Machine Learning to Assess the Fetal Brain from Ultrasound Images

Ana Ineyda Luisa Namburete

Somerville College
University of Oxford

Submitted in partial fulfilment of the degree of Doctor of Philosophy

Hilary 2015

Abstract

Obstetric care decisions fundamentally rely upon accurate estimation of gestational age (GA). Ultrasound- (US) based measurements provide reliable estimates of GA, if performed early in pregnancy. However, in low-income settings, the lack of appropriately trained sonographers and the tendency for women to present for care late in pregnancy are barriers to the use of US for dating purposes. In this thesis, we propose to exploit sonographic image patterns associated with dynamic fetal brain development to predict GA. We designed an algorithm which automatically estimates GA from an US scan collected from a single visit, thereby enabling clinically useful estimates of GA to be made even in the third trimester of pregnancy: a period complicated by biological variation and unreliable size-based estimates.

The presented model was conceived on the basis that fetal brain development follows a precise spatiotemporal pattern, with folds emerging and disappearing on the surface of the brain (cerebral cortex) at fixed time points during pregnancy. This timing is so precise that post-mortem neuroanatomical and MRI evidence suggest that the 'developmental maturation' of the fetal brain may be a better predictor for GA than traditional size-based estimates.

We capitalize on these age-related patterns to develop, for the first time, a unified model which combines sonographic image features and clinical measurements to predict GA and brain maturation. The framework benefits from a manifold surface representation of the fetal head which delineates the inner skull boundary and serves as a common coordinate system based on cranial position. This allows for fast and efficient sampling of anatomically-corresponding brain regions to achieve like-for-like structural comparison of different developmental stages. Bespoke features capture neurosonographic patterns in 3D images, and using a regression forest classifier, we characterize structural brain development both spatially and temporally to capture the natural variation existing in a healthy population ($n = 448$) over an age range of active brain maturation (18 to 34 weeks).

Our GA prediction results on a high-risk clinical dataset ($n = 187$) strongly correlate with true GA ($r = 0.98$, accurate within ± 6.10 days), confirming the

link between maturational progression and neurosonographic activity observable across gestation. Our model also outperforms current clinical methods, particularly in the third trimester. Through feature selection, the model successfully identified regional biomarkers of neurodevelopmental progression over gestation. Guided by these regions, we present a novel approach for defining and testing hypotheses associated with neuropathological deviations.

Machine Learning to Assess the Fetal Brain from Ultrasound Images

Ana Ineyda Luisa Namburete

Somerville College

University of Oxford

Submitted in partial fulfilment of the degree of

Doctor of Philosophy

Hilary 2015

Acknowledgements

Personal

First and foremost, my sincerest gratitude goes to my supervisor, Prof. Alison Noble, for patiently guiding me throughout this project from its earliest (uncertain) days until its current state. You have been a tremendous mentor, always ready to promote my work and to provide me with opportunities to grow as an academic researcher. My gratitude also goes to my collaborators, Prof. Stephen Kennedy and Dr. Bryn Kemp, for adding immense clinical value to my work.

I would also like to thank my colleagues at the Biomedical Image Analysis Laboratory, in particular Richard Stebbing, Sylvia Rueda, Mohammad Yaqub, Gordon Stevenson, Ali Maraci, and Phill Brown. Thank you for taking the time to share ideas, expertise, and words of wisdom, and for making the DPhil experience so much more interesting and enjoyable. You have all inspired and supported me in different ways, and I am lucky to have crossed paths with you.

To my dearest friends, Valerie Babinsky and Marya Bazzi, a heartfelt thank you for being such a strong support network and for all the laughs. You have been my family away from home.

A massive thank you to my wonderful partner, Stephan Uphoff, for being my biggest ally and for always believing in me. My sanity would have been in jeopardy were it not for all the meals, proofreading sessions, practice talks, and for being open and patient to discuss ideas particularly in moments of uncertainty. I deeply cherish the love, support, and encouragement that you continue to provide.

Finally, I would like to thank my fantastic parents, Luisa and Salvador, and siblings, especially Nick and Pedro, for encouraging me at every step of this journey. I am incredibly lucky to be able to count on your love and support, and I cannot thank you enough for championing me through the hard times and celebrating my successes. Your patience and understanding is humbling, particularly when I have been too absorbed to call. I dedicate this work to you.

Funding

I am deeply grateful to the Commonwealth Scholarship Commission for making it possible for me to study and live at Oxford for the past four years. It has truly been an intellectually enriching experience.

Abstract

Obstetric care decisions fundamentally rely upon accurate estimation of gestational age (GA). Ultrasound- (US) based measurements provide reliable estimates of GA, if performed early in pregnancy. However, in low-income settings, the lack of appropriately trained sonographers and the tendency for women to present for care late in pregnancy are barriers to the use of US for dating purposes. In this thesis, we propose to exploit sonographic image patterns associated with dynamic fetal brain development to predict GA. We designed an algorithm which automatically estimates GA from an US scan collected from a single visit, thereby enabling clinically useful estimates of GA to be made even in the third trimester of pregnancy: a period complicated by biological variation and unreliable size-based estimates.

The presented model was conceived on the basis that fetal brain development follows a precise spatiotemporal pattern, with folds emerging and disappearing on the surface of the brain (cerebral cortex) at fixed time points during pregnancy. This timing is so precise that post-mortem neuroanatomical and MRI evidence suggest that the 'developmental maturation' of the fetal brain may be a better predictor for GA than traditional size-based estimates.

We capitalize on these age-related patterns to develop, for the first time, a unified model which combines sonographic image features and clinical measurements to predict GA and brain maturation. The framework benefits from a manifold surface representation of the fetal head which delineates the inner skull boundary and serves as a common coordinate system based on cranial position. This allows for fast and efficient sampling of anatomically-corresponding brain regions to achieve like-for-like structural comparison of different developmental stages. Bespoke features capture neurosonographic patterns in 3D images, and using a regression forest classifier, we characterize structural brain development both spatially and temporally to capture the natural variation existing in a healthy population ($n = 448$) over an age range of active brain maturation (18 to 34 weeks).

Our GA prediction results on a high-risk clinical dataset ($n = 187$) strongly correlate with true GA ($r = 0.98$, accurate within ± 6.10 days), confirming the link between maturational progression and neurosonographic activity observable across gestation. Our model also outperforms current clinical methods, particularly in the third trimester. Through feature selection, the model successfully identified regional biomarkers of neurodevelopmental progression over gestation. Guided by these regions, we present a novel approach for defining and testing hypotheses associated with neuropathological deviations.

Contents

List of Figures

List of Tables

Notation

Throughout this thesis, matrices are denoted by boldface uppercase letters ($\mathbf{X}$), vectors by boldface lowercase letters ($\mathbf{x}$), and vectors extracted from matrices are denoted by indexed boldface letters, such that $\mathbf{x}_i$ is the i-th column of matrix $\mathbf{X}$. The j-th element of vector $\mathbf{x}$ is denoted by x_j. Data sets are denoted by calligraphic notation ($\mathcal{A}$).

Data Notation

Data dimensionality	$D \in \{2, 3\}$, for 2D and 3D data, respectively
Image	$I : \mathbb{N}^D \rightarrow \mathbb{R}$
Pixel/Voxel position	$\mathbf{p} \in \mathbb{N}^D$
Image plane	$\mathbf{Y} \in \mathbb{R}^{3\times2}$, comprising a point and a normal vector
Surface Mesh	$\mathbf{X}$ for complete surface
Data unit (vector)	$\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$
Individual data point (scalar)	$v_i \in \mathbb{R}$
Output data label (continuous)	$\mathbf{a}$
Number of elements in a set	$N_{\mathcal{A}}$ or $\lvert\mathcal{A}\rvert$

Forest Notation

Forest	$\mathcal{F} = \{\mathcal{T}\}$
Tree level	d (d_{max} for maximum tree level)
Number of trees in forest	T
Index of tree within forest	t
Index of tree node within a tree	k
Training dataset (full)	$\mathcal{V}$
Set of data points reaching the k-th node	$\mathcal{V}_k$
Set of data points reaching the k-th children nodes	$\mathcal{V}_{k\in\{L,R\}}$
Number of features tested at each node	m
Number of available features from which to select m	n
Feature response at a given pixel/voxel	$\mathbf{f}(\mathbf{p})$
Feature selector function	$\boldsymbol{\phi} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, where $m \ll n$

1

Introduction

1.1 Clinical Motivation

Accurate gestational age (GA) estimation forms an integral part of obstetric prenatal care. It defines the delivery date, enables expectant parents to make informed decisions about their unborn child, and can influence the success or safety of a clinical intervention. However, in low-income settings, the lack of sophisticated imaging technology, such as magnetic resonance imaging (MRI), and appropriately trained ultrasonographers to analyse the images are barriers to the use of ultrasound (US) for the purposes of accurate fetal monitoring.

Pregnancy dating becomes particularly important in settings where pregnant women typically attend for obstetric care late in pregnancy (after the 24^{th} week of gestation), when menstrual history is unavailable or unreliable. In the absence of clinically-useful history, US measurements provide the most accurate estimation of GA [1]. However, in the third trimester of pregnancy, even US-based dating may produce estimation errors up to ± 3 weeks [2, 3]. Thus, in a setting where screening occurs in the second and third trimesters, the error margins yielded by current methods render them as not clinically useful, potentiating the need for alternative techniques for estimating GA from data collected at a single visit.

In this thesis, we present a machine learning-based tool which exploits the dynamic structural features within US images of the fetal brain to automatically estimate GA. Our tool was conceived on the basis that fetal brain development follows a precise spatial and temporal pattern, with folds emerging on the surface of the brain (cerebral cortex) at fixed time-points during pregnancy (Figure 1.1). Specifically, the fetal brain surface, or *cortex*, rapidly transitions from a relatively smooth surface in the early second trimester to progressively bearing more indentations or folds over the course of pregnancy until it resembles the adult brain at birth (Figure 1.1). This timing is so precise that post-mortem neuroanatomical and MRI evidence suggest that the 'developmental maturation' of the fetal brain is more closely related to GA than currently used fetal size metrics. Deviations from this pattern have been indicative of cortical malformations, and depending on severity, may result in adverse outcomes ranging from developmental delays and retardation to infant mortality [4], which would benefit from early detection during pregnancy. This, in turn, is suggestive of a direct link between healthy brain development and chronological age.

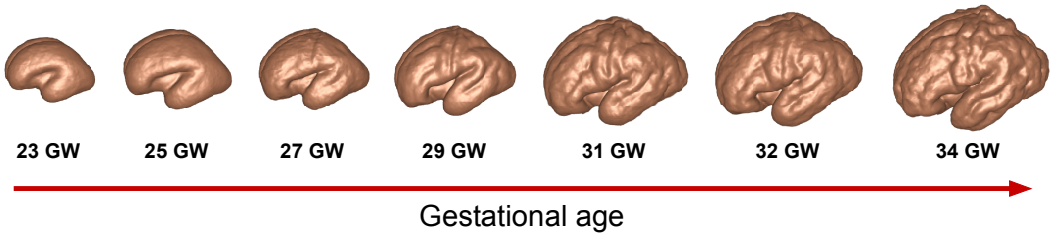


Figure 1.1: **Structural development of the fetal brain.** Indentations appear on the cortical surface at specific time-points during gestation.

Whilst it is possible in specialist centres to quantify fetal brain development subjectively or from MRI data, doing so on a large scale in routine clinical practice is impractical and expensive, and is currently unavailable to expectant mothers in resource-poor settings. An attractive solution is to automate the process of tracking fetal brain development using images already captured during routine US examinations, yet this is a challenging task and has never before been attempted.

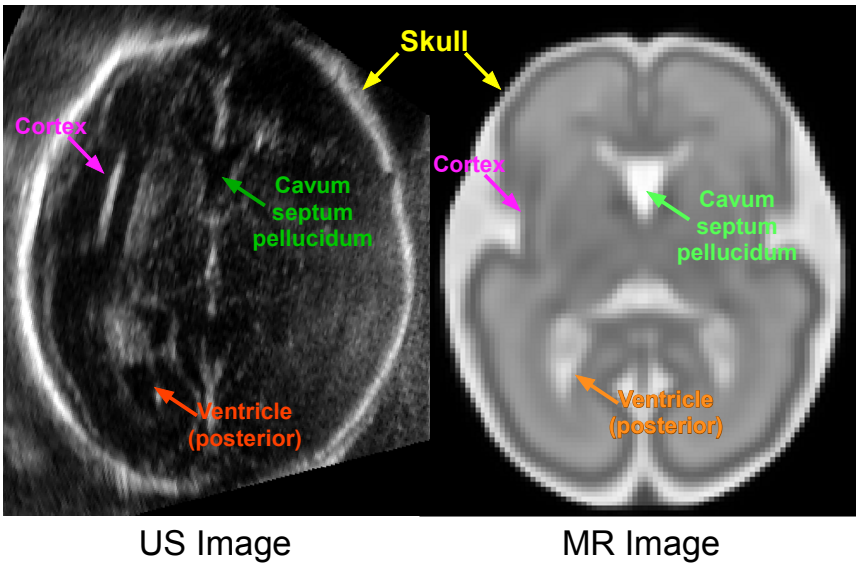


Figure 1.2: **Structures observable in US and MR images.** Despite differences in image intensity, the structures visible in MR images are also present in US images of the fetal brain.

Most prior work on quantifying brain development has used MR images due to their ability to represent the brain in its entirety (e.g. [5, 6, 7]). Al-

though largely overlooked in clinical practice, the process of structural brain development is also observable non-invasively in US images of the fetal head [8], and indeed there is correspondence between the structures visible in MR and US images of the fetal brain, as shown in Figure 1.2. The fact that these image features may be visualised sonographically opens the possibility of using this clinically-available imaging modality to study the brain, and develop appropriate tools which can be translated into clinical practice. This thesis combines image analysis and machine learning to develop a mapping between sonographic image appearance and *in utero* neuroanatomical development to predict GA from a single US scan.

1.2 Contributions: Towards a Predictive Model of Healthy Brain Maturation

The main contribution of this thesis is a framework comprised of a series of sophisticated algorithms to analyse the fetal brain from US images, and to extend the clinical utility of sonographic assessment of fetal maturation, particularly in the third trimester. Figure 1.3 presents a schematic overview of the proposed framework.

Given that the goal of the project is to learn the pattern of morphological changes that the brain undergoes during the fetal period, it is important to first establish a common coordinate frame between images to allow like-for-like anatomical comparisons at different GAs. A typical approach for neuroimage preprocessing would involve skull stripping followed by image registration, in order to align the images into a common image domain. While there has been extensive work in the development of image registration techniques for application to MR images of the brain, these are not directly applicable to US images

due to the unique challenges posed by fetal neurosonography. Specifically, it is a technical challenge to differentiate the fetal brain from maternal soft tissue due to subtle variations in tissue density which results in low contrast resolution. Variations in scanning protocols also affect signal intensities, ultimately complicating the ability to achieve an accurate anatomical alignment between patients.

To overcome these barriers, we describe in Chapter 3 a semi-automated tool for fitting a continuous, parametric surface to the cranial boundaries in the US image. The surface parametrization allows for fast and efficient sampling of anatomically-corresponding brain regions to achieve like-for-like structural comparison of different developmental stages— voiding the need for segmentation or registration of intracranial structures. The user provides an initial alignment of the surface, and the latter is then deformed to adhere to the inner cranial boundary, providing a clear separation between the fetal brain tissue and extracerebral tissues (e.g. bone structures, the placenta, maternal soft tissue, etc). Throughout the automated deformation process, the topological relations between the surface vertices are preserved. This property is fundamental in the creation of a manifold representation of the fetal skull on a predefined coordinate system, ultimately allowing consistent sampling of anatomical regions.

Having successfully established a coordinate system for imaged brain regions, we are able to use the surface domain to sample vertex locations associated with specific anatomies to extract image information. In Chapter 4, we present an automated machine learning-based predictive model to learn the pattern of normal fetal brain maturation. This feature-based model characterizes neuroanatomical appearance both spatially and temporally, capturing the natural variation existing in a healthy fetal population over a period of active brain maturation: 18^{+0} to 33^{+6} gestational week (GW, weeks$^{+\text{ days}}$). Using be-

spoke features designed to capture sonographic activity in the brain, the model exploits the US signal responses and automatically links them to the age-related emergence/disappearance of structures in the brain. Validation experiments on standard clinical US images demonstrate that the model is stable in producing accurate predictions of GA, and that it outperforms current clinical practice.

During the decision-making process, the model selects the most age-discriminating anatomical features from the images. An analysis of these regions is presented in Chapter 5, as we parse the features to discover the structural regions of greatest significance to brain development during the fetal period. We found that through feature selection, the model successfully identified and exploited key regions which are known to undergo the most notable variations in the fetal period. This result clearly indicates that US images contain important information about underlying biological processes, which are not currently used in traditional US-based methods for tracking fetal development.

Finally, in Chapter 6 the model was applied to two datasets of images collected from challenging clinical fetal cohorts to demonstrate potential clinical utility. On these data, our age prediction results strongly correlated with true GA and were accurate within ± 6.1 days, confirming the link between maturational progression and neurosonographic activity observable across gestation. The model also outperformed current clinical methods by ± 4.57 days in the third trimester— a period complicated by biological variations in the fetal population. We then applied the algorithm to select (brain) features in images from healthy fetuses and compared to those selected in images from growth-restricted fetuses. We found that the model does indeed extract healthy neurodevelopmental trajectories, and that it is capable of detecting pathological deviations associated with an adverse intrauterine environment.

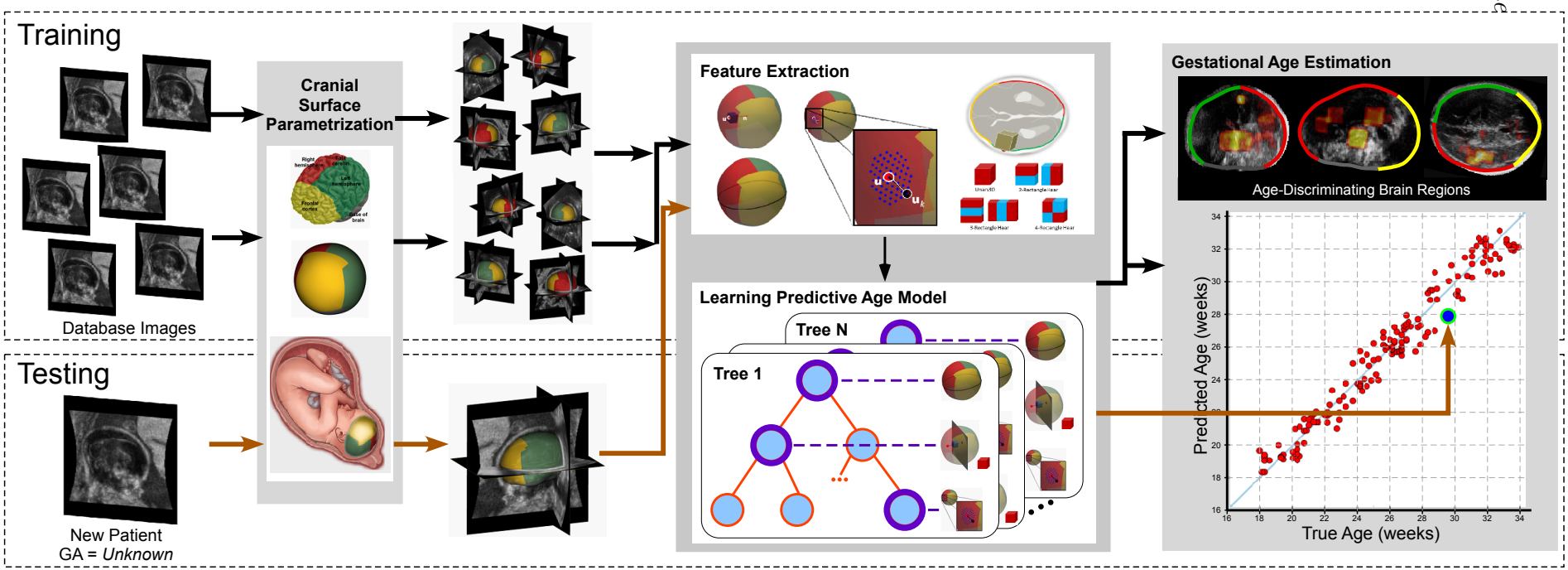


Figure 1.3: **Schematic summarizing the GA prediction framework presented in this thesis.** Given a set of images in the database, the machine learning model is trained to learn age-specific US image appearance and model a mapping between sonographic activity and GA. When given a new test image taken from a fetus of unknown GA, the model can be applied to estimate age and possibly determine clinical outcomes.

1.3 Thesis Structure

The first chapter describes the clinical motivation for developing a model to automatically predict GA from neurodevelopmental information observable in US images, and outlines the structure of the thesis to follow. In Chapter 2, we review the current clinical practice in prenatal obstetric care in tracking fetal growth. We describe the process of brain maturation in early life, as reported in the post-mortem neuroanatomical and MRI literature. Given that this thesis is the first to attempt the use of sonographic image patterns to predict GA and neurodevelopmental maturation, that chapter is, itself, a contribution as it contextualises the work and discusses the contributions and limitations of established quantitative methods for image-based assessment of *in utero* brain development. We also review the pertinent literature on using machine learning to make predictions from medical images, particularly in the context of age estimation.

The original results from this work are presented in Chapters 3, 4, 5, and 6, all of which follow a set structure. At the start of each chapter, an **introduction** section summarizes the pertinent literature and contextualises the challenge to be addressed. This is followed by a description of the **methods** used to analyse the data, and a detailed description of the technical contributions of each chapter to a level that would make experimental reproducibility possible. The experimental outcomes are described in the **results** section, and each chapter ends with a **conclusion** which discusses insights and potential future work.

Chapter 3 introduces a continuous B-spline surface which parametrizes the fetal skull and provides the spatial context for anatomical structures within the brain. The work presented in that chapter is central to the methods, experiments, and analyses discussed in Chapters 4 and 5. It defines the parametric

surface which serves as a surrogate representation of the fetal skull, the basis upon which anatomical regions are sampled, so it forms the initial step in the pipeline of US-based neuroimage analysis presented in this thesis. Thus, validation of the accuracy of surface-based alignment was performed on US images of the fetal head, to verify its ability to extract standard clinical planes from a 3D US volume, and its agreement with manually-extracted planes.

Chapter 4 describes, in detail, the features used in the regression forest model for GA prediction, and through validation experiments, we demonstrate the ability of the model to accurately predict GA from 3D US images of the fetal head. We also provide insights into the functionality and behaviour of the age-predictive model, and examine the effect of feature inclusion on prediction accuracy. In Chapter 5, we parse the trained model to examine the anatomical/image features it has automatically identified as age-discriminating. Through this analysis, we found that the model discovered important neurodevelopmental landmarks, which are associated with biological processes characteristically occurring during the fetal period. This chapter discusses the automated extraction of age/development-related information, in a way that is not currently possible using standard clinical methods.

Each chapter contains experiments which evaluate the performance of a contribution in answering clinical questions. Thus, the analysis of all results was conducted in concert with established clinical studies and, where available, validated against measurements accepted in current clinical practice.

The methods presented in Chapters 4 and 5 were then applied to serial (longitudinal) scans acquired from a cohort of patients classified as small-for-gestational age (SGA), and compared to a group of matched controls. The results, presented in Chapter 6, show that the age-predictive model is capable of characterizing patterns of neurodevelopmental delay associated with intrauter-

ine growth restriction. Finally, Chapter 7 consolidates the work presented in this thesis and discusses the potential role of the contributions of this thesis in monitoring fetal development at the bedside.

1.4 Originality

I declare that I am the sole author of this thesis document, and I produced all the tables and figures included herein. Critique and comments were provided by Prof. J. Alison Noble. Unless otherwise specified, all image preprocessing and manipulation, machine learning algorithm implementation, and data manipulation was my original, individual work. The images were acquired by clinical collaborators as part of the INTERGROWTH-21st and INTERBIO-21st studies.

Chapter 3 Part of this chapter was joint work with Richard V. Stebbing. Specifically, we jointly designed the surface model, Stebbing developed the surface deformation framework, and I manually initialized the cranial surfaces into each image. Manual selection of standard clinical planes was performed by a clinical collaborator: Dr. Raffaele Napolitano. Volumetric segmentation of anatomical structures in the fetal head images was manually performed by myself with assistance from collaborators: Dr. Mohammad Yaqub, Dr. Raffaele Napolitano, Dr. Sylvia Rueda, and Sana Fathima. I independently designed the validation experiments and implemented the MATLAB scripts to analyse the cranial deformation results.

Chapter 4 The labelling of image training data with true (known) GAs was performed by clinical collaborators. Advice on clinical definitions was provided by Dr. Bryn Kemp. Independently, I designed the features

and created a custom implementation of the regression forest algorithm in C++, and optimized it for speed using multi-threading and runtime computation of features. I also made use of the ITK and VTK libraries for importing, exporting, and visualizing image data. Independently, I designed the bespoke features to be extracted from the brain regions, and incorporated them in the regression forest implementation.

Chapter 5 Independently, I created custom software in C++ to parse through the trained regression forests and populate the probability feature maps which indicate the regions pertaining to fetal neurodevelopmental maturation. I also analysed the relevant image features extracted from the algorithm, with respect to GA and the model's decision-making protocol.

Chapter 6 With the help of collaborators, I selected the fetal cohorts which may be of interest to clinical studies. Dr. Bryn Kemp advised on the standard case definitions accepted in clinical practice for the different conditions. I independently generated comparisons between features of the different fetal cohorts (healthy and small-for-gestational-age) to assess regional group differences.

1.5 List of Publications

Journal Articles

Namburete, A. I. L., Stebbing, R. V., Kemp, B., Yaqub, M., Papageorghiou, A., Noble, J. A., "Learning-based prediction of gestational age from ultrasound images of the fetal brain," *Medical Image Analysis*, vol. 21 (1), pp. 72–86, Apr 2015.

Peer-Reviewed Conference Proceedings

Namburete, A. I. L., Yaqub, M., Kemp, B., Papageorghiou, A. T., Noble, J. A., "Predicting fetal neurodevelopmental maturation in ultrasound images," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 2 of *Lecture Notes in Computer Science*, pp. 260–267, Springer-Verlag / Heidelberg, 2014

Namburete, A. I. L., Stebbing, R. V., Noble, J. A., "Diagnostic plane extraction from 3D parametric surface of the fetal cranium," in *Medical Image Understanding and Analysis (MIUA)*, 2014 — Podium presentation, *Student Bursary Award (awarded to top 6 submitted papers)*

Namburete, A. I. L., Stebbing, R. V., Noble, J. A., "Cranial parametrization of the fetal head for 3D ultrasound image analysis," in *Medical Image Understanding and Analysis (MIUA)*, 2013

Conference Presentations

Namburete, A. I. L., Noble, J. A.,"Machine learning-based tool to predict gestational age from ultrasound images of the fetal brain," in *The 2nd Annual*

Oxbridge Women in Computing Science Conference, March 2015, Oxford, UK — Podium presentation, *Best Poster Award*

Kemp, B., **Namburete, A. I. L.**, Noble, J. A., Kennedy, S., Papageorghiou, A. T., "Automated image analysis to estimate gestational age between 18 and 34 weeks using a single 3D brain volume dataset," in *International Society of Ultrasound in Obstetrics and Gynecology (ISUOG)*, September 2014, Barcelona, Spain — Podium presentation

Namburete, A. I. L., Noble, J. A., "Clinical plane selection from 3D parametrized fetal cranial surface," in *The 7th Annual Oxford University Imaging Festival*, October 2013, Oxford, UK

Namburete, A. I. L., Noble, J. A., "Fetal cranial segmentation using super-pixels in 2D ultrasound images," in *The 6th Annual Oxford University Imaging Festival*, October 2012, Oxford, UK

Articles in Preparation

Namburete, A. I. L., Kemp, B., Rueda, S., Papageorghiou, A. T., Noble, J. A., "Analysis of ultrasound images of the fetal brain" (Review article)

Rueda, S., **Namburete, A. I. L.**, Papageorghiou, A. T., Noble, J. A., "Manifold learning to assess the effect of malnutrition on brain development"

2

Review of Fetal Brain Maturation

This chapter contextualises the work presented in this thesis. We review the state-of-art in clinical methods for assessing fetal growth and maturation, and discuss the pertinent literature in automated methods for neuroimage analysis of both MR and US image data.

2.1 Clinical Methods for Estimating Gestational Age

In obstetrics, **gestational age** (GA) is defined as the time elapsed since the last menstrual period (LMP), which coincides with 14 days prior to fertilization. Traditional approaches to GA estimation include: *a*) menstrual dating, in which the LMP serves as a reference point for the expected date of delivery (EDD), and *b*) ultrasound-based biometry, in which linear measurements are collected from two-dimensional (2D) ultrasound (US) images and regressed to population dating charts to assess normality of fetal growth [9]. This section discusses the methods for GA estimation used in current clinical practice, and their limitations.

2.1.1 Menstrual Dating

The first day of the LMP has been accepted as the basis by which to assess GA in the clinic. It provides a proxy indicator for the timing of conception, allowing the EDD to be calculated by adding 280 days (40 weeks) to the first date of menstruation [10]. Once an EDD is established, GA at any point in the interim period can be derived by subtraction of dates. Whilst the LMP approach is simple, it is based on the assumptions that the duration of pregnancy is a constant 280 days and that women ovulate at a fixed time during each cycle, thus expecting that menstrual cyclicity has a constant relation to ovulation and conception [1]. This requires that the methods by which women track their menstrual cycle and determine their LMP must be accurate.

In practice, menstrual dating is limited by the fact that menstrual history is unreliable in 10-45% of pregnant women [11, 12, 13]. Even of the women with 'exact' knowledge of their LMP, the predicted delivery date still differs

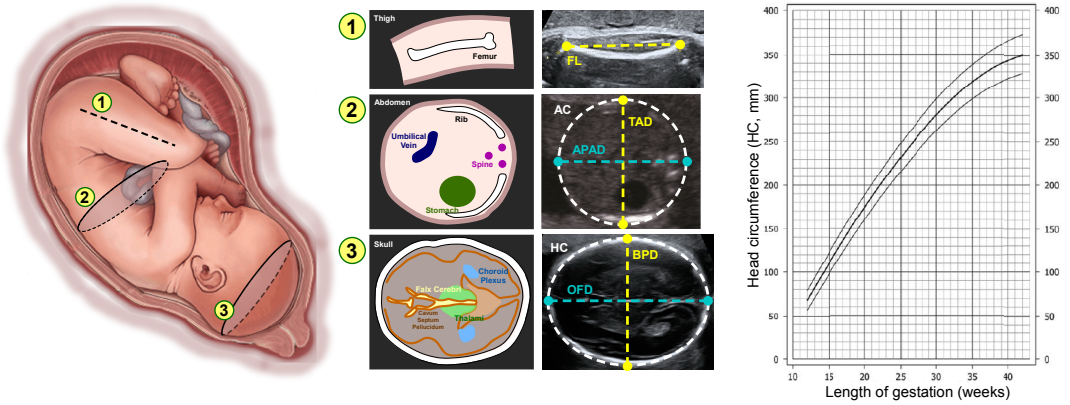


Figure 2.1: **Clinical assessment of fetal growth in the second and third trimesters.** Illustration of the process of clinical US-based GA estimation. a) 2D US scans are collected from (1) the femur, (2) abdomen, and (3) head of the fetus. b) The correct image planes are identified by the presence of anatomical landmarks, and linear measurements are collected from the images. c) These measurements are then regressed to population-based dating charts for GA estimation. (Chart obtained from Figure 5 of the work by Loughna et al [9].)

by approximately 13% days [14]. This is mainly due to *a*) the variability in the length of the ovulation phase of the menstrual cycle, *b*) the inconsistent timing of ovulation and conception in relation to LMP (varying by 8-35 days) [15], *c*) the fact that some women experience postconception bleeding [16], and *d*) the use of oral contraceptives within two months of conception [11]. Despite the biological definition of LMP, these factors contribute to the intrinsic inexactness of using it for GA assessment, rendering it an insecure basis for GA estimation.

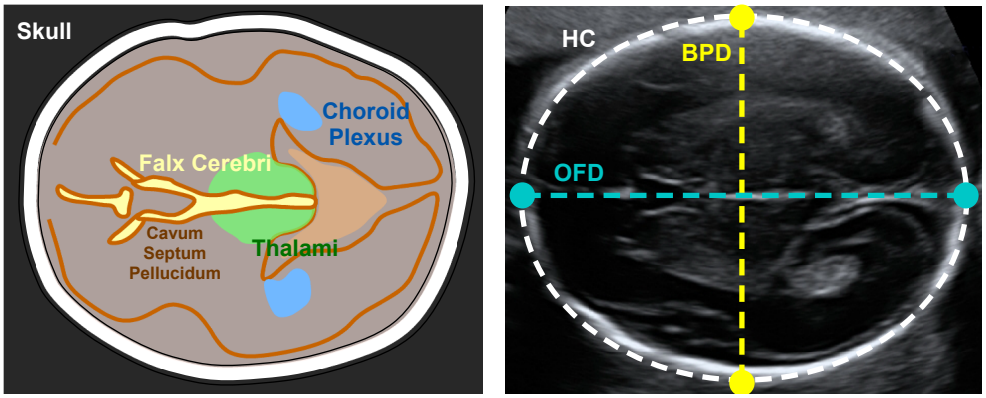
In the absence of clinically-useful menstrual history, US-based measurements provide the most accurate estimate of GA with an accepted accuracy of ± 10 days, if performed before 22 gestational weeks (GW) [3].

2.1.2 Ultrasound-based assessment

Since the late 1970s, 2D US has been routinely used in obstetric care to estimate GA *in utero*, to monitor growth by estimating the fetal size and weight, and to identify growth patterns and abnormalities throughout gestation. During an examination beyond the first trimester of pregnancy, the ultrasonographer collects standard biometric measurements from the fetal head, abdomen, and thigh, as shown in Figure 2.1[1] [17]. These measurements are then regressed to population-based growth charts to assess the normality of fetal growth and to estimate GA. The population-based growth charts are based on a variety of dating formulae derived by quadratic or cubic regression which correlate the biometric measurements with GA (reviewed in [18]).

In the standard neurosonographic examination, an axial plane (the transthalamic plane, Figure 2.1) is used for assessing the fetal head. The ultrasonographer selects this plane by visually identifying anatomical landmarks within the brain, such as the cavum septum pellicidum and choroid plexus (Figure 2.2a), [17]. As shown in Figure 2.2b, the parameters acquired from the fetal head are the biparietal diameter (BPD), occipitofrontal diameter (OFD), and the head circumference (HC), which are used for assessing age beyond 14 weeks of gestation. The BPD is obtained by measuring the diameter from the outer borders of the parietal bones where the skull is widest. The OFD is measured from the outer boundary of the occipital edge to the outer boundary of the frontal edge, where the skull is the longest. The HC is obtained either by approximating the circumference of an ellipse with the BPD and OFD as the minor and major axes, respectively, or by using the following equation:

[1] All illustrations of the fetus used in this thesis are adaptations on a drawing available at http://www.hopkinsmedicine.org/healthlibrary/GetImage.aspx?ImageId=142107



(a) Anatomical landmarks (b) 22 weeks

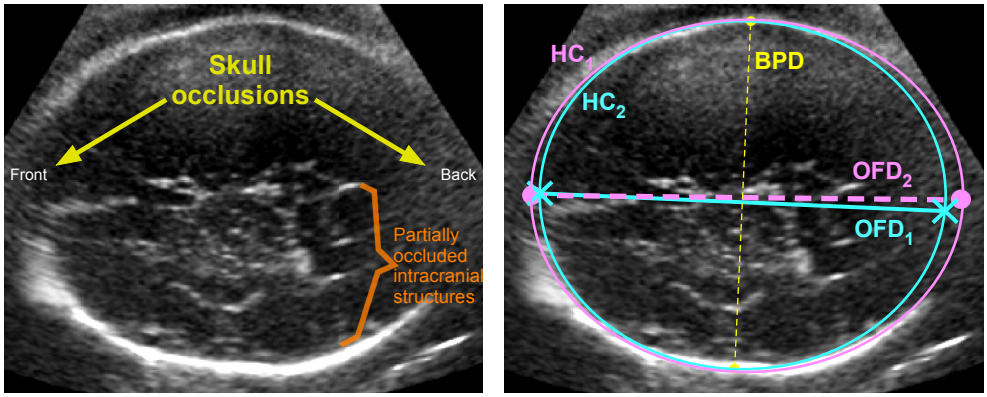
Figure 2.2: **Transthalamic plane for fetal head biometry.** a) Anatomical landmarks necessary for identifying the correct axial plane for extraction of biometric measurements from the fetal head include: Cavum septum pellucidum, falx cerebri, thalami, and choroid plexus. Example calliper placement shown for a fetus at 22 GW. Measurements include: biparietal diameter (BPD), occipito-frontal diameter (OFD), and head circumference (HC).

$$HC = \frac{\pi(BPD + OFD)}{2} \tag{2.1}$$

Finding the correct plane from which to extract these biometric measurements is a challenging and highly subjective task, as the visibility of intracranial structures is confounded by *a*) the experience of the ultrasonographer, particularly in the reliance on his/her knowledge of rapidly changing fetal brain anatomy; *b*) the contrast settings of the US machine; *c*) the thickness of maternal skin and fat tissue; and *d*) fetal pose.

Even in the case of optimal plane selection, the GA prediction is affected by the choice of biometric parameter(s) collected. Studies have shown that HC is a better predictor than BPD, as the latter is dependent on head shape whereas the former is less so [2, 19, 20, 21, 22]. However, in late pregnancy the accuracy of all predictors decreases [22]. In particular, due to varying degrees of partial occlusions of the skull and fetal positioning, the landmarks necessary for assessing the OFD might not be clearly visible, making the OFD measurement

error-prone (Figure 2.3; [21]).

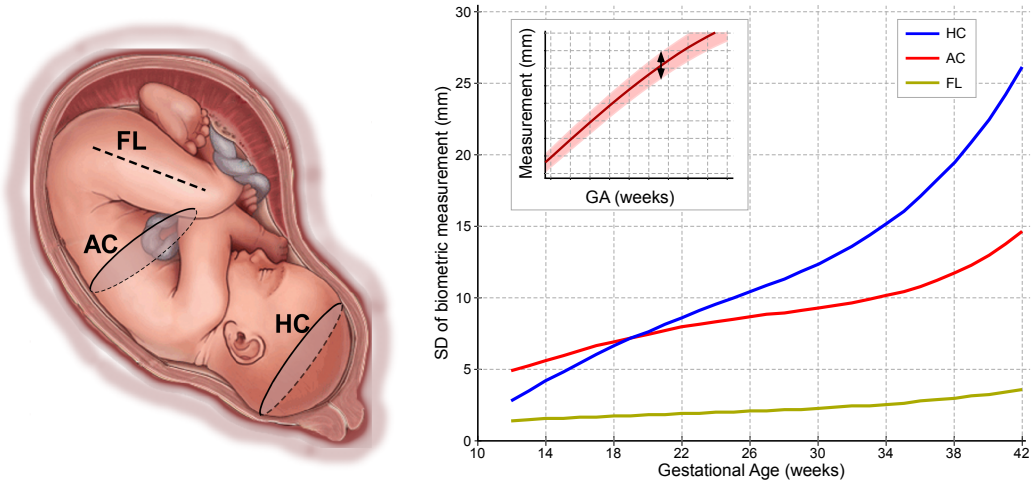


(a) (b)

Figure 2.3: **Occlusions and errors in calliper placement.** Calliper placement for extraction of biometric measurements depends on visibility of the cranial boundaries. Example calliper placement on a 2D slice collected at 34 GW. Note that at this advanced gestation, intracranial structures are less visible and the frontal and rostral parts of the skull are partially occluded, obscuring the task of calliper placement, and hence OFD determination.

Since the HC is a function of OFD, the use of biometry in the third trimester is prone to errors in age estimation, varying in practice by up to ±3 GW in the third trimester (Figure 2.4; [2, 22, 3]). In addition, as gestation progresses, fetuses manifest size variations due to intrinsic factors such as ethnicity [23], genetic growth potential, and variations in growth support from the placenta and mother [24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 19, 34, 35, 36, 37] which would ideally need to be incorporated into the GA estimation as supplementary non-anthropometric fetal and maternal measures, which would be a non-trivial task.

As pregnancy advances and biological heterogeneity amongst normal fetuses increases, so does variation of values of each biometric measurement associated with a specific GA, thus rendering the size-based equations less accurate (Figure 2.4). In practice, this means that whilst the predictive error at 22 weeks' GA (±10 days, [3]) is considered acceptable in the majority of clinical settings, the predictive error at 28-42 weeks (±18 days) is considered to offer little



(a) (b)

Figure 2.4: **Errors in fetal biometry.** Standard deviation (in mm) of three fetal biometry measurements with respect to gestational age (in weeks). Measurements include: head circumference (HC), abdominal circumference (AC), and femur length (FL). Note that as gestation advances, the variation in the range of biometric values increases.

clinical value [38].

To achieve the most accurate GA prediction, growth must be assessed between two known gestational time-points, so fetal biometry cannot be confidently used in the third trimester if not put in direct relation with earlier trimester measurements (e.g. first trimester when fetal growth is exponential and independent of genetic and environmental factors) [1]. However, this may not be a possibility in low-income settings where women only attend for prenatal care late in pregnancy — typically after 24 weeks' gestation (Figure 2.5) — when menstrual history is unavailable or unreliable. Thus, in a setting where screening happens *only once*, the error margins yielded by current methods render them as not clinically useful, potentiating the need for more accurate alternative techniques for personalized **GA predictions from data collected in a *single* clinical visit in the 2nd and 3rd trimesters of pregnancy**.

Neuroanatomical studies have observed that in early life, the human brain

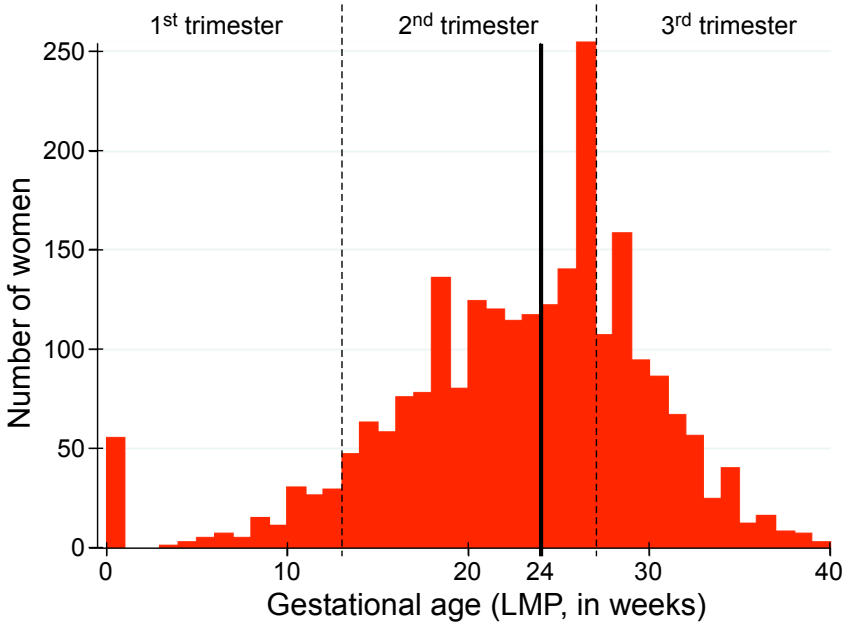


Figure 2.5: **GA at which women present for care.** Histogram of gestational age at which expectant women present for prenatal care at a clinic in Kilifi, Kenya. (Figure used with permission from Dr. Kemp, Nuffield Department of Obstetrics and Gynaecology, University of Oxford.)

follows a predictable trajectory of spatial and temporal changes [16, 39]. This process is observable non-invasively in the form of bright echoes in US images [40] marking the emergence/disappearance of cerebral structures. That being the case, it is conceivable that the developmental state of the brain may serve as an accurate indicator for GA. However, despite the fact that US has been used in prenatal care for four decades and a clear link has been established between brain development and chronological age, these patterns have not been used to predict GA in the clinic.

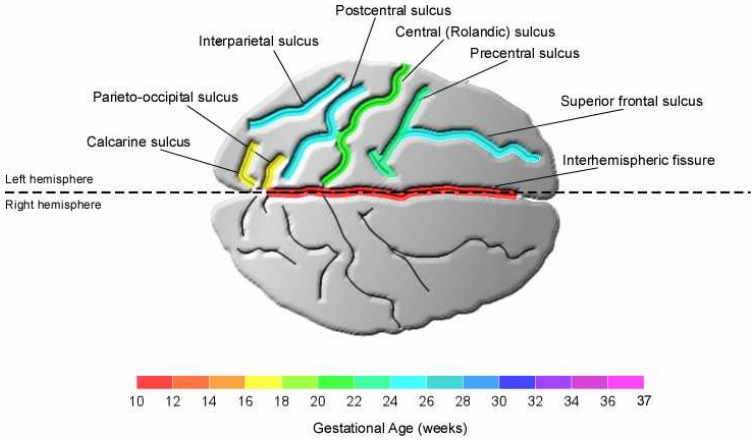
The next section discusses the current understanding of *in utero* neurodevelopmental progression, and outlines the process of fetal brain development, as observed in US images.

2.2 Fetal Neuroanatomical Development

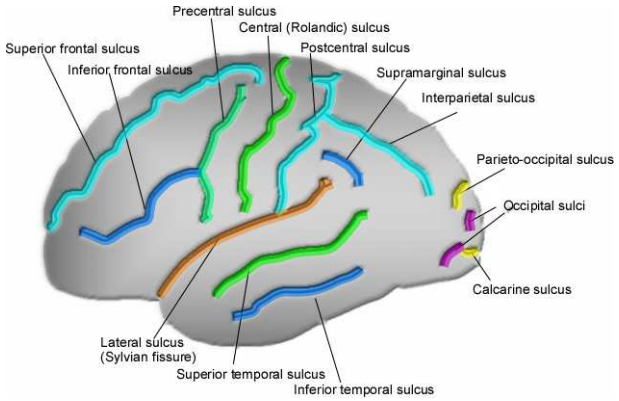
Central to any discussion of fetal development is the concept of GA. Subject to variation in both definition and measurement, differences in GA estimates between studies can introduce bias of up to 2 weeks in reports of times at which structures first emerge or disappear. Caution should be taken when comparing findings from different studies, but for the purposes of this discussion and the remainder of this thesis, GA is defined as the number of post-ovulatory weeks at any specific time in pregnancy.

2.2.1 Process of Fetal Brain Maturation

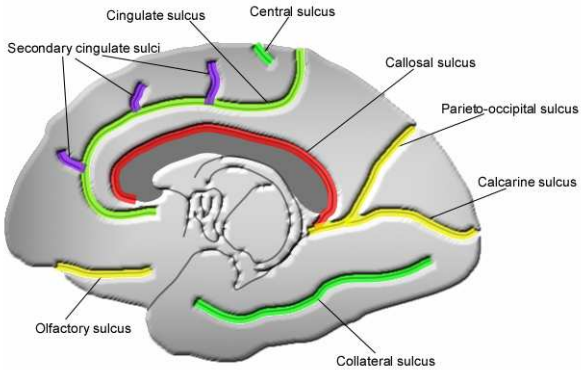
In the late 1970s, neuroanatomical studies using post-mortem fetal brains identified that normal development of the cerebral cortex follows a predictable timetable throughout gestation [16, 39]. This process of *sulcation* (or *gyrification*) is characterized by the time at which indentations, *sulci* and *fissures* (or protrusions known as *gyri*), first appear on the cortical surface and the pattern by which they increase in complexity. Figure 2.6 shows a graphical representation of the spatiotemporal sequence by which sulci emerge on the cortical surface based on the findings of a key post-mortem neuroanatomical study conducted by Chi et al [16]. According to their post-mortem reports, the earliest fissure to appear is the interhemispheric fissure (Figure 2.6a) and is first visualized at the 8^{th} gestational weeks (GW) [16]. It forms the longitudinal separation between the cerebral hemispheres and is fully demarcated by the 10^{th} GW. At this point, the hemispheres remain smooth until the first appearance of a shallow depression on the lateral surface— an indication of the Sylvian fissure— in the 14^{th} GW. As gestation advances, the Sylvian fissure becomes progressively grooved following a characteristic pattern of shape



(a) Dorsal aspect



(b) Lateral aspect



(c) Medial aspect

Figure 2.6: **Pattern of spatiotemporal development of the cortical surface.** Graphical representation of the timeline of *in utero* sulcal appearance demonstrating the sulci present on the (a) dorsal, (b) lateral, and (c) medial aspects of the cortical surface and the corresponding gestational times of first appearance according to the post-mortem neuroanatomical study by Chi et al [16].

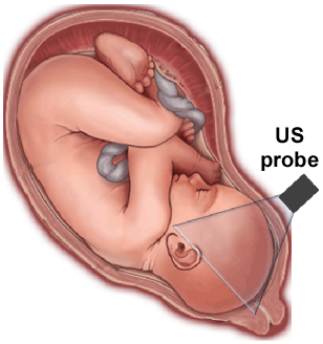
changes, which has also been described using US imaging of the brain (reviewed in [4]).

At the beginning of the second trimester, the hemispheres remain relatively smooth apart from the presence of the Sylvian fissure. As shown in Figure 2.6, more sulci begin to appear as the second trimester progresses. By the 22^{nd} GW, the primary sulci are all apparent on the surface: the Sylvian fissure (14^{th} week), the calcarine sulcus and parieto-occipital sulcus (16^{th} week), the central sulcus (20^{th} week), and the callosal sulcus (22^{nd} week).

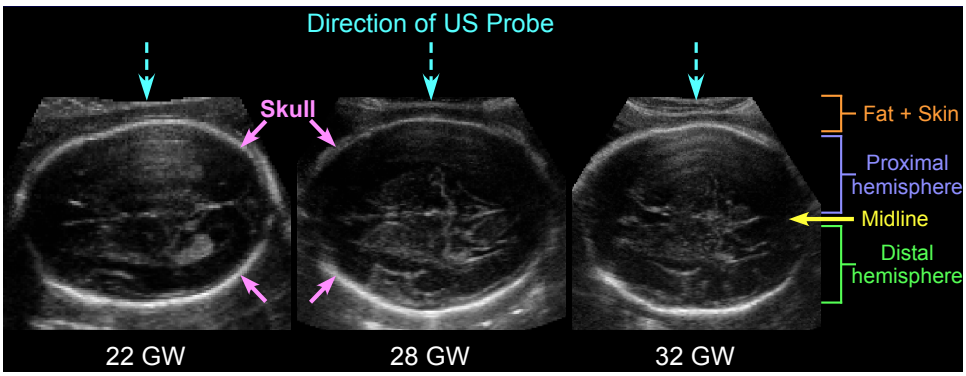
Marked changes during the third trimester of gestation see the brain develop from its primordial relatively smooth and agyric structure into a convoluted structure whose surface morphology closely resembles that of the adult brain (see Figure 1.1; [41, 42]). During this period, neuronal organisation results in the deepening of all sulcal indentations and the emergence of secondary sulci (e.g. secondary cingulate sulci at 32 GW, Figure 2.6c) on the cortical surface.

With the emergence and deepening of the secondary sulci, gyri, and fissures, an increase in bright, linear echoes is seen in US images. Such image signal activity peaks at 28 to 30 GWs [43], corresponding to increased structural changes observed by post-mortem anatomical studies [39]. Figure 2.7 shows the typical appearance of the cortical plate in US images at different GAs. Note that in US images, the cortical plate appears as a hyperechoic (bright) structure with folds at distinct locations: the sulci, which indent the cortical surface and form the boundaries of gyri. However, the intensity of these echoes also decreases with progressing age due to the increasing calcification of the skull bones [44].

All structures follow a predictable pattern of morphological shape changes as gestation progresses and Figure 2.8 demonstrates that these are observable in US images. For instance, the Sylvian fissure (Figure 2.8a) appears in an



(a)



(b)

Figure 2.7: **Ultrasound imaging of the fetal brain.** Due to the interactions between the US beam and maternal tissues and calcification of the bony skull, the proximal hemisphere (above the midline, or *falx cerebri*) is obstructed, and only the structures in the distal hemisphere are observable. This effect is more apparent in later gestation.

axial view as a ridge closely located to the skull at early gestation, but gets progressively pinched due to the overgrowing frontal and temporal lobes. The parieto-occipital fissure (Figure 2.8b) and cingulate sulcus (Figure 2.8c) are located around the midsagittal plane and are often sonographically observed in both cerebral hemispheres. They tend to follow a similar pattern: appearing at first as two smooth arcs, their folds appear and gradually extend into the cerebral hemispheres, approximately mirroring one another. The superior temporal sulcus (Figure 2.8d) starts off smooth and increasingly develops rippling folds which steadily deepen with progressing gestation. From this figure, it is clear that US images of the brain are content-rich and may indeed be used as

a neuroimaging tool to observe cortical folding, yet this imaging modality has never before been used for *in utero* assessment of brain maturation.

2.3 Reports of Cortical Development

To date, visible sulcation of the developing brain has been described by post-mortem neuroanatomical studies [16, 39, 45, 46], MR [47, 48, 49, 50, 51, 52], as well as by US [53, 44, 8, 54, 43, 55, 40] examinations. Most studies have reported the first GA at which a clear indication of sulcal presence could be identified on the brain surface. Reported modality-specific visualization times are summarized and tabulated in Appendix A, and this section discusses the considerations and challenges associated with each method of observing brain maturation.

Neuroanatomical Reports Post-mortem neuroanatomical studies formed the basis for the first descriptions of fetal brain development [56, 16, 39]. Whilst frequently cited as reference standards for imaging studies of the brain, there are a number of reasons to suggest that such comparisons may not be appropriate. Firstly, the processes involved in extraction and chemical fixation of post-mortem brain tissues can influence structures to such an extent that the tissues examined do not accurately represent the structures *in vivo* or, therefore, on imaging studies. These studies inherently assume that during chemical preservation of the brain tissue the resulting shrinkage is *a*) uniform both within the brain and on the cortical surface, *b*) to the same degree in the brains of different subjects at different GAs, and that *c*) it does not affect the observer's ability to detect emerging sulci [41, 57, 58, 59]. Furthermore, since these studies are only possible post-mortem, it is not known whether the processes underlying the interruption of pregnancy are, themselves, associated

with aberrations in cerebral morphology.

MRI Reports Early non-invasive studies of the fetal brain were performed using MRI to image preterm infants (i.e. born prior to 37 GW) and included subjects from pregnancies that miscarried or ended in stillbirth [51, 47, 49, 48, 50]. As with post-mortem studies, these data assume that the processes underlying early delivery do not affect brain structure, which is inaccurate [60, 61, 62]. Thus, it cannot be assumed that the existing descriptive accounts of 'normal brain maturation' that, themselves, are based upon data from abnormal pregnancies, adequately describe the processes of healthy cerebral maturation. Early attempts at imaging the fetal brain *in utero* were challenged by fetal and maternal motion artefacts, which resulted in distorted geometric reconstructions of the image [63]. However, the introduction of sophisticated phased-array receiver coils and rapid acquisition sequences such as single-shot fast spin echo (SSFSE) and half-Fourier acquisition single-shot turbo spin echo (HASTE) make it possible to capture images free of fetal motion artefacts [64, 65, 66]. Though still a challenge, image reconstruction and motion correction are being addressed by super-resolution reconstruction techniques, in which a high-resolution volume is constructed by registration of multiple low-resolution images [67, 68, 69]. Such advances are making possible the study of detailed *in utero* brain development using MR images. However, in clinical practice MR-based assessment of the brain is prompted only by referral from inconclusive US examinations.

