CrossMark

Learning-based prediction of gestational age from ultrasound images of the fetal brain

Ana I.L. Namburete [a,*], Richard V. Stebbing [a], Bryn Kemp [b], Mohammad Yaqub [a], Aris T. Papageorghiou [b], J. Alison Noble [a]

[a] *Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, United Kingdom*
[b] *Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, United Kingdom*

ABSTRACT

We propose an automated framework for predicting gestational age (GA) and neurodevelopmental maturation of a fetus based on 3D ultrasound (US) brain image appearance. Our method capitalizes on age-related sonographic image patterns in conjunction with clinical measurements to develop, for the first time, a predictive age model which improves on the GA-prediction potential of US images. 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. We develop bespoke features which 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 ($N = 447$) over an age range of active brain maturation (18–34 weeks).

On a routine clinical dataset ($N = 187$) our age prediction results 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 by ±4.57 days in the third trimester—a period complicated by biological variations in the fetal population. Through feature selection, the model successfully identified the most age-discriminating anatomies over this age range as being the Sylvian fissure, cingulate, and callosal sulci.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Accurate gestational age (GA) estimation forms an integral part of obstetric prenatal care. It defines the estimated date of delivery (EDD), and can influence the success or safety of a clinical intervention. Moreover, it is essential for the screening of fetal abnormalities. The anomaly scan, which is routinely offered in the early second trimester, forms the legal basis for time-critical care decisions as it enables expectant parents to make informed reproductive decisions about their unborn child (e.g. termination of pregnancy, intrauterine therapy or intervention) (National Collaborating Centre for Women's and Children's Health, 2008).

Traditional approaches to GA estimation include (a) menstrual dating, which makes use of the first day of the last menstrual period (LMP) as a reference point for the EDD and (b) extraction of diameter and circumference measurements from 2D ultrasound

(US) images of the fetal cranium, abdomen, and femur (ISUOG, 2007). These measurements are regressed to population-based dating charts to estimate age and assess normality of fetal growth (Loughna et al., 2009). However, beyond 24 post-menstrual weeks, measurement accuracy is dependent on operator expertise and compromised by increasing biological variation, inconsistencies in skull size approximation, and subjectivity in 2D diagnostic plane finding, all contributing to age approximation errors (Bottomley and Bourne, 2009). As pregnancy advances and biological variation amongst normal fetuses increases, the range of values of each biometric measurement associated with a specific GA also increases and so equations based upon size become less accurate. In practice, this means that whilst the predictive error at 22 weeks' GA (±10 days, Altman and Chitty (1997)) is considered acceptable in the majority of clinical settings, the predictive error at 28–42 weeks (±18 days) is considered to offer little clinical value (Hadlock et al., 1983).

Pregnancy dating becomes particularly important in low-income settings where pregnant women typically attend for

---

obstetric care late in pregnancy, when menstrual history is unavailable or unreliable. In the absence of clinically-useful LMP information, US measurements provide the most accurate estimation of GA (Geirsson, 1991). However, in the third trimester of pregnancy, even US-based dating may produce estimation errors up to ±3 weeks (Hadlock et al., 1984; Altman and Chitty, 1997). 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.

Post-mortem neuroanatomical studies have observed that during early development the fetal brain undergoes dramatic structural changes and have established a spatiotemporal timetable which characterises normal brain development (Chi et al., 1977; Dorovini-Zis and Dolman, 1977). Specifically, the fetal brain surface, or *cortex*, rapidly transitions from a relatively smooth agyric surface in the early second trimester to progressively bearing more indentations or folds (*gyrification*) over the course of pregnancy until it resembles the adult brain at birth. Deviations from this pattern have been indicative of cortical malformations as a result of defective neuronal migration, as is the case of lissencephaly which occurs when gyrification is reduced or stunted. Depending on severity, cortical malformation may result in adverse outcomes ranging from developmental delays and retardation to infant mortality (Ghai et al., 2006). This, in turn, is suggestive of a direct link between healthy gyrification and chronological age. These findings raise the question whether changes in brain morphology could be used as a robust indicator of GA and developmental normality in clinical practice.

1.1. Related work

To date, several methods have been developed to automatically map neuroanatomical structure from MR image data to neonatal or adult age. Using voxel-based morphometry or shape analysis to capture tissue growth (Good et al., 2001; Gholipour et al., 2012), tensor analysis to characterize regional growth patterns (Thompson et al., 2000), or discriminative classifiers to capture characteristics of the developing or ageing brain (Franke et al., 2012; Sabuncu and Van Leemput, 2012; Toews et al., 2012), a clear link between anatomical changes and cerebral progression (or regression) has been demonstrated. With the advent of image preprocessing methods such as slice-to-volume reconstruction and image mosaicing (Jiang et al., 2007; Rousseau et al., 2006), and super-resolution techniques (Gholipour et al., 2010; Kim et al., 2010; Rousseau et al., 2010), 3D MR images with high signal-to-noise ratio and improved spatial resolution are now available and have stimulated studies of fetal (Caldairou et al., 2011; Gholipour et al., 2012; Habas et al., 2010; Habas et al., 2012; Jacob et al., 2011; Rajagopalan et al., 2011; Scott et al., 2011; Scott et al., 2013; Serag et al., 2012; Dittrich et al., 2014; Wright et al., 2014) and neonatal (Kuklisova-Murgasova et al., 2011; Serag et al., 2012) brain development from MR images. However, these techniques are tailored for the challenges affecting MR images and may not be appropriate for application in neurosonography, which continues to be the modality of choice in routine clinical care.

In the clinical literature, the age-related changes in echogeneity of fetal brain structures have been well-described. The timing of emergence of cortical sulci has been observed in US images and described as following a predefined spatiotemporal timetable (Bernard et al., 1988; Monteagudo and Timor-Tritsch, 1997; Toi et al., 2004; Cohen-Sacher et al., 2006; Pistorius et al., 2010), in agreement with MR and post-mortem neuroanatomical findings. In particular, the process of cortical maturation observable in US images of the fetal brain has been detailed by means of simple subjective scoring techniques to define the appearance of sulci and gyri beyond 20 gestational weeks (GW) (Quarello et al., 2008; Pistorius et al., 2010).

Unlike MR images, US images are complicated by intensity artefacts such as signal attenuation, acoustic shadows, and occlusion due to cranial calcification. US probe placement also generates reverberation caused by multiple reflections of the US beam on the fetal skull and other maternal tissues. These factors can affect the visibility of key anatomical landmarks necessary for image registration—the primordial step in image-based brain analysis. However, given that cranial calcification and fusion progress with GA (Malas and Sulak, 2000), the complex image patterns generated by these artefacts may be used along with structural image features to inform on developmental maturation.

Our work is the first to exploit age-related sonographic activity to predict GA and hence neurodevelopmental maturation from US images. We present bespoke appearance-based features designed to capture these age-specific sonographic patterns and use them to develop a model which automatically maps them to GA and hence neurodevelopmental maturation. Learning-based approaches are well-suited for this task due to their ability to take high-dimensional data (i.e. longitudinal images producing 5000+ features representing image appearance at different ages) and establish a compact representation of fetal brain development. In the literature, relevance vector machines (RVM) (Franke et al., 2010) and relevance voxel machines (RVoxM) (Sabuncu and Van Leemput, 2012) have demonstrated the feasibility of learning a mapping between image-based biomarkers and pathologies in adult brains. More recently, Konukoglu et al. (2013) applied neighbourhood approximation forests (NAF) to estimate adult age. While the work of Konukoglu et al. (2013) 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, which results in acoustic shadows, occluded anatomical features, and reverberation artifacts (Kuklisova-Murgasova et al., 2013). 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. Toews et al., 2010).

The advantages of employing decision 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. Thus, decision forests are appropriate for our work in which we seek to identify the structures which are informative for GA decision-making, and aim to apply the model to images from patients at different developmental stages.

1.2. The proposed method

We propose a feature-based model for characterizing 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} GW (weeks^{+days}). Specifically, we present an automated machine learning-based predictive model to learn the pattern of fetal brain changes through dynamic features observable in multiple subject images and apply it to demonstrate successful age estimation from a single unseen scan. Our proposed model comprises of two steps: (i) 3D parametrization of the fetal skull and (ii) feature extraction for learning age-related sonographic patterns from 3D volumes, resulting in the development of an age-predictive model. The model can then be

applied to a new 3D US volume to automatically predict the GA of the fetus from which the volumetric image was collected. Fig. 1 presents an overview of the GA estimation framework.

To tackle the challenges posed by fetal neurosonography, we proposed the use of a 3D continuous B-spline surface to model and parametrize the fetal skull (Namburete et al., 2013). Cranial 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 (Namburete et al., 2014b). The lack of direct registration encourages preservation of age-specific sonographic patterns and image differences, which may be indicative of GA. The methodology for cranial parametrization was first described and validated in our previous conference papers (Namburete et al., 2013; Namburete et al., 2014a), and Section 2.1 describes the process in detail.

It has been demonstrated that the parametric surface is suitable for feature extraction in a supervised age-predictive framework based on regression forests to predict the GA of healthy and routine clinical fetal populations (i.e. containing growth-restricted fetuses) in the second trimester (18^{+0} to 28^{+6} GW) (Namburete et al., 2014b). That approach used age-related changes in US image intensity, distance of near-neighbour points on the cranial surface, and inner head circumference features for the age regression and prediction. We demonstrated that our model successfully estimated GA on a set of 130 healthy fetuses (error of ±3.8 days), outperforming the current best clinical method. Through the feature selection process, the model also identified the Sylvian fissure as a critical age-discriminating region during this age range.

