



OPEN

Fetal face shape analysis from prenatal 3D ultrasound images

Raphael Sivera^{1,4}, Anna E. Clark^{2,4}, Andrea Dall'Asta³, Tullio Ghi³, Silvia Schievano¹ & Christoph C. Lees²✉

3D ultrasound imaging of fetal faces has been predominantly confined to qualitative assessment. Many genetic conditions evade diagnosis and identification could assist with parental counselling, pregnancy management and neonatal care planning. We describe a methodology to build a shape model of the third trimester fetal face from 3D ultrasound and show how it can objectively describe morphological features and gestational-age related changes of normal fetal faces. 135 fetal face 3D ultrasound volumes (117 appropriately grown, 18 growth-restricted) of 24-34 weeks gestation were included. A 3D surface model of each face was obtained using a semi-automatic segmentation workflow. Size normalisation and rescaling was performed using a growth model giving the average size at every gestation. The model demonstrated a similar growth rate to standard head circumference reference charts. A landmark-free morphometry model was estimated to characterize shape differences using non-linear deformations of an idealized template face. Advancing gestation is associated with widening/fullness of the cheeks, contraction of the chin and deepening of the eyes. Fetal growth restriction is associated with a smaller average facial size but no morphological differences. This model may eventually be used as a reference to assist in the prenatal diagnosis of congenital anomalies with characteristic facial dysmorphisms.

Prenatal diagnosis of congenital anomalies is important for patient care and to facilitate parents' counselling, planning of delivery, and postnatal treatment¹. However, prenatal detection of fetal dysmorphisms is challenging because of the wide range of morphological features involved, the limitation of prenatal imaging², and because the phenotypical descriptions of abnormalities seen on 2D ultrasounds are not always linked to easily identifiable genetic mutations. In this context, the analysis of fetal facial morphology in 3D can provide relevant information and serve as a pre-screening tool, facilitating the early detection of developmental disorders and genetic syndromes³⁻⁷.

Ultrasound imaging is routinely performed for fetal assessment, as it is non-invasive and nowadays widely available in middle and higher income countries. Three-dimensional ultrasound (3D US) was introduced clinically in the late 1990s and studies have explored the added value of 3D US for the assessment of fetal abnormalities⁸⁻¹¹. Technical improvements have increased the quality of the images and their visualisation¹². Yet, the use of 3D images has not truly permeated diagnostic fetal medicine and the analysis of these has almost exclusively been confined to qualitative assessments, vulnerable to misinterpretation. Noisy images, possible occlusion of the face, and overall acquisition variability (contrast, fetal position, maternal body habitus, amniotic fluid volume and operator handling) make the analysis of the images difficult. In particular, automatic morphometric approaches need to address several technical challenges to be able to assist the clinical analysis of the fetal face from 3D US.

Segmentation algorithms developed for other image modalities may not be suitable for ultrasound imaging because of the specific intensity, noise and image configuration of ultrasound. Intensity models for classification between fetal tissue and amniotic fluid in prenatal ultrasounds have been designed¹³, and specific approaches that process the images from different view-points to correct for orientation and position changes have been proposed¹⁴. However, these methods are not able to fully reconstruct the fetal face when the fetus is laid on the mother's tissue. Similarly, deep learning approaches have been proposed^{15,16} but the task difficulty and the limited existing ground-truth data that could be leveraged to build and validate a new model limit the quality

¹Institute of Cardiovascular Science, University College London, London, UK. ²Institute of Reproductive and Development Biology, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. ³Department of Medicine and Surgery, University of Parma, Parma, Italy. ⁴These authors contributed equally: Raphael Sivera and Anna E. Clark. ✉email: c.lees@imperial.ac.uk

of the results, and the reconstructions are often rough or incomplete. The method proposed by Alomar et al.¹⁷ demonstrates promising results, is based on a model developed for the faces of newborn babies and requires the manual placement of a large number of landmarks. The reconstructions therefore approximate more closely to a newborn rather than fetal face shape, particularly for earlier gestational ages. Many works still rely on manual slice-by-slice segmentation¹⁸.

Morphable face models have shown their potential to help diagnose congenital abnormalities from 3D scans^{19,20}, including in newborns²¹. Prenatal diagnosis has been however more challenging. Deep-learning-based approaches have been recently developed for the recognition of facial expression associated with fetal brain development stages²² or for the prenatal screening of genetic conditions²³ from 3D images. They are however limited by the interpretability of their features and their generalizability. In 2017, our group demonstrated the potential of shape modelling of the fetal face to assist pre-natal diagnosis and help characterise fetal face morphology from 3D ultrasound in a feasibility study²⁴ with landmark-free morphometry²⁵. However the study required time-consuming manual segmentation of the fetal face which limited the scope of the study and would prevent its clinical use. As such we later developed a semi-automatic processing pipeline to segment 3D US fetal faces volumes²⁶ using an atlas-based segmentation algorithm²⁷: a set of manually labelled images, the atlas, is non-linearly registered on the image to segment, then, a probabilistic labelling map is obtained by merging a selected subset of candidate segmentations²⁸.

This study extends and builds on the methodology proposed in our previous work in two ways. First, a more efficient semi-automatic segmentation workflow to allow the 3D reconstruction of a larger number of cases with a wider spectrum of clinical conditions is presented. Second, the statistical models have been extended to provide 3D description of the normal fetal face in the third trimester including the morphological change with advancing gestational age and the normal shape variability visible in a control population. The aim is to provide a reference range against which individual clinical conditions can be compared and outliers can be identified.

Results

Data and participants

235 participants (238 fetuses) were recruited to the study. Exclusion criteria were based on the absence of suitable 3D US volumes between the 24 and 34 week of gestation, or incomplete follow-up data. We focused on patients without any diagnosed facial dysmorphism with a final data set consisting of 135 fetuses, 117 appropriate grown (AGA) normal fetuses and 18 fetal growth restricted (FGR) normal fetuses. Participants were 18 to 52 years old (mean 32.9 years old), had a body mass index between 17 and 41 kg/m² (mean 24.7 kg/m²), with fetuses of gestational ages (GA) from 24+0 weeks to 34+0 weeks. Mean gestational age is 28+4 weeks (standard deviation 2+6 weeks) for normal AGA, and 29+1 weeks (standard deviation 2+6 weeks) for normal FGR.

Patient population and image acquisition

3D US volumes were obtained in the highest resolution acquisition mode on the ultrasound machine. The volumes were collected by fetal medicine staff experienced in performing ultrasound and no additional training was required. We estimate that an additional 10–15 minutes was required for each clinical appointment to collect the 3D volumes. It was possible to collect 3D facial volumes for 84% of recruited participants (n=204), for the remaining 16% fetal position precluded 3D volume acquisition. 3D US volume was collected during the first clinical visit in 92% of cases, the remaining 8% required between 2–4 clinical appointments in total to obtain a 3D volume.

Semi-automatic segmentation results

At visual inspection, we estimated that 57% would require no or minimal manual refinements, 35% required significant corrections of part of the segmentation, and 8% would need major changes or were judged difficult due to poor visibility. This assessment translated into the time spent on the manual edits that we estimate to be between 0 and 10 min for the first group, less than 45 min for the second and up to 2 hours for the more challenging cases. A median time of around 9 min was measured; the mean time spent by segmentation is significantly higher due to the most difficult cases and was around 20 to 30 min. In addition to this time, 5 to 10 min were spent per case placing the landmarks and setting up the automatic segmentation. The degree of overlap between the automatic and edited segmentations was associated with an average Dice score of 97%. The average symmetric surface distance (ASSD) between pre- and post-manual refinement was 0.14mm (for a 0.6mm image resolution) with 44% of the cohort having an ASSD inferior to one tenth of the voxel size (<0.06mm).

The need for manual refinement was directly related to the quality of the 3D US volume, when the image captured an unencumbered view of all borders of the fetal face, no manual refinement was required. Most of the required manual refinements were associated with missing regions of the face. The regions with low contrast in the images are poorly segmented and the peripheral areas of the face more often required manual editing.

Growth curve model

The estimated growth model of the fetal face allows correction for age-related differences prior to the statistical shape analysis. The scale of the segmentation meshes was measured using an approximation of the face area based on the first two modes of the principal component analysis (PCA) on the points of the reconstructed facial mesh. We compared the resulting growth curve estimation from our dataset to published growth curve for head circumference²⁹(Fig. 1a).