US Reports US offers real-time visualisation of the fetal brain *in utero* [44, 8] using both transabdominal (TA) and transvaginal (TV) approaches. Since the first published studies from the 1980s [53], improvements in spatial and contrast resolution, noise reduction, increased dynamic frequency ranges, and depth-of-

field visualization have led to better image definition of anatomical structures. Thus, variation in sulcal visualization times from early US studies may reflect the lower spatial resolution of earlier machines which affected the visibility of cortical structures (Appendix A, Table A.3).

At present, standard US views of the brain are impaired by acoustic shadows, which are the consequence of increasing calcium density of the cranial bones, a process termed *calcification* ([70]; Figure 2.7). The presence of the fetal skull also causes reverberation artefacts: multiple reflections of the US beam on the fetal skull and maternal tissues. These artefacts result in shadows and missing image features (particularly in the cerebral hemisphere proximal to the US probe), and variable signal strength within the imaged tissues. Due to the partial obstruction of the proximal hemisphere [71, 72], sonographic studies are limited to basing their reports on *one* of the cerebral hemispheres and the artefacts hide potential structural asymmetry within the developing brain [73, 74, 75, 40, 76]. Where relevant, care should be taken to ensure that 'like-for-like' comparisons are made with respect to proximal and distal hemispheres and left-right symmetry.

With the introduction of 3D US, it has become possible to examine structures using reconstructed oblique planes in addition to the traditional orthogonal frame [70, 77, 40]. Offering better views of structures, images can also be reviewed 'offline' thereby reducing the dependence upon operator expertise for diagnosis and reporting [40, 78]. Furthermore, recent developments in 3D US probes address the issues associated with reverberation artefacts, introducing as possibilities the capture of high resolution data with the ability to observe age-specific changes in surface anatomy and structural appearance, *in utero* [79].

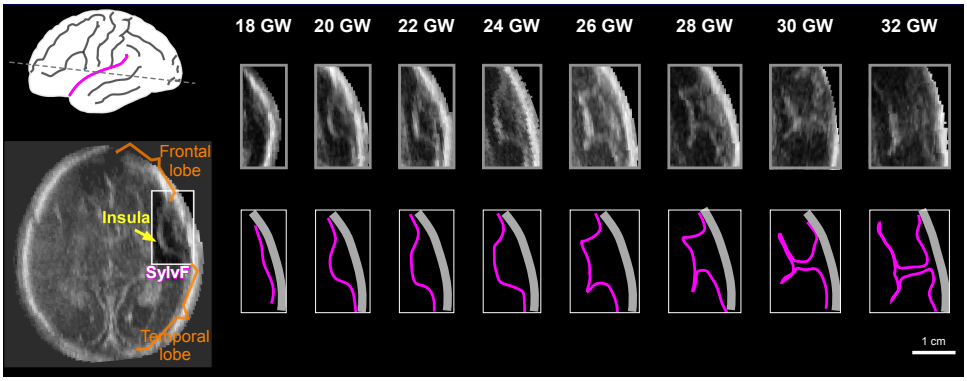
Although MR images are typically used to study the brain, it is possible to

observe the same structures in US images despite differences in signal intensity. Note that in the T_2-weighted MR image in Figure 2.9, the cortical plate has the same overall topographical appearance as US, but due to a low signal intensity it is visualised as a hypointense contour separating the gray cerebral space from the cortical gray matter. In the US image, the cortical plate produces a strong signal response, hence its bright appearance. It is also evident from the figure that while there is correspondence between the structures visible in US and MR images, there may not be a direct mapping between the intensities of each modality [80]. Yet, it has been demonstrated that US may be as useful as MRI in assessing cortical development and diagnosing malformations [81].

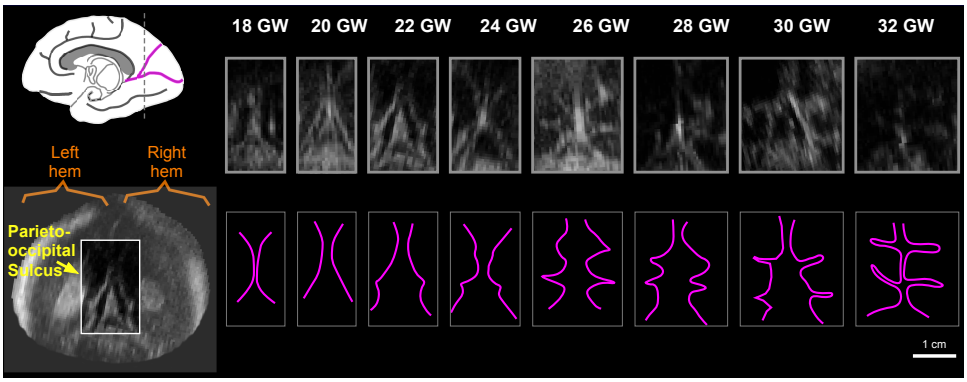
2.3.1 Sulcal Appearance Times

The temporal sequence of sulcal development is summarized in Figure 2.10. As can be seen from this figure, there is correlation between neuroanatomical, MR, and US reports, yet only approximate agreement between them. The sulcal appearance times using US and MR are consistently later than those reported by neuroanatomical studies, with MR being up to 8 GW later [63, 47, 48], and US between 2–4 GW [53, 82, 8]. Such discrepancies may be the consequence of:

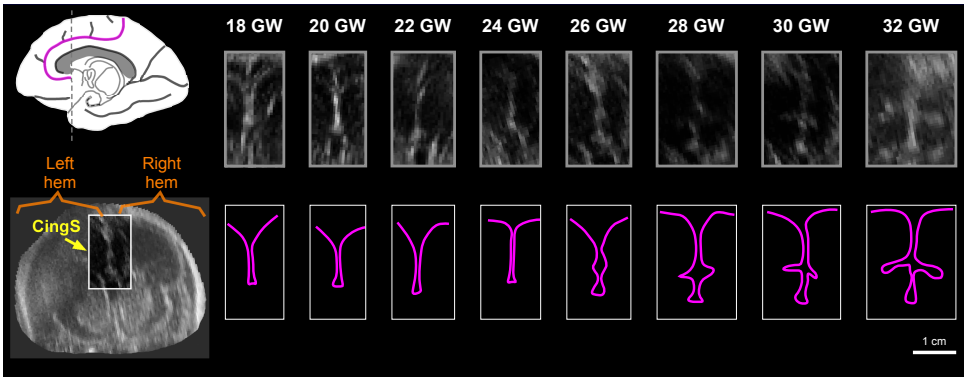
- Methodological inconsistencies between neuroanatomical studies used as reference standards for imaging studies;
- Discrepant or inaccurate GA estimation methods;
- Small sample sizes;
- Inconsistent use of methods to determine sulcal presence/absence;
- Variations in the appearance (in intensity or texture) of structures when



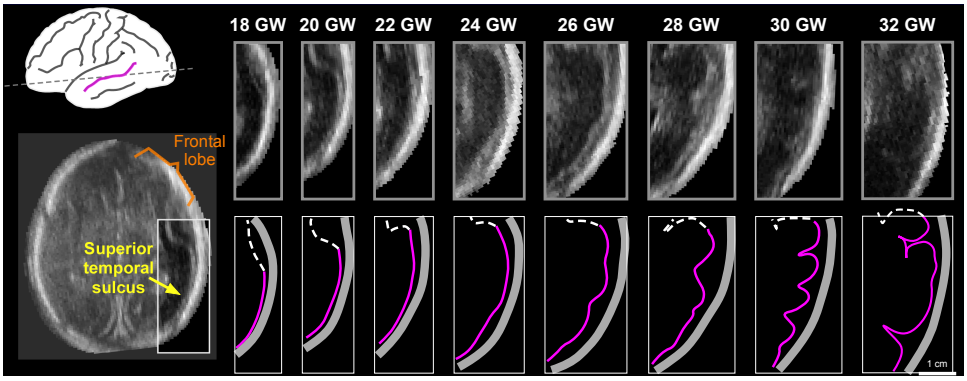
(a) Sylvian fissure



(b) Parieto-occipital fissure



(c) Cingulate sulcus



(d) Superior temporal sulcus

Figure 2.8: **Age-related sonographic patterns of brain development.**

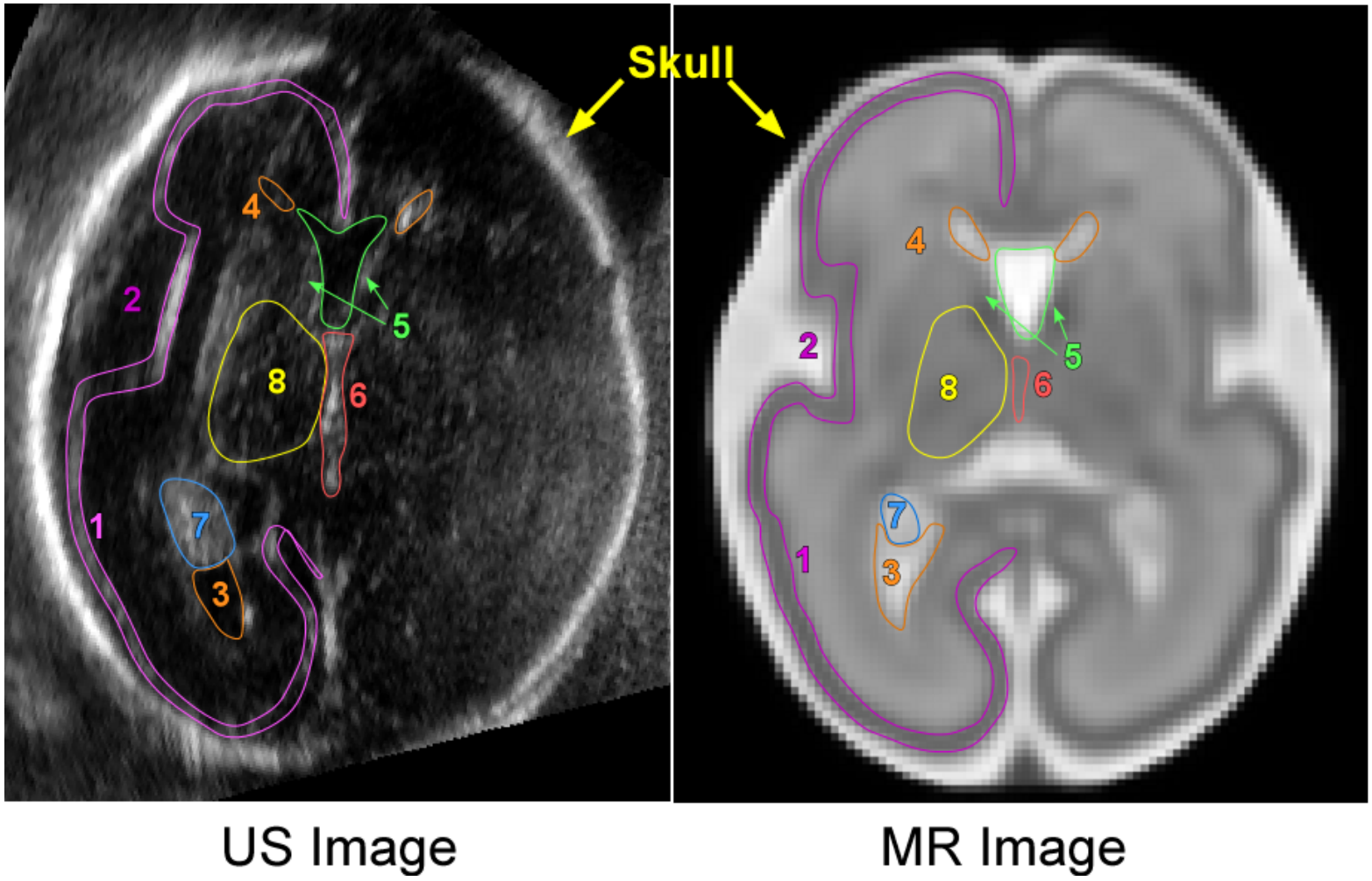


Figure 2.9: **Visual correspondence between fetal brain anatomy in 2D US and MR images.** Example axial sections through the Sylvian fissure of fetal brains at 24 GW as visualized by (a) US and (b) T_2-weighted MR for different subjects. 1. Cortical plate; 2. Open Sylvian fissure; 3. Posterior ventricular cavity; 4. Anterior ventricular cavity; 5. Cavum septum pellucidi; 6. Falx cerebri; 7. Choroid plexus; 8. Thalamus

imaged with different modalities.

It is worth noting, however, that there is high agreement in the first appearance of sulci developing at later gestations (e.g. central, pre/post-central, and superior/inferior frontal sulci), whereas structures which emerge at earlier gestations (e.g. Sylvian fissure, callosal sulcus, cingulate sulcus) have higher discrepancies in appearance times. This may be attributed to the fact that most imaging studies report on sulcal times starting at different GAs (e.g. including infants from as early as 12 GW [49] or as late as 23 GW [50]), which results in unclear distinctions between structures which were already developed by the starting GA, and structures which only emerge at the starting GA.

Given the range of times at which a single sulcus becomes visible, even according to a single imaging modality, it may be appropriate for modality-specific references to be developed, rather than benchmarking one technology as the referential "gold standard". Fetal brain imaging, at present, is highly operator dependent and practice varies considerably in different clinical settings. It must, therefore, be stressed that there is no 'one-size-fits-all' approach, and, that the choice of modality will frequently be determined by a range of factors. However, given that US is the modality of choice in prenatal care and that it lends itself to the visualisation of cortical maturation [40], it is imperative that we develop an understanding of the normal (healthy) trajectory of cortical progression as observed in US images. Guided by an US-specific model for healthy maturation, it may be possible to determine the degree to which a deviation from the normal trajectory may be associated with pathological malformations, or when it is simply representative of normal physiological variation.

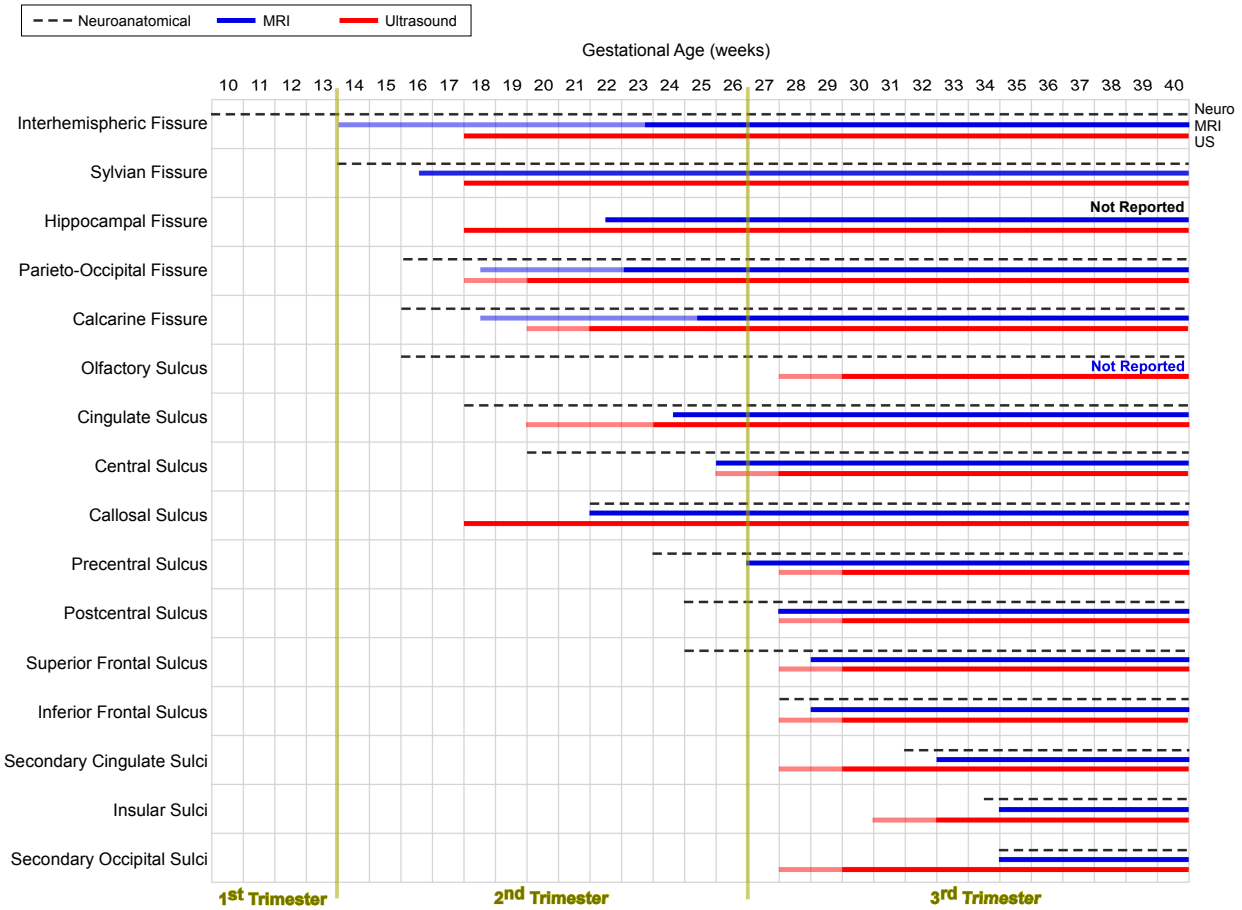


Figure 2.10: **Agreement between sulcal appearance schedules reported by post-mortem neuroanatomy, MR, and US.** Timings (in GW) of the first appearance of each sulcus or gyrus as reported by key neuroanatomical ([16], dashed lines), MRI ([51], dark grey continuous lines), and US ([43], light grey continuous lines) studies. The opaque and solid portions of each line represent the gestational ages at which structures were visualized in 25%-75% and 75% of the images observed, respectively. The time of first visualization of the hippocampal fissure was omitted from the anatomical reports, as was that of the olfactory sulcus in the MRI reports.

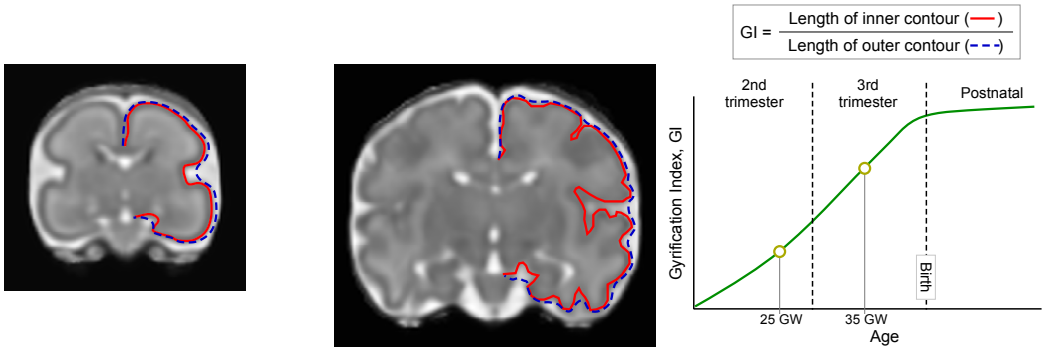
2.4 Qualitative Assessment of Cortical Maturation

Throughout normal pregnancy, sulcal complexity of the cerebral cortex increases (Figure 2.6) and the first methods for quantifying cortical folding were the *gyrification index* and linear sulcal measurements. Macroscopic markers of sulcal maturation include shape, appearance, and depth (i.e. distance between deepest point in fissure and cranial bone or cortical surface) of sulci on the cerebral cortex. Although automated computational techniques to capture these features are currently available, none have been translated into diagnostic aids within a clinical setting. Whilst early studies extracted measurements from 2D US images, recent developments in 3D imaging have motivated more complex computational approaches, discussed in § 2.5. This section reviews the qualitative descriptions and manual measurements of cortical morphology presented in the clinical literature.

Gyrification Index (GI)

The earliest measure of cortical convolutional complexity was the 2D *gyrification* (GI) or sulcation index, first introduced by Zilles et al [83]. It is defined as a ratio between the perimeter of the folded inner cortical surface (including sulci; Figure 2.11a, in red) on the coronal plane and the length of the outer surface (Figure 2.11a, in blue). The GI yields a high value for a brain with a high degree of cortical folding, and vice-versa. Initially, these path lengths were manually measured from photographs of coronal sections of post-mortem brains [84, 83]. Through these studies, it was observed that during the gestational period, the GI steadily increases until the third trimester, then it plateaus at birth and remains constant throughout the lifespan ([41], Figure 2.11c). A

similar measure of cortical complexity is the degree of cortical folding (DCF) introduced by Zhang et al [85]. It differs from the GI in that it is computed on the axial plane as opposed to the coronal plane. However, both measures are limited by the fact that they can only be measured from 2D sections of the brain, inherently not accounting for the 3D nature of the cortical surface. Furthermore, they are not directly applicable to US images due to the partial occlusions and missing image features which would complicate the manual delineation of an, ideally, complete contour of the cortical plate.



(a) 2D GI (25 weeks) (b) 2D GI (35 weeks) (c) GI trends

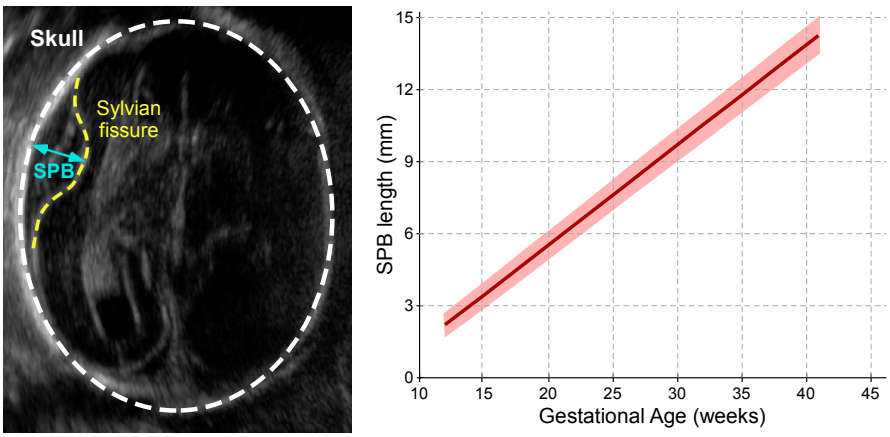
Figure 2.11: **Gyrification index to measure cortical folding.** The 2D gyrification index (GI) on coronal fetal brain MR images at (a) 25 weeks and (b) 35 weeks of gestation. The 2D GI is the ratio between the geodesic path of the folded cortical plate (outlined in red) and its convex hull (outlined by a blue dashed line) (2D MR images were extracted from the atlas database constructed by Serag et al [86]). (c) A theoretical representation of the GI with advancing age.

Sulcal Morphometry

In studying sulcal shape variants, the simplest descriptors comprise length, area, and depth measurements. Early US-based studies of sulcal morphometry have investigated manual image analysis approaches for quantifying sulcal morphometry from US images, typically involving depth measurements. *Sulcal depth* (SD) is defined as the Euclidean distance between the deepest point

of the sulcus (i.e. sulcal basin) and a predetermined reference point, usually the inner cranial contour (Figure 2.12a). For instance, US studies have developed simple, semi-quantitative scoring techniques to correlate sulcal depth and appearance to GA [87, 88, 89, 40, 90]. The most commonly investigated is the Sylvian fissure due to its characteristic pattern of *operculization* (or *opercularisation*: sulcal pinching due to overgrowth of the cerebral lobes) (see Figure 2.8a; [87, 89, 40]), which can be clearly visualized in US images as early as 16 GW [54]. These approaches have focused on selecting axial 2D US slices of the fetal head and analysing the changes in the Euclidean distance between the 'midpoint' of the Sylvian fissure and the parietal bone (SPB, as shown in Figure 2.12, [87, 89]), or by measuring the angle between the insula and the temporal lobe and assigning it a score that indicates the stage of development (as shown in Figure 2.13, [40, 90]). These scores enable the simple quantification of sulcal development and are suggested as suitable predictors of delayed cortical development in abnormal fetuses [90]. However, they are highly operator dependent, with random errors introduced by variation in calliper placement to measure depths and angles. Furthermore, the scoring techniques are limited to the interpretation of sulcal development singly from a 2D plane, also not accounting for the 3D shape of the cortex, most of which is observable with US [44, 8].

Moreover, the studies which attempted to develop US grading scores of sulcal development assign gestational windows which do not agree with the scores generated by their complementary studies. For instance, Figure 2.13 shows that grades 4-6 (corresponding to 22 to 31 GW) proposed by Quarello et al [90] overlap with grades 3-5 which correspond to 24-40 GW according to the scoring method presented by Pistorius et al [40]. Since both studies employed the same age estimation method (i.e. based upon sonographic measurement of

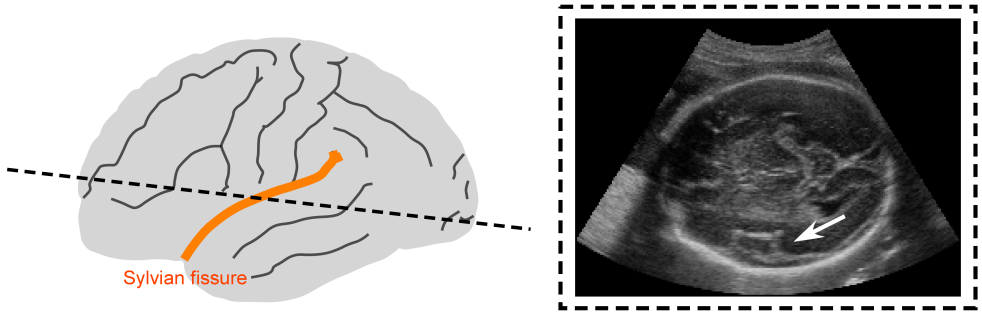


(a) SPB Measurement (b) SPB vs. GA

Figure 2.12: **Sulcal depth measurement in US images.** (a) The distance between the midpoint of the Sylvian fissure and the parietal bone (SPB) is collected from a 2D US axial slice of the fetal brain. (b) This SPB chart for normal development of the Sylvian fissure was proposed in a study by Mittal et al [89] which included 202 fetal patients ranging from 12 to 41 GW. Shown are the 5th and 95th centile curves (shaded in pink), and the mean SPB trajectory (solid red line).

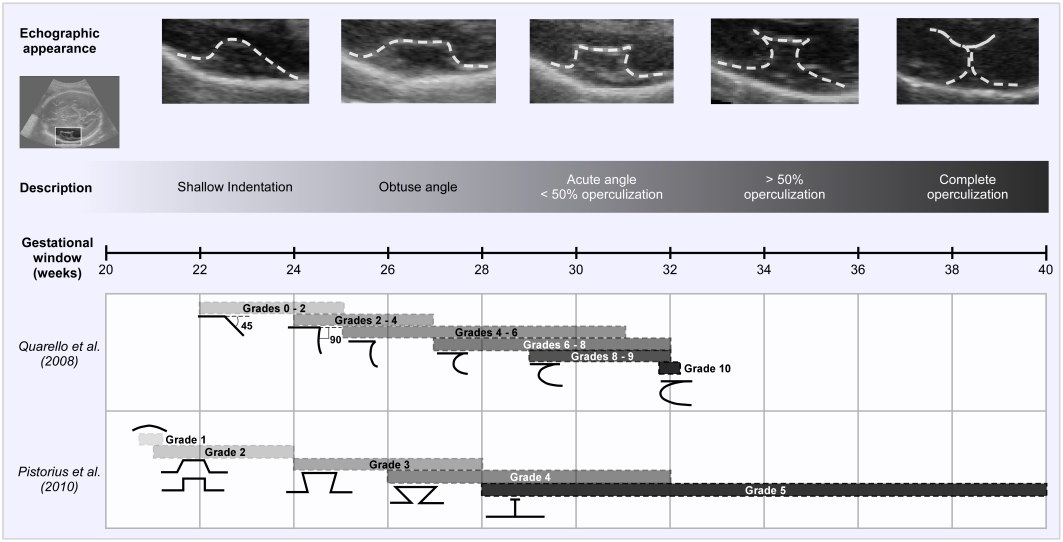
crown-rump length (CRL) before 14 GW), such discrepancies reflect differences in the design of each scoring technique. These observational differences of brain structures in similar developmental states display the subjectivity in the assignment of a score to a level of maturation, particularly in grade scores that fully overlap with adjacent grade scores (e.g. in the method proposed by Quarello et al, grades 8-9 and 10 are embedded within grades 6-8 using the same scoring system). These factors inherently increase the propensity of erroneous assignment of cortical maturation levels.

Though MR-based studies have also investigated sulcal shape and depth measurements on 2D slices [76], their main contribution has been in the development of automated approaches to quantify cortical surface area [91, 52], length [92], sulcal width [93], and cortical thickness [60] from a 3D reconstruction of the cortical surface. These quantitative approaches are discussed in the next section.



(a) Sylvian fissure on cortical surface

(b) US image plane for Sylvian fissure visualization



(c) Operculization process of Sylvian Fissure

Figure 2.13: **Operculization scoring of the Sylvian fissure**. (a) An illustration of the Sylvian fissure on the cortical surface and (b) the plane at which the Sylvian fissure (indicated by white arrow) is typically visualized. (c) The operculization of the Sylvian fissure across gestation as visualized in 2D US images. The progressive grooving of the Sylvian fissure is shown here in the zoomed-in 2D US images displaying the fissure in dashed white lines at different gestational windows in the second and third trimesters of pregnancy. The two studies which have developed scoring methods are shown here: the grayscale intensities correspond to the degree of fissural operculization. Each scoring technique matched a graphical appearance of the operculum to a gestational window.

2.5 Quantifying and Mapping Fetal Brain Development

Automated image analysis methods have been developed to objectively quantify brain growth in an attempt to enhance clinical reporting. In addition to reproducing standard biometry, there are methods designed to extract structural and morphometric data from fetal brain images. These are discussed below.

2.5.1 Automated Biometric Analysis

Since US-based biometric analysis forms the basis for routine prenatal evaluation (discussed in Chapter 2), it is the most prominent in the clinical literature despite developments in MR-based biometry (e.g. [94, 95]). As a result, techniques have been developed to *automate* the extraction of biometric measurements from US images in an attempt to circumvent the inter- and intra-observer variability and to increase work flow efficiency in obstetric examinations [96]. These methods propose a standardization of the measurements to make them objective and reproducible. The automatic methods are mainly designed to measure the size parameters of the head from 2D US images (e.g. BPD, OFD, and HC as shown in Figure 2.2b).

Published US-based biometric algorithms are based on morphological image operators [97, 98, 99], Hough transform ellipse extraction [100, 101, 102, 103], active contour models [104, 105], and deformable models [106, 107]. Table 2.1 summarizes some of the key studies which have automated the extraction of biometric parameters of the fetal head from US images. The table indicates that the goal of these algorithms is to achieve high agreement between the automatically-acquired and the manual expert-annotated measurements. That

being the case, the use of an automated algorithm to obtain the BPD, OFD, and HC addresses the problems associated with operator bias but does not eliminate the errors associated with biometry-based GA estimation, particularly in the third trimester. Despite these proposed methods, accurate automated emulation of clinical biometric measurements still remains a challenge [108]. However, it is worth noting that the more recent studies are working towards the use of 3D image data for biometry assessment [109, 106], and preparation of 3D fetal brain images for automated image analysis including: *a*) the detection and segmentation of intracranial structures [110], and *b*) rigid alignment of brain images to a common reference space [111]. As of yet, none of the research groups have developed models for tracking or detecting developmental changes of *brain* structures using US.

Importantly, it is noteworthy that despite the efforts to automate biometric measurements, only the machine learning-based method developed by Carneiro et al [112] has been offered in a commercial medical Siemens US machine [113]. Philips has developed EPIQ for the provision of quantification and better visualization of US data to assist in the collection of manual biometric measurements. Although these image analysis software packages are now available to clinicians, manual ellipse fitting is still the preferred method for extraction of biometric measurements. This raises the question about the added value of automating extraction of biometric measurements, and whether there is a need to explore novel methods that improve clinical assessment of fetal growth by providing more informative developmental biomarkers which are not easily attainable by manual annotation.

Table 2.1: **Automated methods for fetal head detection.** A summary of the automated methods for fetal cranial segmentation, detection, and biometric analysis in 2D and 3D US images reported in the literature. NR=Not Reported, S=Semi-automated, F=Fully-automated.

Authors	No. Subjects	Age Range (weeks)	Segmentation Technique	GA Reference	GA Estimation Error	Automated	2D/3D
Zador et al (1991) [103]	75	13 - 40	Thresholding, Hough Transform	Manual BPD, OFD, HC	Correlation w/manual: BPD (r=0.986), OFD (r=0.958), HC (r=0.972)	F	2D
Matsopoulos and Marshall (1994) [99]	NR	NR	Morphological Op.	Manual BPD	±8 days (1.050%)	F	2D
Hanna and Youssef (1997) [97]	NR	NR	Thresholding, Hough Transform, Morphological Operators	Manual HC	BPD (r=0.994), HC (r=0.985)	F	2D
Chalana et al (1996) [105]	35	14-37	Active Contour Model	Manual BPD, HC (6 obs.)	± 1 week ; BPD (r=0.999), HC (r=0.994)	S	2D
Chalana and Kim (1997) [104]	30	NR	Active Contour Model (iterative)	Manual BPD, HC (4 obs.)	±2 days	S	2D
Jardim and Figueiredo (2005) [107]	50	NR	Deformable contour, Maximum Likelihood estimation	Manual BPD	± 1 day	S	2D
Lu and Tan (2000) [100]	11	NR	Unsupervised K-means Clustering, Wavelet-based Decomposition, Fuzzy C-Means, Morphological op.	Manual BPD, HC	Error: BPD (5%), HC (2%)	F	2D
Lu et al (2005) [101]	217	12 - 39	K-means, Morphological op., Iterative Randomized Hough Transform	Manual BPD, HC	±1-2 days; BPD (r^2=0.995), HC (r^2=0.988)	S	2D
Yu et al (2008)[102]	215	16-41	Randomized Hough Transform, Fuzzy C-Means, Gradient Vector Flow Snake	Manual contour outline $\rightsquigarrow$ BPD, HC	± 2 days; BPD (r=0.991), HC (r=0.984)	S	2D
Carneiro et al (2007, 2008) [109, 112]	300	12 - 42	Constrained Probabilistic Boosting Tree	Manual BPD, HC (15 obs.)	±2 days if GA<30wks; up to 7 days if GA>30wks; BPD (r=0.985), HC (r=0.996)	F	2D from 3D
Shan and Madheswaran (2009) [114]	NR	5 - 37	Median filtering, Shape sensitive segmentation	LMP and manual BPD, OFD, HC	±3 days	S	2D
Marhaban et al (2010) [98]	20	14 - 38	Morphological ops., Direct Least-Squares Ellipse Fitting	Manual BPD, OFD, HC	0.46-8.49 days; BPD (r=0.997)	F	2D
Chen et al (2012) [106]	11	20-24	3D Snake Deformation, Model-Based Registration	Manual BPD, OFD, HC	± 3 days; BPD (r=0.85), OFD (r=0.81)	F	3D
Namburete et al (2013) [115]	60	25-34	Superpixels (Pixel Clustering); Random Forest Classifier	LMP, First-trimester CRL ($\leq$ 14 GW)	NR (Cranial segmentation only)	F	2D

2.5.2 Morphometric Analysis

In terms of assessing the brain itself rather than the skull as a proxy for brain size/development, the MR imaging community has developed techniques to perform advanced morphometric analysis. Morphometry is the process of characterizing anatomical and physiological differences between anatomies and subjects (e.g. shape and size of structures, or tissue density). In the field of medical imaging a broad range of techniques such as registration, segmentation, and atlasing fall into this category. Morphometric analysis of the brain is a two-stage process: first, the images are registered and spatially normalized to a common reference space, and then the statistics of interest are computed. Current morphometric approaches applied to the brain include:

- Voxel-based (VBM): which aims to identify differences between subject groups by segmenting regions of interest (ROIs) and computing a statistic on a per-voxel basis to reveal information about group membership [116, 117];

- Deformation-based (DBM): which operates on the deformation fields that result from image registration to construct a non-linear model of anatomical differences [93]. Another variant of the DBM is referred to as tensor-based morphometry (TBM) which uses the Jacobian determinant of the deformation matrix to identify group differences [118, 119, 120];

- Feature-based (FBM): computes a probabilistic model which describes local image features in terms of their appearance, geometry, and tissue class or subject-group labels [121, 122, 123].

While VBM generally uses information that is already present in the image to locate statistical differences between spatially normalized images, TBM and DBM use the information from the transformations which align the images to

a reference. Specifically, the images are all non-linearly registered to closely resemble a common reference, and the group-wise differences are encoded within the transformations which map the images to the reference. The term DBM is used when the statistics are calculated on the displacement fields, whereas TBM describes techniques in which the derivatives of the non-rigid transformations are used [119, 120].

Whilst local volume changes of the adult brain are reported based on cortical landmarks such as lobes, it is often challenging to apply VBM techniques to the smooth fetal brain in which some of these landmarks may be absent or underdeveloped [124]. However, TBM methods have been developed by Rajagopalan et al [120] to examine local growth patterns of cortical tissue in fetal MR images. By first performing non-rigid deformations to warp the MR images to a common coordinate system, TBM extracts the statistical properties from the Jacobian maps to measure local tissue volume. This provides a mapping of spatially-varying tissue growth patterns to highlight local tissue expansion relative to overall brain growth [120]. Their results show that different cortical areas grow at different rates, which may be indicative of regional cortical complexity and their respective requirements for functional development at different gestational time-points.

By contrast, FBM does not rely on the registration step to provide a one-to-one correspondence between subject images, which is difficult to achieve in the developing brain. Instead, it is a data-driven learning approach aimed at identifying local patterns of structural anatomy, specifically designed for applications in which accurate alignment of structures is only statistically probable or the data contain high inter-subject variability [121]. Toews et al [122] recently applied FBM to predict the age of infants based on brain features. Their approach modelled structural developmental patterns from image features de-

fined over spatial and temporal windows. Their study demonstrated the ability to learn the widely-accepted spatiotemporal pattern of brain development to automatically make inferences about infant age.

Cortical Shape Analysis

With the progressive emergence of more sulci on the fetal brain surface throughout gestation, cortical complexity increases. Both MR and US studies have attempted to capture the changes in surface morphology and quantify sulcation (discussed in § 2.4), yet MR is the dominant modality in the literature. Reported measures of cortical surface morphology applied to the developing brain include: *a*) global descriptors, such as the 3D GI, brain length, width, height, and volume; and *b*) regional descriptors which quantify surface curvature to extract the convexity (for gyri) and concavity (for sulci) with progressing gyrification.

3D Gyrification Index The 3D GI was introduced in order to address the shortcomings of the 2D GI [125, 126, 127]. Specifically, it is quantified by estimating the ratio of the *area* (as opposed to length) of the folded cortical surface and the *area* of the outer surface. To achieve this, the cortical surface is extracted from 3D MR images and reconstructed using triangulated meshes. The regional GI is either computed at each point as a depth-weighted sum of the neighbouring points on the surface [125, 127], or by computing the ratio between the cortical surface area within a small sphere and the surface area of a disc of the same radius as the sphere [126]. Both of these methods allow for regional GI quantification, extending the GI to become a local descriptor of cortical folding. The 3D GI is used in a number of MR studies of cortical development (e.g. [128, 60, 129]), also confirming a link between cortical complexity

and developmental progression.

Atlases Atlasing techniques have been uniquely applied to MR neuroimage analysis. Brain atlases are a statistical representation of the brain, and contain priors about the tissues and structures present at each stage of development [130]. They are constructed by combining intensity images with their corresponding segmented images consisting of anatomical labels, with the objective of achieving an accurate statistical reference template for a given subject- or age-specific group. MR-based atlases of brains have been established for neonates [131, 132, 133, 134, 135, 136], but it is more challenging to create an accurate age-specific prior for fetal developmental stages due to the fact that the morphological and volumetric changes occur in the order of days or weeks. Nevertheless, some fetal brain MR atlases have been developed [137, 138, 139, 140].

Fetal atlases have progressed towards the development of continuous atlases that parametrically model the changes in size, shape, MR intensity and contrast, and tissue probability of every voxel within the brain tissues as gestation progresses [139, 140, 141]. They are constructed by collecting longitudinal scans from different subjects throughout a range of GAs and constructing a mean growth trajectory that represents the brain changes occurring throughout pregnancy, whilst capturing inter-subject variations. The probabilistic and temporal nature of these atlases is crucial in capturing the dynamic changes in size, shape, and voxel intensity that occur in the developing brain. Recently, Dittrich et al [138] proposed a spatiotemporal latent atlas in which the probabilistic atlas is based on a continuous parameterization of fetal age. Such an approach, if adapted to US, has potential applications in GA estimation from brain structure, which is paramount for making diagnostic decisions in a clinical setting.

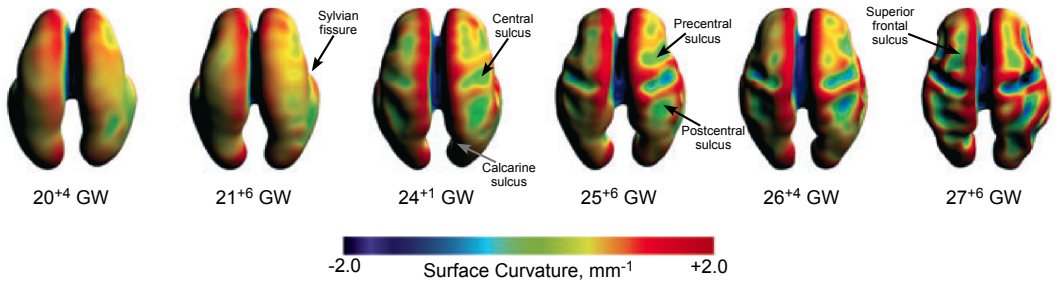


Figure 2.14: **3D maps of cortical surface curvature.** Shown are example cortical surface meshes which were reconstructed from fetal MR images at different GAs. As sulcal folding progresses, the curvature of the cortical plate increases. (Adapted from Figure 7 of the work by Studholme et al [124])

Surface Analysis Regional curvature analysis of cortical surfaces has been achieved by defining a triangulated mesh which defines a layer within the cortical plate (typically the more convoluted inner surface [128]). The tangential differences between neighbouring vectors normal to the triangular isosurfaces yield the principal curvature and direction components from which regional curvatures are quantified [128, 142, 143]. The resulting 3D curvature maps have been used to track local morphological changes on the brain and to study growth trajectories during gestation using MR images (Figure 2.14; [128, 144]), but such a technique may not be fitting to US images due to missing or incomplete visualizations of the cortex. However, the fact that the curvature index can be localized suggests that even with missing parts, one might still be able to approximate local curvatures.

MR studies have also investigated 3D cortical depth using either Euclidean distances [145, 128] or geodesic surface distances [146, 147]. These measure the path length through the sulcus as it extends into the brain, providing a representation of true cortical depth by accounting for circuitous sulcal shapes. In agreement with US reports of sulcal depth, these studies also found that depth increases with GA (as shown in Figure 2.15; [88, 128, 76, 89]), thus proposing cortical depth as a maturational biomarker.

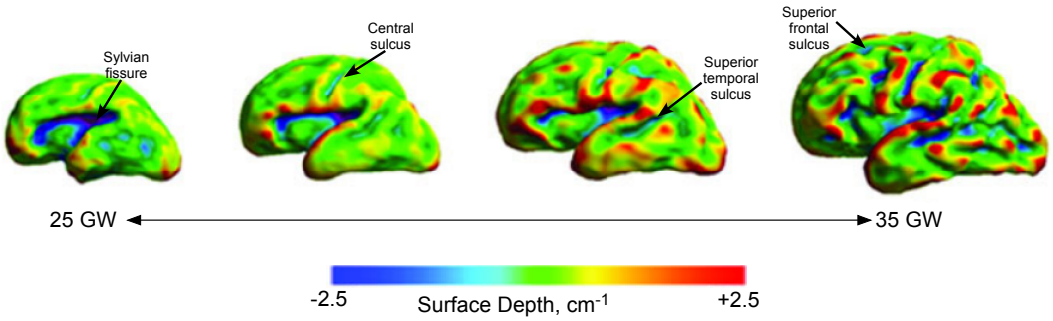
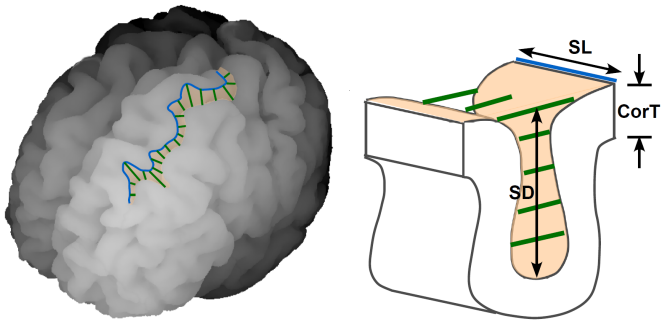


Figure 2.15: **3D maps of cortical surface depth.** Shown are example cortical surface meshes which were reconstructed from fetal MR images at different GAs. As sulcal folding progresses, the depth of the cortical plate increases. (Adapted from Figure 7 of the work by Clouchoux et al [128]).



(a) 3D Sulcal Morphometry (b) Morphometric Measurements

Figure 2.16: **Schematic of 3D sulcal morphometric measurements.** Sulcal length (SL, blue path): the longest path taken by a sulcus. Cortical thickness (CorT): distance between inner and outer cortical surface. Sulcal depth (SD): distance from top to bottom of the sulcus, sampled at a preset number of equidistant points along the length of the sulcus. Sulcal Span (SS, green lines): average width of intra-sulcal space along normal projections on the inner sulcal surface from top to bottom of the sulcus. Sulcal Surface Area (SSA): defined as the largest connected components of negative curvature within a cortical region [91]

.

In addition, the morphometric methods described in Figure 2.16 are also regional descriptors of gyrification, but so far have only been applied to inform on pathological states in MR images of adult brains [92, 93] and gyrogenesis in fetal primates [148], and not yet to the brain of the human fetus.

Indeed, most of the previous work on morphological changes of the cortical surface has been done on MR images, but US remains the modality of choice for perinatal applications despite the limitations associated with image quality. Yet translating these surface-based techniques to US images remains a challenge due to partial occlusions of cortical image features, complicating the necessary task of delineation of the complete cortical contour/surface. Nevertheless, perinatal diagnosis will benefit from the extension or adaptation of MR techniques to inspire bespoke tools which address US-specific challenges. These would enable the quantification of cortical development in US images for direct application to routine clinical practice.

2.5.3 Machine Learning Classifiers for GA Estimation

In recent years, machine learning has played an important role in the field of medical imaging in applications including image registration [149], segmentation [150, 151], and organ detection [152, 153, 154]. Machine learning algorithms for clinical applications are trained on an array of examples that learn from a clinician's prior knowledge about diagnostic analysis of the data [155]. Driven by the fact that medical images contain a wealth of information that is not currently used in the medical field, these algorithms aim to extract relevant and robust biomarkers for a given clinical task, and can achieve higher diagnostic accuracy than current manual methods [156, 123].

The goal of medical image-based prediction is to estimate a clinically-relevant quantity (e.g. GA) directly from an individual scan. A typical approach involves *training data*: a set of images for which the quantity of interest is known. Through the learning process, the prediction model learns its properties from the training data such that it is optimally configured to map the image information to the output quantity. The trained model can then applied

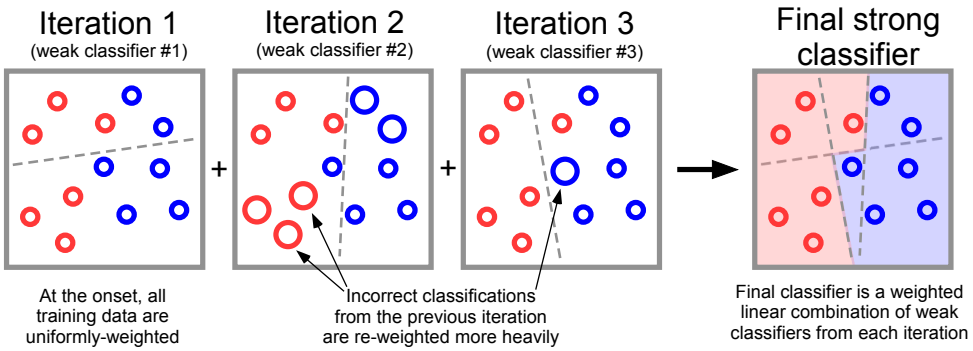


Figure 2.17: **Schematic of AdaBoost algorithm.**

to make predictions on new unseen images.

Learning-based approaches are well-suited for the task of GA estimation due to their ability to take high-dimensional data (i.e. longitudinal images producing 5000+ features representing anatomical appearance at different ages) and establish a compact representation of the image patterns associated with brain development. Though there are numerous machine learning algorithms that have been successfully applied to medical image analysis, this section will highlight the approaches which have been directly applied in the context of GA estimation and making predictions from images of the developing brain.

Object Detection

AdaBoost In the work of Rahmatullah et al [157], an AdaBoost framework was adapted for the task of anatomical object detection in obstetric 2D US images of the fetal abdomen. Their framework was designed to verify whether a 2D US image of the fetal abdomen contains key anatomical landmarks expected to be present in the correct diagnostic image plane (i.e. the stomach and umbilical chord), while remaining robust to size, shape, and position variations of the objects of interest.

AdaBoost (or Adaptive Boosting) is a well-known machine learning algo-

rithm that inherently 'boosts' the classification performance of a simple learning algorithm by combining a weighted sum of weak classification functions (or *weak learners*[2]) to form a stronger classifier [158, 159]. It is adaptive in that, through each iteration, the weighting of subsequent weak learners is modified in favour of the training examples which were misclassified in a previous iteration. This process is summarized in Figure 2.17.

In the early stages of this project, I collaborated with Rahmatullah to extend the object detection framework to fetal US images of the brain [154]. In that work, we developed a method to automatically detect the choroid plexus in 2D US images of the fetal brain (see Figure 2.2a) using the AdaBoost learning algorithm combined with statistical image representations estimated by the Nakagami distribution. Using an empirical model of US image formation to account for the inherent image speckle, we found that the scatter profile of a given tissue can be used as a form of tissue characterization to facilitate the detection task, yielding an accuracy of 87.5%. This result suggests that AdaBoost is capable of determining the absence/presence of other age-specific intracranial structures in US images of the fetal brain, whose timing could be used to predict GA.

Random Decision Forests Random forests have been applied to automatically delineate the myorcardium from 3D [150] and $2D + t$ [151] US images, and for the segmentation of brain tissue from MR images [151, 153]. A random decision forest (or random forest, RF) is a classifier which can be trained for classification or regression tasks [160]. It is a discriminative classifier consisting of a collection of binary decision trees that are constructed by randomly selecting the best feature (from a pre-determined feature set) at each node which will

[2]A learner is dignified as *weak* if its classification accuracy falls slightly below 50%, so it is not expected to achieve the highest classification

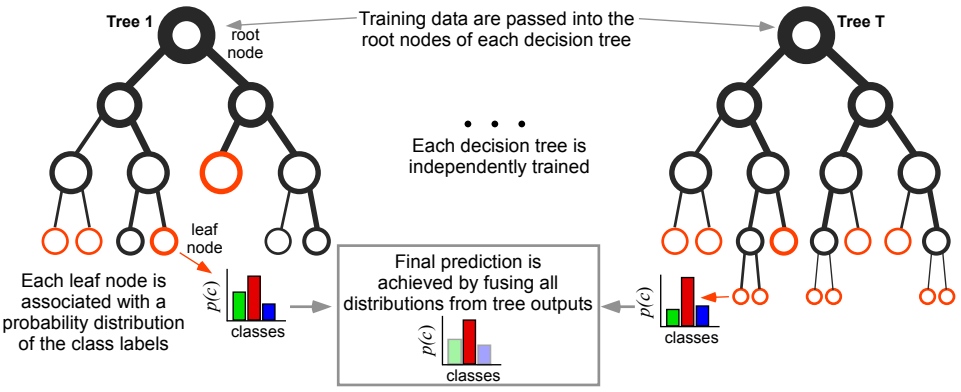


Figure 2.18: **Schematic of Random Forest algorithm.**

optimally partition the data to discriminate between the possible class labels. The data traverse through the nodes in the tree according to a learned binary function of the feature vector, until a leaf (terminal) node is reached. Given an unseen data point, each of the trees casts a vote on the class or label posterior of the input data point, and the final prediction is achieved by fusing all tree outputs (Figure 2.18).

Yaqub et al [161] applied RFs to segment the fetal femur from 3D US images and later to automatically detect local brain structures in 3D fetal US images [110]. By using the RF framework, they were able to correctly localize four cerebral structures (namely the choroid plexus, cavum septum pellucidum, cerebellum, and posterior ventricle cavity). The localization of the structures of interest was based on soft assignment of a label; that is, the probability of a class label being assigned to a particular image voxel. The detection of each structure was achieved by finding the highest probable 3D voxel/neighbourhood within the 3D space, or in every 2D axial slice (mimicking the process performed manually by a clinical expert). They found that training the algorithm on 3D volumes followed by detecting the structures on 2D slices yielded detection accuracies above 90% for all structures in question. This study successfully demonstrated the use of RFs to automatically detect neuroanatomical struc-

tures in US scans.

However, given the discrepancies in reports on the GA at which a structure emerges on the cortical surface (Figure 2.10, discussed in § 2.3) coupled with the inherent biological variation in developmental progression in the fetal cohort, we concluded that although attractive and intuitive, an object-based age categorization approach may not be the most appropriate choice for achieving accurate GA estimations.

Age Regression

Manifold Learning A major challenge in developing predictive models from medical images is related to the "curse of dimensionality". Specifically, medical image data typically contains a large number of voxels, several orders of magnitude higher than the number of images in the training set and the dimensionality of the informative biomarkers which characterize the data. This means that voxel-based predictors need to be strictly controlled in order to avoid *overfitting* to the training data. Thus, a central problem in machine learning and pattern recognition is finding a low-dimensional representation of complex, high-dimensional data (Figure 2.19). Methods that achieve this are collectively referred to as *manifold learning* or *dimensionality reduction* techniques. In particular, manifold learning has been applied to learn a low-dimensional subspace that can then be used to classify unseen images which are mapped to this space.

