

Towards Point-Of-Care Ultrasound (POCUS) Estimation of Fetal Gestational Age from the Trans-Cerebellar Diameter using CNN-based Ultrasound Image Analysis

Mohammad Ali Maraci^a, Mohammad Yaqub^a, Rachel Craik^{b,c}, Sridevi Beriwal^b, Alice Self^b, Peter von Dadelszen^c, Aris Papageorgiou^b, J. Alison Noble^{a,*}

^a Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK

^b Nuffield Department of Women's and Reproductive Health, University of Oxford, UK

^c Department of Women and Children's Health, King's College London, UK

Abstract. Obstetric ultrasound is a fundamental ingredient of modern prenatal care with many applications including accurate dating of a pregnancy, identifying pregnancy-related complications, and diagnosis of fetal abnormalities. However, despite its many benefits, two factors currently prevent wide-scale uptake of this technology for point-of-care clinical decision-making in low- and middle-income country (LMIC) settings. First, there is a steep learning curve for scan proficiency, and second, there has been a lack of easy-to-use, affordable and portable ultrasound devices. In this article, we introduce a framework towards addressing these barriers, enabled by recent advances in machine learning applied to medical imaging. **The framework is designed to be realizable as a point-of-care ultrasound (POCUS) solution** with an affordable wireless ultrasound probe, a smartphone or tablet, and automated machine-learning based image processing. Specifically, we propose a machine learning-based algorithm pipeline designed to automatically estimate the gestational age of a fetus from a short fetal ultrasound scan. We present proof-of-concept evaluation of accuracy of the key image analysis algorithms for automatic head transcerebellar (TC) plane detection, automatic transcerebellar diameter (TCD) measurement, and estimation of gestational age on conventional ultrasound data simulating the POCUS task and discuss next steps towards translation via a first application on clinical ultrasound video from a low-cost ultrasound probe.

Keywords: point-of-care ultrasound (POCUS), prenatal health, machine learning, global health, gestational age.

*Corresponding Author, E-mail: alison.noble@eng.ox.ac.uk

1. Introduction

According to the World Health Organization (WHO), 99% of life-ending maternal complications in pregnancy occur in low- and middle-income countries (LMICs) [1]. Further, optimizing maternal health in pregnancy is designated to be one of the global challenges in 21st Century medicine as diagnosing pregnancy-related conditions can be complex and is compounded by infrastructural constraints imposed by inadequate accessibility, insufficient expertise, and requirement of capital equipment. Ultrasound is the recommended imaging modality for monitoring fetal and maternal health during pregnancy, as it is highly portable and economic,

facilitates real-time data acquisition and visualization, and also benefits from an acceptable safety profile for patients and sonographers alike [2]. Recently, the use of medical ultrasound in limited-resource settings has increased [3]–[6]; in part attributable to the increased portability and affordability of this technology.

Demand for point-of-care ultrasound (POCUS) solutions in first response scenarios and remote locations, the emergency department and at the patient’s bedside has driven recent advances in ultrasound technology portability and affordability. Easily transportable and portable hand-held ultrasound systems can turn either a consumer-grade smartphone or tablet into a portable ultrasound system, thus providing an affordable POCUS solution. Empowering non-experts or minimally-trained healthcare professionals to use such systems is not straightforward as the quality of ultrasound image acquisition is dependent on user skill. Therefore, an important emerging research area is automated image analysis to support non-expert users in performing ultrasound-based diagnostic decisions. This paper contributes to this literature.

Specifically, in this paper we present an image analysis framework suitable for processing ultrasound video in real time for automatic detection and measurement of the trans-cerebellar diameter in the fetal brain, which enables the estimation of fetal gestational age. Fetal gestational age is a key pregnancy biomarker that underlies much of clinical decision-making and is often very inaccurate in LMICs.

1.1. Gestational Age (GA) Estimation

Accurate estimation of gestational age (GA) based on sonographic fetal parameters forms an integral part of modern prenatal care. In addition to defining the expected delivery date, it is also essential for screening of anatomic and growth abnormalities and disturbances, and allows for appropriate scheduling of antenatal care. At a population level, reliable estimates of GA are important to calculate rates of preterm birth and small-for-gestational age (SGA) at birth.

Currently, estimates of preterm birth and SGA rates are mere approximations of true rates, particularly in geographical regions at greatest risk of these conditions [7], [8].

Traditionally, GA is estimated using the first day of the last menstrual period, which assumes that ovulation occurs on day 14 day of a regular 28-day menstrual cycle. Irregular menses, unknown or uncertain dates, oral contraceptive use or recent pregnancy or breastfeeding may all influence the accuracy of this method [9], [10]. In high-income countries (HICs), in such a case, ultrasound measurement of the fetal crown rump length in the first trimester is recommended [11] as this is said to be more accurate than dating in late pregnancy, where growth disturbances result in underestimation of GA for an abnormally-small fetus and an overestimation of GA for a macrosomic fetus. In the second and third trimesters, GA is typically estimated from fetal biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC) or femur length (FL). As the transcerebellar diameter (TCD) is least likely to be affected by fetal growth disturbances, it has been shown to be a single fetal biometric parameter that allows accurate prediction of GA in both singleton and twin gestations [12]. Therefore, the TCD was the parameter chosen for automated GA estimation in our work. To our knowledge, this is the first paper to present an automated approach to TCD-based GA estimation from ultrasound images and video.