In this paper, we further extend the age range of the model to function in the second and, importantly, third trimesters (i.e. 18^{+0} to 33^{+6} GW) to make the method clinically useful. We describe, in detail, the features used in the regression forest model for age prediction (§2.3), and through further validation experiments, we demonstrate the ability of the model to accurately predict GA from 3D US images (§4). We also provide new insight into the functionality and behaviour of the age-predictive model, and examine the effect of feature inclusion on prediction accuracy. Finally, we show, for the first time, that the proposed age estimation framework is capable of identifying age-discriminating regions in the fetal brain, which has never before been demonstrated on clinical US images.

2. Methods

2.1. Cranial parametrization

This section describes the 3D cranial parametrization method in preparation for US image analysis of the fetal brain. Throughout, matrices are denoted by boldface uppercase letters ($\mathbf{X}$), vectors by boldface lowercase letters ($\mathbf{x}$), and column-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.

Cranial parametrization is performed by a semi-automated framework that 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 the anatomically annotated surface is manually
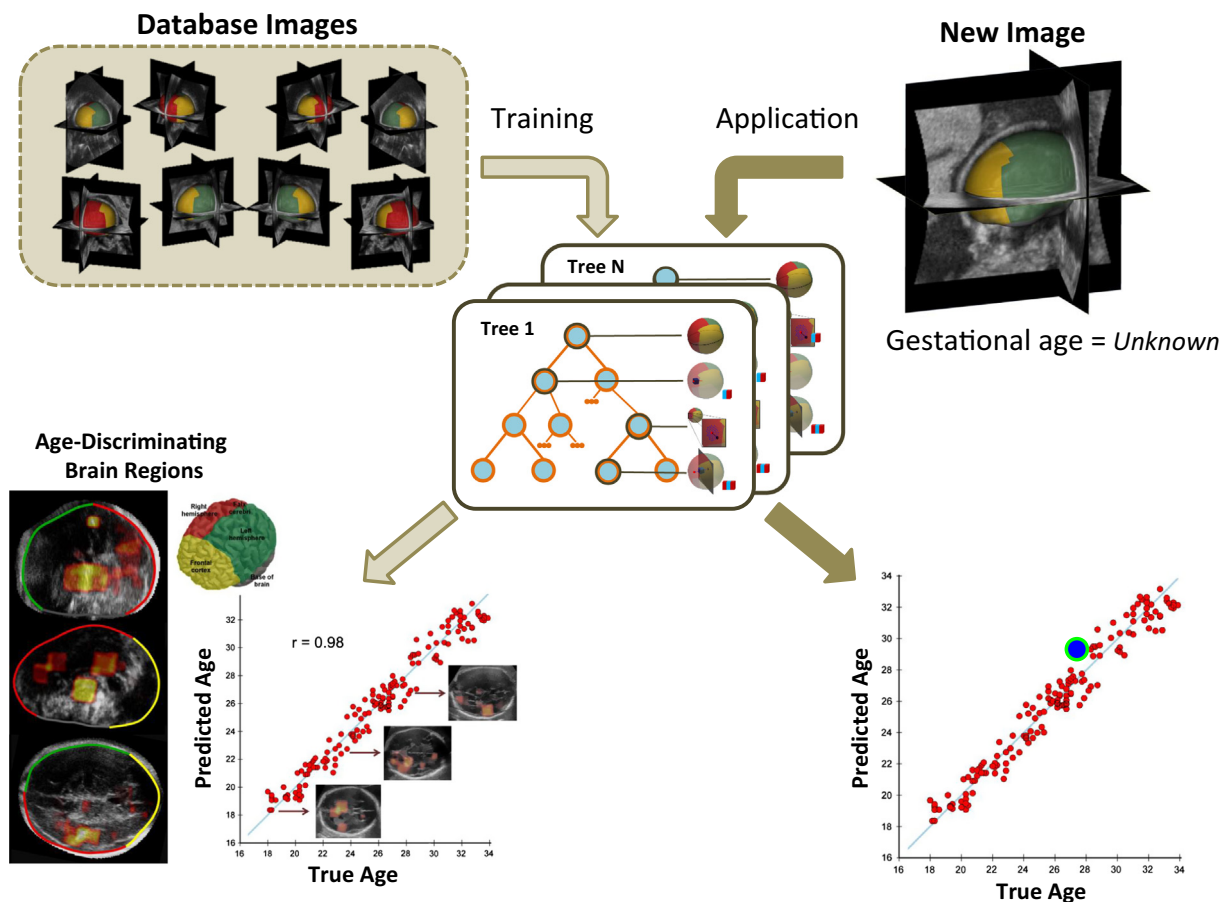

Fig. 1. Schematic flow chart of probabilistic age estimation model. Given a set of images in the database, the machine learning model will be trained to learn age-specific ultrasound image appearance and model a mapping between sonographic activity and age. When given a new test image taken from a fetus of unknown gestational age, the model can be applied to estimate age and possibly determine clinical outcomes.
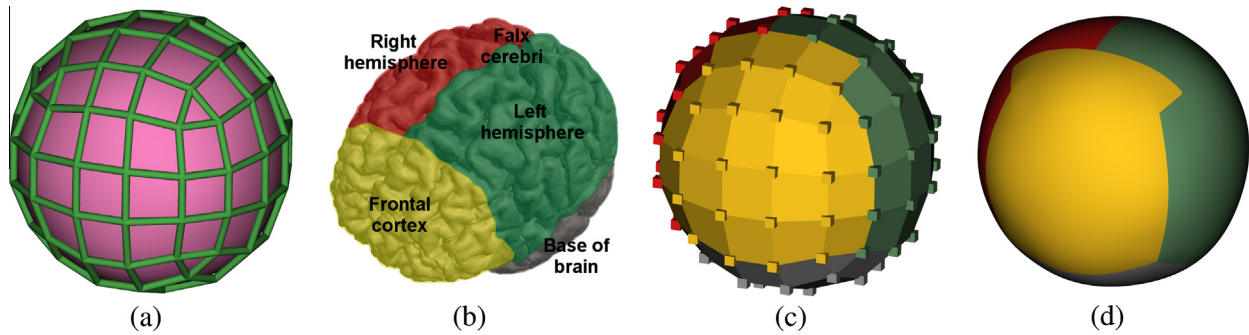
Fig. 2. Surface model: (a) The control mesh and control vertices (**X**, green) defines its underlying surface (pink). (b) Given four coarse anatomical regions, (c) the control vertices and faces are colour-annotated with the anatomical regions with which they ought to align in the image during surface initialization. (d) These annotations are propagated down to the underlying cranial surface. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 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 inner skull 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.

The remainder of this section describes the basic cranial parameterization framework which is further detailed in Namburete et al. (2013) and validated in Namburete et al. (2014a). The surface fitting framework is inspired by "Snakes" (Kass et al., 1988) and Doo-Sabin surface fitting (Orderud et al., 2007) 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 Orderud et al. (2007), 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 Orderud et al. (2007) the coordinates are fixed. The joint optimization is inspired by the works of Cashman and Fitzgibbon (2013) and Marinov and Kobbelt (2004), 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 Orderud et al. (2007) and (b) surface folding is discouraged. Third, in Orderud et al. (2007), 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. This has been done previously for *open* model contours in 2D for which the discrete optimization problem can be solved exactly using dynamic programming (Behiels et al., 2002). For a closed surface in 3D, such as the one presented here, the resulting discrete optimization problem is more complex.

2.1.1. Surface model

We generated a bespoke spherical biquadratic B-spline surface (Piegl and Tiller, 1997) which can be aligned to anatomical regions within the brain. The control mesh $\mathcal{T}$ explicitly represents the underlying cranial surface (Fig. 2(a), in pink), and its geometry is specified by $N_{\mathbf{X}}$ control vertices, $\mathbf{X} \in \mathbb{R}^{3 \times N_{\mathbf{X}}}$ (Fig. 2(a), in green). Each point **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 $M : \Omega \times \mathbb{R}^{3 \times N_{\mathbf{X}}} \to \mathbb{R}^3$ such that $\mathbf{p} = M(\mathbf{u}, \mathbf{X})$. Similarly, the surface normal at **u** is defined by $\mathbf{n} = M_\phi(\mathbf{u}, \mathbf{X})$, with mapping function $M_\phi : \Omega \times \mathbb{R}^{3 \times N_{\mathbf{X}}} \to \mathbb{R}^3$.

2.1.2. Surface Initialization

To initialize the surface control vertices **X** to $\mathbf{X}^0$, the user rigidly aligns the default spherical surface to the imaged skull/brain using a multi-view graphical user interface (GUI) programmed using VTK.[1] As shown in Fig. 3(d), initialization involves displacing the surface to roughly align with the centre of the brain, rotating and anisotropically scaling the surface such that it approximates the cranial dimensions and aligns anatomical regions to surface annotations. This manual interaction is a simple process which takes approximately 1–2 min per 3D image when performed by a non-clinical expert (Namburete et al., 2013).