In medical image analysis, manifold learning has been applied to registration [162], atlas labelling [156], and brain image analysis [163, 164, 165, 156], as well as for diagnostic screening [166]. Relatedly, it has been successfully applied to regress the age of human adults from face images [167], which has potential

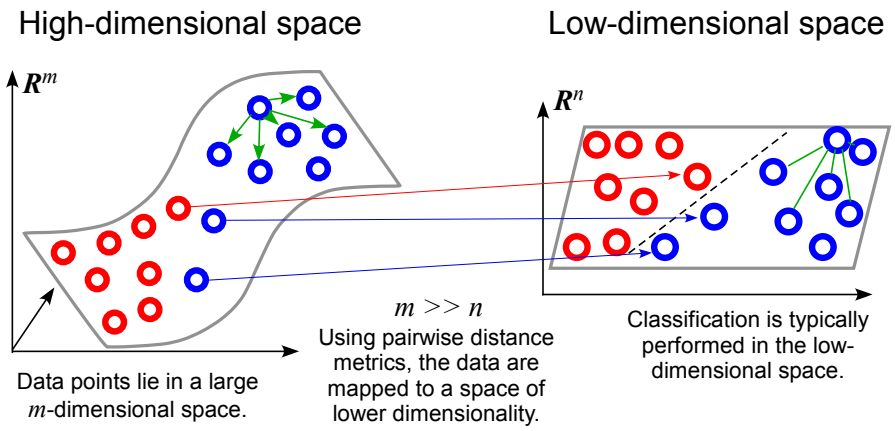


Figure 2.19: **Schematic of Manifold Learning Technique.** An example of dimensionality reduction for classification of data points.

application in GA estimation. In their study, Fu et al [167] collected face images from a broad range of subject ages and learned a low-dimensional discriminative ageing manifold using an Euclidean pairwise correlation metric from which they estimated a statistical age model using linear regression. During testing, an unseen face image was transformed to the low-dimensional ageing feature and mapped to the ageing manifold, where the age label was predicted with high accuracy.

During early fetal and neonatal development, the brain undergoes significant changes at the macroscopic level, such as cortical folding and size changes, which occur in a matter of weeks or days. In their study, Aljabar et al [168] characterized these shape and appearance variations of brain development on MR images of neonates ranging from 29 weeks to 36 weeks. Using an intensity similarity metric and shape descriptors, they captured the degree of morphological deformation at different ages. Each of the measures was learned to produce a manifold, and all manifolds were fused into a single manifold to provide a composite representation of the original data. Their manifold showed significant representational power and consistency with clinical data, partic-

ularly in the association of cortical shape and age at birth which correlates with the rapid changes at this stage of brain development. Wolz et al [169] extended this work to combine imaging information with patient metadata (e.g. gender, age, genotype, and hippocampal volume which is a well-known biomarker for Alzheimer's disease) to assess inter-subject brain variations between healthy control subjects and a group with neurodegenerative impairment. They showed that unifying clinical information with imaging data into a single biomarker yields better classification results than solely using clinical or image-derived biomarkers and has the potential for application to make powerful clinical inferences. Indeed, recent work suggests that manifold learning is a promising technique for GA estimation based on regression [167], representation of subject groups [163, 169], extraction of developmental biomarkers [170, 168], and making inferences by combining imaging data with currently-used clinical data [170, 169]. However, most implementations of dimensionality reduction for medical images have relied on the sensitivity of the pairwise image distance metrics. While these are appropriate for MR images where the brain (i.e. both cerebral hemispheres) is fully represented, they are not directly applicable to US images which are susceptible to partial occlusions and only one visible hemisphere. Thus, in the latter, the pairwise image distances would likely reflect differences in the acquisition protocols as opposed to anatomical structure.

Kernel Methods In the literature, relevance vector machines (RVM) [171] and relevance voxel machines (RVoxM) [172] have demonstrated the feasibility of learning a mapping between image-based biomarkers and pathologies in adult brains. Similar to the Support Vector Machine (SVM) [173], RVMs [174] transform the training data from their input space into a higher-dimensional

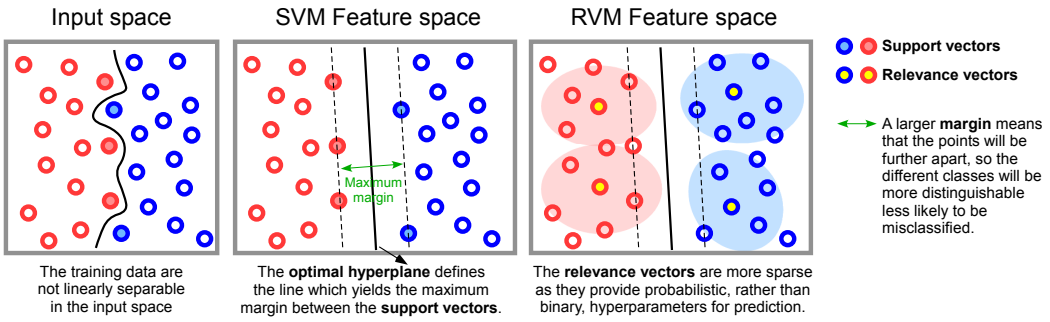


Figure 2.20: **Schematic of Support (SVM) Relevance Vector Machine (RVM) algorithms.** Example kernels method for classification of data points. Both models are sparse in that to determine the optimal separation, they retains only the weights of their support, or relevance, vectors and set the weights of the remaining data points to zero.

feature space in which the separation between subject groups is easier to achieve (Figure 2.20). The hyperplane which best separates the data is selected by maximizing the margin between the two class groups. The key difference between them is that SVMs define the hyperplane on the basis of the difficult-to-classify points located *nearest* to the decision boundary, whereas RVMs are formed by the vectors which are most stereotypical or representative of the two groups, and thus farther away from the decision boundary. For this reason, it has been found that the decision function of the RVM can be sparser than that of the SVM as a result of a lower number of relevance vectors required to describe the groups in the feature space. In the context of image-based regression to age, these kernel methods have been successfully applied to predict adult age from MR images [171, 172]. However, these methods are parametric, which is not ideal when making predictions from a non-linear process such as gyrification, and require the tuning of several parameters to achieve an appropriate separating kernel.

Decision Forests More recently, Konukoglu et al [175] applied neighbourhood approximation forests (NAF)— a variation on random regression forests—

to estimate adult age as a continuous parameter with high accuracy. The advantages of employing regression forests for such tasks are their built-in automatic feature selection, which allows for identification of salient and age-discriminating image features, and their generalizability to images from different age groups and acquisitions. They also aptly handle large training sets, are memory-efficient, fast, and scalable. This suggests that regression forests are an appropriate choice for this work in which I seek to identify the structures which are informative for GA decision-making, and aim to apply the model to images from patients while capturing the variation in sonographic image patterns at different developmental stages.

While the work of Konukoglu et al [175] also presents a forest-based method for predicting age from brain images, their approach relies on accurate alignment and registration of anatomical landmarks, which remains a challenge in US images of the brain. Unlike MR images, the appearance of anatomies in an US image varies with the relative position of the brain with respect to the probe, contributing to acoustic shadows, occluded anatomical features, and reverberation artefacts [80]. Consequently, approaches requiring images of similar intensity appearance and one-to-one inter-subject anatomical correspondence are not, at present, directly applicable to a study of US images of the brain; a local feature-based approach is more appropriate (e.g. [121]).

2.6 Conclusion

US is the imaging modality of choice in prenatal care, and the fact that the spatiotemporal pattern of brain development is directly linked to GA and is observable in US images, opens the possibility of using these data to predict GA *in utero*. However, by virtue of US signal properties and its interactions with

tissues, these images are challenged by attenuation and acoustic shadows, which are unique to US and do not affect MR images. Despite the body of work which exists for assessing brain maturation from MR images, it is inappropriate to apply these techniques to US images. Thus, it is necessary to create US-specific tools to address these challenges to study the brain, and the work presented in this thesis is the first to do this. In this work, we demonstrate the ability of US images to inform on the precise spatiotemporal, though non-linear, evolution of the gyrification process throughout gestation to predict GA from a single US image.

3

Cranial Parametrization of the Fetal Head for 3D Ultrasound Image Analysis

This chapter outlines the image data used in the work presented in this thesis, and provides a detailed description of the source of the images, the criteria for image inclusion, and the image preprocessing steps. In order to develop a predictive model which learns the pattern of fetal brain changes from medical images and to characterize inter-subject anatomical differences, it is first necessary to align the images to a common coordinate system. This chapter presents a bespoke image alignment framework for application to US images

of the brain by parametrizing the cranial surface. The surface parametrization delineates the inner skull boundary and serves as a common coordinate system for efficient sampling of anatomical regions based on cranial position. Early versions of the work presented here have been published in:

Namburete, A. I. L., Stebbing, R. V., Noble, J. A., "Diagnostic plane extraction from 3D parametric surface of the fetal cranium," in *Medical Image Understanding and Analysis (MIUA)*, 2014 — Podium presentation, *Student Bursary Award (awarded to top 6 submitted papers)*

Namburete, A. I. L., Stebbing, R. V., Noble, J. A., "Cranial parametrization of the fetal head for 3D ultrasound image analysis," in *Medical Image Understanding and Analysis (MIUA)*, 2013

3.1 Introduction

The goal of structural neuroimage analysis is to investigate intracranial structures using image information from different subjects and different time points. To achieve this, it is necessary to establish a common coordinate frame between test images in order to perform like-for-like structural comparison of different developmental stages. The typical approach to neuroimage preprocessing involves skull stripping followed by registration, in order to deform the images into a common spatial domain. This inherently relies on the anatomical delineation of internal brain structures within the images.

In developing brains, registration is complicated by absent, underdeveloped, or inconsistent anatomical landmarks for alignment [124]. In particular, analysis of US images is further complicated by increasing calcification of cranial bones which results in the obstruction of the intracranial landmarks necessary for registration. However, regardless of GA, the skull is reliably visualized due

to its echo-bright appearance in comparison to its surrounding tissues. This property of fetal brain US images motivates the need for a method of obtaining image alignment on the basis of a "cranial domain" as opposed to a "voxel domain". To achieve this, we developed a semi-automatic framework to fit a continuous parametric skull surface mesh into each test image. The domain of the surface mesh acts as the cranial domain, allowing image information to be anatomically queried from any subject based on skull position — voiding the need for segmentation or registration of intracranial structures. The lack of direct registration encourages preservation of age-specific sonographic patterns and image differences, which may be indicative of GA.

In our proposed framework, the user provides a rough alignment of the skull surface to the imaged brain (§ 3.3.1 and § 3.3.2). The surface is then minimally deformed to the inner skull boundaries without changing the topology of the surface points (§ 3.3.3). Through validation experiments, the framework is then assessed for its viability in extracting the standard clinical planes used in fetal neurosonography (§ 3.5), with particular emphasis on the ability to extract standard diagnostic planes from the surface domain, verifying the presence of the appropriate anatomical structures within the given plane. We also investigate the ability of the surface to retain anatomical consistency between 2D planes extracted from 3D volumes of different subjects. We compare diagnostic 2D image planes selected by a clinical expert to the planes extracted from the topological surface domain, and evaluate the prospect of using the surface to obtain biometric measurements useful in monitoring fetal growth.

3.2 Image Dataset

Three-dimensional (3D) US images of the fetal brain were obtained from the INTERGROWTH-21st study database, described below. For the work presented in this thesis, images were selected from fetuses ranging from 18^{+0} to 33^{+6} GW, spanning the second and third trimesters of pregnancy: an active period of spatiotemporal sulcal changes visible on the fetal brain surface.

3.2.1 INTERGROWTH-21st Study Database

The International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st[1]) aimed at setting new international standards of fetal and newborn growth with the objective of developing prescriptive standards which define optimal growth in well-nourished, healthy fetuses and its correlation to neonatal outcomes. The project comprised of a collaboration between clinical sites in eight geographically diverse populations. As a prerequisite, the women recruited for the INTERGROWTH-21st study satisfied a set of criteria that define them as having low risk of impairment to fetal growth or diagnosis of fetal abnormality. Exclusion criteria focused on risk factors for intrauterine growth restriction (IUGR, e.g. alcohol use, smoking) and clinical history (e.g. diabetes, chronic illness). Multiple pregnancies were also excluded from the study.

At study entry, a first-trimester scan was collected for the referential crown-rump length (CRL) measurement, along with the last menstrual period (LMP, if known) for GA estimation. The 2D and 3D US image data were obtained using a Philips HD9 linear probe at a transducer frequency of 2-5MHz from over 4000 healthy pregnant women that underwent up to six scanning sessions throughout pregnancy, namely from $<14^{+0}$ GW to term [176].

[1] http://www.intergrowth21.org

Standard biometric measurements of the fetal head (BPD, OFD, HC), the abdomen (TAD, APAD, AC), and femur (FL) were manually measured from the 2D US images (see Figure 2.1), and each measurement was repeated three times. 3D volumetric scans were obtained by scanning *a*) the fetal head at the level of the BPD, *b*) the abdomen at the level of the AC, *c*) and the femur. All 3D fetal head image volumes and clinical metadata (i.e. gestational age, clinical biometry) used in this thesis were extracted from this database [2].

3.2.2 Gestational Age

Central to the topic of this thesis is the definition of the correct baseline GA. In the INTERGROWTH-21st study, GA was defined by the known last menstrual period (LMP) and confirmed by a first-trimester (acquired between 9^{+0} and 13^{+6} GW) ultrasonographic measurement of the fetal crown-rump length. For added confidence, true GA was defined as the LMP- and CRL-based ages agreeing within 7 days, where the CRL-based GA is accurate within 2.7 days, as determined from 3 independent clinical measurements. All women included in the study were screened according to the INTERGROWTH-21st criteria and scanning protocol, and were thus identified as having low risk of impaired fetal growth and absence of fetal brain anomalies. Consequently, it was assumed that there were no delays in cortical development and that the fetal brains included in the study exhibited age-appropriate structural composition.

3.2.3 Image Preparation

Due to the fact that fetal head US images are typically challenged by reverberation artefacts (a consequence of increasing calcification of the cranial bones), the

[2]Ethical approval was provided by the Research Ethics Committee. Reference number: 08/H0606/139

cerebral hemisphere proximal to the ultrasound probe is partially obstructed (see Figure 2.7). Thus, the images selected for this work were included based on the following criteria:

1. Cranium occupies $\geq 50\%$ of the image
2. Distal cranial hemisphere is unobstructed
3. Interhemispheric fissure and falx cerebri are clearly visible in the entire supratentorial region
4. Sylvian fissure is visible in the distal hemisphere
5. Thalami are visible
6. Cavum septum pellucidi are visible

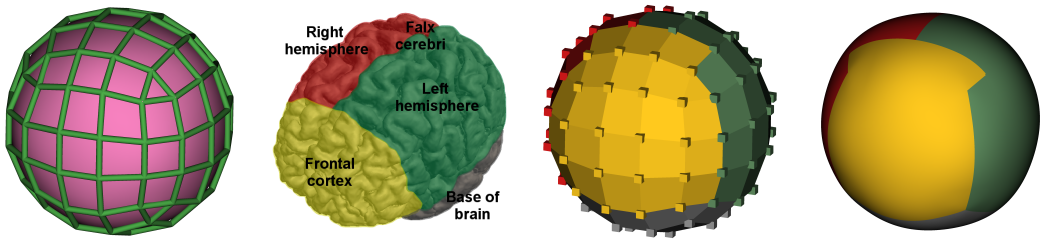
The images were converted to an isotropic spatial resolution whilst preserving anatomical boundaries, speckle, and texture profiles obtained during image acquisition. In order to prepare each image for the experiments presented in Chapters 4, 5, and 6, the cranium of each image was parametrized using the process described in the next section.

3.3 3D Cranial Parametrization

Cranial parametrization is performed by a semi-automated framework which fits a continuous 3D parametric surface into the skull in each US image. The input to the framework is a 3D US image to which an anatomically annotated surface is manually aligned. Though providing coarse inter-subject alignment, the surface does not attempt to achieve one-to-one correspondence between brain regions of different subjects but rather ensures that, when present and

visible, the anatomical regions are approximately aligned for feature extraction. The manually aligned surface serves as the initialization to an energy minimization algorithm which seeks to deform the surface until it finds the best fit to the interior cranial boundaries in the image. During deformation, the surface geometry changes to adhere to the cranial edges, but the topology of the surface points is preserved.

This section describes the cranial parametrization process in preparation for US image analysis of the fetal brain. The surface fitting framework is inspired by "Snakes" [177] and Doo-Sabin surface fitting [178] but differs in three distinct ways. First, surface "fit" is defined as the squared distance between positions on the model surface and a *subset* of detected edge positions in the image. In comparison, "Snakes" define model fit directly against the image and in [178], edge positions are detected at each step when updating the model surface. Second, the coordinates of positions on the surface model — which are used to define the surface fit — are optimized jointly with the surface geometry. In contrast, in "Snakes" and [178] the coordinates are fixed. The joint optimization is inspired by the works of [179] and [180], and although it is more complex, it has two key advantages: (a) selected edge positions are *not* restricted to be perpendicular to the surface model as in [178] and, (b) surface folding is discouraged. Third, in [178], edge positions are detected and selected *independently* for each position on the surface model. The assumption of independence simplifies edge detection and selection, but does not encourage a *coherent* selection of edge positions across the surface. To remedy this, pairwise terms are included here so that the edge positions selected for neighbouring points on the surface model are encouraged to be close.

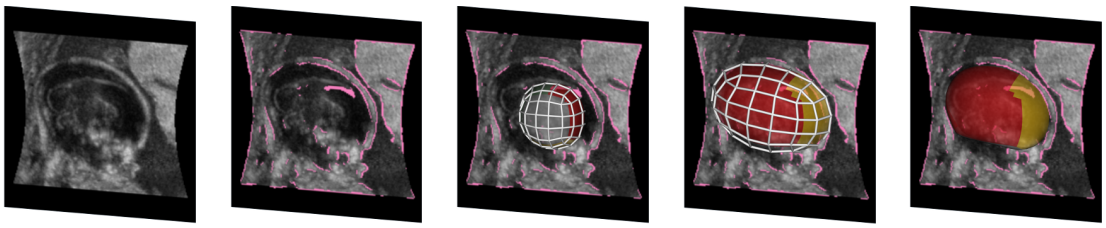


(a) (b) (c) (d)

Figure 3.1: **Cranial surface model.** (a) The control mesh and control vertices ($\mathbf{X}$, green) defines its underlying surface (pink). (b) Given four coarse anatomical regions, (c) the control vertices and faces are colour-annotated with the anatomical regions with which they ought to align in the image during surface initialization. (d) These annotations are propagated down to the underlying cranial surface.

3.3.1 Surface Model

We generated a bespoke spherical biquadratic B-spline surface [181] which can be aligned to anatomical regions within the brain. The control mesh $\mathcal{T}$ explicitly represents the underlying cranial surface (Figure 3.1a, in pink), and its geometry is specified by $N_{\mathbf{X}}$ control vertices, $\mathbf{X} \in \mathbb{R}^{3\times N_{\mathbf{X}}}$ (Figure 3.1a, in green). Each point $\mathbf{p}$ on the cranial surface is parametrized by a surface coordinate $\mathbf{u} \in \Omega$ in the surface domain $\Omega \subset \mathbb{R}^2$ through a mapping function $M\colon \Omega\times\mathbb{R}^{3\times N_{\mathbf{X}}} \to \mathbb{R}^3$ such that $\mathbf{p} = M(\mathbf{u}, \mathbf{X})$. Similarly, the surface normal at $\mathbf{u}$ is defined by $\mathbf{n} = M_\phi(\mathbf{u}, \mathbf{X})$, with mapping function $M_\phi\colon \Omega\times\mathbb{R}^{3\times N_{\mathbf{X}}} \to \mathbb{R}^3$. The cranial control mesh was crafted to be approximately spherical with $N_{\mathbf{X}} = 96$ vertices and 98 faces (Figure 3.1c), which sufficed to describe the degrees of freedom to deform to a realistic skull shape. Exact analytic evaluation of $M(\mathbf{u}, X)$ and $M_\phi(\mathbf{u}, X)$ for *any* control mesh is achieved using Doo-Sabin subdivision [182].



(a) Input (b) Edge Map (c) Default (d) Initialized (e) Deformed

Figure 3.2: **Surface initialization and deformation.** Sagittal view of (a) an example US image of the fetal head with (b) the edge map enhancing ridge-like features, (c) default annotated surface $\mathbf{X}$, (d) user-initialized surface $\mathbf{X}^0$, and (e) 3D rendering of the final deformed surface superimposed on the US image.

3.3.2 Surface Initialization

To initialize the surface control vertices $\mathbf{X}$ to $\mathbf{X}^0$, the user rigidly aligns the default skull surface to the imaged skull/brain using a custom-built multi-view graphical user interface (GUI) programmed using VTK[3]. As shown in Figure 3.2, initialization is achieved by manually *a*) displacing the centre point of the default surface to roughly align with the midline of the brain, and *b*) rotating and anisotropically scaling the surface such that the surface annotations are roughly aligned to their anatomical positions, approximating the cranial dimensions (Figure 3.2d).

To facilitate manual initialization, the faces on the control mesh were colour-annotated with the anatomical regions with which they ought to align in the US image (Figures 3.1b and 3.1c). The colour annotations were based on a coarse parcellation of the brain, including a small number of global landmarks observable in US images which would allow for an approximate anatomical alignment, despite the complications caused by partial occlusions. These anatomical regions include the right and left cerebral hemispheres, the frontal cortex, and the falx cerebri (structure separating the cerebral hemispheres) as displayed in Figure 3.1b. Rigid image alignment using these global landmarks provides in-

[3]http://www.vtk.org

formation about approximate brain size, fetal head pose and orientation within the uterus (e.g. cephalic or breech). The colour annotations on the control faces of $\mathcal{T}$ are propagated down to define colouring of the underlying continuous surface (Figure 3.1d).

3.3.3 Surface Deformation

Determining the ideal deformation for the surface model to fit to the imaged cranial contour first requires selection of relevant boundary candidates in the US image. For the purposes of brain analysis, the ideal candidates lie in the interior skull voxels which we generated using a standard US edge detection technique: Feature Asymmetry (FA), derived from the monogenic signal [183]. The FA image retains important structural information, is contrast-invariant, and allows edge features to be obtained at different scales. In the task of cranial boundary detection, we are interested in the junction between the inner skull boundary and the intracranial soft tissue (e.g. cerebrospinal fluid, subarachnoid space, etc.) which behaves like step-edges or ridge-like structures so we opted for an isotropic log-Gabor filter. Non-maximum suppression was then applied to thin the edges in the FA image. Figure 3.3 shows how cranial image boundaries are extracted by feature asymmetry. It is clear that regardless of GA and image intensity, the inner cranial boundaries are highlighted.

Given an original US image (Figure 3.2a), candidate interior skull positions $\mathbf{C} \in \mathbb{R}^{3\times N_\mathbf{C}}$ and normals $\mathbf{\Phi} \in \mathbb{R}^{3\times N_C}$ are defined from the corresponding FA edge image (Figure 3.3). Thus, given control vertices $\mathbf{X}$, and a matrix $\mathbf{U}$ of $N_\mathbf{U}$ surface points, the energy function defining the fit of the surface to a selection of edge image boundary candidates, $\mathbf{q} \in \mathbb{N}^{N_\mathbf{U}}$, is given by:

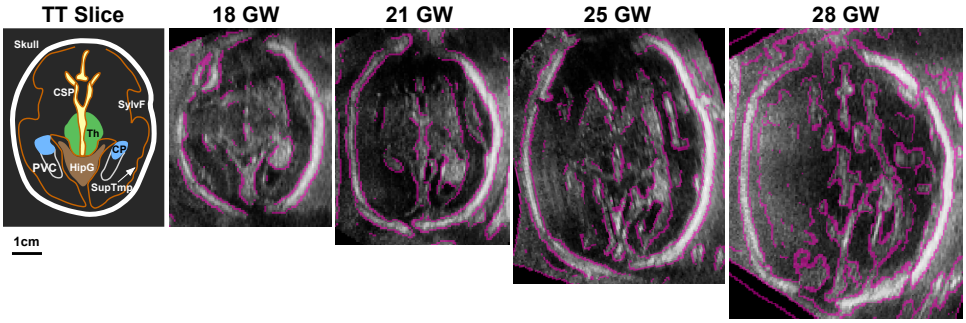


Figure 3.3: **Feature asymmetry edge maps.** Example feature asymmetry edge maps at four different GAs: 18 GW, 21 GW, 25 GW, and 28 GW. Edges of cranial boundaries (in pink) are overlayed on the original US intensity images, shown for TT slices taken at the level of the Sylvian fissure. Kernel of the log-Gabor filter was 4 mm in size.

$$E(\mathbf{q}, \mathbf{U}, \mathbf{X}) = \sum_{i=1}^{N_\mathrm{U}} E_\mathrm{unary}(q_i, \mathbf{u}_i, \mathbf{X}) + \lambda_2 E_\mathrm{pairwise}(\mathbf{q}) + \lambda_3 E_\mathrm{user}(\mathbf{X}) + \lambda_4 E_\mathrm{reg}(\mathbf{X}) \tag{3.1}$$

where $\lambda_2, \lambda_3, \lambda_4$ influence the weight of each term.

Specifically, E_unary quantifies the mismatch between each surface point $\mathbf{u}_i$ and its corresponding boundary point at q_i in terms of position and orientation.

$$E_\mathrm{unary}(q_i, \mathbf{u}_i, \mathbf{X}) = \|\mathbf{c}_{q_i} - M(\mathbf{u}_i, \mathbf{X})\|^2 + \lambda_1 \|\boldsymbol{\phi}_{q_i} - M_\phi(\mathbf{u}_i, \mathbf{X})\|^2 \tag{3.2}$$

where λ_1 penalizes the orientation of the vectors between the surface and boundary points. Note that boundary candidates are not assumed to be perpendicular to the surface model.

Realizing that by virtue of anatomical design, the boundary points are spatially correlated, so the E_pairwise term models the fact that neighbouring surface points should also prefer boundary points which are close (defined as

the weighted sum of position and orientation errors) and spatially coherent.

$$E_{\text{pairwise}}(\mathbf{q}) = \sum_{(i,j)\in\mathcal{N}} \left\|\mathbf{c}_{q_i} - \mathbf{c}_{q_j}\right\|^2 \tag{3.3}$$

where $\mathcal{N}$ is the set of edges over the surface points.

Given that the surface model is manually initialized by an individual who provides approximate anatomical surface orientation and head size, it is desirable that the surface deformation does not deviate too far from its original surface placement. Consequently, E_{user} encourages minimum deformation from the user initialization, and also removes the problem of finding multiple local minima that may arise from the geometric symmetry of the near-ellipsoidal shape of the fetal skull.

$$E_{\text{user}}(\mathbf{X}) = \sum_{i=1}^{N_{\mathbf{X}}} \left\|\mathbf{x}_i - \mathbf{x}_i^0\right\|^2 \tag{3.4}$$

Finally, the regularization term E_{reg} encourages a smooth cranial surface by penalizing large displacements between the surface control vertices.

$$E_{\text{reg}}(\mathbf{X}) = \sum_{(i,j)\in\mathcal{T}} \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2 \tag{3.5}$$

While Equations 3.3 and 3.5 appear similar, they serve different purposes: E_{pairwise} regularizes the selection of boundary points and is a function over the vector of *discrete* boundary point indices; E_{reg} regularizes the surface model and is a function over the matrix of *continuous* model control point positions.

Equation 3.1 is minimized through a series of alternating discrete and continuous optimization steps. By taking the user initialization of $\mathbf{X}$ to $\mathbf{X}^0$ and generating $\mathbf{U}$ to a regular sampling of Ω, we use belief propagation with a subset of edges in E_{pairwise} to solve an approximate $\mathbf{q}$ which is then refined using

Quadratic Pseudo-Boolean Optimisation (`QPBO`) [184]. Next, given $\mathbf{q}$, Equation 3.1 is minimized *jointly* with respect to $\mathbf{X}$ and $\mathbf{U}$ using the Levenberg-Marquardt algorithm [185].

Importantly, as a result of the convexity of the surface, E_{reg}, and due to the *joint* optimisation of $\mathbf{X}$ and $\mathbf{U}$[4], surface self-intersection and stretching is strongly discouraged and was not seen in practice. The resulting surface is an anatomically realistic representation of a 'complete' fetal skull (i.e. filling the gaps of fontanelles and cranial occlusions).

3.4 Qualitative Assessment of Cranial Parametrization

For demonstration purposes, the cranial deformation model framework was applied to 52 randomly-selected 3D US images of the head from healthy fetuses at 22 GW. Each image was typically of dimensions $215 \times 230 \times 151$ voxels with an isotropic resolution of $0.6 \times 0.6 \times 0.6\text{mm}^3$. A surface mesh was manually initialized into each individual image, taking an average of **1.44 minutes** per image. Solving time took approximately **2 - 3 minutes**, dominated by the two discrete optimization steps (Figure 3.4). Model parameters of $\Lambda = (8.0, 3.0, 0.25, 1.0)$, and $N_{\mathbf{U}} = 4704$ were empirically selected and used for all test images. Small changes to Λ did not result in drastically different recovered surfaces, by visual inspection.

3.4.1 Cranial Deformation

The framework updates the control mesh geometry such that the underlying surface matches the cranial boundary. This is evidenced by Figure 3.6 in which

[4]Standard "Snakes" fixes $\mathbf{U}$.

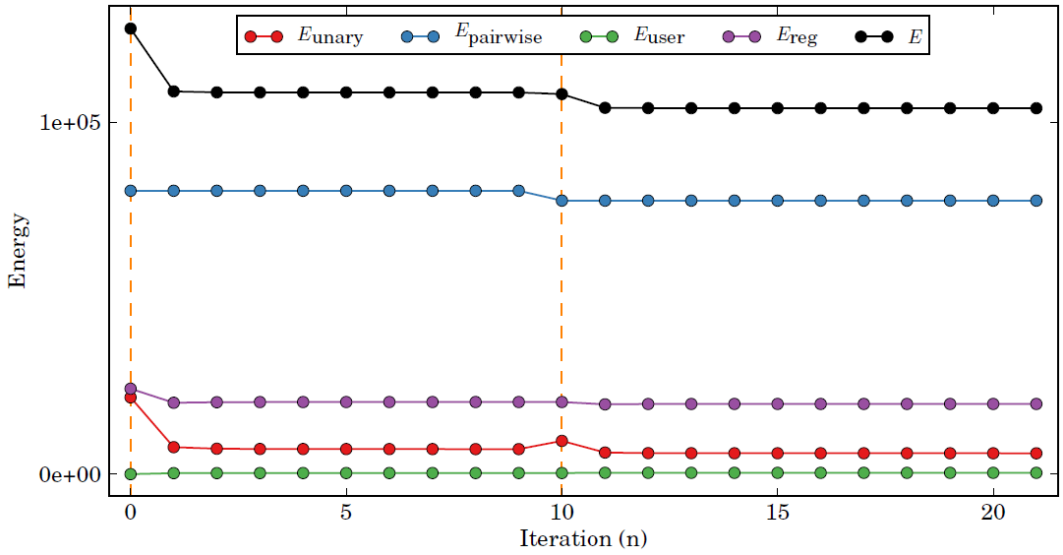


Figure 3.4: **Example convergence of energy minimization.** Each iteration of discrete optimization ($n = 0; 10$; orange lines) solves for $\mathbf{q}$, affecting only E_{unary} and E_{pairwise}. Subsequent continuous optimization steps minimise $\mathbf{U}$ and $\mathbf{X}$ jointly until convergence, affecting all energies except E_{pairwise}. Note the increase in E_{unary} at $n = 10$ is accompanied with a larger decrease in E_{pairwise}.

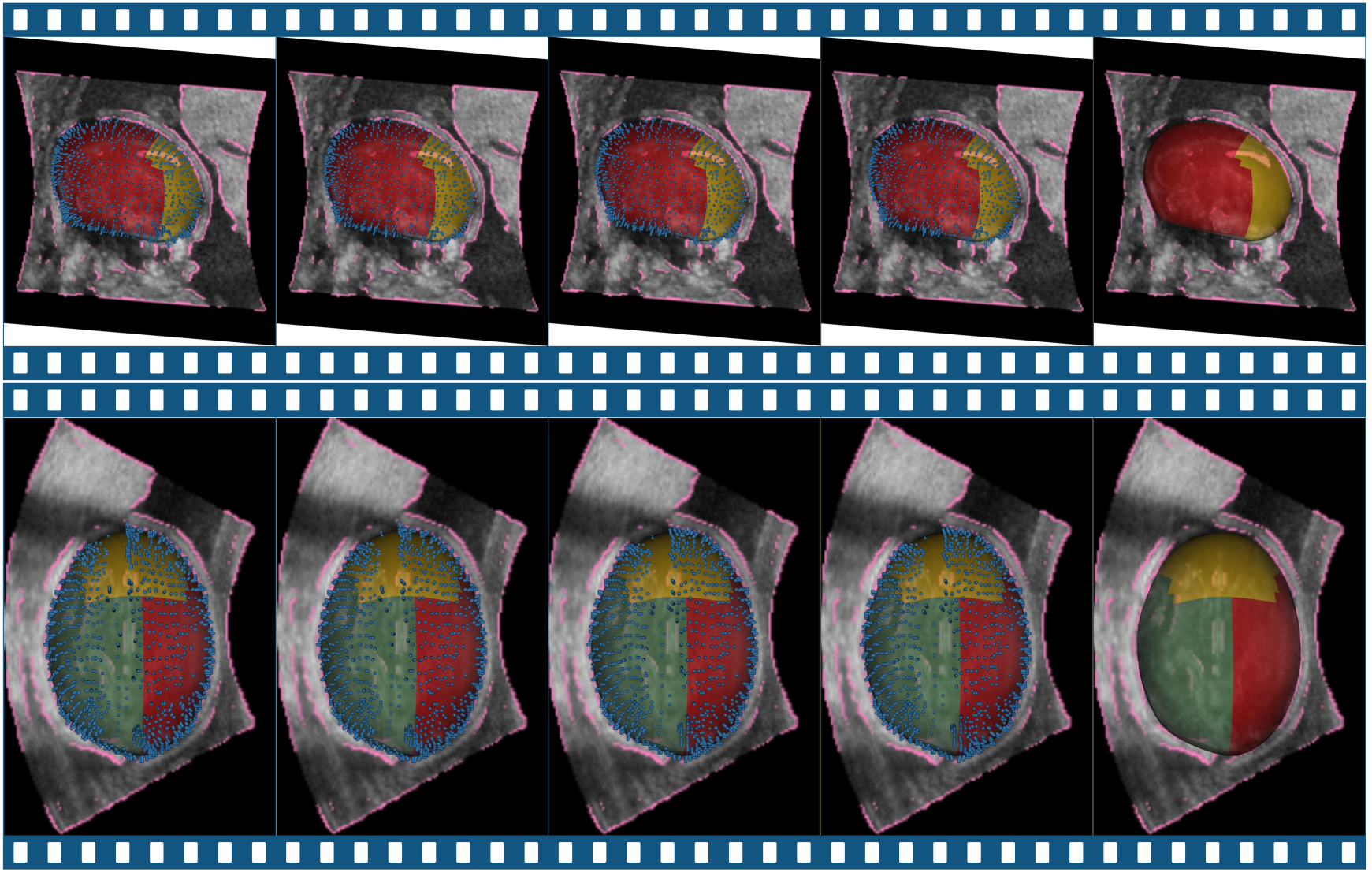


Figure 3.5: **Surface deformation steps.** Coronal (top) and axial (bottom) views of cranial surface deformation. The process occurs is a sequence of discrete (Iterations 1 and 10) and continuous (Iterations 5 and 15) optimization steps. Surface point ($\mathbf{u}$) correspondences (in blue) to cranial boundary candidates ($\mathbf{c}_q$, in pink) are also shown.

the deformed surfaces are displayed on orthogonal image slices for four of the 52 examples, showing the variability of fetal head pose and shape. Each example highlights that the surface is capable of deforming such that it closely adheres to the inner skull boundaries, respecting the skull's non-ellipsoidal shape.

3.4.2 Anatomical Consistency

The deformation process modifies the geometry of the control mesh but preserves topology. To demonstrate this, we first specified cutting planes in the cranial domain. Next, using the deformed surfaces in each image we evaluated these cutting planes in image coordinates (Figure 3.7). It is evident that the same intracranial structures are visible within the different images, with consistent anatomical positioning. The surfaces provide a cranial parametrization which retains anatomical consistency between images, voiding the need for transformation of the images into a common image voxel domain. However, the anatomical positioning is reliant on a correct anatomical alignment provided by the user initialization. In the next section, we investigate the sensitivity of the surface-based alignment to the surface initialisation provided by the user.

3.5 Clinical Plane Extraction

In a clinical setting, one of the main purposes of fetal brain imaging is to identify structural abnormalities at a stage that is sufficiently early in the pregnancy to be confident of prognosis and thereby enable parents to make informed choices relating to their unborn child. During this process, qualitative and quantitative image data from an individual (fetus) are compared to reference standards of normal development to enable a judgement to be made about the likelihood of an abnormality being present. Biometry of the cranium is an obvious starting

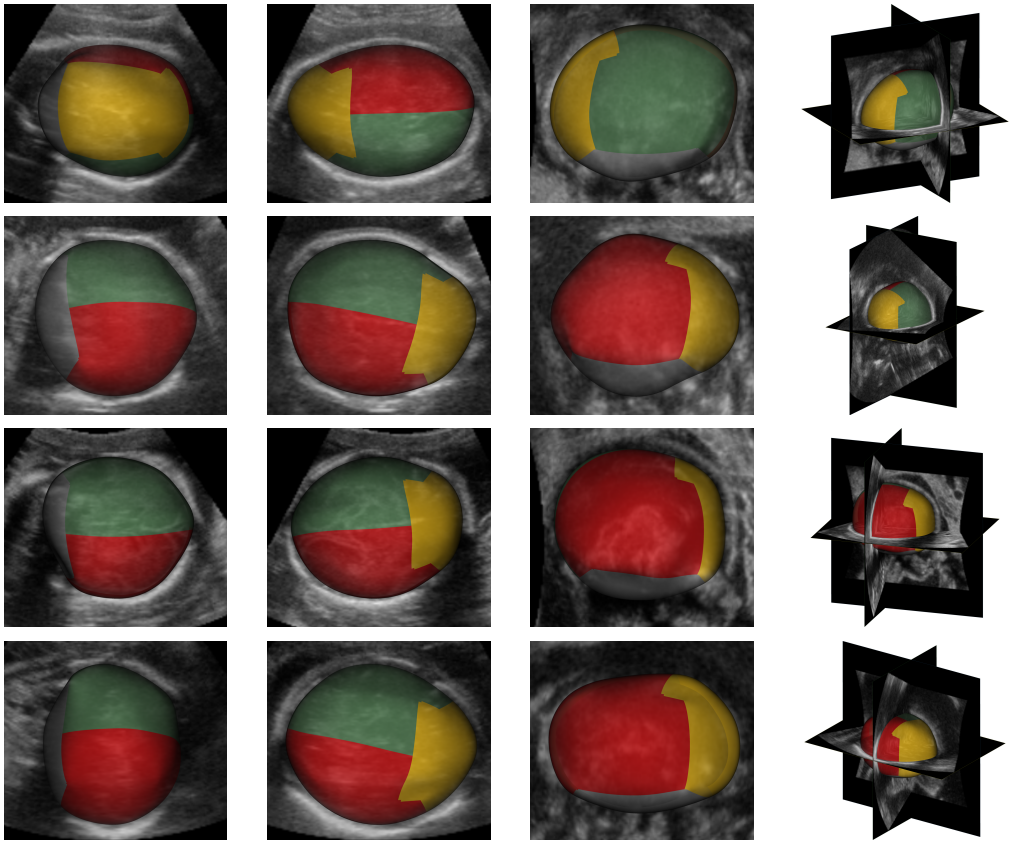


Figure 3.6: **Deformed cranial surface examples.** Four example cranial deformations for the orthogonal image acquisition slices: coronal (first column), transverse (second column), and axial (third column). The resulting deformed surface (right) is shown for each example.

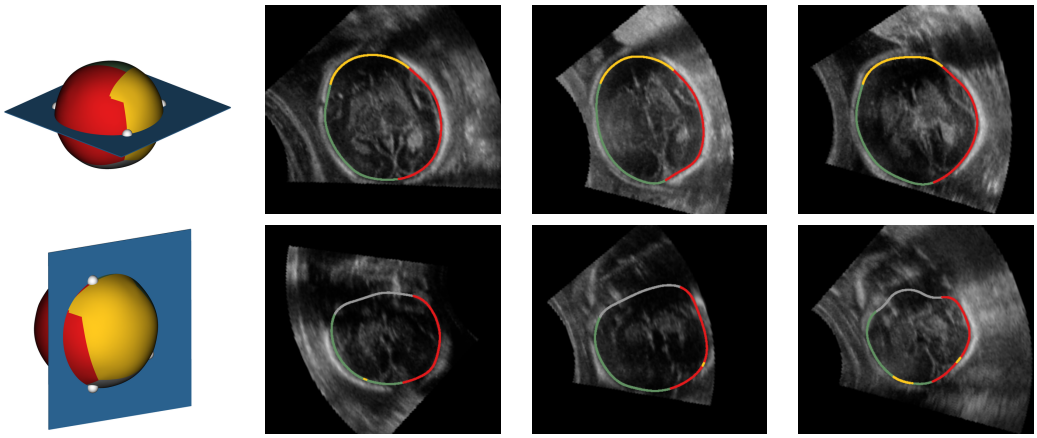


Figure 3.7: **Slice extraction.** Axial (first row) and coronal (second row) slices extracted from deformed surfaces of six different patients at planes defined by 3 points (white). Colour annotations display the anatomical consistency between different slices.

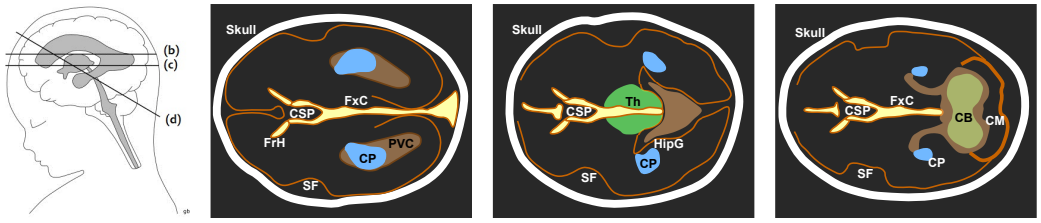
point for such examinations due to its echo-bright appearance on ultrasound. Whilst not directly measuring the brain itself, cranial size is used as a proxy for this and is incorporated into most assessments of brain structure and fetal growth. The International Society of Ultrasound in Obstetrics & Gynaecology (ISUOG) provides specific guidance for US of the central nervous system, outlining a basic screening examination for 'low-risk' patients and a more detailed 'neurosonogram' for complex cases [17]. During the basic examination, the 3 standard axial planes are collected from the fetal brain to assess anatomic integrity (Figure 3.8):

1. The transventricular (TV) plane (Figure 3.8b) includes both horns of the lateral ventricles separated by the cavum of the septum pellucidum (CSP) anteriorly, with the ventricular horns in the posterior portion of the plane. It is used to assess structural integrity and size of the ventricles.

2. The transthalamic (TT) plane (Figure 3.8c) is commonly added specifically to measure the biometry upon which gestational age and fetal growth are assessed. The plane includes the CSP, both bodies of the thalamus in the midline at a level above the cerebellum— a structure which should not be visualised in the plane.

3. The transcerebellar (TC) plane (Figure 3.8d), acquired at a lower level with a posterior 'tilted angle', includes the frontal horns of the anterior ventricles, CSP, cerebellum, and thalami. This plane is used to assess structure of the posterior fossa and aid the diagnosis of neural tube defects.

One of the advantages of a parametrized surface is that it allows for fast and efficient sampling of anatomically-corresponding regions and the extraction of 2D planes from a 3D volume, without requiring registration between



(a) (b) TV Plane (c) TT Plane (d) TC Plane

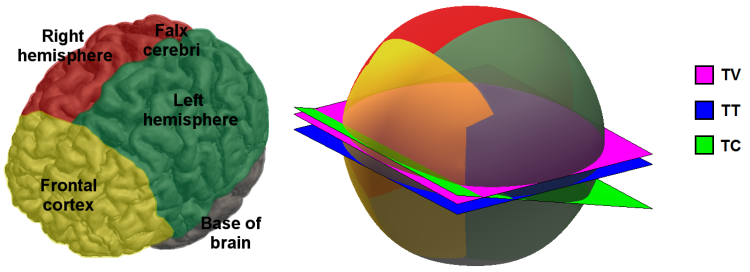
Figure 3.8: **Standard axial planes of fetal head.** (a) Three axial planes used in the standard fetal neurosonography examination (from left to right): (b) transventricular (TV) plane, (c) transthalamic (TT) plane, (d) transcerebellar (TC) plane. The structures of interest in these planes are: choroid plexus (CP), posterior ventricle cavity (PVC), cavum septum pellucidum (CSP), falx cerebri (FxC), Sylvian fissure (SF), frontal horn (FrH), thalami (Th), hippocampal gyrus (HipG), cisterna magna (CM), and cerebellum (CB).

the images, hence preserving age-specific image patterns. In this section, the viability of the surface at extracting the standard fetal head planes through parametrized planes whilst maintaining anatomical consistency between images of different subjects is tested, as well as the accuracy expected of the initial surface alignment provided by the user.

3.5.1 Defining Standard Planes on Cranial Domain

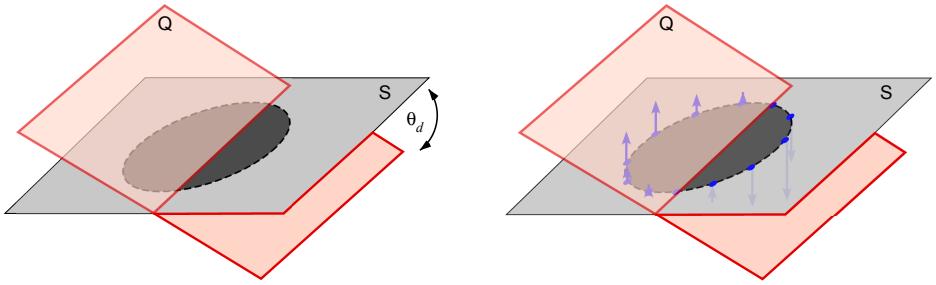
The three standard diagnostic planes were defined on the cranial surface model (Figure 3.9b) based on the regions they ought to intersect on the anatomically-labelled surface (Figure 3.9a). The standard planes were defined on the cranial domain by selecting a single reference 3D fetal brain image and identifying 3 cranial points on the Euclidean image space of each standard plane, $\mathbf{B} \in \mathbb{R}^{3\times3}$. The points were then used to define the center point and normal vector to each standard plane on the reference image, $\mathbf{Y}^j \in \mathbb{R}^{3\times2}$ where $j \in \{\text{TV, TT, TC}\}$.

The fetal cranial surface then underwent the surface deformation process described in § 3.3, and the surface points $\mathbf{U}' \in N_{\mathbf{U}}$ intersected by the standard image plane $\mathbf{Y}^j$ were used to define the standard plane j on the cranial domain.



(a) (b)

Figure 3.9: **Standard axial planes on the cranial surface domain.** (a) Given the anatomical regions corresponding to vertices on the cranial surface model, the standard planes can be defined on the surface (b): tranventricular plane (magenta), transthalamic plane (blue), transcerebellar plane (green).



(a) Dihedral Angle (b) Mean Plane Distance

Figure 3.10: **Plane similarity validation metrics**

Figure 3.9b displays the spatial definition of the standard planes on the cranial surface.

3.6 Validation Experiments

The dihedral angle between the manually-selected ($\mathbf{Y}^Q$) and surface-extracted ($\mathbf{Y}^S$) clinical planes was computed as a measure of the alignment between the planes (Figure 3.10a). The dihedral angle, $\theta_d\left(\mathbf{Y}^Q, \mathbf{Y}^S\right)$, is given by:

$$\theta_d\left(\mathbf{Y}^Q, \mathbf{Y}^S\right) = \arccos\left(\frac{\mathbf{n}_Q \cdot \mathbf{n}_S}{|\mathbf{n}_Q|\,|\mathbf{n}_S|}\right) \tag{3.6}$$

where $\mathbf{n}_Q$ and $\mathbf{n}_S$ are the normal vectors to the manually-selected and surface-extracted planes, respectively, and the norm of a vector is defined as $|\mathbf{n}| = \sqrt{n_x^2 + n_y^2 + n_z^2}$.

It should be noted, however, that although the dihedral angle is indicative of the alignment between planes, it does not translate into a clinical index. More clinically-informative indicators are: *a*) the relationship between the anatomical structures of interest observable in both the surface-extracted and manually-selected planes, and *b*) the geometric distance between the $\mathbf{Y}^Q$ and $\mathbf{Y}^S$ planes. For instance, a good surface-extracted TC plane would demonstrate high overlap in the cerebellar region visualized in the manually-selected plane and small average distance (in mm) from the manually-selected plane. Anatomical structure overlap is also an appropriate measure because the ultrasonographer selects a suitable plane from the 3D brain space on the basis of presence, shape, and size of the structures of interest.

The overlap between the planes was computed using the Dice Coefficient of the manual segmentations of the following anatomical structures: choroid plexus (CP), posterior ventricle cavity (PVC), cavum septum pellucidum (CSP), and cerebellum (CB). The Dice coefficient of an anatomical structure, D_k, is a metric indicative of plane similarity, and was computed by

$$Dice_k = \frac{2\left|N_k\left(\mathbf{Y}^Q\right) \cap N_k\left(\mathbf{Y}^S\right)\right|}{\left|N_k\left(\mathbf{Y}^Q\right)\right| + \left|N_k\left(\mathbf{Y}^S\right)\right|} \tag{3.7}$$

where $k \in \{\text{CP, PVC, CSP, CB}\}$ and the $N_k\left(\mathbf{Y}^j\right)$ is the number of pixels of structure k present in a plane j. Since the planes cannot be assumed to be parallel, it is expected that there may be a shift in the location of corresponding intracranial structures. To account for this, the 2D intensity image planes were first rigidly registered and the transformation was applied to the manual segmentation images prior to computing the Dice coefficient.

The mean distance between $\mathbf{Y}^Q$ and $\mathbf{Y}^S$ planes was computed by extracting the points on the cranial surface which define the contour of the skull in the surface-extracted plane and computing the average Euclidean distance between these points and the manually-extracted plane, as shown in Figure 3.10b. The cranial contour points $\mathbf{U}^S$ are defined as the surface points which overlap with the surface-extracted plane, $\mathbf{Y}^S$, where $N_{\mathbf{U}^S} < N_{\mathbf{U}}$. The Euclidean distance between the planes is defined as the average distance between the surface points overlapping with $\mathbf{Y}^S$ (i.e. $\mathbf{u}_k^S$, for $k \in \{1, \ldots, N_{\mathbf{U}^S}\}$) and the manually-extracted plane, $\mathbf{Y}^Q$:

$$d(\mathbf{Y}^Q, \mathbf{Y}^S) = \frac{1}{N_{\mathbf{U}^S}} \sum_{k=1}^{N_{\mathbf{U}^S}} \hat{\mathbf{n}}_Q \cdot \mathbf{u}_k^S + d_0^Q \tag{3.8}$$

where $\hat{\mathbf{n}_Q}$ is the unit normal vector of the manually-selected plane, $\mathbf{Y}^{\mathrm{Q}}$, and d_0^j is the distance of plane j from the origin.

3.6.1 Plane Selection Results

The mean and standard deviation of the dihedral angles between the manually-selected and surface-extracted planes are summarized in Table 3.1, and Figures 3.11, 3.12, and 3.13 display visual examples of surface-extracted plane selection along with their corresponding dihedral angles. It is evident that in all image slices, the relevant anatomical structures were present in the selected planes, regardless of the dihedral angle.

The dihedral angle between the manually-selected and surface-extracted planes was smallest and least variable for the TT plane: the plane containing the most anatomical landmarks as per the ISUOG guidelines. In contrast, the TC plane displayed the highest variability in plane selection due to the imprecise definition of the angle of measurement and anatomical landmarks expected

to be visible in the plane (Table 3.2). The results tabulated in Table 3.2 exclude 7.69% (4/52 images) of the images which were considered outliers (i.e. $|\theta_d| \geq 25°$). These outliers constituted images satisfying any combination of the following criteria:

1. Clinician selected an incorrect plane
2. Cranial surface was misaligned during the user initialization
3. Intracranial structures were indiscernible

The mean Euclidean distances between the manually-selected and surface-extracted planes were all below 5 mm for all standard planes accounting for approximately 5% of the height of the supratentorial region. This indicates that the surface-extracted planes did not greatly deviate from the planes selected by a clinician for fetal brain assessment.

Table 3.2 summarizes the segmentation overlap between the manually-selected and surface-extracted planes. Low Dice coefficient values are attributable to the difficulty in achieving proper alignment between non-parallel Q and S planes, undefined manual segmentation protocol for each anatomical structure, and the small size of the intracranial structures on each plane. However, the fact that there is overlap between the structures expected at each plane signifies that the surface-extracted planes contain the appropriate intracranial structures of relevance to clinicians. This highlights that although a surface extracted-plane may not exactly match the clinically-selected plane, it still complies with the ISUOG guidelines [17], meaning that it would be an acceptable choice of plane in a fetal neurosonography evaluation. The results also validate the parametrized surface's ability to extract planes with anatomical consistency in different subjects.

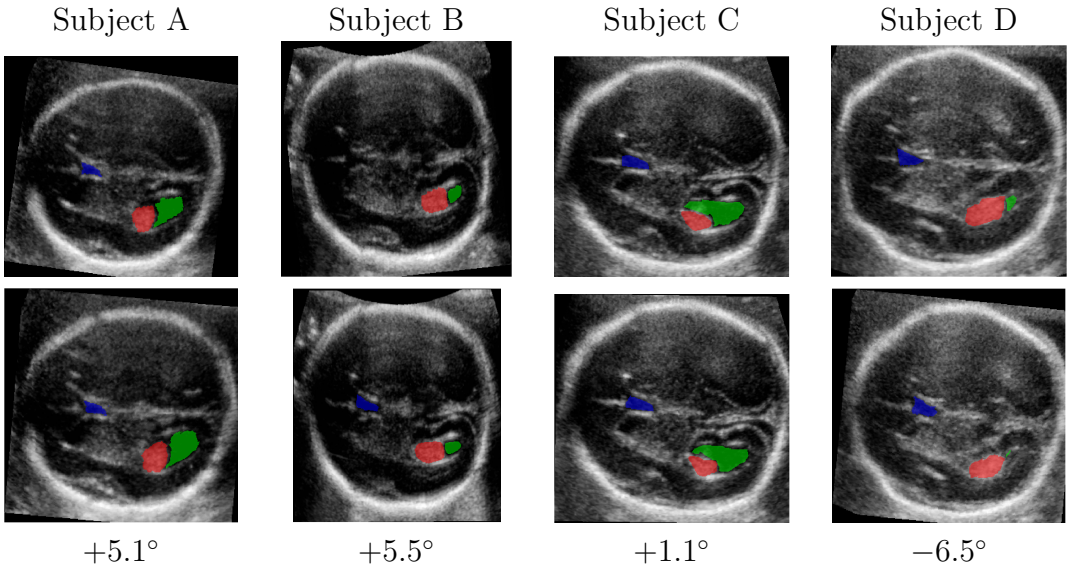


Figure 3.11: **Transventricular (TV) plane selection**. Visual comparison between manually-selected (top row) and surface-extracted TV planes (bottom row) for four different subjects. Manual segmentations of anatomical landmarks are shown: choroid plexus (red), posterior ventricle cavity (green), cavum septum pellucium (blue). Dihedral angle between manually-selected and surface-extracted planes is shown below each image pair.

