This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Exploring the Influence of Prenatal Alcohol Exposure on Face Shape and Brain Development

Daniel Zieger[1]✉[0009−0004−3187−0428], Yana Kuznetcov[1], Maximilian Weiherer[1][0009−0001−4615−4684], Shreya Kapoor[1][0000−0003−2619−8257], Catherine Lebel[2][0000−0002−0344−4032], Marc Stamminger[1][0000−0001−8699−3442], and Bernhard Egger[1][0000−0002−4736−2397]

[1] Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
daniel.zieger@fau.de
[2] University of Calgary, Canada

Abstract. Prenatal alcohol exposure (PAE) leads to fetal alcohol spectrum disorder (FASD) which can manifest as a constellation of cognitive, behavioral and/or emotional challenges, as well as facial dysmorphologies. In this work, we explore how PAE influences the brain and facial dysmorphology and evaluate which facial shape changes reflect the exposure or brain changes. Using MRI scans from exposed and unexposed cohorts, we first preprocess the scans and extract surface meshes of the faces. We then register different statistical face shape models to the extracted meshes using a non-rigid iterative closest point algorithm. The resulting low-dimensional shape parameters are employed to regress the volume of brain regions, as well as to classify the exposure status of a subject. By comparing the regressions fitted using the separate cohorts as well as combined datasets, we analyze the correlations between face deformation and the impact of PAE on the brain volume of interest (VOI). We found that volumes of certain brain regions, such as the right middle cingulate gyrus and right accumbens area, can be better predicted on the combined dataset than on individual exposure splits. This might show PAE-induced effects that potentially lead to a correlation between the two groups due to different levels of exposure. Furthermore, we utilize weights of the fitted classifier as an interpretability metric to identify which facial regions affect the decision the most. Our classification model achieves an F1 score of 0.92 ± 0.04 and the interpretability metric highlights regions like the midface, nasal tip, and perioral region which are consistent with dysmorphologies caused by PAE known from the literature.

Keywords: Prenatal alcohol exposure · Statistical shape model · Face shape analysis· Brain region volume regression · Interpretability

1 Introduction

With an estimated 9.8% of the population, nearly every tenth newborn is subject to PAE [30]. Depending on the amount of alcohol exposure during pregnancy, PAE

manifests as a combination of cognitive, behavioral and/or emotional challenges and often co-occurs with attention-deficit hyperactivity disorder or learning disabilities, which are collectively referred to as FASD. Severe exposure can even result in stillbirth [19]. Providing the affected children with appropriate support early on is immensely helpful, especially for their early development. Early detection of FASD, besides through confirmation of alcohol consumption by the mother, is very limited. The typical facial phenotype is a necessary component for a FASD diagnosis and is in many countries used only in combination with confirmed alcohol exposure [16]. Cognitive or behavioral symptoms start to manifest later in early childhood. In this study, our focus is on confirmed PAE rather than diagnosed FASD, because FASD is harder to diagnose, requires cognitive assessment, and is therefore underdiagnosed [39]. We investigate the relationship between PAE and face shape, as well as the relationship between face shape and brain development. We know from literature that there are facial dysmorphologies and brain region changes associated with FASD caused by PAE, and our goal is to explore correlations between face shape and alterations in brain region volume, thus confirming existing knowledge and establishing a link between the two. Additionally, we aim to detect changes in facial structure to screen whether a subject was potentially exposed to alcohol during pregnancy. To achieve these goals, we analyze face shape and correlate it with brain region size using MRI scans of exposed and unexposed children. First, we extract a face mesh from MRI scans. Then, we use statistical face models to register them to the meshes, which removes small defects on the scan through smoothing and, more importantly, gives us a standardized low-dimensional representation of each face. We later use those shape parameters, which describe how a face is formed, as input for regression and classification methods. Using weights of the classification method, we highlight regions on the registered face meshes that influenced the results the most. We also visualize the deformations correlated to PAE or volume changes of specified brain regions in the datasets. We further investigated if there are sex specific effects in our cohort and repeated our analysis with a face model focusing on inner facial features to reduce effects that might be caused by differences in the specific acquisition setups for the two cohorts. Our key contributions are:

- We present a pipeline to classify exposed vs. unexposed subjects based on parameters from a statistical face model, yielding an F1 score of over 0.91.
- We demonstrate a correlation between face shape and brain volume size for certain brain areas, enabling predictions better than mean predictions.
- We promote interpretability of our method by visualizing deformations and variation in our results on the face; relate those results to findings reported in the FASD literature.

2 Related Work

2.1 Prenatal Alcohol Exposure

PAE is common with a prevalence of 9.8% in pregnancies globally [30], and can cause FASD, which is estimated to be the leading cause of birth defects in Canada before the autism spectrum disorder, Down syndrome and other disorders [29]. To this day it is still often misdiagnosed [39], despite it being mentioned already in 1968 by Lemoine et al. [21]. Since then, the influence of alcohol exposure during pregnancy has been studied with respect to social behavior [9,33], changes in brain development [18,20,22] and dysmorphology [2,6]. Research indicates that exposed subjects exhibit more anti-social behavior, major changes in the white matter, among other parts, and have distinct dysmorphology features, such as a smooth philtrum, shorter palpebral fissure or flat nasal bridge. Current diagnostic criteria for a FASD diagnoses require either maternal comfirmation of alcohol exposure or three detected facial dysmorphological features and in most countries even both. Therefore, we aim to identify more sensitive facial biomarkers based on dysmorphology to help with screening and to provide insight into current brain alterations.

2.2 Statistical Shape Modeling in Medicine

Statistical shape modeling has many applications in the medical context including segmentation [15], surgical planning [7,34] and shape analysis tasks [10,24,36,40]. In the context of PAE and FASD, shape analysis is particularly interesting due to the presence of characteristic dysmorphological features. Fang et al. [10] developed a classifier to identify fetal alcohol syndrome (FAS) subjects using mesh metrics, such as area, flatness and curvature, derived from registered template meshes of 3D scanned faces. Other works [24,36] use the shape model to either show correlated deformations or even establish associations between face shape and specific brain regions for FAS classification. We, on the other hand, use the shape vector of the model to classify the exposure status or regress brain region volume.

2.3 Explainable Shape Analyis

The field of shape analysis can be complex, and the neural networks and analyzing principal component analysis do not offer intuitive ways to interpret or understand them. To address this issue, Prieto et al. [31] incorporated explainable AI (XAI) approaches such as Grad-CAM [35] and RISE [28] into their shape analysis with convolutional neural networks to increase interpretability. Similarly, Biffi et al. [3] identify the most discriminative features for their segmentation and visualize the anatomical variability of the shape based on their ladder variational autoencoder. Using statistical shape models, on the other hand, can provide anatomical measurements directly with the right parameterization [5], increasing interpretability by showing changes instantly. Zieger et al. [41] use a hybrid approach: first, they employ a statistical shape model (SSM) to unify the input

data; then, they classify the condition based on the model, and finally, they visualize the importance, calculated with layer-wise relevance propagation [1] or attention maps, of the input on the SSM. Our pipeline follows this approach of highlighting the important regions on the mesh. However, we use the weights of the regression or classification of the input shape parameters as importance indicator and project it onto the mesh using the SSM's deformation field.