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 (Fig. 2(b) and (c)). The colour annotations were based on a coarse parcellation of the brain, including a minimal number of 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) (Fig. 2(b)). Rigid image alignment using these global landmarks provides information about approximate brain size, fetal head pose and orientation (e.g. cephalic or breech). The colour annotations on the control faces of $\mathcal{T}$ are propagated down to define colouring of the underlying surface (Fig. 2(d)).

2.1.3. Surface deformation

Determining the ideal deformation for the surface model to fit to the imaged cranial boundary first requires selection of relevant boundary candidates in the US image. For the purposes of brain analysis, the ideal voxel candidates lie in the interior skull boundary which we generated using a standard US edge detection technique: Feature Asymmetry (FA) (Kovesi, 1996). 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 which behaves like step-edges or ridge-like structures, so an isotropic log-Gabor filter is appropriate (Boukerroui et al., 2004). Non-maximum suppression was then applied to thin the edges in the FA image.

---

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

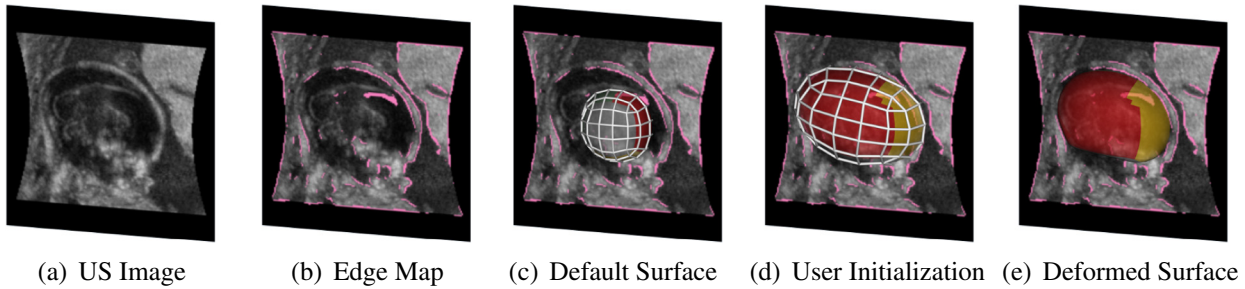(a) US Image   (b) Edge Map   (c) Default Surface   (d) User Initialization   (e) Deformed Surface

Fig. 3. 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 **X**, (d) user-initialized surface $\mathbf{X}^0$, and (e) 3D rendering of the final deformed surface superimposed on the US image.

Given an original US image (Fig. 3(a)), candidate interior skull positions $\mathbf{C} \in \mathbb{R}^{3 \times N_c}$ and normals $\Phi \in \mathbb{R}^{3 \times N_c}$ are defined from the corresponding FA edge image (Fig. 3(b)). Thus, given control vertices **X**, and a matrix **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:

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

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.

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

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 structure boundary points are spatially correlated, so the E_{pairwise} term models the fact that neighbouring surface points should also prefer boundary points which are spatially coherent.

$$E_{\text{pairwise}}(\mathbf{q}) = \sum_{(i,j) \in \mathcal{N}} \left\| \mathbf{c}_{q_i} - \mathbf{c}_{q_j} \right\|^2 \tag{3}$$

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.

$$E_{\text{user}}(\mathbf{X}) = \sum_{i=1}^{N_{\mathbf{X}}} \left\| \mathbf{x}_i - \mathbf{x}_i^0 \right\|^2 \tag{4}$$

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

$$E_{\text{reg}}(\mathbf{X}) = \sum_{(i,j) \in \mathcal{T}} \left\| \mathbf{x}_i - \mathbf{x}_j \right\|^2 \tag{5}$$

While Eqs. (3) and (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.

Eq. (1) is minimized through a series of alternating discrete and continuous optimization steps. By taking the user initialization of **X** to $\mathbf{X}^0$ and generating **U** to a regular sampling of Ω, we use belief propagation with a subset of edges in E_{pairwise} to solve an approximate **q** which is then refined using Quadratic Pseudo-Boolean Optimisation (QPBO) (Kolmogorov and Rother, 2007). Next, given **q**, Eq. (1) is minimized *jointly* with respect to **X** and **U** using the Levenberg–Marquardt algorithm (Gill et al., 1981).

Importantly, as a result of the convexity of the surface, E_{reg}, and due to the *joint* optimisation of **X** and **U**[2], 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).

2.2. Extracting features from the cranial domain

For the task of predicting GA, and hence maturation, from brain images, the features were designed with the objective of extracting morphological changes guided by the findings of post-mortem neuroanatomical studies of early brain development (Chi et al., 1977; Dorovini-Zis and Dolman, 1977). 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 (Pistorius et al., 2010; Toi et al., 2004; Wright et al., 2014); (b) an increase in overall brain size and volume (Chang et al., 2000; Zhang et al., 2013); and (c) changes in brain surface curvature in different cortical regions (Aljabar et al., 2011; Clouchoux et al., 2012; Habas et al., 2012). We designed three 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 (Good et al., 2001; Thompson et al., 2000; Franke et al., 2012; Sabuncu and Van Leemput, 2012) or atlas creation (Dittrich et al., 2014), 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 the image representation to discover distinctive anatomical patterns related to ageing, it may be likened to Feature-Based Morphometry (FBM) (Toews et al., 2010) 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 common coordinate space, voiding the need for a global brain atlas. The features are thus sampled with respect to the parametrized surface

---

[2] Standard "Snakes" fixes **U**.

model, ensuring inter-subject anatomical consistency in the image regions sampled for feature extraction (Namburete et al., 2014a).

2.2.1. 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 thickening, 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 (Fig. 4(a)). For simplicity, the surface is divided by the midsagittal plane $\mathbf{Y}^m \in \mathbb{R}^{3 \times 2}$ and the sparse subset of points $\mathbf{P}^h$ is a matrix of $N_{\mathbf{P}^h} < N_{\mathbf{U}}$ (Fig. 4(a)).

Appearance-based features. This bank of features extracts information from the greyscale US image voxels in order to capture age-related sonographic patterns of an anatomical region. Given the points on the deformed surface $\mathbf{P}^h$ and the 3D US image, a cuboidal volume-of-interest (VOI) is sampled based on the cranial points available on the selected hemispheric subsurface (Fig. 4(a)). 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 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 $l_s \sim \mathcal{U}(0, 0.5)$ is randomly selected during feature evaluation. Scaling allows for characterization of local anatomy independent of cranial surface size such that like-for-like anatomical comparisons can be achieved. The exclusion of age-related brain/cranial growth factors allowed by relative VOI sizing is paramount to retaining the 'pure appearance characterisation' 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 Fig. 4(g). 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 Fig. 4(g)). The cuboids are of arbitrary dimensions and aspect ratio, and they are 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 (Fig. 4(c)). 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. When evaluating sulcal features, the VOI is oriented along the normal vector $\mathbf{n}$ at cranial point $\mathbf{p}$ (Fig. 4(f)).

Intracranial features, on the other hand, are evaluated by displacing the VOI along the vector normal to $\mathbf{Y}^m$ (Fig. 4(b)). The VOI displacement $d_{VOI} = r \|\mathbf{p} - \mathbf{p}'\|$ is determined in proportion to the distance between $\mathbf{p}$ and $\mathbf{p}'$, where $r \sim \mathcal{U}(0, 1)$ allowing 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 brain space. The value of r is also randomly selected during feature evaluation. This relative VOI positioning encourages correspondence in the sampling of intracranial anatomical regions regardless of brain size.
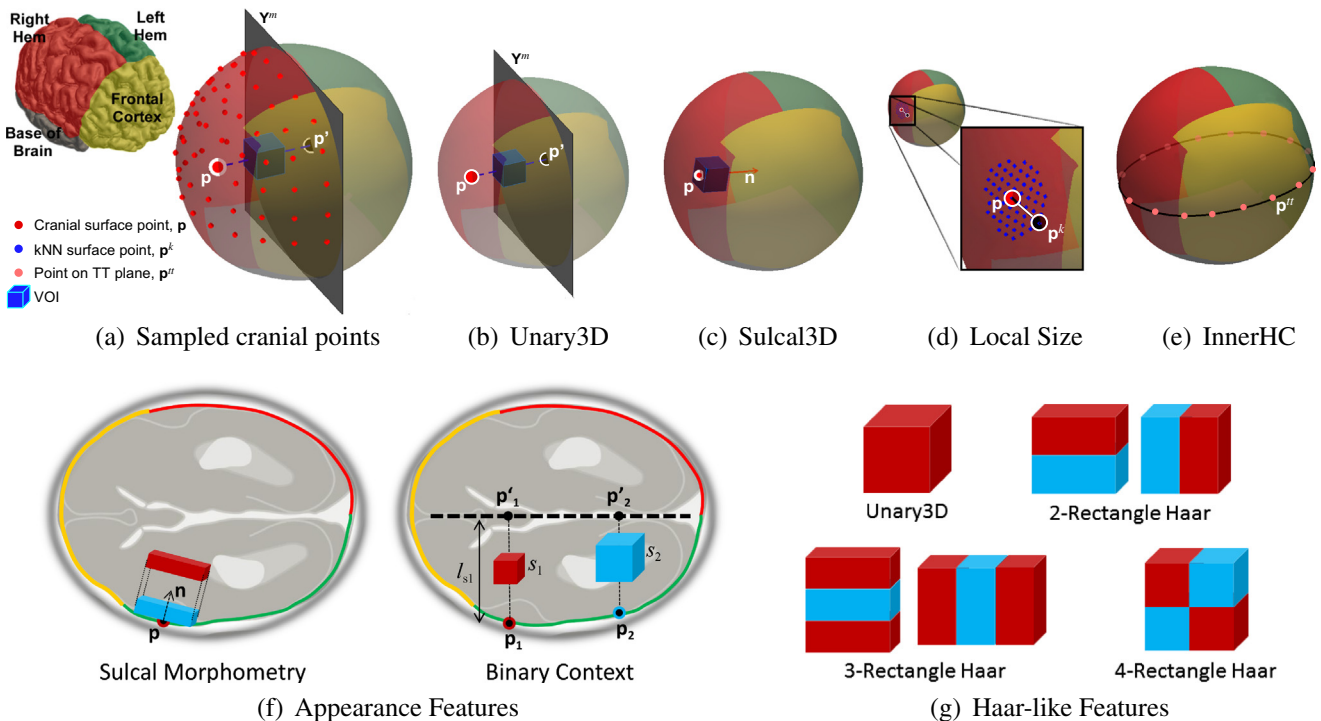