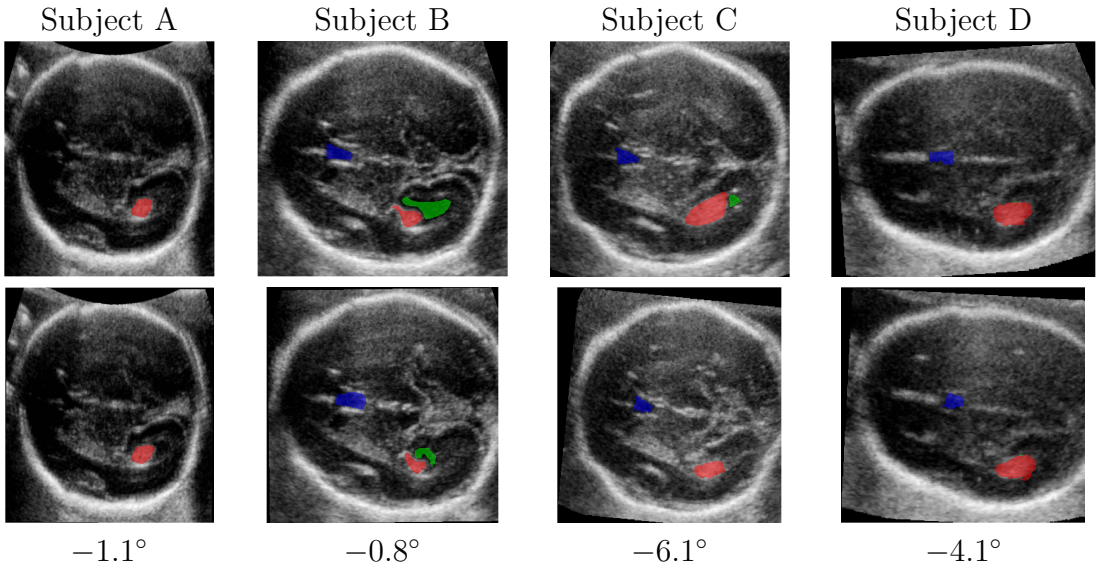


Figure 3.12: **Transthalamic (TT) plane selection**. Visual comparison between manually-selected (top row) and surface-extracted TT planes (bottom row) for four different subjects. Manual segmentations of anatomical landmarks are shown: choroid plexus (red), posterior ventricle cavity (green), cavum septum pellucium (blue). Dihedral angle between manually-selected and surface-extracted planes is shown below each image pair.

Table 3.1: **Plane geometry similarity metrics.** Mean and standard deviation values of dihedral angles between manually-selected and surface-extracted planes ($\theta_d(\mathbf{Y}^Q, \mathbf{Y}^S)$ in degrees), and mean and standard deviation distance values between the planes ($d(\mathbf{Y}^Q, \mathbf{Y}^S)$ in mm).

Plane	n	$\theta_d(\mathbf{Y}^Q, \mathbf{Y}^S)$	$d(\mathbf{Y}^Q, \mathbf{Y}^S)$
TV	50	1.5 ± 8.4	4.31 ± 1.79
TT	51	0.8 ± 6.6	3.55 ± 1.58
TC	48	2.6 ± 7.3	4.63 ± 1.91

Table 3.2: **Plane overlap metrics.** Dice Coefficient for anatomical structures of interest ($Dice_k$). Blank table entries correspond to structures that were consistently absent from the plane.

Plane	$Dice_{\text{CP}}$	$Dice_{\text{PVC}}$	$Dice_{\text{CSP}}$	$Dice_{\text{CB}}$
TV	0.509 ± 0.185	0.334 ± 0.357	0.195 ± 0.265	—
TT	0.285 ± 0.314	0.173 ± 0.349	0.333 ± 0.387	—
TC	—	—	0.136 ± 0.295	0.097 ± 0.156

3.6.2 Extraction of Clinical Measurements from Surface Domain

Having successfully recovered acceptable diagnostic planes using the parametric surface domain, we investigated the ability to use these to extract clinically-relevant measurements, namely head circumference (HC). Figure 3.14 shows a comparison between a HC measurement obtained from a surface-extracted TT plane (Figure 3.14a) and the best standard clinical method: the mean of three HC measurements from independent scans of the same fetus, as shown in Figure 3.14b. The difference between the two measurements demonstrates a tight fit ($r = 0.992; \sigma = \pm 28.9\,\text{mm}, 12.6\%$), and no significant differences ($p < 0.0001$). The bias error is within the acceptable range of interobserver variability for HC measurements ($-11.1\,\text{mm}, 4.8\%$), i.e. below 5% [186]. Note,

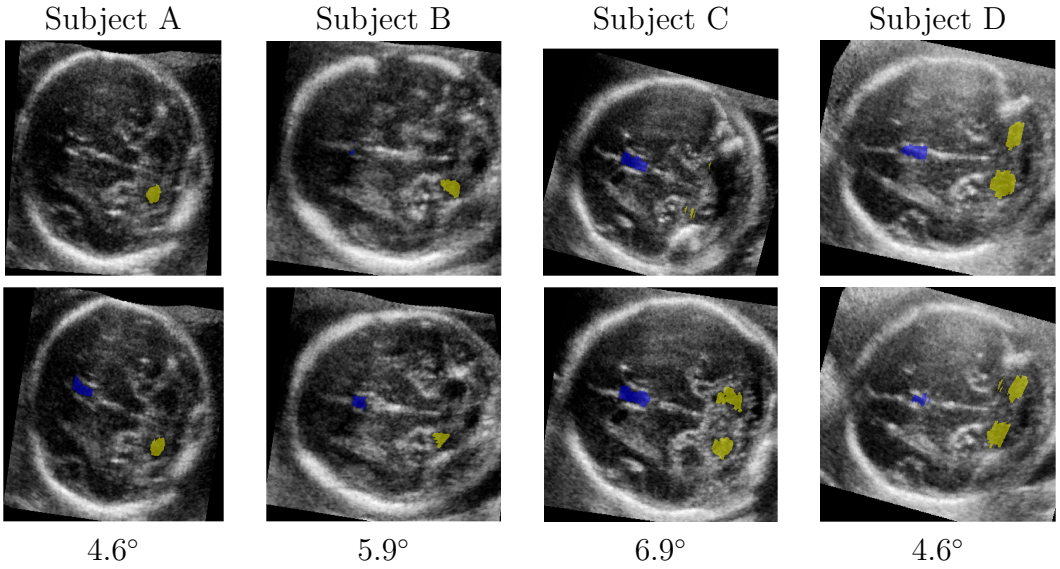
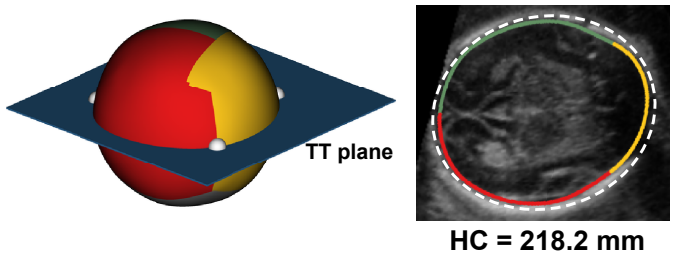


Figure 3.13: **Transcerebellar (TC) plane selection**. Visual comparison between manually-selected (top row) and surface-extracted TC planes (bottom row) for four different subjects. Manual segmentations of anatomical landmarks are shown: cerebellum (yellow), cavum septum pellucium (blue). Dihedral angle between manually-selected and surface-extracted planes is shown below each image pair.

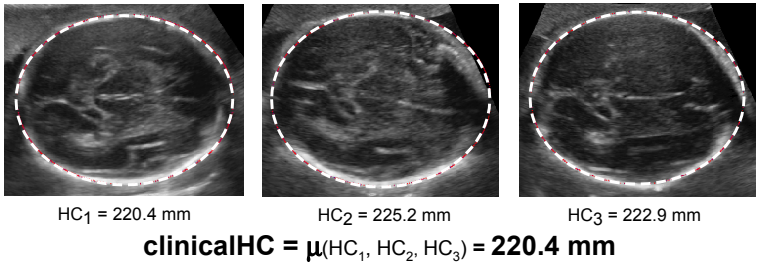
however, that the bias error reflects the fact that the surface-extracted HC is a measurement of the *inner* cranial contour as opposed to the outer contour, so the value of 11 mm may correlate with skull thickness rather than measurement error. This result indicates the clinical viability of the surface-extracted plane measurements.

3.7 Conclusion

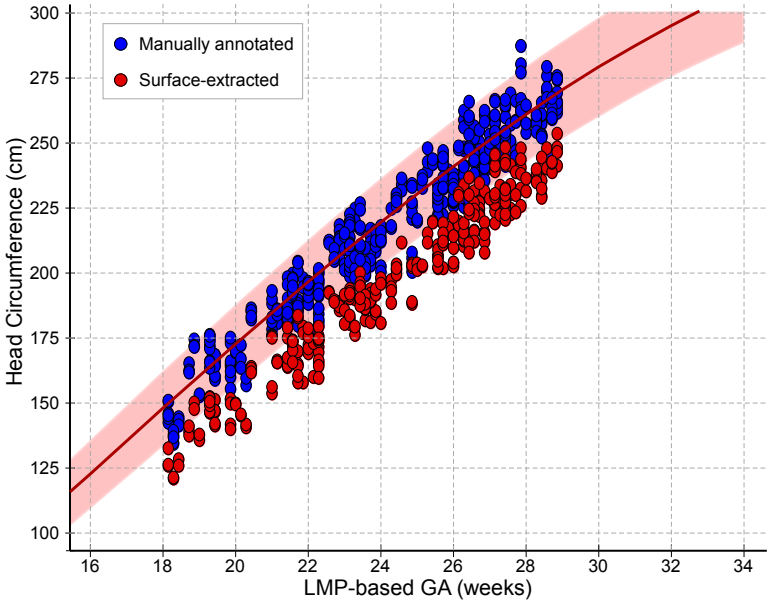
The framework presented in this chapter fits a parametric surface into 3D fetal head US scans. It relies on the user to provide a prior for each model surface which is then deformed to fit the interior cranial boundary. The method recovers detailed structure of the skull and anatomically consistent skull surfaces. However, fine-grained alignment is still desirable and the results show that this should be possible by modelling geometric *and* appearance similarity between



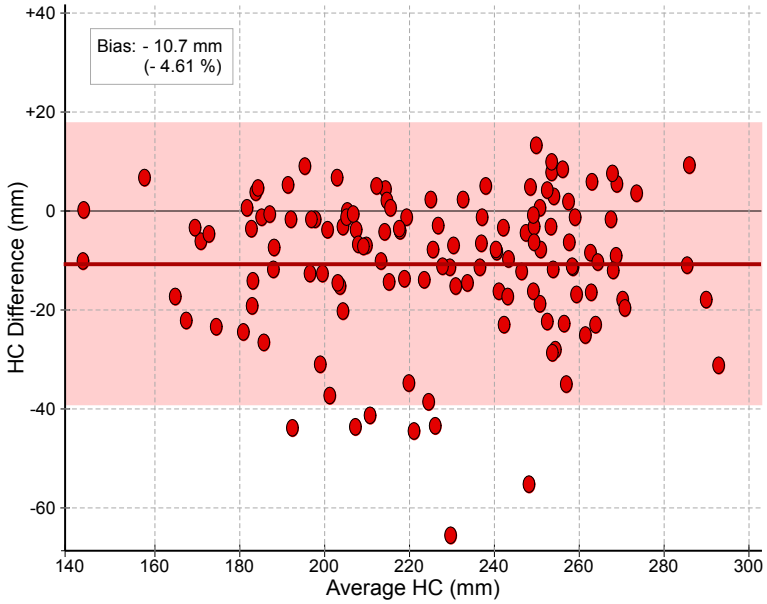
(a) Surface-extracted HC measurement



(b) Clinical manual HC measurement



(c)



(d)

Figure 3.14: **Surface-based HC measurement.** Methods of obtaining head circumference measurement from (a) surface-extracted and (b) manual-selected TT planes. For (b), HC was estimated by the mean of clinical annotations from three different images of the same subject. (c) Plot of the 5th, 50th, and 95th centile population-based curves for head circumference with respect to GA proposed by Papageorghiou et al [176] generated from the INTERGROWTH-21st dataset. It is clear that the surface-extracted HC values follow a similar trajectory to the manually-measured values, but there is an average negative bias of -10.7 mm, displayed by the Bland-Altman plot in (d).

the images of different subjects at different ages.

The fact that the surface deforms to adhere to the inner skull boundaries means that it conveniently separates the brain from extracerebral tissue such as the skull, skin, and maternal tissues such that it may also be employed as a skull-stripping tool. Thus, this framework may prove useful as a preprocessing technique for neuroimage analysis algorithms such as segmentation and registration. In addition, the delineation of the cranial outline allows for the extraction of clinically useful biometric measurements for applications in fetal growth monitoring and detection of craniofacial dysmorphology [106].

For the purposes of this thesis, the parametrization of the cranium with anatomical consistency is of utmost importance. The resulting cranial domain allows neuroanatomical information to be queried from any patient's image, of any age and stage of structural development, based solely on parametrized skull position. This property forms the basis of the feature-based predictive model for neurosonography-based GA estimation presented in Chapter 4.

4

Learning-Based Prediction of Gestational Age

Following from the previous chapter in which we describe the process of defining a "cranial surface domain" for consistently sampling anatomical regions from US images, this chapter describes the use of the surface domain for feature extraction in a learning-based predictive model. The cranial surface serves as a topological manifold representation of the fetal head, and allows for sampling of anatomically-corresponding regions between images of different subjects at different ages. Here, we describe the feature extraction process and detail the

design of bespoke features which capture age-related neurosonographic patterns for use in a regression forest for GA prediction. The work presented in this chapter has been published in:

Namburete, A. I. L., Stebbing, R. V., Kemp, B., Yaqub, M., Papageorghiou, A., Noble, J. A., "Learning-based prediction of gestational age from ultrasound images of the fetal brain," *Medical Image Analysis*, vol. 21 (1), pp. 72–86, Apr 2015.

Namburete, A. I. L., Yaqub, M., Kemp, B., Papageorghiou, A. T., Noble, J. A., "Predicting fetal neurodevelopmental maturation in ultrasound images," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 2 of *Lecture Notes in Computer Science*, pp. 260–267, Springer-Verlag / Heidelberg, 2014

Namburete, A. I. L., Noble, J. A.,"Machine learning-based tool to predict gestational age from ultrasound images of the fetal brain," in *The 2nd Annual Oxbridge Women in Computing Science Conference*, March 2015, Oxford, UK — Podium presentation, *Best Poster Award*

4.1 Introduction

For the task of predicting GA, and hence maturation, from brain images, the features must be designed with the objective of extracting morphological changes guided by the findings of post-mortem neuroanatomical studies of early brain development [16, 39]. Specifically, the second and third trimesters are marked by *a*) increasing cortical complexity with the emergence and developmental progression of sulci and gyri on the fetal brain surface [40, 8, 7]; *b*) an increase in overall brain size and volume [187, 52]; and *c*) changes in brain surface curvature in different cortical regions [168, 128, 144].We designed three

unique features banks, each capitalising on this prior knowledge about anatomical maturation, to be used in our machine learning framework to develop a link between cerebral progression and GA.

Typical approaches to mapping anatomical structure from image data to age have relied on accurate inter-subject segmentation and registration of brain images [117, 118, 5, 172, 122] or atlas creation [6], which remains a challenge in US images of the fetal brain. In particular, when working with US images of fetal brain anatomy, one cannot rely on the assumption that inter-subject alignment is capable of achieving one-to-one anatomical correspondence in subjects of different ages. Given that our data-driven model employs image appearance features to discover distinctive anatomical patterns related to ageing, it may be likened to Feature-Based Morphometry (FBM) [121] as it aims at quantifying feature variability based on cortical appearance, geometry, and occurrence statistics at different GAs. The features capture local patterns of anatomical development, and the parametrized surface model fulfils the role of a common coordinate space, voiding the need for a global brain atlas. The features are thus sampled with respect to the parametrized surface model, ensuring inter-subject anatomical consistency in the image regions sampled for feature extraction [188].

4.2 Feature Design

3D images contain a large amount of information and possibly several neighbouring image patches containing similar information. Reducing the number of surface/image 'points' included in the search space reduces redundancy which in turn improves the computational cost. To this end, the cranial surface is densely evaluated with a preselected number of points to represent anatomical

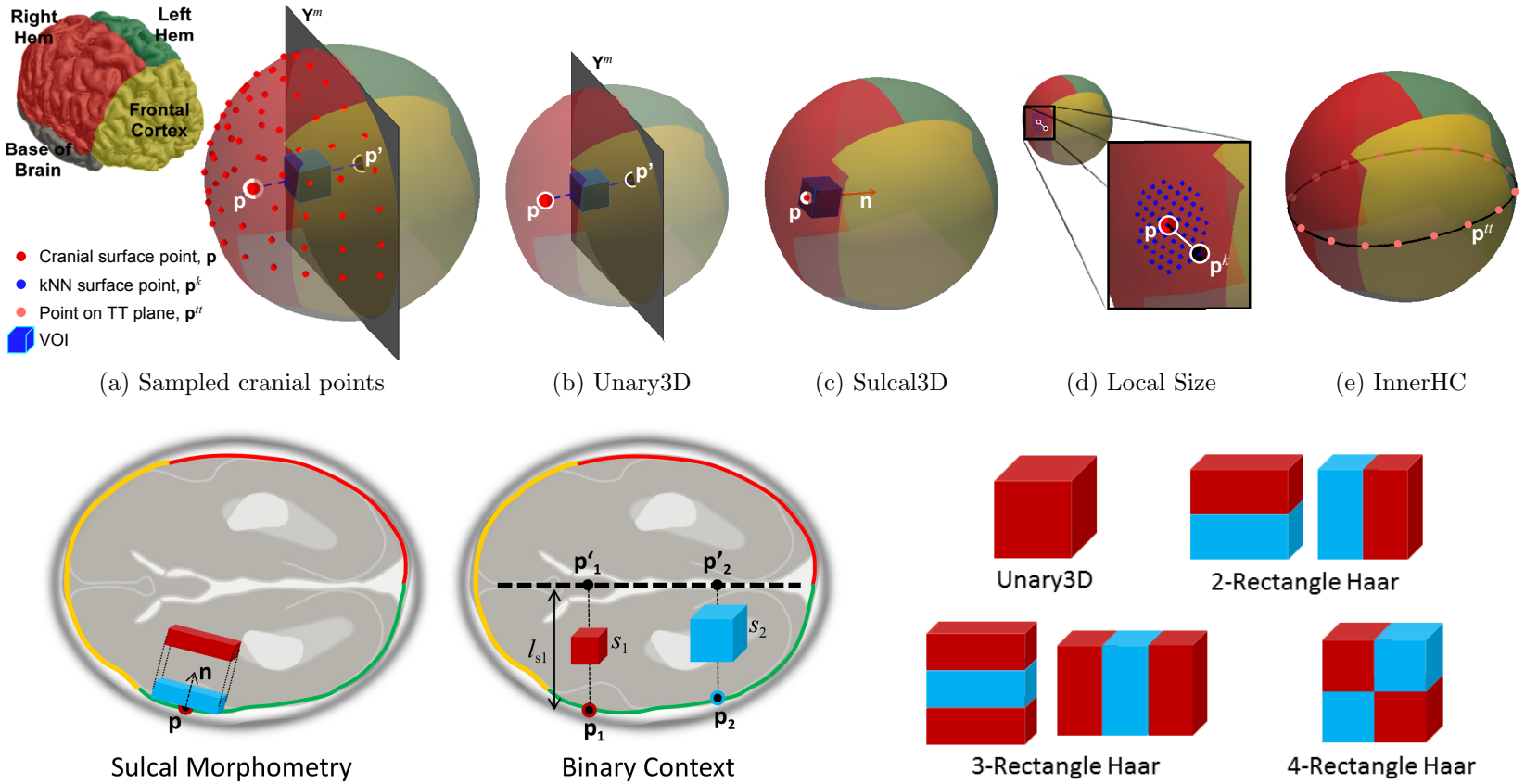
regions of interest, $\mathbf{P}$.

However, due to the effects of cranial calcification, the brain hemisphere proximal to the US probe is typically occluded, leaving only the distal hemisphere with visible and discernible intracranial structures. As such, feature extraction is confined to points on half of the cranial surface $\mathbf{P}^h \in \mathbb{R}^{3\times N_{\mathbf{P}^h}}$ corresponding to the distal cerebral hemisphere in the image volume, such that $h \in \{L, R\}$ denotes the hemisphere in question (L=left, R=right, respectively; c.f. Figure 4.1a). For simplicity, the surface is split by the midsagittal plane—defined by a normal vector and the center point of the plane (i.e. $\mathbf{Y}^m \in \mathbb{R}^{3\times 2}$)—and the sparse subset of points denoted by $\mathbf{P}^h$ is a matrix of $N_{\mathbf{P}^h}$, where $N_{\mathbf{P}^h}$ is a small subset of the total possible number of surface vertices, (i.e. $N_{\mathbf{P}^h} < N_{\mathbf{U}}$; Figure 4.1a).

4.2.1 Appearance-Based Features

This feature bank extracts information from the grayscale US image voxels in order to capture age-related sonographic patterns of a given anatomical region. Given the points on the deformed surface $\mathbf{P}^h$ and the 3D US image $\mathbf{I}$, a cuboidal volume-of-interest (VOI) is sampled based on the cranial points available on the selected hemispheric subsurface (Figure 4.1a). The location of the VOI is determined by the position of the sampled cranial point $\mathbf{p}$, and its normal vector $\mathbf{n}$.

VOI dimensions are defined by a scalar side length, s, which is scaled with respect to the distance between cranial point $\mathbf{p}$ and its projection onto the midsagittal plane, $\mathbf{p}'$, such that $s = l_s\|\mathbf{p} - \mathbf{p}'\|$, where l_s is randomly selected during feature evaluation from a uniform distribution ranging from 0 to 0.5 (i.e. $l_s \sim \mathcal{U}(0, 0.5)$). Scaling allows for characterization of local anatomy independent of cranial surface size such that like-for-like anatomical comparisons



(a) Sampled cranial points (b) Unary3D (c) Sulcal3D (d) Local Size (e) InnerHC

(f) Appearance Features (g) Haar-like Features

Figure 4.1: **Feature design.** Schematic representation of feature sets with respect to the parametrized cranial surface model. (a) Features are extracted by first sampling a cranial point **p** from a set of available surface points. A cuboidal VOI can be used to characterize appearance based on (b) Unary3D or (c) Sulcal3D features sets. Size features characterize (d) local skull deformations or (e) global head growth. (f) Appearance-based features are designed to capture sulcal changes and to compare structural changes between different brain regions. (g) Haar-like features are computed from the cuboidal VOI's by subtracting the voxels within the red cuboid from those in the blue cuboid.

can be achieved. The exclusion of age-related brain/cranial growth factors allowed by relative VOI sizing is paramount to retaining the 'pure appearance characterization' quality of this feature set.

Upon extraction of the image voxels within the cuboidal VOI, the feature score is evaluated by computing one of the Haar-like features shown in Figure 4.1g. In particular, the score is determined by subtracting the sum of voxels in a cuboid from the sum of voxels in an adjacent cuboid of the same dimensions (shown as red and blue cuboids in Figure 4.1g). The cuboids are of arbitrary dimensions and aspect ratio, and the resulting feature score is sensitive to edges and ridge-like structures within the image.

Appearance-based features comprise of two groups: sulcal and intracranial VOIs. Sulcal features are evaluated by affixing the VOI to the inner cranial surface (Figure 4.1c). They are designed to capture the sonographic image appearance related to changes in shape and morphology of the sulci and gyri on the cortical surface across gestation (see Figure 2.8). When evaluating sulcal features, the VOI is oriented along the vector $\mathbf{n}$ normal to the surface at cranial point $\mathbf{p}$ (Figure 4.1f).

Intracranial features, on the other hand, are evaluated by displacing the VOI along the vector normal to $\mathbf{Y}^m$ (Figure 4.1b). The VOI displacement $d_{VOI} = r\|\mathbf{p} - \mathbf{p}'\|$ is determined in proportion to the distance between $\mathbf{p}$ and $\mathbf{p}'$, where $r \sim \mathcal{U}(0, 1)$ allows for the VOI to be placed anywhere in the trajectory between the inner skull and the falx cerebri (or midsagittal plane, $\mathbf{Y}^m$), ultimately covering the entire space in the cerebral hemisphere. The value of r is also randomly selected during feature evaluation. This relative VOI positioning encourages anatomical correspondence in the sampling of intracranial regions, independent of brain size.

4.2.2 Local Size Features

These features capture local skull deformations and hence local growth patterns at different time points in pregnancy. To compute this feature, a cranial point $\mathbf{p}$ is first sampled from the set of available points on the hemisphere of interest, $\mathbf{P}^h$. Using a Ball tree search algorithm [189], the k cranial points nearest to $\mathbf{p}$ are identified. During feature evaluation, one of the k-th nearest neighbours, $\mathbf{p}^k$, is randomly selected and the distance between the sampled cranial point (Figure 4.1d, in red) and its randomly-selected k-th nearest neighbour (Figure 4.1d, in black) is obtained in Euclidean space $\mathbb{R}^3$:

$$d_{\text{Euclid}} = \|\mathbf{p} - \mathbf{p}^k\| \tag{4.1}$$

For our experiments, $k = 9$.

4.2.3 Biometric Features

Guided by prenatal assessment of fetal growth, the biometric feature is akin to the clinical head circumference (HC) measurement acquired from the standard transthalamic (TT) plane of the head (see Figure 2.1; [17]). In this case, the feature is evaluated as the perimeter of the inner contour of the deformed cranial surface at the level of the diagnostic TT plane, $\mathbf{Y}^{tt}$. To define a parametric representation of the TT plane on the cranial surface, the plane is first identified on a single reference brain volume by manual selection of three cranial points, $\mathbf{Z}^{tt} \in \mathbb{R}^{3\times3}$. The surface points defining the TT plane, $\mathbf{P}^{tt}$ (Figure 4.1e, in pink), are the surface vertices which intersect the plane approximated by $\mathbf{Z}^{tt}$.

At feature evaluation, the inner cranial contour (or inner HC) is obtained from each image by extracting the 2D TT plane $\mathbf{Y}^{tt}$ defined by $\mathbf{P}^{tt}$ and com-

puting the closed path length of all the surface faces intersected by the plane. The biometric HC feature captures global changes in head size in a manner similar to the current clinical method of GA estimation [188], emulating global rigid cranial transformations related to fetal head growth.

4.3 Modelling Age-Specific Sonographic Appearance

Our goal is to use age-discriminating image information to predict GA from 3D images of the brain. To develop the predictive age model, we take advantage of the regression forest framework that is well-established in the literature [190] to produce estimates of continuous variables. Given that GA is itself a continuous variable, we opted for training the forest from a large longitudinal dataset of brain US images spanning the entire age range of interest, detailed in § 4.3.1. The process of predictive age regression is explained in § 4.3.2.

4.3.1 Learning the link between age and brain images

The model is trained in a supervised manner on a dataset in which all training examples are annotated with the GAs at scanning. In our framework, each training example $V_i = (\mathbf{I}_i, \mathbf{P}_i^h, a_i)$ comprises of an image $\mathbf{I}_i$ and its corresponding deformed cranial surface parametrization $\mathbf{P}_i^h$, labelled with the GA of the fetus at the time of scanning, a_i. The dataset is denoted by $\mathcal{V} = \{V_i\}$. Figure 4.2 displays a schematic of the forest training process.

During training, we perform bagging with replacement to select a random subset of the dataset, $\mathcal{V}^t \subset \mathcal{V}$, to traverse each tree T_t in the regression forest $\mathcal{F} = \{T_t\}$. All training examples in the subset are passed into the root node $\left(\mathcal{V}_{j=0}^t\right)$ of each tree, and the data are recursively partitioned at each node j in

their path along the tree branches until reaching leaf nodes. At each node, the data are split to the left and right child nodes, $\mathcal{V}^t_{j\in\{L,R\}}$, by age-discriminating binary tests according to the node optimization criteria described below. Data splitting continues until leaf nodes are created due to *a*) a maximum number of tree levels (i.e. tree depth, *d*) has been reached; *b*) the node contains less than a pre-defined number of training data points; *c*) a minimum information gain has been achieved.

Node Optimization

The age-discriminating binary tests are designed to minimize the differential entropy between the data sent to the left and right children nodes. At each node, a feature f_i is randomly-selected from the available feature sets $\mathbf{f} \in \mathbb{N}^n$ (described in § 4.2). The feature is applied to the dataset at the node $\mathcal{V}^t_j \subset \mathcal{V}^t$ and data splitting between children nodes occurs on the basis of a threshold, τ. The optimally age-discriminating node feature is determined by applying m features ($m \ll n$) to $\mathcal{V}^t_j$ to identify the pair of (f_j, τ_j) which minimizes the cost function.

Modelling the ages of the dataset at each node j as a random variable $\mathbf{a}_j$ with univariate Gaussian distribution, $\mathbf{a}_j \sim \mathcal{N}(\bar{\mathbf{a}}_j, \sigma^2_j)$, the cost function can be expressed as

$$I_g = \log\left|\sigma^2_j(\mathcal{V}^t_j)\right| - \sum_{i\in\{L,R\}} \omega_i \log\left|\sigma^2_i\left(\mathcal{V}^{t,i}_j\right)\right| \tag{4.2}$$

where $\bar{\mathbf{a}}_j$ is the mean age of all samples at the j-node, σ^2 denotes the variance of the ages, and ω_i is the ratio between the number of training examples in a child node $\mathcal{V}^{t,i}_j$ and the number of examples in the parent node $\mathcal{V}^t_j$, i.e. $\omega_i = |\mathcal{V}^{t,i}_j|/|\mathcal{V}^t_j|$. Maximizing Equation 4.2 is equivalent to favouring binary tests which

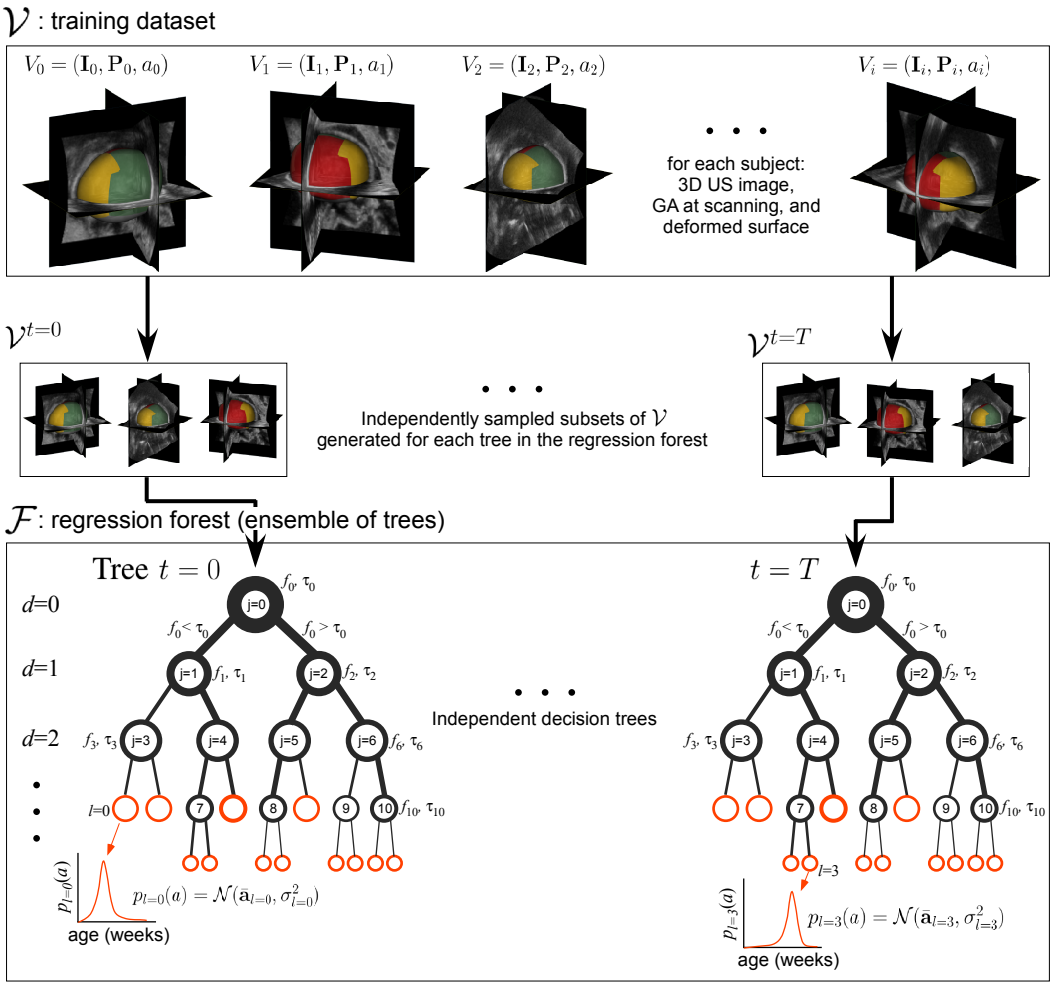


Figure 4.2: **Regression forest training.** Schematic of the forest training procedure. During training, a subset of age-labelled images $\mathcal{V}^t$ is passed into each tree t to optimize binary tests to be used in the prediction process. As the data points traverse the tree from the root down to the leaf nodes, the variance of the ages associated with the data points (modelled as univariate Gaussian distributions) at each specific node decreases, and the confidence in turn increases.

minimize the variance of the ages of the partitioned data, ultimately reducing the uncertainty in the age-discriminating ability of a given test on $\mathcal{V}_j$. The discriminative quality of these binary tests hinges on the assumption that $\mathcal{V}^t$ is representative of the complete dataset, $\mathcal{V}$.

Each leaf node l stores the mean $\bar{a}_l$ and variance σ_l^2 of a Gaussian distribution derived from the vector of ages $\mathbf{a}_l$ to have reached it:

$$p_l(a) = \mathcal{N}\left(\bar{a}_l, \sigma_l^2\right) \tag{4.3}$$

Note that the proposed implementation is *unbiased* in that it weights all training examples equally and independently, as it does not incorporate any longitudinal information about a specific subject.

4.3.2 Prediction of GA

During age prediction, an unseen data point of unknown GA, $W_k = (\mathbf{I}_k, \mathbf{P}_k^h)$, traverses the nodes in each tree of the trained forest model, and the binary test associated with each node evaluates whether to send the data to the left or right child nodes, until W_k eventually reaches a leaf node (Figure 4.3). For each tree, the leaf node reached provides a mean age estimate with an associated variance. Leaf nodes with high variance values have lower age certainty, so they are assumed to be less informative and likely to add noise to the output predictions. Therefore, a single prediction a_k is generated by taking the mean of only the age estimates ($\mathcal{A}$) with associated variance less than $\sigma^2_{\max}$. Specifically:

$$\mathcal{A} = \{a_l | \sigma_l^2 < \sigma^2_{\max}\}, \quad a_k = \tfrac{1}{|\mathcal{A}|} \textstyle\sum_{a_l \in \mathcal{A}} a_l \tag{4.4}$$

where $|\mathcal{A}|$ is the number of ages satisfying the exclusion criteria.

4.3.3 Evaluation of GA Predictions

For the validation experiments which follow, we measured the accuracy of GA predictions using the root-mean-squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{|\mathcal{W}|} \sum_{k=1}^{|\mathcal{W}|} \left(a'_k - a_k\right)^2} \tag{4.5}$$

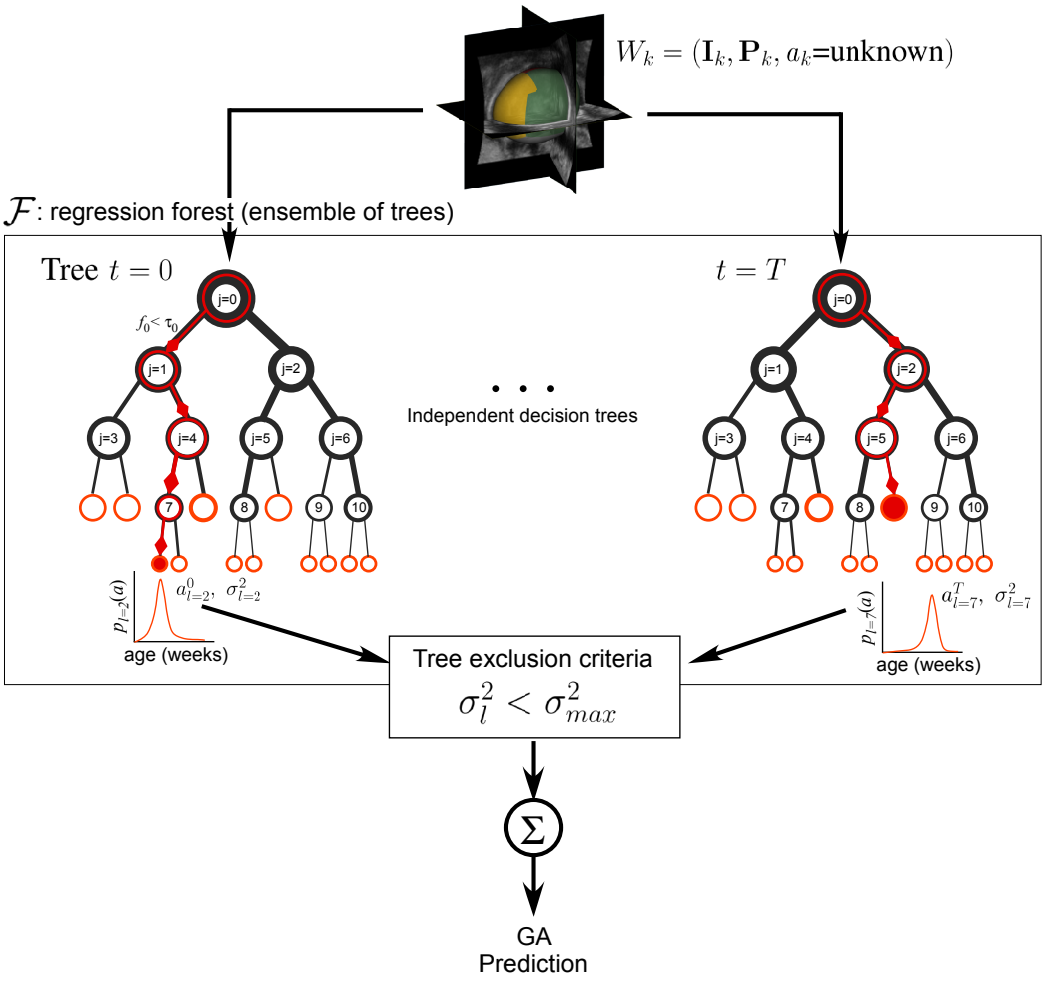


Figure 4.3: **Regression forest testing.** Schematic of the forest testing procedure for an unseen data point, W_k. During testing, W_k is passed into each constituent tree in the forest. At each internal split node, a binary test is applied, and W_k is sent to the appropriate child node (depending on whether its feature result was above or below a predetermined threshold value, τ. This process is repeated until the data point reaches a leaf node, and a reading of the stored leaf prior is taken. The age prediction for W_k is achieved by averaging all the tree posteriors distributions associated with the leaf nodes reached by the data point.

where $|\mathcal{W}|$ is the number of test images, a'_k is the model-predicted GA of the k-th test image, and a_k is the true (known) GA. We also evaluated the Pearson's linear correlation coefficient (r) between the model-predicted and true GAs.

4.4 Training Dataset

The model was developed to characterize all brain regions observable in the distal hemisphere of US images. The input data to the framework are 3D US images of the fetal brain obtained from the INTERGROWTH-21$^{\text{st}}$ study database, which comprises of an optimally healthy group of women with a low-risk of pregnancy complications and fetal abnormalities [176]. The data were acquired to a standard imaging protocol but not specifically for the purposes of brain-based GA estimation, so the acquisition settings were not optimized for visualizing intracranial structures. 'True age' was defined by the last menstrual period (LMP) and confirmed by crown-rump-length (CRL) measurement on US images taken in the first-trimester ($\leq 14^{+0}$ weeks) agreeing within 7 days. For this dataset, CRL-based age is accurate within 2.7 days, determined from 3 independent clinical measurements. For this work, the age prediction model was trained using 448 3D US images of the fetal brain (198 and 250 images with visible left and right hemispheres, respectively) ranging from 18^{+0} to 33^{+6} GW. Figure 4.4 shows the GA distribution of the fetal cohort. Cross-validation experiments were conducted on the same dataset following a leave-10-out protocol, each time reserving 10 images for testing GA prediction performance and using the remaining images for training the regression forest classifier.

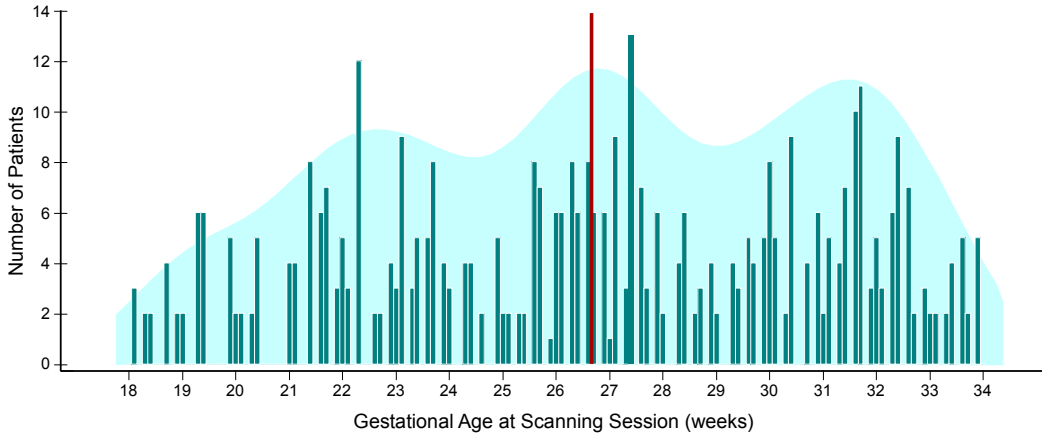


Figure 4.4: **Gestational age spread of the healthy fetal cohort**. Bar chart of the number of subjects for each gestational age with accompanying density trace (estimated using a Gaussian kernel, $\sigma = 1$). The mean age of the cohort is shown in red.

4.5 Validation of Parameter Selection

Regression forests contain several parameters which can be set based on the desired application. For the task of neurosonography-based GA estimation, each regression forest was constructed by training T trees to a maximum depth d_{max}. At each node j, a total of m features was sampled, from which only one feature f_j was selected as possessing the most age-discriminating power. This section discusses the selection of forest parameters which optimize the GA prediction task. The purpose of these experiments is to demonstrate the behaviour of the algorithm with varying parameter values, so for the sake of simplicity and discussion, we trained forests which included only appearance-based features (§ 4.2.1) in the feature vector. In addition, we discuss the results for forests trained only on images with a visible left cerebral hemisphere, but the algorithm demonstrates similar behaviour in images of the right hemisphere.

Variations in Data Preprocessing

All 3D images used for training and validation of the model were preprocessed by first resampling the acquired clinical US images to an isotropic spatial resolution of $0.6 \times 0.6 \times 0.6\text{mm}^3$. Ridge-like structures were then enhanced using a bandpass Gaussian derivative filter. As shown in Figure 4.5, bandpass filtering the images with a Gaussian derivative kernel enhances the ridges of the sulci and soft tissue boundaries (e.g. separating cerebrospinal fluid from white matter), yet it reduces some of the speckle seen in US images. The bandpass kernel size is proportional to the width of the structures preserved by the filtering process. Thus, filtering the images with a small kernel preserves more of the speckle and enhances thinner ridges, whilst large kernels reduce the speckle and enhance thicker structures, at the risk of losing finer structural details.

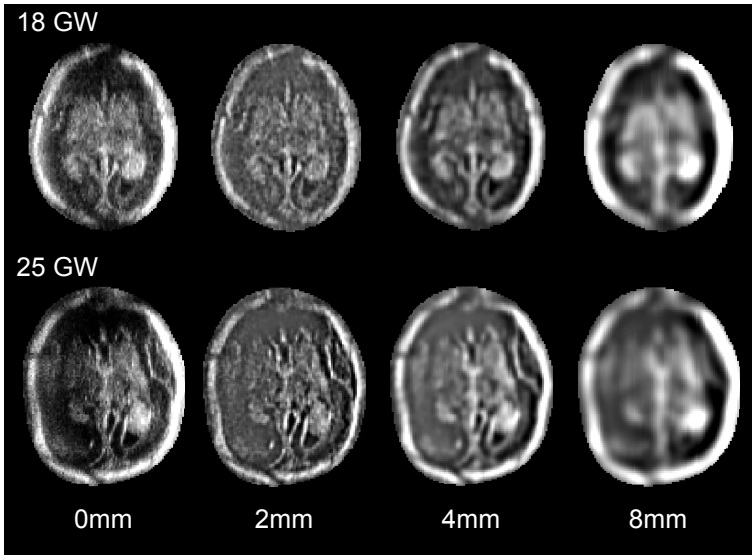


Figure 4.5: **Bandpass filtering of US images.** Comparison of different sizes of Gaussian derivative kernel (from left to right: $\bigtriangledown G_\sigma = 0\,\text{mm}$, $2\,\text{mm}$, $4\,\text{mm}$, and $8\,\text{mm}$) for images acquired from fetuses at 18 (top) and 25 GWs (bottom). Axial slices at the level of the Sylvian fissure are shown.

An exploration into the effect of filter kernel size ($\bigtriangledown G_\sigma$) on the accuracy of GA prediction shows that when the model is trained with images filtered with

a large kernel (i.e. $\bigtriangledown G_\sigma = 8\,\text{mm}$), the error is larger and may be attributed to the loss of finer details and sulci of interest (Figure 4.6). On the other hand, small kernels (i.e. $\bigtriangledown G_\sigma = 0 - 2\text{mm}$) preserve speckle, which contributes to the variations in anatomical appearance — and hence speckle patterns — from image to image due to the position of the US probe relative to the brain. The 'optimal' kernel size enhances the sulci which are typically of a particular width ($\bigtriangledown G_\sigma = 4\,\text{mm}$ in our images), and suppresses details which vary with acquisition protocols.

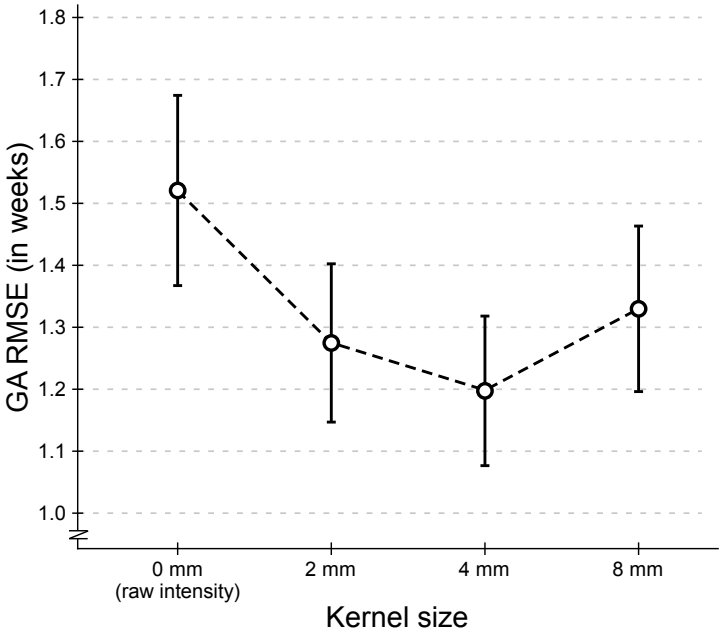


Figure 4.6: **Effect of Gaussian derivative kernel size on age prediction.** Comparison of GA prediction error for different sizes of Gaussian derivative kernel ($\bigtriangledown G_\sigma = 0\,\text{mm}$ to $8\,\text{mm}$). Error bars depict the standard error of the mean (SEM) from the leave-10-out cross-validation experiments. Forest settings: $T = 15$, $m = 100$, $d_{max} = 20$.

Cranial parametrization was then applied to the preprocessed data as detailed in Chapter 3. The bandpassed 3D US images and their respective deformed parametric surfaces were passed into the regression forest algorithm for training and testing.

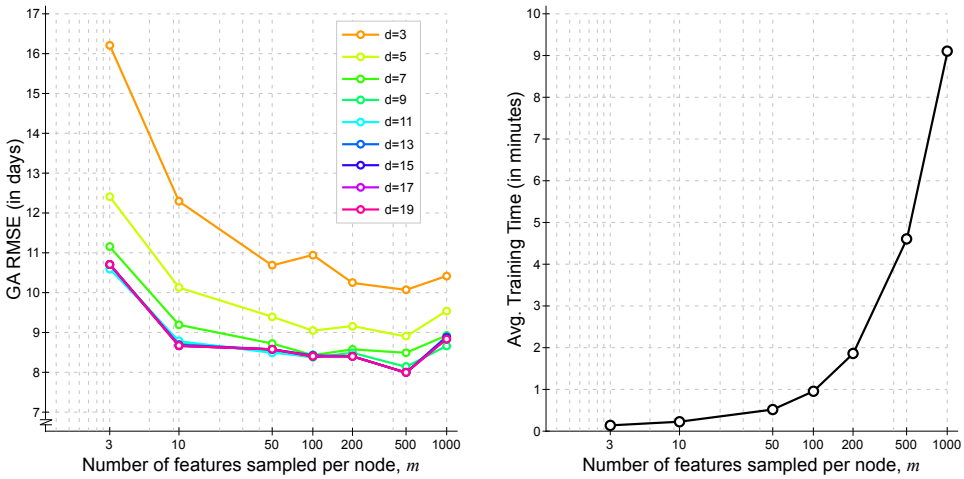
Feature Space Size, m_{opt}

The set of discriminative features which will be relevant to the GA estimation task is not known *a priori*. During training, the goal is to automatically determine which of the features will be important by learning them with respect to the training data and age distributions. At each node, a set of m features is sampled from the available feature space (sampled from feature libraries described in § 4.2) and applied to the training data at the given node. By setting the parameter m to a high value, the trees are exposed to more features during training, essentially increasing the likelihood of finding the most age-discriminating feature for the training data provided. However, a high m may result in trees which are more correlated, which is an undesirable property as it would mean that the trees in the forest would have a high degree of similarity as they would be largely selecting similar features from the feature space. In contrast, a low m would result in highly decorrelated, but less informed trees for the GA prediction task as the sampled features are less likely to be *the* most discriminative features available in the feature space.

In our implementation, the overall feature set was not pre-computed and stored in memory. Rather, the m features were computed on demand at runtime during construction of each node in the tree. This design decision impacts the computational complexity of the training process. Figure 4.7 shows an experiment in which m was set to different values; that is, a comparison between models which were trained with different numbers of features made available to each split node during the training process (i.e. varying m). It is evident from the figure that increasing the number of features available to each node results in high computational costs at training time [1]. Thus, in consideration

[1]The age regression framework was implemented in C++ (3.30GHz quad-core, 12GB RAM), so all computation times and performance metrics are reported accordingly.

with the computational budget and relatively low gain in prediction accuracy between using $m = 200$ and $m = 500$, the rest of the experiments presented in this thesis are based on forests trained with $m_{opt} = 200$.



(a) m versus GA prediction error (b) m versus training time

Figure 4.7: **Influence of the number of features sampled at each node during training.** Prediction error generally decreases with increasing number of features sampled per node, but there is a trade-off between accuracy and the exponentially increasing training time. Forest settings: $T = 15$, $d_{max} = 20$.

Maximum Tree Depth, d_{max}

In the random forest algorithm, the maximum tree depth d_{max} is an important parameter as it relates to overfitting [160]. Setting the maximum tree depth to a low value may limit the predictive power of the forest, as tree construction may stop to create nodes with high sample variance which could have been further split into nodes with higher confidence (i.e. distributions with lower variance). In contrast, a very high d_{max} value may result in the tree construction to stop before the nodes reach d_{max} due to a low number of data points reaching the nodes: a stopping criterion which prompts the creation of a leaf node. However, in our implementation, the influence of a high d_{max} value is dampened by the fact that the leaf distributions at shallower tree levels are propagated to absent

nodes at deeper levels, as shown in Figure 4.8. In doing so, even if the tree is trained to a high d_{max} and leaves are created at shallower levels, overfitting is avoided by assigning the same leaf distributions to nodes at deeper levels if the tree runs out of data points.

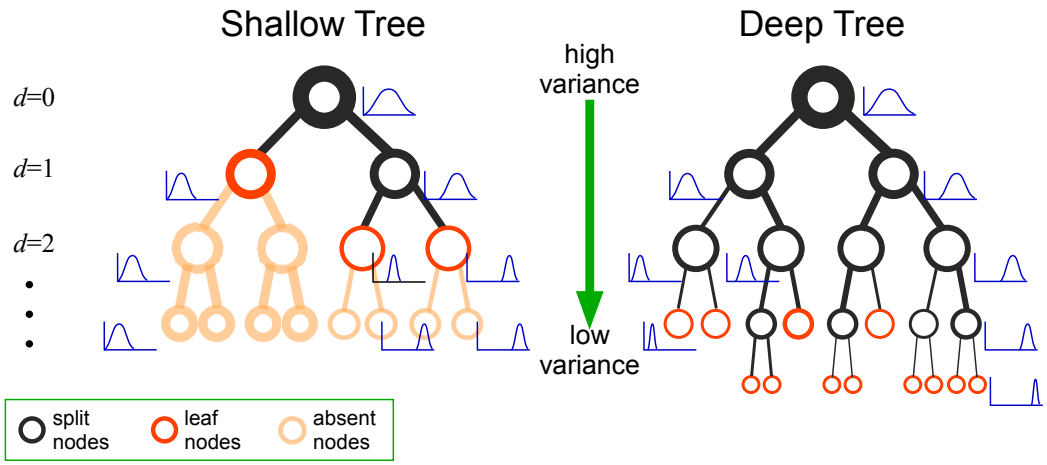


Figure 4.8: **Tree depth and leaf node distributions.** The variance of the node distributions decreases with tree depth. Note that the absent nodes were not constructed during the forest training process. They act as placeholders for posterior probabilities required at deeper depths during testing, and they possess the same probability as their nearest parent leaf node.

This behaviour is demonstrated by Figure 4.9 in which the predictions made at shallower depths were less accurate, and the error saturates at the levels deeper than the optimal depth (i.e. $d = 7$).

The maximum tree depth also influences the computational complexity of tree construction, as a deeper tree has higher memory requirements and takes longer for a test image to traverse to achieve a GA prediction. Thus, identifying the point of saturation is important for determining the point at which the prediction accuracy no longer improves with the creation of a deeper forest and to then set d_{max} such that memory is used more economically.

