 843 thousand parameters. Classification of one test frame with downsampling factor six using the training hardware took only 7.0×10^{-5} seconds. The best performing architecture for the HC measurement contains 1.9 million parameters, taking 5.0×10^{-3} seconds to classify one frame with downsampling factor four. The ultrasound device used in this study acquires 23 frames per second, so real time computation on the training hardware is feasible. Unfortunately, this high-end hardware will be too expensive for developing countries, but implementation of this system on a low-cost GPU—like the Intel Movidius—is feasible and could make this an affordable system for developing countries.

5.4.5 Improvements

It was not possible for an untrained midwife to acquire the data for this study in Ethiopia, since untrained midwives currently do not have access to an ultrasound device. In addition, an experienced sonographer was required to ensure that the standard plane was recorded. This study shows that the OSP can be used to automatically estimate the gestational age. A future study should evaluate this system with the use of data that was acquired by untrained midwives. All data in this study was acquired with one ultrasound device. Future studies should include data from different ultrasound devices to evaluate the system performance for different devices. Using the SonoAce R3, it took one minute to save one sweep; it was therefore considered impractical to acquire all six sweeps of the OSP. Still, all six sweeps are required to be able to estimate the HC for a fetus in shoulder presentation. In the future, an ultrasound device should be used that is able to record all six sweeps consecutively and save the data within a reasonable amount of time. Furthermore, an accurate measurement of the HC was not possible for two fetuses in the validation set, since the head of these fetuses was positioned deeper than 12 cm. We therefore recommend a standard depth of 15 cm for future studies. Next to this, it could be possible that the system measures the HC incorrectly due to a pathology. The user could therefore be presented with the frames that were used to determine the HC. This would give the user the possibility to check if these frames indeed contain the fetal head and therefore determine if the reported HC was correct.

5.5 Conclusion

We presented one deep learning network that detects the fetal head using the OSP and a second deep learning network that estimates the HC from the frames in which the fetal head was detected. This is the first method in literature that automatically measures the fetal HC with the use of a standardized sweep protocol that can be taught to any health care worker within one day. The deep learning network complexity was decreased and the input image was downsampled to decrease hardware demands and make deployment on low-cost hardware possible. The extensive evaluation on data acquired in Ethiopia shows that the head detection system achieves a frame based accuracy of 0.97 on the test set. The MAD between the reference HC and automated HC was 10.3 mm, equivalent to a GA MAD of 8.1 days, for which most estimated GAs fell within the P2.5-P97.5 interval of the Hadlock curve. This demonstrates the feasibility of this approach for a GA between 19 and 40 weeks.

6

Preforming prenatal ultrasound without a trained sonographer

T.L.A. van den Heuvel
H. Petros
S. Santini
C.L. de Korte
B. van Ginneken

Original title:
Performing prenatal ultrasound without a trained sonographer using deep learning and a standardized sweep protocol

Submitted to:
Ultrasound in Obstetrics and Gynecology

Abstract

Obstetric ultrasound imaging is commonly used to detect maternal risk factors, but often remains out of reach for pregnant women in resource-limited settings because of a severe shortage of well-trained sonographers to acquire and interpret the images. The obstetric sweep protocol (OSP), introduced by DeStigter et al. in 2011, consists of six predefined free-hand sweeps with the ultrasound transducer over the abdomen of the pregnant women. The OSP can be taught to any health care worker within a day and thus avoids the need of a trained sonographer to acquire ultrasound images. By combining the OSP with automated image analysis, this study investigates if it is possible to automatically determine maternal risk factors and therefore avoiding the need of a trained sonographer to both acquire and interpret the ultrasound images.

The OSP was acquired from 318 pregnant women using the low-cost MicrUs (TELEMED, Vilnius, Lithuania) in St. Luke's Hospital, Wolisso, Ethiopia. Image analysis systems were evaluated to automatically detect twin preg- nancies, estimate gestational age (GA) and determine fetal presentation. The GA was determined from the head circumference (HC) measurement using the Hadlock curve. The reference HC was obtained using the 2D standard plane, which was acquired by a trained sonographer.

The automated system was able to correctly detect 61% of all twins with a specificity of 99%. The GA was estimate with a median difference of -0.4 days and an interquartile range of 15.2 days compared to the reference GA estimated from the HC obtained in the 2D standard plane. All 31 fetuses in breech presentation were correctly detected and only one of the 216 fetuses in cephalic presentation was incorrectly classified as breech.

The results show that it is possible to automatically detect twins, estimate GA and determine fetal presentation using the OSP. The system therefore shows potential to detect these maternal risk factors without a trained sonographer. This approach could therefore vastly reduce time and costs that are required to train sonographers in resource-limited settings. This could potentially bring ultrasound in reach for pregnant women in resource-limited countries, making it possible to better manage obstetric care and refer pregnant women in time to a health care clinic to receive treatment if necessary.

6.1 Introduction

Worldwide, 99% of all maternal mortality, corresponding to approximately 820 deaths each day, occur in resource-limited settings[1]. Ultrasound imaging is commonly used to manage obstetric care and to detect maternal risk factors. Unfortunately, ultrasound remains out of reach for most women in resource-limited countries, mainly because it is too expensive and requires a trained sonographer to obtain and interpret the ultrasound images. Still, the WHO recommends an ultrasound examination for pregnant women in resource-limited settings to estimate gestational age (GA), determine fetal malpresentation and detect multiple pregnancies[2]. In the recent years, a large number of low-cost and portable ultrasound devices have been introduced to the market. Most of these devices can be connected to a tablet or smartphone, making them a suitable option for use in rural areas where there could be a lack of power supply. These systems still require a trained sonographer to obtain and interpret the ultrasound images. This is unfortunate because, there is a severe shortage of well-trained medical personnel in resource-limited settings[3–5]. Training a sonographer requires a significant investment of time and resources, which impedes the introduction of ultrasound in resource-limited settings. DeStigter et al.[6] introduced the obstetric sweep protocol (OSP), which consists of six predefined free-hand sweeps with the ultrasound transducer over the abdomen of the pregnant women (Figure 6.1). The OSP can be taught to any healthcare worker without knowledge of ultrasound within a day. The OSP therefore enables easy acquisition of prenatal ultrasound data. When there is an internet connection, the data can be send to a remote reading center where the images are interpreted. This remote reading center still requires a trained sonographer to interpret the OSP data. In this paper we combine the OSP with image analysis systems that are sufficiently simple that they can run on a smartphone or tablet, making it possible to automatically analyze the OSP without a trained sonographer. The automated analyzes can be performed on the device which was used to acquire the ultrasound data, which would also obviate the need of an internet connection. In this study we investigate if and which maternal risk factors can be automatically detected using the OSP. Previous work in this area is very limited. To the authors knowledge, only one paper was published which shows feasibility of automatically detecting the fetal hearth and fetal presentation using a single sweep acquired with a mid-range ultrasound device at the University of Oxford[101]. Unfortunately, it was not possible to determine if this system could detect intrauterine fetal death, because the data did not contain fetuses with a non-beating heart. The single sweep also did not contain either the fetal head or abdomen in 31% of the 129 test cases. The OSP, which was used in this study, contains six sweeps and therefore

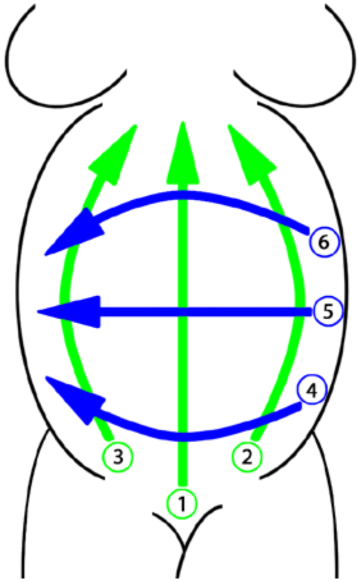

Figure 6.1: Visualization of the obstetric sweep protocol, consisting of six predefined free-hand sweeps with the ultrasound transducer over the abdomen of the pregnant woman.

ensures that the fetal head and abdomen are present in at least one of the six sweeps. In this study we propose a system that can automatically detect twin pregnancy, estimate GA and determine fetal presentation using the OSP, which obviates the need of a trained sonographer for these three tasks.

6.2 Methods

This section is divided into five paragraphs that explain the automated system. The first paragraph describes how each frame within the OSP data was automatically classified in one of six classes that include three fetal body parts and a class for frames where the ultrasound transducer was detached from the abdomen of the pregnant women. Secondly, the frames classified as ultrasound transducer detached were used to separate the OSP cine data into six separate sweeps. The last three paragraphs describe how the system detects twin pregnancies, estimate GA and determine fetal presentation with the use of the automated frame classification.

6.2.1 Data acquisition

The data of this study was acquired at St. Luke's Catholic Hospital and College of Nursing and Midwifery, Wolisso, Ethiopia. The collection of this data for this study was approved by the local ethics committee. The WHO estimated that there were 353 deaths per 100.000 live births in Ethiopia in 2015[1], so the acquired data in this study therefore originates from the target population. A trained gynecologist acquired both the OSP and the 2D standard plane, to obtain the reference HC, for a total of 318 pregnant women, using the low-cost MicrUs EXT-1H in combination with the C5-2R60S-3 transducer (TELEMED, Vilnius, Lithuania). The gynecologist was asked to acquire around 100 frames per sweep. All six sweeps were recorded in a single cine to make the acquisition time as short as possible. This reduces the possibility of fetal movement in between sweeps as much as possible. The imaging depth was set to 15 cm with an imaging angle of 65°. This resulted in a frame rate of 19 frames per second. Figure 6.2 shows a flowchart of the included data. A total of 38 cases had to be excluded for the reasons mentioned in Table 6.1. The remaining 280 cases included 33 twins, 216 fetuses in cephalic presentation and 31 fetuses in breech presentation. The GA was estimated using HC obtained in the 2D standard plane. The reference HC of the study population varied between 139 and 357 mm.

Table 6.1: Exclusion criteria for the obstetric sweep protocol data in this study

Exclusion criterium	**Number of cases**
Image quality was insufficient due to lack of ultrasound gel	12
The sweeps were not acquired in the correct order	6
The transducer was not detached from the abdomen	3
Fetus in transverse presentation	8
Fetus in oblique presentation	9
Total	**38**

6

6.2.2 Frame classification

The OSP data of the 280 included datasets contained a total of 209.642 frames, which were manually labeled in six different classes: fetal head completely present, fetal head partially present, fetal torso present, side view of the fetus present and ultrasound transducer detached. The remaining frames were classified as other. Fetal head completely present meant that the fetal head was full contained within the field of view (FOV) of the frame. Fetal head partially present meant that the fetal

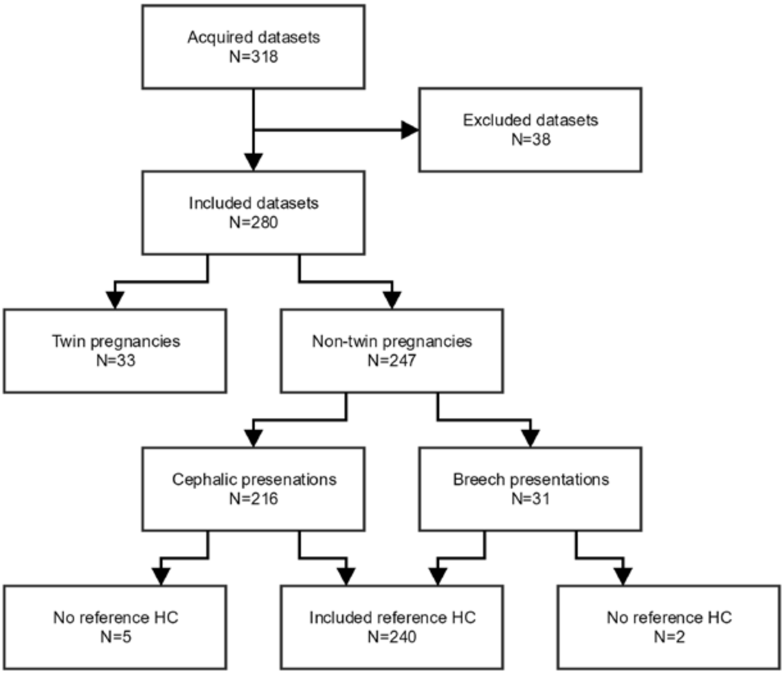

Figure 6.2: Flowchart of the datasets included in this study.

head falls partially outside of the FOV of the frame, which would make a circumference measurement inaccurate. Fetal torso meant a cross section through the fetal torso close to the transversal plane. Side view meant a section where the side of the fetus was visible. The ultrasound transducer was detached from the abdomen of the pregnant women in between the six sweeps, these frames were labeled as detached. A deep learning system inspired on the VGG-net of Simonyan and Zisserman [103] was developed to automatically classify each frame in the correct class.

6.2.3 Sweep separation

The six free-hand sweeps of the OSP were acquired after each other and saved in one cine. The frames classified as detached by the deep learning system were used to automatically separate this cine into six separate sweeps. Since the sweeps were made in free-hand mode, most sweeps did not contain exactly 100 frames. The sweeps were therefore resampled to 100 frames, using nearest-neighbor interpolation, after the sweeps were separated.