1.2. Ultrasound in LMICs

A considerable challenge in LMICs is that women frequently present to a healthcare professional late in pregnancy (later than 20 weeks) and often do not, with confidence, know how many weeks pregnant they are. Therefore, accurate ultrasound-based estimation of GA is essential for implementation of evidence-based interventions such as administration of corticosteroids for fetal lung maturation or induction of labor due to postmaturity.

Obstetrical ultrasound is acceptable to pregnant women and health care workers in many LMIC settings, although, in some settings, it is acknowledged that there is resistance to its use as ultrasound imaging can be used for fetal sex determination [13]. Obstetrical ultrasound training programs reduce measurement variability [14] and improve image quality [15]. Observational studies have reported that measurement reproducibility of trained healthcare workers in limited-resource settings is comparable to that of qualified sonographers from HICs for certain basic diagnostic tasks [16]. However, other work reports that reliable ultrasound-based measurement requires significant pre- and in-service training and is difficult to scale up [13]. In response to this, DiStigter et al [6] proposed a care model by which minimally-trained primary care workers followed a simple scanning protocol for acquiring freehand ultrasound sweeps (cine loops) over the abdomen (called an obstetric sweep protocol (OSP) [17]) and data were transferred to a center where experts reviewed the scans. The OSP essentially allows a healthcare professional to implement the ISUOG basic obstetric ultrasound checklist (Table 1 in [18]). A recent comparison of three low-cost ultrasound devices for manual estimation of GA using both a standard plane and the OSP concluded that manual analysis using low-cost devices and the OSP gave good GA estimates from measurements of HC and AC, but not FL [17]; that paper was the first to report on GA estimation accuracy from ultrasound video sweeps that might be acquired by a minimally-trained healthcare professional rather than assuming that the sonographer has the skill to find the correct biometry plane. However, the selection of the OSP-derived biometry plane and manual GA estimation were performed by an experienced sonographer.

The high prevalence of growth restriction in LMICs makes established methods of late GA assessment, such as measurement of fetal HC, inaccurate. A proposed alternative has been measurement of the fetal TCD. The fetal TCD is a good measure for GA assessment over a wider range of gestational age and is relatively protected from the influence of fetal growth

restriction in the absence of associated anomalies [12]. Thus, a TCD-based GA estimation device could be very useful in LMICs.

1.3. Automated Fetal Ultrasound Image Analysis

We restrict the discussion to relevant literature in 2D ultrasound image analysis as 3D ultrasound is currently too expensive to be considered for GA estimation in LMICs. For a discussion of the latter and a method for automatic 3D ultrasound-based GA estimation we refer the reader to [19].

There is a growing academic literature on automated fetal ultrasound image analysis from 2D ultrasound and video, for instance [20]–[31]. Much of this recent literature is focused on automated estimation of biometry [20], and automatic recognition of standard imaging planes for the second-trimester anomaly scan from ultrasound video frames [21]–[26]. Other research has considered automatic ultrasound video partitioning and automatic video frame description to label key anatomical structures in the fetal heart [27]. Early methods were based on frame-by-frame analysis but recently a number of papers have demonstrated that superior plane recognition accuracy can be achieved with spatio-temporal models built with Recurrent Neural Network (RNN) and long short-term memory (LSTM) based architectures [25], [26]. These make use of frame-to-frame coherence of local structures within the models to overcome ambiguity that can occur in single frame analysis. In all of the previously cited papers proposed methods were motivated by the clinical need to support second-trimester obstetrics screening and fetal growth monitoring in HICs.

There is limited literature on automated fetal ultrasound image analysis aimed at clinical needs in LMIC settings [28]–[31]. Maraci et al [28] was the first paper to consider automation of aspects of the ISUOG basic obstetrical ultrasound scan [18]. Specifically, [28] presents an approach to detect fetal presentation, that the heart is beating (fetal viability), and to estimate

biometry from a single linear ultrasound sweep over the maternal abdomen. Gao et al have further developed that work to propose spatio-temporal models for linear sweep partitioning where the most notable improvement was for fetal heart detection [29], [30]. More recently, van der Heuvel et al [31] proposed deep learning-based methods for automated estimation of fetal presentation, twin detection and HC-based GA estimation from the OSP. While this is the most closely related work to our own, there are key differences: van der Heuvel et al and others [28]–[31] used machine learning-based ultrasound video partitioning for object detection. As the OSP is a whole abdomen scan, the fetal head position with respect to the sweep is unknown. Hence, any given image frame is not guaranteed to be a standardized transventricular (TV) plane to provide the correct biometry plane for HC-based GA estimation. Therefore, in [31] HC-based GA estimation involved empirical combination of HC estimates from a number of frames. It is difficult to judge from that paper how to generalize that strategy to other measurements.

To our knowledge, there is no previous publication on automated TCD-based GA estimation implemented within an automated machine learning-based framework. In the proposed approach we assume (1) a different ultrasound sweep protocol than [31]; and (2) that a minimally-trained healthcare professional can find the fetal head and acquire a short video sweep of the head by moving the ultrasound probe from a starting position of approximately (but not exactly) at the TV position. Then the automated image analysis problem is to find the standardized TC plane and then to estimate the TCD from this cine loop.