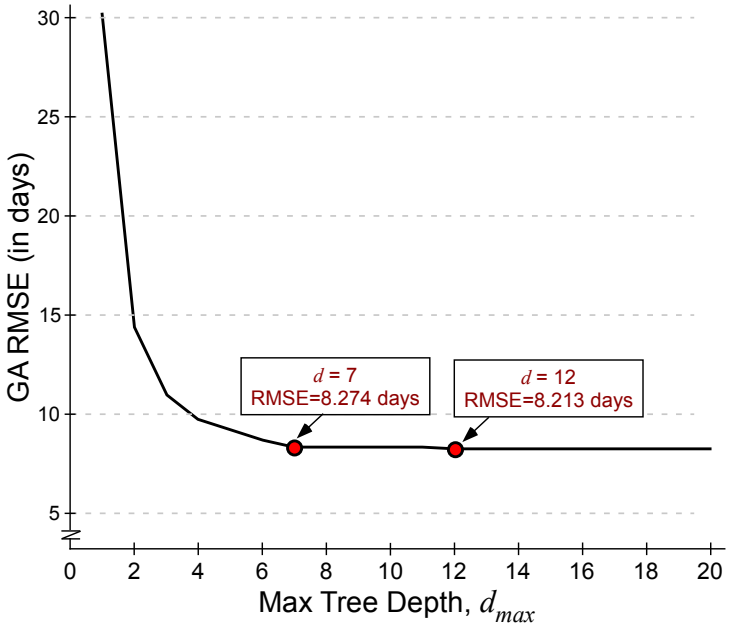


Figure 4.9: **Effect of tree depth on age prediction.** GA prediction error (RMSE) as a function of maximum tree depth (d_{max}). Error saturates at levels deeper than an optimal value (i.e. $d = 7$ for the task of GA prediction). Forest settings: $T = 15$, $m = 100$, $d_{max} = 20$.

Number of Trees, T

Each tree in the forest is independently trained on a randomized subspace of the entire feature representation of the image. During the prediction process described in § 4.3.2, each tree produces a single GA estimate and the overall forest prediction is generated from the mean of the predictions made by *all* trees in the forest. Increasing the number of trees typically results in increased accuracy by virtue of incorporating more information into the prediction. However, this increase in accuracy eventually levels off when the optimal number of trees is reached, which may be application-dependent. Identifying the optimal number of trees at the point when the levelling effect occurs is particularly important when considering that the computational complexity of both training and testing increases with tree count. For the task of GA estimation, the error begins to level off after $T = 20$, but the turning point (or local minimum error)

was found to be $T = 36$ (Figure 4.10).

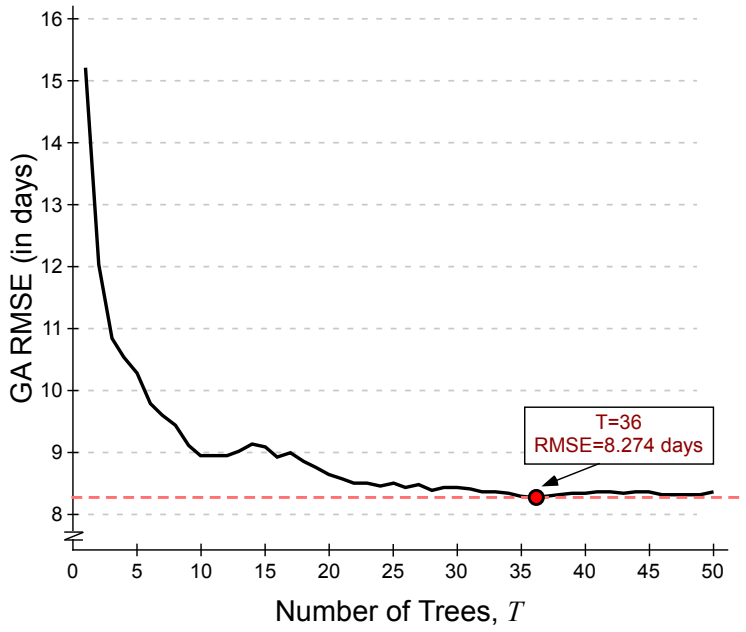


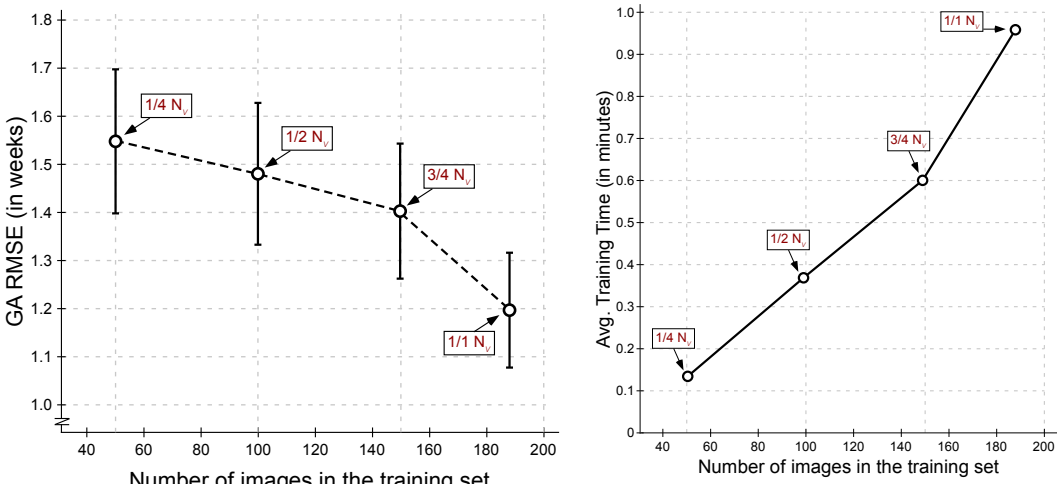
Figure 4.10: **Effect of number of trees on age prediction.** GA prediction error as a function of the number of trees in the forest. As the tree count increases, the accuracy generally improves, but as the number of trees increases the marginal benefit of additional trees decreases. Forest settings: $m = 50$, $d_{max} = 20$.

Dataset Size, $N_{\mathcal{V}}$

In order to explore the scalability of our framework to larger datasets, we explored the influence of the size of the training set ($N_{\mathcal{V}}$) on the accuracy of age prediction. Figure 4.11a illustrates that $N_{\mathcal{V}}$ indeed affects the accuracy of age prediction. The accuracy of age prediction increases when the model is trained with a larger dataset, and it appears that the RMS error is likely to continue to decrease as more data is provided for model training.

In addition, Figure 4.11b shows that the average training time increases linearly with data size from the 1/4 to 3/4 of the full dataset. However, there is a slight decrease in speed when using the full dataset, which may be attributed to memory limitations. The typical training time is most relevant when consid-

ering the scalability and likely behaviour of training a model on a much larger dataset (i.e. $\sim$ 1000 images) as there is a trade-off between training time and prediction accuracy. However, it is worth emphasizing that the heavy computational burden is incurred only *once* for a given training dataset, and that it is possible to train the model on a high-resolution dataset from hundreds of subjects in a matter of hours. In practice, the time required to achieve a GA prediction for a single image is unaffected by the size of the training dataset.



(a) $N_{\mathcal{V}}$ versus GA prediction error (b) $N_{\mathcal{V}}$ versus training time

Figure 4.11: **Influence of the size of training dataset.** (a) GA prediction error as a function of the number of images included in the training set. The error steadily decreases as the training set gets larger. (b) In contrast, the higher the number of images in the training set, the longer it takes to train each tree. Forest settings: $T = 15$, $m = 100$, $d_{max} = 20$.

Variance Threshold, σ^2_{max}

During GA prediction (§ 4.3.2), a test image traverses the nodes in a tree until it reaches a leaf node which is associated with an age and a variance (σ^2_l). Each tree in the forest produces a single age estimate, and the variance threshold (σ^2_{max}) determines whether the estimate of a tree will be included in the overall age prediction (i.e. the mean value of all tree estimates whose leaf nodes are

associated with a variance below the variance threshold (i.e. $\sigma_l^2 < \sigma_{max}^2$). A low variance threshold results in the exclusion of trees from the GA prediction, selecting only the trees with high prediction confidence. In contrast, setting σ_{max}^2 to a high value includes a larger number of trees in the GA prediction, at the risk of including trees which may be noisy and ultimately contribute to a lower prediction accuracy. The optimal variance threshold will include the trees with high enough confidence, but exclude noisy trees to achieve accurate GA predictions. Figure 4.12 demonstrates this behaviour. For the task of GA prediction, the optimal value for the variance threshold is $\sigma_{max}^2 = 1.0$.

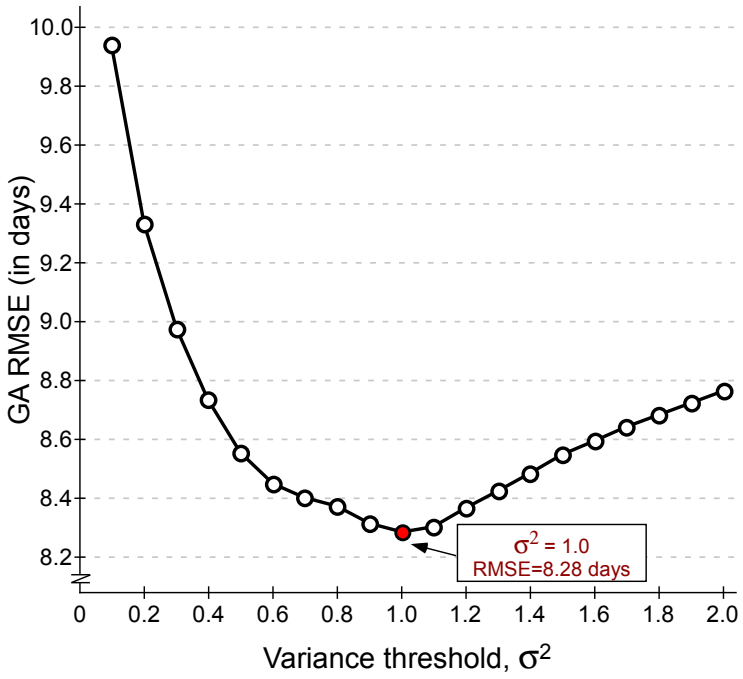


Figure 4.12: **Effect of variance threshold σ_{max}^2 on age prediction.** There is a trade-off between the variance threshold σ^2 and the number of trees included in the output prediction: low σ^2 is more constrained as it excludes more trees, whereas high σ^2 is more susceptible to noise as it includes more trees.

Application-Specific Parameter Settings: GA Prediction

For the experiments included in the rest of this chapter, the parameters were set to the values summarized in Table 4.1.

Table 4.1: **Forest parameter settings.** Parameter values for the regression forest algorithm for the task of GA estimation from US images. These values are based on the cross-validation experiments discussed in § 4.5 and applied to the rest of the experiments presented in this chapter.

Parameter	
Gaussian kernel size	$\bigtriangledown G_\sigma = 4\,\text{mm}$
Number of trees in forest	$T = 40$
Number of features tested at each node	$m_{opt} = 200$
Maximum tree depth	$d_{max} = 10$
Variance threshold	$\sigma^2_{max} = 1.0$

4.6 Validation of GA Prediction

In this section, we describe the experiments conducted in order to demonstrate the performance and functionality of the age regression forest for the task of estimating GA from 3D US images of the fetal brain.

4.6.1 Selection of Age-Discriminating Features

To gain insight into the functionality of the GA prediction model, we assessed the relative importance of each feature set presented in § 4.2 in its age discriminating ability. This was achieved by plotting forest level (d) against the number of times a particular feature f_i was selected at each level in the forest, normalized by the total number of nodes present at the given level across all trees in the forest (N_d), i.e. $\frac{N_d(f_j)}{N_d}$. This analysis is shown in Figures 4.13, 4.14, and 4.15; it is based on the understanding that decision trees select more general binary tests to split the data at the shallower levels and progressively select

more specific tests as the data traverses to deeper levels. In our work, such an analysis informs on which features contain global GA discriminating ability, and conversely which provide information about subtler or more detailed age-related differences. Three separate forests were trained for comparison (Table 4.2).

Table 4.2: **Types of feature sets.** Description of the different types of regression forests trained for model selection, and their constituent feature sets. Graphical explanation of each feature set is provided in Figure 4.1 and discussed in § 4.2.

Forest Type	Feature Vector
$\mathcal{F}_{app}$	Appearance only
$\mathcal{F}_{app+lSz}$	Appearance and local size
$\mathcal{F}_{app+lSz+HC}$	Appearance, local size, and biometry

We demonstrate the feature selection profiles for forests $\mathcal{F}_{app}$ (Figure 4.13), $\mathcal{F}_{app+lSz}$ (Figure 4.14), and $\mathcal{F}_{app+lSz+HC}$ (Figure 4.15) for the leave-10-out cross-validation forests, displaying each appearance and size feature subset separately. It is evident from Figure 4.13 that when the forest is trained exclusively with appearance-based features ($\mathcal{F}_{app}$), Haar-like features are selected more frequently than Unary3D or Binary3D context features. At all forest levels, Haar3D was consistently selected in more than 60% of the nodes as the most powerful feature. This indicates that the algorithm selected binary tests which capture ridge-like image features more frequently than those capturing differences in homogeneous image regions, suggesting that sulcal and fissural development were salient in the brain during the second and third trimester.

Figure 4.14 shows that when the feature vector comprises of appearance and local size features, the latter are more age-discriminating in the shallower forest levels, but anatomical appearance is dominant from level $d = 2$, and optimal prediction with minimal RMSE occurs at level $d = 11$.

However, when the innerHC feature is also included in the feature vector

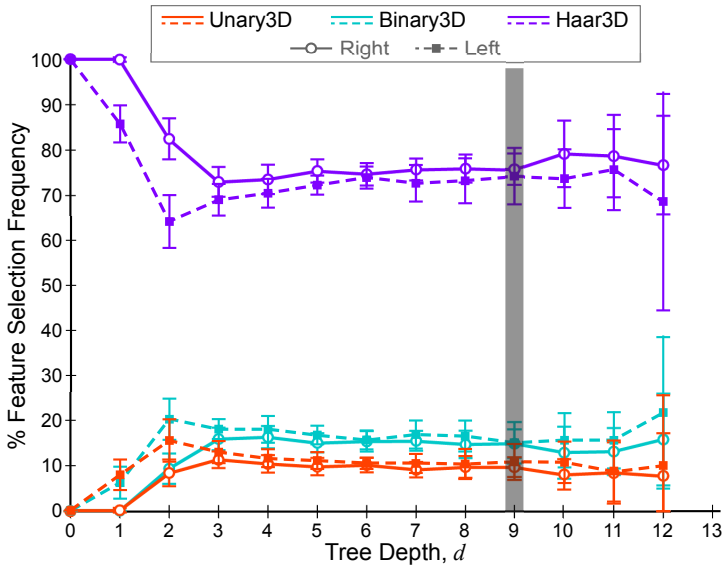


Figure 4.13: **Feature selection profile for $\mathcal{F}_{app}$.** Feature selection versus tree depth for the leave-10-out cross validation forest $\mathcal{F}_{app}$. Lines denote mean feature selection frequency at a given tree depth (solid: right distal hemisphere; dashed: left distal hemisphere). Error bars indicate standard deviation. Gray vertical lines indicate the tree depth of optimal GA prediction.

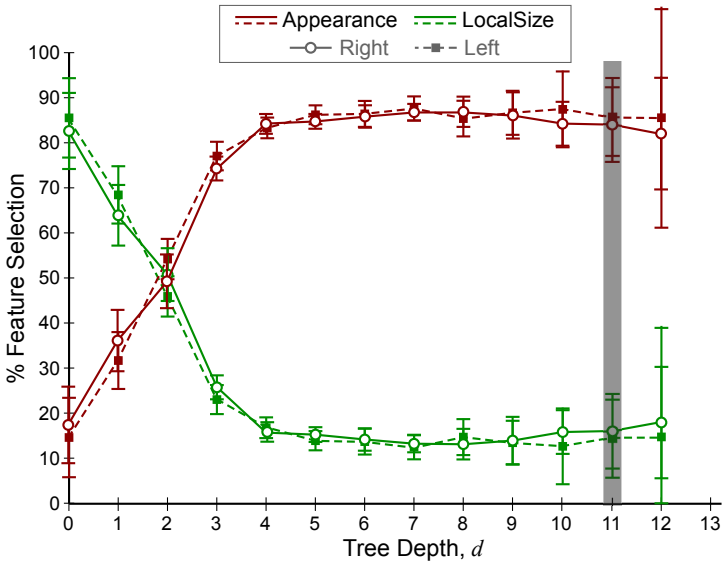


Figure 4.14: **Feature selection profile for $\mathcal{F}_{app+lSz}$.** Feature selection versus tree depth for the leave-10-out cross validation forest $\mathcal{F}_{app+lSz}$. Lines denote mean feature selection frequency at a given tree depth (solid: right distal hemisphere; dashed: left distal hemisphere). Error bars indicate standard deviation. Gray vertical lines indicate the tree depth of optimal GA prediction.

(Figure 4.15), it is preferentially selected in the shallower tree levels, superseding even local size features, before anatomical appearance features are rendered

more important also at level $d = 2$. **This clearly demonstrates that global and local head size provide an important first estimate of GA, but sonographic anatomical appearance is valuable in refining age predictions.**

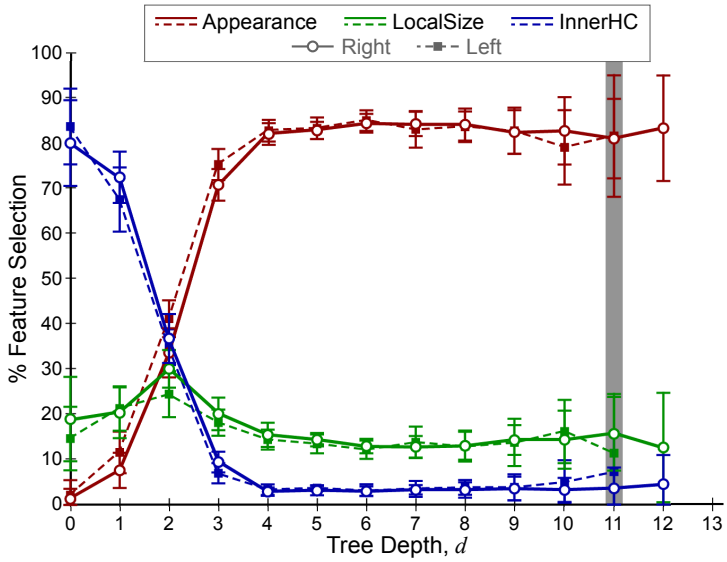


Figure 4.15: **Feature selection profile for $\mathcal{F}_{app+lSz+HC}$.** Feature selection versus tree depth for the leave-10-out cross validation forest $\mathcal{F}_{app+lSz+HC}$. Lines denote mean feature selection frequency at a given tree depth (solid: right distal hemisphere; dashed: left distal hemisphere). Error bars indicate standard deviation. Gray vertical lines indicate the tree depth of optimal GA prediction.

To assess the stability of feature selection, cross-validation experiments were conducted using a leave-10-out protocol, each time partitioning the dataset such that 10 images were kept for testing the model, whilst the remaining images (i.e. 188 and 240 images with visible left and right distal hemispheres, respectively) were used for training the corresponding regression forests. Figures 4.13, 4.14, and 4.15 display error bars indicating the standard deviation of feature selection frequency for all forests trained during cross-validation. The fact that the error bars tightly follow the curves for all feature profiles demonstrates that the model is stable in its selection of age-discriminating features at the different forest levels. It is also clear from the figure that regardless of which

cerebral hemisphere was used to train the model, the feature selection profiles were similar, and their age predictions are not statistically different at the 5% significance level. This demonstrates the ability of the model to simply capture anatomical progression of the brain, respecting cerebral hemispheric symmetry.

4.6.2 Feature Selection Paths

Figure 4.16 illustrates the process of feature selection and demonstrates the tree traversal paths for two data examples from fetuses at 19^{+3} and 28^{+3} GW for a typical tree from forest $\mathcal{F}_{app+lSz+HC}$. Note that although the examples traverse the tree along different paths before reaching a leaf node, the binary tests applied to achieve GA estimation follow a similar pattern. Specifically, the inner HC is tested at the root node ($d = 0$), followed by local size features in the shallow levels ($d = 1 - 2$), and lastly image appearance features were applied as the final tests before arriving at a leaf node where a decision about GA is achieved. This corroborates the findings of the feature selection profiles plotted in Figure 4.15, and most test examples followed a similar tree traversal pattern.

Also, the example from the younger fetus had a shorter tree traversal path than the older fetus. Guided by this indication, we found that on average, the trend was for tree traversal path length to increase with GA (Figure 4.17), which may be indicative of the longitudinal behaviour of the model as the brain increases in complexity.

4.6.3 Model-Based Prediction of GA

To demonstrate the predictive quality of each combination of feature sets across gestation and hence identify the best GA predictor, we plotted the age predictions from the leave-10-out cross-validation experiments on the healthy dataset

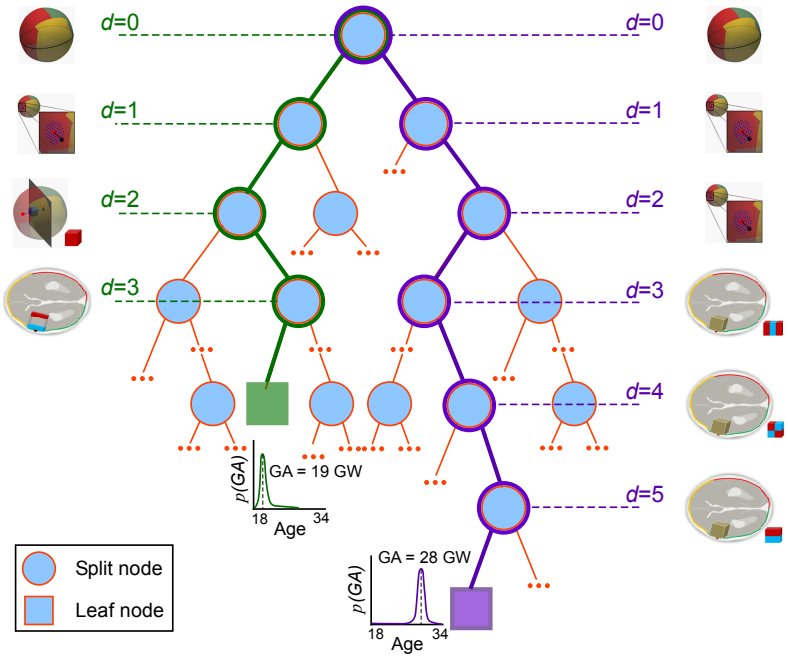


Figure 4.16: **Example feature selection.** Illustration of feature selection paths for two different fetuses at 19 and 28 GWs, demonstrating typical tree traversal during testing.

Table 4.3: **Subject demographics for training data and cross-validation results.** Age group demographics for the data used to train the regression forest algorithm. Cross-validation results of age predictions are presented, and $p < 0.001$ for all age groups.

Age Group (weeks)	No. Subjects (n)	Mean (SD) GA (weeks)	RMS Prediction Error (r)
18-20	22	19.00 (0.53)	± 0.65 (0.90)
20-22	51	21.13 (0.59)	± 0.43 (0.86)
22-24	59	22.98 (0.59)	± 0.35 (0.83)
24-26	46	25.04 (0.61)	± 0.62 (0.76)
26-28	85	26.92 (0.57)	± 0.68 (0.71)
28-30	46	29.04 (0.60)	± 0.55 (0.82)
30-32	73	30.94 (0.62)	± 0.47 (0.81)
32-34	57	32.80 (0.61)	± 0.89 (0.67)

(Figure 4.18). In each cross-validation session, 188 and 240 for the left and right hemisphere, respectively, were used for training a regression forest model with different feature sets. The model was then applied to the remaining 10 images for testing. Figure 4.18 shows the predicted versus the true GA for each subject. It is important to note that each fetus in the database was treated as a test subject in only one of the cross-validation sessions, and the figure shows the GA

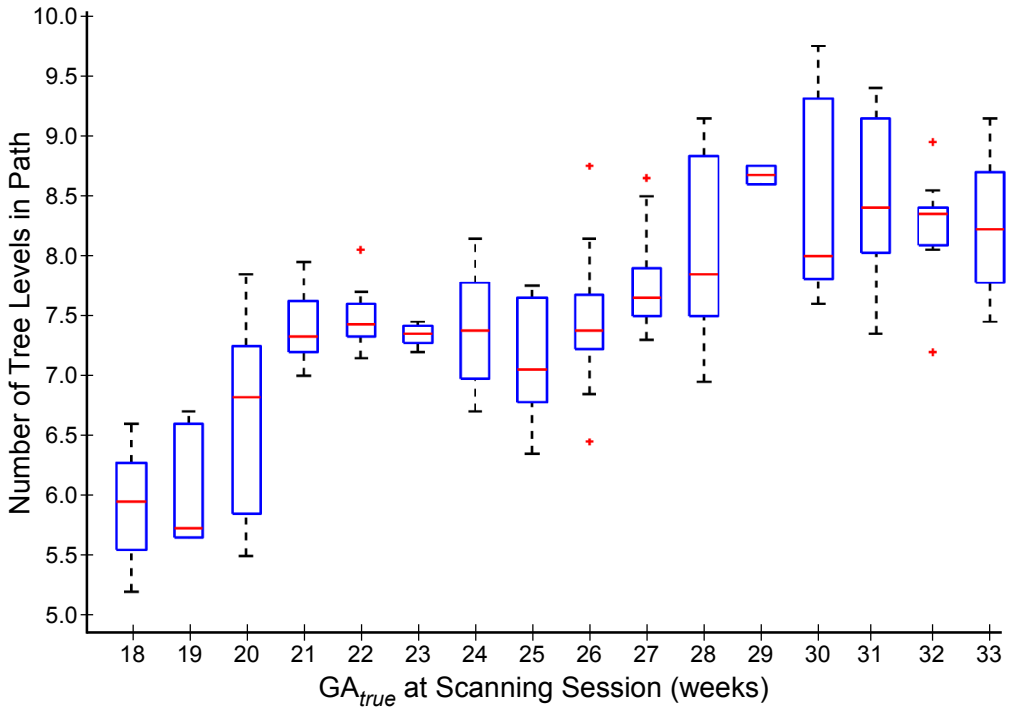
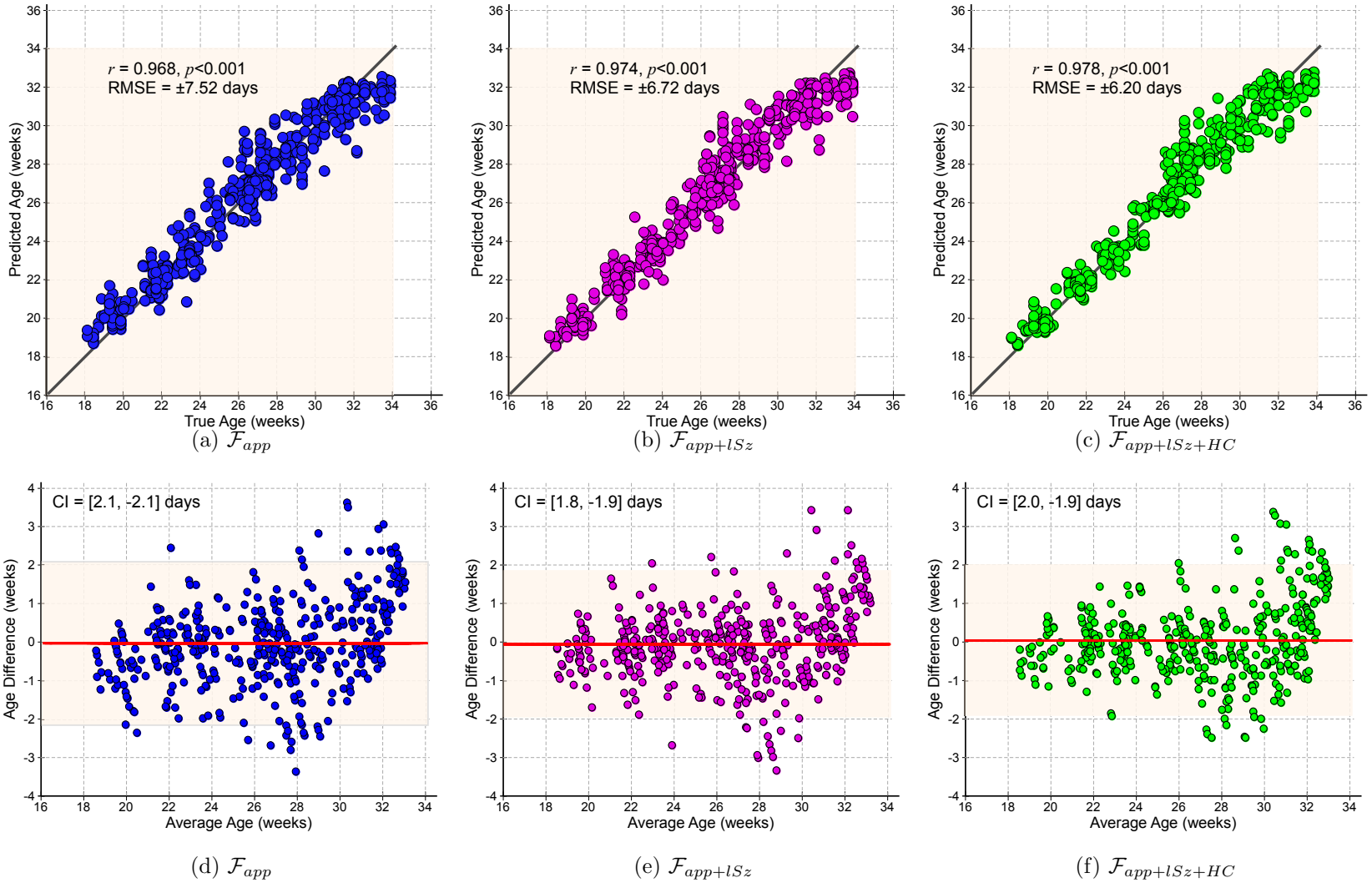


Figure 4.17: **Tree traversal path length versus GA.** Box plot of the average number of tree levels in the decision forest path (from root to leaf) for the test examples at each GA. Median is denoted by a red line in each box, and the edges indicate the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles. Outliers are plotted individually.

predictions pooled across all sessions. All forest models demonstrated strong correlations between the true and estimated GAs, ranging between $r = 0.97$ and $r = 0.98$ ($p < 0.001$). Our results also show that although the predictions from the three forest models are not significantly different from one another, the forest which combines all available features ($\mathcal{F}_{app+lSz+HC}$) yields the lowest RMSE ($\pm$6.2 days) and hence the most accurate age predictions. However, it is worth highlighting that even the forest trained exclusively with anatomical appearance features ($\mathcal{F}_{app}$) was capable of generating predictions with a high correlation between predicted and true GAs ($r = 0.97, p < 0.001$) and root-mean squared error less than 7.52 days. This indicates that the model is capable of describing GA based on the existing patterns of maturation. Importantly, it suggests the existence of age-dependent brain maturation patterns in US images of the fetal brain.

As shown in Figure 4.19, we exemplarily compared age predictions in eight pre-defined age groups. We found that the estimated ages predicted by the $\mathcal{F}_{app+lSz+HC}$ forest model differed between neighbouring age groups ($F = 20.68,\ p < 0.001$), with post-hoc t-tests revealing that the groups are significantly different from one another ($p < 0.005$ between groups 30-32 GW and 32-34 GW, $p < 0.001$ for all others). Even when the predictions were achieved with the brain-only forest model ($\mathcal{F}_{app}$), the neighbouring age groups were significantly different ($F = 20.68,\ p < 0.001$), supporting the idea that there exist age-specific, identifiable brain maturation patterns.

Table 4.3 summarizes the age prediction error for the different GA groups. It is interesting to note that the largest deviation from the line of equality occurs in the third trimester, after 28 GW. This may be attributed either to increased biological variation in later gestation [2], and also to the fact that more image regions are occluded due to increasing cranial calcification [191].



(a) $\mathcal{F}_{app}$ (b) $\mathcal{F}_{app+lSz}$ (c) $\mathcal{F}_{app+lSz+HC}$

(d) $\mathcal{F}_{app}$ (e) $\mathcal{F}_{app+lSz}$ (f) $\mathcal{F}_{app+lSz+HC}$

Figure 4.18: **Regression results from cross-validation experiments.** Graphs plot regression forest predicted age versus true LMP-based age for each image in the training dataset using a leave-10-out protocol on a cohort of 448 normal fetal subjects. Shown are the age predictions achieved using forests trained with different feature sets: from left to right, (a) F_{app}, (b) $F_{app+lSz}$, (c) $F_{app+lSz+HC}$. The r value and root-mean-squared error (RMSE) are provided for each plot. The rectangular overlay indicates the age range of data included in this work (i.e. 18^{+0} to 33^{+6} GW).

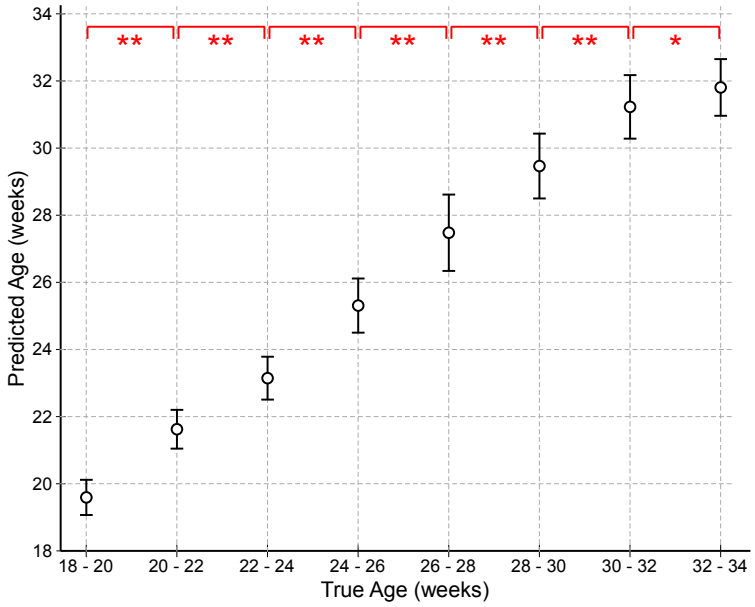


Figure 4.19: **Age group differentiation using $\mathcal{F}_{app+lSz+HC}$** Mean and standard deviation of estimated ages in eight pre-defined gestational age groups. Post-hoc t-tests resulted in significant differences between all groups (red lines: one asterisk $p < 0.005$, two asterisks $p < 0.001$)

We also plotted curves denoting the error margin of the GA prediction (δ_w) against GA, shown in Figure 4.21. The different types of forests were applied to a healthy dataset and each was compared to GA_{true} on the basis of δ_w. Although the forests were set to train to a maximum depth of $d_{max} = 15$, none of the trees reached this depth during training, and optimal results were achieved at $d = 10$ for $\mathcal{F}_{app}$, and $d = 12$ for both $\mathcal{F}_{app+lSz}$ and $\mathcal{F}_{app+lSz+HC}$. To compute the δ_w, the GA predictions (with respect to the line of equality) were separated into two groups: over-estimations made by the model $\mathbf{R}^+ \in \mathbb{R}^{2\times N_+}$, and the under-estimations $\mathbf{R}^- \in \mathbb{R}^{2\times N_-}$ in GA prediction. Quadratic fit functions were generated to model each of the positive (ζ^+) and negative (ζ^-) bounds of the GA predictions. The δ_w for each GA prediction model was defined as the absolute difference between the fit functions for the upper and lower centiles, $\delta_w = |\zeta^+ - \zeta^-|$. We illustrate this procedure in Figure 4.20.

A good age prediction model would demonstrate low δ_w values, indicating

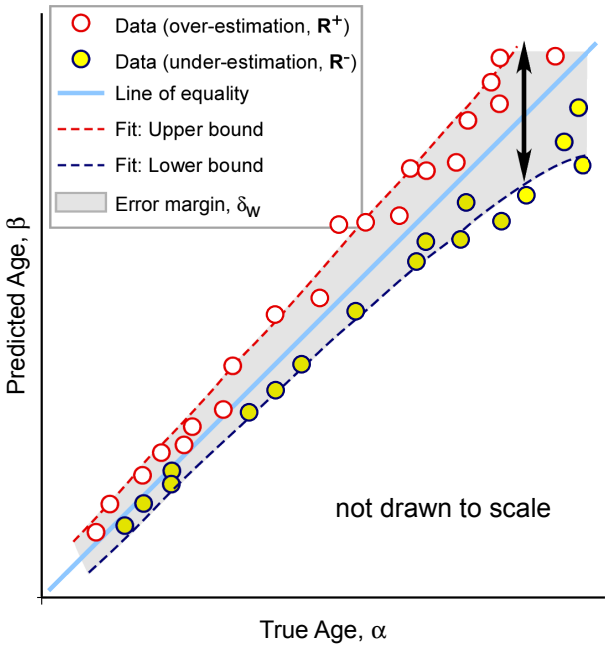


Figure 4.20: **Error margin of GA predictions,** δ_w. Illustration of the ω_{CI}. The positive and negative GA predictions (relative to the line of equality) are marked as red and yellow circles, respectively. Quadratic fitting functions approximating the positive (ζ^+, dashed red line) and negative (ζ^-, dashed blue line) bounds of the confidence interval are shown, and the difference between these functions denotes the value at δ_w when sampled at each age value (gray region).

a narrow error range (i.e. narrow spread). According to Figure 4.21, all predictors produce the lowest errors in the second trimester, and error increases steadily with GA. However, the error margins differed between predictors and clinical HC across the entire GA range ($F = 1550.5,\ p < 0.0001$), as did the rate of error increase. For instance, the δ_w's of both $\mathcal{F}_{app}$ and $\mathcal{F}_{app+lSz}$ increase almost linearly at a rate of +0.226 days/GW and +0.215 days/GW, respectively. The error margin of the clinical HC predictions displays a sharp quadratic increase such that the error nearly doubles from 18^{+0} to 33^{+6} GW, rendering the GA predictions in the third trimester approximately 10 days more erroneous. On the other hand, the forest containing all features ($\mathcal{F}_{app+lSz+HC}$) produces the lowest errors in the second trimester, but mimics the behaviour of clinical HC predictions in the third trimester, as revealed by insignificant dif-

ferences between their error margins of ($\mathcal{F}_{app+lSz+HC}$) and clinical predictions ($p = 0.890$).

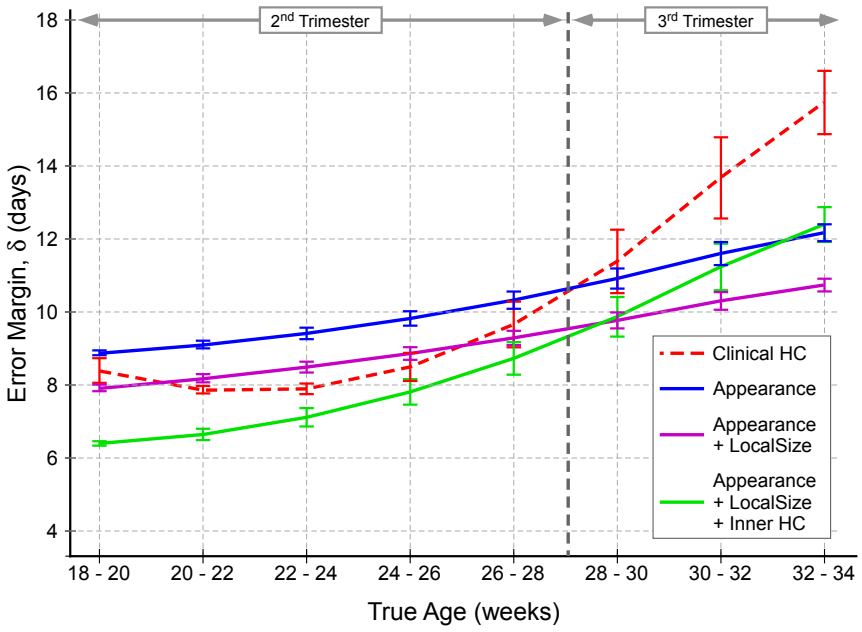


Figure 4.21: **GA prediction model comparison.** Error of GA predictions, δ_w, expressed in days, for the best current clinical method (dashed red line), and for each trained forest model (solid lines) during the second and third trimesters of pregnancy.

We also observed that model-based predictions improved with each feature set that was incorporated into the feature vector. Specifically, at 18 GW, $\mathcal{F}_{app}$ yields the widest error margin ($\delta_w = 8.2$ days) whilst $\mathcal{F}_{app+lSz+HC}$ has the narrowest ($\delta_w = 6.2$ days). Although this is not strictly the case in the third trimester, $\mathcal{F}_{app+lSz}$ continues to yield lower GA errors than $\mathcal{F}_{app}$. Nevertheless, this experiment identifies forest $\mathcal{F}_{app+lSz+HC}$ as yielding the most accurate age estimates overall, and hence the highest predictive power of the GA estimation model. In addition, the fact that the error margins of model-based predictors were always narrower than that of clinical HC, even when trained exclusively with appearance-based features. This demonstrates the ability of the model to outperform the best current clinical method for the entire GA range considered

in this work.

4.7 Conclusion

This chapter presents and validates a semi-automated learning-based framework for discovering age-related sonographic patterns in the images and linking them to neurodevelopmental maturation. The model benefits from a surface manifold representation of the fetal skull which allows for fast and efficient sampling of anatomically-corresponding brain regions to achieve like-for-like structural comparison of different developmental stages and serves as a skull-stripping tool (presented in Chapter 3). We demonstrated that the model is capable of characterizing neuroanatomical appearance both spatially and temporally, modelling GA as a continuous variable from 18^{+0} to 33^{+6} GW, capturing the natural variation existing in a healthy fetal population over an age range of active brain maturation.

Clinically relevant metadata (i.e. fetal head circumference) was provided as input to the machine learning framework, and in addition, we extended canonical features sets (e.g. Haar-like features [159]) to capture structural changes within the fetal brain. The results indicate that features which inform on cranial (and hence brain) size are selected as the first tests for discriminating between age groups, but structural image appearance features inform on finer age-related differences. This suggests that the additional information about structural brain development reduces errors in GA estimation and improves the confidence of predictions throughout the entire age range.

Validation of the model revealed that *a*) the model will generalize to an independent dataset; and *b*) that feature sets are consistently selected by different forests, hence indicating that the model represents a stable solution for

GA estimation. It is worth noting that due to the visualization of only one cerebral hemisphere in every given brain US image, each forest was trained *separately* to be applied to its respective hemisphere. That is, if only the left cerebral hemisphere is clearly discernible in a test image, age prediction will be obtained by applying the forest trained using only left-hemisphere images in order to ensure developmental and anatomical hemispheric likeness between training and testing data. We found that regardless of which cerebral hemisphere is observable in the US image and thus used for training the regression forest, an accurate age prediction can be achieved using our model, reflecting its ability to capture developmental symmetry.

In the next chapters, the presented model will be applied to independent clinical datasets to demonstrate its predictive power and additional information which can be automatically extracted by the algorithm. Specifically, Chapter 5 analyses the ability of the model the extract regions of active neuroanatomical development in the gestational period, and Chapter 6 demonstrates the generalizability of the algorithm in predicting GA in diverse fetal populations.

5

Automated Discovery of Discriminative Neurodevelopmental Landmarks

A key advantage of random forests (over other algorithms such as neural networks) is in their ease of interpretability [160]. Specifically, it is possible to examine each node in the trained forest to determine what the forest has learned and motivate why it may have learned it. For instance, in the task of age estimation using a regression forest trained with appearance-based features, each node represents salient *anatomical landmarks* which discriminate between GA groups. By virtue of being selected during the training process, these

anatomical regions have powerful age-discriminating ability and may represent biomarkers of fetal brain development which are currently overlooked by traditional US-based dating methods. This chapter presents a qualitative analysis of these automatically selected regions to *a*) identify key neurodevelopmental landmarks during the fetal period, and *b*) examine how the selection of these regions relates to GA.

5.1 Introduction

The aim of image-based prediction is to achieve high performance accuracy, so the goal of the training process is to minimize a cost function which favours regions displaying anatomical variates which relate to low prediction errors. In doing so, the algorithm is inherently learning to identify the image voxel information yielding the highest predictive power. Thus, it is expected that the model automatically identifies the anatomical regions of most relevance to the prediction task, and uses *only* those regions to make predictions on a new unseen image.

The predictive power of image-based prediction models is derived from *simultaneously* considering all image voxels for the task. That is, by using many voxel-level measurements or features which would be weakly predictive if considered on their own, the predictive models combine them to achieve high prediction accuracy. This differs from traditional "mapping" approaches such as voxel- or deformation-based morphometry [192] or cortical thickness analysis [193] which also relate image content to clinically relevant information by generating maps of the affected anatomical areas but are limited in that they rely on *a priori* information and generally consider each region independently. When using a learning-based approach, we capitalize on all existing voxel infor-

mation and as a result, it may be possible to obtain contributions from several anatomical regions to achieve a holistic prediction.

In this chapter, we interpret the complex patterns of association between anatomical image regions which the model has learnt for the age prediction task. Interpreting the patterns of anatomical regions selected by the model in a manner that is intuitive and biologically informative is a challenge particularly if it appears that *all* regions contribute to the prediction, or if the regions contributing the most to the predictions are randomly scattered within the image space. However, in medical US image data, it is expected that not all image locations may be equally relevant to the age prediction task. So, to facilitate the interpretation of region selection results, we exploit the parametric surface of the brain presented in Chapter 3 which provides a spatial structure of the brain image data. We examine the spatially-connected surface vertices and their corresponding image patches from which the appearance-based features were extracted, and quantify the frequency with which the algorithm selected these regions to generate a model which best represents the training data. Through this analysis, we gain insight into the regions demonstrating the important anatomical variations which drive the accurate GA predictions achieved in § 4.6.3.

5.2 Surface-Based Analysis of Anatomical Development

In order to assess the relative importance of each cortical region, we first evaluated the frequency with which each vertex location on the surface domain was selected by the model. To achieve this, we explored the nodes in the tree traversal paths of the training data used in the cross-validation experiments

discussed in § 4.6. In doing so, we took into consideration the generalization and stability of the regression forest model in identifying relevant anatomical regions. For all the experiments presented in this chapter, the data consisted of $N_L = 198$ and $N_R = 250$ images with visible left and right cerebral hemispheres, respectively, ranging from 18^{+0} to 33^{+6} GW (mean $\pm$ stdev: 26.84 ± 4.20 GW for the left hemisphere; 26.73 ± 4.18 GW for the right hemisphere) acquired from the INTERGROWTH-21$^{\text{st}}$ dataset of normal fetal subjects.

In each of the the cross-validation sessions, the regression forest comprised of $T = 40$ trees trained to a maximum depth of $d_{max} = 10$. A total of $m = 500$ features were sampled at each tree node in order to ensure that the model has searched a large enough feature space to select only the most age-discriminating feature relevant for GA prediction. Since the aim of this study was to identify anatomical biomarkers of brain development, size-based features were not provided to the algorithm during the training process, and the forests were trained exclusively with appearance-based features (i.e. $\mathcal{F}_{app}$). This section summarizes the results of the frequency with which the surface vertices were automatically selected by the algorithm. Firstly, we demonstrate the vertex selection frequency with respect to the locations on the surface domain, and further examine the selected *intracranial* structures within the image space. These regions are then analysed with respect to GA in § 5.3 to investigate whether the algorithm identified and exploited a longitudinal pattern of brain development during the fetal period.

5.2.1 Automated Selection of Relevant Surface Vertices

In a regression forest classifier, the decision-making process occurs in a hierarchical fashion as the data traverses the decision trees from the nodes at shallow levels down to the leaf nodes at deeper levels. In this process, it is likely that

more general binary tests are posed to the data at shallow tree nodes (associated with higher prediction variance), and more GA group-specific discriminating tests are performed at deeper tree nodes, which are associated with lower age prediction variance (see Figure 4.8). In this section, we assess the relative importance of different brain regions with respect to forest level (d), and hence their hierarchical relevance in the decision-making process. We also investigate the spatial distribution of landmark selection.

For each training session, only a small subset of cranial surface points were made available for selection to forests during the training process, as shown in Figure 5.1 for the right hemisphere. This matrix of points $\mathbf{P}^h \in \mathbb{R}^{3\times 61}$ is defined from a sparse evaluation of the cranial surface domain. We selected regularly-spaced vertices, positioned in a manner that allows the cuboidal feature banks to capture information within the entire brain space whilst reducing the redundancy of neighbouring image patches which may contain similar appearance/statistical information. Since these vertices are on a pre-defined coordinate system, we assume that the vertices will correspond to similar anatomical structures in the deformed surfaces of different subjects, allowing us to then compare the behaviour of anatomical region selection on the basis of the surface domain.

Distribution of the relevant vertex selection

In order to assess the stability and recurrence of vertex selection, we plotted the number of surface vertices (i.e. surrogates for features) which were selected by a majority of the cross-validation sessions. Since our implementation applied random data partitioning (through the leave-10-out protocol) and bagging of the training set used for constructing each cross-validation forest, it is expected that the trees in each forest will be decorrelated enough for the model to be

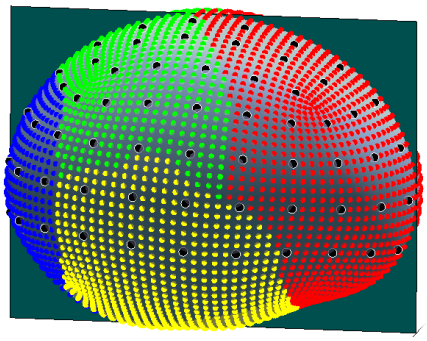


Figure 5.1: **Available surface vertices.** Illustration of the $N_{\mathbf{P}^h} = 61$ points on the surface domain which were made available during the forest training process (shown as black circles). For visualisation purposes, the total number of vertices evaluated on the surface domain are also displayed, and annotated by the colour of the cerebral lobe on which they are found: frontal (red), parietal (green), temporal (yellow), occipital (blue).

generalizable, and that the trained forests will in turn be dissimilar enough to capitalize on the most informative image features to achieve high prediction accuracy. Motivated by this intuition, if the independently trained forests are generally selecting similar/overlapping features, it can be concluded that these commonly-selected features reflect the most age-discriminating anatomies present in US images.

Figure 5.2 shows a histogram of the percentage of the available cranial surface vertices $\mathbf{P}^h$ (out of a total of $N_{\mathbf{P}^h} = 61$ vertices) which were selected by at least 75% of the cross-validation sessions for each forest level, d, normalized by the total number of nodes present at the given forest level. Due to the presence of hemispheric occlusions resulting from reverberation artefacts, the forests to be applied to each cerebral hemisphere were trained exclusively on images in which the hemisphere was distal to the US probe and observable; consequently, their results were analysed separately in this section.

From the figure, it is evident that although the forests for the left and right hemispheres were trained on separate independent datasets, their overall vertex selection results exhibited similar behaviour. Specifically, it is clear that

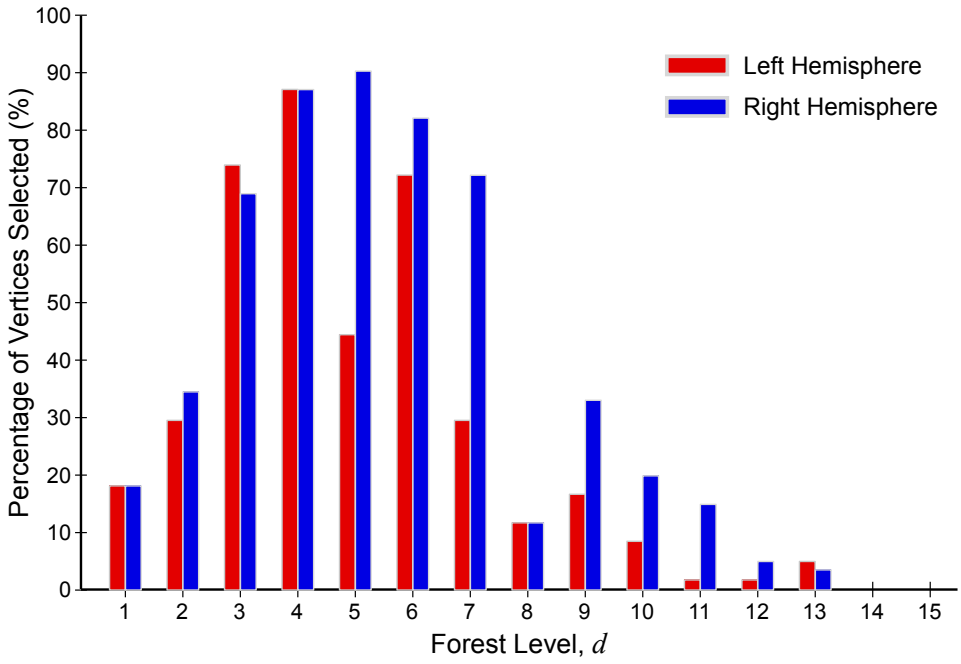


Figure 5.2: **Percentage of surface vertices selected versus forest level.** The percentage of vertices which were selected by at least 75% of the forests from the cross-validation sessions trained on data from the left (19 sessions) and right (25 sessions) hemispheres.

at $d = 1$ only a few vertices ($< 20\%$) were selected by a majority of the forests from different training sessions. As the data traverses the trees to the mid-range levels (i.e. $d = 3-6$), the number of selected vertices gradually increases to more than 80% of $N_{\mathbf{P}^h}$. Since nearly all available points are selected at these forest levels, it suggests that the feature locations are dispersed around the brain space. This increase in sampled points from $d = 1$ to $d = 6$ may suggest that at mid-range levels the forest is likely to query different regions which may be of relevance to some GAs and not others. That is, at shallow nodes, the binary tests are coarsely splitting the data into subsets containing a wider spread of ages (i.e. high variance) whereas the nodes in mid-range levels receive these subsets of data and split them into even more age-specific subsets (i.e. lower variance). Given that there are more nodes present at $d = 3 - 6$, each with progressively more specific age-groups to discriminate, it

is likely that the optimally-selected binary tests show that *most* brain features are of relevance to the GA prediction task, but perhaps different age groups are more powerfully discriminated by specific regions which are not as relevant to other ages. Thus, in order to discriminate between all the age groups in the full range of the training set, the model has to query all regions in the brain space. However, at deep levels (i.e. $d \geq 9$), the number of relevant vertices gradually decreases to below 20%. It is at these forest levels that the features are selected to distinguish between even more age-specific differences. Thus, we can speculate that, though fewer, these highly age-specific features are also dispersed at the deep levels.

This behaviour of vertex selection frequency can only be confirmed by analysing the spatial distribution of the relevant features mapped onto the cranial surface. The next section explores this.

Spatial distribution of the relevant vertices

To complement Figure 5.2 and the above discussion, Figure 5.3 shows a cortical surface representation of the frequency with which each vertex location was selected across the cross-validation sessions at levels $d = 1, 3, 6, 9$ for forests trained on images of each hemisphere. The colour map corresponds to surface vertices which were consistently selected by a majority of the cross-validation forests to have reached a given level d (coloured in red), and the vertices which were least relevant to the GA prediction task at the same forest level (coloured in blue). Transparent vertices had a zero contribution to the prediction. The vertices which were consistently selected by different forests (i.e. in red) are likely to correspond to developmentally informative anatomical regions.