6.2.4 Detect twin pregnancies

Previous studies have shown that twin pregnancies increase the risk of significant maternal morbidity and mortality[113–115]. The detection of a twin pregnancy is the first step that makes it possible to monitor these pregnancies and refer these women in time to a health care clinic for delivery. The frame classification of the six sweeps was used to automatically detect twin pregnancies using a Random forest classifier[69]. It was assumed that the head and torso of the two fetuses could be separated using the frame classification, making it possible to automatically detect twins.

6.2.5 Estimate gestational age

Estimation of GA and fetal growth is essential for optimal obstetric management[39]. Ultrasound could especially be useful in resource-limited settings to accurately estimate gestational age, as menstrual dates may be incorrect or unknown[116]. In previous work it was shown that the OSP can be used to manually select an optimal frame of the fetal head from the sweep data to estimate GA[117]. In this work, a deep learning system inspired on the U-net of Ronneberger et al.[110] was used to automatically estimate the HC in all frames classified as fetal head completely present by the first deep learning system. The frames classified as fetal head partially present were not included, because an accurate HC measurement was not possible in these frames. The HC measured in the standard plane is one of the largest circumference one can measure from the fetal head, so the final HC was determined by taking the 75th percentile of all estimated HCs. The GA was determined from the HC using the curve of Hadlock et al.[112]. All 33 twin pregnancies were excluded for evaluation of the GA. Additionally, seven fetuses had to be excluded for evaluation of the GA, because the 2D reference standard was not acquired and therefore no ground truth HC available for these cases.

6.2.6 Determine fetal presentation

It was investigated if it is possible to automatically distinguish fetuses in cephalic presentation from fetuses in breech presentation. Although the published work in literature is still ambiguous whether a cesarean section should[118–121] or should not[122,123] be performed when the fetus is in breech presentation, it is clear that detection of breech presentation is required to plan the delivery. Early detection of breech presentation is especially important in resource-limited settings, where medical attention is not easily accessible[31,33,124]. In this work, the automated frame classification was used to distinguish fetuses in cephalic presentation from fetuses in breech presentation using a Random forest classifier[69]. The frame classification shows the location of the fetal head in relation to the fetal torso. A cephalic or breech presentation was detected when the fetal head was located below or above the fetal torso, respectively.

6.2.7 Statistical analysis

A five-fold cross-validation was used to be able to evaluate the results on all scans. Each fold, the algorithms were trained on 80% of the data and the remaining 20% was used for evaluation of the results. The number of twins, cephalic, breech and gestational age were balanced across the folds. The performance of the automated frame classification was evaluated using the accuracy score, which shows the percentage of the frames that were classified in the correct class, according to the manual labels. The mean and standard deviation were used to describe the results when the data was normally distributed. Otherwise the median and interquartile range (IQR) were used to describe the results.

6.3 Results

6.3.1 Frame classification

The deep learning network classifies each frame of the cine into one of six classes. The network was able to correctly classify 92.6% of all frames into the correct class. Table 6.2 shows the accuracy for all six classes separately. Figure 6.3.A. shows an example image for five frames from the OSP data of one patient which were classified in the correct class by the deep learning network. Figure 6.3.B. shows the frame classification for all frames of the OSP data of one patient. The six classes are depicted in a different color. The frames classified as detached (depicted in black) show where the six sweeps start and end. The five parts classified as detached were used to separate the cine into six separate sweeps. In this example, the first frame of sweep 1 was classified as fetal head completely present. The six separated and resamples sweeps are shown in Figure 6.3.C. From this figure it is possible to determine that this fetus lies in cephalic presentation, because the fetal head (depicted in dark and light blue) is located below the fetal torso (depicted in green).

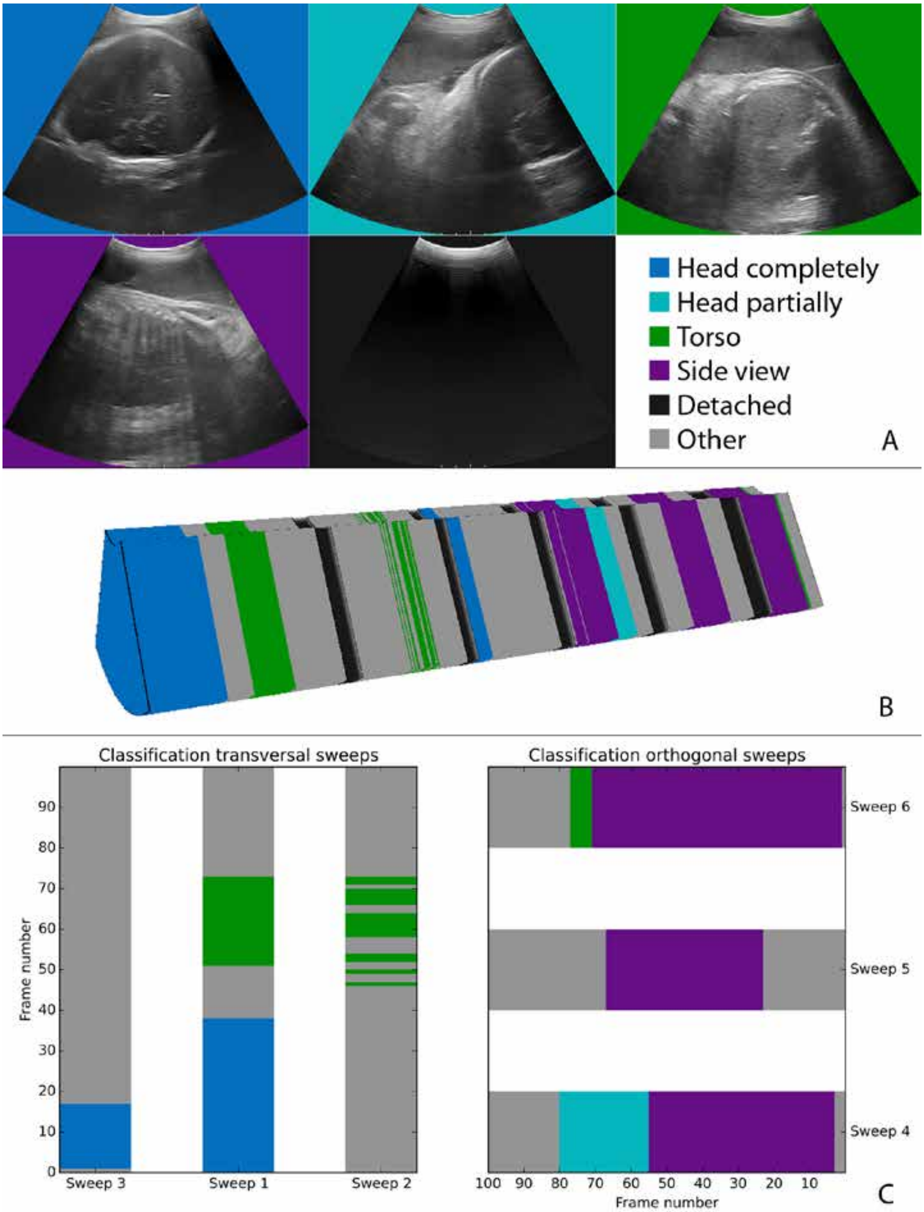

Figure 6.3: All three subfigures contain information extracted from the OSP of one pregnant woman. A: Five example frames, showing each a different class. B: The result of the automated frame classification for this pregnant woman, including all six sweeps and the detached frames in between (depicted in black). C: Frame classification after separation and resampling of the six sweeps. The is a fetus in cephalic presentation, which could be automatically detected because the fetal head (depicted in dark and light blue) was detected below the fetal torso (depicted in green).

6

Table 6.2: Result of the deep learning system for the frame classification of the sweep data

Class	Accuracy
Head completely	90.7%
Head partially	79.1%
Torso	81.1%
Side view	85.3%
Background	96.3%
Detached	97.3%

6.3.2 Detect twin pregnancies

The data for this study included 33 twins, of which 20 were correctly detected by the automated system (sensitivity of 61%). Figure 6.4 shows the automated frame classification of two correctly detected twins. For both correctly detected twins, one fetus lies in cephalic presentation and the other fetus in breech presentation, which makes it possible to distinguish the two fetuses with the use of the automated frame classification.

Figure 6.5.A shows the frame classification of a twin that was not detected by the automated system. In this example, both fetuses lie in cephalic presentation. The frame classification only shows the fetal head at the start sweep 1 and sweep 3 and the system was therefore not able to distinguish the two fetuses and detect the twin pregnancy. Figure 6.5.B. shows frame 32 of sweep 3 of this undetected twin. This frame clearly shows the torsos of two fetuses. This frame was therefore

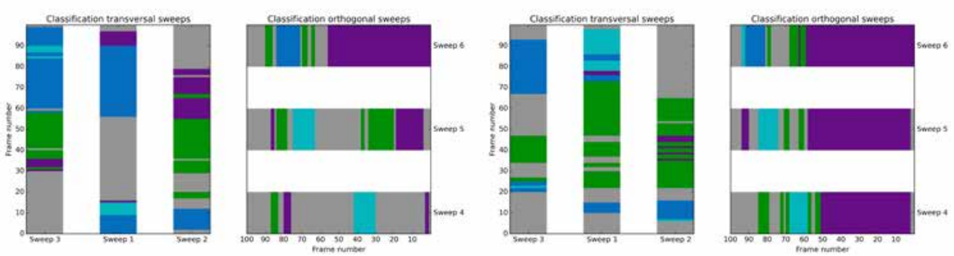

Figure 6.4: Frame classification of two correctly detected twins. For both twins, one fetus lies in cephalic presentation and the other fetus in breech presentation, which makes it possible to distinguish the two fetuses with the use of the automated frame classification.

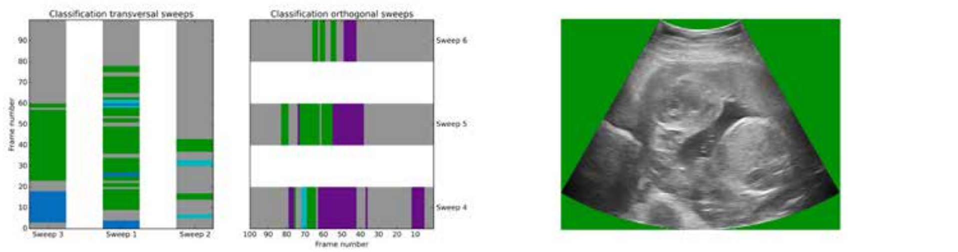

Figure 6.5: Left: Frame classification of an undetected twin pregnancy. Right: Frame 32 of sweep 3, which shows the torsos of two fetuses that both lie in cephalic presentation.

correctly classified as torso, but this was not sufficient to separate the two fetuses in order to detect the twin. Only three of the 247 non-twins were classified as twin, corresponding to a specificity of 99%. All these three fetuses showed an abnormality: one fetus in breech presentation, one fetus with lower urinary tract obstruction and one fetus with severe ascites (shown in Figure 6.8.A.).

6.3.3 Estimate gestational age

The standard plane—used to measure the reference HC—was acquired by the gynecologist for 240 fetuses (Figure 6.2). The HC of 15 fetuses fell outside the curve of Hadlock et al.[112], so the GA was estimated for the remaining 225 fetuses. Figure 6.6 shows a scatterplot with the results. The x-axis shows the reference GA that was determined from the HC obtained in the standard plane. The y-axis shows the GA that was determined from the automatically estimated HC utilizing the OSP. The curve of Hadlock has a confidence interval, which results from variations in the fetal head size and observer variability for measuring the HC. This confidence interval increases with GA which can also be seen in Figure 6.6. From all automatically estimated GAs, 87.6% fell within the p2.5%-p97.5% interval of the curve of Hadlock. A total of six automatically estimated GAs were larger than the p97.5% curve and a total of 22 automatically estimated GAs were smaller than the P2.5% curve. The median difference between the reference GA and the automated estimated GA was -0.4 days with an IQR of 15.2 days. This median difference was -0.4 days with an IQR of 9.3 days in the second trimester (14+0 until 27+6 week of gestation) and -0.4 days with an IQR of 16.7 days in the third trimester (28+0 until 40+0 week of gestation).

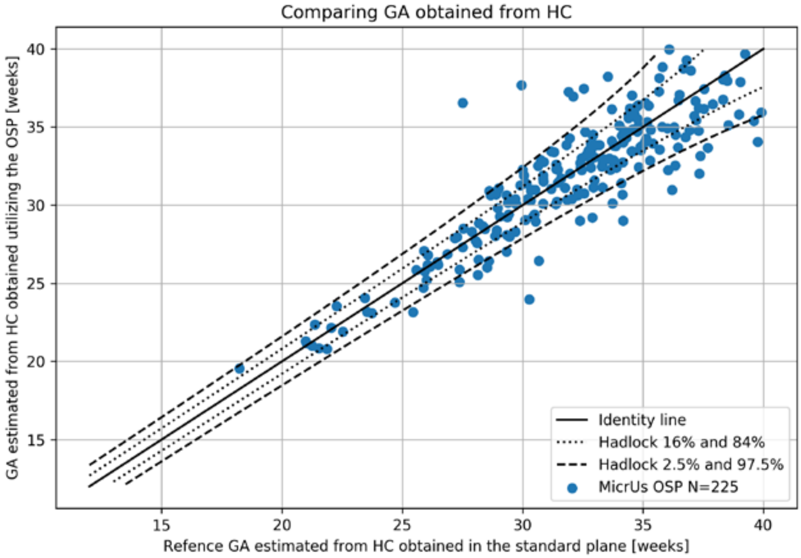

Figure 6.6: . Scatterplot with the reference GA on the x-axis and the automatically estimated GA on the y-axis. Two confidence intervals of the Hadlock curve are shown in black dashed lines.