3 Method

We now describe our pipeline to classify PAE and predict brain region volumes from facial shape features in detail; see Fig. 1 for an overview. In Section 3.1, we explain the preprocessing of MRI data to ensure a high-quality image segmentation and subsequent face mesh extraction. We then review statistical shape models and methods to register them to the extracted meshes in Section 3.2. Finally, in Section 3.3, we explain the usage of the resulting shape parameters to classify the exposure status of subjects, estimate brain region volumes and identify correlations between facial regions and PAE using interpretability techniques.

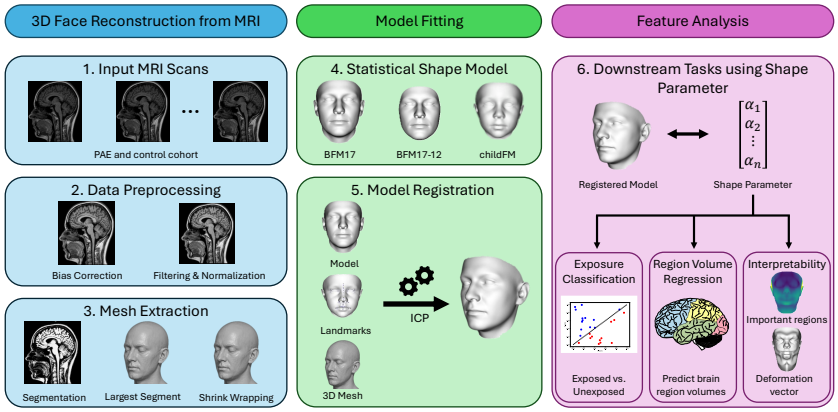

Fig. 1. Schematic overview of our pipeline. The three major stages, face reconstruction from MRI, shape model registration, and downstream analysis, are color-coded. We start by taking the input MRI scans (Step 1) and preprocessing them (Step 2), followed by the extraction of face meshes from the processed scans in Step 3. In Step 4, we show different statistical shape models, fitted to the extracted meshes in Step 5. Based on the low-dimensional shape parameters obtained from model fitting, we perform downstream tasks in Step 6, including exposure classification and regression of brain region volumes. We further apply interpretability techniques to highlight correlations on the registered meshes. All MRI scans and reconstructed MRI faces are representative image generated by OpenAI's ChatGPT 5.6 Luna image generation tool.

3.1 3D Face Extraction from MRI

The following two steps are performed in 3D Slicer [11].

Data Preprocessing. We follow a standard three-stage pipeline for preprocessing MRI scans to mitigate blurring and ghosting artifacts, ensuring as high-quality as possible face meshes for further processing. First, a bias field correction is performed using the N4ITK [37] module to correct non-uniform intensities across the MRI images caused by the scanner hardware. We then apply a gradient anisotropy diffusion filter to remove noise without smoothing sharp edges. The final step involves a global intensity normalization across all scans to account for variations in intensity levels between different scanners.

Mesh Extraction. We threshold the preprocessed MRI images to extract a solid face mesh using Otsu's method [25]. Meshes obtained from these segmentations may contain holes and small artifacts. To remove outliers, we retain only the segmented region with the largest enclosed area (via thresholding with an empirically determined threshold), discarding all other regions. We then employ the Wrap Solidify [38] module to automatically fill holes in segments and generate a 3D surface mesh via a shrink-wrapping approach.

3.2 Model Fitting

Having extracted face meshes at hand, we now fit a statistical face model to these meshes. This establishes dense correspondence across subjects, *meaningfully* fills in missing regions, reduces noise, and provides a compact parametrization of each mesh. Statistical shape models, such as the Basel Face Model [26], are learned using Principal Component Analysis (PCA) applied to a collection of face meshes in dense correspondence, and represent a face shape $\mathbf{S}_{\boldsymbol{\alpha}}$ as

$$\mathbf{S}_{\boldsymbol{\alpha}} = \boldsymbol{\mu} + \mathbf{U}\boldsymbol{\alpha}, \tag{1}$$

where $\boldsymbol{\mu}$ denotes the mean computed over all training shapes (the average face), and $\mathbf{U}$ contains the principal components scaled by their respective standard deviation. We use the shape parameters $\boldsymbol{\alpha}$ as a low-dimensional parametrization of a face mesh for subsequent downstream analyses.

To obtain the corresponding shape parameters for meshes extracted from the MRI scans, we fit a statistical shape model to each face mesh using a custom non-rigid Iterative Closest Point (ICP) algorithm implemented in Scalismo [14]. We first perform a rigid registration using Procrustes alignment on a set of corresponding landmarks defined on the model and each target face mesh; an overview of these landmarks is provided in the supplementary material. This initial alignment is followed by non-rigid ICP that incorporates the same landmark constraints. It iteratively computes a *posterior model* by interpreting the shape model as a Gaussian Process Morphable Model [23] and based on closest-point

correspondences established between model and target face mesh. We prune correspondences based on normal deviation and distance to account for missing facial regions, and implemented and tested the following two strategies to establish them:

- Fixed: correspondences are searched *once prior* to non-rigid alignment and fixed for all iterations of the non-rigid ICP algorithm.
- Dynamic: correspondences are established after every iteration (additionally, we lower the distance threshold by 50% every second iteration until a minimum value of 0.5mm is reached).

3.3 Feature Analysis

We investigate possible correlations between face shape and brain region volume and exposure status, which may act as non-invasive biomarkers, by using the shape parameters in regression and classification tasks, and to calculate deformation vectors.

