Predicting Fetal Neurodevelopmental Age from Ultrasound Images

Ana I. L. Namburete[1], Mohammad Yaqub[1], Bryn Kemp[2], Aris T. Papageorghiou[2], and J. Alison Noble[1]

[1] Institute of Biomedical Engineering, Dept of Engineering, University of Oxford, UK
[2] Nuffield Department of Obstetrics and Gynaecology, University of Oxford, UK
ana.namburete@eng.ox.ac.uk

Abstract. We propose an automated framework for predicting age and neurodevelopmental maturation of a fetus based on 3D ultrasound (US) brain image appearance. A topology-preserving manifold representation of the fetal skull enabled design of bespoke scale-invariant image features. Our regression forest model used these features to learn a mapping from age-related sonographic image patterns to fetal age and development. The Sylvian Fissure was identified as a critical region for accurate age estimation, and restricting the search space to this anatomy improved prediction accuracy on a set of 130 healthy fetuses (error $\pm$**3.8 days**; r=0.98), outperforming the best current clinical method. Our framework remained robust when applied to a routine clinical population.

1 Introduction

Accurate gestational age (GA) estimation is essential to prenatal care. It can influence the success and safety of a clinical intervention, defines the estimated date of delivery, and forms the legal basis for a range of time-critical decisions throughout pregnancy. Traditional approaches to GA estimation include menstrual dating and extraction of diameter measurements from 2D US images of the fetal cranium [1]. These measurements are regressed to population-based dating charts to estimate age and assess normality of fetal growth. However, beyond the first trimester of pregnancy, measurement accuracy is compromised by increasing biological variation, inconsistencies in skull size approximation, and subjectivity in 2D diagnostic plane finding, all contributing to errors up to $\pm$7 days [2]. Moreover, pregnancy dating is a particular problem in low-income settings where pregnant women typically attend for obstetric care late in pregnancy, when menstrual history is unavailable or unreliable, and current methods do not provide clinically useful dating information.

Post-mortem neuroanatomical studies have established that the fetal brain follows a predictable spatiotemporal timetable of morphological changes [3]. In particular, the fetal brain surface (cortex) transitions from relatively smooth in the early second trimester to progressively bearing more indentations over the course of pregnancy until it resembles the adult brain at birth. This is suggestive of a direct link between neuroanatomical development and chronological age.

P. Golland et al. (Eds.): MICCAI 2014, Part II, LNCS 8674, pp. 260–267, 2014.

Several methods have been developed to map anatomical structure from MR image data to neonatal or adult age. Using voxel-based morphometry to capture tissue growth [4], tensor analysis to characterize regional growth patterns [5], or discriminative classifiers [6, 7, 8], a clear link between anatomical changes and cerebral progression (or regression) has been demonstrated. However, these methods generally rely on accurate segmentation and registration, which remain a challenge due to absent or inconsistent anatomical landmarks in the developing brain, regardless of imaging modality. While analysis of US images is indispensable in prenatal care, it is further complicated by partial occlusions due to increasing cranial ossification, variations in fetal position, and maternal factors, all contributing to poor image resolution.

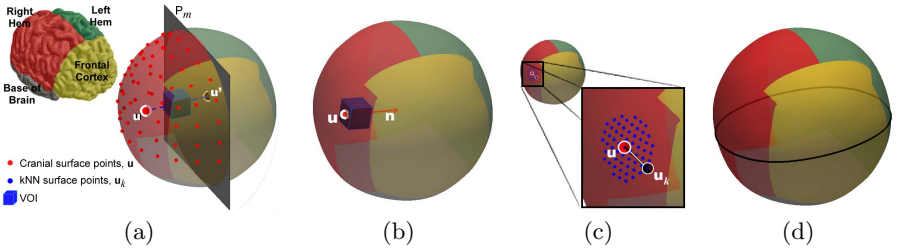
Throughout pregnancy, there are notable changes in the bright echoes in US images [9] marking the emergence/disappearance of cerebral structures. This paper is the first to exploit age-related sonographic activity to predict neurodevelopmental maturation. Specifically, we present an automated, machine learning-based predictive model to learn the pattern of fetal brain changes, and demonstrate successful age estimation. We have developed bespoke features which capture sonographic patterns in 3D images, taking into consideration the challenges posed by US image data. Our model benefits from a manifold surface representation of the skull which 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— and also serves as a skull-stripping tool [10]. We characterize 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 ranging from 18^{+0} to 27^{+6} gestational weeks (GW, $weeks^{+days}$). Cross-validation studies indicate that this model can identify the most age-discriminating anatomies over this range (i.e. the Sylvian Fissure), and achieve high age estimation accuracy within ± 3.8 days.