6.3.4 Determine fetal presentation

A total of 216 cephalic and 31 breech presentations were evaluated in this study (Figure 6.2). All 31 fetuses in breech presentation were correctly detected by the automated system. Figure 6.7 shows two examples of correctly classified breech presentations, where the frames classified as fetal head is located above the fetal torso. Figure 6.3.C. shows an example of a fetus in cephalic presentation, where the frames classified as fetal head are located below the fetal torso. Only one fetus in cephalic presentation was misclassified as breech. This fetus showed severe ascites, which resulted in a misclassification of fetal abdomen and therefore breech presentation was incorrectly inferred. One frame of the OSP of this fetus is shown in Figure 6.8.A, which closely resembles a fetal head with hydrocephalus. Figure 6.8.B. shows a frame of one of the four fetuses with hydrocephalus. This frame was correctly classified as fetal head and shows resemblance with the Figure 6.8.A. Table 6.3 gives an overview of all abnormalities present in 33 of the 247 non-twin datasets (13%).

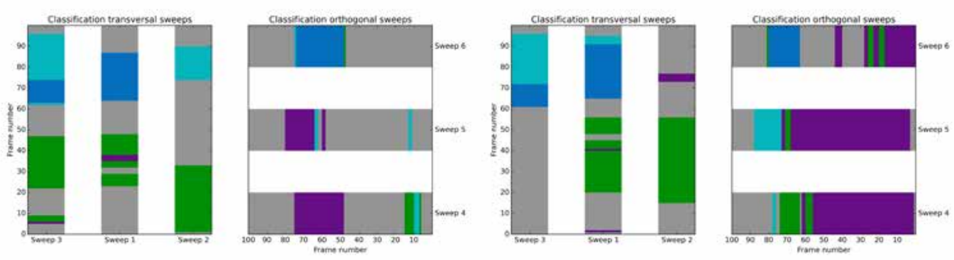

Figure 6.7: Frame classification of two fetuses in breech presentation that were correctly detected. The fetal head (depicted in dark and light blue) is located above the torso (depicted in green).

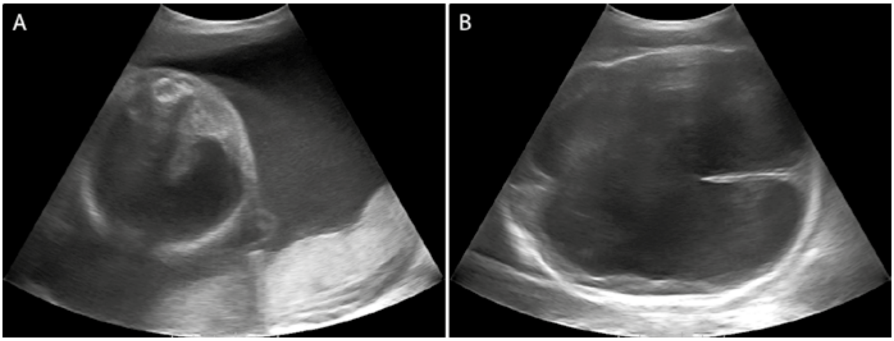

Figure 6.8: A: frame of the OSP data which shows the bladder abnormality of a fetus in cephalic presentation that was incorrectly classified as breech. B: Frame of the OSP data that shows a fetus with hydrocephalus that was correctly classified as head.

Table 6.3: Abnormalities that are present in the 247 non-twins

Abnormality	**Occurrence**
Antepartum hemorrhage	2
Placenta previa	8
Low-lying placenta	5
Abruption	1
Premature rupture of membranes	3
Oligohydramnios	5
Hydrocephalus	4
Intrauterine growth restriction	1
Macrosomia	1
Fetal ascites	1
Fetal lower urinary tract obstruction	1
Intrauterine Fetal Death	1
Total	**33**

6.4 Discussion

In this study we show that it is feasible to automatically detect twin pregnancies, determine GA and distinguish fetuses in cephalic presentation from fetuses in breech presentation with the use of six predefined free-hand sweeps acquired with a low-cost ultrasound device. The system therefore shows potential to detect maternal risk factors without a trained sonographer to both acquire and interpret the ultrasound images for these tasks, making wide spread use of this method feasible in resource-limited settings.

6.4.1 Detection twin pregnancies

The system was able to detect 61% of all twins. Even though not all twins were detected, this automated detection of twins makes it possible to detect the majority of the twin pregnancies which enables follow-up over time and the possibility to refer these women in time to a hospital for delivery. Further improvement of the system is required to be able to detect the other 39%. This could potentially be achieved by segmentation of the fetus within the frames. The result of such an algorithm will not only report which fetal body part is present in a frame, but will also segment the body part. This would make it possible to distinguish two fetuses when they lie in the same presentation, since it is possible to count the number of fetal body parts in one frame. The three non-twins that were classified as twin all presented an abnormality as mentioned in the Results section. More data from fetuses showing these abnormalities are required to resolve these misclassifications in the future. It should be noted that the included 280 datasets had a very high prevalence of twin pregnancies (12%).

6.4.2 Estimate gestational age

The GA could be automatically estimated with a median difference of -0.4 days with an IQR of 15.2 days compared to the reference GA. The confidence interval of the Hadlock curve shows that the GA can be estimated more accurately in the second trimester compared to the third trimester. The results also show that the IQR of the automated estimated GA in the second trimester of 9.3 days is smaller than the IQR of 16.7 days in the third trimester. This means that the system can automatically estimate the GA with a difference of less than five days compared to the reference HC in fifty percent of the cases. In previous work we have manually estimated the GA using the OSP acquired with the MicrUs from fifteen pregnant women in the Netherlands. The result showed a median difference of -0.2 days with an IQR of 4.4

days in the second trimester compared to the CRL measurement which was obtained in the first trimester[117].The CRL measurement was not performed for the studied population in this paper, so a direct comparison is not possible, but it indicates that there is room for improvement. It has to be investigated if these improvements will be made with a more accurate automated system, a better reference standard or with an improved OSP data acquisition.

6.4.3 Determine fetal presentation

The system was able to detect all 31 fetuses in breech presentation. Only one fetus in cephalic presentation was misclassified as breech. This fetus showed severe ascites, which resulted in a misclassification of fetal abdomen. The automated system is robust for other abnormalities, like hydrocephalus. The presentation was determined for all 247 non-twins with a GA variating from 17-40 weeks. The fetal presentation is not relevant at an earlier point in pregnancy since the fetus can freely move around, so the automatically estimated GA could potentially be used to determine whether the estimated fetal presentation is presented to the user.

6.4.4 Clinical implications

The present system can automatically detect three maternal risk factors. Any healthcare worker can be trained to obtain this data within a day, which would make widespread maternal risk screening with low-cost ultrasound much cheaper and easier to implement. When implementing such a system in resource-limited settings, it is very important to take into consideration how the patient management and follow-up is taken care of after an increased maternal risk is detected. This is out of the scope of this study. But detection is the first step to be able to set up obstetric management, which would include patient referral to health care centers or hospital to receive follow-up. Future improvements could bring the automated system close to the level of a trained sonographer for these tasks, but the automated system is at this moment not of added value when a trained sonographer is present. A trained sonographer can estimate the GA more accurately and will likely make less mistakes in determining the fetal presentation and in the detection of twins. Future research is required to investigate if it is possible to detect the heartbeat of the fetus in order to determine if the fetus is alive[101],if it is possible to estimate the amount of amniotic fluid, or if polyhydramnios could be detected using the OSP. The presented frame classification can be computed in real-time. This would make it possible to give real-time feedback to the user during the acquisition of the OSP. For example, the user could be notified which sweep to acquire next, when the transducer is detached from the abdomen of the pregnant women and when there is not enough gel used during acquisition of the data. This could further improve data acquisition of the OSP and decrease the number of scans that need to be excluded in future studies.

6.4.5 Study limitations

The GA in our study population varied between 17 and 40 weeks. It was therefore not possible to test the feasibility of this system in the first trimester. However, it is very difficult to obtain data of the first trimester in resource-limited settings since most pregnant women do not visit a health care clinic or hospital during the first trimester. We had to exclude 38 cases as described in Table 6.2. The automated detection of fetuses in transverse and oblique presentation was not possible, because not enough samples were available. More data would be required to ensure robust detection of fetuses in transverse and oblique presentation.

7

General discussion

In this thesis, we show that it is possible to automatically detect maternal risk factors with the use of a low-cost ultrasound device in a resource-limited setting. The algorithms presented in this thesis show the potential to automatically detect twin pregnancies, estimate gestational age and determine fetal presentation with the use of the obstetric sweep protocol (OSP). The OSP consists of six predefined free-hand sweeps over the abdomen of a pregnant woman, which can be taught to any health care worker within one day. The combination of the OSP and the computer algorithms developed in this thesis therefore obviates the need to extensively train a sonographer to obtain the correct images and detect these three maternal risk factors. This would make widespread obstetric ultrasound affordable and fast to implement in resource-limited settings, where there is a severe shortage of well-trained medical personnel.

Contrary to the teleradiology solution advocated by DeStigter et al.[6], this approach does not require an internet connection since the algorithms can run on the device itself. Together with the use of a low-cost ultrasound device that can be connected to a laptop or tablet makes this a portable solution that can be used in rural areas. When maternal risks are detected at an early stage in pregnancy, it allows pregnant women to be referred in time to a hospital to receive the medical attention they need. This is especially of importance in resource-limited settings, where health care services are limited and therefore could require a substantial amount of travel time to reach. This solution would give pregnant women enough time to travel to a medical facility to receive treatment and therefore has the potential to decrease the number of maternal deaths in resource-limited settings.

Deployment in resource-limited settings

We have shown feasibility of automated detection of three maternal risk factors with the use of the OSP that was acquired by a gynecologist in Wolisso, Ethiopia. Deployment of these algorithms in resource-limited settings is not an easy task that will include many challenges. In this paragraph we will only focus on the technical challenges of such a deployment.

The first step for deployment of this system in resource-limited settings would be integration of the automated image analysis algorithms with the ultrasound device on a laptop or tablet. This integration should include a robust encapsulation of the ultrasound device with the laptop or tablet. The algorithms were developed on high-end graphics cards after all the data was acquired. All algorithms were designed to run with low hardware requirements and it should therefore be possible to run the algorithms during acquisition of the ultrasound data. This would make it possible to give real-time feedback to the user when acquiring the OSP. An easy to use user interface should be designed to enable guided acquisition of the OSP. For example, the user should be notified which sweep to acquire next, get feedback when the ultrasound transducer does not make proper contact with the skin of the pregnant women, get a notification when a total of 100 frames is acquired and show the results of the algorithm on the screen after acquisition of all six sweeps is completed. The automated evaluation should also include quality control to ensure that the acquired sweep data is of sufficient quality to perform the measurements. Otherwise the user should be instructed immediately to acquire the OSP again.

A next step would be the introduction of this prototype to midwives in resource-limited settings. During this process it is very important to check if the midwives are able to correctly use the device. DeStigter et al.[6] mentioned that this OSP could be taught to any health care worker within a day. During my Ph.D. research period, I have visited the maternal health care clinic in Wolisso, Ethiopia, and provided training to local midwives who had never used ultrasound before. This training showed us that it was indeed possible to explain the basic procedures to them in an on-the-spot training session of a few hours.

If health care workers are allowed by local authorities to obtain the OSP, there should be a proper patient referral system that takes care of pregnant women when the system detects maternal risk factors. By closely monitoring the referred patients it would be possible to monitor the performance of the automated system and to improve it if necessary. Ultimately, it could be possible to follow-up referred pregnant women and check if the number of maternal mortality decreases as a consequence of the automated detection and inclusion of ultrasound in maternal care.

It is important to manage the expectations of the users when such a system is introduced. The system is not (yet) able to detect all maternal risk factors, and it is also not perfect in the tasks it can perform. It should therefore be clear to a health care worker what the system cannot do. For example: the automated system will not report the gender of the fetus. This is an important recommendation of the WHO on antenatal care for ultrasound[2]: "ultrasound sexing of the fetus in some low-income countries has a negative impact on gender equity and needs to be monitored". Since there is no trained sonographer required to use this system, it will ensure that there is no negative impact on gender equity.

Future research

The computer algorithms presented in this thesis did not perform perfectly and they do not detect all maternal risk factors that were recommended by the WHO[2], which leaves room for improvements in the future.

The computer algorithm presented in Chapter 6 was not able to detect twins when they are positioned in the same presentation. This could potentially be solved in the future by replacing the frame classification with a fetal segmentation. The result of such an algorithm will not only report which fetal body part is present in a frame, but should also segment the body parts. This would make it possible to distinguish two fetuses when they lie in the same presentation, since it is possible to count the number of fetal body parts in one frame.

Future research could investigate how accurate the fetal growth could be monitored when the gestational age is automatically estimated multiple times during one pregnancy. When the gestational age can be estimated very accurately, it would be possible to detect intra-uterine growth restrictions.