(a) Sampled cranial points (b) Unary3D (c) Sulcal3D (d) Local Size (e) InnerHC

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

Fig. 4. 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 $\mathbf{p}$ 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 cranial subsurface of interest. Using a Ball tree search algorithm (Omohundro, 1989), k cranial points nearest to $\mathbf{p}$ are identified ($k = 9$). During feature evaluation, one of the k-th nearest neighbours, $\mathbf{p}^k$, is randomly selected and the Euclidean or orthogonal distance between the sampled cranial point (Fig. 4(d), in red) and its randomly-selected k-th nearest neighbour (Fig. 4(d), in black) is obtained in Euclidean space $\mathbb{R}^3$:

$$d_{Euclid} = \|\mathbf{p} - \mathbf{p}^k\| \tag{6}$$

and the orthogonal size features are the x, y, z-component distances given by:

$$d_\perp^j = \left| p_j - p_j^k \right|, \quad \text{where } j = \{1, 2, 3\} \tag{7}$$

Biometric features. Guided by current clinical 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 (ISUOG, 2007). 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, $\mathbf{Z}^{tt} \in \mathbb{R}^{3\times 3}$. The surface points defining the TT plane, $\mathbf{P}^{tt}$, are the surface points closest to $\mathbf{Z}^{tt}$.

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 computing the closed path length of all the surface edges which the plane intersects. The biometric HC feature captures global changes in head size in a manner similar to the current clinical method of GA estimation (Namburete et al., 2014a), emulating rigid cranial transformations related to fetal growth.

2.3. 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 (Breiman, 2001; Criminisi et al., 2012) 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 Section 2.3.1. The process of predictive age regression is explained in Section 2.3.2.

2.3.1. Learning the link between age and brain images

The model is trained on a supervised 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 a 3D US 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 $\mathcal{V} = \{V_i\}$.

We train our trees following the regression forest framework proposed by Criminisi et al. (2012). During training, a random subset of the dataset $\mathcal{V}^t \subset \mathcal{V}$ is traversed through 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 is recursively partitioned at each node j in its path along the tree branches until reaching leaf nodes. At each node, the data is split to the left and right child nodes, $\mathcal{V}_{j\in\{L,R\}}^t$, 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 defined number of training data points and (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 §2.2.1). The feature is applied to the dataset at the node $\mathcal{V}_j^t \subset \mathcal{V}^t$ 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 $\mathcal{V}_j^t$ to identify the pair of (f_i, τ) 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, $\mathbf{a}_j \sim \mathcal{N}(\bar{\mathbf{a}}_j, \sigma_j^2)$, the cost function can be expressed as

$$I_g = \log \left| \sigma_j^2(\mathcal{V}_j^t) \right| - \sum_{i \in \{L,R\}} \omega_i \log \left| \sigma_i^2\left(\mathcal{V}_j^{t,i}\right) \right| \tag{8}$$

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 $\mathcal{V}_j^{t,i}$ and the number of examples in the parent node $\mathcal{V}_j^t$, i.e. $\omega_i = \left| \mathcal{V}_j^{t,i} \right| / \left| \mathcal{V}_j^t \right|$. Maximizing Eq. 8 is equivalent to favouring binary tests which 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:

$$p_l(a) = \mathcal{N}\left(\bar{a}_l, \sigma_l^2\right) \tag{9}$$

2.3.2. Prediction of GA

During age prediction, an unseen data point of unknown GA, $W_k = (\mathbf{I}_k, \mathbf{P}_k^h)$, traverses through 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. 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. Therefore, a single prediction a_k is generated by taking the mean of only the age estimates with associated variance less than σ_{max}^2. Specifically:

$$\mathcal{A} = \{a_l | \sigma_l^2 < \sigma_{max}^2\}, \quad a_k = \frac{1}{|\mathcal{A}|} \sum_{a_l \in \mathcal{A}} a_l \tag{10}$$

where $|\mathcal{A}|$ is the number of ages satisfying the exclusion criteria. For the experiments that follow $\sigma_{max}^2 = 1.0 \, \text{GW}^2$.

3. Dataset and implementation details

3.1. US data and preprocessing

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 two study databases: (a) INTERGROWTH-21st,[3] an optimally healthy group of women with a low risk of pregnancy complications

---
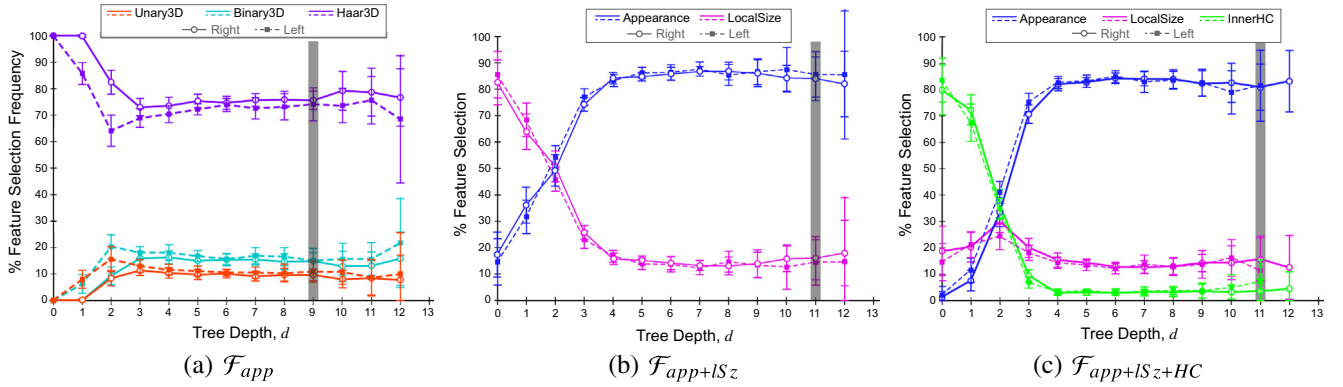
[3] http://www.intergrowth21.org.uk.

Fig. 5. Feature selection profiles. Feature selection versus tree depth for the leave-10-out cross validation forests (a) $\mathcal{F}_{app}$, (b) $\mathcal{F}_{app+lSz}$, and (c) $\mathcal{F}_{app+lSz+HC}$. Lines denote mean feature selection frequency at a given tree depth (solid: right distal hemisphere; dashed: left distal hemisphere). Error bars indicate standard deviation. Grey vertical lines indicate the tree depth of optimal GA prediction.

and fetal abnormalities (Papageorghiou et al., 2014) (Dataset A); and (b) INTERBIO-21st,[4] an unselected, routine clinical cohort (Dataset B). Throughout the text, we refer to these as Dataset A and Dataset B, respectively. 'True age' (GA_{true}) was defined by the last menstrual period (LMP) and confirmed by crown-rump-length (CRL) measurement on US images taken in the first-trimester ($\leqslant 14^{+0}$ weeks) agreeing within 7 days. For both datasets, 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) acquired from Dataset A ranging from 18^{+0} to 33^{+6} GW. Cross validation experiments were conducted on the same dataset following a leave-10-out protocol. The image inclusion criteria were:

- Cranium occupies $\geqslant 50\%$ of the image.
- Distal cerebral hemisphere is not occluded.
- Interhemispheric fissure is clearly visible in the entire supratentorial region.
- Intracranial structures (namely, Sylvian fissure, thalami, ventricles, and cavum septum pellucidum) are clearly visible.

3D US images of the fetal head were collected using a Philips HD9 curvilinear probe at a 2–5 MHz wave frequency. All images used for training and validation of the model were preprocessed by first resampling the acquired images to an isotropic spatial resolution of $0.6 \times 0.6 \times 0.6$ mm³. Ridge-like structures were then enhanced using a bandpass Gaussian derivative filter (kernel size $\bigtriangledown G_\sigma = 4$ mm). Cranial parametrization was then applied to the preprocessed data as detailed in Section 2.1. The bandpassed 3D US images and their respective deformed parametric surfaces were passed into the regression forest algorithm for training and testing.