2 Gestational Age Estimation Framework

Our model relies on a manifold representation of the cranial surface for efficient and consistent sampling of anatomical regions in different patients. A B-spline spherical surface is manually aligned to key cerebral landmarks visible in the 3D US image (Fig. 1), a simple process which can be performed in 1-2 minutes by a non-clinical professional [10]. The surface automatically deforms to and parametrizes the cranial surface, excluding all extracerebral tissues while preserving the topology of the surface points. The resulting surface defines a coordinate system for feature extraction from anatomically-corresponding regions.

2.1 Feature Design

Our objective is to predict the age, and hence maturation, of a fetus from brain structures using a machine learning approach. We extended canonical feature



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

Fig. 1. Schematic representation of feature sets with respect to the parametrized cranial surface model, color-annotated according to key anatomical landmarks.

sets (e.g. Haar) to capture sulcal changes within the fetal brain, as well as development of intracranial structures. A bespoke bank of pure US appearance- (Figs. 1(a), 1(b)) and size-based (Figs. 1(c), 1(d)) features was created, defined with respect to a point $\mathbf{u} \in U$ on the parametric cranial surface ($U \subset \mathbb{R}^3$).

Due to cranial ossification, the brain hemisphere proximal to the US probe is typically occluded in fetal brain images, leaving only the distal hemisphere with discernible intracranial structures. As a result, it was decided to confine the search space to the hemisphere bounded by the midsagittal plane P_m and the cranial subsurface of the distal hemisphere (left or right), $\mathcal{B}^h$, where $h \in \{L, R\}$.

Appearance-Based Features. A cuboidal volume-of-interest (VOI) was sampled in 3D image space, whose location was defined with respect to a surface point $\mathbf{u}$, to describe the intensity around the corresponding anatomical region. The side length, s, of the VOI was scaled with respect to distance between cranial point $\mathbf{u}$ and its projection onto the midsagittal plane, $\mathbf{u}'$, such that $s = l_s\|\mathbf{u}-\mathbf{u}'\|$ where $l_s \in [0, 0.5]$, and was randomly selected during forest training. Scaling allowed for the characterization of local anatomy independent of brain size. The VOI was displaced along the normal vector of the midsagittal plane, as shown in Fig. 1(a), to detect changes in intracranial structures. Alternatively, the VOI was affixed to the surface (Fig. 1(b)) to capture sulcal appearance patterns.

Local Size Features. These features were obtained by computing the Euclidean, or orthogonal, distances between the cranial point $\mathbf{u}$ and its k-th nearest neighbour on the cranial surface, $d(\mathbf{u}, \mathbf{u}_k)$ (Fig. 1(c)). They represent local skull deformations and hence local growth patterns at different stages of gestation.

Biometric Features. Guided by current clinical assessment of fetal growth, the biometric feature is similar to the clinical head circumference (HC) measurement acquired from the transthalamic plane of the head [1]. In this case, we obtained the perimeter of the deformed cranial surface at the level of the transthalamic plane (defined by 3 points on the manifold), as shown in Fig. 1(d). This feature captures global changes in head size, emulating rigid cranial transformations.

3 Gestational Age Estimation with Regression Forests

Given that the goal is to discriminate subjects by age, and that age is itself a continuous variable, we opted for a regression forest classifier [11]. A regression forest is an ensemble of T trees trained on a set of training examples, $\mathbf{V}(I, S)$, each defined by an image I and its corresponding deformed cranial surface S. During training, a random subset of the data $\mathbf{v}^t \subseteq \mathbf{V}$ is traversed through each tree t in the forest $\mathcal{T} = \{T_t\}$, being recursively split at each node Q in its path by age-discriminating binary tests. These binary tests are designed to maximize the differential entropy between the data sent to the left and right children nodes, $Q_{\{L,R\}}$. Modelling the ages of dataset $\mathbf{v}_a$ as a random variable with univariate Gaussian distribution, $p(\mathbf{v}_a) = \mathcal{N}(\mathbf{v}_a; \bar{\mathbf{v}}_a, \boldsymbol{\Sigma}^a)$, the cost function is expressed as:

$$I_g = \log|\boldsymbol{\Sigma}^a(Q)| - \sum_{i\in\{L,R\}} \omega_i \log|\boldsymbol{\Sigma}^a_i(Q_i)| \tag{1}$$

Maximizing equation (1) is equivalent to minimizing the determinant of the age covariance matrix $\boldsymbol{\Sigma}^a$, ultimately reducing the uncertainty in the age-discriminating ability of a given binary test on $\mathbf{v}_a$. The discriminative quality of these binary tests hinges on the assumption that $\mathbf{v}^t$ is representative of the complete dataset, $\mathbf{V}$. A leaf node l is created when the maximum tree depth, D, is reached or if a node contains a minimum number of points $\mathbf{v}_{min}$. Each leaf node stores the Gaussian distribution of the ages of the data to have reached it, $p(\mathbf{v}_a|l) = \mathcal{N}(\mathbf{v}^l_a; \bar{\mathbf{v}}^l_a)$.