Improvements could also be made by automatically detecting other maternal mortality risks apart from the three risks that were investigated in this thesis. The report from the WHO[2] mentioned four other important maternal risk factors that are of importance in resource-limited settings which could be detected using ultrasound imaging: fetal anomalies, placenta previa, polyhydramnios and fetal viability. Maraci et al.[101] showed feasibility of automatically detecting the fetal heart using a single sweep acquired with a mid-range ultrasound device at the University of Oxford. However, they could not investigate if it was possible to detect intrauterine fetal death (IUF), because the data used in that study did not contain fetuses with a non-beating heart. The OSP data of the 247 included non-twins of Chapter 6 only included one IUF, which shows that it will be difficult to gather enough data for a robust evaluation of the detection of IUF.

Future research could also investigate if it is possible to use the OSP to automatically estimate the amount of amniotic fluid to detect polyhydramnios or oligohydramnios and to automatically determine the position of the placenta to detect placenta previa.

In this thesis we made use of the OSP as introduced by DeStigter et al.[6], which consists of six predefined sweeps over the abdomen of the pregnant woman. We did not investigate any improvements for this acquisition protocol. The automated frame classification enables real-time feedback to the user. This feedback could be used to optimize the OSP or to explore different strategies to acquire the data. The sweeps can be made quickly, so acquiring a few more sweeps is certainly feasible.

It would also be interesting to investigate if it is possible to create a 3D volume from the six sweeps. Prevost et al.[125] showed that it is possible to use deep learning to reconstruct 3D volumes from free-hand ultrasound sweeps. The six sweeps of the OSP contain some overlap, which could also be exploited to reconstruct a 3D volume and additional sweeps could be made as well. Such a 3D volume could potentially allow the 3D reconstruction of the complete fetus, which might make more accurate measurements of the fetus possible.

Low-cost ultrasound devices

In Chapter 2 of this thesis, the development of a very low-cost ultrasound device was described. The production cost of this device was aimed less than $100 and showed promising *in vivo* results for obstetric care. Unfortunately, Chapter 3 showed that the low frame rate impede the use of this ultrasound device to obtain the gestational age from the acquired OSP. It was therefore not possible to combine this device with the automated image processing algorithms developed in this thesis. Luckily, two other low-cost devices were able to acquire the OSP. Chapter 6 showed the use of one of these two devices in combination with the developed algorithms to automatically determine maternal risk factors using OSP data that was acquired from 280 pregnant women in Ethiopia. Unfortunately, this still leaves a gap between the aimed costs for ultrasound in resource-limited settings as introduced in Figure 1.4.

This gap might be closed in the future through the use of a capacitive micromachined ultrasonic transducer (CMUT). CMUTs have been introduced in the 1990s[126], but it took two decades before this technique reached the medical field. The advantage of CMUT is that the production costs are lower compared to the conventional transducers that use piezo-electric elements, because they can be integrated with the electronics on a single chip. An additional advantage is that these transducers have a wide frequency bandwidth. This makes it possible to image the complete human body with the use of one ultrasound probe, where the conventional piezo-electric element-based transducers require a different transducer when the imaging depth varies. In 2017, Butterfly Network introduced the iQ, which is a CMUT based ultrasound device that can be purchased for $2k. With this device they are not only the first company that can image the complete human body with one transducer, but the device can also be connected to a smartphone, which makes it possible to carry it in your pocket.

Ultrasound devices have become more portable in the last decades due to the innovations in portable devices like laptops, tablets and smartphones. This enables the development of more applications for ultrasound imaging, since you can bring the ultrasound device to any desirable location. Apart from the Butterfly Network iQ, there are many other vendors that have recently introduced portable ultrasound devices to the market: Philips Lumify, GE Vscan Extend, Telemed MicrUs, Interson SiMPLi and Clarius Wireless Scanner. This shows a trend which could lead to an increase in the point-of-care use of ultrasound imaging. The consequence of such a development is that either a lot more sonographers need to be trained to be able to use these devices, or sophisticated computer algorithms need to be developed to automatically obtain the correct image and interpret it for diagnosis.

Computer-aided diagnosis

During this PhD project, the field of computer-aided diagnosis in medical imaging has made a transition from classical machine learning towards deep learning. The classical machine learning approach required the process of manually crafting features for a specific task. Especially for ultrasound it is a difficult task to develop hand-crafted features that are robust to several artifacts commonly present in the images. This made the introduction of widely used algorithms for ultrasound images more challenging compared to other medical imaging modalities like computed tomography and magnetic resonance imaging. The introduction of deep learning made it possible to create algorithms that are robust to ultrasound artifacts which could be developed in a more time efficient way. This could boost the number of CAD systems for ultrasound images in the future.

In Chapter 4 we still used the classical machine learning approach for the measurement of the fetal head circumference, but this algorithm was replaced by a deep learning approach in Chapter 5. It took much less time to optimize the deep learning system compared to the classical machine learning approach. Another advantage was that the post-processing steps—like dynamic programming—were not required for robust estimation of the head circumference using the deep learning approach, since the deep learning network was more robust to artifacts compared to the random forest classification. The classical machine learning approach was still used to detect twin pregnancies and determine the fetal presentation given the frame classification in Chapter 6, since there was not enough data to train a deep learning network to perform these tasks.

Data availability is a very important to be able to train a deep learning network. In this thesis we have developed several deep learning networks that give high performance on an independent test set, using OSP data of a few hundred patients. But the amount of required data is very dependent on the task that is given to the system. All developed deep learning networks were only trained on data acquired with one ultrasound device, so it is unknown if these systems need to be retrained on data from different vendors. Transfer learning was not applied in this thesis, but this could be applied when the data limitation plays a limiting factor in future research.

High-end GPU cards were used for the development of the deep learning networks presented in this thesis, but all presented networks were reduced in size as much as possible to be able to use them with low-end GPU cards. It is important to realize that the training of deep learning networks takes much more time compared to the time for application of a trained network (*test time* is typically very low). But a GPU is probably required to be able to run these networks in real-time during acquisition of the data. Low-cost laptops and tablets currently do not include a GPU that can be used to run deep learning networks. Fortunately, Intel® introduced the Movidius™ Neural Compute Stick in 2017. This stick has the size of a USB stick and contains GPU card that is specially designed to run deep learning applications with low power consumption. Such a stick makes it possible to turn each low-cost device into a system for deep learning computations in a cheap way. With the additional advantage that it would not be required to send data into the cloud, which preserves patient privacy. Similar devices from other manufacturers are also available, and, in the near future, low-cost chips that can run reasonably simple networks may be included in low-cost hardware.

Beyond maternal care

The work described in this thesis is limited to automated image analysis for maternal care in resource-limited settings with 2D ultrasound. But the field of medical ultrasound is not limited to maternal care. In this paragraph we will look beyond maternal care to touch upon future developments of ultrasound and the role of CAD systems in this field.

In recent years, electronics have become smaller and devices have therefore become more portable. A good example is the smartphone, which was introduced around the beginning of this century. In less than two decades, the smartphone is present in the pocket of almost every individual each day. Computer processors have become more powerful and the well-known Moore's law[127] states that computer processors will double in power every two year. When combining more portable and powerful electronics with new techniques that make ultrasound transducers cheaper, it may be possible to use ultrasound everywhere. Roy Filly wrote an Editorial in Radiology in 1988 entitled Ultrasound: The Stethoscope of the Future, Alas[128] in which he states:

> *As we look to the proliferation of US instruments into the hands of untrained physicians, we can only come to the unfortunate realization that diagnostic sonography truly is the next stethoscope: used by many, understood by few.*

Today, 30 years later, there is an entire field called point-of-care ultrasound (POCUS), that brings ultrasound imaging to wherever a patient is being treated. This shows that ultrasound imaging is more widely used than ever[129,130]. Even though there are more physicians these days that know how to use ultrasound devices compared to 30 years ago, we can agree with Roy Filly that it is still very important to train a sonographer, because they need to know how to obtain and interpret the ultrasound images. In this thesis we show that it is possible to use ultrasound imaging without the use of a trained sonographer with the use of CAD systems. This shows that it will be possible to guide untrained users in obtaining useful images and aiding the user in making the correct diagnosis. With the use of deep learning, this approach could be applied for several other tasks. The iQ ultrasound transducer from Butterfly Network already uses computer algorithms to guide the user in obtained the correct imaging plane using guiding arrows on the screen of the smartphone and the software informs the user in real-time how useful the acquired image is.

The trend to make ultrasound devices more portable could result in new areas where ultrasound imaging could be useful for making a diagnosis. The football club PSV started to use the Philips Lumify in 2017 to make early diagnosis of injuries and provide acute care to their players. This gives the medical staff the option to make a point-of-care diagnosis. CAD systems could play a very important role in the next step for wide application of these portable ultrasound devices where easy to use ultrasound imaging could be beneficial. These systems could potentially aid untrained personnel to acquire ultrasound image with sufficient quality which could be automatically analyzed to give a diagnosis. One can think of a general practitioner, use at home, or even in a standard emergency kit. This could only be achieved when easy to use acquisition protocols can be performed by any user under the guidance of a CAD systems to acquire the ultrasound images in combination with algorithms that automatically interpret the images to form a diagnosis on the spot.

The decrease in cost of ultrasound devices could make 3D ultrasound more widely available. CAD systems could play a role in analyzing this data, to simplify the use of ultrasound and improve intra- and inter-observer variability. A reduction in human interaction could make ultrasound less observer dependent and make it more user friendly, which are currently two important limiting factors of ultrasound imaging. The 3D field of view would not require the observer to navigate to the correct cross section, but only require the user to place the ultrasound traducer close to the structure of interest. The computer algorithms could then be used to extract relevant information, or even to extract information that cannot be extracted with array transducers. This could potentially lead to improvements in clinical care as we know it today.

Summary

Worldwide, 99% of all maternal deaths occur in developing countries. In absolute numbers, this corresponds to approximately 820 deaths per day[1]. Ultrasound imaging can be used to detect maternal risk factors, but too often remains out of reach for pregnant women in developing countries. This is mainly caused by two reasons: ultrasound is too expensive for resource-limited countries and it requires a trained sonographer to acquire and interpret the ultrasound images, while there is a severe shortage of well-trained medical personnel in these countries[3–5]. In this thesis we aim to solve this problem by combining low-cost ultrasound devices with the obstetric sweep protocol (OSP) and automated image analysis. The OSP consists of six predefined free-hand sweeps with the ultrasound transducer over the abdomen of the pregnant woman. The OSP can be taught to health care workers without prior knowledge of ultrasound within one day and thus avoid the need of a trained sonographer to acquire ultrasound images. By combining the OSP with low-cost ultrasound devices and combining it with automated image analysis, it was investigated if it is possible to automatically determine maternal risk factors and therefore avoid the need of a trained sonographer to both acquire and interpret the ultrasound images.

Chapter 2 describes the development of a very low-cost medical ultrasound device with an aimed production cost of less than $100, which is an order of magnitude lower than any other ultrasound system on the market today. The hardware costs were reduced by replacing the array of piezo-electric elements by a single piezo-electric element, which was mechanically swept across the target scene. Next to this, synthetic aperture focusing was used instead of fixed focusing to form the ultrasound image. This approach was evaluated using simulations, phantom experiments and *in vivo* experiments. A prototype—the single element synthetic aperture scanner (SESAS)—was built to perform the *in vivo* measurements. The simulations and phantom experiments show that the achievable lateral resolution of the presented approach is superior compared to the fixed focus approach but also reveal a lower signal to noise ratio. Consequently, the *in vivo* acquisitions show limited application of the SESAS for clinical diagnostics in prenatal care.

Chapter 3 compares the SESAS to the SeeMore and the MicrUs which can be purchased around $2k to $3k. It is investigated how accurately a sonographer can estimate the gestational age (GA) using these low-cost ultrasound devices by measuring the fetal head circumference (HC), abdominal circumference (AC) and femur length (FL) using both the standard plane and the OSP. The GA was estimated with the curve of Verburg et al.[39]. The results show that the HC, AC and FL can be used to estimate the GA using all three low-cost ultrasound devices from the standard plane within the inter-observer variability presented in literature. The OSP can be used to estimate the GA with the HC and AC, but not the FL using the SeeMore and the MicrUs. The frame rate of the SESAS was too low to estimate the GA from the OSP.

Chapter 4 presents a computer aided detection (CAD) system for automated measurement of the fetal HC in 2D ultrasound images obtained in the standard plane for all trimesters of the pregnancy. The data for this study was acquired with a high-end ultrasound device in the Netherlands. The CAD system makes use of a classical machine learning approach, in which Haar-like features were computed from the ultrasound images to train a random forest classifier to locate the fetal skull. The HC was extracted using the Hough transform, dynamic programming and an ellipse fit. The results show that the CAD system performs comparable to an experienced sonographer and shows similar or superior results compared to systems published in literature. This is the first automated system for HC assessment that is evaluated on a large test set which contains data of all trimesters of the pregnancy.

Chapter 5 presents two deep learning networks for automated estimation of the GA by estimating the HC with the use of the OSP. The data for this study was acquired with a mid-range ultrasound device in Ethiopia. First, a VGG-Net[103] inspired network was trained to automatically detect the frames in the OSP data which contained the fetal head. Second, a U-net[110] inspired network was trained to automatically measure the HC for all frames in which the first network detected a fetal head. The HC was estimated from these frame measurements and the curve of Hadlock et al.[112] was used to determine the GA. The results show that it is possible to automatically estimate the GA from the OSP data that was obtained in Ethiopia.