In agreement with the above discussion, the figure shows that at $d = 1$ fewer than half of the vertices are selected by a majority of the cross-validation

forests (i.e. $\geq 75\%$ of forests). It is noteworthy that these vertices are generally clustered around a specific area: the boundary between the temporal and frontal lobes, and between the frontal and parietal lobes. This cluster is likely to correspond to the area around the Sylvian fissure and the superior temporal sulcus, both of which are known to undergo distinct developmental changes during the fetal period, and are clearly visible in US images (see Figure 2.8; [40]). Interestingly, these clustered vertices correspond to regions which are directly incident to the US probe angle used to acquire the images used in this study. It is possible that the model queries these regions at the start of the decision-making process either because *a*) the acquisition protocol renders it the most clearly discernible anatomical area within the imaged brain space; or *b*) the region undergoes such distinct structural changes throughout the GA range of interest that it provides the most developmentally-informative patterns that quickly and effectively guide the age prediction task.

Figure 5.3 also confirms that at the mid-range levels ($d = 3, 6$), almost all vertices are selected by a majority of the forests, and that they are all dispersed around the brain space. This supports the notion that all brain regions need to be queried in order to examine differences between subjects within an age group. At the deep levels (e.g. $d = 9$), most vertices had a zero contribution or were only occasionally selected by the forests (i.e. $< 50\%$ selection). Although there is only a small number of vertices selected at $d = 9$ just as there are at $d = 1$ (as also shown in Figure 5.2), the key difference is that at $d = 9$ the vertices are not arranged in a cluster but are rather more dispersed. This suggests that the selected vertices are only of interest to very age-specific datasets (with low GA variance) to distinguish finer differences between subjects within narrow GA bands, and may not be relevant to other GA groups.

It is also worth noting that Figure 4.17 in Chapter 4 showed that depending

on GA not all images will reach leaf nodes at deep forest levels before a prediction has been achieved. In particular, most images from fetuses younger than 28 GW will not reach $d = 9$ during the tree traversal before they reach a leaf node. Thus, it is also possible that the regions selected in the forests at deep forest levels correspond only to the anatomical regions relevant to discriminating more complex anatomical differences associated with advanced GAs (i.e. older than 28 GW).

In summary, the fact that the model selected a small cluster of regions at shallow forest levels suggests that at the onset of the decision-making process, there is only a low number of key regions which will coarsely discriminate between developmental groups. The model aims to verify this cluster prior to examining other regions within the entire brain space which will inform on finer anatomical differences between subjects within a progressively narrower GA band at deeper forest levels. This observation clearly indicates that rather than randomly querying regions in an unstratified manner, the model exhibits hierarchical behaviour in the selection of relevant regions for GA prediction from US images.

5.3 Discovery of Relevant Brain Regions and Maturation Maps

Having established a method which examines the entire image volume of the cerebral hemisphere, in order to fully understand the trained model for predicting GA and neurodevelopmental maturation, we must assess and visualise the key regions selected by the model *within* the brain image voxels. Following from the above discussion, it is expected that only some regions within the entire 3D cerebral volume will accurately report on brain maturation, while

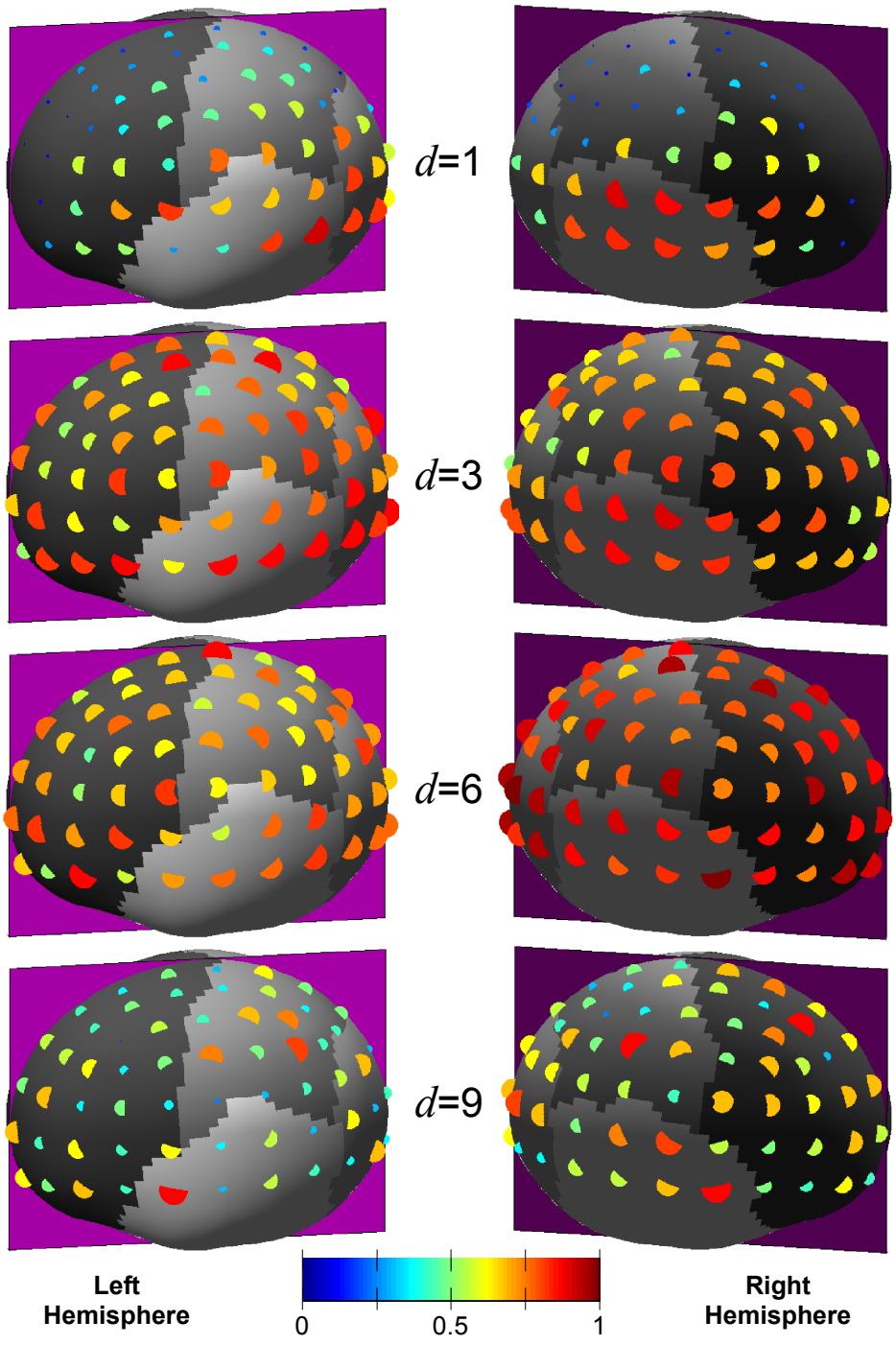


Figure 5.3: **Relevant surface vertices for GA prediction task.** The frequency at which each surface vertex had a non-zero contribution in the regression forest model across all cross-validation sessions for the left (19 sessions) and right (25 sessions) hemispheres. Transparent regions correspond to vertex locations which had no contribution to the age prediction task.

low-contrast or highly variable regions could obscure this task. We hypothesized that the unguided training framework could by itself identify the relevant brain regions that provide the best age discrimination.

Similar to the works of Sabuncu and Van Leemput [172], Toews et al [121], and Konukoglu et al [175] on MR brain images, we examined the most age-discriminating brain regions selected by the algorithm. To achieve this, heat maps were generated to demonstrate the image appearance features selected in each level of a forest. To generate these maps, the nodes in each forest level d were parsed, and the associated feature was used to populate the voxels of the heat map. A heat map at forest level d associated with the features selected at given GA a is denoted as $\mathbf{B}_a^d$. Specifically, each identified feature $f_j^d(\mathbf{x}_i, s_i)$ at node j of level d corresponds to a cuboidal patch defined by a location $\mathbf{x}_i$, a scale s_i which informs on the scalar side length of the cuboid, and the image measurements computed within the given region (selected from the feature banks described in § 4.2.1). The image voxels of the identified feature were incremented by a value of 1 each time they were sampled by a node. The resulting image was then normalized by the number of nodes present at level d, i.e. N_d [1]. A composite heat map $\mathbf{H}^d$ was generated by averaging the level-specific heat maps for all k test images left out of the cross-validation sessions (i.e. accumulating all 10 images from each session), such that it was defined as follows:

$$\mathbf{H}^d = \frac{1}{N_k} \sum_{k=1}^{N_k} \mathbf{B}_{\forall a}^d(\mathbf{I}_k) \tag{5.1}$$

Figure 5.4 demonstrates the composite heat maps for four different levels of the forest, showing the locations of age-discriminating image features selected

[1] Note that during testing, the image traverses each forest level only once in each tree, so N_d is less than or equal to the number of trees in the forest T

by the model. Our investigation showed that the algorithm consistently selected a small number of key age-discriminating regions. These included:

a) regions around the midsagittal plane such as the calcarine sulcus, corpus callosum, cingulate sulcus, and parieto-occipital fissure at $d = 1, 3, 6$, and thalami at $d = 9$;

a) the posterior and anterior ventricles at $d = 1, 9$; and

a) sulcal areas such as the Sylvian fissure ($d = 1, 3, 6$) and central sulcus ($d = 6, 9$).

These are regions which experience substantial change over the gestational period [8, 40], reflecting the ability of the model to query developmentally-informative anatomies during the GA prediction process. We also observed that at $d = 1$ and $d = 9$, the model queried fewer brain regions than in the mid-range levels $d = 3, 6$. For instance, at $d = 6$ the model queried several regions (e.g. Sylvian fissure, cingulate sulcus, central sulcus, callossal sulcus, parieto-occipital fissure, calcarine sulcus), while it focused primarily on the Sylvian fissure and calcarine sulcus at $d = 1$, and on the thalami and cingulate sulcus at $d = 9$. Given that the nodes at deeper levels have lower variance values, it indicates that the binary tests associated with these nodes improve the confidence of GA estimates, highlighting finer differences between age-related sonographic appearance. In theory, it might be expected that the regions queried by the model would be spatially distributed in the brain space. However, in practice, we found that the same key anatomical regions were consistently selected by different forests at the same forest levels, identifying them as salient landmarks for GA estimation and confirming the hierarchical nature of the prediction process.

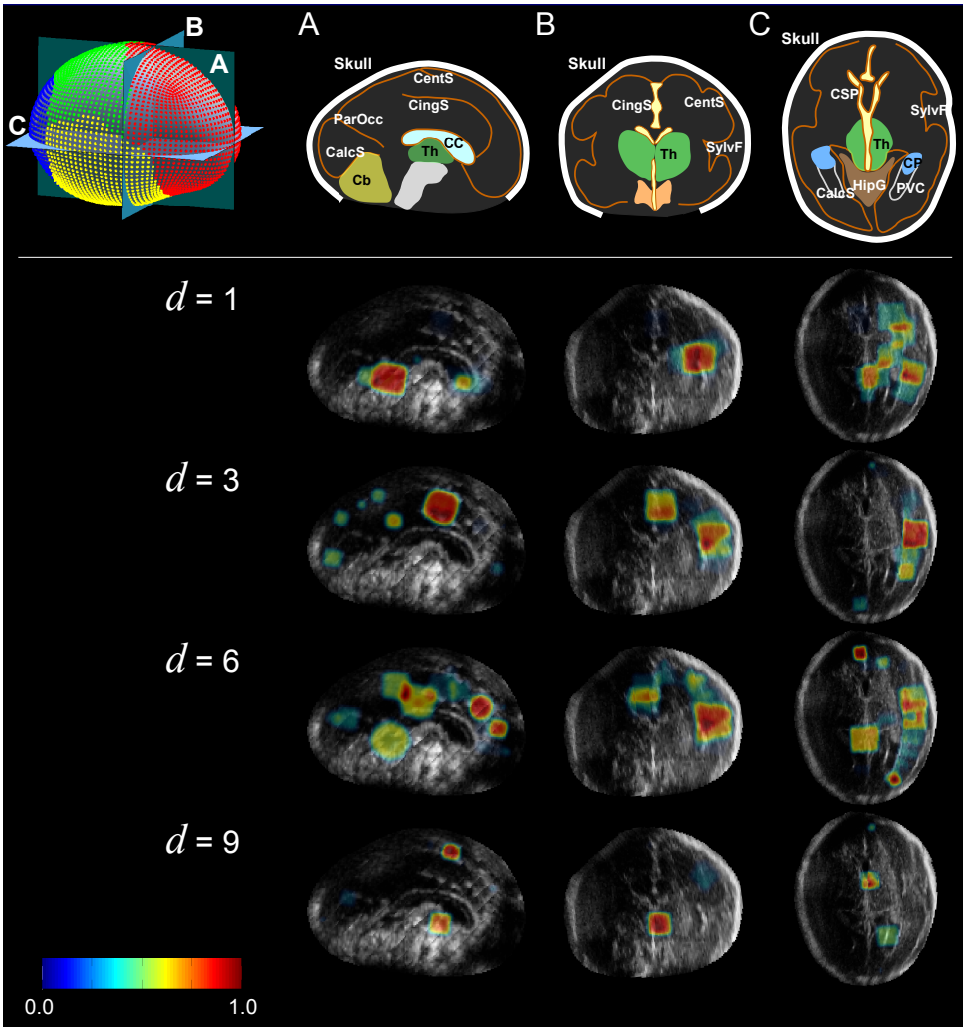


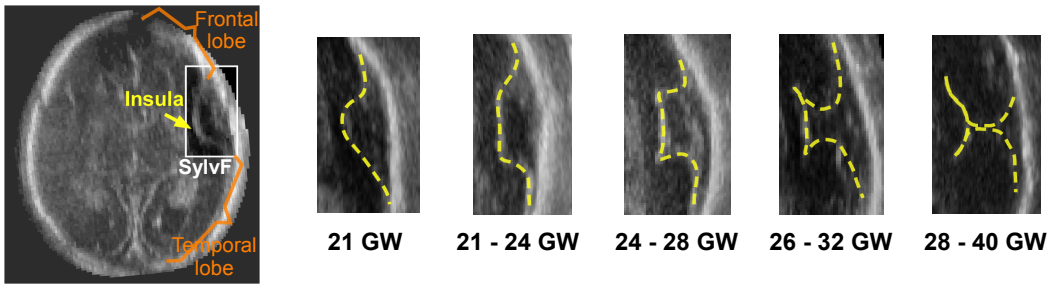
Figure 5.4: **GA-discriminating brain regions.** Visual example of the feature locations selected at four different levels of the forest ($d = 1, 3, 6, 9$) superimposed on sagittal (left column), coronal (centre column), and axial (right column) slices intersecting at the middle of the brain surface space. The heat map reflects the relative feature importance, such that orange/red regions correspond to brain regions frequently selected by the model to estimate GA from US images. Colour scale indicates the weight. *sylF*: Sylvian Fissure, *calS*: Callossal Sulcus, *centS*: Central Sulcus, *parOcc*: Parieto-occipital Fissure, *cingS*: Cingulate Sulcus, *calcF*: Calcarine Fissure, *posV*: Posterior Ventricle, *antV*: Anterior Ventricle, *thal*: Thalami

5.3.1 Focus on the Sylvian Fissure

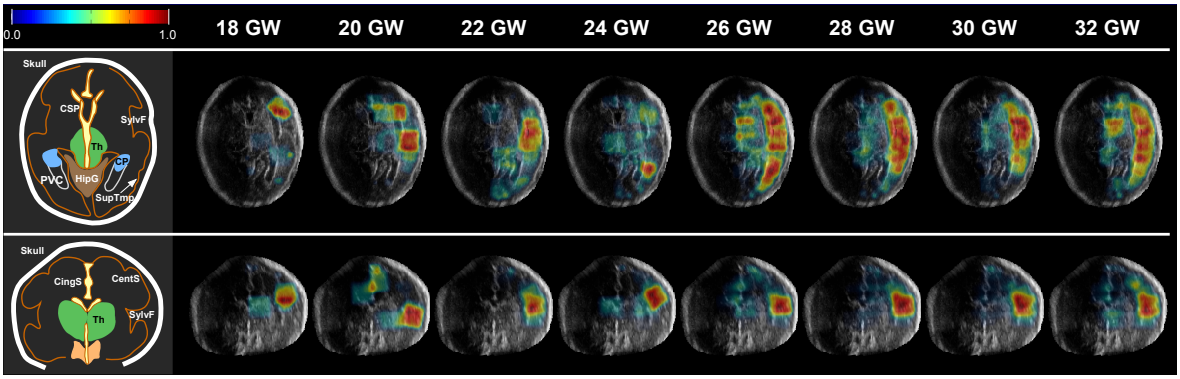
Realising that the Sylvian fissure is consistently selected by the model at different forest levels and has been reported in the literature as being a structure

which reliably informs on gyrification states [8, 89, 90, 40], we further investigated the behaviour of the model in this region. As shown in Figure 5.5a, the dynamic process of operculization is observable in US images as a progression of the Sylvian fissure from its primordial shallow indent on the cortical surface into an insula/operculum pinched by the overgrown frontal and temporal lobes. This characteristic pattern of morphological changes mainly occurs in the GA range considered in this work, so it is expected that the model would utilize information from this critical region in the GA prediction. Also, it is likely that the model is learning the *age-specific* appearance of the operculization process in this region. To assess this, we analysed axial and coronal slices of the heat maps taken at the level of the Sylvian fissure at all the different GAs, displayed on the same reference US image for ease of comparison. Following from the discussion on forest level d and its relation to prediction-refining region selection in § 5.3, we decided to incorporate only the information of brain regions selected at the mid-range forest levels: i.e. $d = 5-9$. This allows us to examine only the regions which were relevant to refining the GA predictions within narrower GA ranges, as opposed to the regions which discriminate between broad GA ranges (as in shallow forest levels) or in too specific age groups (as in deep levels). To achieve this, we computed the heat maps of feature selection for images at a given GA for each level of interest $d = 5 - 9$ separately, and averaged these heat maps to obtain the overall map of relevant brain regions for the given GA. Figure 5.5b shows that the model indeed exploits information from the operculization process by consistently selecting a feature in the Sylvian fissure region for all GAs with only slight variations associated with age. These are discussed below.

18 GW: The model selects a region in the frontal lobe near the Sylvian fissure region (or insula), but not quite at its standard location. This may either



(a) Operculization of the Sylvian fissure



(b) Selection frequency of Sylvian fissure at different GAs

Figure 5.5: **Spatiotemporal selection frequency of the Sylvian fissure.** (a) Illustration of the operculization process of the Sylvian fissure during the gestational period as observable by US imaging as reported in [40]. Note that the angle between the insula and the overgrowing frontal and temporal lobes becomes progressively more pronounced with advancing age. (b) Heat maps illustrating the feature locations selected at different GAs superimposed on axial (top row) and coronal slices (bottom row) taken at the level of the Sylvian fissure. The heat map corresponds to the relative feature importance, such that orange/red regions correspond to brain regions frequently selected by the model to estimate GA from US images. Colour scale indicates the weight. *sylF*: Sylvian fissure, *calS*: Callossal sulcus, *centS*: Central sulcus, *parOcc*: Parieto-occipital fissure, *cingS*: Cingulate sulcus, *calcF*: Calcarine fissure, *posV*: Posterior ventricle, *antV*: Anterior ventricle, *supTemp*: Superior temporal sulcus, *thal*: Thalami

be due to the fact that at this GA, the Sylvian fissure is not yet clearly indented and may not be distinguishable from the inner skull in all images. However, the position of the identified feature is near the back of the frontal lobe which aligns with reports of operculization occurring in a frontal-to-temporal direction. That is, the model may have identified the initial folding of the Sylvian fissure, starting around the rear of the frontal

lobe before progressing towards the temporal lobe.

20 GW: At this GA, the model has identified two key features which correspond to an underlying increase in the angles of folding at which both the frontal and temporal lobes are overriding the insula.

22 GW: At this GA, the model has identified a feature centred on the insula. This feature encloses all the image voxels constituting the entire Sylvian fissure.

24 GW: The model did not select the Sylvian fissure as the most salient feature. However, an interesting feature appears in the coronal slice which shows that the model identified a key region located around the central sulcus. This is an important observation because the central sulcus has been reported to appear on the cortical surface at approximately 24 GW in post-mortem [39] and MR-based findings [49, 50, 51]. Such a result indicates that the model is also capable of detecting the emergence of a new structure and its consistent presence in most images at the particular GA.

26 - 32 GW: The Sylvian fissure is frequently selected by the model, and it appears that the informative feature regions are formed in the shape of the ridges on the cortical surface (particularly the Sylvian fissure). There is also an increase in activity in the temporal lobe region, possibly due to the emergence and ensuing gyrification of the superior temporal sulcus (reported to emerge after 26 GW [39, 53, 49, 51, 8]).

Overall, the model was able to identify a feature consistently located in a stable position around the Sylvian fissure for all GAs, but also around the frontal and temporal lobes in later gestation. This suggests that the model

has exploited age-related morphological changes associated with the emergence/disappearance of structures and cortical folding in this region by adhering to the temporal schedule of the operculization process.

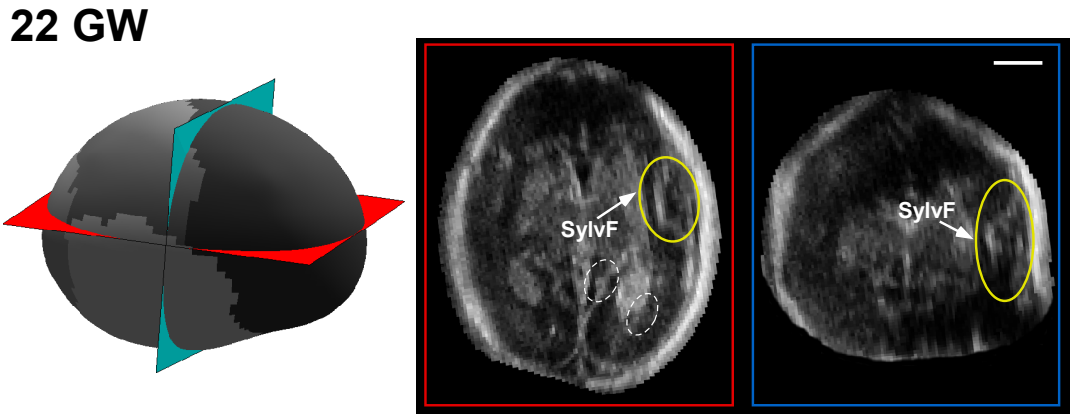
5.3.2 Age-Specific Landmarks of Neurodevelopmental Maturation

From Figure 5.5b it is evident that the model is exhibiting longitudinal behaviour by selecting regions which are most relevant to specific GAs. Realising that there are age-related differences between the features selected by the model, we probed the relevance heat maps of different GAs to identify the *age-specific* regions of highest importance to the decision-making process. Again, we examined heat maps which demonstrated only the brain regions selected by the model at $d = 5-9$. Here, we show the key anatomical landmarks at 22 GW and 32 GW, but we refer the reader to Appendix B for further illustrations of relevant landmarks spanning from 18 to 32 GW. It is important to note that for all GAs, the scales of the salient features reflects the spatial extent of the underlying anatomical structures. For instance, at 22 GW the identified feature region includes the entire Sylvian fissure (Figure 5.6a). This is the age at which the Sylvian fissure displays a fully defined 'plateau' shape when the angles between the insula and the overgrowing lobes become pronounced. At 32 GW (Figure 5.6b) the Sylvian fissure is also selected, but most interestingly, it appears that the features are also outlining the individual folds along the temporal lobes— possibly signalling the reduction in sulcal width, increase in sulcal depth, and branching of the superior temporal sulcus observed in later gestation [40]. In addition, the cingulate and callosal sulci are selected, and despite their spatial proximity, the model was capable of identifying them as *separate* neuroanatomical features of interest.

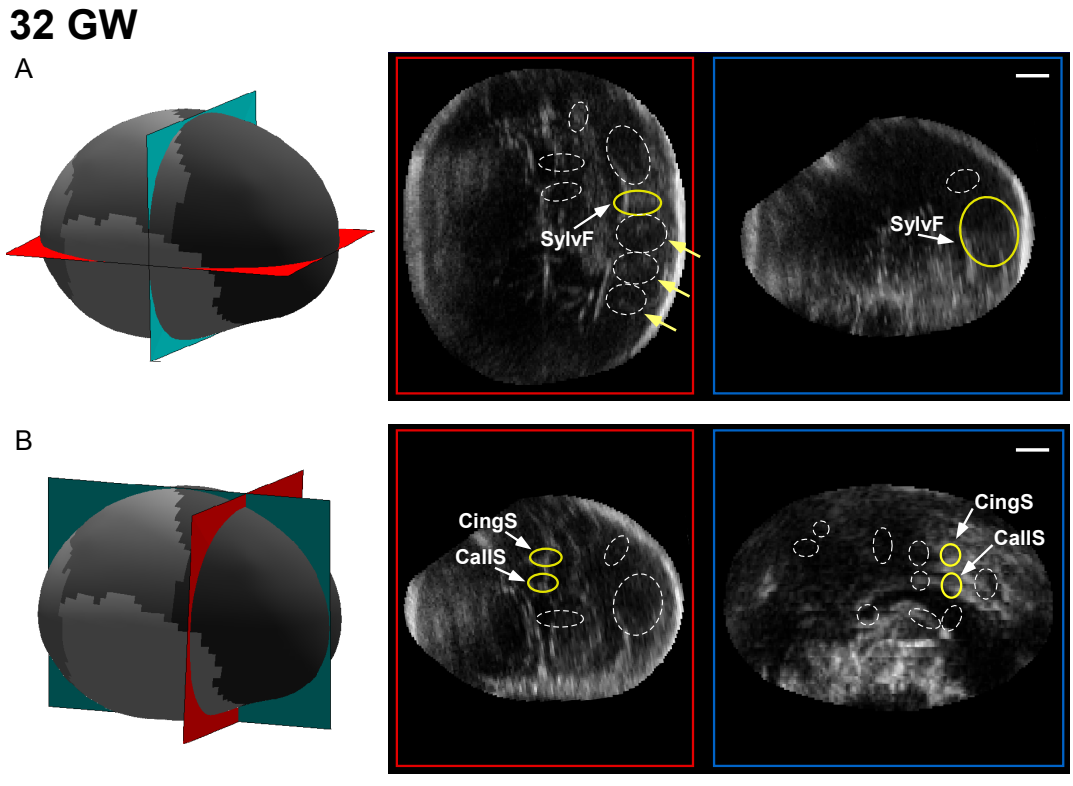
5.4 Conclusion

In this chapter, we demonstrated the feasibility of automatically learning the pattern of sonographic activity and linking it to GA. In particular, the model was able to identify relevant brain regions such as the Sylvian fissure, cingulate sulcus, and callosal sulcus as powerful image regions in the age-discrimination task. This corroborates findings in the clinical literature about these cerebral structures following a characteristic pattern of development [8, 89, 40]. We also demonstrated that the model uses information from the operculization process in decision-making. This process has been described for US images used in longitudinal observation studies of the fetal brain [8, 89, 90, 40]. The model also queried this anatomical information in a stratified, hierarchical manner, rather than randomly searching the brain space for signs of maturational progression. The fact that the model has utilized temporal information about brain development, it is conceivable that it is also capable of capitalizing on information about other underlying biological processes, provided that they are observable in US images.

In summary, we have shown that beyond GA prediction, our model is capable of identifying anatomical regions which undergo significant changes during gestation. This indicates that our GA estimation framework has the potential to provide clinically relevant information from US images which is not currently exploited. In the next chapter, we demonstrate potential clinical utility of the model on independent datasets of clinical US images.



(a)



(b)

Figure 5.6: **GA-specific developmental landmarks at 22 GW and 32 GW.** Illustration of the features extracted from 3D US images of fetuses at (a) 22 GW and (b) 32 GW. Ovals represent the locations and scales of features in the axial (red) and coronal (blue) slices. Solid yellow lines represent image features of highest importance ($\geq 80\%$ selection frequency) at the given GA and are present in all slices, and dashed white lines represent regions present within the given slice. At 22 GW key regions include: Sylvian fissure, calcarine sulcus, and rostral part of the choroid plexus. At 32 GW key regions include: Sylvian fissure, cingulate sulcus, callosal sulcus, superior temporal sulcus (indicated by three arrows). Scale bar: 1 cm.

6

Clinical Application of GA Prediction Model: Two Pilot Studies

6.1 Introduction

This chapter describes the application of the age estimation model presented in Chapter 4 in the context of two clinical pilot studies. Through our experiments, we demonstrate the potential clinical utility of the model by applying it to challenging clinical datasets. In the first application, we test the ability of the model to predict GA in a high-risk clinical population, containing fetuses from

pregnancies at risk of malnutrition and infections, and compare the result to predictions by the method used in current clinical practice. In the second pilot study, we apply the model to a set of growth-restricted fetuses and compare these to a healthy population to determine whether the algorithm is capable of detecting and characterizing intrauterine growth restriction and its effect on the developing brain.

6.2 Pilot Study 1: High-Risk Clinical Cohort

6.2.1 Experiments

For this experiment, we applied the model to 187 fetuses from the INTERBIO-21st study database[1] ranging from 18^{+0} to 33^{+6} GWs, spanning the second and early third trimesters of pregnancy. These fetuses represent a high-risk cohort of complicated pregnancies from six different ethnic populations. Notably, the subjects included in this study have different risk profiles, as some are at risk of preterm delivery or intrauterine growth restriction (IUGR) due to malnutrition and complications associated with infection such as HIV or malaria. Similar to the fetuses considered in Chapter 4, 'true age' was defined by the last menstrual period (LMP) and confirmed by first-trimester ($\leq 14^{+0}$ weeks) sonographic measurement. Figure 6.1 shows the distribution of ages in the fetal cohort included in this experiment. The 3D images of the fetal head were collected using a Philips HD9 curvilinear probe at a wave frequency of 2-5 MHz, and prior to being applied to the predictive age model, they were resampled to an isotropic resolution of $0.6 \times 0.6 \times 0.6\text{mm}^3$ and filtered with a bandpass Gaussian derivative kernel of size $\bigtriangledown G_\sigma = 4\,\text{mm}$.

[1] www.interbio21.org.uk/

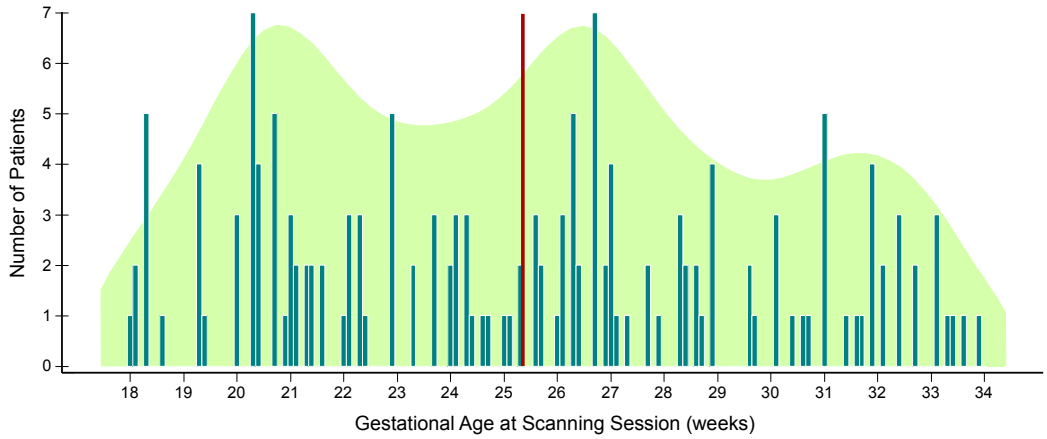


Figure 6.1: **Gestational age spread of the high-risk fetal cohort**. Bar chart of the number of subjects at each gestational age with accompanying density trace (estimated using a Gaussian kernel, $\sigma = 1$). The mean age of the cohort is shown in red.

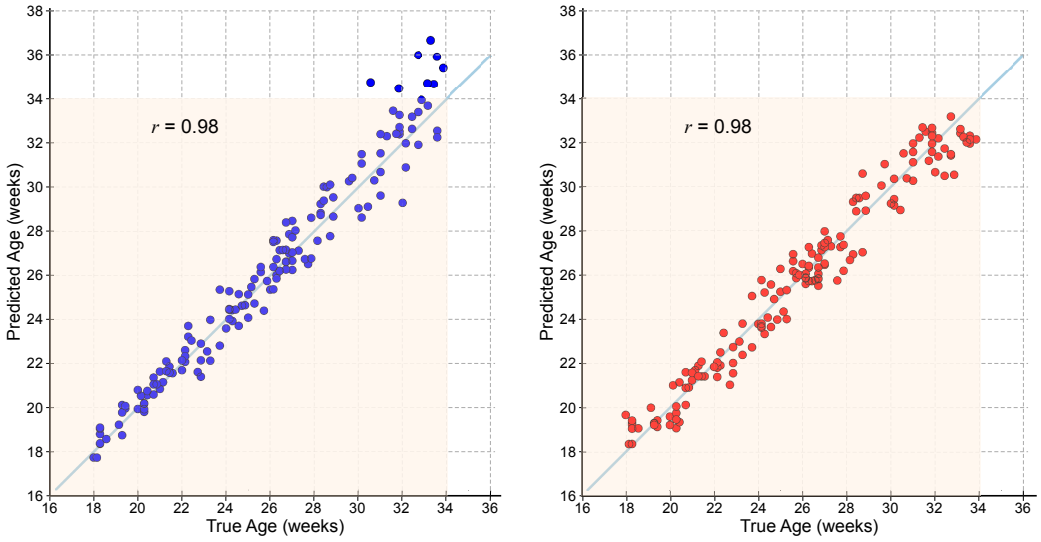
6.2.2 Results

Predicting Gestational Age

Model-based age predictions ($\mathbf{a}^{\text{pred}}$) were compared against the current best clinical method for estimating gestational age, $\mathbf{a}^{\text{clin}}$: the mean of three HC measurements regressed to population growth charts (see Figure 3.14b; [9]). Table 6.1 summarizes the GA estimation results from applying $\mathcal{F}_{app+lSz+HC}$ and the clinical HC method to the high-risk fetuses, and Figure 6.2 plots the GA predictions against true GA. As a comparison, a predictor using random guessing would result in a RMS error of ±2.76GW (approximately 19.4 days). A paired t-test reveals that the clinical and model-based prediction errors (when compared to $\mathbf{a}^{\text{true}}$) are significantly different from one another (assuming that the errors follow a normal distribution with zero mean, $p < 0.001, t = 5.77$).

Our results demonstrate that although $\mathbf{a}^{\text{clin}}$ yields the lowest prediction errors in the second trimester (7.2 days), $\mathbf{a}^{\text{pred}}$ has a lower overall error throughout the entire age range, outperforming $\mathbf{a}^{\text{clin}}$ by ±0.91 days and reducing the CI by ±1.68 days. However, the real benefit of $\mathbf{a}^{\text{pred}}$ becomes apparent in the third

trimester in which the CI is reduced by approximately $\pm\mathbf{4.56}$ days and GA predictions are improved by $\pm\mathbf{2.51}$ days. This observation is best visualized by a Bland-Altman plot (Figure 6.3), which shows that the difference between the clinical and model-based predictions increases quadratically with GA, with $\mathbf{a}^{\text{clin}}$ yielding higher predictions than $\mathbf{a}^{\text{pred}}$, particularly in the third trimester (Table 6.1, $p < 0.05$). Furthermore, visual inspection of Figure 6.2 indicates that the results from $\mathbf{a}^{\text{pred}}$ had a tighter fit to the line of equality for the entire age range, whereas results from $\mathbf{a}^{\text{clin}}$ diverged with progressing GA.



(a) Clinical HC GA Estimation (b) Model-Based GA Estimation

Figure 6.2: **GA prediction results for the high-risk dataset.** GA estimation results using (a) the best current clinical method (blue), and (b) the $\mathcal{F}_{app+lSz+HC}$ forest model (red). The rectangular overlay indicates the true age range of data included in this experiment (i.e. 18^{+0} to 33^{+6} GW).

Predicting Developmental Trajectories

In order to determine whether the model produces chronologically consistent age predictions between scanning sessions, we assessed the developmental trajectories for a subset of fetuses scanned at multiple time points in pregnancy. Specifically, we assessed whether the model is capable of predicting the elapsed

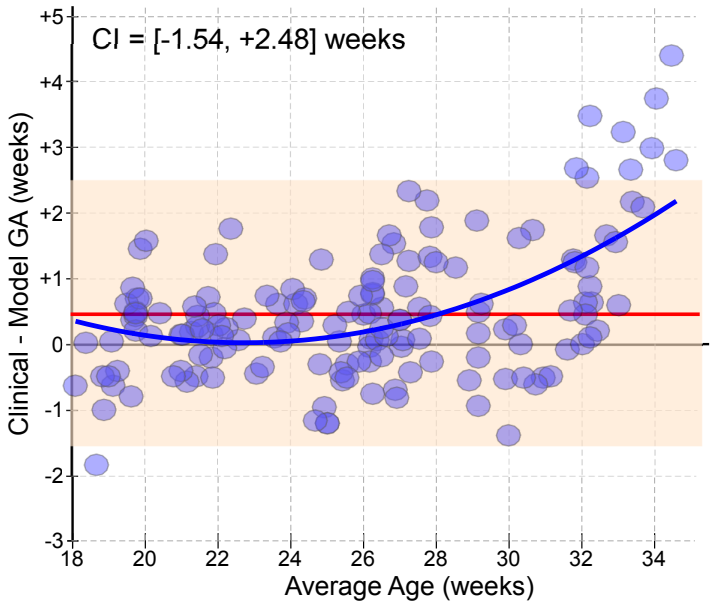


Figure 6.3: **Bland-Altman plot for clinical vs. model-based predictions**. Absolute difference in predictions, in weeks: $\mathbf{a}^{\text{clin}} - \mathbf{a}^{\text{pred}}$. The 95% confidence intervals are shaded in light red.

Table 6.1: **Performance comparison between automated model and clinical HC.** GA prediction performance measures for the clinical HC method and the best model-based predictor ($\mathcal{F}_{app+lSz+HC}$) applied to the high-risk dataset for the 2nd and 3rd trimesters of pregnancy. RMSE and confidence interval (CI) values expressed in days.

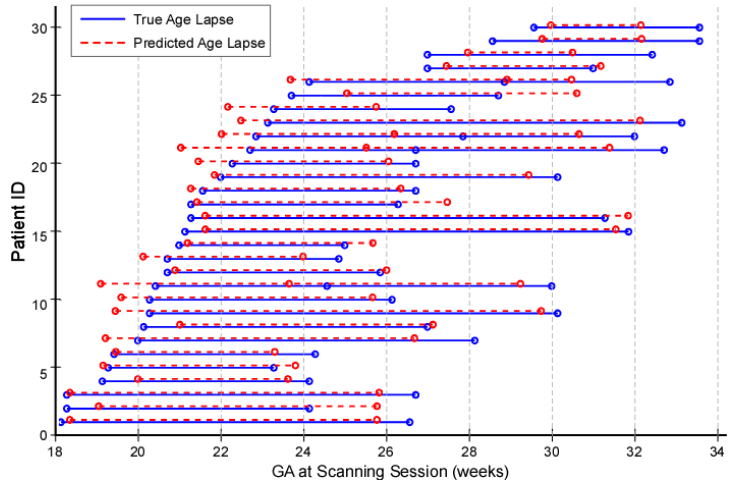
Trimester	2nd		3rd		2nd and 3rd	
Age Range, weeks	$18^{+0} - 27^{+6}$		$28^{+0} - 33^{+6}$		$18^{+0} - 33^{+6}$	
Age Mean (SD), weeks	23.45 (2.94)		31.07 (1.76)		25.83 (4.41)	
No. subjects, n	108		49		157	
Performance Measure	RMSE (r)	CI	RMSE (r)	CI	RMSE (r)	CI
Clinical HC Prediction, $\mathbf{a}^{\text{clin}}$	4.86 (0.94)	±9.46	10.28 (0.76)	±18.57	7.01 (0.98)	±13.32
Model-Based Prediction, $\mathbf{a}^{\text{pred}}$	5.18 (0.97)	±10.10	7.77 (0.83)	±14.01	6.10 (0.98)	±11.64
t-test significance	$p = 0.60$		$\mathbf{p < 0.05}$		$p = 0.35$	

time between separate scanning sessions. Figure 6.4a compares the GAs predicted for longitudinal scans collected from 31 subjects from this high-risk cohort. The lines denote the time lapses between subsequent scans (marked by circles). It is evident that there is high agreement between the true GA at a scanning session ($a_{t_i}^{\text{clin}}$) and the corresponding model-predicted GA ($a_{t_i}^{\text{pred}}$), with a correlation value of $r = 0.98$ ($p < 0.01$). We also found that there is

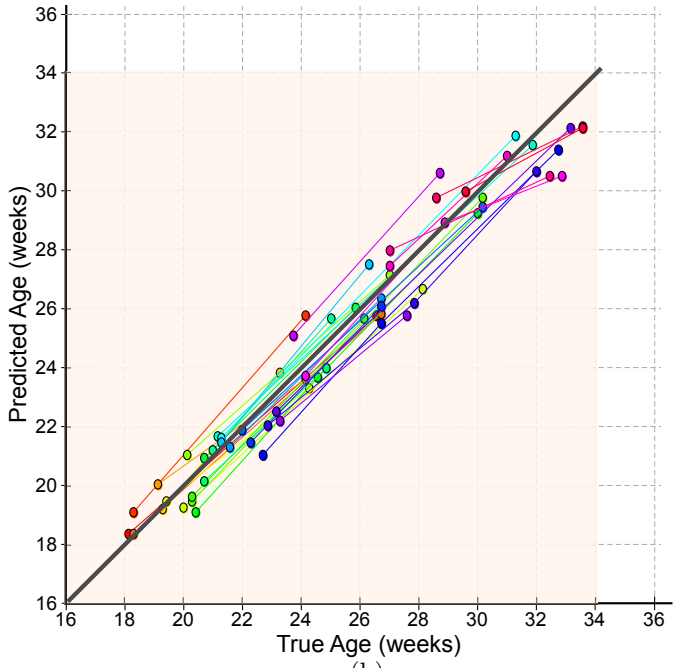
high correlation between the time lapses of subsequent visits extracted from true GAs and from predicted GAs ($r = 0.90, p < 0.001$). This demonstrates the ability of the model to consistently approximate GA with monotonically consistent time lapses between the predictions of each scanning session. However, a paired t-test reveals that although the time lapses are not significantly different when the last scan is conducted prior to 32 GW ($p = 0.32$), the time lapses do differ significantly if the subsequent scan is collected after 32 GW ($p < 0.05$). This is also visible in Figure 6.4a, where predicted time lapses tend to be shorter than true time lapses after 32 GW, and in Figure 6.4b where it is also evident that the model tends to underestimate GA after 32 GW. Figure 6.4b shows the developmental trajectories of each patient throughout the period of longitudinal tracking/scanning. In the ideal case, a fetus' developmental trajectory should follow a linear progression, adhering closely to the line of equality which marks ideal healthy maturation. For our dataset, we found that this plot more clearly displays the consistently tight fit of the longitudinal age predictions for the same patient along the line of equality, and that predictions increase monotonically with $\mathbf{a}^{\text{true}}$. Such a plot informs that even in complicated pregnancies the developmental trajectory followed a near-linear progression. This is also indicative of the potential for using the model to extract personalized maturational progression across different time points in pregnancy.

6.2.3 Discussion

This section presented an application of the framework described in Chapter 4 to a high-risk clinical fetal cohort of 187 fetuses. Our results show that the algorithm improved the confidence of age predictions provided by the clinical HC method by ± 0.64 days and ± 4.57 days in the second and third trimesters,



(a)



(b)

Figure 6.4: **Developmental trajectories for a high-risk dataset**. The time difference between predicted ages of a fetus at multiple separate scanning sessions. For each patient, lines demonstrate age difference between subsequent sessions, and individual scanning sessions are denoted by circles. In (b), each individual patient is displayed in a unique colour.

respectively. Specifically, our third trimester predictions within ±7.77 days are a notable improvement on the ±18 days reported in the literature [38]. Moreover, the fact that model-based prediction errors increased with GA may be

reflective of the inherent biological variation as gestation progresses, ultimately making the task of GA estimation more challenging. In fact, the largest errors were seen in the third trimester when it is likely that the model relies more on clinical HC than on image appearance. That is, due to progressive skull calcification and increasing structural occlusions, there is a reduction in the image-based support available to the GA estimation task. This, in conjunction with the fact that tree traversal path lengths generally increase with GA (as shown in Figure 4.17), demonstrate the longitudinal behaviour of the model as it is able to respond to the fact that US-based GA estimation is more complex in later gestation.

6.3 Pilot Study 2: Small-for-Gestational-Age (SGA) Cohort

Low birth weight (BW) is known as a major cause of morbidity in early infancy, and being born small-for-gestational-age (SGA) has been associated with increased mortality from disorders such as coronary heart disease and stroke in childhood [194]. The most common cause of intrauterine growth restriction (IUGR) is placental insufficiency [195], and clinical studies have demonstrated a link between an adverse intrauterine environment and alterations in brain tissue volume [196]. Using Doppler US imaging, it has been observed that in the presence of placental insufficiency, the fetus adapts to preferentially perfuse the central nervous system to maintain oxygen supply to the brain as best as possible [197]. This fetal adaptive mechanism refers to the relative protection of the brain as compared with other organs during fetal development, and is termed the "brain-sparing effect." In this section, we apply our algorithm to investigate whether our predictive age model is capable of identifying neurodevelopmental indications of the brain-sparing effect.

6.3.1 Experiments

Central to this pilot study is the definition of SGA. We followed the definition accepted by most neonatologists, by which a newborn whose birth weight falls below the 10^{th} centile of the population chart is classified as SGA [198]. As a reference, we used the charts provided by the recently published newborn size charts presented by Villar et al [199] generated from the INTERGROWTH-21^{st} dataset, a subset of which was used in this work. For this pilot study, we selected cross-sectional scans from 91 fetuses from the INTERGROWTH-21^{st} database whose newborn size measurements (i.e. in the form of birth weight

and GA at birth) were available. Of the selected fetuses, $n = 50$ were classified as SGA (BW $\leq 10^{\text{th}}$ centile), and the remaining $n = 41$ fetuses were treated as healthy controls (BW $> 10^{\text{th}}$ centile). Figure 6.5 displays the birth weights of the fetuses included in this experiment, and detailed characteristics of both groups can be found in Table 6.2.

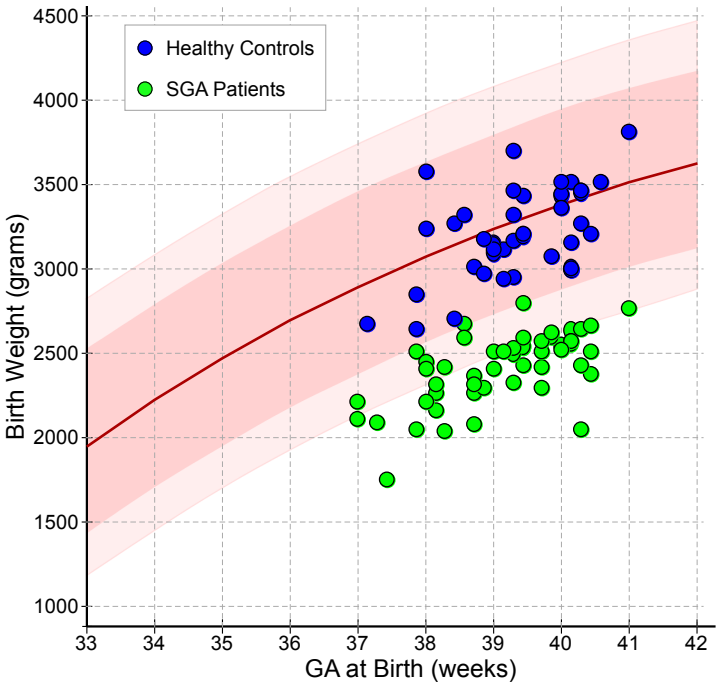


Figure 6.5: **Birth weights from cross-sectional study: SGA vs. Healthy cohorts.** Shaded regions (in pink) indicate the 3^{rd}, 10^{th}, 50^{th}, 90^{th}, and 97^{th} smoothed centile curves for the birth weight versus GA at birth as presented by Villar et al [199] for the INTERGROWTH-21^{st} database. Each circle represents a single subject. Plotted are the birth weights versus the GA at birth for the healthy (blue circles) and SGA (green circles) fetuses included in Pilot Study 2. Note that we adhere to the definition of SGA fetuses as those whose birth weight falls below the 10^{th} centile.

6.3.2 Results

Predicting Gestational Age

To demonstrate the predictive power of the age estimation model, true GA was plotted against the model-predicted GA for the healthy controls and the SGA

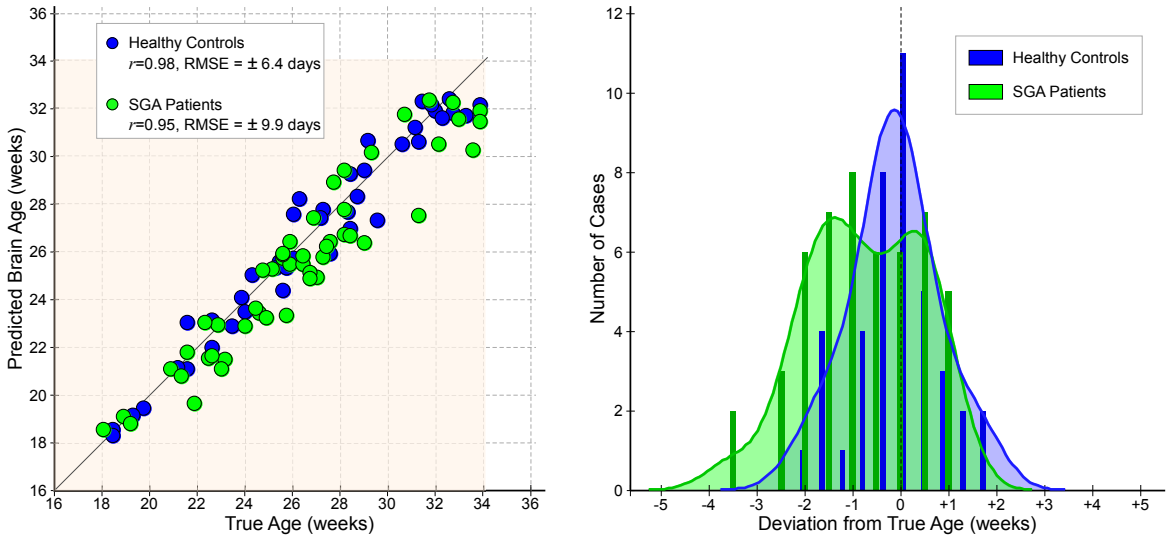
Table 6.2: **Subject demographics for cross-sectional study** of healthy controls and SGA patients.

	Healthy controls (BW > 10^{th} centile)	SGA patients (BW ≤ 10^{th} centile)
No. of subjects	41	50
Mean (SD) age at birth, weeks	39.35 (0.86)	39.09 (0.99)
Mean (SD) birth weight, kg	3.20 (0.27)	2.40 (0.22)
Mean (SD) age at scan, weeks	26.77 (4.36)	26.28 (3.96)
Correlation coefficient, r	0.98	0.95
RMSE (days)	6.4	9.9

patients (Figure 6.6a). In both groups, prediction accuracy was high, resulting in correlation coefficients between the estimated and true GAs of $r = 0.978$ ($p < 0.001$) and $r = 0.953$ ($p < 0.001$) for the healthy and SGA fetuses, respectively. However, through the RMS error, it is evident that the model produced higher errors in predicting the age of the SGA fetuses (RMSE= ±9.9 days) than of the healthy controls (RMSE= ±6.4 days), with post-hoc t-tests resulting in differences between the prediction errors observed in the two cohorts ($p < 0.001$). This may be attributed to the complexity and varying permutations of conditions which may cause an individual fetus to be diagnosed as SGA, and consequently how the brain is affected. Nevertheless, it is worth noting that our algorithm did not yield significantly age predictions when applied to these fetal cohorts ($p = 0.186$), and that only the *errors* produced differed.

To illustrate the relative likelihood of the prediction model over- or under-estimating the GAs of fetuses in each of the fetal cohorts, Figure 6.6b shows the probability density estimate of the signed root-mean-squared (RMS) error for the GA predictions. For these data, the algorithm tends to underestimate GA predictions on the SGA dataset, hence the negative bias in the density function (in green). By contrast, the curve associated with predictions for the healthy controls is centred around the zero-RMS error region, has a narrower error range, and has a higher peak. This indicates that the model is capable of more accurately predicting GA exclusively from brain maturation, particularly

in cases where the brain is following a trajectory of healthy brain development. The fact that the bias associated with SGA predictions is *negative* suggests that there may be neurodevelopmental delay in these cases.



(a) Age regression results

(b) Deviations from GA_{true}

Figure 6.6: **GA regression results from cross-sectional study – SGA vs. Healthy Cohorts.** (a) GA estimation results for the patient group identified as small-for-gestational age (SGA) in comparison to a healthy control group. (b) The probability density distributions of the RMS age error are shown for the healthy controls (blue) and the SGA patients (green). Shaded regions correspond to the density traces estimated using a Gaussian kernel, $\sigma = 0.5$)

Tracking Longitudinal Brain Maturation

To assess whether the algorithm is capable of identifying intrauterine neurodevelopmental retardation (i.e. slow brain development based on two or more US scans), we applied the model to scans collected from a fetus at two to four separate time-points in pregnancy. Of the subjects considered in the previous experiment, we selected $n = 10$ healthy subjects and $n = 14$ SGA patients which were tracked longitudinally. Figure 6.7 shows the birth weights of these subjects. Note that the SGA dataset includes some severe cases in which BW falls below the 3$^{\text{rd}}$ centile ($n = 10$).