During testing, GA estimation is achieved by averaging the single age predictions from each tree in the forest $p(\mathbf{V}_i(a)|L) = \frac{1}{T}\sum_{t\in\mathcal{T}} p(\mathbf{v}_a|l_t)$. In order to reduce noise, we omit the 'uncertain' leaf nodes whose Gaussian distributions exceed an empirically-selected variance, Var_{max}, from age predictions.

4 Experiments

In order to characterize the model for all visible intracranial regions, the input data to our framework are 3D US images of the fetal brain obtained from two study databases[1]: *a*) INTERGROWTH-21st, an optimally healthy group of women with a low risk of pregnancy complications and fetal abnormalities, and *b*) INTERBIO-21st, an unselected, routine clinical cohort. 'True age' for the data included in this study was defined by the last menstrual period (LMP) and first-trimester US measurement (collected $\leq 14^{+0}$ GW, accurate within 2.7 days) agreeing within 7 days. For this work, 130 images were selected from fetuses ranging from **18^{+0} to 27^{+6} GW**, spanning the second and early third trimesters of pregnancy: an active period of spatiotemporal sulcal changes visible on the fetal brain cortex. 3D images of the fetal head were collected using a Philips HD9 curvilinear probe at a 2-5MHz wave frequency. All images were preprocessed by enhancing ridge-like structures (bandpass filtering using a Gaussian derivative filter, $\nabla G_{\sigma=4mm}$), rendered to an isotropic spatial resolution of $0.6mm$.

[1] `http://www.intergrowth21.org.uk/` and `http://www.interbio21.org.uk/`

Cross-validation experiments were conducted using a leave-10-out protocol, each time using 120 US images for training the regression forest, and the remaining 10 images for testing, all taken from the healthy database. The model was then applied to 32 fetal brain US images from the clinical population.

Parameter Selection and Implementation Details. Each regression forest comprised of $T = 20$ trees with a maximum depth of $D_{max} = 15$, although none of the trees reached this depth during training. A total of 500 features was sampled at each tree node, from which only the most age-discriminating feature was selected. For a training dataset of 120 3D images, neurodevelopmental progression in the entirety of the left cerebral hemisphere $\mathcal{B}^L$ was characterized by sampling from a total of 61 cranial surface points. The age regression framework took an average of **6.1min** to train each tree with a C++ implementation (3.30GHz quad-core, 12GB RAM), and **0.1sec** to predict GA from a test image. Three separate forests were trained for comparison:

(i) $\underline{\mathbf{f}_{app}}$ Feature vector consisting of only pure appearance features
(ii) $\underline{\mathbf{f}_{app+lSz}}$ Appearance and local size features included in feature vector
(iii) $\underline{\mathbf{f}_{app+lSz+HC}}$ Appearance, local size, and biometric features in feature vector

Results were compared against the best clinical standard for estimating fetal age: the mean of three HC measurements taken from 2D US images of the transthalamic plane and regressed to population growth charts.

5 Results

Experiment 1: Feature Selection. To provide insight into the functionality of age discrimination on the input dataset, the relative importance of each feature was assessed. To achieve this, tree depth D was plotted against the number of times a particular feature was selected at each tree depth in the forest, normalized by the total number of nodes present at the given depth (Fig. 2). From Fig. 2, it is evident that when the feature vector comprises of appearance and local size features, the latter is more age-discriminating in the shallower tree depths, but anatomical appearance is dominant from $D = 3$, where prediction with minimal root-mean square (RMS) error occurs at $D = 8$. However, when the inner HC feature is also included in the feature vector, it is preferentially selected in the shallower tree depths, superseding even local size features, before appearance features are rendered more important also at $D = 3$. This clearly demonstrates that global and local head size provide an important first estimate of age, but sonographic anatomical appearance is valuable in refining age predictions.