3.2. Parameters and training

Regression forests consist of 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 $T = 20$ trees to a maximum depth of $d_{max} = 15$. At each node j, a total of $m = 200$ features was sampled, from which only one feature f_j was selected as possessing the most age-discriminating power. The age regression framework was implemented in C++ (3.30 GHz quad-core, 12 GB RAM), and (as expected) training time proved to be related to the size of the training dataset (data not shown). For each tree, training took an average of **4.6 min** and

---

7.2 min for left and right distal hemispheres, respectively, and approximately **0.1 s** to predict GA from a single volume.

3.3. Evaluation of GA predictions

To measure the accuracy of GA predictions, we used the root-mean-squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{|\mathcal{W}|} \sum_{k=1}^{|\mathcal{W}|} \left(a'_k - a_k\right)^2} \qquad (11)$$

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 GA_{true}.

4. Results

In the following, we describe 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.

4.1. Model parameter and feature selection

To gain insight into the functionality of the GA prediction model, we assessed the relative importance of each feature set presented in Section 2.2.1 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 result is shown in Fig. 5 and 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 which provide information about subtler or more detailed age-related differences. Three separate forests were trained for comparison (Table 1).

Fig. 5 demonstrates the feature selection profiles for forests (a) $\mathcal{F}_{app}$, (b) $\mathcal{F}_{app+lSz}$, and (c) $\mathcal{F}_{app+lSz+HC}$ for the leave-10-out cross-validation forests, displaying each appearance and size feature subset separately. It is evident from Fig. 5(a) that when the forest is trained exclusively with appearance-based features ($\mathcal{F}_{app}$), Haar-like features are selected more frequently than Unary or Binary 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

Table 1
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 Fig. 4 and discussed in Section 2.2.1.

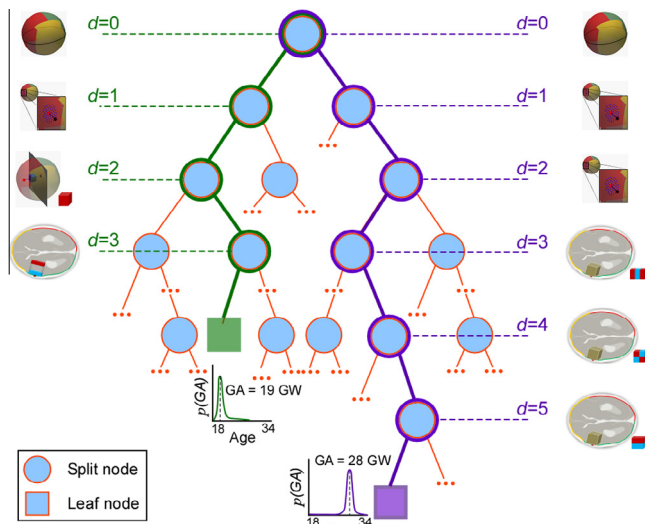| Forest type	Feature vector
$\mathcal{F}_{app}$	Appearance only
$\mathcal{F}_{app+lSz}$	Appearance and local size
$\mathcal{F}_{app+lSz+HC}$	Appearance, local size, and biometry

Fig. 6. Example feature selection. Illustration of feature selection paths for two different fetuses at 19 and 28 GW, demonstrating typical tree traversal.

which capture ridge-like image features more frequently than those capturing differences in homogeneous image regions, suggesting that sulcal and fissural development was salient in the brain during the second and third trimester. Fig. 5(b) 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 (Fig. 5(c)), 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.

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. Fig. 5 displays 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. This demonstrates the ability of the model to simply

capture anatomical progression of the brain, respecting cerebral developmental symmetry.

Fig. 6 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 Fig. 5, 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 the tree traversal path length to increase with GA (Fig. 7), which may be indicative of the longitudinal behaviour of the model as the brain increases in complexity.

4.2. Brain maturation maps

Having established the method by assessing the entire image volume of each distal hemisphere, it is expected that only some regions within the entire 3D cerebral volume would accurately report on brain maturation, while 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 (2012), Toews et al. (2010) and Konukoglu et al. (2013) 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 image voxels pertaining to the VOI selected at a node were incremented each time they were sampled. The resulting image was then normalized by the number of nodes present at level d, i.e. N_d.

Fig. 8 demonstrates the locations of age-discriminating image features selected by the model at three different levels of the forest. 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 callossal sulcus, thalami, and parieto-occipital fissure at $d = 3$; (b) the posterior and anterior ventricles at $d = 6$; and (c) sulcal areas such as the Sylvian fissure ($d = 3, 6, 9$) and central sulcus ($d = 3$). These are regions which show substantial change during gestation (Toi et al., 2004; Pistorius et al., 2010), reflecting the ability of the model to query developmentally-informative anatomical regions during the GA prediction process. We also observed that at shallow forest levels, the model queried more brain regions than at deeper levels. For instance, at $d = 3$ the model queried six key regions (i.e. callossal sulcus, thalami, parieto-occipital fissure, central sulcus, cingulate sulcus, and Sylvian fissure), while it focused primarily on the Sylvian fissure 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 these nodes would be spatially distributed in the brain. However, in practice, we found that key anatomical regions were consistently selected by different trees, identifying them as salient landmarks for GA estimation.
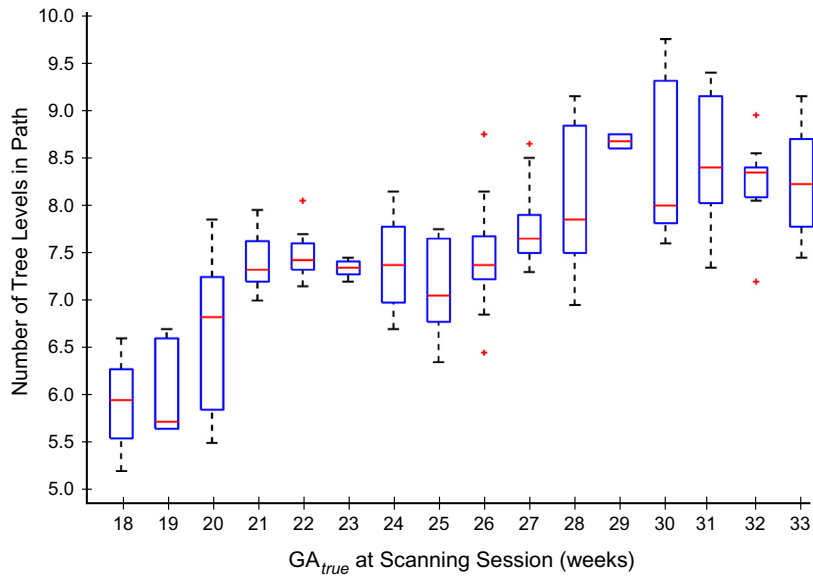
Fig. 7. Tree traversal path length versus GA. Box plot of the average number of tree levels in the decision forest path (from root to leaf) for the test examples at each GA. Median is denoted by a red line in each box, and the edges indicate the 25th and 75th percentiles. Outliers are plotted individually. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
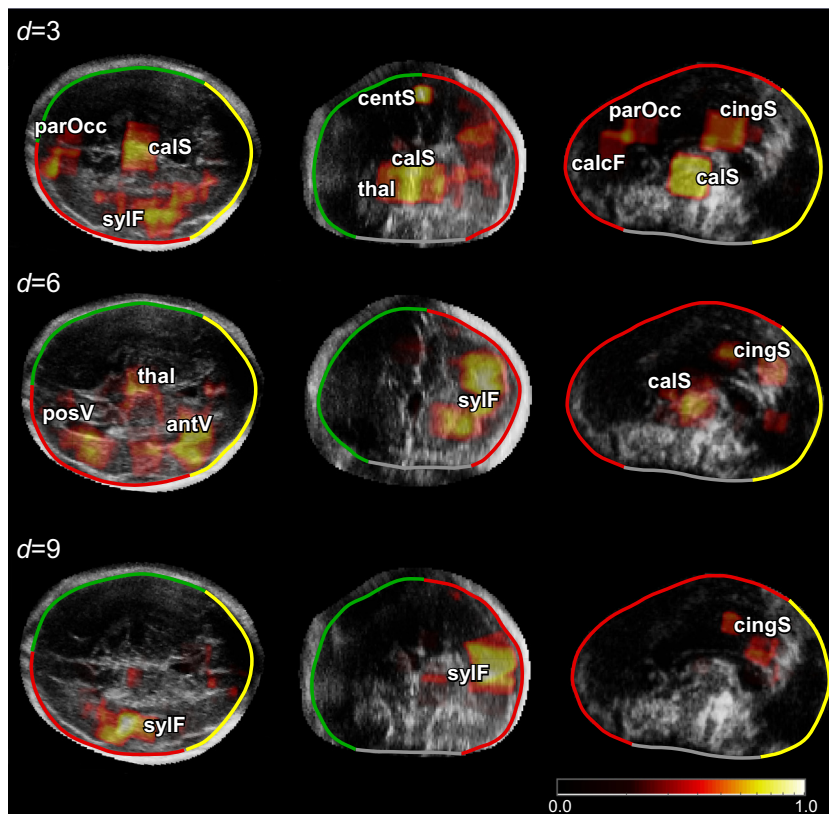

Fig. 8. GA-discriminating brain regions. Visual example of the feature locations selected at three different levels of the forest ($d = 3, 6, 9$) superimposed on axial (left column), coronal (centre column), and sagittal slices (right column). The heat map corresponds to the relative feature importance, such that bright regions correspond to frequently selected brain regions. *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.