The two curves are almost proportional, describing a similar growth rate, and would lead to a similar rescaling. The difference between both scale factor would be on average less than 1% and always inferior to 3% (maximum at 34 weeks gestation). The variability is higher for our measurement, probably due to the noise associated

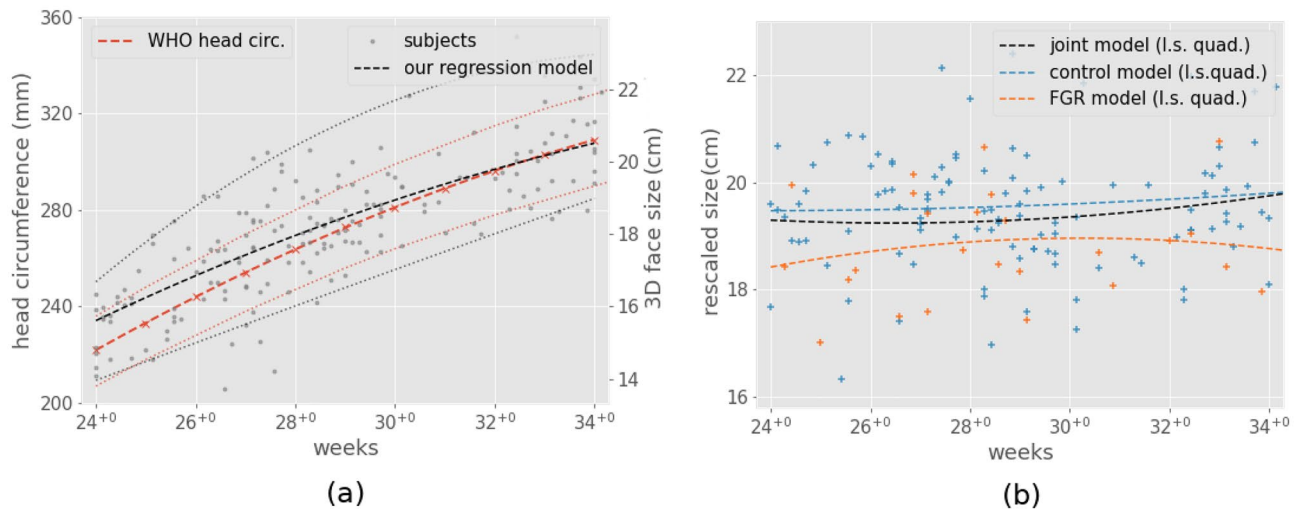


Figure 1. Size measurements of the fetal face. (a) face and head sizes with respect to age before rescaling: our regression model (black), reference model for head circumference from Kiserud et al.²⁹ (red). The quadratic growth curves are estimated using quantile regression (medians are represented by dashed lines, 5th and 95th percentiles by dotted lines). (b) face size measurements after rescaling. Regression models are shown for AGA cases (blue), FGR cases (orange) and both combined (black).

to the segmentation and automatic metric. We chose to rescale the meshes using the growth curve estimated from our dataset because it describes the face size directly rather than head size.

After size correction, using the growth curve, the average face size did not show any visible correlation with age (Fig. 1b). When comparing the FGR cases (orange line) to the AGA cases (blue line) a consistent difference is visible between the two groups.

Statistical shape model of the fetal face

The estimated morphological reference and shape variability in the population are both shown in Fig. 2. The reference shape, also called template, represents the anatomical mean shape of the population (Fig. 2a,b). As a result of averaging, the template morphology is symmetrical in appearance with smooth facial features and a loss of definition of fine detail.

The reconstruction error is relatively small with an average distance between the model mesh nodes and the original segmented surface of 0.29mm (highest = 0.79mm at a single point). Only 5 cases have an average distance superior to 0.4mm, these cases are associated with extreme phenotypes with overly large or small faces. The reconstruction error is partially associated with the smoothing of small noisy defects from the segmentations but is more pronounced in the areas with more fine details such as the lips and the nose (Fig. 2c).

The shape variability captured by the model (Fig. 2d) is one order of magnitude higher (2.10mm in average) than the reconstruction error, showing evidence that the shape model is able to capture most of the variation between subjects. The highest variability of the model is at the edge of the mesh while the forehead is the most stable part.

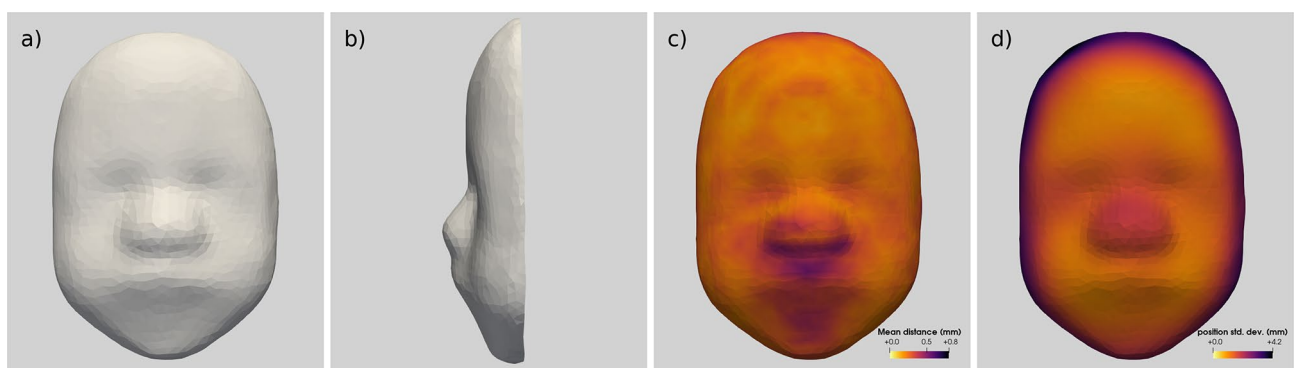


Figure 2. (a, b) Mean template of the fetal face, the template is warped by subject-specific deformations to match each individual morphology. (c) Average reconstruction error at each point, measured as the distance between the reconstructions and the original models. Small errors are yellow/orange and larger errors purple. (d) Shape variability measured using the standard deviation of the mesh node positions over all of the reconstruction. Purple denotes the areas with the highest variability and yellow/orange the most stable areas.

Some of the variations could be related to the pre-processing steps: the segmentation and mesh cutting may increase the variance near the edges while the iterative closest point (ICP) rigid registration may artificially reduce the variability in the central areas. In addition, any size differences would be more apparent near the edges than closer to the centre. To a lesser extent, some variability is noted around the nose, the nasal arch and the eyes. This effect could be associated with small differences in eyes and nose position, size or with more local differences in their shape. Less variability is measured on the forehead, other than the upper-edge, and around the mouth despite the mouth being a potentially well-defined anatomical features.

The main axes of variations computed using a PCA decomposition of the momenta are shown in Fig. 3. The first two modes capture 41.5% and 7.6% of the total variance respectively. 26 modes are needed to describe 90% of the total variance. The first mode (represented horizontally) shows the size of the face, specifically facial width to be the most important contributor to shape variability. The overall shape of the face varies between an oval to a more 'squared' facial shape. In this mode, while the whole face gets larger the centre of the face gets relatively smaller; with eyes getting closer and the nose getting a bit flatter.

The second mode (represented vertically) illustrates a strong emphasis on the width of the cheeks and forehead length with negative values showing a trend towards a longer thinner face. The second mode is positively correlated with age (Pearson's correlation coefficient test: $r = 0.28$, $p < 0.001$). A similar correlation with age is also noted with the 3rd, 5th and 6th modes. We can note that the first mode is not correlated with age ($r = 0.009$), reflecting the correction for age-related size changes.

Gestational age-related changes in the control population and growth restriction

Changes associated with GA were estimated using linear regression on the momenta. These changes complement the size change that was modelled in the previous section using the scalar growth curve. Figure 4a and b shows the model of the average face at 24 and 34 weeks. Advanced GA is associated with wider fuller cheeks, flatter foreheads, deeper eyes and slightly bigger noses compared to earlier GA.

Figure 4c is the projection on the face of the test statistic (z-value) used in the permutation-based likelihood test to assess where the age-related changes can be considered significant. The z-values are computed at each of the deformation control points and then represented on the template surface. Here the z-values is higher on the cheeks and in the centre of the face, especially around the nasal arch. This age effect was associated with a p -value $p < 0.001$ for 10000 random permutations.

The divergence of the deformation (momenta) vector fields on the surface (Fig. 4d) can help visualize the changes by focusing on the local volume changes. We see here an expansion of the cheekbones, the posterior aspect of the jaw, and around the nose, and a contraction of the chin, the left side of the face and near the eyes.

Describing fetal growth restriction with a shape model

Cases of fetal growth restriction (FGR) without any underlying genetic condition ($n=18$) were compared to the normal controls ($n=117$). The main difference is a difference of size with smaller average for the FGR cases (Fig. 5b) than for the control cases (Fig. 5a). The facial features do not show any other clinically meaningful difference between the two groups.