2. Methods

2.1. Experimental Data

For the proof-of-principle purposes of this paper we used two exemplar datasets.

Dataset A: Our underlying assumption is that the healthcare professional can find the head and approximately locate the TV plane. They then move the probe around slowly (small rotations and translations) while acquiring a short cine loop that includes the TC plane. The image analysis challenge is to find the TC plane when many candidates are actually TV planes. The difference between the two imaging planes is illustrated schematically in Fig. 1.

At the time of this research we did not have access to a large real-world video dataset so we simulated this case by using a dataset of 5000 TC and TV images. This dataset was divided into a 3000 image training dataset consisting of 1500 images of the TC plane as positive examples and 1500 images of the TV plane as negative examples. Subsequently, the validation set consisted of 1000 unseen data samples with 500 TC (positive) and 500 TV (negative) images. This dataset was used to evaluate the plane finder.

Dataset B: Image dataset B was used to develop and evaluate the automatic TCD measurement algorithm. 3736 retrospective TC images were obtained from a second-trimester fetal ultrasound screening database. 500 images were randomly selected for the training set and the TC structure manually annotated, and the remaining 3236 images were used in testing. The TCD had been manually measured by a sonographer for each image during the scanning session. The manually-measured TCD value was written on the image. Optical character recognition (OCR) was used to read the manually-measured TCD (mm) from the image. This provided the TCD measurement ground truth. According to the manual TCD measurements the GA range for these images covered 16-26 weeks gestational age.

These images were in DICOM format and hence pixel size in millimeters was available. This meant that we could directly convert the automatic TCD from pixels to millimeters.

FIGURE 1 INSERT HERE

2.2. Acquisition and Automated Analysis Pipeline

A schematic summarizing the acquisition and automated analysis pipeline principle is shown in Fig. 2. We briefly explain the automated analysis components in this subsection. The user (sonographer) would acquire a short video sweep of the fetal head using an ultrasound probe. Such an acquisition is easy to acquire with minimal training. In Fig. 2 Engine A extracts TC frames from the video and Engine B localizes the cerebellum and measures the TCD to enable estimation of the GA. Importantly, the user does not need to identify the TC plane or make a manual TCD measurement.

2.2.1. Automated Image Frame Detection (Engine A)

There are various ways to perform automatic fetal ultrasound standard plane detection reported in the literature including methods based on random forests [22], and convolutional neural networks (CNNs) [21, 23, 25, 26, 33]. For fetal TC plane detection accuracies of the order of 90% and higher have been reported in a number of papers [22], [23]. [In this work we implemented a standard CNN, based on a variation of the well-known AlexNet architecture \[34\] that poses TC plane detection as a binary classification problem.](#) A schematic of the CNN architecture is shown in Fig. 3. The ultrasound image is replicated three times as input. The CNN architecture consists of four convolutional layers followed by a fully-connected layer. Subsequently a softmax layer is used to produce the class probabilities. Network weights are initialized with values trained on the ImageNet dataset [35]. Using a pre-trained network for initialization has been shown to significantly speed up training and improve classification accuracy of a network relative to choosing random values in other medical imaging applications [36] as well as fetal ultrasound image plane classification [37].

FIGURE 2 INSERT HERE

FIGURE 3 INSERT HERE

2.2.2. Automated Image Analysis Module (Engine B)

To perform gestational age estimation based on the TCD, an accurate diameter must be automatically measured. We formulate this as an image segmentation problem and train a fully convolutional neural network (FCN) [38] to segment the cerebellum, hereafter called the TC structure, from which the TCD can be estimated.

An FCN is a variant of a CNN where the last fully-connected layer(s) are typically substituted by one or more convolution layers. In addition, transposed convolution layers are added to up-sample the output of the convolutional layers and hence the output of a FCN is a segmentation mask which is the same size as the original image. In such an architecture, the entire network can be trained end-to-end. Details of the FCN architecture design and parameters employed are given in Fig. 4. [A softmax loss function was used.](#) The output of the pool3 layer is summed with the 2x upsampled prediction layer, before a transposed convolution layer with stride 8 is applied on the resulting feature maps to obtain the segmentation map. In the experiments reported in Sect. 3.2 we compare FCN architectures with different input image sizes and with and without transfer learning initialization.

The training set used to derive the FCN segmentation model consisted of 500 images from Dataset B. This is a relatively small dataset for training a deep learning model. Data augmentation was employed to boost segmentation performance. This included random translation by 25% of image dimensions, random zooming by up to 20%, random rotations between -10 to 10 degrees, and changes of image intensity to mimic random brightness and contrast variations.

A typical input image, manually segmented TC structure (ground truth), and automatically segmented TC structure mask are shown in Fig. 5.

The TCD can be automatically measured from the automatic segmentation mask. This is achieved by finding the major axes of the segmented cerebellum structure assuming it is elliptical in shape. Although this is an approximation, experimentally this assumption turns out to lead to good TCD estimates compared to manual measurement (see the results in Sect. 3). Given the size of a pixel in mm, the location of the two pixels at the extremes of the major axis are transformed from pixel space to physical space. The Euclidean distance is then computed between these two points to estimate the TCD in mm.