4.3. Model-based prediction of age

To demonstrate the predictive quality of each combination of feature sets across gestation and hence identify the best predictor, we plotted the age predictions from the leave-10-out cross-valida-tion experiments on Dataset A (Fig. 9). Our results show that the forest which combines all available features ($\mathcal{F}_{app+lSz+HC}$) yields the most accurate age predictions. However, even the forest trained exclusively with anatomical appearance features (F_{app}) was capable of generating predictions with a high *r*-value
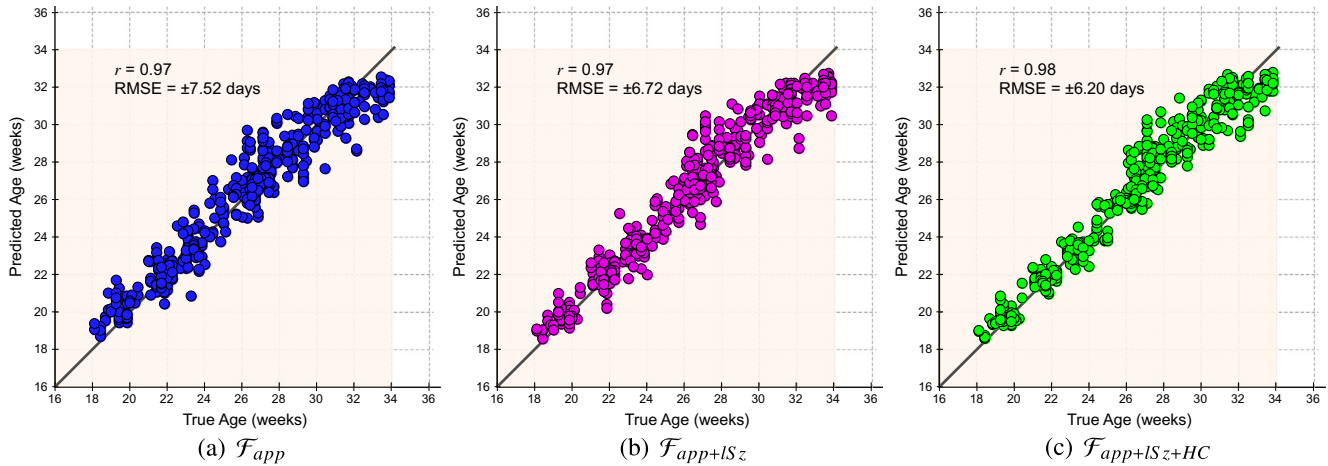
Fig. 9. Regression results from cross-validation experiments. Graphs plot true age versus predicted age for each image in the training database (Dataset A) using a leave-10-out protocol. Shown are the age predictions achieved using forests trained with different feature sets: from left to right, (a) 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).

($r = 0.97$) and low RMSE. We also plotted the curves denoting the the error margin of the GA predictions (δ_w) against GA, shown in Fig. 11. The different types of forests were applied to Dataset A 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 refer to Fig. 10.
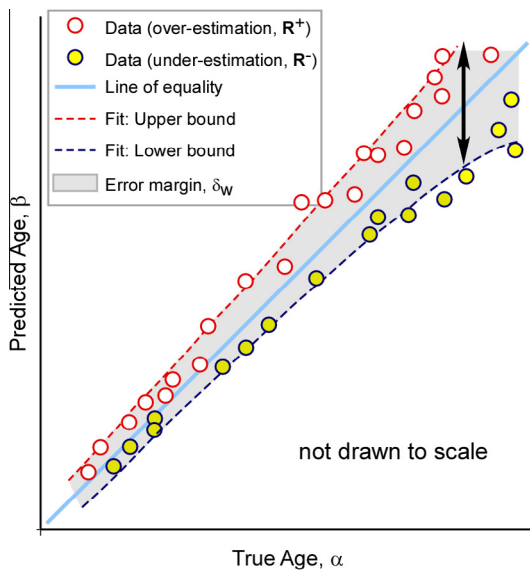

Fig. 10. Error margin of the GA predictions, δ_w. Illustration of the ω_{CI}. The positive and negative GA predictions (relative to the line of equality) are marked as red and yellow circles, respectively. Quadratic fitting functions approximating the positive (ζ^+, dashed red line) and negative (ζ^-, 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 (grey region). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

A good age prediction model would demonstrate low δ_w values, indicating a narrow error range (i.e. narrow spread). According to Fig. 11, all predictors produce the lowest errors in GA estimation during the second trimester, and error increases steadily with GA. However, the rate of increase differs for the predictors. For instance, the δ_w of both $\mathcal{F}_{app}$ and $\mathcal{F}_{app+lSz}$ increases almost linearly at a rate of $+0.226$ days/GW and $+0.215$ days/GW, respectively. The width of the clinical HC predictions shows 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 ≈ 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.

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
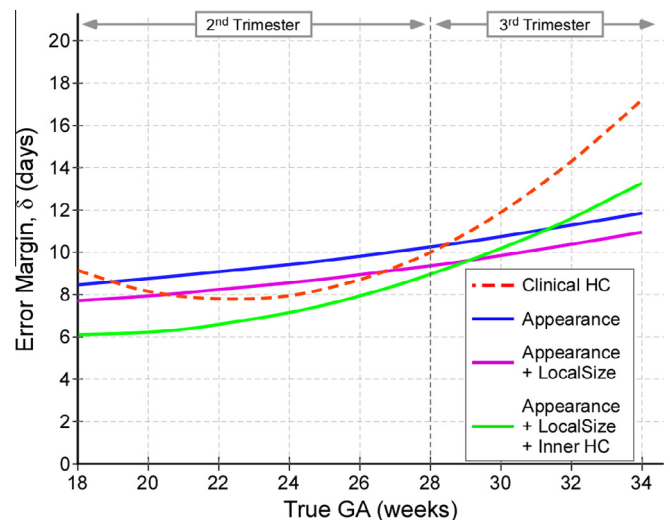

Fig. 11. Model comparison. Error margin of GA prediction, δ_w, expressed in days, for the best current clinical method (dashed red line), and for each trained forest model (solid lines) during the second and third trimesters of pregnancy. (Results shown for Dataset A.)

Table 2
GA prediction performance measures for the clinical HC method and the best model-based predictor ($\mathcal{F}_{app+ISz+HC}$) applied to Dataset B 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	4.86 (0.94)	±9.46	10.28 (0.76)	±18.57	7.01 (0.98)	±13.32
Model-based prediction	5.18 (0.97)	±10.10	7.77 (0.83)	±14.01	6.10 (0.98)	±11.64

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 demonstrates the ability of the model to outperform the best current clinical method for the entire GA range considered in this work.

4.4. Estimation and validation of age predictions

Model-based age predictions (GA_{model}) were compared against the best clinical method for estimating gestational age, $GA_{clinical}$: the mean of three HC measurements taken from independent 2D US images of the TT plane and regressed to population growth charts (Loughna et al., 2009). Table 2 summarizes the GA estimation results from applying $\mathcal{F}_{app+ISz+HC}$ and the clinical HC method to Dataset B, and Fig. 12 plots the GA predictions against true GA. As a comparison, a predictor using random guessing would result in a RMS error of ±2.76 GW ($\approx$ 19.4 days). Our results demonstrate that although $GA_{clinical}$ yields the lowest prediction errors in the second trimester (7.2 days), GA_{model} has a lower overall error throughout the entire age range, outperforming $GA_{clinical}$ by ±0.91 days and reducing the CI by ±1.68 days. However, the real benefit of GA_{model} becomes apparent in the third trimester where the CI is reduced by approximately ±**4.56** days and GA predictions are improved by ±**2.51** days. Furthermore, visual inspection of Fig. 12 indicates that the results from GA_{model} had a tighter fit to

the line of equality for the entire age range, whereas results from $GA_{clinical}$ diverged with progressing GA.

4.5. Developmental trajectories

In order to determine whether the model produces chronologically consistent age predictions between scanning session, we assessed the developmental trajectories for a subset of fetuses scanned at multiple time points in pregnancy. Fig. 13 compares the GA trajectories as predicted by separate scanning sessions for 31 subjects from Dataset B. The lines denote the time lapses between subsequent scans (marked by circles), and connected circles correspond to a single patient. It is evident that there is high agreement between the true age at a scanning session, and the model-predicted GA. This demonstrates the ability of the model to consistently approximate GA with monotonically consistent time lapses between the predictions of each scanning session. This is also indicative of the potential for using the model to extract personalized maturational progression across different time points in pregnancy.