Brain Region Volume Prediction. To predict the volume of a specific brain region, we perform linear regression using the extracted shape model parameters and ground truth volumina. To account for differences in age and overall size, we scale each subject's brain volume of interest (VOI) by their intracranial volume (ICV), $\overline{\mathrm{VOI}}_i = {}^{\mathrm{VOI}_i}/_{\mathrm{ICV}_i}$, and further scale it relative to the maximum $\overline{\mathrm{VOI}}$ observed within the cohort, $\hat{\mathrm{VOI}}_i = {}^{\overline{\mathrm{VOI}}_i}/_{\max\{\overline{\mathrm{VOI}}_0,\ldots,\overline{\mathrm{VOI}}_{m-1}\}}, \forall i \in \{0,\ldots,m-1\}$, where m is the number of scans. We employ a linear regression model of the form $\boldsymbol{y} = \mathbf{A}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{y} \in \mathbb{R}^m$ is a vector containing ground truth $\hat{\mathrm{VOI}}$s of the subjects, $\mathbf{A}$ represents the collection of the shape model parameters, $\boldsymbol{\beta}$ are the regression weights, and $\boldsymbol{\epsilon}$ is assumed (normally-distributed) observation noise. We compare different regularization strategies:

- Ordinary least squares (OLS): $\hat{\boldsymbol{\beta}} = \left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \boldsymbol{y}$
- Ridge regularization: $\hat{\boldsymbol{\beta}}_\lambda = \left(\mathbf{A}^\top \mathbf{A} + \lambda \mathbf{I}\right)^{-1} \mathbf{A}^\top \boldsymbol{y}$
- Elastic Net regularization: $\hat{\boldsymbol{\beta}}_{\lambda_1,\lambda_2} = \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \mathbf{A}\boldsymbol{\beta}\|^2 + \lambda_1 \|\boldsymbol{\beta}\|_1 + \lambda_2 \|\boldsymbol{\beta}\|^2$

Exposure Classification. We use binary logistic regression to classify the exposure status. Logistic regression models the probability of a subject being exposed to alcohol during pregnancy as

$$p(\boldsymbol{y} = 1) = \frac{1}{1 + e^{-\mathbf{A}\boldsymbol{\beta}}}, \tag{2}$$

where, this time, $\boldsymbol{y} \in \{0,1\}^m$ is a vector of ground truth binary labels, which is 1 if a subject has been exposed and 0 otherwise. $\boldsymbol{A}$ is defined as above. We solve for the coefficients $\boldsymbol{\beta}$ by minimizing the negative log likelihood.

Interpretability. To improve the interpretability of our pipeline, we employ two complementary approaches.

First, we quantify the importance of the deformation for the model's decision. Because the parameters of the statistical shape model influence the resulting facial deformation linearly, the optimized weights from the volume prediction and exposure classification reveal which specific deformations have the most significant impact the results. We compute the vertex importance vector $\boldsymbol{I} \in \mathbb{R}^v$, with v being the number of vertices of the face mesh, from the regression or classification weight vector $\boldsymbol{\beta}$ by projecting the weights onto the principal component matrix $\mathbf{U}$ and computing the L_1 norm of the resulting 3D deformation for each vertex: $\boldsymbol{I}_i = \left|\left|(\mathbf{U}\boldsymbol{\beta})_{i,:}\right|\right|_1, i = 0, ..., v-1$. The normalized importance is then visualized on the mean dataset face mesh using a colormap.

Second, we show the facial deformations, which are correlated to PAE or brain region volume differences using the method proposed by Blanz & Vetter [4]. The deformation vector $\Delta\mathbf{S}$ for the vertices is calculated like:

$$\Delta\mathbf{S} = \sum_{j=0}^{m-1} \gamma_j \left(\mathbf{S}_j - \boldsymbol{\mu}\right) \tag{3}$$

In this equation, we subtract the mean face $\boldsymbol{\mu}$ from the face of each subject and weight the difference with γ. For the PAE deformation vector γ is set to 1 for exposed subjects and -1 otherwise. For the brain region volume deformation vector, γ corresponds to the subject's brain volume normalized within the cohort. This deformation vector $\Delta\mathbf{S}$ can then be added to the mean face with a scaling factor to visualize the deformation associated with PAE or the brain region volume.

4 Evaluation

To evaluate our pipeline, we first introduce the datasets that are used during the evaluation in Section 4.2. In Section 4.3, we then evaluate the classification performance quantitatively and qualitatively. We identify which regions are important for the decision using the method described in Section 3.3, analyze them, and compare them to the observed deformations using the PAE deformation vector. To assess the regression for brain region volumes in Section 4.4, we quantitatively compare the results of our models trained on either the control, the exposed, or combined cohort to determine whether merging the datasets improves or impairs regression performance. Our evaluation was implemented with scikit-learn [27]. For all our evaluations, we split the data into 80% training and 20% evaluations sets based on subject ids in a seeded random manner to perform a five fold cross-validation, which we use to calculate the mean and standard deviation. Furthermore, the mean is removed and we scale the parameter to unit variance.

4.1 Statistical Shape Models

We used five distinct models during the evaluation and varied the maximal amount of used shape parameters $[20, 50, 200]$ for the regression and classification, which can be more than the available amount of parameters depending on the model. The first model is the Basel Face Model 2017 (BFM17) [13] with 199 shape parameters and a face mask incorporating the neck and ears. The second one is the Basel Face Model 2017 with the face12 mask (BFM17-12) with 199 shape parameters, which, does not include the neck and ear region. The remaining are custom models were created for this work using Scalismo. These include a child face model (childFM) with 72 shape parameters built from 31 child faces, from subjects between 2 and 8 years old from the Headspace dataset [8]. The next two are augmentations of BFM17, one (augBFM) with 255 shape parameters augmenting BFM17 with additional model flexibility and symmetric bias for the deformation field [23]. The other is symBFM with 199 shape parameters; it is a symmetrized variant of the BFM17 by sampling 750 faces from it, which were symmetrized and used to build the new model.

4.2 Datasets

For the following evaluations, we use two datasets with similar demographics and collection methodologies. The first dataset, the Calgary preschool MRI dataset, was collected by Reynolds et al. [32] and is used as control. The second dataset, from the paper by Kar et al. [17,18], only contains subjects with confirmed PAE.

Calgary preschool MRI Dataset. The dataset consists of a variety of MRI scans of 126 children and we use the T1-weighted image. These scans were collected as part of a longer MRI protocol at the Alberta Children's Hospital in Calgary, Canada. All children were free from genetic disorders and significant intellectual or motor impairments and were born $\geq$35 weeks' gestation [32]. From the dataset we reconstructed all faces from the 279 T1-weighted MRI scans using the methodology described in Section 3.1. Furthermore, we have the corresponding volume information for 116 brain areas for 240 of the reconstructed scans. These 240 scans were collected from 90 of the 126 children, with up to eight scans taken for an individual at intervals of half a year to a year. The subjects are aged $1.97 - 6.9$ years (4.46 ± 1 years), with 133 male and 107 female scans (48 male and 42 female subjects, respectively). Parent/guardian written informed consent and child verbal assent were obtained for each subject. The University of Calgary Conjoint Health Research Ethics Board (CHREB) approved this study (REB13-020).

Prenatal Alcohol Exposure Dataset. The dataset consists of a cohort of 57 subjects with confirmed PAE, but without a diagnosis of FASD. Children with attention-deficit hyperactivity disorder, learning disabilities, language delays, and/or mental health diagnoses were included, as these frequently co-occur with

PAE [18]. Of the 57 children, 54, aged $2.49 - 6.97$ years (5.21 ± 1.11 years; 27 males/27 females) underwent a total of 116 MRI scans. For 75 of these scans is the corresponding brain volume information available. Scanning was performed approximately every six months. Parent/guardian written informed consent and child verbal assent were obtained for each subject. The University of Calgary CHREB approved this study (REB14-2266).