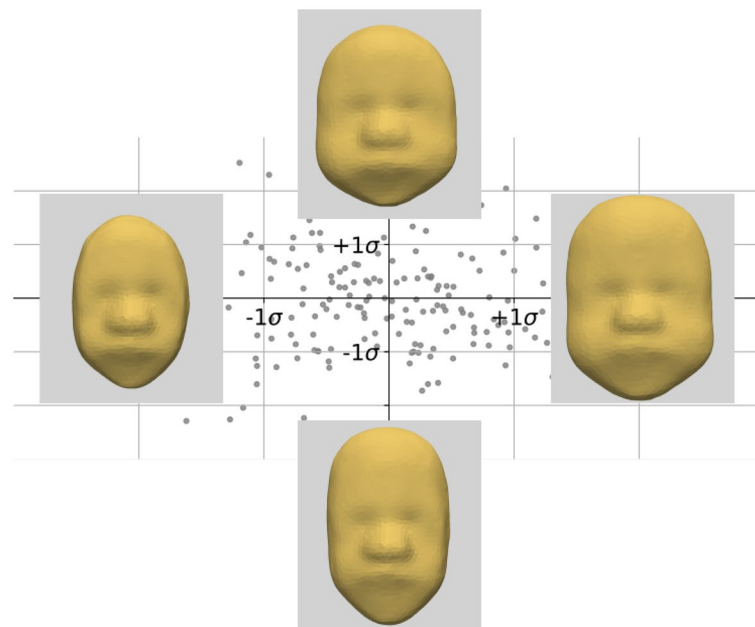


Figure 3. PCA projection of the first two modes of the SSM model. The first mode is represented horizontally, the second vertically. Each grey dot represents one subject. The faces illustrate the model at $-/+ 2$ standard deviations.

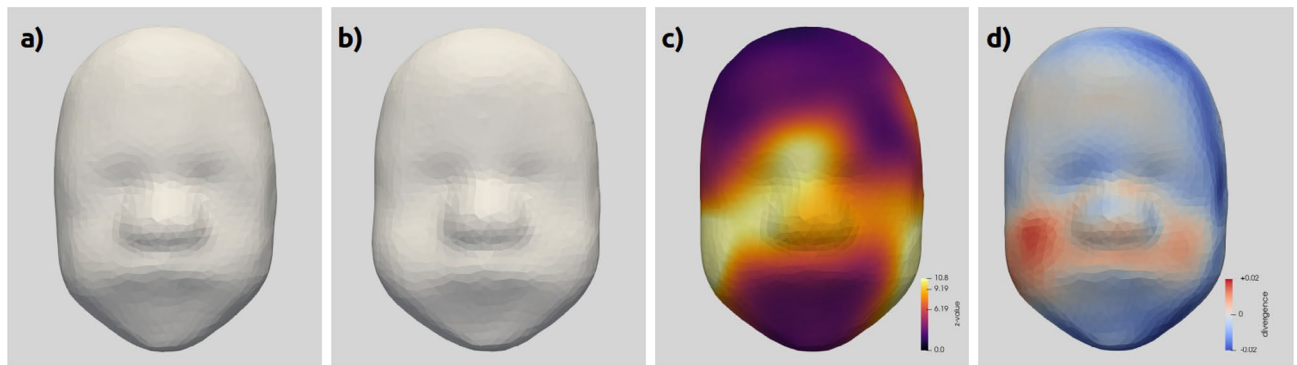


Figure 4. The effect of age on fetal facial morphology. (a, b) Model of the average morphology at (a) 24 weeks and (b) 34 weeks of gestational age. (c) Illustration of the local z-values assessing the effect of gestational age on morphology. The colour bar is normalised to show levels corresponding to $p=0.5$ (purple), $p=0.1$ (orange) and $p=0.05$ (yellow). (d) divergence of the deformation vector field (momenta).

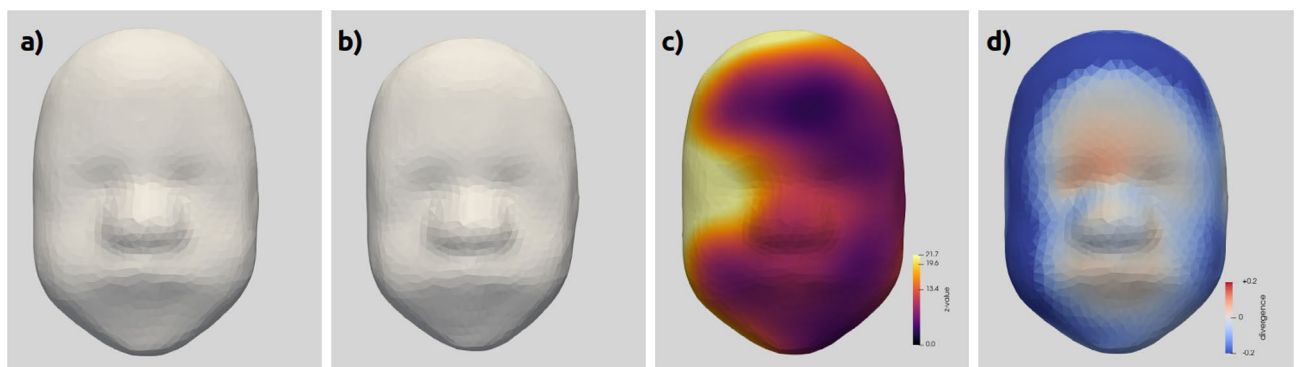


Figure 5. Differences associated with FGR: (a) average control face, (b) average FGR face (normal diagnosis only), (c) group comparison z-value assessing for the effect of FGR on the morphology, the colour-bar ticks are normalized to show values corresponding to $p=0.5$ (purple), $p=0.1$ (orange), and $p=0.05$ (light yellow) for the t^2 statistic. d) divergence of the deformation vector field (momenta).

The statistical significance of these differences is assessed using two-sample Hotelling's t^2 -test on the deformation parameters at each control point (Fig. 5c), highlighting differences over the mid-face with larger differences asymmetrical differences seen on the right side at the level of the eye. The p -value estimated using 1000 random permutations was equal to $p=0.008$.

The visualization of the divergence field (Fig. 5d) confirms these observations. A strong contraction in a ring all around the face, the effect is stronger on the right side. The centre of the face is on the contrary quite preserved, with an hint of a small expansion near the top of the nasal arch.

Variability within a normal face

Descriptive statistics, such as mean, variance or PCA modes, are useful to describe the population variability and the differences associated with clinical variables such as gestational age or diagnosis. In this section, we propose to use the shape model to inform us about a specific subject, and its deviation from the control cohort. Compared to the statistical tests of the previous sections, we do not focus here on the evaluation of a specific hypothesis but on a qualitative description of the differences. Our approach is a generalization, to morphological differences, of the reporting of z-scores (also called standard scores). The standard shape scores are computed at each point on the face based on the momenta values. The shape distribution parameters (mean and covariance) are estimated using the control AGA samples only. We propose two scores.

The first one, the *deformation score*, is based on the Hotelling distribution previously used, and takes into account the full 3D deformation vectors. The second one, the *orthogonal score*, only considers the projection of the deformation orthogonally to the surface. The second score is more directly interpretable as it describes changes in and out of the surface, the first one however can capture tangential morphological differences.

We show in Fig. 6 the results for two normal fetal faces. Case A, selected randomly, is a 24+3 gestational week fetus. Case B is the normal case with the highest average deformation score (32+3 weeks of gestation). In case A, the deformation score (Fig. 6c) shows small differences on the right upper lateral aspect of the forehead but overall no meaningful difference throughout the face. The orthogonal score is more variable over the face with high and low areas, the negative values on the upper-left demonstrate that this fetal face is characterized by receding forehead in this area (Fig. 6d). The deformation score for case B (Fig. 6g) demonstrates large differences at upper middle of the forehead, across the eyes and on the left aspect of the mouth and lateral border of the