Chapter 6 presents an extension of the system presented in Chapter 5. This system cannot only estimate the GA, but also detects twin pregnancies and determine fetal presentation using the OSP which was acquired with a low-cost ultrasound device in Ethiopia. The automated system is able to correctly detect 61% of all twins with a specificity of 99%. The GA can be estimated with a median difference of -0.4 days and an interquartile range of 15.2 days. The system was able to correctly detect all 31 fetuses in breech presentation and only incorrectly classified one of the 216 fetuses in cephalic presentation as breech.

The results of this thesis show the potential use of image analysis to automatically detect twins, estimate GA and determine fetal presentation using the OSP, making it possible to detect these risk factors without a trained sonographer. This approach can therefore vastly reduce time and costs that are required to train sonographers in resource-limited settings. The presented system could potentially bring ultrasound in reach for pregnant women in resource-limited countries, making it possible to better manage obstetric care and refer pregnant women in time to a health care clinic to receive required treatment when maternal risk factors are detected.

Samenvatting

Wereldwijd vindt 99% van alle moedersterfte plaats in ontwikkelingslanden. In absolute getallen komt dit overeen met ongeveer 820 sterfgevallen per dag[1]. Ultrageluid kan worden gebruikt om zwangerschapsrisicos te detecteren, maar deze techniek blijft vaak buiten het bereik van zwangere vrouwen in ontwikkelingslanden. Dit wordt voornamelijk veroorzaakt door twee problemen. Enerzijds is echografie te duur voor landen met beperkte middelen. Daarnaast is er een echografist nodig om de echobeelden te maken en deze te interpreteren, terwijl er in deze landen een ernstig tekort is aan goed opgeleid medisch personeel. In dit proefschrift willen we deze twee problemen oplossen door het combineren van goedkope echoapparatuur met het verloskundig opname protocol (VOP) en geautomatiseerde beeldanalyse. Het VOP bestaat uit zes vooraf gedefinieerde opnamebewegingen met de echotransducer over de buik van de zwangere vrouw. Het VOP kan binnen één dag aan een vroedvrouw worden geleerd die geen voorafgaande kennis van ultrageluid heeft, waardoor het niet nodig is om een echografist op te leiden voor het maken en het interpreteren van de echobeelden.

Hoofdstuk 2 beschrijft de ontwikkeling van een zeer goedkoop medisch echoapparaat met een beoogde kostprijs van minder dan 100 USD per stuk. Dit is een order van grootte lager dan enig ander echoapparaat dat momenteel op de markt verkrijgbaar is. De kosten zijn gereduceerd door het vervangen van de rij piëzo-elektrische elementen door één enkel piëzo-elektrisch element dat mechanisch wordt bewogen over het af te beelden weefsel. Daarnaast is er gebruik gemaakt van *synthethic aperture focusing* in plaats van een vast focus om het echobeeld te vormen. Deze aanpak is geëvalueerd met behulp van simulaties, en opnames van fantomen en van zwangere vrouwen. Voor deze opnames is een prototype, genaamd *single element synthethic aperture scanner* (SESAS), ontwikkeld. De simulaties en de opnames van de fantomen tonen aan dat de laterale resolutie van de gepresenteerde aanpak beter is dan een vast focus, maar laat ook een lagere signaal-ruisverhouding zien. De SESAS toonde gelimiteerde toepassingen voor het gebruik van dit apparaat voor prenatale diagnostiek.

Hoofdstuk 3 vergelijkt de SESAS met twee andere goedkoper echoapparaten, de SeeMore en de MicrUs, die voor ongeveer twee- tot drieduizend USD kunnen worden aangeschaft. Er is onderzocht hoe nauwkeurig een echografist met behulp van deze goedkope echoapparatuur de zwangerschapsduur (ZD) kan bepalen door het meten van de foetale hoofdomtrek (HO), de buikomtrek (BO) en de femurlengte (FL) met behulp van zowel de standaarddoorsnede als het VOP. De ZD is bepaald met behulp van de curve van Verburg et al.[39]. De resultaten tonen aan dat de HO, BO en FL kunnen worden gebruikt om de ZD binnen de inter-observer variabiliteit af te schatten met behulp van de standaarddoorsnede die is opgenomen met een goedkoop echoapparaat. Het VOP kan worden gebruikt om de ZD te schatten met de HO en de BO, maar niet de FL met behulp van de SeeMore en de MicrUs. De beeld verversingsfrequentie van de SESAS was te laag om de ZD te schatten met het VOP.

Hoofdstuk 4 presenteert een computeralgoritme die de foetale HO automatisch kan meten in 2D echobeelden in alle trimesters van de zwangerschap. De beelden voor deze studie zijn opgenomen met een high-end echoapparaat in Nederland. Het algoritme maakt gebruik van een klassieke *machine learning* aanpak waarbij *Haar-like features* werden berekend op basis van de echobeelden om een *random forest classifier* te trainen om de schedel van de foetus te lokaliseren. De HO werd geëxtraheerd met behulp van de *Hough transform*, *dynamic programming* en het fitten van een ellips. Uit de resultaten blijkt dat de HO gemeten door het algoritme vergelijkbaar is met de HO gemeten door een ervaren echografist. Het algoritme toont vergelijkbare of betere resultaten in vergelijking met de systemen die in de literatuur zijn gepubliceerd. Dit is bovendien het eerste geautomatiseerde systeem voor het meten van de HO dat is geëvalueerd op een grote onafhankelijke test-set van 335 beelden afkomstig van alle trimesters van de zwangerschap.

Hoofdstuk 5 presenteert twee *deep learning* netwerken voor een geautomatiseerde schatting van de ZD door het schatten van de HO met behulp van het VOP. De data voor deze studie is opgenomen met een middelklasse echoapparaat in Ethiopië. Het eerste *deep learning netwerk* is geïnspireerd op het VGG-Net[103] en is getraind om automatisch de beelden in de VOP te detecteren die het hoofd van de foetus bevatten. Het tweede *deep learning netwerk* is geïnspireerd op het U-net[110] en is getraind om automatisch de HO te meten in alle beelden waar het eerste netwerk het hoofd van foetus detecteerde. De ZD is vanuit deze metingen geschat met behulp van de curve van Hadlock et al.[112]. De resultaten tonen aan dat het mogelijk is om de ZD automatisch te schatten met behulp van het VOP dat is opgenomen in Ethiopië.

Hoofdstuk 6 presenteert een uitbreiding van het systeem dat in hoofdstuk 5 is gepresenteerd. Dit systeem kan niet alleen de ZD schatten, maar kan ook tweelingen detecteren en de ligging van de foetus bepalen met behulp van het VOP dat was opgenomen met een goedkoop echoapparaat in Ethiopië. Het geautomatiseerde systeem is in staat om 61% van alle tweelingen te detecteren met een specificiteit van 99%. De ZD kan worden geschat met een verschil in de mediaan van -0,4 dagen en een interkwartielafstand van 15,2 dagen. Het systeem was in staat om alle 31 foetussen in stuitligging correct te detecteren en classificeerde slechts één foetus in achterhoofdsligging ten onrechte als stuitligging.

De resultaten van dit proefschrift tonen de dat het mogelijk is om met beeldanalyse algoritme automatisch tweelingen te detecteren, de ZD te schatten en de ligging van de foetus te bepalen met behulp van het VOP. Hierdoor is het mogelijk om deze risicofactoren te detecteren zonder een echografist te hoeven opleiden. Deze aanpak kan de tijd en kosten die nodig zijn om echografisten op te leiden in ontwikkelingslanden aanzienlijk verminderen. Het gepresenteerde systeem heeft de potentie om ultrageluid te introduceren bij zwangere vrouwen in ontwikkelingslanden, waardoor het mogelijk wordt om de verloskundige zorg te verbeteren en zwangere vrouwen op tijd door te verwijzen naar een ziekenhuis indien er zwangerschapsrisicos gedetecteerd zijn.

Publications

Papers in international journals

T.L.A. van den Heuvel, H. Petros, S. Santini, C.L. de Korte and B. van Ginneken. ”Performing prenatal ultrasound without a trained sonographer using deep learning and a standardized sweep protocol”, *Submitted to Ultrasound in Obstetrics and Gynecology*, 2018.

T.L.A. van den Heuvel, H. Petros, S. Santini, C.L. de Korte and B. van Ginneken. ”Automated fetal head detection and circumference estimation from free-hand ultrasound sweeps using deep learning in resource-limited countries”, *Accepted for publication in Ultrasound in Medicine and Biology*, 2018.

T.L.A. van den Heuvel, D. de Bruijn, C.L. de Korte and B. van Ginneken. ”Automated measurement of fetal head circumference using 2D ultrasound images”, *PLoS One*, 13(8):e0200412, 2018.

T.L.A. van den Heuvel, D. de Bruijn, D. Moens-van de Moesdijk, A. Beverdam, B. van Ginneken and C.L. de Korte. ”Comparison study of low-cost ultrasound devices for estimation of gestational age in resource-limited countries”, *Ultrasound in Medicine and Biology*, 44(11):2250-2260, 2018.

T.L.A. van den Heuvel, D.J. Graham, K.J. Smith, C.L. de Korte and J.A. Neasham. ”Development of a low-cost medical ultrasound scanner using a monostatic synthetic aperture”, *IEEE Transactions on Biomedical Circuits and Systems*, 11(4):849-857, 2017.

T.L.A. van den Heuvel, A.W. van der Eerden, R. Manniesing, M. Ghafoorian, T. Tan, T.M.J.C. Andriessen, T.V. Vyvere, L. van den Hauwe, B.M. ter Haar Romeny, B.M. Goraj and B. Platel. ”Automated detection of cerebral microbleeds in patients with Traumatic Brain Injury”, *NeuroImage: Clinical*, 12:241-251, 2016.

Papers in conference proceedings

K. Standvoss, T. Crijns, L. Goerke, D. Janssen, S. Kern, T. van Niedek, J. van Vugt, N.A. Burgos, E.J. Gerritse, J. Mol, D. van de Vooren, M. Ghafoorian, **T.L.A. van den Heuvel** and R. Manniesing. ”Cerebral microbleed detection in traumatic brain injury patients using 3D convolutational neural networks”, In: *Medical Imaging*, volume 10575 of Proceedings of the SPIE, 2018.

T.L.A. van den Heuvel, H. Petros, S. Santini, C.L. de Korte, and B. van Ginneken. ”A step towards measuring the fetal head circumference with the use of obstetric ultrasound in a low resource setting”, In: *Medical Imaging*, volume 10139 of Proceedings of the SPIE, 2017.

T.L.A. van den Heuvel, H. Petros, S. Santini, C.L. de Korte, and B. van Ginneken. ”Combining automated image analysis with obstetric sweeps for prenatal ultrasound imaging in developing countries”, In: *MICCAI Workshop: Point-of-Care Ultrasound*, pages 105-112 of Lecture Notes in Computer Science, 2017.

T.L.A. van den Heuvel, M. Ghafoorian, A.W. van der Eerden, B.M. Goraj, T.M.J.C. Andriessen, B.M. ter Haar Romeny and B. Platel. ”Computer aided detection of brain micro-bleeds in traumatic brain injury”, In: *Medical Imaging*, volume 9414 of Proceedings of the SPIE, 2015.

Abstracts in conference proceedings

T.L.A. van den Heuvel, B. van Ginneken, and C.L. de Korte. ”Automated detection of fetal presentation and gestational age using low-cost ultrasound and deep learning in a resource-limited setting”, In: *IEEE International Ultrasonics Symposium*, 2018.

A.W. van der Eerden, **T.L.A. van den Heuvel**, B.H. Geurts, B. Platel, T.V. Vyveree, L. van den Hauwee, T.M. Andriessen, B.M. Goraj and R. Manniesing. ”Automatic versus human detection of traumatic cerebral microbleeds on susceptibility weighted imaging”, In: *European Congress of Radiology*, 2018.

T.L.A. van den Heuvel, C.L. de Korte and B. van Ginneken. ”Automated measurement of fetal head circumference in ultrasound images”, In: *Dutch Biomedical Engineering Conference*, 2017.

Awards

Best Oral Presentation. ”Automated detection of maternal risk factors in developing countries”, In: *Radboud Institute for Health Sciences PhD Retreat*, 2017

Robert F. Wagner Best Student Paper Award, Ultrasonic Imaging and Tomography Conference Finalist. ”A step towards measuring the fetal head circumference with the use of obstetric ultrasound in a low resource setting", In: *Medical Imaging*, volume 10139 of Proceedings of the SPIE, 2017.

Best Poster Award Finalist. "Automated measurement of fetal head circumference in ultrasound images", In: *Dutch Biomedical Engineering Conference*, 2017.