5. Discussion

For the task of GA estimation from US scans of the fetal brain, this paper presents and validates a semi-automated learning-based framework for discovering age-related sonographic patterns in the images and linking them to neurodevelopmental maturation. Our
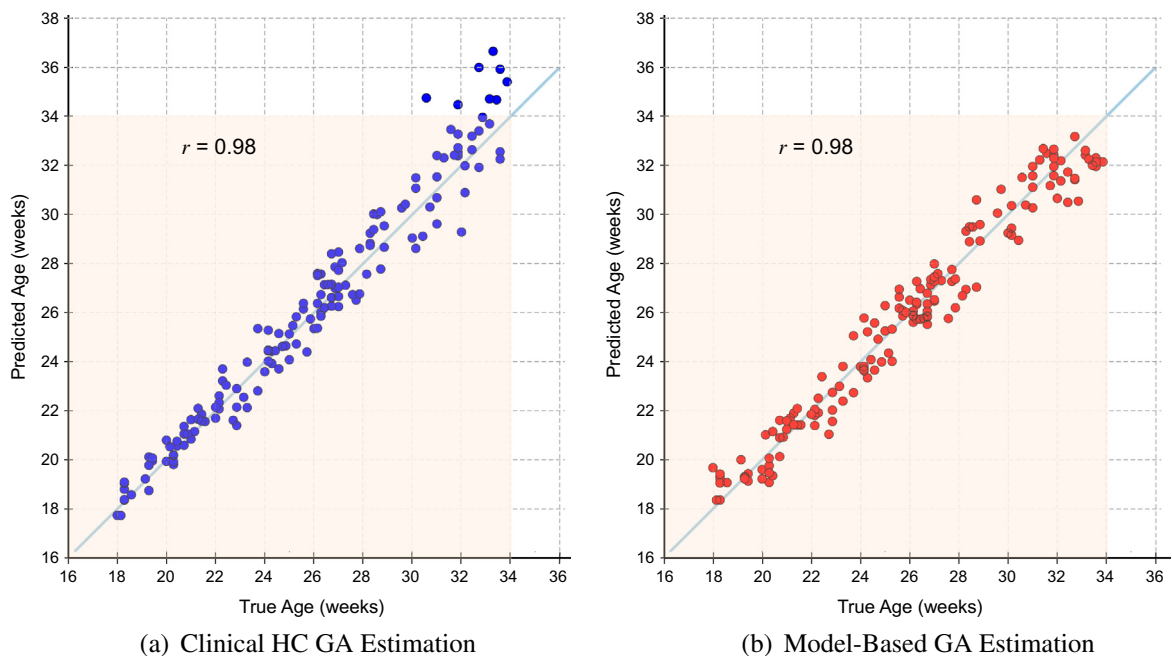

(a) Clinical HC GA Estimation

(b) Model-Based GA Estimation

Fig. 12. GA regression results. GA estimation results for a routine clinical dataset, using (a) the best current clinical method (blue), and (b) the $\mathcal{F}_{app+ISz+HC}$ forest model (red). The rectangular overlay indicates the age range of data included in this work (i.e. 18^{+0} to 33^{+6} GW). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
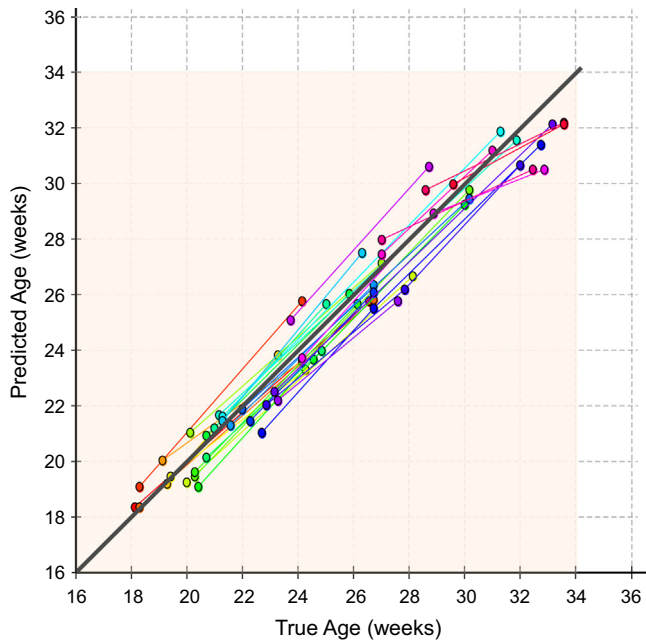
Fig. 13. Developmental trajectories. Longitudinal age predictions of images taken from fetuses at multiple separate scanning sessions. Connected line segments correspond to a single patient, and each individual scanning session is denoted by a circle.

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. 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 (Viola and Jones, 2004)) to capture structural changes within the fetal brain. Our 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 of finer age-related differences. This suggests that the method that is currently used in the clinic provides a good first prediction of GA, but 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 the 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.

Our model 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 (Toi et al., 2004; Mittal et al., 2007; Pistorius et al., 2010). Beyond the extraction of anatomical regions which undergo significant changes during gestation, our GA estimation framework has the potential to provide clinically relevant information.

Finally, our GA estimation framework has the potential to provide clinically relevant information. On dataset B, the presented 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, respectively. Specifically, our third trimester predictions within ±7.77 days are a notable improvement on the ±18 days reported in the literature (Hadlock et al., 1983). 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 the model relied more on clinical HC than on image appearance. This behaviour may be attributed to progressive skull thickening which results in increasing structural occlusions, reducing the image-based support available to the GA estimation task. This, in conjunction with the fact that decision tree traversal path lengths increased with GA, demonstrate the longitudinal behaviour of the model as it is able to respond to the fact that GA estimation is more complex in later gestation.

In conclusion, we have developed a novel feature-based model which regresses brain development to GA using US images, which has never before been attempted. Validation on a healthy dataset of 448 fetuses demonstrated the ability of the model to accurately approximate true chronological age. Application of the framework to a routine clinical fetal cohort of 187 fetuses resulted in GA prediction accuracy of ±6.1 days, particularly outperforming the current clinical method of GA estimation in the third trimester. Our preliminary analysis identified age-discriminating brain regions observable in fetal US images, namely the Sylvian fissure, callosal sulcus, and parieto-occipital fissure, all of which have been reported as undergoing significant change during the gestational period.

Acknowledgements

References

Aljabar, P., Wolz, R., Srinivasan, L., Counsell, S.J., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D., 2011. A combined manifold learning analysis of shape and appearance to characterize neonatal brain development. IEEE Trans. Med. Imaging 30, 2072–2086.

Altman, D.G., Chitty, L.S., 1997. New charts for ultrasound dating of pregnancy. Ultrasound Obstet. Gynecol. 10, 174–191.

Behiels, G., Maes, F., Vandermeulen, D., Suetens, P., 2002. Evaluation of image features and search strategies for segmentation of bone structures in radiographs using active shape models. Med. Image Anal. 6, 47–62.

Bernard, C., Droulle, P., Didier, F., GTrard, H., Larroche, J., Plenat, F., Bomsel, F., Roland, J., Hoeffel, J., 1988. Echographic aspects of cerebral sulci in the ante- and perinatal period. J. Radiol. 69, 521–532.

Bottomley, C., Bourne, T., 2009. Dating and growth in the first trimester. Best Pract. Res. Cl. Ostet. Gynecol. 23, 439–452.

Boukerroui, D., Noble, J.A., Brady, M., 2004. On the choice of band-pass quadrature filters. J. Math. Imaging Vis. 21, 53–80.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Caldairou, B., Passat, N., Habas, P., Studholme, C., Koob, M., Dietemann, J.L., Rousseau, F., 2011. Segmentation of the cortex in fetal MRI using a topological model. In: IEEE International Symposium on Biomedical Imaging (ISBI), pp. 2045–2048.

Cashman, T.J., Fitzgibbon, A.W., 2013. What shape are dolphins? Building 3D morphable models from 2D images. IEEE TPAMI 35, 232–244.

Chang, C.H., Chang, F.M., Yu, C.H., Ko, H.C., Chen, H.Y., 2000. Assessment of fetal cerebellar volume using three-dimensional ultrasound. Ultrasound Med. Biol. 26, 981–988.

Chi, J.G., Dooling, E.C., Gilles, F.H., 1977. Gyral development of the human brain. Ann. Neurol. 1, 86–93.

Clouchoux, C., Kudelski, D., Gholipour, A., Warfield, S.K., Viseur, S., Bouyssi-Kobar, M., Mari, J.L., Evans, A.C., du Plessis, A.J., Limperopoulos, C., 2012. Quantitative in vivo MRI measurement of cortical development in the fetus. Brain Struct. Funct. 217, 127–139.

Cohen-Sacher, B., Lerman-Sagie, T., Lev, D., Malinger, G., 2006. Sonographic developmental milestones of the fetal cerebral cortex: a longitudinal study. Ultrasound Obstet. Gynecol. 27, 494–502.

Criminisi, A., Shotton, J., Konukoglu, E., 2012. Decision forests: aunified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Found. Trends Comp. Graph. Vision 7 (2–3), 81–227.

Dittrich, E., Raviv, T.R., Kasprian, G., Donner, R., Brugger, P.C., Prayer, D., Langs, G., 2014. A spatio-temporal latent atlas for semi-supervised learning of fetal brain segmentations and morphological age estimation. Med. Image Anal. 18, 9–21.

Dorovini-Zis, K., Dolman, C.L., 1977. Gestational development of brain. Arch. Pathol. Lab. Med. 101, 192–195.

Franke, K., Ziegler, G., Klppel, S., Gaser, C., ADNI, 2010. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. NeuroImage 50, 883–892.

Franke, K., Luders, E., May, A., Wilke, M., Gaser, C., 2012. Brain maturation: predicting individual BrainAGE in children and adolescents using structural MRI. NeuroImage 63, 1305–1312.