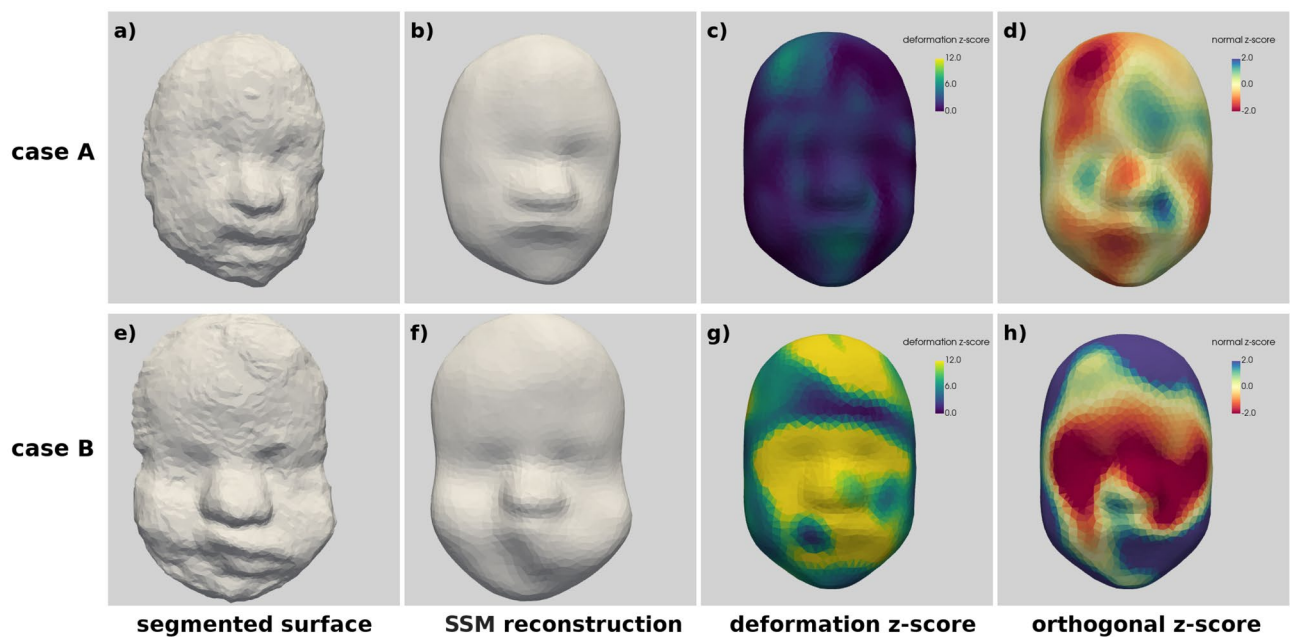


Figure 6. Two cases with normal diagnosis: case A top row and B (the most “abnormal normal”) bottom row. From left to right: original mesh, model reconstruction, deformation z-score, orthogonal z-score.

chin. The orthogonal score (Fig. 6h) shows that these differences in the forehead and mouth/lateral aspect of the chin represent values above the mean and the differences across the eyes and mid-face represent values below the mean. When inspecting this face visually, the lower half of the face appears to be slightly distorted as though the lower right aspect of the chin is pressed against something, distorting both that side of the mouth as well as the contralateral side which may have caused the differences identified and further highlights the importance of high-quality US volumes with clear borders.

Discussion

The aims of this study are to propose a shape model of the normal population and its variability, and to showcase a morphometric approach that objectively describe fetal face features and could assist the diagnosis of complex fetal phenotypic variants.

Segmentation algorithm

The analysis of a fetal face from a 3D volume generally require several manual processing steps, from the image acquisition, to the landmark annotation, or the image segmentation and creation of the 3D model. These steps are time consuming and challenging to automate due to imaging noise, fetal movements, low soft-tissue contrast, and occlusions. Fast and reproducible processing methods are required to conduct large studies and to develop clinical tools.

The introduction of two additional steps to our segmentation algorithm (4 manual landmarks and monogenic signal map) improved the performance compared to our previous work²⁶ reducing the need for manual refinement. This improvement enabled the processing of a large number of 3D US and the reconstruction of a population of 3D surface models of the fetal face. In order to allow the inclusion of the maximum number of subjects within this work we included cases requiring some manual refinement. It is however unrealistic in clinical practice to expect clinicians to undertake manual refinement of the segmentations prior to analysis.

Recent algorithmic contributions in fetal ultrasound image processing especially using deep learning^{16,30,31} are promising, but the development of a reliable segmentation of the 3D fetal face remains a challenge and publicly available data-sets very limited^{32,33}. The validation of our algorithm and the comparison to other existing approaches^{14,16} would benefit from experiments on an independent set of images, in particular if we want to extend the study to more diverse settings. For example, at earlier gestational ages, the signal to noise ratio would be weaker and smaller fetuses would mean lower resolution. The effect of segmentation variability induced by the use of different segmentation algorithms, imaging methods, and manual edits would also need to be assessed.

Growth model

The analysis of the morphological variability of the fetal face cannot ignore the growth associated with gestational age. The use of growth-curves and age-specific morphological references is an established standard in many clinical applications but shape modelling approaches generally consider these changes in a latter stage of the statistical analysis²⁰.

We proposed in the work to include an initial model to approximate these changes using an isotropic size increase. This original approach allows the 3D shape model to describe the inter-subject size variability and to

fully characterize subtle age-related changes without being dominated by the large differences between less and more advanced gestational age cases that could hinder the estimation of the statistical model and the analysis of the results. This approach is directly generalizable to other settings where size is an important factor and inter-subject size differences should be accounted for.

In our case, we showed that the reference growth rate could be directly estimated on the included sample or based on established measurements. In particular, we found similar results with a model based on facial surface area or on tabulated fetal head circumference. The highlighted size differences between AGA and FGR cases were an additional motivation to look, using computerized morphometry, for more extensive information about the associated morphological features or development delay that could be associated with this diagnosis.

Shape reference

In the last decades, statistical shape modelling has seen many potent development to describe population of 3D objects. In particular, morphable face models have been used to describe facial dysmorphisms^{19,20} with potential applications such as the early detection of genetic syndromes²¹. The quantitative analysis of the fetal face has however been limited by the difficult reconstruction of the 3D surface¹⁷.

We previously demonstrated the feasibility of using landmark-free diffeomorphic morphometry to model fetal faces²⁴. We show here that this approach is scalable to a large number of cases, and that the model is able to reconstruct every face with high accuracy at a relatively fine scale. In particular, the reconstruction error is small in comparison to the shape variability captured.

The model is also generative and the shape space can be used to sample synthetic cases that are realistic, as shown by the visualization of the 2 first PCA modes. The unsupervised statistical model is able to describe some localized shape features (shape and size of the nose, position of the eyes, etc.) and more global change of the overall shape of the face (oval, square, etc.) that are difficult to quantify otherwise. The correlation of PCA mode 2 with age illustrates some potential application of these descriptions to analyse clinical correlates. As such, the model, using the mean template and the distribution of shape vectors, appears to be a valid reference point for comparison.

Age-related changes and fetal growth restriction

Using multivariate statistical analysis of the shape model parameters, we highlighted morphological changes associated with age that were complementary to the uniform size increase. We described wider cheeks and more delineated face features of the fetus, not to our knowledge previously described in detail. These changes were assessed to be significant in our statistical model. To the best of our knowledge, these gestational age-related changes have not been previously described in details in the literature, however analysis of fetal facial bone growth using post-mortem CT scans³⁴ demonstrated that an increase in the anteroposterior mandibular diameter, maxillary width and bimaxillary distance is strongly correlated with an increase in gestational age. A weaker correlation to zygomatic bone growth was also identified. These findings are consistent with the age-related differences found in our analysis.

Similarly, we also showed significant differences associated between AGA and FGR cases. FGR diagnosis was associated with a smaller face but with no other facial feature was apparent in our analysis, aside from a slight asymmetrical difference at the level of the right eye.

It is difficult to compare the effect of gestational age and the differences AGA/FGR. First, age-related size differences have been partially accounted for by the growth curve model while no uniform scaling have been applied depending on the FGR diagnosis. And, overall, even if some similarities with the age-related changes could be suggested (on the cheek width or jaw line for example), no clear face features significantly differ between AGA and FGR cases and the global differences do not match changes that would be associated with any gestational age gap. We should however keep in mind that it is difficult to guarantee that the model is able to represent as accurately FGR cases than AGA cases due to the size difference.

Another hypothesis is that the suggested differences in the fetal cheeks may represent a degree of fetal malnutrition with the loss of buccal fat and sharply defined jaw and chin that can be seen in FGR and is a key part of the clinical assessment of nutritional status (CAN) score to assess newborns³⁵.

Our sample size is also relatively small and the localization of the difference is not the main outcome of the statistical analysis. For example, the asymmetrical difference may be explained by the capturing of the original 3D ultrasound volume; 8 of the 18 cases were captured with the fetal face deviated to the right resulting in slightly less clarity in the right side of the fetal face, 2 cases were slightly deviated to the left and the remaining cases were positioned centrally. This may have contributed to the asymmetrical difference noted in the analysis. Validation of these explanatory results on an independent dataset would be valuable.

This analysis also highlights the difficulty in visualising, identifying and characterising the morphological differences at the very edge of the facial mesh.

Shape z-scores

The analysis of individual cases provide one illustration of potential use of the models to extract, from the three-dimensional ultrasound, more detailed information about the facial morphology than the conventional 2D US analysis. Were this technique to be used in clinical practice in the future, we envisage that it would be used by clinicians to describe and analyse cases in which an underlying syndrome is suspected but in which the diagnosis remains elusive despite invasive tests, rather than being used as an unrestricted diagnostic tool.