Gestational age can be calculated from the TCD estimate in millimeters using established parametric equations. In this work, we used the equation from the Intergrowth-21st study [39]:

$$GA = 4.777048 + (0.724321 * TCD) \quad (1)$$

Here GA is the gestational age in weeks and TCD is the transcerebellar diameter in millimeters.

FIGURE 4 INSERT HERE

FIGURE 5 INSERT HERE

3. Experimental Evaluation

Experiments were designed to evaluate the accuracy of: (1) TC plane detection; (2) TC structure segmentation for different FCN variants, and (3) automatic versus manual TCD-based GA estimation. We also present a first result of applying the developed image analysis methods to low-cost ultrasound probe video.

3.1. TC Plane Detection

The TC plane classification algorithm was trained for 1000 iterations with 128 images in each batch, on a training dataset of 3000 images using an NVIDIA TitanXp GPU. The training dataset consisted of 1500 images of the TC plane as positive examples and 1500 images of the TV plane as negative examples. The resulting model had around 6.4 million parameters. Subsequently, the validation set consisted of 1000 unseen data samples with 500 positive and 500 negative images. On the validation set, the algorithm precision and recall for TC plane detection were both 0.99, and the average accuracy 99%. Figure 6 shows the first 500 iterations of the training loss and of the validation accuracy.

FIGURE 6 INSERT HERE

3.2. Trans-Cerebellum Segmentation/Detection

The accuracy of the TC structure segmentation algorithm was evaluated on, and compared with, images of varying size and FCN architectures with and without transfer learning to see if this made a difference. Where employed, transfer learning was used to initialize the first two layers of the network prior to training. It is important to note that the cerebellum is a small structure in the fetal brain. This translates to a high imbalance between the pixels that correspond to the class of interest (the TC structure) versus the background pixels (Fig. 5). Therefore, considering solely the pixel segmentation accuracy would not be a suitable metric. Instead, we utilize the metrics introduced in [34], namely the mean accuracy and the mean intersection-over-union, that takes class imbalance into account. In essence, the intersection-over-union metrics are preferred since they control for the total number of pixels for a class. Referring to Table 1, Overall Accuracy measures the correctly predicted pixels of the foreground class, class c (the cerebellum), divided by the total number of pixels. Mean

Accuracy measures the correctly predicted pixels of class c divided by the total number of pixels of that class. Mean Intersection-Over-Union (IOU) is the IOU averaged over all classes.

Table 1 presents TC structure segmentation results for input images of varied size where the frames are down-sampled from 128x192 pixels to 96x128 pixels to 64x96 pixels. In addition, in each case the effect of transfer learning (using a pre-trained network) is shown.

Table 1. Cerebellum (Trans-cerebellum structure) segmentation accuracy. The input to the network is Height x Width x 3. Overall Accuracy measures the correctly predicted pixels of class c (the cerebellum) divided by the total number of pixels, Mean Accuracy measures the correctly predicted pixels of class c divided by the total number of pixels of that class, and Mean Intersection Over Union is the Intersection Over Union averaged over all classes. TF: TransFer learning.

Frame Size (Height x Width)	Overall Accuracy (%)	Mean Accuracy (%)	Mean Intersection Over Union (%)
(64x96x3)NoTF	96.78	79.59	73.25
(64x96x3)WithTF	96.95	80.86	77.57
(96x128x3)NoTF	97.38	83.62	77.66
(96x128x3)WithTF	97.52	85.07	78.93
(128x192x3)NoTF	97.82	86.30	81.00
(128x192x3)WithTF	97.89	86.99	81.62

As can be seen from Table 1, the highest mean accuracy is achieved for an input size of 128x192x3 pixels. Furthermore, as the size of the input image decreases to 96x128, and 64x96, the mean accuracy achieved is also decreased to 85.07% and 80.86% respectively. In each case, and for both metrics, transfer learning had a small positive impact in boosting the accuracy.

The best segmentation task model had around 38 million parameters.

3.3. TCD Measurement and GA Estimation

Automatically-derived TCD estimates were compared with manual measurements (the ground truth) for dataset B (n=3236 images).

Segmentation failure associated with low image quality will lead to an invalid measurement. Based on visual inspection, a measurement difference threshold of 5mm was empirically selected to capture this source of error. Using this value the algorithm failed in 35 cases only (1% of the total cases). For the full dataset, the mean of the manual measurements was 20.6mm

with a standard deviation (stdev) of 1.2mm, compared with the mean of the automated measurement of 20.3mm (stdev 2.9mm) (p-value <0.01). When excluding the 35 failure cases, the mean manual measurements was 20.6mm (stdev 1.2mm) while the mean of the automatic measurement was 20.3mm (stdev 1.6mm) (p-value <0.01), suggesting automated measurement accuracy is very good and that the algorithm is stable on the majority of cases.

Equation 1 was used to convert the TCD estimates in millimeters to GA in weeks. This allowed any error to be expressed in terms of gestational age units (weeks). The mean manual GA estimation was found to be 19.7 weeks (stdev 0.9) compared with automated estimation of 19.5 weeks (stdev 2.1) (p-value <0.1) and, when excluding the 35 failure cases, 19.5 weeks (stdev 1.2) (p-value <0.1). Figure 7 shows how manual and automated GA estimates are correlated (n=3201 estimates excluding the 35 failure cases). Figure 8 shows a Bland-Altman plot to visualize the agreement between the manual and automated GA estimation methods. Again, we conclude that there is a good agreement between the two estimates.

