Online Statistical Inference for Quantifying Mandible Growth in CT Images

Moo K. Chung[1,2], Ying Ji Chuang[2], Houri K. Vorperian[2]

[1]Department of Biostatistics and Medical Informatics
[2]Vocal Tract Development Laboratory, Waisman Center
University of Wisconsin-Madison, USA

mkchung@wisc.edu

Abstract. We present a unified online statistical framework for quantifying a collection of binary segmentation of mandibles in CT images. Since the segmentation was done semi-automatically, the processed binary images were available in a sequential manner. Thus, there was a need to develop an iterative analysis framework where the final statistical maps are updated sequentially based on an additional image. We present a method using a unified online algorithm for performing statistical inference. This proposed method is then applied to identify and characterize regions of mandible growth during the first two decades of life.

1 Introduction

The typical implementation of statistical inference in medical imaging requires that all the images are available in advance. That is the usual premise of existing medical image analysis tools such as ImageJ, SPM, SFL and AFNI. However, there are many situations where the entire imaging data set is not available and parts of imaging data is obtained in a timely sequential fashion. This is a common problems in medical imaging, where not every subject is scanned and processed at the same time.

When the image size itself is large, it may not be possible to fit all of the imaging data in a computer's memory for statistical analysis, making it necessary to perform the analysis by adding one image at a time in a sequential manner. In another situation, the imaging data set may be so large that it is not practical to use all the images in the dataset but use of a subset of the images. In these situations, we need to incrementally add stratified datasets one at a time to see if we are achieving reasonable statistical results. In all the above situations, we need a way to incrementally update the statistical analysis without repeatedly running the entire analysis whenever new images are added.

An *online algorithm* is one that processes its inputted data in a sequential manner [14]. Instead of processing the entire set of imaging data from the start, an online algorithm processes one image at a time. That way, we can bypass the memory requirement and reduce numerical instability. Online algorithms and

machine learning are both concerned with problems of making decisions about the present based on knowledge of the past [5]. Thus, online algorithms are often encountered in machine learning literature but there are somewhat limited number of studies in medical imaging.

Motivated by the concept of online algorithm in machine learning, we propose to advance online statistical inference procedures for the t- and F-tests. The online version of the F-test requires the development of the online version of multiple linear regression. The online methods are then used to characterize the mandible growth using 3D CT images.

2 Probabilistic model of binary segmentation

Let $p(x)$ be the probability of voxel x belonging to some region of interest (ROI) $\mathcal{M}$ such as mandible segmentation. Let $\mathbf{1}_{\mathcal{M}}$ be an indicator function defined as

$$\mathbf{1}_{\mathcal{M}}(x) = \begin{cases} 1 \text{ if } x \in \mathcal{M}, \\ 0 \text{ otherwise.} \end{cases}$$

We assume that the shape of ROI $\mathcal{M}$ is *random* (due to noise) and we associate it with probability $p(x)$:

$$P(x \in \mathcal{M}) = p(x), \quad P(x \notin \mathcal{M}) = 0.$$

Unless we use probabilistic segmentation techniques such as Gaussian mixture models [2,3], the probability is simply given as a Bernoulli distribution, i.e., 0 or 1. Then the volume of $\mathcal{M}$ is given by

$$vol(\mathcal{M}) = \int_{\mathbb{R}^3} \mathbf{1}_{\mathcal{M}}(x)\, dx.$$

Since the shape of $\mathcal{M}$ is random, the volume is also random. The mean volume of $\mathcal{M}$ is

$$\mathbb{E}\, vol(\mathcal{M}) = \int_{\mathbb{R}^3} \mathbb{E}\, \mathbf{1}_{\mathcal{M}}(x)\, dx = \int_{\mathbb{R}^3} p(x)\, dx.$$

The integral of the probability map can thus be used as an estimate for the volume of ROI. Unfortunately, mandible segmentations often have holes and cavities that have to be patched topologically for accurate volume estimation (Fig. 1). Such topological defects can be easily patched by Gaussian kernel smoothing (Fig. 2) without resorting to more complicated topology correction methods [11].