The high scores of case B (above) demonstrated the limits of this approach. From its high scores, this subject seems to be abnormal or even an outlier. It is normal for some subjects to be on the extreme of the distribution. The surface mesh do not reflect the uncertainty of the segmentation in regions where the contrast is low and the

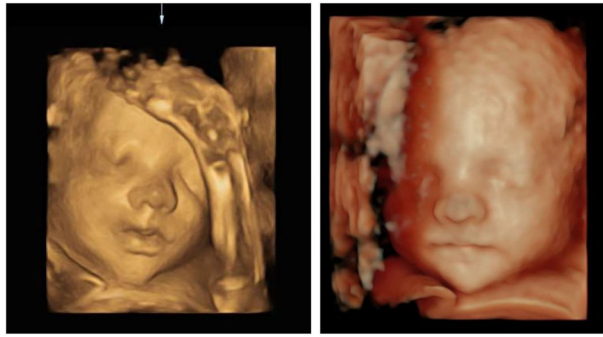


Figure 7. Two 3D US volumes collected from the same participant during the same appointment. The volume on the right was selected as the facial border were clearer and there was no fetal limb in contact with the fetal face as seen in the left image.

face border poorly visible. In this case, it is possible to precisely locate the face limits in an area near the mouth. Moreover, the bump on the forehead is arguably a segmentation error. The head is very near the edge of the US field of view, and this sharp edge tends to deform the image and pull the segmentation at the expense of the correct head border.

General limitations

The methodology that is described is not yet applicable to clinical practice, chiefly because of the amount of time required for segmentation and analysis (1 to 1.5 hour for one case with, in average, around 30 min of manual work). Also the study only includes two imaging centres and the segmentation process has been centrally and uniformly conducted. A more scattered clinical deployment would lead to more heterogeneous results.

A single surface mesh is a very simple representation of the morphological information available in a 3D ultrasound. It is only a summary of the interactive and dynamic signal visible during the examination. The selection of one volume to describe the entire face is already a limitation. In addition, meshing and smoothing lose fine, and less visible features.

A related limitation comes from the fact that we are trying to model the shape of soft tissues in a complex environment. In particular, the face is deformed by contact with fetal limbs, or maternal tissues, and these contacts are not only not modelled but also hidden to the final analysis based on the surface reconstruction only.

Perspectives

In this work, we describe a semi-automatic algorithm to build a shape model of the third trimester fetal face from 3D ultrasound and demonstrate how it can objectively describe morphological features and gestational-age related changes of normal fetal faces. This study represents an important step for objective assessment and characterisation of normal fetal face development and variability in the third trimester. Using these templates, future work will allow the model to be assessed for its discriminatory ability in the diagnosis of congenital anomalies and dysmorphic or syndromic facial features.

Methods

Patient population and image acquisition

A prospective cohort study of women attending for a fetal medicine scan between 24–34 weeks gestation was performed. Recruitment took place at 2 tertiary fetal medicine centres between January 2019 and May 2021. The study was approved at both units (REC approval 18/WM/0370 by the National Research Ethics Service Committee West Midlands – Edgbaston Research Ethics Committee and 579/2019/OSS/AOUPR by Comitato Etico dell'Area Vasta Emilia Nord and written informed consent from study participants was obtained.

Diagnosis information was extracted from clinical reports, and confirmed after 12 months follow-up. FGR diagnosis is based on the Delphi criteria³⁶.

Participants for whom it was not possible to collect a facial volume between the 24 and 34 week of gestation were excluded. Volumes that did not show the complete face or in which the fetal position did not allow for good-quality segmentation were also excluded from analysis. When several volumes were available for the same patient, the best one was subjectively selected based on image contrast, clarity and clearness of the facial borders and facial field (see Fig. 7).

Participants with incomplete follow up data or those with an uncertain diagnosis were also excluded from analysis. Finally in this study, we focus on patients without any diagnosed facial dysmorphism. The inclusion process is described in Fig. 8.

The ultrasound images included in this study are 3D static acquisitions obtained in the midsagittal plane of the fetal face, using the highest resolution acquisition mode with a low frequency probe (4–8 MHz) using a Voluson E8 or E10 (GE Healthcare) or a Hera W10 (Samsung) ultrasound machine. The volumes were aligned in the multiplanar view along the x, y and z axes and subsequently exported as 3D images.

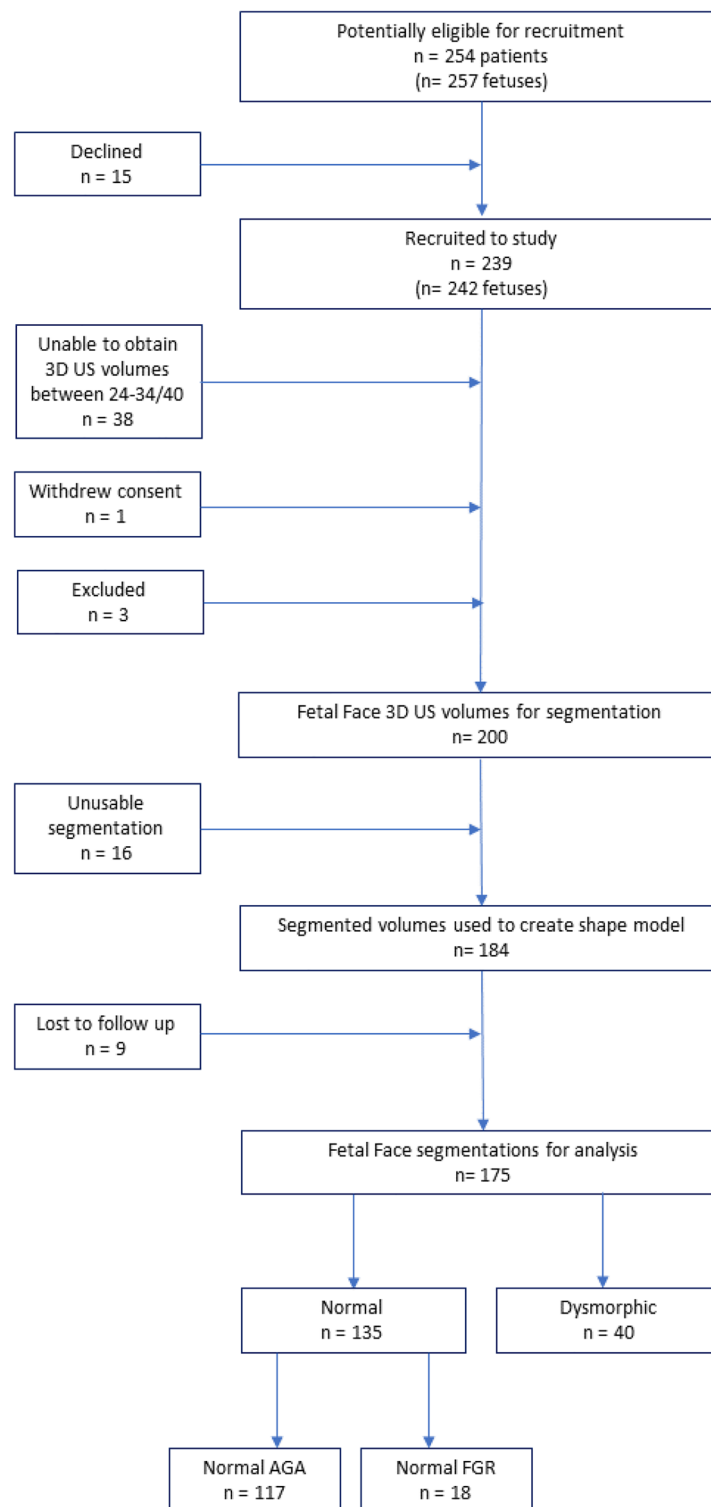


Figure 8. Flowchart of recruitment, 3D US volume acquisition, segmentation quality and final numbers for analysis.

Acquisition and processing of the patient data were carried out in accordance with clinical guidelines and data protection regulations.

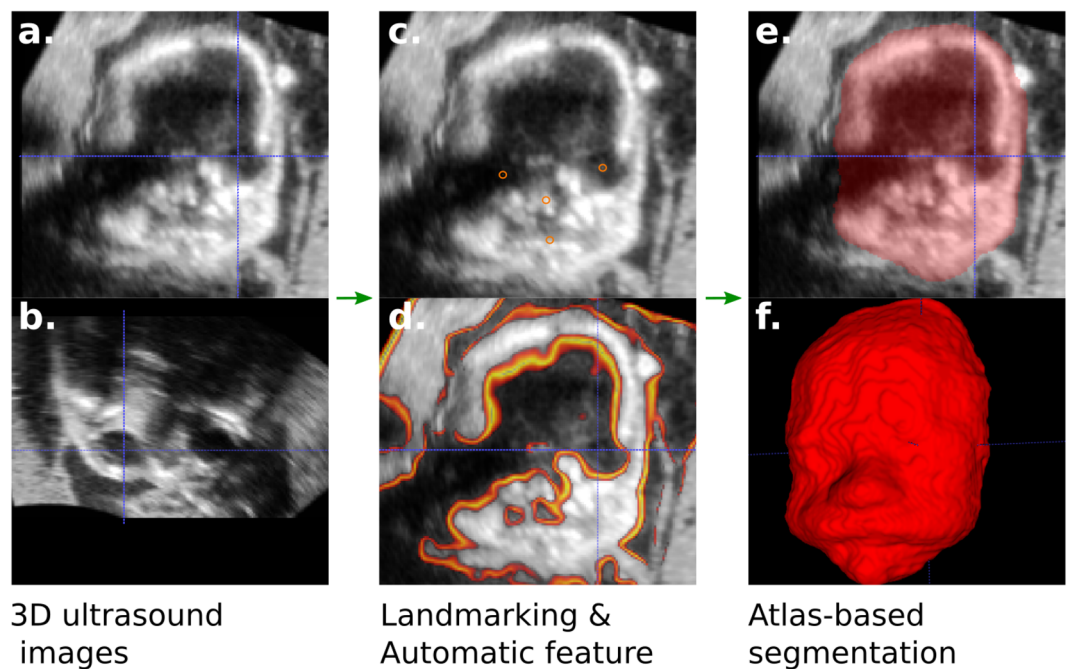


Figure 9. Overview of the atlas-based segmentation algorithm steps: (a, b) 3D US volume (c) 4 manual landmarks (eyes, nose, midpoint of the lips) (d) monogenic signal map (e) segmentation mask (f) 3D surface mesh.