3.4. Application of Methods to a Video from a Low-Cost Ultrasound Probe

In this subsection we report a first result of applying the machine-learning based GA estimation methods to a single ultrasound video from a low-cost ultrasound probe. Note that the algorithms were applied “as is” with no attempt to refine the deep learning models to the local properties of the low-cost ultrasound video.

For this experiment we used the Konted Gen1 C10R (3.5MHz convex) wireless ultrasound probe for ultrasound video acquisition (Konted, Beijing, China). A single ultrasound video was acquired following the acquisition protocol advocated for use with the machine-learned methods, namely the sonographer found the TV plane and then slowly moved the probe from the TV to include the TC over approximately 8 seconds. Retrospectively, an experienced sonographer viewed the video frame-by-frame and located the frame they considered to be the

best TC plane based on clinical guidelines for TC standard plane definition [40]. A sonographer manually measured the TCD in this frame. For comparison, we also had access to the clinical hospital scan TCD measurement on the same subject made using a high-end ultrasound machine (GE Voluson E8).

The TC plane detection algorithm reported in this paper was applied to all video frames. The TC structure segmentation and TCD estimation algorithms reported in this paper were applied to the frame identified by the sonographer as the best TC plane.

The TC plane detection algorithm found the correct TC plane with a high class probability, but TC plane detection had a high false positive rate overall. TCD automated measurement (26.2mm) was an underestimation compared with manual measurement (34.9mm) and the hospital measurement (36.2mm). The underestimation was found to be due to under-segmentation of the TC structure, see Fig. 9.

FIGURE 7 INSERT HERE

FIGURE 8 INSERT HERE

FIGURE 9 INSERT HERE

4. Discussion

We have presented [a pipeline of the key image analysis tasks](#) suitable for a novel point-of-care low-cost ultrasound-based fetal gestational-age solution. The solution design is intended to be (1) usable by a minimally-trained, LMIC healthcare professional using a simple scanning protocol and (2) suitable for a wide range of GA by using the TCD as the basis for GA estimation.

Although the primary objective is to provide a point-of-care GA estimate the proposed GA estimation device is designed to enable the user to separate data acquisition from the automated

analysis stage, if desired. This could be helpful if either multiple scans are taken or multiple subjects are scanned, as analysis can be carried out at a later time.

The deep learning architectures were chosen as they are readily available CNN architectures for proof-of-concept demonstration of the complete pipeline for this novel application. However, they are large CNN models (millions of parameters), and are not well suited for deployment on smartphones or tablets with modest memory. More recent state-of-the-art deep learning architectures that result in compact models, such as MobileNet [41] and ShuffleNet [42], will be investigated in future work but the trade off between accuracy and model size will need to be understood.

We have proposed a simple realization of the tool although thus far the analysis components have only been thoroughly tested separately as a proof of principle due to the current lack of availability of large volumes of low-cost ultrasound probe data. However, in Section 3.3 we reported on the application of the developed models to one of the early ultrasound videos we have acquired using a low-cost probe. These results are encouraging and show that our proposed automated methodology works in principle on images and video from this low-cost probe but, perhaps not unexpectedly, the machine learning models will require fine tuning on video from this domain to achieve clinically accepted results. For the reported result, we chose a video for which the TC was visible to the human eye in the video. The image appearance of the low-cost probe ultrasound video was different from the higher end ultrasound probe, but the sonographer was confident that they could see the TC structure in the video. In general, the visibility of the TC may be lower than with a standard clinical probe, but research needs to be conducted to either confirm or refute this possible limitation.

More work is likely to be required on standard plane verification if the tool is to be used by healthcare professionals who do not have the skill to find an accurate standard plane. Attention or saliency maps may be useful here as suggested in [30], [24].

The automated TCD measurement algorithm has been well-validated on a large data set and achieved very good accuracy compared with manual measurement on the same images. We believe this to be the first time automated TCD estimation has been demonstrated at scale.

A natural question to ask is whether the accuracy is good enough for clinical use. The clinically acceptable age difference (error) depends on clinical use of a GA estimation so there is no single answer to the question of what is clinically acceptable. Indeed, the improvement of GA estimation is an intense field of clinical research.

Therefore, it seems reasonable to say that an “acceptable” difference in GA estimation is what is acceptable compared to current standards – clinical measurement of the fetus. According to [43] these are +/- 4-7 days at around 12 weeks; 10-12 days at around 20 weeks; +/- 14 days at 28 weeks and +/- 18-21 days at 34 weeks. Our results, which are on conventional ultrasound images, are within these bounds for the second trimester but it is yet to be proven whether the same accuracy can be achieved using a low-cost ultrasound probe, and across a larger gestational age range. The contributing factors to accuracy in our particular case will be firstly the image plane finding (sonographer skill), secondly structure appearance (fetal and maternal factors), and thirdly the accuracy of automated segmentation. The relative importance of these contributions will be investigated in future work when we have a working low-cost ultrasound-based device in the field.