Consider a 3-dimensional Gaussian kernel

$$K(x) = \frac{1}{(2\pi)^{3/2}} \exp\Big(-\frac{\|x\|^2}{2}\Big),$$

where $\|\cdot\|$ is the Euclidean norm of $x \in \mathbb{R}^3$. The rescaled kernel K_t is defined as

$$K_t(x) = \frac{1}{t^3} K\big(\frac{x}{t}\big).$$

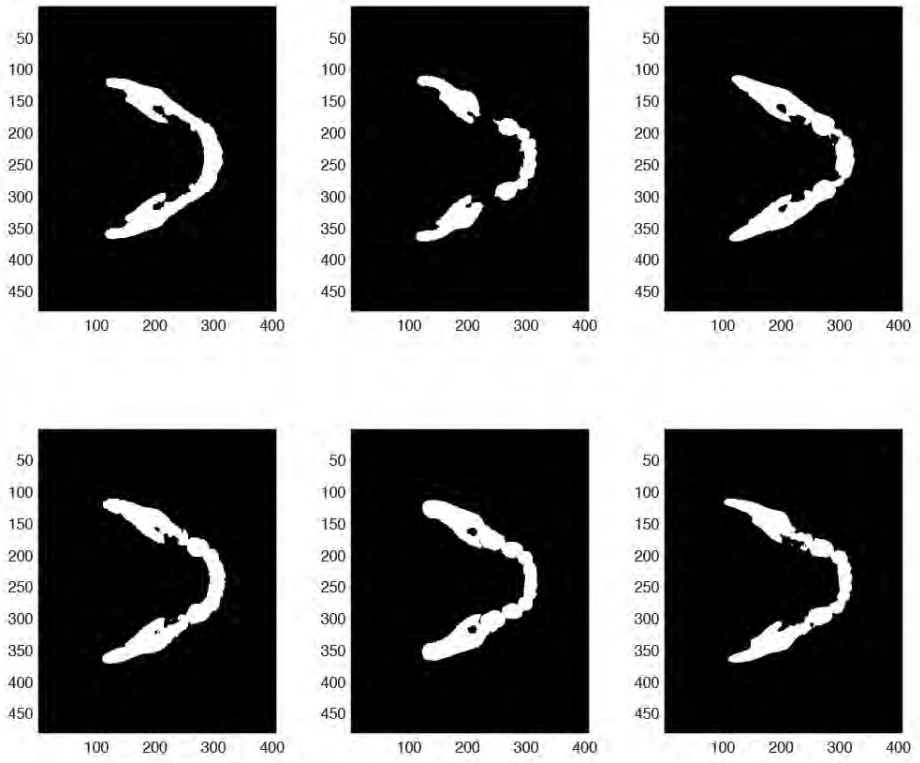

Fig. 1: Six representative mandible binary segmentations that are affine registered to the template space.

Then Gaussian kernel smoothing applied to the probability map $p(x)$ is given by

$$K_t * p(x) = \int_{\mathbb{R}^3} K_t(x-y)p(y)\ dy, \tag{1}$$

which is the scale-space representation of probability map $p(x)$ [17,18,19,29,25].

$$\begin{aligned}\int_{\mathbb{R}^3} K_t * p(x)\ dx &= \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} K_t(x-y)p(y)\ dy\ dx \\ &= \int_{\mathbb{R}^3} p(y)\ dy.\end{aligned}$$