4.3 PAE Classification

Quantitative Evaluation. For the of PAE classifcation based on the shape parameters, we keep the class balance between the classes using stratified splits. As described in Section 3.3, we use logistic regression and optimized the weights according to negative log likelihood loss. We then evaluate the performance on both datasets using the F1 score. We also performed a hyperparameter search to improve the results with a search space for the regularization penalty $[OLS, Ridge, ElasticNet]$, the maximum optimization iterations $[1000, 5000, 10000]$, the inverse regularization strength C $\left[10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^{2}\right]$, and the Elastic Net mixing parameter ρ $[0.1, 0.5, 0.9]$. These are used in the regularization part of scikit-learn's [27] logistic regression cost function $r(\boldsymbol{\beta})/SC$, where $S = \sum_{i=0}^{m-1} y_i$ and $r(\boldsymbol{\beta})$ is 0 for OLS, $||\boldsymbol{\beta}||^2/2$ for $Rigde$ and $1-\rho/2||\boldsymbol{\beta}||^2 + \rho||\boldsymbol{\beta}||_1$ for $ElasticNet$. For the following comparisons, we use the best result of the hyperparameter search.

Figure 2 shows the results of the different statistical shape models and the registration methods, from Section 3.2, across the sex-based splits. We observe that the classification with symBFM with fixed registration yields the best results regardless of the parameter count or sex of the subjects. We expected dynamic correspondences perform better that fixed correspondences, but the resampling can lead to anatomically invalid correspondences in regions where the neck and chin are absent from the MRI scan due to the limited field of view, resulting in truncated anatomical structures. Overall, the classification of PAE when using only female subjects performs the best, reaching an F1 score of 0.96 ± 0.06 using up to 200 parameters with Elastic Net regularization. In contrast, the performance of the classification is the lowest for male subjects only, with its best F1 score being 0.87 ± 0.13 using up to 200 parameters with Elastic Net regularization. The classification for the full cohort was the only one that performed better when not using all the available parameters. Instead, the best F1 score was 0.94 ± 0.05 with 50 parameters with Elastic Net regularization. While performance generally improves as the number of parameters increase for the sex-specific groups, the opposite was observed for the full cohort, where the F1 score decreases to 0.92 ± 0.04 at higher parameter counts. Although sex-related differences are known to exist in brain development [22], there are no significant differences between the sexes when it comes to a diagnosis of FASD [12].

Qualitative Evaluation. Using the method to calculate the importance for each vertex described in Section 3.3, we visualize the importance based on the

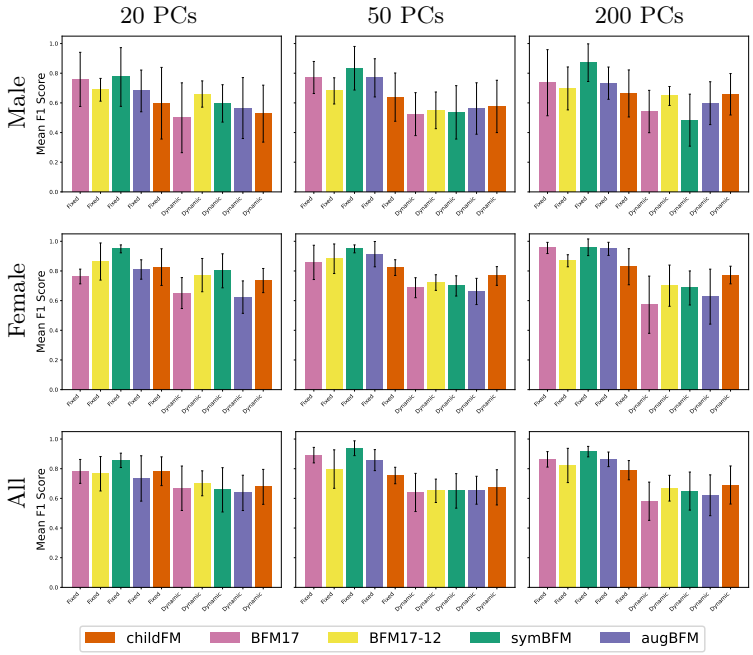

Fig. 2. Quantitative comparison of different setups of our pipeline. The columns show up to how many parameters are used for the classification and the rows show which subjects were used for the classification. On the x-axis of each graph are the different versions with different models and registration methods. On the y-axis is the mean F1 score over the cross validation for the best hyperparameter setup with the corresponding bar.

classification weights for the three sex-based splits. The averaged importance across the cross-validation for the best-performing classification with up to 200 parameters for each split is visualized in Figure 3. There are clear hotspots of importance across all splits, located between the ear and the temple. Those are most likely due to a combination of overfitting, despite regularization, pressure indentations around the ears present in many MRI scans as well as possible differences in during acquisition of the datasets. These indentations come either from the foam cushions used to keep the head still or from the headphones that were provided for watching movies during the scan. During registration, these pressure indentations, which could have systemic differences based on the exact setup, impact the resulting mesh, thereby serving as indicator of the different datasets. Another artifact is at the neck, where the ICP algorithm is less constrained, because the area below the chin is not reconstructible for all scans and the ICP can deform the model more freely in this region. This can aid the classification, because more chins are missing in the control dataset in terms of percentage. Nevertheless, other regions can be linked to known dysmorphologies in the literature [16]. A flat midface is part of the FASD facial phenotype and the

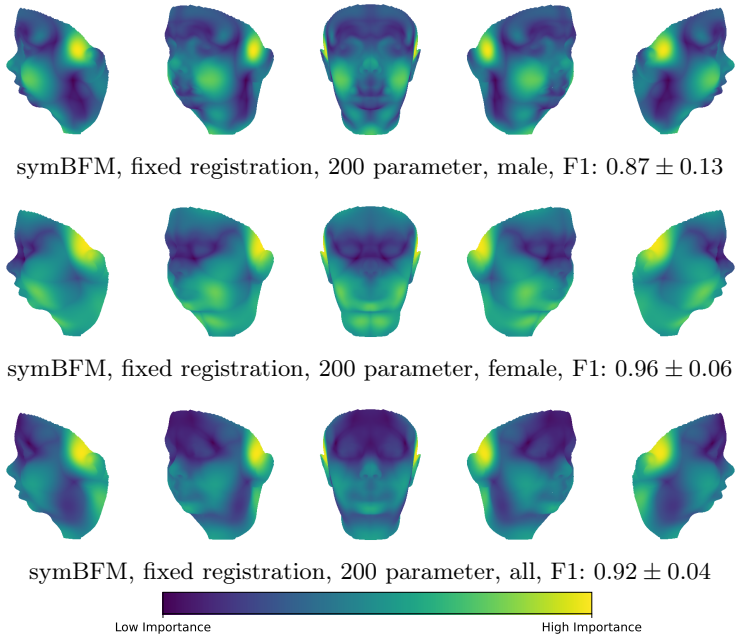

Fig. 3. Importance comparison for PAE classification separated by sex. Each row shows the averaged importance across the cross-validation for the different sex splits of the symBFM model (best performing configuration). In the top row the importance using only the male subjects resulting in an F1-score of 0.87 ± 0.13 is displayed. In the middle row illustrates the importance using only the female subjects resulting in an F1-score of 0.96 ± 0.06 and in the bottom shows the importance using all subjects with an F1-score of 0.92 ± 0.04. The same face is shown in each row five times from different angles.