Semi-automatic segmentation workflow

Several steps are required to create a 3D reconstruction mesh representing the face shape from the ultrasound volume (Fig. 9a,b). We previously adapted the atlas-based segmentation algorithm developed by Zuluaga et al. (2013)²⁷ to the fetal faces with good result²⁶ however it still required significant manual refinement.

We introduced two further improvements to the algorithm in this work. Firstly, the images are initially rigidly aligned using four manual landmarks (eyes, nose, midpoint of the lips, see Fig. 9c) to correct for head position. Secondly, the image registration similarity, key element of the segmentation algorithm, includes an additional term designed to guide the segmentation with US-specific signal properties. This term is defined by the normalized cross-correlation between monogenic phase asymmetry map computed for each image (Fig. 9d). The asymmetry map is defined by:

$$R(x) = \frac{\max(0, |f_o(x)| - |f_e(x)| - T)}{A(x) + \varepsilon}$$

where f_e is the radial part of the image signal, f_o is the odd part, A is the local amplitude, and T a user-defined threshold between 0 and 1 (that we set to 0.5). We refer to Bridge's introduction to monogenic signals³⁷ for more details.

Monogenic signals have proven to be useful to capture edge features in clinical ultrasound images³⁸. A similar idea has been used by Qiu et al. (2017)³⁹ who computes the phase congruency maps of ultrasound volumes to drive the registration of an atlas-based segmentation of neonatal cerebral ventricles.

Every resulting segmentation was visually checked and manually edited in ITK-snap (itksnap.org) to correct for mis-segmentation and small missing parts (Fig. 9e). If the boundary could not be reasonably estimated, the volume was excluded. This step is crucial to correct for errors in the automatic segmentation caused by image artefacts or obstruction of the face and to select the best image for each subject. A closed surface mesh was then exported (Fig. 9f).

Mesh processing and size normalisation

To assess meaningful features in the shape analysis, the 3D meshes have to be relatively regular and all orientated in the same position. This is done in two steps. Firstly, remeshing is performed using pygalmesh software⁴⁰ to optimise the cell shapes and provide a more uniform cell density. This step can also be used to control for the average number of mesh elements, an important factor in the trade-off between computational cost and resolution. We set the target triangle radius to 2mm and the facet distance to 0.1mm, leading to approximately 6700 cells per mesh. Secondly, meshes are rigidly aligned using the 4 previously defined manual landmarks and automatic iterative closest point registration (ICP). Meshes are then automatically cut based upon a predefined plane and the orientation of cell normals.

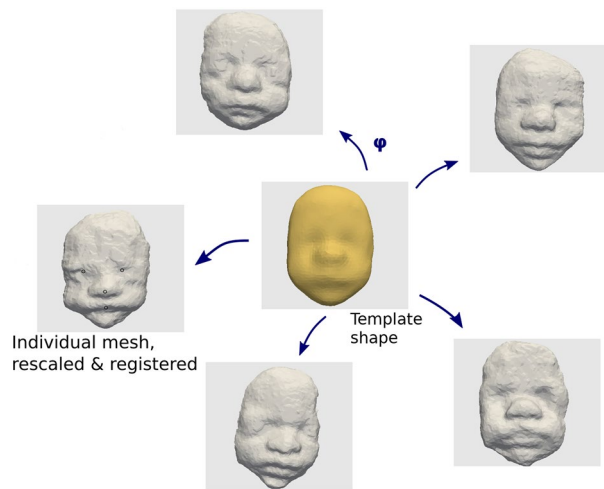


Figure 10. Schematic representation of the statistical shape model: template shape to individual cases/ deformations.

An important question to address is the size normalization. The dominant change between 24 and 34 weeks is associated with the fetal growth and, if unaccounted for, this size variability can hide more subtle, interesting, and potentially clinically significant differences, in the estimation of the statistical shape model. The growth can be, in a first step, approximated by an isotropic size increase. Morphometric measurements are often normalized for organ or body size (see^{41,42} for example), however we are here interested in modelling the change with age and size differences are not only caused by the fetal growth but also by inter-subject variability. Rescaling the geometries through similarity registration or using measurements of the fetal face (such as head circumference) would erase this information despite it being clinically relevant (for the study of fetal growth restriction (FGR) for example). We chose an intermediate approach where the faces are rescaled according to a growth model giving the average size at every gestational age. Younger cases are enlarged, and older cases are shrunk towards an average size corresponding to the median age in our population (29 weeks).

In absence of ground-truth size measurements (such as a manual head circumference measurement), we chose to use the square root of the area of the face estimated using the variance along the two principal directions of the point cloud. This point cloud-variance metric has been chosen rather than a more direct measurement of the mesh surface area because of the strong, albeit consistent, bias introduced by the mesh processing steps on the direct measurement (the mesh exported from the segmentation is a close surface with a noisy back side while the pre-process mesh only preserves the open surface corresponding to the face, this process reduces the total mesh surface by around 30 percent). The selected measure is also fully automatic, fast and easily adaptable to various settings.

The growth model is defined by a quadratic curve fitted by quantile regression to this surrogate size in our complete population (cases with and without pathologies). This model will be compared to tabulated growth curves for head circumference²⁹.

Statistical shape modelling

The shape analysis approach used in this work relies on the Deformetrica framework²⁵. In this framework, geometrical objects are compared through the dense deformations of space that can transform one shape into another. The aim is to enable the comparison of morphologically distinct objects without requiring any point correspondence. The space of valid deformations is parametrized by infinitesimal displacements, called momenta. These momenta are defined in the whole 3D space and belong to a reproducing kernel Hilbert space (RKHS) built using a Gaussian kernel with fixed width and a set of control points. The advantage is that, by construction, the deformations are diffeomorphic, fully parametrized by 3D vectors defined on the set of spatial control points, and that linear operations, and usual Euclidean statistics, can be used in this space. The position of the control points is optimized to efficiently sample the ambient space occupied by the 3D surfaces.

This framework is then used to build a statistical model of the shape population. The most common approach is called atlas and is constituted of two main components: a reference shape called template that is representative of the anatomical mean shape in the population, and individual deformations that map this template to each individual shape and capture the morphological variability in the population (Fig. 10). Once the atlas (the template and the individual deformations) has been estimated, the statistical analysis is conducted directly on the deformation parameters (the momenta vectors defined on every control points).

Shape atlas estimation

The shape model is controlled by two main parameters: the resolution λ_W that controls the scale at which the objects are described, and the stiffness of the deformations λ_V that controls how smooth the deformations between subjects will be⁴³. It is possible to intuit reasonable values for these parameters based on the regularity

of the objects under study. To inform our choice, we conducted preliminary experiments to test for a range of values between 3mm and 20mm (using only the subset of normal cases to speed up the process). The best λ_W value was selected using the Akaike information criteria (AIC) to balance for the increase of number of deformation control points when the resolution decrease. The average reconstruction error was used to select λ_V .

No major difference was observed for $\lambda_W \in 5 : 10$ with a minimum at 7mm. Smaller values were unable to be evaluated due to RAM limitations. We then set $\lambda_W = 7$. The reconstruction error was monotonically increasing for $\lambda_V \geq 5$, but slowly for $\lambda_V \in 5 : 9$ and with numerical instabilities for smaller values (not converging for $\lambda_V = 4$ and poor result for $\lambda_V = 3$). As a compromise, we chose to set $\lambda_V = 7$ mm.

The initialisation of the optimization could also play a role; mainly, the initial template shape, and the position of the deformation control points. To address these two questions, we iterate over several model estimations focusing on these aspects first. Designing an approximate average face is not too difficult as there are no major topological differences between subjects; from a random initialisation (one average-sized subject), a first template is estimated and then manually smoothed and remeshed. The second optimization run focuses on control point positions, with a fixed template. Points associated with almost no momenta variance (less than 0.01% the maximum) were then filtered out in order to remove the ones that were outside the space covered by the faces and speed up the model estimation.