PhD portfolio

Name PhD candidate:	T.L.A. van den Heuvel	
Department:	Radiology and Nuclear Medicine	
Graduate School:	Radboud Institute for Health Sciences	
PhD period:	16-03-2015 - 16-08-2018	
Promotors:	Prof. dr. ir. B. van Ginneken	
	Prof. dr. ir. C.L de Korte	

	Year(s)	ECTS
Training activities		
a) Courses & Workshops		
- Basiscursus Regelgeving en Organisatie voor Klinisch onderzoekers	2016	2.0
- Scientific Integrity	2016	1.0
- Scientific Writing for PhD Candidates	2016	3.0
- ASCI Advanced Pattern Recognition Course	2016	5.0
- Deep Learning 101 Workshop	2016	3.0
- Opfriscursus Statistiek voor Promovendi	2016	1.5
- RIHS introduction course for PhD students	2016	1.0
- NFBIA Summer School	2015	1.5
- Principes en kwaliteit van medische ultrageluid apparatuur	2015	1.0
- General introduction day Radboudumc	2015	0.3
b) Seminars & lectures		
- NFBIA symposium	2017	0.3
- NVKF conference	2017	0.3
- Medical Imaging Symposium for PhD Students	2016	0.5
- Seminar Machine Learning and Health	2016	0.3
- NVPHBV meeting	2015	0.5
- Radboud New Frontiers	2016,2018	1.0
- Annual DIAG-FME symposium	2015-2018	1.5
- Bianual NVMU meeting	2015-2018	3.5
c) Symposia & congresses		
- IEEE International Ultrasonics Symposium†	2018	3.0
- Radboud Institute for Health Sciences PhD Retreat†	2017	1.0
- POCUS satellite event of MICCAI Conference†	2017	3.0
- SPIE Medical Imaging Conference†	2017	3.0
- Dutch Bio-Medical Engineering Conference‡	2017	1.0
d) Other		
- Contestant Radboud Talks Grand Finale	2017	1.5
- Weekly research meeting	2015-2018	7.0
- Weekly DIAG discussion hour	2015-2018	7.0
- Weekly MUSIC research meeting	2015-2018	7.0
Teaching activities		
e) Supervision		
- Supervisor of the course Intelligent Systems in Medical Imaging	2017, 2018	4.0
Total		**64.7**

†Indicates an oral presentation

‡Indicates a poster presentation

Bibliography

[1] WHO, UNICEF, UNFPA, The World Bank, and the United Nations Population Division. Trends in maternal mortality: 1990 to 2013: estimates by WHO, UNICEF, UNFPA, The World Bank and the United Nations Population Division, 2014.

[2] World Health Organization. *WHO recommendations on antenatal care for a positive pregnancy experience*. World Health Organization, 2016.

[3] LaGrone L. N., Sadasivam V., Kushner A. L., and Groen R. S. A review of training opportunities for ultrasonography in low and middle income countries. *Tropical Medicine and International Health*, 17(7):808–819, 2012.

[4] Carrera J. M. Obstetric ultrasounds in africa: Is it necessary to promote their appropriate use? *Donald School Journal of Ultrasound in Obstetrics and Gynecology*, 5(3):289–296, 2011.

[5] Hofmeyr G. Routine ultrasound examination in early pregnancy: is it worthwhile in low-income countries? *Ultrasound in Obstetrics and Gynecology*, 34(4):367–370, 2009.

[6] DeStigter K. K., Morey G. E., Garra B. S., Rielly M. R., Anderson M. E., Kawooya M. G., Matovu A., and Miele F. R. Low-cost teleradiology for rural ultrasound. In *IEEE Global Humanitarian Technology Conference*, pages 290–295, 2011.

[7] van Ginneken B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiological Physics and Technology*, 10(1):23–32, 2017.

[8] Litjens G., Kooi T., Ehteshami Bejnordi B., Setio A. A. A., Ciompi F., Ghafoorian M., van der Laak J., van Ginneken B., and Sánchez C. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017.

[9] Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 1980.

[10] Goodfellow I., Bengio Y., and Courville A. *Deep Learning*. MIT Press, 2016.

[11] Li M., McGuire M., Ho K. S., and Hayward G. Array element failure correction for robust ultrasound beamforming and imaging. In *IEEE International Ultrasonics Symposium*, pages 29–32, 2010.

[12] Fort J. R., Neidell N. S., Morgan D. J., and Landmeier P. C. Synthetic aperture ultrasound imaging system, November 14 1995. U.S. Patent 5465722.

[13] Bae M. H. and Jeong M. K. A study of synthetic-aperture imaging with virtual source elements in b-mode ultrasound imaging systems. *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, 47(6):1510–1519, 2000.

[14] Jensen J. A., Nikolov S. I., Gammelmark K. L., and Pedersen M. H. Synthetic aperture ultrasound imaging. *Ultrasonics*, 44:e5–e15, 2006.

[15] Korukonda S. and Doyley M. M. Estimating axial and lateral strain using a synthetic aperture elastographic imaging system. *Ultrasound in Medicine and Biology*, 37(11):1893–1908, 2011.

[16] Bottenus N., Long W., Zhang H. K., Jakovljevic M., Bradway D. P., Boctor E. M., and Trahey G. E. Feasibility of swept synthetic aperture ultrasound imaging. *IEEE Transactions on Medical Imaging*, 35(7):1676–1685, 2016.

[17] Burckhardt C. E., Grandchamp P. A., and Hoffmann H. An experimental 2 mhz synthetic aperture sonar system intended for medical use. *IEEE Transactions on Sonics and Ultrasonics*, 21

(1):1–6, 1974.

[18] Kortbek J., Jensen J. A., and Gammelmark K. L. P2b-1 synthetic aperture focusing applied to imaging using a rotating single element transducer. In *IEEE International Ultrasonics Symposium*, pages 1504–1507, 2007.

[19] Opretzka J., Vogt M., and Ermert H. A synthetic aperture focusing technique with optimized beamforming for high-frequency ultrasound. In *IEEE International Ultrasonics Symposium*, pages 2303–2306, 2010.

[20] Opretzka J., Vogt M., Maschauer S., Prante O., and Ermert H. High frequency ultrasonic imaging based on a combination of synthetic aperture focusing and limited angle compounding: Small animals imaging results. In *European Conference on Synthetic Aperture Radar*, pages 38–41. VDE, 2012.

[21] Andresen H., Nikolov S. I., and Jensen J. A. Synthetic aperture focusing for a single-element transducer undergoing helical motion. *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, 58(5):935–943, 2011.

[22] Neasham J. A. and Graham D. J. Ultrasound imaging apparatus, October 20 2015. U.S. Patent 9161742.

[23] Schueler C. F., Lee H., and Wade G. Fundamentals of digital ultrasonic imaging. *IEEE Transactions on Sonics and Ultrasonics*, 31(4):195–217, 1984.

[24] Smith K. J., Graham D. J., and Neasham J. A. Design and optimization of a voice coil motor with a rotary actuator for an ultrasound scanner. *IEEE Transactions on Industrial Electronics*, 62 (11):7073–7078, 2015.

[25] Bloomfield P. *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons, 2004.

[26] Czerwinski R. N., Jones D. L., and O'Brien W. D. Detection of lines and boundaries in speckle images-application to medical ultrasound. *IEEE Transactions on Medical Imaging*, 18(2):126–136, 1999.

[27] Jensen J. A. and Svendsen N. B. Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, 39(2):262–267, 1992.

[28] Jensen J. A. Field: A program for simulating ultrasound systems. In *Nordicbaltic Conference on Biomedical Imaging*, pages 351–353. Citeseer, 1996.

[29] Thijssen J. M., Weijers G., and de Korte C. L. Objective performance testing and quality assurance of medical ultrasound equipment. *Ultrasound in Medicine and Biology*, 33:460–471, 2007.

[30] Hadlock F. P., Harrist R. B., Sharman R. S., Deter R. L., and Park S. K. Estimation of fetal weight with the use of head, body, and femur measurements-a prospective study. *American Journal of Obstetrics and Gynecology*, 151(3):333–337, 1985.

[31] Wanyonyi S. Z., Mariara C. M., Vinayak S., Stones W., Ostensen H., Sippel S., Muruganandan K., Levine A., Shah S., and de Bernis L. Opportunities and challenges in realizing universal access to obstetric ultrasound in sub-saharan africa. *Ultrasound International Open*, 3(02):E52–E59, 2017.

[32] Aliyu L. D., Kurjak A., Wataganara T., de Sá R. A. M., Pooh R., Sen C., Ebrashy A., Adra A.,

and Stanojevic M. Ultrasound in africa: what can really be done? *Journal of Perinatal Medicine*, 44(2):119–123, 2016.

[33] Stanton K. and Mwanri L. Global maternal and child health outcomes: the role of obstetric ultrasound in low resource settings. *World Journal of Preventive Medicine*, 1(3):22–29, 2013.

[34] Sippel S., Muruganandan K., Levine A., and Shah S. Use of ultrasound in the developing world. *International Journal of Emergency Medicine*, 4(1):72, 2011.

[35] Gladstone M., White S., Kafulafula G., Neilson J. P., and Broek N. V. D. Post-neonatal mortality, morbidity, and developmental outcome after ultrasound-dated preterm birth in rural malawi: a community-based cohort study. *PLoS Medicine*, 8(11):e1001121, 2011.

[36] Shah S., Noble V. E., Umulisa I., Dushimiyimana J. M. V., Bukhman G., Mukherjee J., Rich M., and Epino H. Development of an ultrasound training curriculum in a limited resource international setting: successes and challenges of ultrasound training in rural rwanda. *International Journal of Emergency Medicine*, 1(3):193–196, 2008.

[37] Harris R. D. and Marks W. M. Compact ultrasound for improving maternal and perinatal care in low-resource settings. *Journal of Ultrasound in Medicine*, 28(8):1067–1076, 2009.

[38] Kotlyar S. and Moore C. L. Assessing the utility of ultrasound in liberia. *Journal of Emergencies Trauma and Shock*, 1(1):10, 2008.

[39] Verburg B. O., Steegers E. A. P., Ridder M. D., Snijders R. J. M., Smith E., Hofman A., Moll H. A., Jaddoe V. W. V., and Witteman J. C. M. New charts for ultrasound dating of pregnancy and assessment of fetal growth: longitudinal data from a population-based cohort study. *Ultrasound in Obstetrics and Gynecology*, 31(4):388–396, 2008.

[40] AIUM. Aium practice guideline for the performance of obstetric ultrasound examinations. *Journal of Ultrasound in Medicine*, 32(6):1083, 2013.

[41] van den Heuvel T. L. A., Graham D. J., Smith K. J., de Korte C. L., and Neasham J. A. Development of a low-cost medical ultrasound scanner using a monostatic synthetic aperture. *IEEE Transactions on Biomedical Circuits and Systems*, 11(4):849–857, 2017.

[42] Salomon L. J., Alfirevic Z., Berghella V., Bilardo C., Hernandez-Andrade E., Johnsen S. L., Kalache K., Leung K. Y., Malinger G., Munoz H., Prefumo F., Toi A., and Lee W. Practice guidelines for performance of the routine mid-trimester fetal ultrasound scan. *Ultrasound in Obstetrics and Gynecology*, 37(1):116–126, 2011.

[43] Butt K., Lim K., Bly S., Cargill Y., Davies G., Denis N., Hazlitt G., Morin L., Ouellet A., and Salem S. Determination of gestational age by ultrasound. *Journal of Obstetrics and Gynaecology Canada*, 36(2):171–181, 2014.

[44] Hayes A. F. and Krippendorff K. Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1):77–89, 2007.

[45] Sarmandal P., Bailey S. M., and Grant J. M. A comparison of three methods of assessing interobserver variation applied to ultrasonic fetal measurement in the third trimester. *British Journal of Obstetrics and Gynaecology*, 96(11):1261–1265, 1989.

[46] Perni S. C., Chervenak F. A., Kalish R. B., Magherini-Rothe S., Predanic M., Streltzoff J., and Skupski D. W. Intraobserver and interobserver reproducibility of fetal biometry. *Ultrasound in*

Obstetrics and Gynecology, 24(6):654–658, 2004.

[47] Rijken M. J., Lee S. J., Boel M. E., Papageorghiou A. T., Visser G. H. A., Dwell S. L. M., Kennedy S. H., Singhasivanon P., White N. J., Nosten F., and McGready R. Obstetric ultrasound scanning by local health workers in a refugee camp on the thai–burmese border. *Ultrasound in Obstetrics and Gynecology*, 34(4):395–403, 2009.

[48] Lima J. C., Miyague A. H., Nastri C. O., and Martins W. P. Biometry and fetal weight estimation by two-dimensional and three-dimensional ultrasonography: an intraobserver and interobserver reliability and agreement study. *Ultrasound in Obstetrics and Gynecology*, 40(2):186–193, 2012.

[49] Chang T. C., Robson S. C., Spencer J. A. D., and Gallivan S. Ultrasonic fetal weight estimation: Analysis of inter-and intra-observer variability. *Journal of Clinical Ultrasound*, 21(8):515–519, 1993.

[50] Sarris I., Ioannou C., Chamberlain P., Ohuma E., Roseman F., Hoch L., Altman D. G., and Papageorghiou A. T. Intra-and interobserver variability in fetal ultrasound measurements. *Ultrasound in Obstetrics and Gynecology*, 39(3):266–273, 2012.

[51] Verburg B. O., Mulder G. H. P., Hofman A., Jaddoe V. W. V., Witteman J., and Steegers E. A. P. Intra-and interobserver reproducibility study of early fetal growth parameters. *Prenatal Diagnosis*, 28(4):323–331, 2008.

[52] Napolitano R., Donadono V., Ohuma E. O., Knight C. L., Wanyonyi S. Z., Kemp B., Norris T., and Papageorghiou A. T. Scientific basis for standardization of fetal head measurements by ultrasound: a reproducibility study. *Ultrasound in Obstetrics and Gynecology*, 48(1):80–85, 2016.