Current work is integrating the components together into a single android-based device suitable for a minimally-trained user, as well as there are plans to conduct a usability study of the next generation solution in the field which will be reported on in a future publication.

5. Conclusion

The main contribution of this paper is to describe and implement key components of a computational framework for estimating the fetal gestational age from an ultrasound video

sweep of the fetal brain as the basis of a point-of-care ultrasound-based GA estimation solution for LMICs. Future work will implement the fully integrated device and conduct translational research usability studies within the context of the sub-Saharan African PRECISE Network (www.precisenetwork.org).

Acknowledgments

This research was funded by the University of Oxford EPSRC Impact Acceleration Award (EP/K503769/1), and a Bill & Melinda Gates Foundation Grand Challenges Exploration award (project 1128591). MAM acknowledges the RCUK Centre for Doctoral Training in Healthcare Innovation for a doctoral research studentship (EP/G036861/1). JAN acknowledges EPSRC grants EP/M013774/1, EP/L505316/1, EP/R013853/1 and ERC Advanced Grant 694581 (PULSE). Aris T. Papageorgiou is supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The low cost ultrasound video was acquired as part of the MRC GCRF Programme PRECISE Network grant (MR/P027938/1).

Statement on Use of Data from Human Subjects

The authors confirm that the appropriate institutional approval was received to use the clinical images used in the research described in this paper.

Conflict of Interest Statement

The authors declare no conflicts of interest with the research described in the paper.

References

- [1] World Health Organization, Fact sheets: Maternal mortality, 16 February 2018.
<https://www.who.int/news-room/fact-sheets/detail/maternal-mortality>.

- [2] R. Hershkovitz, E. Sheiner, and M. Mazor, Ultrasound in obstetrics: a review of safety, *Eur. J. Obstet. Gynecol. Reprod. Biol.*, vol. 101, no. 1, pp. 15–18, 2002.
- [3] R. Anderson *et al.*, The midwife’s assistant: designing integrated learning tools to scaffold ultrasound practice, in *Proceedings of the Fifth International Conference on Information and Communication Technologies and Development*, 2012, pp. 200–210.
- [4] W. Brunette *et al.*, Portable antenatal ultrasound platform for village midwives, in *Proceedings of the First ACM Symposium on Computing for Development*, 2010, p. 23:1–23:10.
- [5] 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, *SPIE Medical Imaging 2017*, vol. 10139, no., p. 101390V–10139–6.
- [6] K. K. DeStigter *et al.*, Low-cost teleradiology for rural ultrasound, in *2011 IEEE Global Humanitarian Technology Conference*, 2011, pp. 290–295.
- [7] H. Blencowe *et al.*, National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications, *Lancet*, vol. 379, no. 9832, pp. 2162–2172, 2012.
- [8] M. de Onis, M. Blössner, and J. Villar, Level and patterns of intrauterine growth retardation in developing countries, *Eur. J. Clin. Nutr.*, vol. 52 Suppl 1. 1998.
- [9] S. Campbell, S. L. Warsof, D. Little, and D. J. Cooper, Routine ultrasound screening for the prediction of gestational age., *Obstet. Gynecol.*, vol. 65, no. 5, pp. 613–620, May 1985.
- [10] T. H. Nguyen, T. Larsen, G. Engholm, and H. Møller, Evaluation of ultrasound-estimated date of delivery in 17 450 spontaneous singleton births: do we need to modify Naegele’s rule?, *Ultrasound Obstet. Gynecol.*, vol. 14, no. 1, pp. 23–28, Jul. 1999.
- [11] R. Napolitano *et al.*, Pregnancy dating by fetal crown-rump length: a systematic review of charts., *BJOG*, vol. 121, no. 5, pp. 556–565, Apr. 2014.
- [12] M. R. Chavez, C. V Ananth, J. C. Smulian, and A. M. Vintzileos, Fetal transcerebellar diameter measurement for prediction of Gestational Age at the extremes of fetal growth, *J. Ultrasound Med.*, vol. 26, no. 9, pp. 1167–1171, Sep. 2007.
- [13] E. T. Kim, K. Singh, A. Moran, D. Armbruster, and N. Kozuki, Obstetric ultrasound use in