Statistical analysis of the momenta

The atlas model and the compact parametrization of deformations by momenta facilitate the study of morphological features and of their relations to external independent variables. However, the statistical analysis is still challenging. One of the main challenge lies in the dimensionality of the data, even if it has been strongly reduced compared to the original surface meshes, it is still too large for naive statistical approaches and visualization. For reference, our final model involved 412 control points, by consequence the shapes are characterized by $D = 412 \times 3 = 1236$ parameters. We summarize here the key elements of the main statistical methods that will be used to analyse the momenta distribution and their relation to clinical variables.

Principal component analysis (PCA) is used to describe the shape distribution using a reduced number of modes. Singular value decomposition is performed on the momenta matrix and the first modes are extracted. It is a useful tool for visualization and interpretation purposes as it gives a low-dimensional representation of the shape population. It can also hint towards important variability factors that could be explicitly considered in the model. For instance, size is, in many applications, the first or one of the first modes⁴⁴.

Multivariate discriminant statistical testing is used to assess the existence and localization of a correlation between clinical variables and morphological features. A statistic is computed at each deformation control point and a permutation scheme is used to estimate an empirical distribution of the maximum of this statistic under H_0 . This approach gives a non-parametric p -value that is both a global assessment and an indicator of the localization of the effect. This approach avoids the problem of multiple testing⁴⁵.

A specificity of our work is that the statistics, computed at each control points, are also multivariate. Indeed, we need to account for the 3 directions of the momenta vectors. In this context, we use the Hotelling t^2 -statistic to assess the difference of mean value between two groups. If the independent variable is continuous (such as age for example) or if multiple variables are jointly considered (multiple diagnoses and age correction for example), we used log-likelihood testing for multivariate linear regression⁴⁶. We can note that this approach is equivalent to the Hotelling test when there is only one binary variable. Unless specified, p -values are calculated using $N=1000$ random permutations and significance level is chosen to $p=0.05$.

Shape-related z-scores and assessment of the deviation from the control population of individual subjects

We propose to use *standardized shape scores* to describe and visualize the deviation of individual morphologies from the population distribution characterized by the statistical shape model. This approach is a generalization of the reporting of a z-score (also called standard score) in the case of a scalar variable. Standard scores could be computed directly on the momenta parameters at every control points; to get a more visual and detailed description we propose to compute them at each cell center of the template mesh instead. The momentum vector m for a subject k at any point x in space is directly computed from these momenta parameters $(a_i)^k$ defined at each control point $(c_i)_i$ (see Durrleman et al. (2014)²⁵ for more details):

$$m^k(x) = \sum_i K(x, c_i) a_i^k$$

Parameters a^k for a new subjects are estimated through registration to the template face.

We propose two scores: the *deformation score* based on the Hotelling statistic for 3D vectors, and the *orthogonal score* for changes locally normal to the surface. We first compute the average $\bar{m}(x)$ and the empirical covariance $\bar{\Sigma}(x)$ of the momenta vectors on the set of control subjects used to build the model. Then: we define at each point the *deformation score* of subject k by

$$z_d^k = (m^k - \bar{m})^T \bar{\Sigma}^{-1} (m^k - \bar{m})$$

The *orthogonal score* uses the local surface normals $n(x)$ and can then only be computed on the template surface mesh. We define $\bar{\sigma}_n^2 = \frac{1}{N-1} \sum ((m - \bar{m}) \cdot n)^2$ the empirical standard deviation, on the control samples, of the projection of the momenta along these normals. The orthogonal score writes

$$z_o^k = \frac{(m^k - \bar{m}) \cdot n}{\bar{\sigma}_n}$$

A signature score, similar to the *orthogonal score* has been used for paediatric patient by Matthews et al. (2021)²⁰. The changes associated with this second score are more directly interpretable but do not take into account the complete deformation vectors and can overlook more complex morphological differences.

If we assume the momenta vector $m(x)$ follows a multivariate Gaussian distribution $\mathcal{N}(\bar{m}, \bar{\Sigma})$ (only true in first approximation), the *orthogonalscore* follows a normalized Gaussian distribution $\mathcal{N}(0, 1)$ with the usual 95% confidence interval $[-2; +2]$. The *deformationscore* would however follow a chi-squared distribution with 3 degrees of freedom χ_3^2 (confidence interval $[0; +7.8]$).

Data availability

The ultrasound images analysed in the current study are not publicly available due to privacy concern. The segmented 3D surface meshes are however unidentifiable and are available, with corresponding diagnoses, on the UCL research data repository under the DOI 10.5522/04/23717376. For review purposes, the data is available at <https://figshare.com/s/861ec1626bc36a562c4d> and the shape model parameters at <https://figshare.com/s/4bbaf08ac95523f5b86>.

Code availability

Access to the implementation of the segmentation workflow and of the statistical analysis is available on demand for review at <https://gitlab.com/rsivera/segment-faces>. The repository will be made publicly available before publication.

Received: 26 June 2023; Accepted: 19 December 2023

Published online: 22 February 2024

References

- Whitworth, M., Bricker, L., Neilson, J. P. & Dowswell, T. Ultrasound for fetal assessment in early pregnancy. *Cochrane Database Syst. Rev.* <https://doi.org/10.1002/14651858.CD007058.pub2> (2010).
- Pooh, R. K., Shiota, K. & Kurjak, A. Imaging of the human embryo with magnetic resonance imaging microscopy and high-resolution transvaginal 3-dimensional sonography: Human embryology in the 21st century. *Am. J. Obstet. Gynecol.* **204**(77), e1–16. <https://doi.org/10.1016/j.ajog.2010.07.028> (2011).
- Levaillant, J.-M. et al. Prenatal phenotypic overlap of Costello syndrome and severe Noonan syndrome by tri-dimensional ultrasonography. *Prenat. Diagn.* **26**, 340–344. <https://doi.org/10.1002/pd.1412> (2006).
- David, A. L. et al. Diagnosis of apert syndrome in the second-trimester using 2d and 3d ultrasound. *Prenat. Diagn.* **27**, 629–632. <https://doi.org/10.1002/pd.1758> (2007).
- Chen, C.-P. et al. Abnormally flat facial profile on two- and three-dimensional ultrasound and array comparative genomic hybridization for the diagnosis of pallister-killian syndrome. *Taiwan. J. Obstet. Gynecol.* **49**, 124–128. [https://doi.org/10.1016/S1028-4559\(10\)60027-7](https://doi.org/10.1016/S1028-4559(10)60027-7) (2010).
- Clark, D. M. et al. Identification of a prenatal profile of Cornelia de Lange syndrome (CdLS): A review of 53 CdLS pregnancies. *Am. J. Med. Genet. A* **158A**, 1848–1856. <https://doi.org/10.1002/ajmg.a.35410> (2012).
- Theulier, E., Levaillant, J. M., Roume, J., Quarello, E. & Bault, J. P. Cornelia de Lange syndrome: Specific features for prenatal diagnosis. *Ultrasound Obstet. Gynecol.* **49**, 668–670. <https://doi.org/10.1002/uog.15788> (2017).
- Merz, E. & Abramowicz, J. S. 3D/4D ultrasound in prenatal diagnosis: Is it time for routine use?. *Clin. Obstet. Gynecol.* **55**, 336. <https://doi.org/10.1097/GRF.0b013e3182446ef7> (2012).
- Araujo Júnior, E. Three-dimensional ultrasound in fetal medicine after 25 years in clinical practice: Many advances and some questions. *Radiol. Brasil.* **49**, 5–6. <https://doi.org/10.1590/0100-3984.2016.49.5e1> (2016).
- Merz, E. & Pashaj, S. Advantages of 3D ultrasound in the assessment of fetal abnormalities. *J. Perinat. Med.* **45**, 643–650. <https://doi.org/10.1515/jpm-2016-0379> (2017).
- Werner, H. et al. Prenatal diagnosis of Apert syndrome using ultrasound, magnetic resonance imaging, and three-dimensional virtual/physical models: Three case series and literature review. *Childs Nerv. Syst.* **34**, 1563–1571. <https://doi.org/10.1007/s00381-018-3740-y> (2018).
- Raga, F. et al. High definition real-time ultrasound (HDLive) of embryonic and fetal malformations before week 16. *Donald School J. Ultrasound Obstet. Gynecol.* **7**, 1–8. <https://doi.org/10.5005/jp-journals-10009-1266> (2013).
- Anquez, J., Angelini, E. D., Grangé, G. & Bloch, I. Automatic segmentation of antenatal 3-D ultrasound images. *IEEE Trans. Biomed. Eng.* **60**, 1388–1400. <https://doi.org/10.1109/TBME.2012.2237400> (2013).
- Bonacina, L., Froio, A., Conti, D., Marcolin, F. & Vezzetti, E. Automatic 3D foetal face model extraction from ultrasonography through histogram processing. *J. Med. Ultrasound* **24**, 142–149. <https://doi.org/10.1016/j.jmu.2016.08.003> (2016).
- Feng, S., Zhou, S. K., Good, S. & Comaniciu, D. Automatic fetal face detection from ultrasound volumes via learning 3D and 2D information. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2488–2495. <https://doi.org/10.1109/CVPR.2009.5206527> (2009).
- Singh, T., Kudavelly, S. R. & Venkata Suryanarayana, K. Deep learning based fetal face detection and visualization in prenatal ultrasound. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 1760–1763. <https://doi.org/10.1109/ISBI48211.2021.9433915> (2021).
- Alomar, A. et al. 3D Fetal Face Reconstruction from Ultrasound Imaging. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 615–624. <https://doi.org/10.5220/0010340306150624> (2021).
- Speranza, D. et al. Additive manufacturing techniques for the reconstruction of 3D fetal faces. *Appl. Bionics Biomech.* <https://doi.org/10.1155/2017.9701762> (2017).
- Hallgrímsson, B. et al. Automated syndrome diagnosis by three-dimensional facial imaging. *Genet. Med.* **22**, 1682–1693. <https://doi.org/10.1038/s41436-020-0845-y> (2020).
- Matthews, H. S. et al. Large-scale open-source three-dimensional growth curves for clinical facial assessment and objective description of facial dysmorphism. *Sci. Rep.* **11**, 12175. <https://doi.org/10.1038/s41598-021-91465-z> (2021).