To demonstrate the relative likelihood of a GA prediction assuming a certain error value, Fig. 2(c) shows the probability density estimate of the signed RMS error for the best clinical HC method and for each of the trained forests. For our data, the clinical method tends to underestimate GA (data not shown), hence the negative bias in its density function. It is also evident that the forest containing all features, $\mathbf{f}_{app+lSz+HC}$, has the highest peak around the zero-RMS

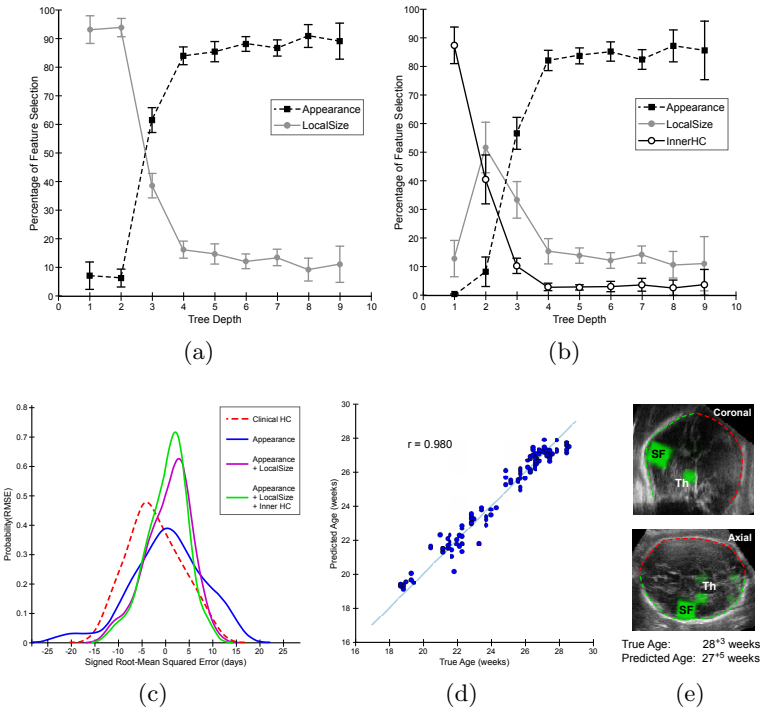


Fig. 2. Feature selection versus tree depth for the 13 leave-10-out cross-validation forests (a) $\mathbf{f}_{app+lSz}$ and (b) $\mathbf{f}_{app+lSz+HC}$. Lines denote mean feature selection frequency at a given tree depth; error bars: standard deviation. (c) The probability density distributions of the RMS age error are shown for the best standard clinical HC method and for each forest. (d) $\mathbf{f}_{app+lSz+HC}$ age predictions plotted against true GA for the healthy dataset (line of equality shown). (e) Representative example of age-discriminating regions selected by the predictive model on (i) a coronal and (ii) an axial slice through the 3D US brain volume (in green: SF=Sylvian Fissure, Th=Thalamus). Non-discriminating voxels are transparent.

error region and a narrower error range, meaning that it is the best method for estimating GA, outperforming the current best clinical method.

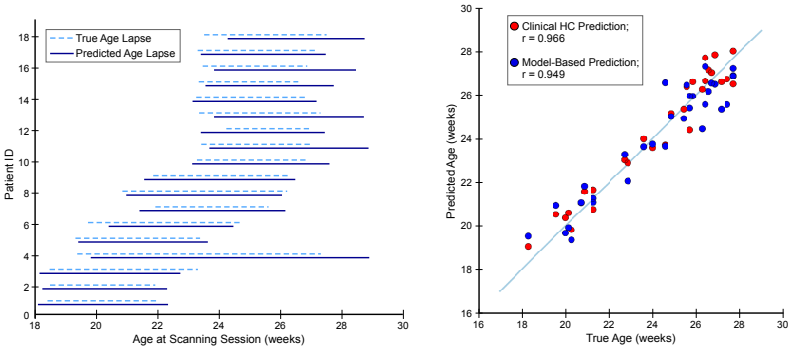
Experiment 2: Clinical Prediction of Age. To demonstrate the predictive power of the age estimation model, true GA was plotted against the model-predicted GA for the healthy dataset (Fig. 2(d)). As a comparison, a linear age predictor using random guessing would result in a RMS error of ±2.76 weeks (∼19.4 days). Figure 2(d) and Table 1 demonstrate that our model is capable of approximating the true GA with an error of ±**4.12 days**, outperforming the best clinical method by ±1.79 days for the cross-validation dataset.

Having established a method by assessing the entire volume of the left hemisphere ($\mathcal{B}^L$), we expected that only some regions within the entire 3D cerebral hemisphere 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 particular brain regions that provide the best age discrimination at different developmental stages. Further investigation showed that the algorithm consistently selected a small number of key age-discriminating regions (Fig. 2(e)). One such region is the Sylvian Fissure, which is known to follow a characteristic pattern of development [9]. Guided by the algorithm's behaviour, we next investigated if the accuracy of age determination could be further improved by restricting the search space to the Sylvian Fissure area ($\mathbf{u}_{\mathrm{sf}}$). As summarized in Table 1, confining the search space to a high confidence region yielded higher accuracy age predictions.