midface shows moderate importance in all three splits. Another notable phenotype is having a short nose and low nasal bridge, and the nasal region is highlighted in the male and full cohort splits. The male split also shows another region around the eyes that can be linked to the variations in the epicanthal folds. Meanwhile, the other two splits highlight the general mouth region which includes phenotypes such as smooth philtrum, thin upper lip and prognathism [16]. However, they do not highlight the regions separately.

To eliminate the influence of the hotspots in the temple region, we created an additional model with the fitted faces of the best classification, symBFM, and applied a mask only retaining the broader eye, nasal, and mouth regions (Figure 4). Except for the female split, the F1 score decreased relative to the unmasked model and the optimal number of parameters is lower. However, we do not have the importance hotspots at the regions with known artifacts; instead, the importance is mainly on the eyes, nose tip and lips. Nevertheless, the highlights

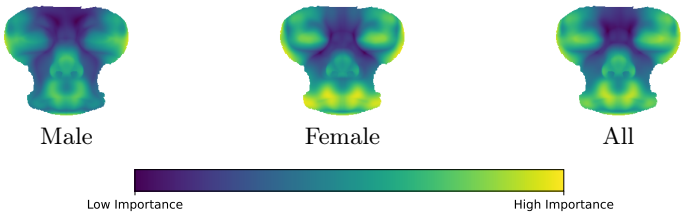

Fig. 4. Importance comparison for masked model of classifying PAE using the sexes separately. We show the importance of a masked version of symBFM with fixed registration all with 50 parameter with Elastice Net regularization. In the first column is the classification for the male split with an F1 score of 0.79 ± 0.24. For the female split the classification results is an F1 score of 0.97 ± 0.03 and for the full cohort the F1 score is 0.87 ± 0.07.

on the upper lip and corners of the mouth might come from overfitting resulting from the presence of landmarks there.

When examining the faces created using the PAE deformation vector, described in Section 3.3, the most notable deformations are observed at the cheeks, chin and neck (see Figure 5). The exposed face is more rotund with a small chin and receding neck. The unexposed face on the other hand has a wider chin, thinner face and protruding neck. By comparing these areas to the highlighted regions from the importance visualization, we can identify similar regions of deformation. Especially, the midface and cheek regions' deformations match the regions of importance from the classification. However, the unexposed face demonstrates more signs of midface hypoplasia with cheek and upper jaw region sitting further posterior than in the exposed face. This is the opposite of what we expected based on the literature [16], which could be due to a bias in the dataset caused by the small number of subjects. It is also noteworthy that the region of highest importance, between the ear and temple, has no visible deformation, also suggesting that the hotspots are likely due to overfitting.

4.4 Brain Region Volume Prediction

Quantitative Evaluation. We compare the three regularization methods of linear regression to predict the VOI and different registration methods, mentioned in Section 3, according to the mean squared error. For the hyperparameter search we had the search space for only positive coefficients β $[true, false]$, convergence tolerance $\left[10^{-5}, 10^{-6}\right]$. For all regularizations, the regularization strength α $[0.01, 0.1, 1, 10]$ and additionally for Elastic Net regression is the mixing parameter ρ $[0.2, 0.5, 0.8]$. These are use for scikit-learn's [27] rigde regression, $||\boldsymbol{y} - \mathbf{A}\boldsymbol{\beta}||^2 + \alpha||\boldsymbol{\beta}||^2$, and Elastic Net regression, $1/2m||\boldsymbol{y} - \mathbf{A}\boldsymbol{\beta}||^2 + \alpha\rho||\boldsymbol{\beta}||_1 + 0.5\alpha\,(1-\rho)\,||\boldsymbol{\beta}||^2$, objective functions. We first compare the different regularizations with the mean baseline, which predicts a constant mean value of the training target data, and notice in Figure 6 that only using Elastic Net regularization is always at least as

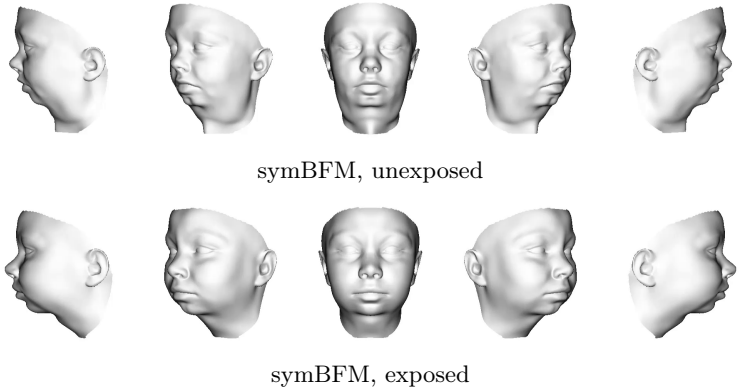

Fig. 5. PAE face deformation showing the unexposed and exposed face. The top row shows the unexposed face calculated according to the PAE deformation vector and the bottom row shows the exposed face. The same face is shown in each row five times from different angles.

good as the baseline. This holds true regardless of model, registration method, used parameters, exposure and sex split. Therefore we only use Elastic Net regularization for the remaining evaluations. Furthermore, the regression with BFM17-12 with dynamic registration with 20 parameters is evaluated, because it is the best performing model combination across exposure and sex for most brain regions.

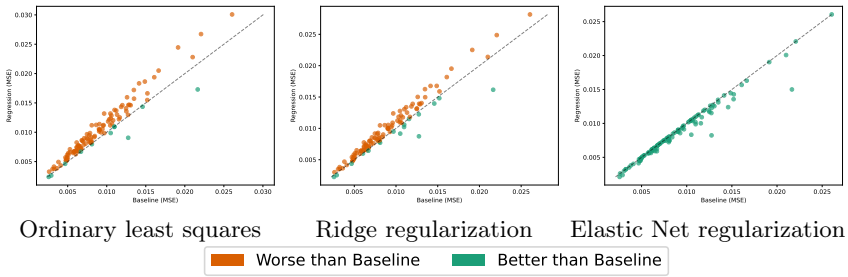

Fig. 6. Scatterplots of the three regularizations comparing the shape parameter regression with the mean baseline for the combined dataset. The figure shows the regression results for childFM using up tp 200 parameters with dynamic registration on the full cohort. Each dot in the scatterplot is one brain region and is placed according to the mean baseline and the regression MSE along the x-and y-axis respectively. The diagonal lines show where the error is same.