Here, we used the fact that Gaussian kernel is a probability density, i.e.,

$$\int_{\mathbb{R}^3} K_t(x,y)\ dx = 1$$

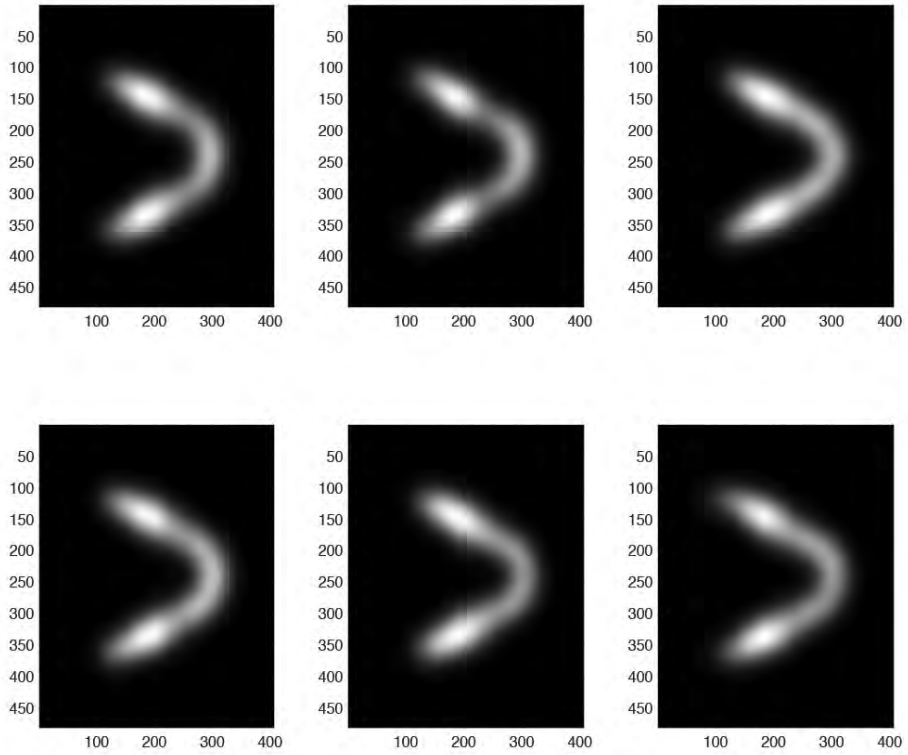

Fig. 2: Gaussian kernel smoothing of the six representative binary segmentations in Fig. 1 with bandwidth $\sigma = 20$. Smoothing can easily patches cavities and handles that were present in Fig. 1.

for any $y \in \mathbb{R}^3$. Thus, the volume estimate $\mathbb{E}\ vol(\mathcal{M})$ is invariant under different smoothing scales and we have

$$\mathbb{E}\ vol(\mathcal{M}) = \int K_t * p(x)\ dx. \tag{2}$$

The smoothed probability map $K_t * p(x)$ can be taken as a more robust probability map of whether a voxel belongs to a mandible and can be used as a response variable in modeling the growth of mandible (Fig. 2).

Note that

$$|K_t * p(x)| \leq \int_{\mathbb{R}^3} K_t(x-y)\Big[\sup_{y\in\mathbb{R}^3} p(y)\Big]\ dy = 1.$$

The smoothed probability map is still bounded between 0 and 1 so it is not exactly normally distributed. In order to set up most statistical routines such as t- or F -tests, which assumes a Gaussian noise model, it is necessary to make to

make the smoothed probability map more normal. One way of doing this is to apply the Fisher or logit transforms. However, since we performed the Gaussian kernel smoothing with reasonably large bandwidth, which tends to make the data more Gaussian due to the central limit theorem, it was not necessary to perform these transforms [2,20].

3 Online algorithm for *t*-test

By taking one smoothed image at a time as an input, we perform the incremental statistical inference. Smoothing tends to increase the robustness of statistical procedures.