21. Tu, L. *et al.* Three-Dimensional Face Reconstruction from Uncalibrated Photographs: Application to Early Detection of Genetic Syndromes. In Greenspan, H. *et al.* (eds.) *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging and Clinical Image-Based Procedures*, Lecture Notes in Computer Science, 182–189. https://doi.org/10.1007/978-3-030-32689-0_19 (2019).
22. Miyagi, Y., Hata, T., Bouno, S., Koyanagi, A. & Miyake, T. Recognition of facial expression of fetuses by artificial intelligence (AI). *J. Perinat. Med.* **49**, 596–603. <https://doi.org/10.1515/jpm-2020-0537> (2021).
23. Tang, J. *et al.* A deep-learning-based method can detect both common and rare genetic disorders in fetal ultrasound. *Biomedicines* **11**, 1756. <https://doi.org/10.3390/biomedicines11061756> (2023).
24. Dall'Asta, A. *et al.* Quantitative analysis of fetal facial morphology using 3D ultrasound and statistical shape modeling: A feasibility study. *Am. J. Obstet. Gynecol.* **217**(76), e1–76.e8. <https://doi.org/10.1016/j.ajog.2017.02.007> (2017).
25. Durrleman, S. *et al.* Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *Neuroimage* **101**, 35–49. <https://doi.org/10.1016/j.neuroimage.2014.06.043> (2014).
26. Clark, A. E. *et al.* Developing and testing an algorithm for automatic segmentation of the fetal face from three-dimensional ultrasound images. *R. Soc. Open Sci.* **7**, 201342. <https://doi.org/10.1098/rsos.201342> (2020).
27. Zuluaga, M. A., Cardoso, M. J., Modat, M. & Ourselin, S. Multi-atlas propagation whole heart segmentation from MRI and CTA using a local normalised correlation coefficient criterion. In Ourselin, S., Rueckert, D. & Smith, N. (eds.) *Functional Imaging and Modeling of the Heart*, Lecture Notes in Computer Science, 174–181. https://doi.org/10.1007/978-3-642-38899-6_21 (2013).
28. Cardoso, M. J., Modat, M., Ourselin, S., Keihaninejad, S. & Cash, D. Multi-STEPS: Multi-label similarity and truth estimation for propagated segmentations. In *2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 153–158. <https://doi.org/10.1109/MMBIA.2012.6164748> (2012).
29. Kiserud, T. *et al.* The world health organization fetal growth charts: A multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. *PLoS Med.* **14**, e1002220. <https://doi.org/10.1371/journal.pmed.1002220> (2017).
30. Cerrolaza, J. J. *et al.* Deep learning with ultrasound physics for fetal skull segmentation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 564–567. <https://doi.org/10.1109/ISBI.2018.8363639> (IEEE, 2018).
31. Chen, C. *et al.* Region Proposal Network with Graph Prior and IoU-Balance Loss for Landmark Detection in 3D Ultrasound. <https://doi.org/10.48550/arXiv.2004.00207> (2020).
32. Fiorentino, M. C., Villani, F. P., Di Cosmo, M., Frontoni, E. & Moccia, S. A review on deep-learning algorithms for fetal ultrasound-image analysis. *Med. Image Anal.* **83**, 102629. <https://doi.org/10.1016/j.media.2022.102629> (2023).
33. Ramirez Zegarra, R. & Ghi, T. Use of artificial intelligence and deep learning in fetal ultrasound imaging. *Ultrasound Obstet. Gynecol.* **62**, 185–194. <https://doi.org/10.1002/uog.26130> (2023).
34. Lesieur, E. *et al.* Fetal facial bone growth: Post-mortem CT analysis. *Morphologie* **107**, 116–126. <https://doi.org/10.1016/j.morpho.2022.05.006> (2023).
35. Metcalf, J. Clinical assessment of nutritional status at birth: Fetal malnutrition and SGA are not synonymous. *Pediatr. Clin. North Am.* **41**, 875–891. [https://doi.org/10.1016/S0031-3955\(16\)38836-8](https://doi.org/10.1016/S0031-3955(16)38836-8) (1994).
36. Gordijn, S. J. *et al.* Consensus definition of fetal growth restriction: A Delphi procedure. *Ultrasound Obstet. Gynecol.* **48**, 333–339. <https://doi.org/10.1002/uog.15884> (2016).
37. Bridge, C. P. Introduction To The Monogenic Signal. *arXiv:1703.09199 [cs]* (2017).
38. Hacihaliloglu, I., Abugharbieh, R., Hodgson, A. J. & Rohling, R. N. Bone surface localization in ultrasound using image phase-based features. *Ultrasound Med. Biol.* **35**, 1475–1487. <https://doi.org/10.1016/j.ultrasmedbio.2009.04.015> (2009).
39. Qiu, W. *et al.* Automatic segmentation approach to extracting neonatal cerebral ventricles from 3D ultrasound images. *Med. Image Anal.* **35**, 181–191. <https://doi.org/10.1016/j.media.2016.06.038> (2017).
40. Schlömer, N. *et al.* Pygalmesh at github.com/nenschloe/pygalmesh. <https://doi.org/10.5281/zenodo.5564819> (2021).
41. Buckner, R. L. *et al.* A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: Reliability and validation against manual measurement of total intracranial volume. *Neuroimage* **23**, 724–738. <https://doi.org/10.1016/j.neuroimage.2004.06.018> (2004).
42. Simone, G. d. *et al.* Normalization for body size and population-attributable risk of left ventricular hypertrophy*: The Strong Heart Study. *American Journal of Hypertension* **18**, 191–196. <https://doi.org/10.1016/j.amjhyper.2004.08.032> (2005).
43. Bruse, J. L. *et al.* A statistical shape modelling framework to extract 3d shape biomarkers from medical imaging data: Assessing arch morphology of repaired coarctation of the aorta. *BMC Med. Imaging* **16**, 40. <https://doi.org/10.1186/s12880-016-0142-z> (2016).
44. Heutinck, P. *et al.* Statistical shape modelling for the analysis of head shape variations. *Journal of Cranio-Maxillofacial Surgery* **49**, 449–455. <https://doi.org/10.1016/j.jcms.2021.02.020> (2021).
45. Nichols, T. E. & Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Hum. Brain Mapp.* **15**, 1–25. <https://doi.org/10.1002/hbm.1058> (2002).
46. Sivera, R. *et al.* Voxel-based assessments of treatment effects on longitudinal brain changes in the Multidomain Alzheimer Preventive Trial cohort. *Neurobiol. Aging* **94**, 50–59. <https://doi.org/10.1016/j.neurobiolaging.2019.11.020> (2020).

Acknowledgements

This work was supported by the National Institute for Health Research (NIHR) Biomedical Research Centres based at Imperial College Healthcare NHS Trust and Imperial College London (grant no. P74419), and Great Ormond Street Hospital for Children NHS Trust and University College London. CL is supported by the NIHR Biomedical Research Centre (BRC) based at Imperial College Healthcare NHS Trust and Imperial College London. The research was funded by the Engineering and Physical Sciences Research Council (EP/N02124X/1), the European Research Council (ERC-2017-StG-757923) and Great Ormond Street Hospital for Children Charity (Grant no. 12SG15). The views expressed are those of the authors and not necessarily those of the NHS, Imperial College London, University College London, the NIHR or the Department of Health.

Author contributions

A.C., A.D.A., S.S., C.L. conceptualised the study. A.C. and A.D.A. were instrumental in patient recruitment and the collection of data. R.S. and S.S. designed the analytical methodology. R.S. implemented and performed the analyses. A.C. and R.S. wrote the first draft of the manuscript. A.C., A.D.A., T.G. and C.L. contributed commentary on the clinical implications of the methods developed. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.C.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024