In addition to the mean baseline, we also use an ICV baseline, that predicts the region's volume based only on the ICV, and both are used to identify brain

regions that correlate with face shape and to determine how PAE affects these correlations. We state the number of of brain regions which were better predicted using the parameter regression than the two baselines in Table 1, showing a higher correlation between face shape and brain region than between ICV and brain region. For the full cohort, the regression outperformed both the mean baseline and the ICV baseline in 70 of the 116 brain regions. We also can see that the number is generally higher on the female split than on the male split similar to the classification, but using both sexes outperforms both splits separately for the control and combined dataset. For the combined dataset it is notable that less brain regions perform better than the baselines on the female and male splits separately than using the full cohort. For 26 out of the 70 brain regions the

Table 1. Comparison of number of brain regions where the parameter regression is better than the mean baseline and the regression based on the ICV baseline for the Elastic Net regression based on parameters of the BFM17-12 model with dynamic registration with 20 parameters . We compare across the exposure and sex splits for the 116 brain regions using the mean MSE across the cross validation.

	Exposed	Control	Combined
Male	34	50	48
Female	46	61	59
All	26	62	**70**

regression error is smaller using the combined dataset than it is for the datasets separately, meaning that a better overarching correlation was found than the separate correlations. This number goes down to 16 and 11 for the female and male split respectively, meaning there are correlations across the sexes which help with the regression. Out of the 26 regions regions like the right frontal pole, right putamen, right middle cingulate gyrus, or left planum temporale are especially interesting, because there is a general agreement that they are affected by PAE [20]. However, regions like the accumbens area or right middle occipital gyrus are also part of this subset. These regions are known to have minimal to no correlation [20], meaning that the separate exposure splits augment each other to perform better regardless of exposure.

Qualitative Evaluation. We compare the deformation vectors of brain regions to identify similarities and differences to determine the impact of PAE. Specifically, we show the deformation of faces according to the deformation vector of the right accumbens area and right middle cingulate gyrus in Figure 7. Both brain regions are in the subset of the regions which performed better in the parameter regression with the combined dataset, but we chose one region with consensus of PAE correlation, the right middle cingulate gyrus, and one without, the right accumbens area [20].

The figure shows no discernible difference, if the scaling factor is within the regular range $(0-1)$. Meaning the observable deformation according to the brain

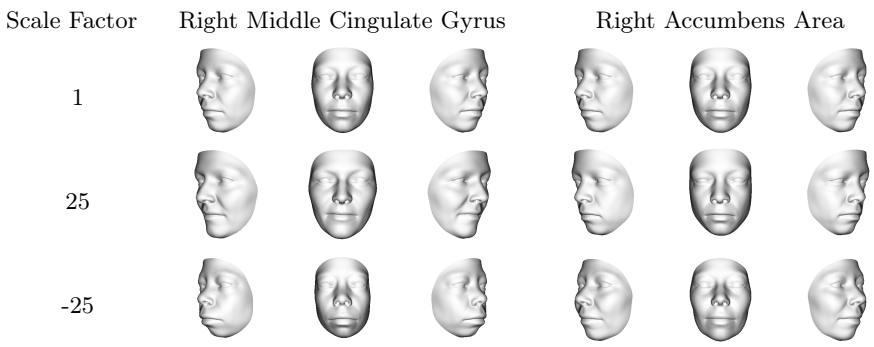

Fig. 7. Brain region deformation showing the faces according to the deformation correlated to the right middle cingulate gyrus size in the middle columns and right area accumbens in the right column from different angles. Each row demonstrates a different scaling factor for the deformation vectors. We also selected values outside the regular range to accentuate the differences.

region volume change is very subtle. Setting the scale factor to ± 25, we see deformation mainly in the chin, nose and forehead area. For the right middle cingulate gyrus the deformations are overall more pronounced than for the right area accumbens. The deformation vector resembles PAE dysmorphologies [16] with a protruding chin, full midface and a longer and down-turned nasal tip. Using a negative scale factor visualizes dysmorphologies, such as prognathism, hypoplastic midface and anteverted nares. The deformation for the right accumbens area on the other hand is small compared to the right middle cingulate gyrus. Its major deformations are at the forehead, which is not in the list of dysmorphologies for FASD diagnosis [16], and subtle changes in the midface and nose.

5 Conclusion

In this paper, we explored the influence of PAE on face shape and brain development. For this purpose, we presented a pipeline which first extracts a face mesh from MRI scans, registers a statistical shape model to those meshes and uses the shape parameters for machine learning. We showed that the approach of using the shape parameters of statistical shape models is capable of assessing the presence of PAE in children, with the best-performing model combination achieving an average F1 score greater than 0.91 for the entire cohort across the cross-validation. We applied interpretability techniques to highlight the important areas for the decision process and noticed that areas which correspond to known dysmorphologies from the literature are important, such as the midface and nasal area. However, the main influence stems from the area between the temple and ear, which can be attributed to systematic artifacts in the datasets. To remove

this influence we used a masked version only consisting of the eye, nasal and mouth region. The classification still results in a F1 score of over 0.86 without clear important highlights due to artifacts. We also explored the ability of linear regression to predict brain region volumes based on model parameters. During our regression tests of brain region volumes, we found that the regression of regions such as the right frontal pole or right middle cingulate gyrus, which are known to have a correlation with PAE, yielded better results using the combined dataset than using the control and PAE dataset separately. This suggests an overarching correlation. However, we also found regions, such as the accumbens area and right middle occipital gyrus, which have better results using the full cohort, that are not known to be clearly correlated with PAE. Additionally, we visually inspected the face deformation for PAE and specific brain region volumes calculated directly from the dataset and found the deformation for the right middle cingulate gyrus shares similarities to FASD dysmorphologies, such as prognathism, nasal and midface deformations.

The major limitation of this work is the presence of artifacts due to missing facial regions and possible systematic differences during the scan, which are exploited by classification and regression. However, we showed that our approach is valid nevertheless using explainability to link the decision process to known deformation in the literature. Another limitation is the limited dataset used; using another dataset with a different modality like RGB for the mesh extraction could also make this approach more accessible without the requirement of a MRI scanner and long capturing sessions.