Given images $x_1, \cdots, x_m$, an *online algorithm* for computing the sample mean image μ_m is given by

$$\mu_m = \frac{1}{m}\sum_{i=1}^{m} x_i = \mu_{m-1} + \frac{1}{m}(x_m - \mu_{m-1}) \tag{3}$$

for any $m \geq 1$. This algorithm avoids accumulating large sums and tend to be more numerically stable than the following algorithm [13]:

$$\mu_m = \frac{m-1}{m}\mu_{m-1} + \frac{x_m}{m}.$$

An online algorithm for computing the sample variance image σ_m^2 is algebraically involved [8,16]. Let

$$\sigma_m^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \mu_m)^2.$$

Then, it can be shown that

$$(m-1)\sigma_m^2 - (m-2)\sigma_{m-1}^2 = m(m-1)(\mu_m - \mu_{m-1})^2.$$

From (3), it can be iteratively written in terms of previous estimates $\mu_{m-1}, \sigma_{m-1}^2$ and new data x_m as

$$\sigma_m^2 = \frac{m-2}{m-1}\sigma_{m-1}^2 + \frac{1}{m}(x_m - \mu_{m-1})^2$$

for $m \geq 2$. The algorithm starts with the initial estimate $\sigma_1^2 = 0$. Figure 3 displays the result of mean and variance computation using the online algorithms.

For comparing a collection of images within a group, one-sample t-statistic is used. The segmentation probability $p(x)$ is smoothed using Gaussian kernel K_t and modeled as a Gaussian with mean $\mu(x)$ and variance $\sigma^2(x)$ at each voxel x. For one sample case, we are often testing

$$H_0(x) : \mu(x) = \mu_0 \text{ vs. } H_1(x) : \mu(x) > \mu_0 \tag{4}$$

for some predetermined value μ_0. The one-sample t-statistic is then

$$T_m(x) = \frac{\mu_m - \mu_0}{\sqrt{\sigma_m^2/m}} \tag{5}$$

at each voxel x. Subsequently, we compute T_m iteratively

$$T_1 \to \cdots \to T_m$$

in m steps.

For comparing a collection of images between groups, two-sample t-statistic is used. Given measurements $x_1, \cdots, x_m \sim N(\mu^1, (\sigma^1)^2)$ in one group and $y_1, \cdots, y_n \sim N(\mu^1, (\sigma^2)^2)$ in the other group, the two-sample t-statistic for testing

$$H_0 : \mu^1(x) = \mu^2(x) \;\; vs. \;\; H_1 : \mu^1(x) > \mu^2(x)$$

at each voxel x is given by

$$T_{m,n}(x) = \frac{\mu_m^1 - \mu_n^2 - (\mu^1 - \mu^2)}{\sqrt{(\sigma^1)_m^2/m + (\sigma^2)_n^2/n}}, \tag{6}$$

where μ_m^1, μ_n^2, $(\sigma^1)_m^2, (\sigma^2)_m^2$ are sample means and variances in each group respectively. Following the proposed online algorithm for computing means and variances iteratively, we compute $T_{m,n}$ iteratively

$$T_{1,0} \to T_{2,0} \to \cdots \to T_{m,0} \to T_{m,1} \to \cdots \to T_{m,n}$$

in $m + n$ steps.

4 Online algorithm for linear regression

In addition to the above described online algorithm for t-tests, it is possible to have an online algorithm for linear regression as described below. The online algorithm for linear regression is itself useful but additionally more useful in constructing an online algorithm for F-tests in the next section.

Given data vector $\mathbf{y}_{m-1} = (y_1, \cdots, y_{m-1})'$ and design matrix Z_{m-1}, consider linear model

$$\mathbf{y}_{m-1} = Z_{m-1}\boldsymbol{\lambda}_{m-1} \tag{7}$$

with unknown parameter vector $\boldsymbol{\lambda}_{m-1} = (\lambda_1, \lambda_2, \cdots, \lambda_k)'$. Z_{m-1} is a matrix of size $(m-1) \times k$. Multiplying Z'_{m-1} on the both sides we have

$$Z'_{m-1}\mathbf{y}_{m-1} = Z'_{m-1}Z_{m-1}\boldsymbol{\lambda}_{m-1} \tag{8}$$

Let $W_{m-1} = Z'_{m-1}Z_{m-1}$, which is a $k \times k$ matrix. In most applications, there are substantially more data than the number of parameters, i.e., $m \gg k$, and