[53] Lu W. and Tan J. Detection of incomplete ellipse in images with strong noise by iterative randomized hough transform (irht). *Pattern Recognition*, 41:1268–1279, 2008.

[54] Satwika I. P., Habibie I., Ma'sum M. A., Febrian A., and Budianto E. Particle swarm optimation based 2-dimensional randomized hough transform for fetal head biometry detection and approximation in ultrasound imaging. In *IEEE International Conference on Advanced Computer Science and Information Systems*, pages 468–473, 2014.

[55] Carneiro G., Georgescu B., Good S., and Comaniciu D. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. *IEEE Transactions on Medical Imaging*, 27(9):1342–1355, 2008.

[56] Zalud I., Good S., Carneiro G., Georgescu B., Aoki K., Green L., Shahrestani F., and Okumura R. Fetal biometry: a comparison between experienced sonographers and automated measurements. *The Journal of Maternal-Fetal and Neonatal Medicine*, 22(1):43–50, 2009.

[57] Ni D., Yang Y., Li S., Qin J., Ouyang S., Wang T., and Heng P. A. Learning based automatic head detection and measurement from fetal ultrasound images via prior knowledge and imaging parameters. In *IEEE International Symposium on Biomedical Imaging*, pages 772–775, 2013.

[58] Jatmiko W., Habibie I., Ma'sum M. A., Rahmatullah R., and Satwika I. P. Automated telehealth system for fetal growth detection and approximation of ultrasound images. *International Journal on Smart Sensing and Intelligent Systems*, 8(1):697–719, 2015.

[59] Ponomarev G. V., Gelfand M. S., and Kazanov M. D. A multilevel thresholding combined with edge detection and shape-based recognition for segmentation of fetal ultrasound images.

Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images, ISBI, pages 17–19, 2012.

[60] Sun C. Automatic fetal head measurements from ultrasound images using circular shortest paths. *Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images, ISBI*, pages 13–15, 2012.

[61] Stebbing R. V. and McManigle J. E. A boundary fragment model for head segmentation in fetal ultrasound. *Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images, ISBI*, pages 9–11, 2012.

[62] Ciurte A., Bresson X., and Cuadra M. B. A semi-supervised patch-based approach for segmentation of fetal ultrasound imaging. *Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images, ISBI*, pages 5–7, 2012.

[63] Foi A., Maggioni M., Pepe A., and Tohka J. Head contour extraction from the fetal ultrasound images by difference of gaussians revolved along elliptical paths. *Proceedings of Challenge US: Biometric Measurements from Fetal Ultrasound Images, ISBI*, pages 1–3, 2012.

[64] Perez-Gonzalez J., Muńoz J. B., Porras M. R., Arámbula-Cosío F., and Medina-Bańuelos V. Automatic fetal head measurements from ultrasound images using optimal ellipse detection and texture maps. In *VI Latin American Congress on Biomedical Engineering*, volume 49, pages 329–332. Springer, 2015.

[65] Anto E. A., Amoah B., and Crimi A. Segmentation of ultrasound images of fetal anatomic structures using random forest for low-cost settings. In *IEEE International Conference on Engineering in Medicine and Biology Society*, pages 793–796, 2015.

[66] Zhang L., Ye X., Lambrou T., Duan W., Allinson N., and Dudley N. J. A supervised texton based approach for automatic segmentation and measurement of the fetal head and femur in 2d ultrasound images. *Physics in Medicine and Biology*, 61(3):1095, 2016.

[67] Lienhart R. and Maydt J. An extended set of Haar-like features for rapid object detection. In *International Conference on Image Processing*, volume 1, pages I–900 – I–903, 2002.

[68] Viola P. and Jones M. J. Robust real-time object detection. *International Journal of Computer Vision*, 57:137–154, 2004.

[69] Breiman L. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[70] Hough P. V. C. Method and means for recognizing complex patterns, December 18 1962. US Patent 3,069,654.

[71] Bellman R. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences of the United States of America*, 38(8):716–719, 1952.

[72] Fitzgibbon A., Pilu M., and Fisher R. B. Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:476–480, 1999.

[73] Huttenlocher D. P., Klanderman G. A., and Rucklidge W. J. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:850–863, 1993.

[74] Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Biol. Skr.*, 5:1–34, 1948.

[75] Rueda S., Fathima S., Knight C. L., Yaqub M., Papageorghiou A. T., Rahmatullah B., Foi A., Maggioni M., Pepe A., Tohka J., Stebbing R. V., McManigle J. E., Ciurte A., Bresson X., Cuadra M. B., Sun C., Ponomarev G. V., Gelfand M. S., Kazanov M. D., Wang C. W., Chen H. C., Peng C. W., Hung C. M., and Noble J. A. Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge. *IEEE Transactions on Medical Imaging*, 33(4):797–813, 2014.

[76] Carneiro G., Amat F., Georgescu B., Good S., and Comaniciu D. Semantic-based indexing of fetal anatomies from 3-D ultrasound data using global/semi-local context and sequential sampling. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[77] Cuingnet R., Somphone O., Mory B., Prevost R., Yaqub M., Napolitano R., Papageorghiou A., Roundhill D., Noble J. A., and Ardon R. Where is my baby? A fast fetal head auto-alignment in 3D-ultrasound. In *IEEE International Symposium on Biomedical Imaging*, pages 768–771, 2013.

[78] Sofka M., Zhang J., Good S., Zhou S. K., and Comaniciu D. Automatic detection and measurement of structures in fetal head ultrasound volumes using sequential estimation and integrated detection network (IDN). *IEEE Transactions on Medical Imaging*, 33(5):1054–1070, 2014.

[79] Yaqub M., Rueda S., Kopuri A., Melo P., Papageorghiou A. T., Sullivan P. B., McCormick K., and Noble J. A. Plane localization in 3-D fetal neurosonography for longitudinal analysis of the developing brain. *IEEE Journal of Biomedical and Health Informatics*, 20(4):1120–1128, 2016.

[80] Ryou H., Yaqub M., Cavallaro A., Roseman F., Papageorghiou A., and Noble J. A. Automated 3D ultrasound biometry planes extraction for first trimester fetal assessment. In *International Workshop on Machine Learning in Medical Imaging*, pages 196–204, 2016.

[81] Namburete A., Xie W., Yaqub M., Zisserman A., and Noble J. Fully-automated alignment of 3d fetal brain ultrasound to a canonical reference space using multi-task learning. *Medical Image Analysis*, 2018.

[82] Dudley N. J. and Chapman E. The importance of quality management in fetal measurement. *Ultrasound in Obstetrics and Gynecology*, 19(2):190–196, 2002.

[83] Schluter P. J., Pritchard G., and Gill M. A. Ultrasonic fetal size measurements in Brisbane, Australia. *Journal of Medical Imaging and Radiation Oncology*, 48(4):480–486, 2004.

[84] International Society of Ultrasound in Obstetrics & Gynecology Education Committee. Sonographic examination of the fetal central nervous system: guidelines for performing the'basic examination'and the'fetal neurosonogram'. *Ultrasound in Obstetrics and Gynecology*, 29(1):109, 2007.

[85] American Institute of Ultrasound in Medicine. AIUM practice guideline for the performance of obstetric ultrasound examinations. *Journal of Ultrasound in Medicine*, 29(1):157, 2010.

[86] Li J., Wangi Y., Lei B., Cheng J.-Z., Qin J., Wang T., Li S., and Ni D. Automatic fetal head circumference measurement in ultrasound using random forest and fast ellipse fitting. *IEEE Journal of Biomedical and Health Informatics*, 22(1):215–223, 2018.

[87] Rahmatullah B., Papageorghiou A., and Noble J. A. Automated selection of standardized planes from ultrasound volume. In *International Workshop on Machine Learning in Medical Imaging*, pages 35–42, 2011.

[88] Abuhamad A., Falkensammer P., Reichartseder F., and Zhao Y. Automated retrieval of standard diagnostic fetal cardiac ultrasound planes in the second trimester of pregnancy: a prospective evaluation of software. *Ultrasound in Obstetrics and Gynecology*, 31(1):30–36, 2008.

[89] Chykeyuk K., Yaqub M., and Noble J. A. Class-specific regression random forest for accurate extraction of standard planes from 3d echocardiography. In *International MICCAI Workshop on Medical Computer Vision*, pages 53–62, 2013.

[90] Nie S., Yu J., Chen P., Wang Y., and Zhang J. Q. Automatic detection of standard sagittal plane in the first trimester of pregnancy using 3-D ultrasound data. *Ultrasound in Medicine and Biology*, 43(1):286–300, 2017.

[91] Feng S., Zhou S., Good S., and Comaniciu D. Automatic fetal face detection from ultrasound volumes via learning 3D and 2D information. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2488–2495, 2009.

[92] Dall'Asta A., Schievano S., Bruse J. L., Paramasivam G., Kaihuranta C. T., Dunaway D., and Lees C. C. Quantitative analysis of fetal facial morphology using 3D ultrasound and statistical shape modeling: a feasibility study. *American Journal of Obstetrics and Gynecology*, 2017.

[93] Roy-Lacroix M. E., Moretti F., Ferraro Z. M., Brosseau L., Clancy J., and Fung-Kee-Fung K. A comparison of standard two-dimensional ultrasound to three-dimensional volume sonography for routine second-trimester fetal imaging. *Journal of Perinatology*, 37(4):380, 2017.

[94] Kumar A. M. C. and Shriram K. S. Automated scoring of fetal abdomen ultrasound scan-planes for biometry. In *IEEE International Symposium on Biomedical Imaging*, pages 862–865, 2015.

[95] Yu Z., Tan E., Ni D., Qin J., Chen S., Li S., Lei B., and Wang T. A deep convolutional neural network based framework for automatic fetal facial standard plane recognition. *IEEE Journal of Biomedical and Health Informatics*, 2017.

[96] Bridge C. P., Ioannou C., and Noble J. A. Automated annotation and quantitative description of ultrasound videos of the fetal heart. *Medical Image Analysis*, 36:147–161, 2017.

[97] Lei B., Tan E., Chen S., Zhuo L., Li S., Ni D., and Wang T. Automatic recognition of fetal facial standard plane in ultrasound image via Fisher vector. *PLoS One*, 10(5):e0121838, 2015.

[98] Chen H., Wu L., Dou Q., Qin J., Li S., Cheng J., Ni D., and Heng P. Ultrasound standard plane detection using a composite neural network framework. *IEEE Transactions on Cybernetics*, 47(6): 1576–1586, 2017.

[99] Baumgartner C., Kamnitsas K., Matthew J., Fletcher T., Smith S., Koch L., Kainz B., and Rueckert D. Sononet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound. *IEEE Transactions on Medical Imaging*, 36(11):2204–2215, 2017.

[100] Kwitt R., Vasconcelos N., Razzaque S., and Aylward S. Localizing target structures in ultrasound video–a phantom study. *Medical Image Analysis*, 17(7):712–722, 2013.

[101] Maraci M. A., Bridge C. P., Napolitano R., Papageorghiou A., and Noble J. A. A framework for analysis of linear ultrasound videos to detect fetal presentation and heartbeat. *Medical Image Analysis*, 37:22–36, 2017.

[102] van den Heuvel T. L. A., Petros H., Santini S., de Korte C. L., and van Ginneken B. A step towards measuring the fetal head circumference with the use of obstetric ultrasound in a low

resource setting. In *Medical Imaging*, volume 10139 of *Proceedings of the SPIE*, page 101390V, 2017.

[103] Simonyan K. and Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[104] He K., Zhang X., Ren S., and Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *International Conference on Computer Vision*, pages 1026–1034, 2015.

[105] The Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv preprint arXiv:1605.02688*, May 2016.

[106] Dieleman S., Schlüter J., Raffel C., Olson E., Sønderby S. K., Nouri D., , et al. Lasagne: First release., August 2015. URL `http://dx.doi.org/10.5281/zenodo.27878`.

[107] Li M., Zhang T., Chen Y., and Smola A. J. Efficient mini-batch training for stochastic optimization. In *SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 661–670. ACM, 2014.

[108] Kingma D. and Ba J. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[109] Srivastava N., Hinton G., Krizhevsky A., Sutskever I., and Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1): 1929–1958, 2014.

[110] Ronneberger O., Fischer P., and Brox T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241, 2015.

[111] Tukey J. *Exploratory data analysis*, volume 2. Reading, Mass., 1977.

[112] Hadlock F. P., Deter R. L., Harrist R. B., and Park S. K. Estimating fetal age: computer-assisted analysis of multiple fetal growth parameters. *Radiology*, 152(2):497–501, 1984.

[113] Santana D. S., Cecatti J. G., Surita F. G., Silveira C., Costa M. L., Souza J. P., Mazhar S. B., Jayaratne K., Qureshi Z., Sousa M. H., and Vogel J. P. Twin pregnancy and severe maternal outcomes: the world health organization multicountry survey on maternal and newborn health. *Obstetrics and Gynecology*, 127(4):631–641, 2016.

[114] Young B. C. and Wylie B. J. Effects of twin gestation on maternal morbidity. In *Seminars in Perinatology*, volume 36, pages 162–168. Elsevier, 2012.

[115] Conde-Agudelo A., Belizán J. M., and Lindmark G. Maternal morbidity and mortality associated with multiple gestations. *Obstetrics and Gynecology*, 95(6):899–904, 2000.