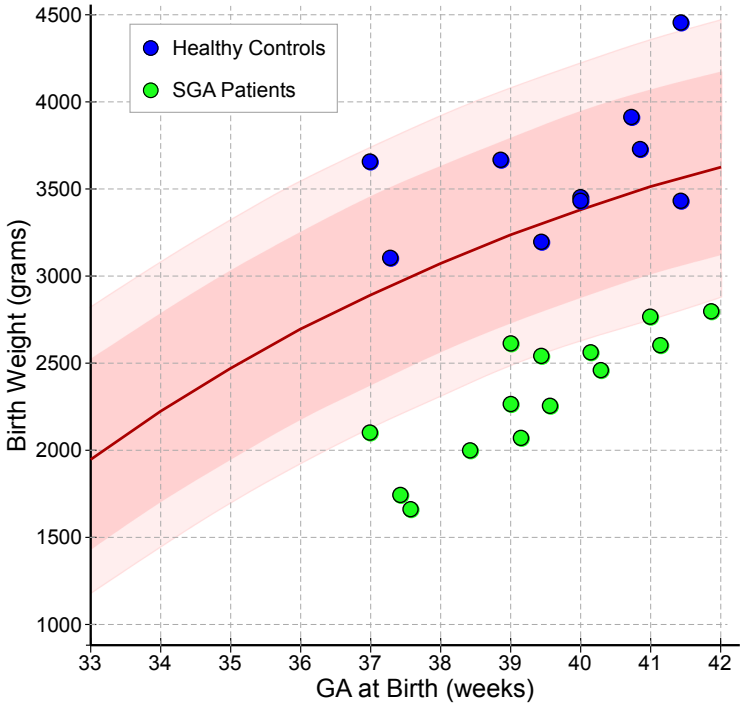


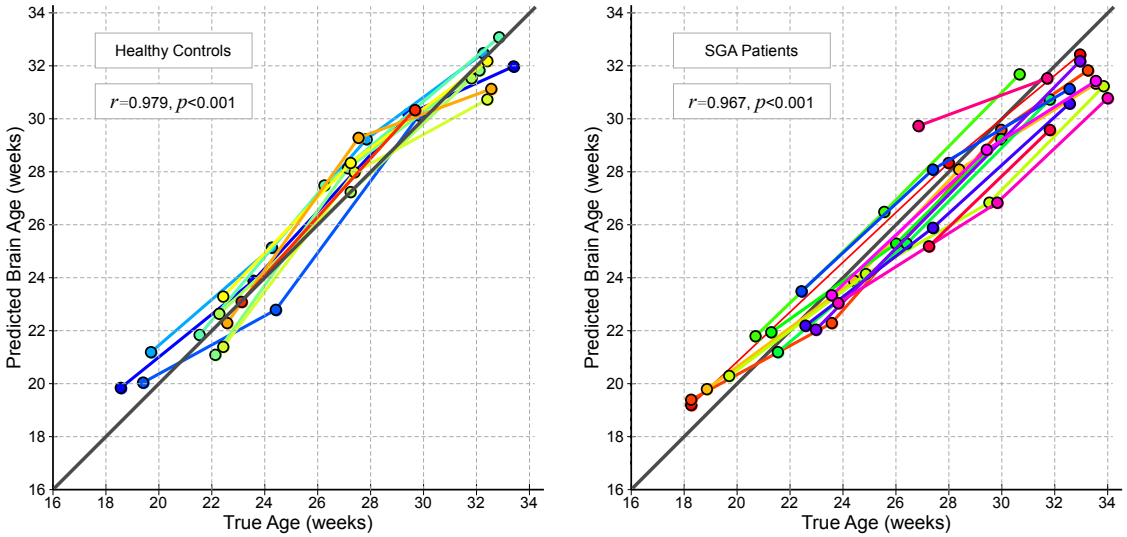
Figure 6.7: **Birth weights for longitudinal SGA study: SGA vs. Healthy cohorts.** Shaded regions (in pink) indicate the 3^{rd}, 10^{th}, 50^{th}, 90^{th}, and 97^{th} smoothed centile curves for the birth weight versus gestational age at birth as presented by Villar et al [199] for the INTERGROWTH-21^{st} database. Each circle represents a single subject. Plotted are the birth weights versus the GA at birth for the healthy (blue circles) and SGA (green circles) fetuses included in the longitudinal experiment of Pilot Study 2. Note that we adhere to the definition of SGA fetuses as those whose birth weight falls below the 10^{th} centile.

Figure 6.8 displays the GA predictions from the longitudinal scans of the healthy and SGA groups. From the figure, it is evident that in the second trimester, the algorithm behaves in a similar manner for the two groups. Specifically, the prediction error is not significantly different for the two groups between 18 and 28 GW ($p = 0.172$). However, in the third trimester, larger deviations from the line of equality are observed in the SGA group ($p < 0.05$), which could be indicative of developmental delays. This suggests that both groups have a similar developmental starting point and that the effect of an adverse intrauterine environment is only clearly observable from approximately

28 GW when deviations from the trajectory of normal maturation become more pronounced.

By following each individual SGA patient, we notice that once their GA is underestimated and the trajectory deviates below the line of equality particularly in later gestation, it *remains* below the line even in subsequent scans. A Kolmogorov-Smirnov (KS) test reveals that SGA predictions are consistently below the line ($p < 0.001,\ k = 0.311$), emphasising that the errors do not follow a standard normal distribution, in contrast to the errors of healthy controls ($p = 0.305,\ k = 0.169$). This is an important result because clinically, if a trajectory crosses the line of equality more than once, it would confound a diagnosis of potential neurodevelopmental delay, rendering the model an insecure basis for informing on neurodevelopmental states. The fact that all SGA fetuses in our dataset were consistently below the line is important for clinical tracking of neurodevelopmental progression because it means that the model confidently and consistently confirmed the state of delayed brain maturation in each scanning session.

In order to verify whether there are group-specific differences in the overall trends of neurodevelopmental maturation, we plotted the mean trajectories for the healthy (Figure 6.9a) and SGA (Figure 6.9b) groups approximated by a quadratic fit function. It is clear that the mean trajectory of the healthy group adheres closely to the normative trajectory of healthy brain development, whereas the SGA group generally follows a trajectory which deviates below this line. Again, this indicates the presence of developmental delays in the SGA group, particularly in the third trimester. However, to better observe group-specific differences in developmental trajectories, we plotted the GA prediction in the form of a brain maturation score (BMS; Figure 6.9c). The BMS is calculated as the mean difference between the model-based GA predictions



(a) Healthy controls (b) SGA patients

Figure 6.8: **Developmental trajectories for (a) Healthy and (b) SGA cohorts**. The time difference between predicted ages of a fetus at multiple separate scanning sessions. For each patient, lines demonstrate age difference between subsequent sessions, and individual scanning sessions are denoted by circles. Each individual patient is displayed in a unique colour.

and the true LMP-based GA. In this experiment, we exemplarily compared fetuses scanned within four GA ranges (i.e. 18-22 GW, 22-26 GW, 26-30 GW, and 30-34 GW) and computed the mean GA difference and standard deviation for the subjects within each GA group, $\mathcal{G}_a$:

$$\text{BMS} = \frac{1}{|\mathcal{G}_a|} \sum_{k \in \mathcal{G}_a} a_k^{\text{pred}} - a_k^{\text{true}} \tag{6.1}$$

where a_k^{true} is the true GA of the k-th test image in age group $\mathcal{G}_a$, and a_k^{pred} is the model-predicted GA.

From the figure, we can observe the longitudinal trend of brain maturation for both fetal groups. The ideal score associated with normative development would be approximately zero for all GA groups, deviations above zero indicate accelerated brain development, and conversely, deviations below zero are indicative of delayed maturation. Note that the brain scores for normal fetuses

hover around zero, whereas the SGA fetuses increasingly deviate below the normative trajectory as gestation progresses. In particular, the brain maturation scores between the two fetal cohorts differed significantly at 26-30 GW ($p < 0.01$; $df = 13$) and at 30-34 GW ($p < 0.05$; $df = 18$), which corroborates the finding that the developmental deviations become more apparent after 28 GW. Table 6.3 summarises the demographics of each age group, and the statistical significance results between SGA and control subjects for each GA.

Table 6.3: **Subject demographics for longitudinal study** of healthy controls (labelled 'H', BW $> 10^{\text{th}}$ centile) and SGA patients (labelled 'S', BW $\leq 10^{\text{th}}$ centile), for four pre-defined age groups.

	Group	18 - 22 GW	22 - 26 GW	26 - 30 GW	30 - 34 GW
No. of subjects	H	4	9	10	8
	S	7	10	12	13
Mean (SD) age at scan, weeks	H	19.8 (1.26)	23.0 (0.87)	28.0 (1.23)	32.5 (0.48)
	S	19.8 (1.39)	24.0 (1.21)	28.4 (1.32)	32.7 (0.98)
Mean (SD) error of prediction, days	H	+6.48 (3.93)	-1.31 (6.13)	+5.74 (3.71)	-4.71 (5.36)
	S	+4.96 (3.69)	-2.69 (5.38)	-5.26 (11.14)	-10.58 (7.87)
Significance (unpaired t-test)		$p = 0.55$ (df=6)	$p = 0.61$ (df=16)	$\mathbf{p < 0.01}$ (df=13)	$\mathbf{p < 0.05}$ (df=18)

Assessing Patient-Specific Neurodevelopmental Delays

In Chapter 5, we demonstrated the use of model-extracted features to map an individual's brain maturation, and identify regions of relevance to fetal neurodevelopment as observed in US images. By using these feature probability maps of brain maturation, we can also compare the behaviour of the model on SGA cases to healthy populations to assess the anatomical regions in which the model responds differently— possibly due to the presence of developmental delays. In this experiment, we selected two SGA cases and obtained the feature probability map for each of their scanning sessions, $\mathbf{H}_i^{a_i}$. The individual's probability map from a given scanning session ($\mathbf{H}_i^{a_i}$) was then subtracted from

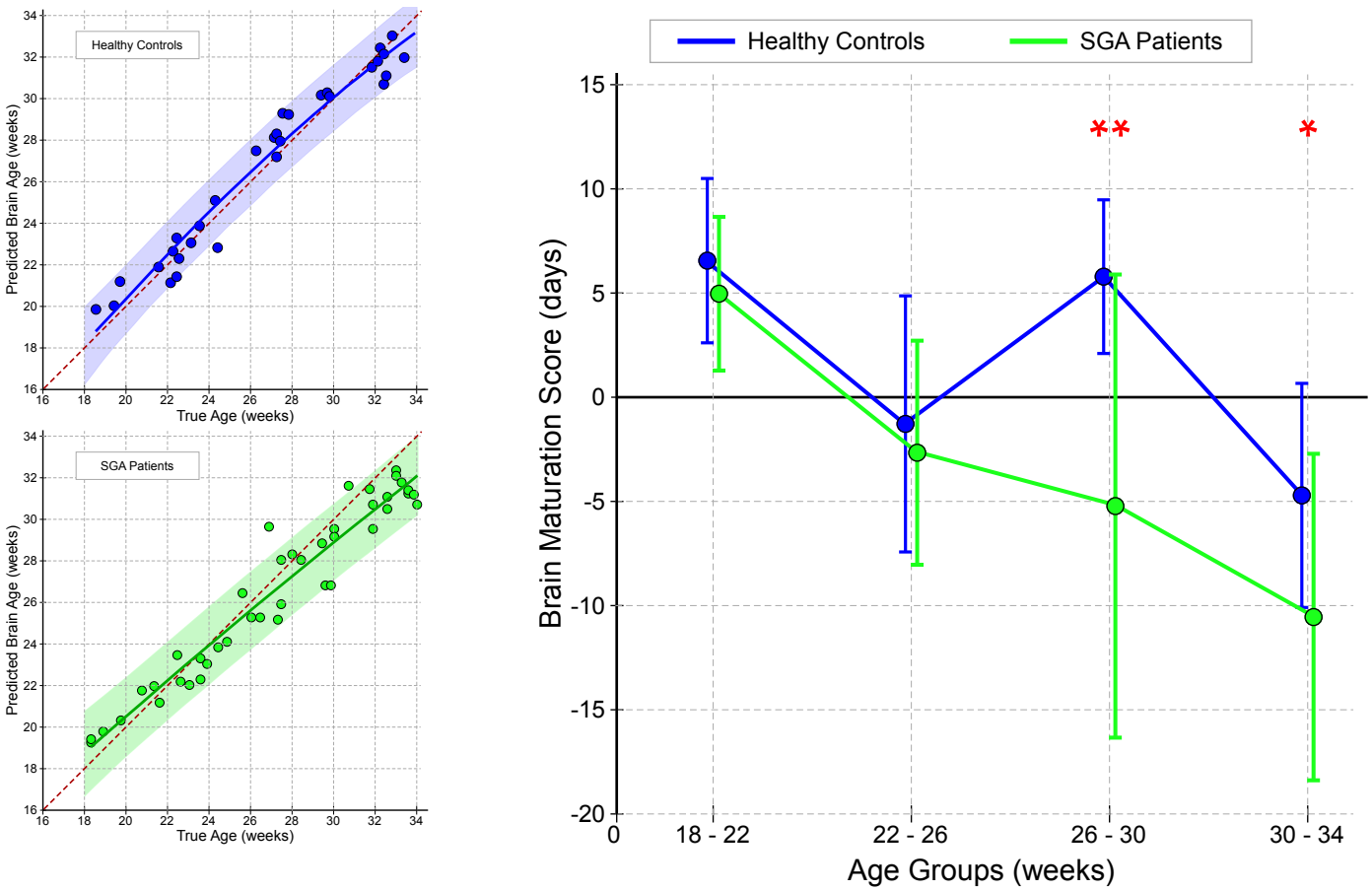


Figure 6.9: **Confidence bounds and brain maturation score for longitudinal study.** The 95% confidence intervals of GA prediction are shown for the (a) healthy controls and (b) SGA fetuses considered in the longitudinal study. The shaded regions contain the values between the 5th and 95th percentiles of the samples, including the mean trajectory (solid line). The line of equality (dashed red line) illustrates the normative trajectory of brain maturation. To demonstrate the longitudinal trend in neurodevelopmental progression, (c) shows the brain maturation scores for the SGA (green) and healthy cohorts (blue) for fetuses grouped into four pre-defined GA groups. Error bars denote standard deviation, and post-hoc t-tests resulted in significant differences at 26-30 GW (p¡0.01, two asterisks) and at 30-34 GW (p¡0.05, one asterisk).

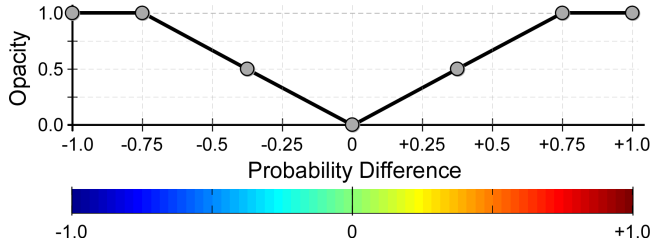
the population probability map representative of the GA of the scanning session ($\mathbf{B}^{d}_{a=a_i}$, generated in § 5.3), such that a difference probability map ($\mathbf{H}_{\text{diff}}$) of features associated with maturation can be calculated as follows:

$$\mathbf{H}_{\text{diff}} = \mathbf{H}_{i}^{a_i} - \frac{1}{(d_2 - d_1)} \sum_{d=d_1}^{d_2} \mathbf{B}^{d}_{a=a_i} \tag{6.2}$$

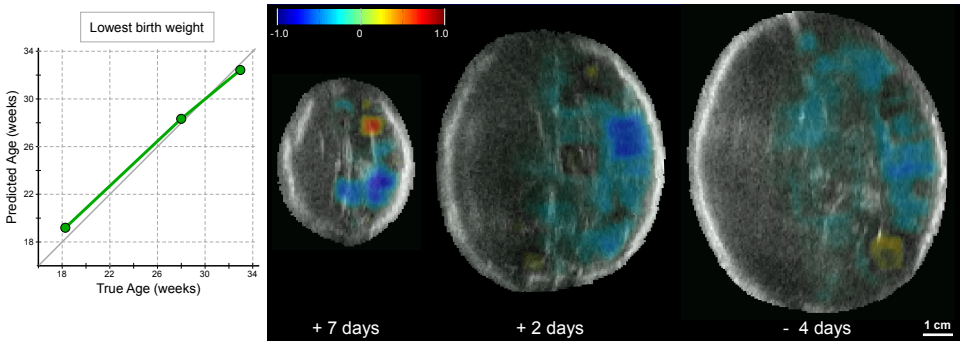
where a_i is the true GA of individual i at a particular scanning session. For this experiment, only the features selected as the image traversed between forest levels $d_1 = 5$ and $d_2 = 9$ were included as salient, age-specific regions of interest.

We first assessed the fetus with the lowest birth weight (born at 37^{+4} GW, weighing 1.66 kgs) to verify whether any effect of brain-sparing was detectable. We found that this fetus yielded low errors in GA predictions and the probability values of the difference map were generally low, as marked by light blue regions in Figure 6.10b. This suggests that the feature responses of this fetus' scans were similar to the responses observed in the normal population brain maps, which indicates that although this fetus was born small, its brain development was age-appropriate at each scanning session. Such a result suggests that birth weight does not directly indicate delayed brain maturation. It may also suggest that brain sparing is in effect, ensuring that the brain is preferentially perfused and developed in this severe SGA case. One might even speculate that the blue regions (low response in the individual compared to the population map) seen in the probability difference map at 18 GW are compensated for by the orange regions which may be over-developed in this case to balance the SGA fetus' brain maturation.

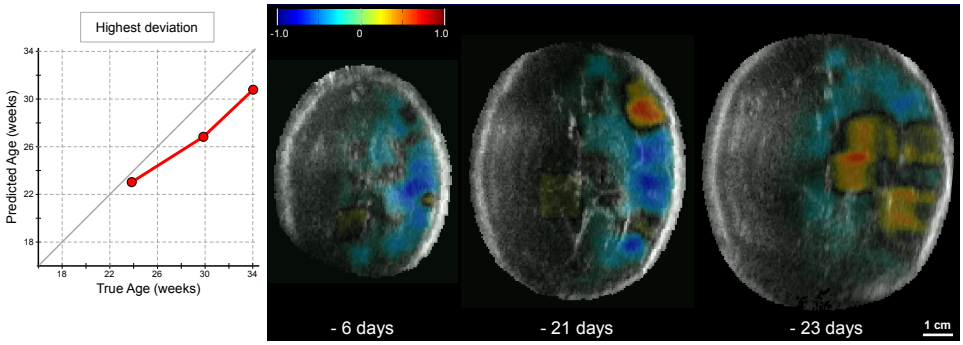
By contrast, the fetus with the largest deviations in its developmental trajectory (i.e. associated with the lowest BMS), demonstrated larger positive (Figure 6.10c, in orange/red) and negative (Figure 6.10c, in blue) feature responses than the population map in all scanning sessions. We can infer from



(a) Colour function for probability difference maps



(b) Lowest birth weight



(c) Lowest brain maturation score

Figure 6.10: **Difference probability maps for two SGA patients tracked longitudinally**. (a) Colour function demonstrating the opacity of the colours used for visualization of the probability difference maps, $\mathbf{H}_{\text{diff}}$. Colour range shows negative values (blue) denoting underestimations in GA predictions made the model, ranging to positive values (red) which correspond to overestimations. Shown are the difference maps for (b) an SGA patient whose birth weight was the lowest in the cohort, and (c) an SGA patient who demonstrated the largest deviation from the trajectory of normal brain maturation. Each fetus was scanned twice, and shown is the brain maturation score (in days), which is calculated by subtracting the model-predicted GA ($\mathbf{a}^{\text{pred}}$) from true GA ($\mathbf{a}^{\text{true}}$). Scale bar: 1 cm.

this that the age-appropriate features were not present in this example, and so at each GA the model examined and detected image regions which were associ-

ated with a different GA. Since the probability values and spread of the negative (blue) difference maps increased with GA and its BMS became progressively negative throughout pregnancy, it is possible in this case that brain-sparing did not enter into effect as there seems to be little or insufficient recovery in steering the brain towards the normal trajectory.

6.3.3 Discussion

In this section, we demonstrated that the model is capable of using only appearance-based image information to detect developmental delays in SGA fetuses. We found that although healthy and SGA fetuses display a similar neurodevelopmental starting point, the latter tend to deviate from the trajectory of normal development, particularly in the third trimester. It is important, however, to note that while we found indications for developmental delays through the feature responses extracted by the model, this section is merely a speculative discussion on the presence/absence of the brain-sparing effect in two SGA cases. Fetal growth restriction is a complex condition which is subject to several definitions and associated with a myriad of pathologies, and its impact on the fetus and its developing brain is a highly complex phenomenon. So at this time we cannot draw strong conclusions about fetal size and how it relates to brain maturation. In order to fully, deterministically do so, we may need to conduct further orthogonal experiments involving nutritional markers, infections, and other fetal conditions, which are beyond the scope of this project and are reserved for future work.

6.4 Conclusion

In this chapter, we applied the algorithm to unique and challenging clinical datasets to demonstrate its ability to outperform current clinical practice in predicting gestational age, particularly in the third trimester when size-based approaches to GA estimation are not clinically useful. By applying our model to a high-risk clinical dataset, we demonstrated that our model successfully predicts GA within ±7.77 days, which reduces the typical error of third trimester age predictions by 11 days [38]. In clinical practice, a predictive error of ±10 days is considered acceptable at 22 GW [3]; the fact that our results fall below this error places our third trimester predictions within an acceptable range, thereby extending the clinically useful range of US image-based GA prediction. Moreover, this result suggests that our model may be applied to more accurately predict GA in resource-poor settings in which women present late for care (i.e. after 24 GW), and thus to advise on the potential success or safety of a clinical intervention, particularly in high-risk pregnancies such as those considered in this chapter.

Our preliminary experiments on SGA fetuses demonstrated that using only image-based support during the decision-making process, the algorithm can singly predict GA and detect indications for neurodevelopmental delays. By comparing the trajectory of GA predictions from longitudinal scans collected from a single patient, we were able to detect and quantify the individual's deviation from normal maturation. In addition, by analysing the brain maturation maps for each individual we showed that the model is sensitive to developmental differences within the imaged brain. Clinically, this information could be useful in assessing whether the brain-sparing effect is capable of recovering from IUGR-related delays whilst *in utero*, and quantifying how well the fetal body is responding to protect the brain and possibly 'catch up' to restore normal brain

development in such complicated cases. In order to deterministically quantify the degree to which the patient is delayed, a large-scale clinical study would need to be carried out and population-based trajectories of SGA maturational progression would be generated. A large-scale clinical trial is beyond the scope of this thesis project, but here, we have presented the tools and conducted a pilot study to demonstrate the type of analysis which would lead to a better understanding of brain states in an adverse intrauterine environment.

7

Conclusion

This thesis has presented an automated learning-based framework to predict gestational age, and hence developmental maturation, from a single 3D ultrasound image of the human fetal brain. We have developed a method for aligning US images despite being challenged by partial occlusions and reverberation artefacts, and we have demonstrated its usefulness in— and the feasibility of— extracting biologically informative content from sonographic image features. Such tools are required for non-invasively and routine quantification of *in utero* brain development. This chapter summarizes the main contributions of this thesis, and discusses the limitations of the present work and opportunities for

future research directions.

7.1 Contributions

The main contributions of this thesis are fourfold:

1. **Development of a cranial parametrization surface domain for sampling brain regions**

   A semi-automatic framework which fits a continuous parametric surface to the skull boundaries in an US image of the fetal head was presented in this work. The surface domain acts as a coordinate system on the basis of which image information can be anatomically queried from any patient image based on skull position. The user provides a rough alignment of the skull surface to the imaged brain area, and the surface is then minimally deformed to the inner skull boundaries without changing the topology of the surface points. By selecting surface points, intracranial regions may be sampled and image planes may be extracted on different surfaces associated with different patients at different GAs, without requiring registration between images. In addition, it conveniently separates the fetal brain from extracerebral tissue such as the skull, skin, and maternal tissues.

2. **A learning-based model to predict gestational age from 3D US images of the fetal brain**

   Realising that throughout pregnancy there are notable changes in the echoes visible in US images marking the emergence and disappearance of cerebral structures, this work was the first to exploit age-related sonographic activity to predict neurodevelopmental maturation. To achieve this, we developed an automated, machine learning-based predictive model

to learn the pattern of fetal brain changes, and demonstrated successful age estimation. We developed features which capture sonographic patterns in 3D images, taking into consideration the challenges posed by US image data. The model characterized neuroanatomical appearance both spatially and temporally, capturing the natural variation existing in a healthy fetal population over an age range of active brain maturation, modelling age as a continuous variable. Our features captured image information about structural appearance and spatial extent, and were capable of characterizing age-specific biomarkers within the fetal population. The framework incorporated clinically-relevant subject metadata, which is often available in clinical studies. The combination of imaging and non-imaging information allowed for a unified representation of the fetus' neurodevelopmental state. Our approach outperformed the methods used in current clinical practice, particularly in the third trimester when biometric measurements offer little clinical value, thereby extending the clinically useful range of US-based tracking of fetal development beyond the second trimester.

3. **Discovery of age-specific anatomical landmarks**

   By analysing the features automatically selected by the predictive age model during the decision-making task, we discovered key regions which experience significant changes over gestation and are observable in US images. Our analysis showed that the model was sparse in its identification of salient biomarkers, and that the query of different anatomies was conducted in a stratified, hierarchical manner— rather than by random selection— during the decision-making process. The biomarker identification was achieved using a data-driven approach, as opposed to *a priori* biomarkers defined by manual segmentation. We found that the auto-

matically selected regions of interest correspond to anatomies which have been reported in the literature to undergo characteristic changes, thus demonstrating that US-images are content-rich and that they may feasibly be used to study underlying biological processes in the brain.

4. **Detection of neurodevelopmental delays associated with growth restriction**

   We have shown that the model is applicable to real, and challenging, clinical datasets to predict GA and to obtain information about an individual fetus' neurodevelopmental state. Our accurate GA predictions offer the potential to track the maturational progression of a single patient over a series of independent scanning sessions. By applying our model to longitudinal tracking of a fetus, we found that neurodevelopmental delays associated with intrauterine growth restriction are generally detectable in the third trimester, whereas healthy patients do not demonstrate large deviations from the normative trend of brain maturation.

7.2 Limitations and Future Work

There are limitations of the current work which provide the opportunity for future research to improve on the performance, medical applications, and user interaction required by the GA estimation framework. The areas of potential improvement may be categorized as follows:

Surface-Based Image Alignment

1. **Validation of sensitivity to manual user initialisation**

   A major drawback of the surface-based image alignment is that it relies on the user to provide a prior initialisation for each model surface which

is then deformed to fit the interior skull boundary. Our validation experiments demonstrated that the surface model (from a single manual alignment) is able to recover standard clinical images slices within 5 mm of the slices manually selected by a clinician. However, we did not test the sensitivity of the surface deformation process to the degree of misalignment in the manually initialised surface. A simple experiment in which a single volume is initialised more than once, by independent clinical experts, would inform on the inter- and intra-user variability in user alignment and its effect on GA prediction results. It may also be of interest to determine the tolerance levels for user-provided initial misalignment which would allow acceptable inter-subject alignment and anatomical correspondences to result in sufficiently accurate GA predictions.

2. **Inclusion of appearance-based information into surface deformation energy function**

   We have shown that our method recovers detailed structure of the skull and anatomically consistent skull surfaces. However, fine-grained alignment is still desirable to ensure direct anatomical correspondence between images. Using only geometric information, we were able to recover decent anatomical alignment between images and patients. However, supplementing the model with additional appearance (intensity) information may improve the fit of the model surface to the imaged brain structures. Specifically, it would be possible to modify Equation 3.1 to include an image appearance term such that the equation is expressed as follows:

$$E(\mathbf{q},\mathbf{U},\mathbf{X},\mathbf{I}) = \sum_{i=1}^{N_{\mathrm{U}}} E_{\mathrm{unary}}(q_i,\mathbf{u}_i,\mathbf{X}) + \lambda_2 E_{\mathrm{pairwise}}(\mathbf{q}) + \lambda_3 E_{\mathrm{user}}(\mathbf{X}) + \lambda_4 E_{\mathrm{reg}}(\mathbf{X}) + \lambda_5 E_{\mathrm{app}}(\mathbf{I}) \quad (7.1)$$

where λ_5 is the weight of the appearance term in the energy function. The appearance term could simply penalise the texture differences between image patches in the test image ($\mathbf{I}$) and a predefined reference image ($\mathbf{I}_{\mathrm{ref}}$):

$$E_{\mathrm{app}}(\mathbf{I}) = \|\mathbf{I}(\mathbf{p}) - \mathbf{I}_{\mathrm{ref}}(\mathbf{p})\|^2 \quad (7.2)$$

3. **Creating an automated image alignment tool**

   The major limitation of this work is that the parametric surface presented in Chapter 3 serves only as a surrogate representation of the fetal skull such that images from different patients at different ages can be co-aligned to a common coordinate space. In this prototype, the surface is *manually* aligned to the brain pixels, which does not guarantee anatomical correspondence and is ineffective for processing large datasets. Despite our proposed surface-based approach, accurate isolation of the imaged brain area, or 'skull-stripping', and inter-patient image alignment remains a challenge in US images as we have not, for instance, automatically detected landmarks for use as input into registration algorithms. Thus, an ultrasound-specific registration framework would need to be developed. Directly following from this work, it may be possible to use develop a landmark detector (e.g. using AdaBoost learning as we have shown to be possible in US images of the brain in our previous work [154]) to extract the image features identified by our model (discussed in Chapter 5)

as input to a feature-based alignment model, as in [200]. The detected feature locations may then be used to re-align the images on the basis of feature correspondence. Such an alignment tool would transform the current framework into a fully-automated tool, and would open the door for the development of more sophisticated algorithms to study US-based brain maturation.

Modelling Brain Maturation

1. **Extracting structural growth patterns and trajectories**

   In this work, we have analysed the model-selected image patches/feature patterns to generate probability maps of neuroanatomical development. A limitation of this approach is that the probability map is only a proxy for the underlying structural development, as the feature patches do not exactly *delineate* each anatomy of interest, but rather *encompass* the entire region of interest. Consequently, with such a result we are unable to quantify growth trajectories of individual structures or tissue volume changes occurring as the brain matures. Segmenting the intracranial structures and the cerebral cortex from the US images may provide new insights about the *morphological* evolution of the brain, and allow for direct quantification of local growth patterns. Such a study would also enable us to capture the appearance, shape, and spatial variation of brain structures over gestation.

2. **Construction of a population-based brain atlas**

   Provided that an automated US-specific skull-stripping and accurate image registration tools are available, it would open the possibility of co-aligning the brain images to an unbiased inter-subject template coordinate space. In this work, we had access to the INTERGROWTH-21st

database of images, and if we capitalize on this unique database of 4321 healthy patients scanned six times during pregnancy, we could aggregate the images to make a probabilistic atlas of healthy brain maturation. By accurately aligning the images to a coordinate space, it could be possible to average the images to generate maps of average (normative) brain anatomy (e.g. [201]). Such an atlas could serve as a population-based reference against which individual patients may be compared.

Applications

The tool presented in this thesis was designed for the purpose of gestational age prediction from standard clinical US images of the fetal brain, and we have demonstrated its accuracy and stability in that application. However, the proposed framework now has the potential to be applied to answer a myriad of clinical questions. In Chapter 6, we demonstrated its potential in detecting neurodevelopmental delays based on the presence and selection of image features uncharacteristic of the fetus' GA. We also demonstrated that it has the potential to track personalised patient-specific neurodevelopmental trajectories from subsequent scans, to inform on an individual patient's maturation. The fact that the model is capable of establishing the normal, healthy trajectory of brain development, it may also be used to compare the (normal) developmental states between different fetal populations (e.g. ethnic, pathological, etc) to classify patients into different clinical categories, or to provide indicators of pathological severity or progression. In addition, the fact that the cranial surface delineates the skull and recovers the three-dimensional shape of the skull may allow for the extraction of clinically useful biometric measurements for applications in fetal growth monitoring [106] and detection of craniofacial dysmorphology [202, 203].

7.3 Summary

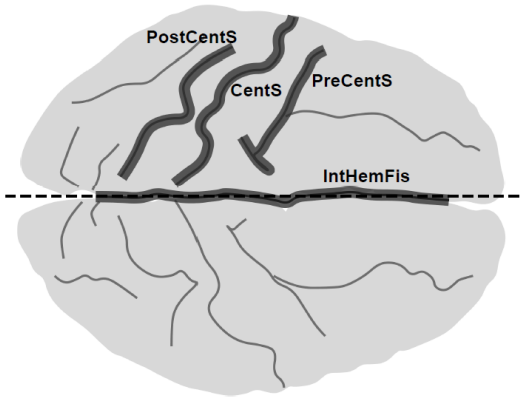
The work presented in this thesis represents a cross-disciplinary effort in the combination of machine learning algorithms, structural brain development, and ultrasound imaging to develop a tool with potential for clinical use. Specifically, we have developed a novel feature-based model which regresses brain development to GA using US images, which has never before been attempted. Using standard clinical US images, this tool predicts gestational age with better accuracy than current clinical practice even in high-risk datasets (Chapter 6). It was designed to automatically map the US signal responses to age-related emergence/disappearance of folds on the brain surface. The model is based on a regression forest classifier which takes as input a 3D US image and a parametric surface which provides the spatial context for anatomical structures (Chapter 3); using statistical image measurements it outputs a real-valued age prediction (Chapter 4). Through the feature-extraction process, the model is capable of automatically discovering brain regions associated with neurodevelopmental maturation during the gestational period (Chapter 5). Using these anatomical features, the model is capable of tracking *normal* brain development, and hence detect pathological deviations from the healthy trajectory (Chapter 6).

A

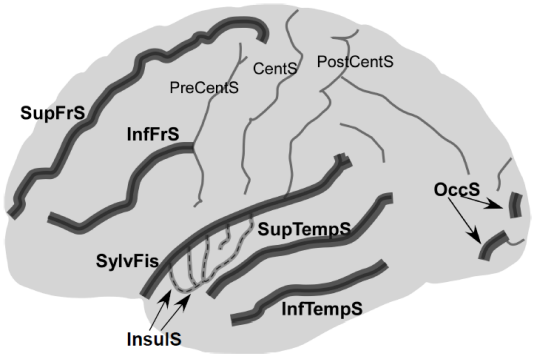
Reports of fetal brain gyrification

This section tabulates the sulcal appearance times, as reported in the literature, specific to post-mortem studies, and MR and US imaging modalities. For all tables, the structures are labelled as shown overleaf.

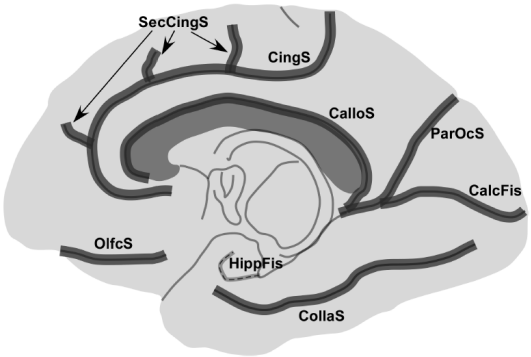
The ages of first visualization of a sulcus are expressed in gestational week (GW) and pertains to successful visualizations in more than 75% of images. Where applicable, the number inside the brackets corresponds to successful visualizations in 25%-75% of the images observed in the study.



1st Row: Dorsal aspect IntHemFis = Interhemispheric fissure; CentrS = Central sulcus; PreCentrS = Precentral sulcus; PostCentrS = Postcentral sulcus



2nd Row: Lateral aspect SylvFis = Sylvian fissure; SupFrS = Superior frontal sulcus; InfFrS = Inferior frontal sulcus; SupTempS = Superior temporal sulcus; InfTempS = Inferior temporal sulcus; OccS = Occipital sulci; InsulS = Insular sulci



3rd Row: Medial aspect ParOcS = Parieto-occipital sulcus; CalcFis = Calcarine fissure; CalloS = Callosal Sulcus; CingS = Cingulate sulcus; SecCingS = Secondary cingulate sulci; CollaS = Collateral sulcus; OlfS = Olfactory sulcus; HippFis = Hippocampal fissure

Table A.1: **Timeline of sulcal development reported by post-mortem neuroanatomical studies.** The ages of first visualization are expressed in *weeks*. NR=Not reported, NS=Not studied.

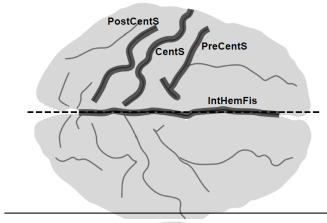
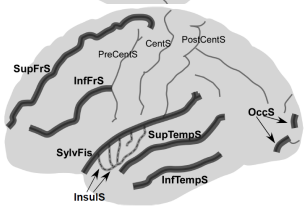
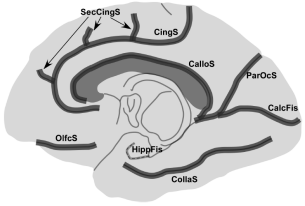
		Post-mortem Neuroanatomical Appearance Times			
		Chi et al (1977) [16]	Dorovini-Zis & Dolman (1977) [39]	Nishikuni (2006) [45]	Afif et al (2007) [46]
	No. of Fetuses (n)	507	80	214	42
	Age Range (weeks)	10 weeks – 1 month (post-natal)	22 – 41	12 weeks – 8 months (post-natal)	13 – 28
	GA Estimation Method	LMP	HC, CRL, body weight, brain weight	Body weight	LMP corrected with US biometry
	Data Source	Coronal sections; whole brain	Separated hemispheres	Separated hemispheres	Separated hemispheres
	Interhemispheric Fissure	10	NR	12	NS
	Central Sulcus	20	24	21	18-21
	Postcentral Sulcus	25	24	26±3	24
	Precentral Sulcus	24	NR	26±3	22
	Sylvian Fissure	14	22	17	18
	Superior Frontal Sulcus	25	NR	25±2	NS
	Inferior Frontal Sulcus	28	NR	30±3	NS
	Superior Temporal Sulcus	23	28	26±3	NS
	Inferior Temporal Sulcus	30	NR	31±3	NS
	Occipital Sulci	34	NR	30-40	NS
	Insular Sulci	34-35	NR	29±2	28
	Parieto-Occipital Fissure	16	22	17	NS
	Calcarine Fissure	16	22	17	NS
	Callosal Sulcus	14	NS	12	NS
	Cingulate Sulcus	18	24	19	18
	Secondary Cingulate Sulci	32	NR	30-40	NS
	Collateral Sulcus	23	NR	24±2	NS
	Olfactory Sulcus	16	NS	27	NS
	Hippocampal Fissure	NR	NR	15	NS

Table A.2: **Timeline of sulcal development reported by MR studies.** The ages of first visualization are expressed in GW, the number outside the brackets pertains to successful visualizations in more than 75% of images, and in brackets is the successful visualization in 25%-75% of images observed. NR=Not reported, NS=Not studied.

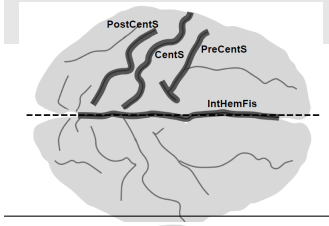
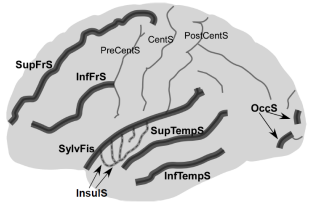
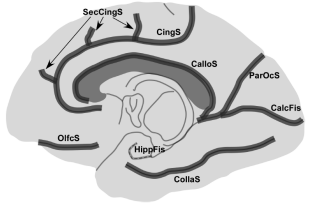
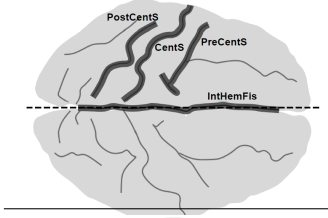
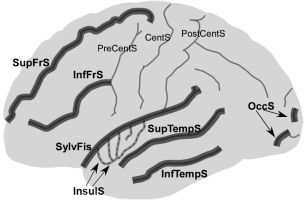
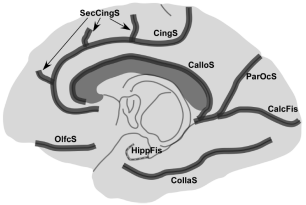
		MR Imaging Sulcal Appearance Times					
		Hansen et al (1993) [47]	Levine & Barnes (1999) [48]	Lan et al (2000) [49]	Ruoss et al (2001) [50]	Garel et al (2003) [51]	Zhang et al (2013) [52]
	No. of Fetuses (n)	NR	53	25	51	173	69
	Age Range (weeks)	6-28	14-38	12-38	23-43	22-38	12-22
	GA Estimation Method	CRL + BPD	LMP	LMP + CRL	LMP + Biometry	Biometry (12 wks)	NR
	Data Source (In/Ex vivo)	Ex vivo	In vivo	In vivo	In vivo	In vivo	Ex vivo
	Field Strength (Tesla, T)	1.5T	1.5T	1.5T	1.5T	0.5T	7T
	Interhemispheric Fissure	8-13	14-15, *already present*	NR, *already present*	12	22-23	12
	Central Sulcus	18-22	26-27	24-26	24.5	26	16 (15)
	Postcentral Sulcus	22-25	28-29		24.5	28	NR
	Precentral Sulcus	22-25	26-27			27	18 (16)
	Sylvian Fissure	13-14	16-17	15	NR	NR	14 (12)
	Superior Frontal Sulcus	22-25				29	16 (15)
	Inferior Frontal Sulcus	26-29	32-33			29	22 (18)
	Superior Temporal Sulcus	18-22	28-29	24-26		27-32	18 (16)
	Inferior Temporal Sulcus	26-29	32-33			29	
	Occipital Sulci	22-25				34	NR
	Insular Sulci	NS	32-35			34	NR
	Parieto-Occipital Fissure	14-17	18-19		30	22-23	16-15
	Calcarine Fissure	14-17	18-19	29		24-25	15 (13)
	Callosal Sulcus	8-13	NR		32	22-23	14 (12)
	Cingulate Sulcus	14-17	26-27	NR	28	24-25	14 (12)
	Secondary Cingulate Sulci	NR				33	NR
	Collateral Sulcus	18-22	NR	NR		27	
	Olfactory Sulcus	14-17	NR	NR		NR	16 (15)
	Hippocampal Fissure	18-22	NR	NR		22-23	14 (12)

Table A.3: **Timeline of sulcal development reported by US studies.** The ages of first visualization are expressed in GW, the number outside the brackets pertains to successful visualizations in more than 75% of images, and in brackets is the successful visualization in 25%-75% of images observed. NR=Not reported, NV=Not visualized, NS=Not studied, TA=Transabdominal, TV=Transvaginal.

		Ultrasonographic Appearance Times						
		Bernard et al (1988) [53]	Monteagudo & Timor-Tritsch (1997) [44]	Toi et al (2004) [8]	Correa et al (2006) [54]	Cohen-Sacher et al (2006) [43]	Ruiz et al (2006) [55]	Pistorius et al (2010) [40]
	No. of Fetuses (n)	70	262	50	202	22	158	215
	Age Range (weeks)	10-37	14-40	15-29	16-24	18-38	21-34	20-40
	GA Estimation Method	LMP	LMP + Biometry	Biometry	LMP confirmed by biometry (≤15wks)	1^{st} trimester Biometry (≤14wks)		LMP confirmed by CRL
	Data Source (In/Ex vivo)	Ex vivo	In vivo	In vivo	In vivo	In vivo	In vivo	In vivo
	Imaging (2D/3D)	TA 2D	TV 2D (orthogonal and oblique)	TA/TV 2D	TA 2D from 3D (offline)	TV 2D	TA 2D from 3D (offline)	TA 2D and 3D
	Interhemispheric Fissure	10-12	NS	NR		18 *present*	21 *present*	NR
	Central Sulcus	NV	NR	23.2-27.9	NR	28 (26)	26	27 (26)
	Postcentral Sulcus	NV	NR	23.2-27.9	NR	30 (28)	28	NR
	Precentral Sulcus	NV	NR	NR	NR	30 (28)	27	NR
	Sylvian Fissure	19	18-22	17	18-19 (16-17)	18 *present*	21 *present*	22
	Superior Frontal Sulcus	NR	NR	NR	NR	30 (28)	28	NR
	Inferior Frontal Sulcus	NR	NR	NR	NR	30 (28)	30	NR
	Superior Temporal Sulcus	26-28	NR	27	NR	30 (28)	27	26 (25)
	Inferior Temporal Sulcus	NR	NR	NR	NR	30 (28)	32	NR
	Occipital Sulci	NR	NR	NR	NR	30 (28)	34	NR
	Insular Sulci	NS	NS	NR	NR	32 (30)	34	NR
	Parieto-Occipital Fissure	23	18	18.5-20.5	(22-24)	20 (18)	NR	23 (22)
	Calcarine Fissure	23	18	18.5-21.9	22-24 (20-21)	22 (20)	24	23 (22)
	Callosal Sulcus	NR	14	NR	NR	18, *already present*	24	23 (22)
	Cingulate Sulcus	22-24	26	23.2-24.3	20-21 (18-19)	24 (20-22)	NR	26 (23)
	Secondary Cingulate Sulci	NR	NR	NR	NR	32 (30)	NV	NR
	Collateral Sulcus	NR	NR	NR	NR	NR	26	NR
	Olfactory Sulcus	NR	NR	NR	NR	30 (28)	NR	NR
	Hippocampal Fissure	23-25	NR	NR	NR	18, *already present*	21, *already present*	NR

B

Age-Specific Landmarks of Neurodevelopmental Maturation

This section provides illustrations of the features automatically extracted from 3D US images of fetuses at six different gestational weeks: 18 GW, 20 GW, 22 GW, 24 GW, 26 GW, 28 GW, and 30 GW. Ovals represent the locations and scales of features in the axial (red), coronal (green), and sagittal (blue) slices. Solid yellow lines represent image features of highest importance ($\geq 80\%$ selection frequency) at the given GA and are present in all slices, and dashed white lines represent regions present within the given slice.

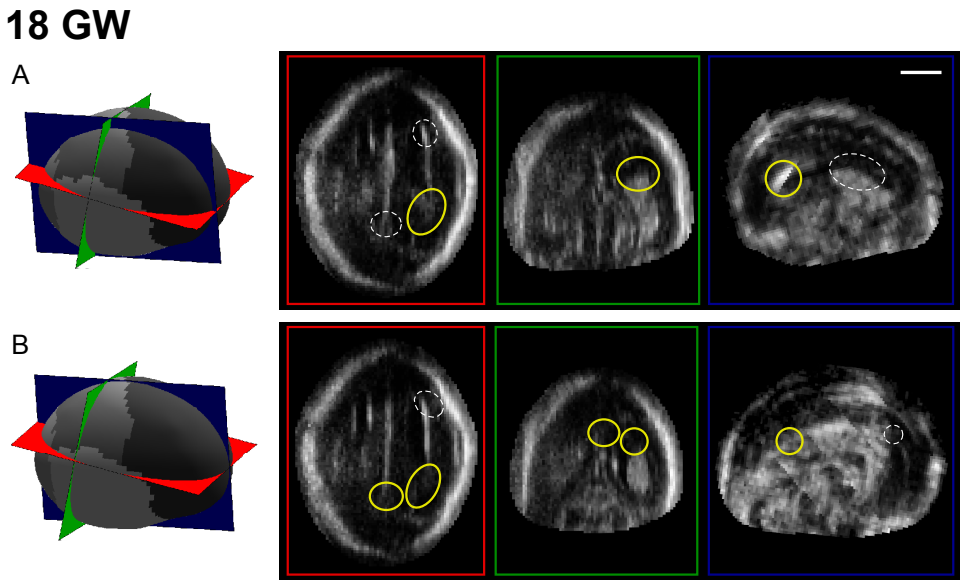


Figure B.1: **GA-specific developmental landmarks at 18 GW.** Key regions include (in order of importance): choroid plexus, posterior ventricle cavity. Scale bar: 1 cm.

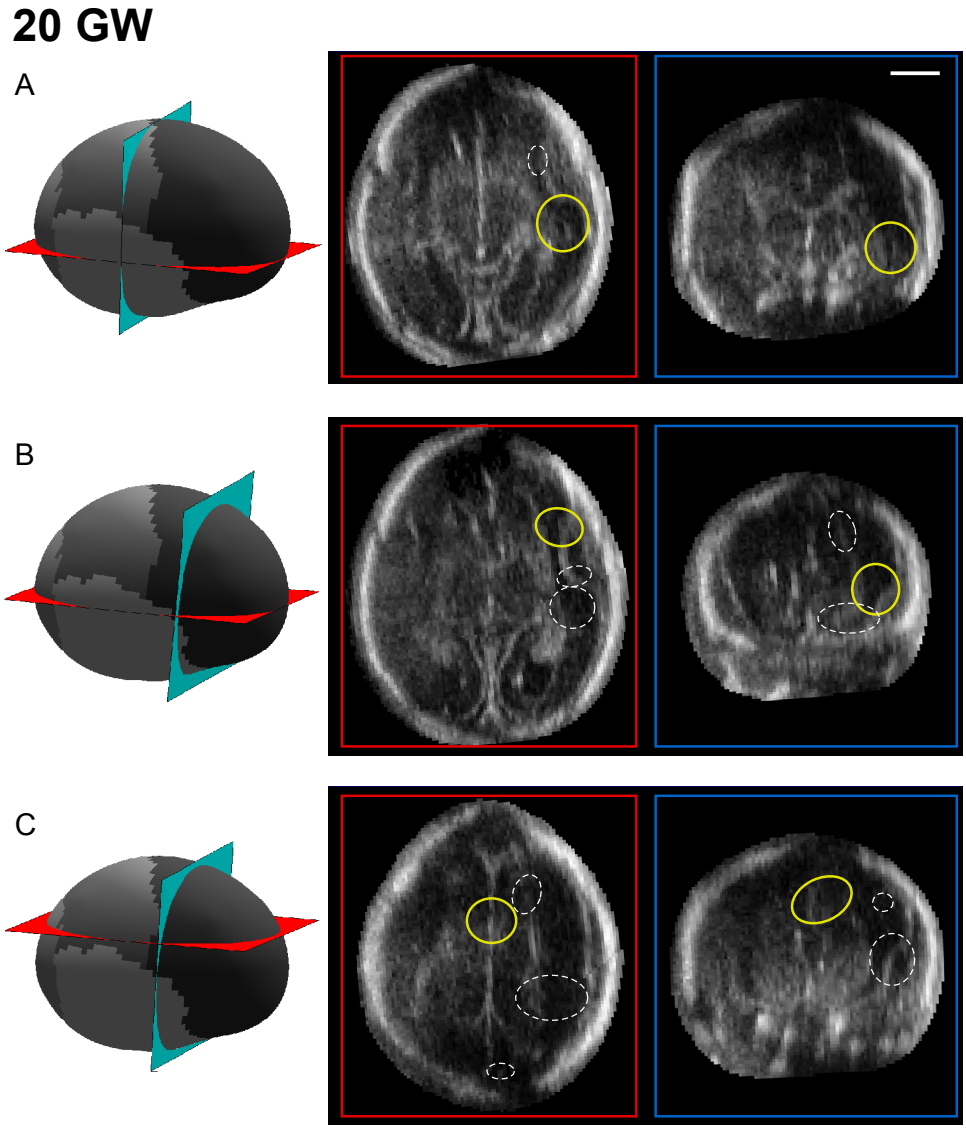


Figure B.2: **GA-specific developmental landmarks at 20 GW.** Key regions include (in order of importance): cingulate sulcus, front part of Sylvian fissure. Scale bar: 1 cm.

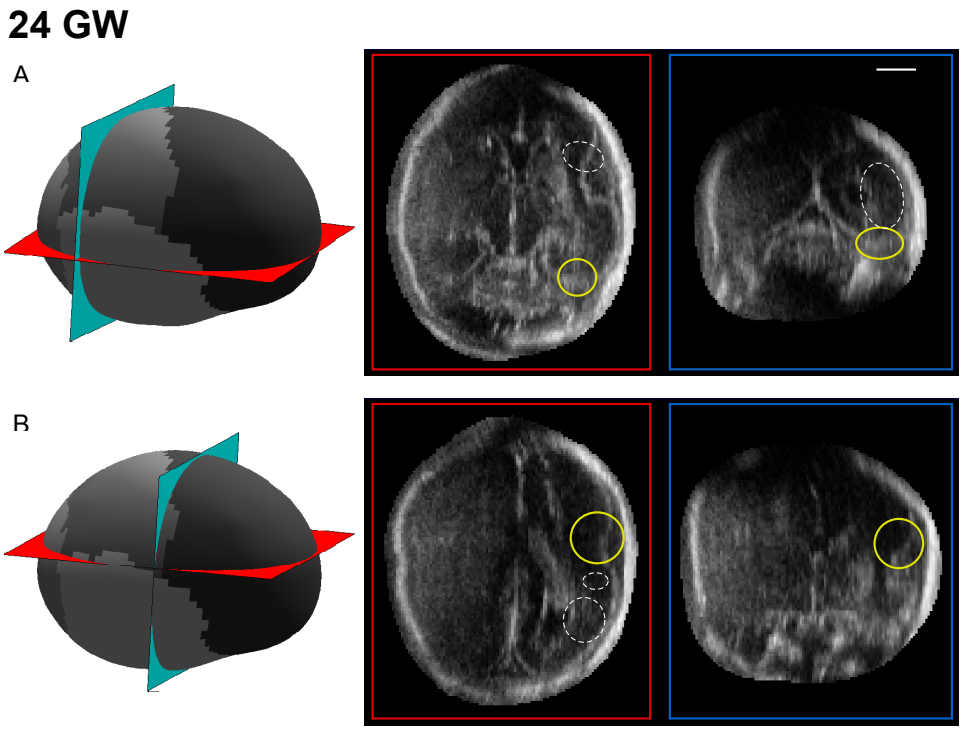


Figure B.3: **GA-specific developmental landmarks at 24 GW.** Key regions include (in order of importance): front part of Sylvian fissure, calcarine sulcus. Scale bar: 1 cm.

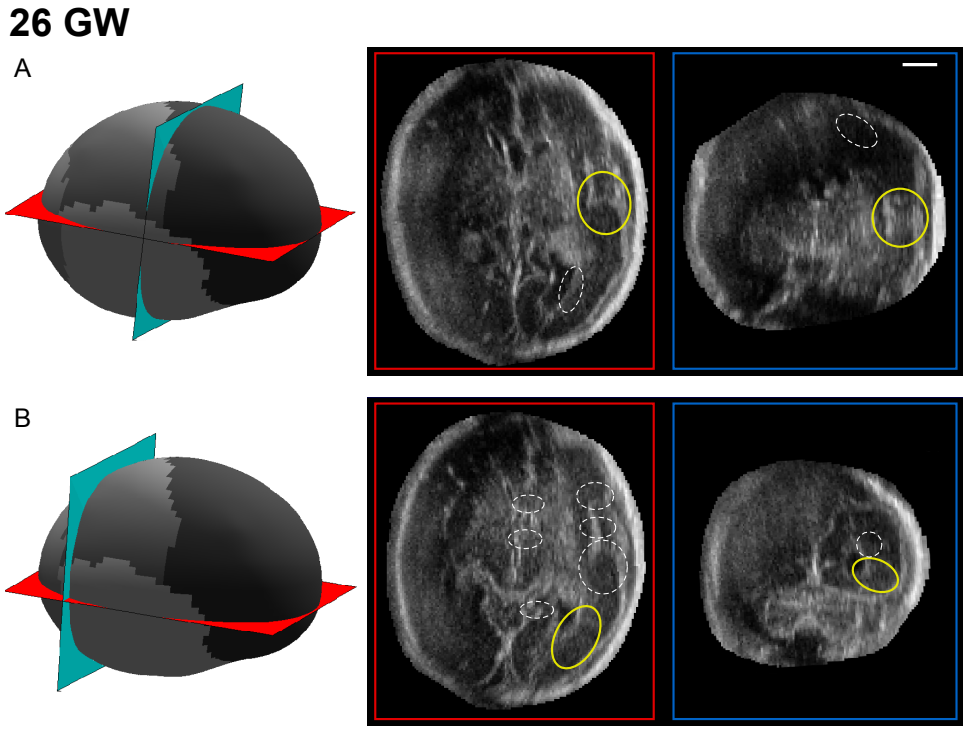


Figure B.4: **GA-specific developmental landmarks at 26 GW.** Key regions include (in order of importance): Sylvian fissure, posterior ventricle. Scale bar: 1 cm.