- low and middle income countries: a narrative review, *Reprod. Health*, vol. 15, no. 1, p. 129, 2018.
- [14] I. Sarris *et al.*, Standardization of fetal ultrasound biometry measurements: improving the quality and consistency of measurements, *Ultrasound Obstet. Gynecol.*, vol. 38, no. 6, pp. 681–687, Dec. 2011.
 - [15] S. Z. Wanyonyi, R. Napolitano, E. O. Ohuma, L. J. Salomon, and A. T. Papageorghiou, Image-scoring system for crown–rump length measurement, *Ultrasound Obstet. Gynecol.*, vol. 44, no. 6, pp. 649–654, Dec. 2014.
 - [16] M. J. Rijken *et al.*, Obstetric ultrasound scanning by local health workers in a refugee camp on the Thai-Burmese border, *Ultrasound Obstet. Gynecol.*, vol. 34, no. 4, pp. 395–403, Oct. 2009.
 - [17] 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 Med. Biol.*, vol. 44, no. 11, pp. 2250–2260, 2018.
 - [18] ISUOG Education Committee recommendations for basic training in obstetric and gynecological ultrasound, *Ultrasound Obstet. Gynecol.*, vol. 43, no. 1, pp. 113–116, Jan. 2014.
 - [19] A. I. L. Namburete, R. V. Stebbing, B. Kemp, M. Yaqub, A. T. Papageorghiou, and J. A. Noble, Learning-based prediction of gestational age from ultrasound images of the fetal brain, *Med. Image Anal.*, vol. 21, no. 1, 2015.
 - [20] S. Rueda *et al.*, Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: A grand challenge, *IEEE Trans. Med. Imaging*, vol. 33, no. 4, 2014.
 - [21] H. Chen *et al.*, Standard plane localization in fetal ultrasound via domain transferred deep neural networks., *IEEE J. Biomed. Heal. informatics*, vol. 19, no. 5, pp. 1627–1636, Sep. 2015.
 - [22] M. Yaqub, B. Kelly, A. T. Papageorghiou, and J. A. Noble, *Guided Random Forests for identification of key fetal anatomy and image categorization in ultrasound scans*, *Medical Image Computing and Computer-Assisted Interventions -- MICCAI 2015*, pp 687-694, 2015.
 - [23] C. F. Baumgartner *et al.*, SonoNet: real-time detection and localisation of fetal standard scan

- planes in freehand ultrasound, *IEEE Trans. Med. Imaging*, vol. 36, no. 11, pp. 2204–2215, 2017.
- [24] J. Schlemper *et al.*, Attention-gated networks for improving ultrasound scan plane detection, *CoRR*, vol. abs/1804.05338, 2018.
 - [25] H. Chen *et al.*, Automatic fetal ultrasound standard plane detection using knowledge transferred Recurrent Neural Networks, *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*, pp. 507–514.
 - [26] H. Chen *et al.*, Ultrasound standard plane detection using a composite neural network framework, *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1576–1586, 2017.
 - [27] C. P. Bridge, C. Ioannou, and J. A. Noble, Automated annotation and quantitative description of ultrasound videos of the fetal heart, *Med. Image Anal.*, vol. 36, pp. 147–161, 2017.
 - [28] M. A. Maraci, C. P. Bridge, R. Napolitano, A. Papageorgiou, and J. A. Noble, A framework for analysis of linear ultrasound videos to detect fetal presentation and heartbeat, *Med. Image Anal.*, vol. 37, pp. 22–36, 2017.
 - [29] Y. Gao, M. A. Maraci, and J. A. Noble, Describing ultrasound video content using deep convolutional neural networks, in *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 787–790.
 - [30] Y. Gao and J. A. Noble, Detection and characterization of the fetal heartbeat in free-hand ultrasound sweeps with weakly-supervised two-streams convolutional networks, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017*, pp. 305–313.
 - [31] 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, Imaging for Patient-customized Simulations and Systems for Point-of-Care Ultrasound, *Proceedings of BIVPS 2017 and POCUS 2017*, pp. 105–112.
 - [32] H. Chen, Y. Zheng, J.-H. Park, P.-A. Heng, and S. K. Zhou, Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pp. 487–495.
 - [33] M. Yaqub, B. Kelly, A. T. Papageorgiou, and J. A. Noble, A deep learning solution for

- automatic fetal neurosonographic diagnostic plane verification using clinical standard constraints, *Ultrasound Med. Biol.*, vol 43, number 12, pp. 2925-2933, 2017.
- [34] A. Krizhevsky, I. Sutskever, G.A. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems 25*, pp. 1097-1105, 2012.
- [35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, *International Journal of Computer Vision (IJCV)*, 2015, 115(3):211-252, doi 10.1007/s11263-015-0816-y
- [36] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, et. al, Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning, *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285-1298, 2016.
- [37] Y. Gao, M.A.M. Maraci, J.A. Noble, Describing ultrasound video content using deep convolutional neural networks, *IEEE International Symposium on Biomedical Imaging*, pp. 787-790, 2016.
- [38] J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, *CoRR*, vol. abs/1411.4038, 2014.
- [39] V. Donadono, R. A. S. Napolitano et. al, Development of normal fetal brain structures: standards based on ultrasound measurements from the fetal growth longitudinal study of the Intergrowth-21st project, <https://doi.org/10.1002/uog.17595>, (First published: 16 September 2017)
- [40] Sonographic examination of the fetal central nervous system: guidelines for performing the "basic examination" and the "fetal neurosonogram", *Ultrasound Obstet. Gynecol.*, vol. 29, pp. 109-116, 2007.
- [41] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, MobileNets: efficient convolutional neural networks for mobile vision applications, [arXiv:1704.04861](https://arxiv.org/abs/1704.04861), 2017.
- [42] X. Zhang, X. Zhou, M. Lin, J. Sun, ShuffleNet: an extremely efficient convolutional neural

network for mobile devices, arXiv:1707.01083, 2017.

- [43] A.T. Papageorgiou, B. Kemp, W. Stones, E.O. Ohuma, S.H. Kennedy, M. Purwar, L.J. Salomon, D.G. Altman, J.A. Noble, E. Bertino, M.G. Gravett, R. Pang, L. Cheikh Ismail, F.C. Barros, A. Lambert, Y.A. Jaffer, C.G. Victora, Z.A. Bhutta, J. Villar; International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st). Ultrasound-based gestational-age estimation in late pregnancy. *Ultrasound Obstet Gynecol.* 2016 Dec;48(6):719-726. doi: 10.1002/uog.15894.