[116] Boamah E. A., Asante K. P., Ae-Ngibise K. A., Kinney P. L., Jack D. W., Manu G., Azindow I. T., Owusu-Agyeie S., and Wylie B. J. Gestational age assessment in the ghana randomized air pollution and health study (GRAPHS): ultrasound capacity building, fetal biometry protocol development, and ongoing quality control. *JMIR Research Protocols*, 3(4), 2014.

[117] van den Heuvel T. L. A., de Bruijn D., Moens-van de Moesdijk D., Beverdam A., van Ginneken B., and de Korte C. L. Comparison study of low-cost ultrasound devices for estimation of gestational age in resource-limited countries. *Ultrasound in Medicine and Biology*, 44(11):2250–2260, 2018.

[118] Rietberg C. C. T., Elferink-Stinkens P. M., Brand R., van Loon A. J., Hemel O. J. S. V., and Visser G. H. A. Term breech presentation in the netherlands from 1995 to 1999: mortality and morbidity in relation to the mode of delivery of 33,824 infants. *British Journal of Obstetrics and Gynaecology*, 110(6):604–609, 2003.

[119] Hannah M. E., Hannah W. J., Hewson S. A., Hodnett E. D., Saigal S., Willan A. R., and Term Breech Trial Collaborative Group. Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial. *Lancet*, 356(9239):1375–1383, 2000.

[120] Roman J., Bakos O., and Cnattingius S. Pregnancy outcomes by mode of delivery among term breech births: Swedish experience 1987–1993. *Obstetrics and Gynecology*, 92(6):945–950, 1998.

[121] Kunzel W. Recommendations of the FIGO committee on perinatal health on guidelines for the management of breech delivery. *International Journal of Gynecology and Obstetrics*, 44(3):297–300, 1994.

[122] Liu S., Liston R. M., Joseph K. S., Heaman M., Sauve R., and Kramer M. S. Maternal mortality and severe morbidity associated with low-risk planned cesarean delivery versus planned vaginal delivery at term. *Canadian Medical Association Journal*, 176(4):455–460, 2007.

[123] Nwosu E. C., Walkinshaw S., Chia P., Manasse P. R., and Atlay R. D. Undiagnosed breech. *British Journal of Obstetrics and Gynaecology*, 100(6):531–535, 1993.

[124] Kongnyuy E. J. and Broek N. V. D. The use of ultrasonography in obstetrics in developing countries. *Tropical Doctor*, 37(2):70–72, 2007.

[125] Prevost R., Salehi M., Sprung J., Ladikos A., Bauer R., and Wein W. Deep learning for sensorless 3D freehand ultrasound imaging. In *Medical Image Computing and Computer-Assisted Intervention*, pages 628–636, 2017.

[126] Haller M. I. and Khuri-Yakub B. T. A surface micromachined electrostatic ultrasonic air transducer. *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, 43(1):1–6, 1996.

[127] Moore G. E. Cramming more components onto integrated circuits. *Electronics*, pages 114–117, 1965.

[128] Filly R. A. Ultrasound: the stethoscope of the future, alas. *Radiology*, 167(2):400–400, 1988.

[129] Solomon S. D. and Saldana F. Point-of-care ultrasound in medical education–stop listening and look. *New England Journal of Medicine*, 370(12):1083, 2014.

[130] Abu-Zidan F. M. Point-of-care ultrasound in critically ill patients: Where do we stand? *Journal of Emergencies, Trauma, and Shock*, 5(1):70, 2012.

Dankwoord

Gedurende mijn promotie heb ik veel mensen leren kennen die inhoudelijk hebben bijgedragen aan het werk dat te lezen is in dit proefschrift en daarnaast hebben bijgedragen aan alle gezelligheid gedurende mijn promotie tijd. Ik had de eer om in twee onderzoeksgroepen te mogen werken, waardoor ik twee keer zoveel wekelijkse vergaderingen had, maar natuurlijk ook twee keer zoveel gezellige collega's. Ik heb een hele leuke en interessante promotietijd gehad en daarvoor wil ik een aantal mensen in het bijzonder bedanken.

Allereerst wil ik mijn promotoren **Bram van Ginneken** en **Chris de Korte** bedanken voor alle begeleiding en hulp die ik gedurende mijn promotie heb gekregen. Ik vond het een mooie uitdaging om met dit project beide onderzoeksgroepen te combineren. Het project sloot precies aan op wat ik interessant vond, namelijk het combineren van medische beeldvormende en beeldverwerkings technieken met als doel het ontwikkelen van een product. Ondanks dat mijn project relatief klein was, kon ik altijd rekenen op jullie hulp. Hierdoor heb ik veel kunnen leren en heb ik mijzelf zowel op professioneel en persoonlijk gebied kunnen ontwikkelen.

Daarnaast wil ik **Guido Geerts**, **Frank Vijn**, **Fleur Posthumus** en **Susanne van Gils** van Delft Imaging Systems bedanken. Zonder jullie had dit project niet bestaan en zonder jullie hulp en inzet had ik dit project niet zo goed kunnen uitvoeren. De reizen naar Newcaslte en Ethiopië waren hoogtepunten in mijn promotie die ik samen met jullie heb gemaakt. Ik vind het zeer bewonderingswaardig dat Delft Imaging Systems in dit onderzoek investeert om zwangere vrouwen in ontwikkelingslanden te helpen.

I would like to thank **Jeff Neasham** and **Dave Graham** for the development of the SESAS device. This device was the reason that this project started in the first place. I really enjoyed working with you and the possibility to discuss the technical details which resulted in the article described in Chapter 2. It was a great pleasure to work with you and I also enjoyed our visits to the pubs for some food and drinks during my visits to NewCastle.

Veronique Wolbers, jij was degene die de allereerste beelden van een foetus heeft gemaakt met de SESAS. De bereidwilligheid en hulp die je hebt geboden om het prototype te kunnen testen en de inhoudelijke feedback die je met jouw klinische ervaring hebt gegeven waren van groot belang aan het begin van dit project.

Dagmar de Bruijn, heel erg bedankt voor alle keren dat ik heb mogen meelopen op de afdeling prenatale diagnostiek, alle kennis die je met mij hebt gedeeld en alle vragen die je hebt beantwoord hebben ervoor gezorgd dat mijn kennis over dit vakgebied erg is verbeterd. Ik vond onze samenwerking was heel prettig en gezellig. Natuurlijk ook dank voor het testen van de SESAS en alle scans en annotaties die gemaakt zijn voor de vergelijkingsstudie die staat omschreven in Hoofdstuk 3. Voor deze laatste studie wil ik **Desirée Moens** en **Anette Beverdam** ook bedankten voor hun hulp bij het verzamelen van de data. Jullie waren altijd vriendelijk en dachten inhoudelijk mee vanuit een klinisch perspectief. Dit maakt dat ik altijd met veel plezier richting de afdeling prenatale diagnostiek ging.

Daarnaast wil ik alle andere mensen bedanken die het mogelijk hebben gemaakt dat ik de SESAS heb kunnen gebruiken in het Radboudumc: **Tijs Rensen** en **Stef Mientki** voor alle hulp vanuit de instrumentele dienst, **Roland Loeffen** en **Frank de Lange** voor alle hulp bij het maken van het IMDD. I would like to extend my gratitude to **Rik Vos, Martin Verweij** and **Shreyas Ragunathan** who helped with the measurements of the acoustic output power of the SESAS.

A special thanks to **Hezkiel Petros**, **Stefano Santini** and **Yonas Desta** who made it possible that the data from 500 pregnant women was acquired for the studies described in Chapter 5 and Chapter 6. This would not have been possible without your help. This data made the work described in this thesis much more important and increased its impact. Next to this, I would like to thank you for your hospitality during my visits to the St. Luke's Hospital in Wolisso. I learned a lot during my visits and thank you for taking time out of your busy schedule for me.

Wessel Eijkman en **Frank van Doren** van MANGO consult. Dankzij jullie heb ik veel geleerd van een hele interessante reis naar Ethiopië. Bedankt voor alle hulp en ondersteuning die jullie hebben gegeven om de data acquisitie mogelijk te maken. Daarnaast natuurlijk ook bedankt voor alle leuke en nuttige informatie over lokale gebruiken. Dit maakte mijn reis naar Ethiopië een stuk interessanter en leuker.

De hulp van **Nathalie Popken** vanuit CORDAID aan het begin van mijn promotie traject en de hulp van **Simone Soeters** tijdens mijn tweede reis naar Ethiopië. Door de reorganisatie hebben we elkaar niet heel vaak gesproken tijdens dit project, maar desondanks wil ik jullie hartelijk bedanken voor al jullie hulp.

I would like to thank **Jan Moltz**, **Michael Schwier** and **Henning Kost** from Fraunhofer Mevis. You performed the feasibility study that showed it was possible to automatically measure fetal biometrics from 2D ultrasound images. This feasibility study was included in the funding proposal and gave me a good starting point for the development of the algorithms described in this thesis.

Alle collega's van DIAG waarmee ik deze jaren een kamer heb mogen delen en een hele gezellige tijd mee heb gehad. Mijn kamergenoten (kantoor 32&23!), **Midas, Ajay, Freerk, Mohsen, Babak, Sil, Bart, Cristina, Suzan** en **Kevin** waarmee ik altijd kon sparren over de inhoud van mijn onderzoek, maar misschien nog wel belangrijker alle gezellige borrels, activiteiten en persoonlijke verhalen die tot een mooie vriendschap heeft geleidt die hopelijk ook na mijn promotie zal blijven bestaan. Ook dank aan het **RSE team** en dan speciale dank voor alle hulp van **Paul Gerke** die altijd klaar stond om mijn vragen vriendelijk te beantwoorden en voor alle support bij de deep learning experimenten. **James Meakin**, thank you for all your help with grand-challenge for both my project and the ISMI course. Alle andere collega's van DIAG wil ik natuurlijk heel erg bedanken voor de gezellige tijd in de groep, de borrels en etentjes, de leuke tijd bij conferenties, de DIAG-weekenden en de feestjes **Bram Platel**, **Rashindra**, **Clarisa**, **Henkjan**, **Nico**, **Geert**, **Francesco**, **Colin**, **Sven**, **Sjoerd**, **Jonas**, **Charlotte**, **Jan-Jurre**, **Rick**, **Albert**, **Leticia**, **Mark**, **Kaman**, **Arnoud**, **Gabriel**, **Anton**, **Ecem**, **Péter**, **Meyke**, **David**, **Wouter**, **John-Melle** en **Oscar**.

Collega's van MUSIC, ook al zat ik gedurende mijn promotie niet bij jullie op de kamer heb ik me wel altijd betrokken gevoeld bij de hechte club met alle leuke uitjes, gezellige (kerst-)borrels en feestjes. Speciale dank voor **Gert Weijers** en **Han Thijsen** die me hebben geholpen met de QA4US voor de SESAS. Daarnaast natuurlijk ook alle andere MUSICI die me inhoudelijk hebben geholpen bij alle zaken die te maken hebben met de beeldacquisitie en de gezellige tijd. Hierdoor voelde ik me helemaal thuis in de groep **Gijs**, **Kaj**, **Anne**, **Stein**, **Jan**, **Rik**, **Maartje**, **Khalid**, **Sonja**, **Sandra**, **Jeroen**, **Shreya**, **Chuan**, **Anton**, **Leon** en **Roel**.

Mijn familie, vrienden en in het bijzonder **Harrie**, **Nicole**, **Tim**, **Karlijn**, **Franske**, **Maarten**, **Erik**, **Mariette**, **Robert** en **Iboya**. Jullie hadden altijd een luisterend oor en lieten mij (soms met wat veel woorden) enthousiast over mijn onderzoeksproject vertellen. Ook dachten jullie met mij mee en gaven jullie handige tips en suggesties. Maar boven al kon ik altijd bij jullie terecht en hebben jullie mij veel steun en vertrouwen gegeven tijdens mijn promotie.

Tot slot wil ik **Eline** natuurlijk heel erg bedanken voor je oneindige support, je luisterend oor en alle hulp en liefde die je me de afgelopen jaren hebt gegeven. Inhoudelijk kon ik altijd rekenen op je suggesties en ideeën die vaak net uit een andere invalshoek kwamen waardoor ik met een frisse blik naar mijn onderzoek kon kijken. Maar waarschijnlijk nog veel belangrijker was dat jij ervoor zorgde dat we samen heel veel leuke dingen ondernamen die ervoor zorgden dat ik niet bezig was met mijn promotie, waaronder alle mooie reizen die we in de afgelopen jaren hebben gemaakt. Zonder jou was mijn promotie tijd nooit zo leuk en gezellig geweest.

Curriculum Vitae

Thomas Leon Adrian van den Heuvel was born in Roermond, the Netherlands, on May 15th, 1990. He received his BSc. and MSc. degree in biomedical engineering from Eindhoven University of Technology, Eindhoven, the Netherlands, in 2012 and 2014, respectively. In March 2015, he started working as a Ph.D. student at the Diagnostic Image Analysis Group and the Medical Ultrasound Imaging Center at the Radboud university medical center in Nijmegen, the Netherlands. His PhD project focuses on the improvement of antenatal care in resource-limited settings by development of a low-cost ultrasound device and computer aided detection software to automatically detect maternal mortality risk factors, supervised by Bram van Ginneken and Chris de Korte. The results of these works are described in this thesis.

Institute for Health Sciences
Radboudumc