Geirsson, R.T., 1991. Ultrasound instead of last menstrual period as the basis of gestational age assignment. Ultrasound Obstet. Gynecol. 1, 212–219.

Ghai, S., Fong, K.W., Toi, A., Chitayat, D., Pantazi, S., Blaser, S., 2006. Prenatal US and MR imaging findings of lissencephaly: review of fetal cerebral sulcal development. Radiographics 26, 389–405.

Gholipour, A., Estroff, J., Warfield, S., 2010. Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI. IEEE Trans. Med. Imaging 29, 1739–1758.

Gholipour, A., Akhondi-Asl, A., Estroff, J., Warfield, S., 2012. Multi-atlas multi-shape segmentation of fetal brain MRI for volumetric and morphometric analysis of ventriculomegaly. NeuroImage 60, 1819–1831.

Gill, P.E., Murray, W., Wright, M.H., 1981. Practical Optimization. Academic Press.

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

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

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

Hadlock, F.P., Deter, R.L., Harrist, R.B., Park, S.K., 1983. Computer assisted analysis of fetal age in the third trimester using multiple fetal growth parameters. J. Clin. Ultrasound 11, 313–316.

Hadlock, F.P., Deter, R.L., Harrist, R.B., Park, S.K., 1984. Estimating fetal age: computer-assisted analysis of multiple fetal growth parameters. Radiology 152, 497–501.

ISUOG, 2007. Sonographic examination of the fetal central nervous system. Ultrasound Obstet. Gynecol. 29, 109–116.

Jacob, F.D., Habas, P.A., Kim, K., Corbett-Detig, J., Xu, D., Studholme, C., Glenn, O.A., 2011. Fetal hippocampal development: analysis by magnetic resonance imaging volumetry. Pediatr. Res. 69, 425–429.

Jiang, S., Xue, H., Glover, A., Rutherford, M., Rueckert, D., Hajnal, J., 2007. MRI of moving subjects using multislice snapshot images with volume reconstruction (SVR): application to fetal, neonatal, and adult brain studies. IEEE Trans. Med. Imaging 26, 967–980.

Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: active contour models. IJCV 1, 321–331.

Kim, K., Habas, P., Rousseau, F., Glenn, O., Barkovich, A., Studholme, C., 2010. Intersection based motion correction of multislice MRI for 3-D in utero fetal brain image formation. IEEE Trans. Med. Imaging 29, 146–158.

Kolmogorov, V., Rother, C., 2007. Minimizing nonsubmodular functions with graph cuts-a review. IEEE Trans. Pattern Anal. 29, 1274–1279.

Konukoglu, E., Glocker, B., Zikic, D., Criminisi, A., 2013. Neighbourhood approximation using randomized forests. Med. Image Anal. 17, 790–804, Special Issue on the 2012 Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI).

Kovesi, P., 1996. Invariant Measures of Image Features from Phase Information, Ph.D. thesis. Department of Psychology, University of Western Australia.

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

Kuklisova-Murgasova, M., Cifor, A., Napolitano, R., Papageorghiou, A., Quaghebeur, G., Rutherford, M.A., Hajnal, J.V., Noble, J.A., Schnabel, J.A., 2013. Registration of 3d fetal neurosonography and MRI. Med. Image Anal. 17, 1137–1150.

Loughna, P., Chitty, L., Evans, T., Chudleigh, T., 2009. Fetal size and dating: charts recommended for clinical obstetric practice. Ultrasound 17, 160–166.

Malas, M.A., Sulak, O., 2000. Measurements of anterior fontanelle during the fetal period. J. Obstet. Gynaecol. 20, 601–605.

Marinov, M., Kobbelt, L., 2004. Optimization techniques for approximation with subdivision surfaces. In: Symposium on Solid Modeling and Applications. ACM, pp. 113–122.

Mittal, P., Gontalves, L.F., Kusanovic, J.P., Espinoza, J., Lee, W., Nien, J.K., Soto, E., Romero, R., 2007. Objective evaluation of Sylvian fissure development by multiplanar 3-dimensional ultrasonography. J. Ultrasound Med. 26, 347–353.

Monteagudo, A., Timor-Tritsch, I.E., 1997. Development of fetal gyri, sulci and fissures: a transvaginal sonographic study. Ultrasound Obstet. Gynecol. 9, 222–228.

Namburete, A.I.L., Stebbing, R.V., Noble, J.A., 2013. Cranial parametrization of the fetal head for 3D ultrasound image analysis. In: Medical Image Understanding and Analysis (MIUA), pp. 196–201.

Namburete, A.I.L., Stebbing, R.V., Noble, J.A., 2014a. Diagnostic plane extraction from 3D parametric surface of the fetal cranium. In: Medical Image Understanding and Analysis (MIUA), pp. 27–32.

Namburete, A.I.L., Yaqub, M., Kemp, B., Papageorghiou, A.T., Noble, J.A., 2014b. Predicting fetal neurodevelopmental age from ultrasound images. In: Medical Image Computing and Computer Assisted Intervention (MICCAI). Springer-Verlag, pp. 260–267.

National Collaborating Centre for Women's and Children's Health, 2008. Antenatal Care Routine Care for the Healthy Pregnant Woman, vol. 29. NICE/RCOG Press (pp. 109–116).

Omohundro, S.M., 1989. Five Balltree Construction Algorithms. Technical Report. International Computer Science Institute.

Orderud, F., Hansgård, J., Rabben, S.I., 2007. Real-time tracking of the left ventricle in 3d echocardiography using a state estimation approach. In: MICCAI. Springer, pp. 858–865.

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

Piegl, L.A., Tiller, W., 1997. The NURBS Book. Springer, Verlag.

Pistorius, L.R., Stoutenbeek, P., Groenendaal, F., de Vries, L., Manten, G., Mulder, E., Visser, G., 2010. Grade and symmetry of normal fetal cortical development: a longitudinal two- and three-dimensional ultrasound study. Ultrasound Obstet. Gynecol. 36, 700–708.

Quarello, E., Stirnemann, J., Ville, Y., Guibaud, L., 2008. Assessment of fetal Sylvian fissure operculization between 22 and 32 weeks: a subjective approach. Ultrasound Obstet. Gynecol. 32, 44–49.

Rajagopalan, V., Scott, J., Habas, P.A., Kim, K., Corbett-Detig, J., Rousseau, F., Barkovich, A.J., Glenn, O.A., Studholme, C., 2011. Local tissue growth patterns underlying normal fetal human brain gyrification quantified in utero. J. Neurosci. 31, 2878–2887.

Rousseau, F., Glenn, O.A., Iordanova, B., Rodriguez-Carranza, C., Vigneron, D.B., Barkovich, J.A., Studholme, C., 2006. Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images. Acad. Radiol. 13, 1072–1081.

Rousseau, F., Kim, K., Studholme, C., Koob, M., Dietemann, J.L., 2010. On super-resolution for fetal brain MRI. In: Medical Image Computing and Computer Assisted Intervention (MICCAI). Springer-Verlag, pp. 355–362.

Sabuncu, M., Van Leemput, K., 2012. The relevance voxel machine (RVoxM): a self-tuning Bayesian model for informative image-based prediction. IEEE Trans. Med. Imaging 31, 2290–2306.

Scott, J.A., Habas, P.A., Kim, K., Rajagopalan, V., Hamzelou, K.S., Corbett-Detig, J.M., Barkovich, A.J., Glenn, O.A., Studholme, C., 2011. Growth trajectories of the human fetal brain tissues estimated from 3D reconstructed in utero MRI. Int. J. Dev. Neurosci. 29, 529–536.

Scott, J.A., Habas, P.A., Rajagopalan, V., Kim, K., Barkovich, A.J., Glenn, O.A., Studholme, C., 2013. Volumetric and surface-based 3D MRI analyses of fetal isolated mild ventriculomegaly. Brain Struct. Funct. 218, 645–655.

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

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

Toews, M., Wells 3rd, W., Collins, D.L., Arbel, T., 2010. Feature-based morphometry: discovering group-related anatomical patterns. NeuroImage 49, 2318–2327.

Toews, M., Wells 3rd, W.M., Zöllei, L., 2012. A feature-based developmental model of the infant brain in structural MRI. In: Medical Image Computing and Computer Assisted Intervention (MICCAI). Springer Verlag, pp. 204–211.

Toi, A., Lister, W.S., Fong, K.W., 2004. How early are fetal cerebral sulci visible at prenatal ultrasound and what is the normal pattern of early fetal sulcal development? Ultrasound Obstet. Gynecol. 24, 706–715.

Viola, P., Jones, M., 2004. Robust real-time face detection. Int. J. Comp. Vis. 57, 137–154.

Wright, R., Kyriakopoulou, V., Ledig, C., Rutherford, M., Hajnal, J., Rueckert, D., Aljabar, P., 2014. Automatic quantification of normal cortical folding patterns from fetal brain MRI. NeuroImage 91, 21–32.

Zhang, Z., Hou, Z., Lin, X., Teng, G., Meng, H., Zang, F., Fang, F., Liu, S., 2013. Development of the fetal cerebral cortex in the second trimester: assessment with 7T postmortem MR imaging. Am. J. Neuroradiol. 34, 1462–1467.