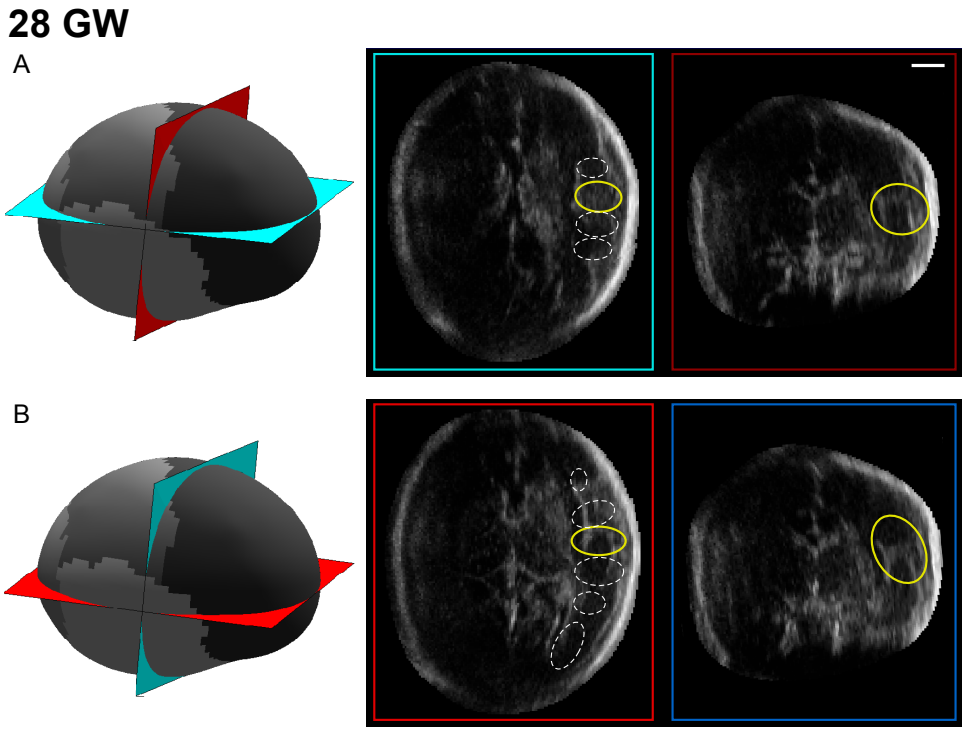


Figure B.5: **GA-specific developmental landmarks at 28 GW.** Key regions include: Sylvian fissure, particularly the rostral part. Scale bar: 1 cm.

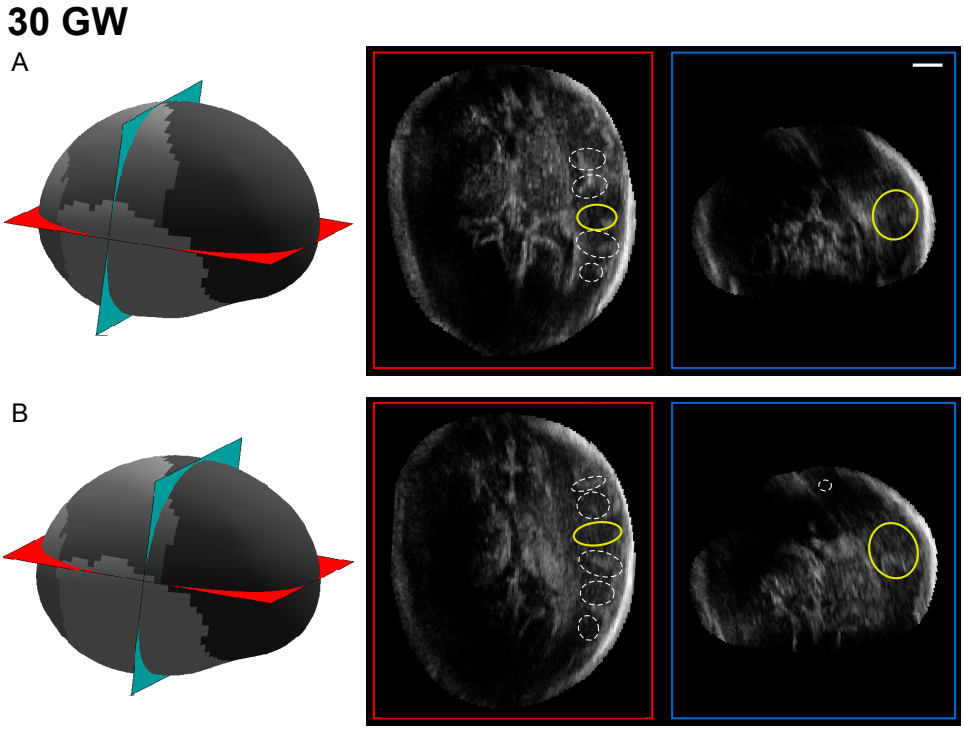


Figure B.6: **GA-specific developmental landmarks at 30 GW.** Key regions include (in order of importance): superior temporal sulcus, Sylvian fissure. Scale bar: 1 cm.

Bibliography

[1] R. T. Geirsson, "Ultrasound instead of last menstrual period as the basis of gestational age assignment," *Ultrasound in Obstetrics & Gynecology*, vol. 1, pp. 212–219, May 1991.

[2] F. P. Hadlock, R. L. Deter, R. B. Harrist, and S. K. Park, "Estimating fetal age: Computer-assisted analysis of multiple fetal growth parameters," *Radiology*, vol. 152, pp. 497–501, Aug 1984.

[3] D. G. Altman and L. S. Chitty, "New charts for ultrasound dating of pregnancy," *Ultrasound in Obstetrics & Gynecology*, vol. 10, pp. 174–191, Sep 1997.

[4] S. Ghai, K. W. Fong, A. Toi, D. Chitayat, S. Pantazi, and S. Blaser, "Prenatal US and MR imaging findings of lissencephaly: Review of fetal cerebral sulcal development," *RadioGraphics*, vol. 26, pp. 389–405, Apr 2006.

[5] K. Franke, E. Luders, A. May, M. Wilke, and C. Gaser, "Brain maturation: Predicting individual BrainAGE in children and adolescents using structural MRI," *NeuroImage*, vol. 63, pp. 1305–1312, Nov 2012.

[6] E. Dittrich, T. R. Raviv, G. Kasprian, R. Donner, P. C. Brugger, D. Prayer, and G. Langs, "A spatio-temporal latent atlas for semi-supervised learning of fetal brain segmentations and morphological age estimation," *Medical Image Analysis*, vol. 18, pp. 9–21, Jan 2014.

[7] R. Wright, V. Kyriakopoulou, C. Ledig, M. A. Rutherford, J. V. Hajnal, D. Rueckert, and P. Aljabar, "Automatic quantification of normal cortical folding patterns from fetal brain MRI," *NeuroImage*, vol. 91, pp. 21 – 32, May 2014.

[8] A. Toi, W. S. Lister, and K. W. Fong, "How early are fetal cerebral sulci visible at prenatal ultrasound and what is the normal pattern of early fetal sulcal development?," *Ultrasound in Obstetrics & Gynecology*, vol. 24, pp. 706–715, Dec 2004.

[9] P. Loughna, L. Chitty, T. Evans, and T. Chudleigh, "Fetal size and dating: Charts recommended for clinical obstetric practice," *Ultrasound*, vol. 17, pp. 160–166, Aug 2009.

[10] C. Bottomley and T. Bourne, "Dating and growth in the first trimester," *Best Practice & Research Clinical Obstetrics & Gynaecology*, vol. 23, pp. 439 – 452, Aug 2009.

[11] S. Campbell, S. L. Warsof, D. Little, and D. J. Cooper, "Routine ultrasound screening for the prediction of gestational age," *Obstetrics & Gynecology*, vol. 65, pp. 613–620, May 1985.

[12] M. H. Hall, R. A. Carr-Hill, C. Fraser, D. Campbell, and M. L. Samphier, "The extent and antecedents of uncertain gestation," *British Journal of Obstetrics & Gynaecology*, vol. 92, pp. 445–451, May 1985.

[13] P. Bergsjø, D. W. Denman III, H. J. Hoffman, and O. Meirik, "Duration of human singleton pregnancy. A population-based study," *Acta Obstetricia et Gynecologica Scandinavica*, vol. 69, pp. 197–207, Jan 1990.

[14] D. K. Waller, W. D. Spears, Y. Gu, and G. C. Cunningham, "Assessing number-specific error in the recall of onset of last menstrual period," *Paediatric and Perinatal Epidemiology*, vol. 14, pp. 263–267, Jul 2000.

[15] E. M. Walker, M. Lewis, W. Cooper, M. Marnie, and P. W. Howie, "Occult biochemical pregnancy: Fact or fiction?," *British Journal of Obstetrics & Gynaecology*, vol. 95, pp. 659–663, Jul 1988.

[16] J. G. Chi, E. C. Dooling, and F. H. Gilles, "Gyral development of the human brain," *Annals of Neurology*, vol. 1, pp. 86–93, Jan 1977.

[17] International Society of Ultrasound in Obstetrics and Gynecology (ISUOG), "Sonographic examination of the fetal central nervous system: Guidelines for performing the 'basic examination' and the 'fetal neurosonogram'," *Ultrasound in Obstetrics & Gynecology*, vol. 29, pp. 109–116, Jan 2007.

[18] C. Ioannou, K. Talbot, E. Ohuma, I. Sarris, J. Villar, A. Conde-Agudelo, and A. T. Papageorghiou, "Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size," *British Journal of Obstetrics & Gynaecology*, vol. 119, pp. 1425–1439, Nov 2012.

[19] W. J. Ott, "Accurate gestational dating: Revisited," *American Journal of Perinatology*, vol. 11, pp. 404–408, Nov 1994.

[20] P. M. Doubilet and C. B. Benson, "Improved prediction of gestational age in the late third trimester," *Journal of Ultrasound in Medicine*, vol. 12, pp. 647–653, Nov 1993.

[21] F. Chervenak, G. Isaacson, and J. Lorber, *Anomalies of the Fetal Head, Neck and Spine: Ultrasound Diagnosis and Management.* Philadelphia: W B Saunders Co, 1989.

[22] C. B. Benson and P. M. Doubilet, "Sonographic prediction of gestational age: Accuracy of second- and third-trimester fetal measurements," *American Journal of Roentgenology*, vol. 157, pp. 1275–1277, Dec 1991.

[23] M. Mongelli, N. G. Yuxin, A. Biswas, and S. Chew, "Accuracy of ultrasound dating formulae in the late second-trimester in pregnancies conceived with in-vitro fertilization," *Acta Radiologica*, vol. 44, pp. 452–455, Jul 2003.

[24] S. Campbell and G. B. Newman, "Growth of the fetal biparietal diameter during normal pregnancy," *Journal of Obstetrics & Gynaecology of the British Commonwealth*, vol. 78, pp. 513–519, Jun 1971.

[25] B. Hearn-Stebbins, "Normal fetal growth assessment: A review of literature and current practice," *Journal of Diagnostic Medical Sonography*, vol. 11, pp. 176–187, Jul 1995.

[26] A. B. Kurtz, R. J. Wapner, R. J. Kurtz, D. D. Dershaw, C. S. Rubin, C. Cole-Beuglet, and B. B. Goldberg, "Analysis of biparietal diameter as an accurate indicator of gestational age," *Journal of Clinical Ultrasound*, vol. 8, pp. 319–326, Aug 1980.

[27] L. S. Chitty, D. G. Altman, A. Henderson, and S. Campbell, "Charts of fetal size: 2. Head measurements," *British Journal of Obstetrics & Gynaecology*, vol. 101, pp. 35–43, Jan 1994.

[28] L. S. Chitty, D. G. Altman, A. Henderson, and S. Campbell, "Charts of fetal size: 4. Femur length," *British Journal of Obstetrics & Gynaecology*, vol. 101, pp. 132–135, Feb 1994.

[29] L. S. Chitty, D. G. Altman, A. Henderson, and S. Campbell, "Charts of fetal size: 3. Abdominal measurements," *British Journal of Obstetrics & Gynaecology*, vol. 101, pp. 125–131, Feb 1994.

[30] A. Selbing and B. Kjessler, "Conceptual dating by ultrasonic measurement of the fetal biparietal diameter in early pregnancy," *Acta Obstetricia et Gynecologica Scandinavica*, vol. 64, no. 7, pp. 593–597, 1985.

[31] J. T. Queenan, G. D. O'Brien, and S. Campbell, "Ultrasound measurement of fetal limb bones," *American Journal of Obstetrics & Gynecology*, vol. 138, pp. 297–302, Oct 1980.

[32] P. Jeanty, F. Rodesch, D. Delbeke, and J. E. Dumont, "Estimation of gestational age from measurements of fetal long bones," *Journal of Ultrasound in Medicine*, vol. 3, pp. 75–79, Feb 1984.

[33] P. H. Persson and B. M. Weldner, "Reliability of ultrasound fetometry in estimating gestational age in the second trimester," *Acta Obstetricia et Gynecologica Scandinavica*, vol. 65, no. 5, pp. 481–483, 1986.

[34] T. Mul, M. Mongelli, and J. Gardosi, "A comparative analysis of second-trimester ultrasound dating formulae in pregnancies conceived with artificial reproductive techniques," *Ultrasound in Obstetrics & Gynecology*, vol. 8, pp. 397–402, Dec 1996.

[35] F. P. Hadlock, R. B. Harrist, Y. P. Shah, D. E. King, S. K. Park, and R. S. Sharman, "Estimating fetal age using multiple parameters: A prospective evaluation in a racially mixed population," *American Journal of Obstetrics & Gynecology*, vol. 156, pp. 955–957, Apr 1987.

[36] C. W. Hohler, "Ultrasound estimation of gestational age," *Clinical Obstetrics & Gynecology*, vol. 27, pp. 314–326, Jun 1984.

[37] M. Mongelli, S. Chew, N. G. Yuxin, and A. Biswas, "Third-trimester ultrasound dating algorithms derived from pregnancies conceived with artificial reproductive techniques," *Ultrasound in Obstetrics & Gynecology*, vol. 26, pp. 129–131, Aug 2005.

[38] F. P. Hadlock, R. L. Deter, R. B. Harrist, and S. K. Park, "Computer assisted analysis of fetal age in the third trimester using multiple fetal growth parameters," *Journal of Clinical Ultrasound*, vol. 11, pp. 313–316, Aug 1983.

[39] K. Dorovini-Zis and C. L. Dolman, "Gestational development of brain," *Archives of Pathology & Laboratory Medicine*, vol. 101, pp. 192–195, Apr 1977.

[40] L. R. Pistorius, P. Stoutenbeek, F. Groenendaal, L. de Vries, G. Manten, E. Mulder, and G. Visser, "Grade and symmetry of normal fetal cortical development: a longitudinal two- and three-dimensional ultrasound study," *Ultrasound in Obstetrics & Gynecology*, vol. 36, pp. 700–708, Dec 2010.

[41] E. Armstrong, A. Schleicher, H. Omran, M. Curtis, and K. Zilles, "The ontogeny of human gyrification," *Cerebral Cortex*, vol. 5, pp. 56–63, Jan-Feb 1995.

[42] T. P. Naidich, J. L. Grant, N. Altman, R. A. Zimmerman, S. B. Birchansky, B. Braffman, and J. L. Daniel, "The developing cerebral surface. Preliminary report on the patterns of sulcal and gyral maturation–anatomy, ultrasound, and magnetic resonance imaging," *Neuroimaging Clinics of North America*, vol. 4, pp. 201–240, May 1994.

[43] B. Cohen-Sacher, T. Lerman-Sagie, D. Lev, and G. Malinger, "Sonographic developmental milestones of the fetal cerebral cortex: A longitudinal study," *Ultrasound in Obstetrics & Gynecology*, vol. 27, pp. 494–502, May 2006.

[44] A. Monteagudo and I. E. Timor-Tritsch, "Development of fetal gyri, sulci and fissures: A transvaginal sonographic study," *Ultrasound in Obstetrics & Gynecology*, vol. 9, pp. 222–228, Apr 1997.

[45] K. Nishikuni, *Estudo do desenvolvimento morfológico fetal e pós-natal dos sulcos cerebrais.* PhD thesis, Universidade de São Paulo, 2006.

[46] A. Afif, R. Bouvier, A. Buenerd, J. Trouillas, and P. Mertens, "Development of the human fetal insular cortex: Study of the gyration from 13 to 28 gestational weeks," *Brain Structure & Function*, vol. 212, pp. 335–346, Dec 2007.

[47] P. E. Hansen, M. C. Ballesteros, K. Soila, L. Garcia, and J. M. Howard, "MR imaging of the developing human brain," *RadioGraphics*, vol. 13, pp. 21–36, Jan 1993.

[48] D. Levine and P. D. Barnes, "Cortical maturation in normal and abnormal fetuses as assessed with prenatal MR imaging," *Radiology*, vol. 210, pp. 751–758, Mar 1999.

[49] L. M. Lan, Y. Yamashita, Y. Tang, T. Sugahara, M. Takahashi, T. Ohba, and H. Okamura, "Normal fetal brain development: MR imaging with a half-Fourier rapid acquisition with relaxation enhancement sequence," *Radiology*, vol. 215, pp. 205–210, Apr 2000.

[50] K. Ruoss, K. Lövblad, G. Schroth, A. C. Moessinger, and C. Fusch, "Brain development (sulci and gyri) as assessed by early postnatal MR imaging in preterm and term newborn infants," *Neuropediatrics*, vol. 32, pp. 69–74, Apr 2001.

[51] C. Garel, E. Chantrel, M. Elmaleh, H. Brisse, and G. Sebag, "Fetal MRI: Normal gestational landmarks for cerebral biometry, gyration and myelination," *Child's Nervous System*, vol. 19, pp. 422–425, Aug 2003.

[52] Z. Zhang, Z. Hou, X. Lin, G. Teng, H. Meng, F. Zang, F. Fang, and S. Liu, "Development of the fetal cerebral cortex in the second trimester: Assessment with 7T postmortem MR imaging," *American Journal of Neuroradiology*, Feb 2013.

[53] C. Bernard, P. Droulle, F. Didier, H. Gérard, J. Larroche, F. Plenat, F. Bomsel, J. Roland, and J. Hoeffel, "[Echographic aspects of cerebral sulci in the ante- and perinatal period]," *Journal of Radiology*, vol. 69, pp. 521–532, Aug-Sept 1988.

[54] F. F. Correa, C. Lara, J. Bellver, J. Remohí, A. Pellicer, and V. Serra, "Examination of the fetal brain by transabdominal three-dimensional ultrasound: Potential for routine neurosonographic studies," *Ultrasound in Obstetrics & Gynecology*, vol. 27, pp. 503–508, May 2006.

[55] A. Ruiz, C. Sembely-Taveau, C. Paillet, and D. Sirinelli, "Repeères échographiques de gyration cérébrale foetale normale," *Journal of Radiology*, vol. 87, pp. 49–55, Jan 2006.

[56] G. Retzius, *Das Menschenhirn. Studien in der makroskopischen Morphologie.* Stockholm: P.A. Norstedt, 1896.

[57] A.-M. Guihard-Costa and J.-C. Larroche, "Differential growth between the fetal brain and its infratentorial part," *Early Human Development*, vol. 23, pp. 27 – 40, Jun 1990.

[58] H. Kretschmann, *Brain Growth.* Bibliotheca Anatomica, Karger, 1986.

[59] W. R. Shankle, B. H. Landing, M. S. Rafii, A. Schiano, J. M. Chen, and J. Hara, "Evidence for a postnatal doubling of neuron number in the developing human cerebral cortex between 15 months and 6 years," *Journal of Theoretical Biology*, vol. 191, pp. 115–140, Mar 1998.

[60] J. Dubois, M. Benders, C. Borradori-Tolsa, A. Cachia, F. Lazeyras, R. Ha-Vinh Leuchter, S. V. Sizonenko, S. K. Warfield, J.-F. Mangin, and P. S. Hüppi, "Primary cortical folding in the human newborn: An early marker of later functional development," *Brain*, vol. 131, pp. 2028–2041, Aug 2008.

[61] T. E. Inder, S. K. Warfield, H. Wang, P. S. Hüppi, and J. J. Volpe, "Abnormal cerebral structure is present at term in premature infants," *Pediatrics*, vol. 115, pp. 286–294, Feb 2005.

[62] O. Kapellou, S. J. Counsell, N. Kennea, L. Dyet, N. Saeed, J. Stark, E. Maalouf, P. Duggan, M. Ajayi-Obe, J. Hajnal, J. M. Allsop, J. Boardman, M. A. Rutherford, F. Cowan, and A. D. Edwards, "Abnormal cortical development after premature birth shown by altered allometric scaling of brain growth," *Public Library of Science (PLoS) Medicine*, vol. 3, p. e265, Aug 2006.

[63] C. Garel, E. Chantrel, H. Brisse, M. Elmaleh, D. Luton, J. F. Oury, G. Sebag, and M. Hassan, "Fetal cerebral cortex: Normal gestational landmarks identified using prenatal MR imaging," *American Journal of Neuroradiology*, vol. 22, pp. 184–189, Jan 2001.

[64] B. Ertl-Wagner, A. Lienemann, A. Strauss, and M. Reiser, "Fetal magnetic resonance imaging: Indications, technique, anatomical considerations and a review of fetal abnormalities," *European Radiology*, vol. 12, pp. 1931–1940, Aug 2002.

[65] O. Glenn and A. Barkovich, "Magnetic resonance imaging of the fetal brain and spine: An increasingly important tool in prenatal diagnosis, part 1," *American Journal of Neuroradiology*, vol. 27, pp. 1604–1611, Sept 2006.

[66] A. Perrone, S. Savelli, C. Maggi, L. Pietro, M. Maurizio, J. Tesei, L. Ballesio, C. Felice, A. Giancotti, R. Iorio, and L. Manganaro, "Magnetic resonance imaging versus ultrasonography in fetal pathology," *Radiologia Medica*, vol. 113, pp. 225–241, Apr 2008.

[67] A. Gholipour, J. A. Estroff, and S. K. Warfield, "Robust super-resolution volume reconstruction from slice acquisitions: Application to fetal brain MRI," *IEEE Transactions on Medical Imaging*, vol. 29, pp. 1739–1758, Oct 2010.

[68] M. Kuklisova-Murgasova, G. Quaghebeur, M. A. Rutherford, J. V. Hajnal, and J. A. Schnabel, "Reconstruction of fetal brain MRI with intensity matching and complete outlier removal," *Medical Image Analysis*, vol. 16, pp. 1550–1564, Dec 2012.

[69] F. Rousseau, K. Kim, C. Studholme, M. Koob, and J.-L. Dietemann, "On super-resolution for fetal brain MRI," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, vol. 2 of *Lecture Notes on Computer Science*, pp. 355–362, Springer-Verlag, 2010.

[70] A. Monteagudo and I. E. Timor-Tritsch, "Normal sonographic development of the central nervous system from the second trimester onwards using 2D, 3D and transvaginal sonography," *Prenatal Diagnosis*, vol. 29, pp. 326–339, Apr 2009.

[71] E. Merz, *Ultrasound in Obstetrics and Gynecology*, vol. 2. Thieme, 2007.

[72] A. Monteagudo, "Fetal neurosonography: Should it be routine? Should it be detailed?," *Ultrasound in Obstetrics & Gynecology*, vol. 12, pp. 1–5, Jul 1998.

[73] R. Hering-Hanit, R. Achiron, S. Lipitz, and A. Achiron, "Asymmetry of fetal cerebral hemispheres: In utero ultrasound study," *Archives of Disease in Childhood. Fetal & Neonatal Edition*, vol. 85, pp. F194–F196, Nov 2001.

[74] J. Dubois, M. Benders, F. Lazeyras, C. Borradori-Tolsa, R. Ha-Vinh Leuchter, J.-F. Mangin, and P. Hüppi, "Structural asymmetries of perisylvian regions in the preterm newborn," *NeuroImage*, vol. 52, pp. 32 – 42, Aug 2010.

[75] Z. Kivilevitch, R. Achiron, and Y. Zalel, "Fetal brain asymmetry: In utero sonographic study of normal fetuses," *American Journal of Obstetrics & Gynecology*, vol. 202, pp. 359.e1–359.e8, Apr 2010.

[76] G. Kasprian, G. Langs, P. C. Brugger, M. Bittner, M. Weber, M. Arantes, and D. Prayer, "The prenatal origin of hemispheric asymmetry: An in utero neuroimaging study," *Cerebral Cortex*, vol. 21, pp. 1076–1083, May 2011.

[77] A. Monteagudo, I. E. Timor-Tritsch, and P. Mayberry, "Three-dimensional transvaginal neurosonography of the fetal brain: 'Navigating' in the volume scan," *Ultrasound in Obstetrics & Gynecology*, vol. 16, pp. 307–313, Sept 2000.

[78] F. Yang, K. Y. Leung, Y. P. Lee, H. Y. Chan, and M. H. Y. Tang, "Fetal biometry by an inexperienced operator using two- and three-dimensional ultrasound," *Ultrasound in Obstetrics & Gynecology*, vol. 35, pp. 566–571, May 2010.

[79] D. Pugash, G. Hendson, C. P. Dunham, K. Dewar, D. M. Money, and D. Prayer, "Sonographic assessment of normal and abnormal patterns of fetal cerebral lamination," *Ultrasound in Obstetrics & Gynecology*, vol. 40, pp. 642–651, Dec 2012.

[80] M. Kuklisova-Murgasova, A. Cifor, R. Napolitano, A. Papageorghiou, G. Quaghebeur, M. A. Rutherford, J. V. Hajnal, J. A. Noble, and J. A. Schnabel, "Registration of 3D fetal neurosonography and MRI," *Medical Image Analysis*, vol. 17, pp. 1137–1150, Dec 2013.

[81] L. R. Pistorius, P. M. Hellmann, G. H. A. Visser, G. Malinger, and D. Prayer, "Fetal neuroimaging: Ultrasound, MRI, or both?," *Obstetrical & Gynecological Survey*, vol. 63, pp. 733–745, Nov 2008.

[82] P. Govaert, R. Swarte, A. De Vos, and M. Lequin, "Sonographic appearance of the normal and abnormal insula of Reil," *Developmental Medicine & Child Neurology*, vol. 46, pp. 610–616, Sept 2004.

[83] K. Zilles, E. Armstrong, A. Schleicher, and H. J. Kretschmann, "The human pattern of gyrification in the cerebral cortex," *Anatomy & Embryology*, vol. 179, pp. 173–179, Nov 1988.

[84] E. Armstrong, M. Curtis, D. P. Buxhoeveden, C. Fregoe, K. Zilles, M. F. Casanova, and W. F. McCarthy, "Cortical gyrification in the rhesus monkey: A test of the mechanical folding hypothesis," *Cerebral Cortex*, vol. 1, pp. 426–432, Sept-Oct 1991.

[85] Z. Zhang, S. Liu, X. Lin, B. Sun, T. Yu, and H. Geng, "Development of fetal cerebral cortex: Assessment of the folding conditions with post-mortem magnetic resonance imaging," *International Journal of Developmental Neuroscience*, vol. 28, pp. 537–543, Oct 2010.

[86] A. Serag, P. Aljabar, G. Ball, S. J. Counsell, J. P. Boardman, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert, "Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression," *NeuroImage*, vol. 59, pp. 2255–2265, Feb 2012.

[87] I. Alonso, M. Borenstein, G. Grant, I. Narbona, and G. Azumendi, "Depth of brain fissures in normal fetuses by prenatal ultrasound between 19 and 30 weeks of gestation," *Ultrasound in Obstetrics & Gynecology*, vol. 36, pp. 693–699, Dec 2010.

[88] C. M. S. Alves, E. A. Jr., L. M. M. Nardozza, P. S. de Oliveira, S. M. Goldman, S. A. Ajzen, and A. F. Moron, "[Fetal brain fissures development a three-dimensional ultrasonography study]," *Revista Brasileira de Ginecologia e Obstetrícia*, vol. 33, pp. 111–117, Mar 2011.

[89] P. Mittal, L. F. Gonalves, J. P. Kusanovic, J. Espinoza, W. Lee, J. K. Nien, E. Soto, and R. Romero, "Objective evaluation of Sylvian fissure development by multiplanar 3-dimensional ultrasonography," *Journal of Ultrasound in Medicine*, vol. 26, pp. 347–353, Mar 2007.

[90] E. Quarello, J. Stirnemann, Y. Ville, and L. Guibaud, "Assessment of fetal Sylvian fissure operculization between 22 and 32 weeks: A subjective approach," *Ultrasound in Obstetrics & Gynecology*, vol. 32, pp. 44–49, Jul 2008.

[91] J. Dubois, M. Benders, A. Cachia, F. Lazeyras, R. Ha-Vinh Leuchter, S. V. Sizonenko, C. Borradori-Tolsa, J.-F. Mangin, and P. S. Hüppi, "Mapping the early cortical folding process in the preterm newborn brain," *Cerebral Cortex*, vol. 18, pp. 1444–1454, Jun 2008.

[92] P. Kochunov, W. Rogers, J.-F. Mangin, and J. Lancaster, "A library of cortical morphology analysis tools to study development, aging and genetics of cerebral cortex," *NeuroInformatics*, vol. 10, pp. 81–96, Jan 2012.

[93] J.-F. Mangin, F. Poupon, E. Duchesnay, D. Rivière, A. Cachia, D. L. Collins, A. C. Evans, and J. Régis, "Brain morphometry using 3D moment invariants," *Medical Image Analysis*, vol. 8, pp. 187–196, Sept 2004.

[94] M. R. Hatab, S. W. Kamourieh, and D. M. Twickler, "MR volume of the fetal cerebellum in relation to growth," *Journal of Magnetic Resonance Imaging*, vol. 27, pp. 840–845, Apr 2008.

[95] B. Tilea, C. Alberti, C. Adamsbaum, P. Armoogum, J. F. Oury, D. Cabrol, G. Sebag, G. Kalifa, and C. Garel, "Cerebral biometry in fetal magnetic resonance imaging: New reference data," *Ultrasound in Obstetrics & Gynecology*, vol. 33, pp. 173–181, Feb 2009.

[96] J. Espinoza, S. Good, E. Russell, and W. Lee, "Does the use of automated fetal biometry improve clinical work flow efficiency?," *Journal of Ultrasound in Medicine*, vol. 32, pp. 847–850, May 2013.

[97] C. Hanna and A. Youssef, "Automated measurements in obstetric ultrasound images," in *International Conference on Image Processing (ICIP)*, vol. 3, pp. 504–507, Oct 1997.

[98] M. H. Marhaban, R. S. Kaid, and S. B. Mohd Noor, "Automatic estimation of gestational age in ultrasound images based on direct least-squares fitting of ellipse," in *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 5, pp. 569–573, Wiley Subscription Services, Inc., Sept 2010.

[99] G. K. Matsopoulos and S. Marshall, "Use of morphological image processing techniques for the measurement of a fetal head from ultrasound images," *Pattern Recognition*, vol. 27, pp. 1317–1324, Oct 1994.

[100] W. Lu and J. Tan, "Segmentation of ultrasound fetal images," in *SPIE: International Society for Optics and Photonics*, vol. 4203 of *Biological Quality and Precision Agriculture*, pp. 81–90, SPIE, Dec 2000.

[101] W. Lu, J. Tan, and R. Floyd, "Automated fetal head detection and measurement in ultrasound images by iterative randomized Hough transform," *Ultrasound in Medicine & Biology*, vol. 31, pp. 929 – 936, Jul 2005.

[102] J. Yu, Y. Wang, and P. Chen, "Fetal ultrasound image segmentation system and its use in fetal weight estimation," *Medical and Biological Engineering and Computing*, vol. 46, pp. 1227–1237, Dec 2008.

[103] I. E. Zador, V. Salari, L. Chik, and R. J. Sokol, "Ultrasound measurement of the fetal head: Computer versus operator," *Ultrasound in Obstetrics & Gynecology*, vol. 1, pp. 208–211, May 1991.

[104] V. Chalana and Y. Kim, "A methodology for evaluation of boundary detection algorithms on medical images," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 642–652, Oct 1997.

[105] V. Chalana, T. C. Winter III, D. R. Cyr, D. R. Haynor, and Y. Kim, "Automatic fetal head measurements from sonographic images," *Academic Radiology*, vol. 3, pp. 628 – 635, Aug 1996.

[106] H.-C. Chen, P.-Y. Tsai, H.-H. Huang, H.-H. Shih, Y.-Y. Wang, C.-H. Chang, and Y.-N. Sun, "Registration-based segmentation of three-dimensional ultrasound images for quantitative measurement of fetalcraniofacial structure," *Ultrasound in Medicine & Biology*, vol. 38, pp. 811 – 823, May 2012.

[107] S. M. G. V. B. Jardim and M. A. T. Figueiredo, "Segmentation of fetal ultrasound images," *Ultrasound in Medicine & Biology*, vol. 31, pp. 243–250, Feb 2005.

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

[109] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, "Automatic fetal measurements in ultrasound using constrained probabilistic boosting tree," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, vol. 4792 of *Lecture Notes in Computer Science*, pp. 571–579, Springer Berlin / Heidelberg, 2007.

[110] M. Yaqub, R. Napolitano, C. Ioannou, A. T. Papageorghiou, and J. A. Noble, "Automatic detection of local fetal brain structures in ultrasound images," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1555–1558, May 2012.

[111] R. Cuingnet, O. Somphone, B. Mory, R. Prevost, M. Yaqub, R. Napolitano, A. T. Papageorghiou, D. Roundhill, J. A. Noble, and R. Ardon, "Where is my baby? a fast fetal head auto-alignment in 3D-ultrasound," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 768–771, Apr 2013.

[112] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree," *IEEE Transactions on Medical Imaging*, vol. 27, pp. 1342–1355, Sept 2008.

[113] G. Carneiro, B. Georgescu, and S. Good, "Knowledge-based automated fetal biometrics using syngo auto OB measurements: ACUSON S2000 Ultrasound System," tech. rep., Siemens Medical Solutions, 2008.

[114] B. Shan and M. Madheswaran, "Extraction of fetal biometrics using class separable shape sensitive approach for gestational age estimation," in *International Conference on Computer Technology and Development (ICCTD)*, vol. 2, pp. 376–380, Nov 2009.

[115] A. I. L. Namburete and J. A. Noble, "Fetal cranial segmentation in 2D ultrasound images using shape properties of pixel clusters," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 720–723, Apr 2013.

[116] J. Ashburner, *Computational Neuroanatomy*. PhD thesis, University College London, 2000.

[117] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. Henson, K. J. Friston, and R. S. Frackowiak, "A voxel-based morphometric study of ageing in 465 normal adult human brains," *NeuroImage*, vol. 14, pp. 21 – 36, Jul 2001.

[118] P. M. Thompson, J. N. Giedd, R. P. Woods, D. MacDonald, A. C. Evans, and A. W. Toga, "Growth patterns in the developing brain detected by using continuum mechanical tensor maps," *Nature*, vol. 404, pp. 190–193, Mar 2000.

[119] M. K. Chung, K. M. Dalton, and R. J. Davidson, "Tensor-based cortical surface morphometry via weighted spherical harmonic representation," *IEEE Transactions on Medical Imaging*, vol. 27, pp. 1143–1151, Aug 2008.

[120] V. Rajagopalan, J. Scott, P. A. Habas, K. Kim, J. Corbett-Detig, F. Rousseau, A. J. Barkovich, O. A. Glenn, and C. Studholme, "Local tissue growth patterns underlying normal fetal human brain gyrification quantified in utero," *Journal of Neuroscience*, vol. 31, pp. 2878–2887, Feb 2011.

[121] M. Toews, W. Wells III, D. L. Collins, and T. Arbel, "Feature-based morphometry: Discovering group-related anatomical patterns," *NeuroImage*, vol. 49, pp. 2318–2327, Feb 2010.

[122] M. Toews, W. M. Wells III, and L. Zöllei, "A feature-based developmental model of the infant brain in structural MRI," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)* (N. Ayache, H. Delingette, P. Golland, and K. Mori, eds.), vol. 7511 of *Lecture Notes in Computer Science*, pp. 204–211, Springer Berlin Heidelberg, 2012.

[123] A. I. L. Namburete, M. Yaqub, B. Kemp, A. Papageorghiou, and J. A. Noble, "Predicting fetal neurodevelopmental age from ultrasound images," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)* (P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, eds.), vol. 8674 of *Lecture Notes in Computer Science*, pp. 260–267, Springer International Publishing, 2014.

[124] C. Studholme, "Mapping fetal brain development in utero using magnetic resonance imaging: The big bang of brain mapping," *Annual Review of Biomedical Engineering*, vol. 13, pp. 345–368, Aug 2011.

[125] M. Schaer, M. Cuadra, L. Tamarit, F. Lazeyras, S. Eliez, and J.-P. Thiran, "A surface-based approach to quantify local cortical gyrification," *IEEE Transactions on Medical Imaging*, vol. 27, pp. 161–170, Feb 2008.

[126] R. Toro, M. Perron, B. Pike, L. Richer, S. Veillette, Z. Pausova, and T. Paus, "Brain size and folding of the human cerebral cortex," *Cerebral Cortex*, vol. 18, pp. 2352–2357, Feb 2008.

[127] M. Schaer, M. B. Cuadra, N. Schmansky, B. Fischl, J.-P. Thiran, and S. Eliez, "How to measure cortical folding from MR images: A step-by-step tutorial to compute local gyrification index," *Journal of Visualized Experiments*, p. e3417, Jan 2012.

[128] C. Clouchoux, D. Kudelski, A. Gholipour, S. K. Warfield, S. Viseur, M. Bouyssi-Kobar, J.-L. Mari, A. C. Evans, A. J. du Plessis, and C. Limperopoulos, "Quantitative in vivo MRI measurement of cortical development in the fetus," *Brain Structure & Function*, vol. 217, pp. 127–139, Jan 2012.

[129] H.-H. Hu, W.-Y. Guo, H.-Y. Chen, P.-S. Wang, C.-I. Hung, J.-C. Hsieh, and Y.-T. Wu, "Morphological regionalization using fetal magnetic resonance images of normal developing brains," *European Journal of Neuroscience*, vol. 29, pp. 1560–1567, Apr 2009.

[130] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. B. Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," *Computer Methods and Programs in Biomedicine*, vol. 104, pp. e158–e177, Dec 2011.

[131] M. Altaye, S. K. Holland, M. Wilke, and C. Gaser, "Infant brain probability templates for MRI segmentation and normalization," *NeuroImage*, vol. 43, pp. 721–730, Dec 2008.

[132] A. Gholipour, A. Akhondi-Asl, J. A. Estroff, and S. K. Warfield, "Multi-atlas multi-shape segmentation of fetal brain MRI for volumetric and morphometric analysis of ventriculomegaly," *NeuroImage*, vol. 60, pp. 1819–1831, Apr 2012.

[133] M. Murgasova, L. Dyet, D. Edwards, M. Rutherford, J. Hajnal, and D. Rueckert, "Segmentation of brain MRI in young children," *Academic Radiology*, vol. 14, pp. 1350–1366, Nov 2007.

[134] M. Prastawa, J. H. Gilmore, W. Lin, and G. Gerig, "Automatic segmentation of MR images of the developing newborn brain," *Medical Image Analysis*, vol. 9, pp. 457 – 466, Oct 2005.

[135] F. Shi, P.-T. Yap, G. Wu, H. Jia, J. H. Gilmore, W. Lin, and D. Shen, "Infant brain atlases from neonates to 1- and 2-year-olds," *Public Library of Science (PLoS) One*, vol. 6, p. e18746, Apr 2011.

[136] N. I. Weisenfeld and S. K. Warfield, "Automatic segmentation of newborn brain MRI," *NeuroImage*, vol. 47, pp. 564–572, Aug 2009.

[137] E. Dittrich, T. Riklin-Raviv, G. Kasprian, P. Brugger, D. Prayer, and G. Langs, "Learning a spatio-temporal latent atlas for fetal brain segmentation," *Workshop on Image Analysis of Human Brain Development at MICCAI 2011*, pp. 9–16, Sept 2011.

[138] E. Dittrich, G. Kasprian, D. Prayer, and G. Langs, "Atlas learning in fetal brain development," *Topics in Magnetic Resonance Imaging*, vol. 22, pp. 107–111, Jun 2011.

[139] P. A. Habas, K. Kim, J. M. Corbett-Detig, F. Rousseau, O. A. Glenn, A. J. Barkovich, and C. Studholme, "A spatiotemporal atlas of MR intensity, tissue probability and shape of the fetal brain with application to segmentation," *NeuroImage*, vol. 53, pp. 460 – 470, Nov 2010.

[140] P. A. Habas, K. Kim, F. Rousseau, O. A. Glenn, A. J. Barkovich, and C. Studholme, "Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses," *Human Brain Mapping*, vol. 31, pp. 1348–1358, Sep 2010.

[141] M. Kuklisova-Murgasova, P. Aljabar, L. Srinivasan, S. J. Counsell, V. Doria, A. Serag, I. S. Gousias, J. P. Boardman, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert, "A dynamic 4D probabilistic atlas of the developing brain," *NeuroImage*, vol. 54, pp. 2750–2763, Feb 2011.

[142] E. Luders, P. M. Thompson, K. L. Narr, A. W. Toga, L. Jancke, and C. Gaser, "A curvature-based approach to estimate local gyrification on the cortical surface.," *NeuroImage*, vol. 29, pp. 1224–1230, Feb 2006.

[143] V. A. Magnotta, N. C. Andreasen, S. K. Schultz, G. Harris, T. Cizadlo, D. Heckel, P. Nopoulos, and M. Flaum, "Quantitative in vivo measurement of gyrification in the human brain: Changes associated with aging," *Cerebral Cortex*, vol. 9, pp. 151–160, Mar 1999.

[144] P. A. Habas, J. A. Scott, A. Roosta, V. Rajagopalan, K. Kim, F. Rousseau, A. J. Barkovich, O. A. Glenn, and C. Studholme, "Early folding patterns and asymmetries of the normal human brain detected from in utero MRI," *Cerebral Cortex*, vol. 22, pp. 13–25, Jan 2012.

[145] G. Lohmann, "Extracting line representations of sulcal and gyral patterns in MR images of the human brain," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 1040–1048, Dec 1998.

[146] C.-Y. Kao, M. Hofer, G. Sapiro, J. Stem, K. Rehm, and D. A. Rottenberg, "A geometric method for automatic extraction of sulcal fundi," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 530–540, Apr 2007.

[147] M. E. Rettmann, X. Han, C. Xu, and J. L. Prince, "Automated sulcal segmentation using watersheds on the cortical surface," *NeuroImage*, vol. 15, pp. 329–344, Feb 2002.

[148] P. Kochunov, C. Castro, D. Davis, D. Dudley, J. Brewer, Y. Zhang, C. D. Kroenke, D. Purdy, P. T. Fox, C. Simerly, and G. Schatten, "Mapping primary gyrogenesis during fetal development in primate brains: High-resolution in utero structural MRI of fetal brain development in pregnant baboons," *Frontiers in Neuroscience*, vol. 4, p. 20, May 2010.

[149] S. Powell, V. A. Magnotta, H. Johnson, V. K. Jammalamadaka, R. Pierson, and N. C. Andreasen, "Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures," *NeuroImage*, vol. 39, pp. 238–247, Jan 2008.

[150] V. Lempitsky, M. Verhoek, J. A. Noble, and A. Blake, "Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography," in *Workshop on Functional Imaging and Modeling of the Heart at MICCAI* (N. Ayache, H. Delingette, and M. Sermesant, eds.), vol. 5528 of *Lecture Notes in Computer Science*, pp. 447–456, Springer Berlin Heidelberg, 2009.

[151] Z. Yi, A. Criminisi, J. Shotton, and A. Blake, "Discriminative, semantic segmentation of brain tissue in MR images," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, eds.), vol. 5762 of *Lecture Notes in Computer Science*, (Berlin, Heidelberg), pp. 558–565, Springer-Verlag, 2009.

[152] B. Rahmatullah, A. T. Papageorghiou, and J. Noble, "Integration of local and global features for anatomical object detection in ultrasound," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (N. Ayache, H. Delingette, P. Golland, and K. Mori, eds.), vol. 7512 of *Lecture Notes in Computer Science*, pp. 402–409, Springer Berlin / Heidelberg, 2012.

[153] M. Yaqub, M. K. Javaid, C. Cooper, and J. A. Noble, "Improving the classification accuracy of the classic RF method by intelligent feature selection and weighted voting of trees with application to medical image segmentation," in *Workshop on Machine learning in medical imaging at MICCAI*, vol. 7009 of *Lecture Notes in Computer Science*, (Berlin, Heidelberg), pp. 184–192, Springer-Verlag, 2011.

[154] A. I. L. Namburete, B. Rahmatullah, and J. A. Noble, "Nakagami-based AdaBoost learning framework for detection of anatomical landmarks in 2D fetal neurosonograms," in *Annals of the British Machine Vision Association (BMVA)*, vol. 2013, pp. 1–16, Jul 2013.

[155] K. Suzuki, P. Yan, F. Wang, and D. Shen, "Machine learning in medical imaging," *International Journal of Biomedical Imaging*, vol. 2012, p. 2, 2012.

[156] R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, D. Rueckert, and for the Alzheimer's Disease Neuroimaging Initiative (ADNI), "LEAP: Learning embeddings for atlas propagation," *NeuroImage*, vol. 49, pp. 1316–1325, Jan 2010.

[157] B. Rahmatullah, I. Sarris, A. Papageorghiou, and J. Noble, "Quality control of fetal ultrasound images: Detection of abdomen anatomical landmarks using AdaBoost," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 6–9, Mar-Apr 2011.

[158] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, pp. 119 – 139, Aug 1997.

[159] P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision (IJCV)*, vol. 57, pp. 137–154, May 2004.

[160] A. Criminisi and J. Shotton, eds., *Decision Forests for Computer Vision and Medical Image Analysis*. Springer London Heidelberg, 2013.

[161] M. Yaqub, P. Mahon, M. K. Javaid, C. Cooper, and J. Noble, "Weighted voting in 3D random forest segmentation," in *Medical Image Understanding and Analysis (MIUA)*, pp. 261–266, Jul 2010.

[162] J. Hamm, D. H. Ye, R. Verma, and C. Davatzikos, "GRAM: A framework for geodesic registration on anatomical manifolds," *Medical Image Analysis*, vol. 14, pp. 633–642, Oct 2010.

[163] S. Gerber, T. Tasdizen, S. Joshi, and R. Whitaker, "On the manifold structure of the space of brain images," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, eds.), vol. 5761 of *Lecture Notes in Computer Science*, pp. 305–312, Springer Berlin Heidelberg, 2009.

[164] R. Verma, P. Khurd, and C. Davatzikos, "On analyzing diffusion tensor images by identifying manifold structure using isomaps," *IEEE Transactions on Medical Imaging*, vol. 26, pp. 772–778, Jun 2007.

[165] P. Khurd, S. Baloch, R. C. Gur, C. Davatzikos, and R. Verma, "Manifold learning techniques in image analysis of high-dimensional diffusion tensor magnetic resonance images," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 1–7, Jun 2007.

[166] S. Klein, M. Loog, F. van der Lijn, T. den Heijer, A. Hammers, M. de Bruijne, A. van der Lugt, R. Duin, M. Breteler, and W. Niessen, "Early diagnosis of dementia based on intersubject whole-brain dissimilarities," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 249 –252, Apr 2010.

[167] Y. Fu and T. Huang, "Human age estimation with regression on discriminative aging manifold," *IEEE Transactions on Multimedia*, vol. 10, pp. 578 –584, Jun 2008.

[168] P. Aljabar, R. Wolz, L. Srinivasan, S. J. Counsell, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert, "A combined manifold learning analysis of shape and appearance to characterize neonatal brain development," *IEEE Transactions on Medical Imaging*, vol. 30, pp. 2072–2086, Dec 2011.

[169] R. Wolz, P. Aljabar, J. V. Hajnal, J. Lötjönen, D. Rueckert, and for the Alzheimer's Disease Neuroimaging Initiative (ADNI), "Nonlinear dimensionality reduction combining MR imaging with non-imaging information," *Medical Image Analysis*, vol. 16, pp. 819–830, May 2012.

[170] Z. Y. Sun, M. Perrot, A. Tucholka, D. Rivière, and J.-F. Mangin, "Constructing a dictionary of human brain folding patterns," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, eds.), vol. 5762 of *Lecture Notes in Computer Science*, pp. 117–124, Springer Verlag Heidelberg, 2009.

[171] K. Franke, G. Ziegler, S. Klöppel, C. Gaser, and for the Alzheimer's Disease Neuroimaging Initiative (ADNI), "Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters," *NeuroImage*, vol. 50, pp. 883–892, Apr 2010.

[172] M. Sabuncu and K. Van Leemput, "The relevance voxel machine (RVoxM): A self-tuning Bayesian model for informative image-based prediction," *IEEE Transactions on Medical Imaging*, vol. 31, pp. 2290–2306, Dec 2012.

[173] V. N. Vapnik, *Statistical Learning Theory*. New York: John Wiley & Sons, Inc., 1998.

[174] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, Jun 2001.

[175] E. Konukoglu, B. Glocker, D. Zikic, and A. Criminisi, "Neighbourhood approximation using randomized forests," *Medical Image Analysis*, vol. 17, pp. 790 – 804, Oct 2013. Special Issue on the 2012 Conference on Medical Image Computing and Computer Assisted Intervention.

[176] A. T. Papageorghiou, E. O. Ohuma, D. G. Altman, T. Todros, L. C. Ismail, A. Lambert, Y. A. Jaffer, E. Bertino, M. G. Gravett, M. Purwar, J. A. Noble, R. Pang, C. G. Victora, F. C. Barros, M. Carvalho, L. J. Salomon, Z. A. Bhutta, S. H. Kennedy, and J. Villar, "International standards for fetal growth based on serial ultrasound measurements: The fetal growth longitudinal study of the INTERGROWTH-21st Project," *The Lancet*, vol. 384, pp. 869–879, Sept 2014.

[177] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision (IJCV)*, vol. 1, pp. 321–331, Jan 1988.

[178] F. Orderud, J. Hansgård, and S. I. Rabben, "Real-time tracking of the left ventricle in 3D echocardiography using a state estimation approach," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (N. Ayache, S. Ourselin, and A. Maeder, eds.), vol. 4791 of *Lecture Notes in Computer Science*, pp. 858–865, Springer, 2007.

[179] T. J. Cashman and A. W. Fitzgibbon, "What shape are dolphins? Building 3D morphable models from 2D images," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 35, pp. 232–244, Jan 2013.

[180] M. Marinov and L. Kobbelt, "Optimization techniques for approximation with subdivision surfaces," in *Symposium on Solid Modeling and Applications*, pp. 113–122, ACM, 2004.

[181] L. A. Piegl and W. Tiller, *The NURBS Book*. Springer Verlag, 1997.

[182] D. Doo and M. Sabin, "Behaviour of recursive division surfaces near extraordinary points," *Computer-Aided Design*, vol. 10, pp. 356–360, Nov 1978.

[183] P. Kovesi, *Invariant measures of image features from phase information.* PhD thesis, Department of Psychology, University of Western Australia, 1996.

[184] V. Kolmogorov and C. Rother, "Minimizing nonsubmodular functions with graph cuts-a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 29, pp. 1274–1279, Jul 2007.

[185] P. E. Gill, W. Murray, and M. H. Wright, *Practical optimization*. Academic press, 1981.

[186] I. Sarris, C. Ioannou, P. Chamberlain, E. Ohuma, F. Roseman, L. Hoch, D. G. Altman, A. T. Papageorghiou, and for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st), "Intra- and interobserver variability in fetal ultrasound measurements," *Ultrasound in Obstetrics & Gynecology*, vol. 39, pp. 266–273, Mar 2012.

[187] C. H. Chang, F. M. Chang, C. H. Yu, H. C. Ko, and H. Y. Chen, "Assessment of fetal cerebellar volume using three-dimensional ultrasound," *Ultrasound in Medicine & Biology*, vol. 26, pp. 981–988, Jul 2000.

[188] A. I. L. Namburete, R. V. Stebbing, and J. A. Noble, "Diagnostic plane extraction from 3D parametric surface of the fetal cranium," in *Medical Image Understanding and Analysis (MIUA)*, pp. 27–32, Jul 2014.

[189] S. M. Omohundro, "Five balltree construction algorithms," tech. rep., International Computer Science Institute, 1989.

[190] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, Jan 2001.

[191] M. A. Malas and O. Sulak, "Measurements of anterior fontanelle during the fetal period," *Journal of Obstetrics & Gynaecology*, vol. 20, pp. 601–605, Nov 2000.

[192] J. Ashburner and K. J. Friston, "Voxel-based morphometry–the methods," *NeuroImage*, vol. 11, pp. 805–821, Jun 2000.

[193] B. Fischl and A. M. Dale, "Measuring the thickness of the human cerebral cortex from magnetic resonance images," *Proceedings of the National Academy of Sciences of the USA (PNAS)*, vol. 97, pp. 11050–11055, Sep 2000.

[194] World Health Organization (WHO), "The World Health Report: Reducing risks, promoting healthy life," *The World Health Report*, pp. 1–230, 2002.

[195] J. Kingdom, B. Huppertz, G. Seaward, and P. Kaufmann, "Development of the placental villous tree and its consequences for fetal growth," *European Journal of Obstetrics and Gynecology and Reproductive Biology*, vol. 92, pp. 35–43, Sept 2000.

[196] S. Rees and T. Inder, "Fetal and neonatal origins of altered brain development," *Early Human Development*, vol. 81, pp. 753–761, Sept 2005.

[197] S. Chauhan and E. Magann, "Screening for fetal growth restriction," *Clinical Obstetrics & Gynecology*, vol. 49, pp. 284–294, Jun 2006.

[198] P. J. Thureen, M. S. Anderson, and W. W. Hay, "The small-for-gestational age infant," *NeoReviews*, vol. 2, pp. e139–e149, Jun 2001.

[199] J. Villar, L. Cheikh Ismail, C. G. Victora, E. O. Ohuma, E. Bertino, D. G. Altman, A. Lambert, A. T. Papageorghiou, M. Carvalho, Y. A. Jaffer, M. G. Gravett, M. Purwar, I. O. Frederick, A. J. Noble, R. Pang, F. C. Barros, C. Chumlea, Z. A. Bhutta, S. H. Kennedy, and for the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st), "International standards for newborn weight, length, and head circumference by gestational age and sex: The newborn cross-sectional study of the INTERGROWTH-21st Project," *The Lancet*, vol. 384, pp. 857–868, Sept 2014.

[200] M. Toews, L. Zöllei, and W. M. Wells III, "Feature-based alignment of volumetric multi-modal images," in *Information Processing in Medical Imaging (IPMI)*, vol. 23 of *Lecture Notes in Computer Science*, pp. 25–36, 2013.

[201] K. K. Bhatia, P. Aljabar, J. P. Boardman, L. Srinivasan, M. Murgasova, S. J. Counsell, M. A. Rutherford, J. V. Hajnal, A. D. Edwards, and D. Rueckert, "Groupwise combined segmentation and registration for atlas construction," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (N. Ayache, S. Ourselin, and A. Maeder, eds.), vol. 4791 of *Lecture Notes in Computer Science*, pp. 532–540, Springer Berlin Heidelberg, 2007.

[202] P. Hammond, "The use of 3D face shape modelling in dysmorphology," *Archives of Disease in Childhood*, vol. 92, pp. 1120–1126, Dec 2007.

[203] Y. Heuzé, S. A. Boyadjiev, J. L. Marsh, A. A. Kane, E. Cherkez, J. E. Boggan, and J. T. Richtsmeier, "New insights into the relationship between suture closure and craniofacial dysmorphology in sagittal nonsyndromic craniosynostosis," *Journal of anatomy*, vol. 217, pp. 85–96, Jun 2010.