Table 1. RMSE of Fetal Age Prediction (in days)

Dataset	Clinical HC	Appearance Model	Appearance + Local Size	Appearance + Local Size + innerHC
Healthy (n=130)	**5.91**	$\mathcal{B}^L$ = 7.36, $\mathcal{B}^R$=5.83 $\mathbf{u}_{\mathrm{sf}}$ =7.20	$\mathcal{B}^L$ = 4.62,$\mathcal{B}^R$ =4.92 $\mathbf{u}_{\mathrm{sf}}$ = 4.59	$\mathcal{B}^L$ = 4.12,$\mathcal{B}^R$ =4.62 $\mathbf{u}_{\mathrm{sf}}$ **= 3.84**
Clinical (n=32)	5.08	–	–	$\mathcal{B}^L$ = 6.11, $\mathcal{B}^R$ = 6.66



(a) (b)

Fig. 3. (a) Developmental trajectories are shown as the time difference between predicted ages of a fetus at two separate scanning sessions. For each patient, lines demonstrate age difference between subsequent sessions. (b) True age versus predicted ages for the routine clinical dataset (line of equality shown).

Longitudinal Maturation. The developmental trajectories for a subset of fetuses from the healthy dataset (scanned at multiple time points in pregnancy) are shown on Fig. 3(a), comparing the GAs predicted at separate scanning sessions. We found that there was high agreement between the true GA at a scanning session, and the model-predicted GA, and that there were monotonically consistent time lapses between ages at subsequent sessions. This demonstrates the ability of the model to consistently approximate age of the same subject at different time points and, more notably, to extract personalized maturational progression.

Experiment 3: Predicting Maturation. The model was trained on images with visible left ($\mathcal{B}^L$) and right ($\mathcal{B}^R$) cerebral hemispheres. It was found that high

accuracy was achieved regardless of which cerebral hemisphere was observable, and hence assessed, in the US image (Fig. 3(b), Table 1).

6 Discussion and 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 demonstrates the ability of the model to accurately approximate true GA. Our preliminary analysis on a small number of cases suggests that the model may remain robust in a realistic fetal population containing abnormal growth phenotypes (e.g. growth-restricted, small-for-gestational age); formal testing of this hypothesis is planned on a larger clinical dataset. Nonetheless, this result suggests that the model does not simply estimate chronological age, but rather the neurodevelopmental stage of a fetus based on brain structure. Ongoing work focuses on extracting other anatomical regions of relevance for assessment of maturation, and determining local neurodevelopmental patterns.

References

[1] ISUOG: Sonographic examination of the fetal central nervous system. Ultrasound Obstet. Gynecol. 29(1), 109–116 (2007)
[2] Bottomley, C., Bourne, T.: Dating and growth in the first trimester. Best Pract. Res. Cl Ostet. Gynecol. 23(4), 439–452 (2009)
[3] Chi, J.G., Dooling, E.C., Gilles, F.H.: Gyral development of the human brain. Ann. Neurol. 1(1), 86–93 (1977)
[4] Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S.: A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage 14(1 pt. 1), 21–36 (2001)
[5] Thompson, P.M., Giedd, J.N., Woods, R.P., MacDonald, D., Evans, A.C., Toga, A.W.: Growth patterns in the developing brain detected by using continuum mechanical tensor maps. Nature 404(6774), 190–193 (2000)
[6] Franke, K., Luders, E., May, A., Wilke, M., Gaser, C.: Brain maturation: predicting individual BrainAGE in children and adolescents using structural MRI. NeuroImage 63(3), 1305–1312 (2012)
[7] Sabuncu, M.R., Van Leemput, K.: The relevance voxel machine (RVoxM): A self-tuning Bayesian model for informative image-based prediction. IEEE Trans Med Imaging 31(12), 2290–2306 (2012)
[8] Toews, M., William, W.3., Louis, C.D., Tal, A.: Feature-based morphometry: discovering group-related anatomical patterns. NeuroImage 49(3), 2318–2327 (2010)
[9] Toi, A., Lister, W.S., Fong, K.W.: 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(7), 706–715 (2004)
[10] Namburete, A.I.L., Stebbing, R.V., Noble, J.A.: Cranial parametrization of the fetal head for 3D ultrasound image analysis. In: MIUA, pp. 196–201 (2013)
[11] Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)