Author Biographies

Mohammad Ali Maraci has an MSc in databases and web-based systems, and subsequently completed a PhD in Healthcare Innovation at the Institute of Biomedical Engineering (IBME), University of Oxford, UK, on the topic of ultrasound image analysis methods for resource-constrained environments. He maintained an active interest in affordable healthcare activities while a PhD student and subsequently held postdoctoral research fellow positions at the IBME considering translation of obstetric ultrasound images analysis methods into LMIC settings.

Mohammad Yaqub has a PhD on machine-learning based fetal ultrasound image analysis from the IBME, University of Oxford, UK, and then worked as a postdoctoral research fellow at the IBME on cardiac and fetal ultrasound image analysis. He left in 2016 to work at Intelligent Ultrasound Ltd where he is currently a Vice President of Engineering. He has maintained an academic interest in teaching and currently holds an academic visitor appointment at Oxford University.

Rachel Craik has a first degree in Biology from York University and holds research coordinator posts at King's College London and Oxford University. She coordinated the

Oxford-led INTERBIO-21st Study, a global initiative investigating the growth and development of infants from conception to the age of 2 years in high risk settings, and currently co-ordinates the KCL-led PRECISE study in sub-Saharan Africa and a number of Oxford-led global ultrasound clinical and translational projects including this one.

Sridevi Beriwal, is a clinical research fellow in the Nuffield Department of Women's and Reproductive Health at the University of Oxford. She works on the PRECISE global health ultrasound project with IBME engineers and clinical teams in the Gambia, Kenya and Mozambique.

Alice Self, is a clinical research fellow in the Nuffield Department of Women's and Reproductive Health at the University of Oxford with a special interest in global health. She primarily works on the CALOPUS project; an inter-disciplinary collaboration between NDWRH, IBME and THSTI, New Delhi, aiming to develop more affordable, point-of-care, automated ultrasound technology for use in LMICs and is an Oxford clinical team member on the PRECISE study.

Peter von Dadelszen, an obstetrician (maternal-fetal medicine), is Professor of Global Women's Health, King's College London. He trained, is certified, and has practiced in New Zealand, Canada and the UK. With Professor Laura Magee, he leads and collaborates on research into pregnancy hypertension, fetal growth and stillbirth, from basic science to clinical epidemiology and health services research to global health, primarily through the Gates Foundation-funded PRE-EMPT and UKRI-funded PRECISE initiatives.

Aris Papageorgiou is a Clinician Scientist in Fetal Medicine based in Oxford and London. His research is in the areas of maternal, fetal and perinatal health using diverse methods - basic science, epidemiology, trials and implementation science. This has included large global projects that aim to improve phenotyping and assess fetal / infant growth and development. Improving pregnancy care in low resource settings with AI-assisted ultrasound imaging is a major research line.

J. Alison Noble is the Oxford University Technikos Professor of Biomedical Engineering. Her research group is best known for its inter-disciplinary research at the interfaces of ultrasound imaging, computational (machine-learning based) image analysis, and clinical medicine. She has published widely on the topic, and holds a European Research Council Advanced Grant on next generation ultrasound systems (PULSE). Professor Noble is a former President of the MICCAI Society and a Fellow of the UK Royal Society.

List of Figure and Table Captions

Figure 1. Schematic of (a) the transventricular (TV) standard imaging plane; and (b) the transcerebellar (TC) standard imaging plane.

Figure 2. Analysis framework schematic. A short ultrasound video is acquired using a low-cost ultrasound device. Next, transcerebellar (TC) plane frames are extracted from the ultrasound video in Engine A. Engine B then localizes the TC structure and performs automated trans-cerebellar diameter (TCD) estimation from which the gestational age (GA) can be estimated via an equation.

Figure 3. Transcerebellar (TC) plane classification method schematic. An ultrasound frame fed to the convolutional neural network (CNN) architecture is classified as a TC plane or otherwise.

Figure 4. Schematic representation of the transcerebellar (TC) structure segmentation FCN-based network architecture.

Figure 5. (a) A fetal ultrasound image where the transcerebellar (TC) structure is visible; (b) the ground truth segmentation label for the TC structure overlaid on the image; (c) the automated TC structure segmentation result.

Figure 6. TC plane classification training loss and validation accuracy for the first 500 iterations.

Figure 7. Manual versus automatic gestational age estimates in weeks of gestational age (n=3201).

Figure 8. Bland-Altman plot to show the agreement between the manual and automatic gestational age estimation (n=3201).

Figure 9. A low-cost ultrasound probe TC plane image (a) original image, and (b) image with automated segmentation overlaid – the TC structure is under-segmented.

Table 1. TC structure segmentation task accuracy. The input image to the network has dimensions height x width x 3 pixels. Overall Accuracy measures the correctly predicted pixels of class c divided by the total number of pixels, Mean Accuracy measures the correctly predicted pixels of class c (cerebellum) divided by the total number of pixels of that class, and Mean Intersection Over Union is the Intersection Over Union averaged over all classes. TF: TransFer learning.