Acknowledgments. We would like to thank **Paul Himmler** for the support during the creation of the pipeline and paper.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.R., Samek, W.: On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLOS ONE **10**(7), 1–46 (07 2015). https://doi.org/10.1371/journal.pone.0130140, https://doi.org/10.1371/journal.pone.0130140
2. Bandoli, G., Jones, K., Wertelecki, W., Yevtushok, L., Zymak-Zakutnya, N., Granovska, I., Plotka, L., Chambers, C., CIFASD: Patterns of prenatal alcohol exposure and alcohol-related dysmorphic features. Alcohol. Clin. Exp. Res. **44**(10), 2045–2052 (Oct 2020)
3. Biffi, C., Cerrolaza, J.J., Tarroni, G., Bai, W., de Marvao, A., Oktay, O., Ledig, C., Le Folgoc, L., Kamnitsas, K., Doumou, G., Duan, J., Prasad, S.K., Cook, S.A., O'Regan, D.P., Rueckert, D.: Explainable anatomical shape analysis through deep hierarchical generative models. IEEE Trans. Med. Imaging **39**(6), 2088–2099 (Jun 2020)

4. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. p. 187–194. SIGGRAPH '99, ACM Press/Addison-Wesley Publishing Co., USA (1999). https://doi.org/10.1145/311535.311556, https://doi.org/10.1145/311535.311556
5. Boutillon, A., Salhi, A., Burdin, V., Borotikar, B.: Anatomically parameterized statistical shape model: Explaining morphometry through statistical learning. IEEE Transactions on Biomedical Engineering **69**(9), 2733–2744 (2022). https://doi.org/10.1109/TBME.2022.3152833
6. Breuer, L., Greenmyer, J.R., Wilson, T.: Clinical diagnosis and management of fetal alcohol spectrum disorder and sensory processing disorder in children. Children **11**(1) (2024). https://doi.org/10.3390/children11010108, https://www.mdpi.com/2227-9067/11/1/108
7. Cai, X., Wu, Y., Huang, J., Wang, L., Xu, Y., Lu, S.: Application of statistical shape models in orthopedics: a narrative review. Intelligent Medicine **4**(4), 249–255 (2024). https://doi.org/https://doi.org/10.1016/j.imed.2024.05.001, https://www.sciencedirect.com/science/article/pii/S2667102624000573
8. Dai, H., Pears, N., Smith, W., Duncan, C.: Statistical modeling of craniofacial shape and texture. International Journal of Computer Vision **128**(2), 547–571 (Feb 2020). https://doi.org/10.1007/s11263-019-01260-7, https://doi.org/10.1007/s11263-019-01260-7
9. de Water, E., Rockhold, M.N., Roediger, D.J., Krueger, A.M., Mueller, B.A., Boys, C.J., Schumacher, M.J., Mattson, S.N., Jones, K.L., Lim, K.O., Wozniak, J.R.: Social behaviors and gray matter volumes of brain areas supporting social cognition in children and adolescents with prenatal alcohol exposure. Brain Research **1761**, 147388 (2021). https://doi.org/https://doi.org/10.1016/j.brainres.2021.147388, https://www.sciencedirect.com/science/article/pii/S000689932100113X
10. Fang, S., McLaughlin, J., Fang, J., Huang, J., Autti-Rämö, I., Fagerlund, Å., Jacobson, S., Robinson, L., Hoyme, H., Mattson, S., Riley, E., Zhou, F., Ward, R., Moore, E., Foroud, T., on Fetal Alcohol Spectrum Disorders, C.I.: Automated diagnosis of fetal alcohol syndrome using 3D facial image analysis. Orthodontics & Craniofacial Research **11**(3), 162–171 (2008). https://doi.org/https://doi.org/10.1111/j.1601-6343.2008.00425.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1601-6343.2008.00425.x
11. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J.V., Pieper, S., Kikinis, R.: 3D slicer as an image computing platform for the quantitative imaging network. Magnetic Resonance Imaging **30**(9), 1323–1341 (2012). https://doi.org/https://doi.org/10.1016/j.mri.2012.05.001, https://www.sciencedirect.com/science/article/pii/S0730725X12001816, quantitative Imaging in Cancer
12. Flannigan, K., Poole, N., Cook, J., Unsworth, K.: Sex-related differences among individuals assessed for fetal alcohol spectrum disorder in canada. Alcohol, Clinical and Experimental Research **47**(3), 613–623 (2023). https://doi.org/https://doi.org/10.1111/acer.15017, https://onlinelibrary.wiley.com/doi/abs/10.1111/acer.15017
13. Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Luthi, M., Schoenborn, S., Vetter, T.: Morphable face models - an open framework. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). pp. 75–82 (2018). https://doi.org/10.1109/FG.2018.00021

14. Graphics and Vision Research Group, University of Basel: Scalismo - scalable image analysis and shape modelling. https://scalismo.org/, accessed 2026-06-11
15. Heimann, T., Meinzer, H.P.: Statistical shape models for 3d medical image segmentation: A review. Medical Image Analysis **13**(4), 543–563 (2009). https://doi.org/https://doi.org/10.1016/j.media.2009.05.004, https://www.sciencedirect.com/science/article/pii/S1361841509000425
16. Hoyme, H.E., Kalberg, W.O., Elliott, A.J., Blankenship, J., Buckley, D., Marais, A.S., Manning, M.A., Robinson, L.K., Adam, M.P., Abdul-Rahman, O., Jewett, T., Coles, C.D., Chambers, C., Jones, K.L., Adnams, C.M., Shah, P.E., Riley, E.P., Charness, M.E., Warren, K.R., May, P.A.: Updated clinical guidelines for diagnosing fetal alcohol spectrum disorders. Pediatrics **138**(2), e20154256 (08 2016). https://doi.org/10.1542/peds.2015-4256, https://doi.org/10.1542/peds.2015-4256
17. Kar, P., Reynolds, J.E., Gibbard, W.B., McMorris, C., Tortorelli, C., Lebel, C.: Trajectories of brain white matter development in young children with prenatal alcohol exposure. Human Brain Mapping **43**(13), 4145–4157 (2022). https://doi.org/https://doi.org/10.1002/hbm.25944, https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.25944
18. Kar, P., Reynolds, J.E., Grohs, M.N., Gibbard, W.B., McMorris, C., Tortorelli, C., Lebel, C.: White matter alterations in young children with prenatal alcohol exposure. Developmental Neurobiology **81**(4), 400–410 (2021). https://doi.org/https://doi.org/10.1002/dneu.22821, https://onlinelibrary.wiley.com/doi/abs/10.1002/dneu.22821
19. Kesmodel, U., Wisborg, K., Olsen, S.F., Henriksen, T.B., Secher, N.J.: Moderate alcohol intake during pregnancy and the risk of stillbirth and death in the first year of life. American Journal of Epidemiology **155**(4), 305 – 312 (2002). https://doi.org/10.1093/aje/155.4.305, https://www.scopus.com/inward/record.uri?eid=2-s2.0-0037083266&doi=10.1093%2faje%2f155.4.305&partnerID=40&md5=d1f5e7e1987e0cfa50138471ba2d40f5, cited by: 200; All Open Access, Bronze Open Access
20. Lebel, C., Roussotte, F., Sowell, E.R.: Imaging the impact of prenatal alcohol exposure on the structure of the developing human brain. Neuropsychol. Rev. **21**(2), 102–118 (Jun 2011)
21. Lemoine, P., Harousseau, H., Borteyru, J.P., Menuet, J.C.: Les enfants de parents alcooliques. anomalies observées. In: Ouest-Médical. vol. 21, pp. 476–482 (1968)
22. Long, M., Kar, P., Forkert, N.D., Landman, B.A., Gibbard, W.B., Tortorelli, C., McMorris, C.A., Huo, Y., Lebel, C.A.: Sex and age effects on gray matter volume trajectories in young children with prenatal alcohol exposure. Frontiers in Human Neuroscience **Volume 18 - 2024** (2024). https://doi.org/10.3389/fnhum.2024.1379959, https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2024.1379959
23. Lüthi, M., Jud, C., Gerig, T., Vetter, T.: Gaussian process morphable models. TPAMI **40** (2018)
24. Muggli, E., Matthews, H., Suttie, M., Halliday, J., Penington, A., Elliott, E.J., Thompson, D., Spittle, A., Hearps, S., Anderson, P.J., Claes, P.: Low to moderate prenatal alcohol exposure and facial shape of children at age 6 to 8 years. JAMA Pediatrics **179**(4), 438–446 (04 2025). https://doi.org/10.1001/jamapediatrics.2024.6151, https://doi.org/10.1001/jamapediatrics.2024.6151
25. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics **9**(1), 62–66 (Jan 1979). https://doi.org/10.1109/TSMC.1979.4310076

26. Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3D face model for pose and illumination invariant face recognition. In: 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance. pp. 296–301 (2009). https://doi.org/10.1109/AVSS.2009.58
27. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
28. Petsiuk, V., Das, A., Saenko, K.: RISE: randomized input sampling for explanation of black-box models. In: British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018. p. 151. BMVA Press (2018), http://bmvc2018.org/contents/papers/1064.pdf
29. Popova, S., Lange, S., Poznyak, V., Chudley, A.E., Shield, K.D., Reynolds, J.N., Murray, M., Rehm, J.: Population-based prevalence of fetal alcohol spectrum disorder in canada. BMC Public Health **19**(1), 845 (Jun 2019). https://doi.org/10.1186/s12889-019-7213-3, https://doi.org/10.1186/s12889-019-7213-3
30. Popova, S., Lange, S., Probst, C., Gmel, G., Rehm, J.: Estimation of national, regional, and global prevalence of alcohol use during pregnancy and fetal alcohol syndrome: a systematic review and meta-analysis. The Lancet Global Health **5**(3), e290–e299 (2017). https://doi.org/https://doi.org/10.1016/S2214-109X(17)30021-9, https://www.sciencedirect.com/science/article/pii/S2214109X17300219
31. Prieto, J.C., Miranda, F., Gurgel, M., Anchling, L., Hutin, N., Barone, S., Al Turkestani, N., Aliaga, A., Yatabe, M., Bianchi, J., Cevidanes, L.: ShapeAXI: Shape analysis explainability and interpretability. Proc. SPIE **12931**, 1293116 (Feb 2024)
32. Reynolds, J.E., Long, X., Paniukov, D., Bagshawe, M., Lebel, C.: Calgary preschool magnetic resonance imaging (MRI) dataset. Data in Brief **29**, 105224 (2020). https://doi.org/https://doi.org/10.1016/j.dib.2020.105224, https://www.sciencedirect.com/science/article/pii/S2352340920301189
33. Rockhold, M.N., Krueger, A.M., de Water, E., Lindgren, C.W., Sandness, K.E., Eckerle, J.K., Schumacher, M.J., Fink, B.A., Boys, C.J., Carlson, S.M., Fuglestad, A.J., Mattson, S.N., Jones, K.L., Riley, E.P., Wozniak, J.R.: Executive and social functioning across development in children and adolescents with prenatal alcohol exposure. Alcohol. Clin. Exp. Res. **45**(2), 457–469 (Feb 2021)
34. Rodriguez-Florez, N., Bruse, J.L., Borghi, A., Vercruysse, H., Ong, J., James, G., Pennec, X., Dunaway, D.J., Jeelani, N.U.O., Schievano, S.: Statistical shape modelling to aid surgical planning: associations between surgical parameters and head shapes following spring-assisted cranioplasty. Int. J. Comput. Assist. Radiol. Surg. **12**(10), 1739–1749 (Oct 2017)
35. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision **128**(2), 336–359 (Feb 2020). https://doi.org/10.1007/s11263-019-01228-7, https://doi.org/10.1007/s11263-019-01228-7
36. Suttie, M., Wozniak, J.R., Parnell, S.E., Wetherill, L., Mattson, S.N., Sowell, E.R., Kan, E., Riley, E.P., Jones, K.L., Coles, C., Foroud, T., Hammond, P., CIFASD: Combined face-brain morphology and associated neurocognitive correlates in fetal alcohol spectrum disorders. Alcohol. Clin. Exp. Res. **42**(9), 1769–1782 (Sep 2018)
37. Tustison, N., Gee, J.: N4ITK: Nick's N3 ITK implementation for MRI bias field correction. Insight J pp. 1–8 (01 2009). https://doi.org/10.54294/jculxw

38. Weidert, S., Andress, S., Linhart, C., Suero, E.M., Greiner, A., Böcker, W., Kammerlander, C., Becker, C.A.: 3D printing method for next-day acetabular fracture surgery using a surface filtering pipeline: feasibility and 1-year clinical results. International Journal of Computer Assisted Radiology and Surgery **15**(3), 565–575 (Mar 2020). https://doi.org/10.1007/s11548-019-02110-0, https://doi.org/10.1007/s11548-019-02110-0
39. Wozniak, J.R., Riley, E.P., Charness, M.E.: Clinical presentation, diagnosis, and management of fetal alcohol spectrum disorder. Lancet Neurol. **18**(8), 760–770 (Aug 2019)
40. Zhang, M., Golland, P.: Statistical shape analysis: From landmarks to diffeomorphisms. Medical Image Analysis **33**, 155–158 (2016). https://doi.org/https://doi.org/10.1016/j.media.2016.06.025, https://www.sciencedirect.com/science/article/pii/S1361841516301001, 20th anniversary of the Medical Image Analysis journal (MedIA)
41. Zieger, D., Henningson, J.O., Egger, B., Stamminger, M.: A unified pipeline for explainable gait analysis. In: Wachinger, C., Luijten, G., Elhabian, S., Gopinath, K., Egger, J. (eds.) Shape in Medical Imaging. pp. 29–43. Springer Nature Switzerland, Cham (2026)