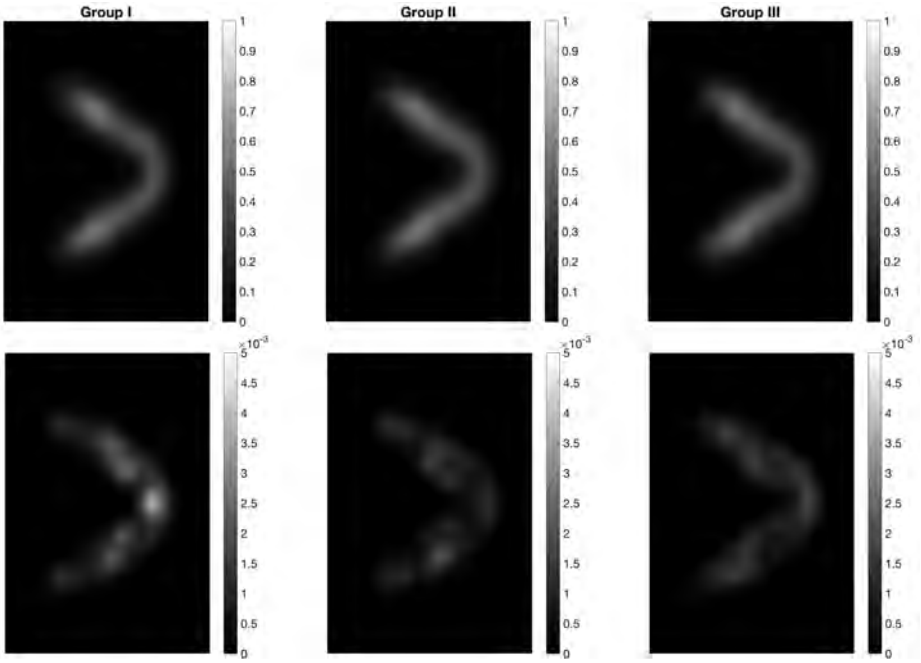

Fig. 3: The mean (top) and variance (bottom) of probability map in each age group. We are interested in localizing the regions of probability difference between the age groups.

W_{m-1} is invertible. Then the least squares estimation (LSE) of $\boldsymbol{\lambda}_{m-1}$ is given by

$$\boldsymbol{\lambda}_{m-1} = W_{m-1}^{-1} Z'_m \mathbf{y}_{m-1}.$$

When new data y_m is introduced to the linear model (7), the model is updated to

$$\begin{pmatrix} \mathbf{y}_{m-1} \\ y_m \end{pmatrix} = \begin{pmatrix} Z_{m-1} \\ z_m \end{pmatrix} \boldsymbol{\lambda}_m,$$

where z_m is $1 \times k$ row vector. Then multiplying the transpose of the design matrix, we have

$$\begin{aligned} (Z'_{m-1}\ z'_m) \begin{pmatrix} \mathbf{y}_{m-1} \\ y_m \end{pmatrix} &= (Z'_{m-1}\ z'_m) \begin{pmatrix} Z_{m-1} \\ z_m \end{pmatrix} \boldsymbol{\lambda}_m \\ Z'_{m-1}\mathbf{y}_{m-1} + z'_m y_m &= (W_{m-1} + z'_m z_m)\boldsymbol{\lambda}_m \\ W'_{m-1}\boldsymbol{\lambda}_{m-1} + z'_m y_m &= (W_{m-1} + z'_m z_m)\boldsymbol{\lambda}_m. \end{aligned}$$

Using Woodbury formula [12],

$$(W_{m-1} + z'_m z_m)^{-1} = W_{m-1}^{-1} - c_m W_{m-1}^{-1} z'_m,$$

where $c_m = 1/(1 + z_m W_{m-1} z_m')$ is scalar. Then we have the explicit online algorithm for updating the parameter vector:

$$\boldsymbol{\lambda}_m = (I - W_{m-1}^{-1} z_m' y_m - c_m W_{m-1}^{-1} z_m' W_{m-1}')\boldsymbol{\lambda}_{m-1} - c_m W_{m-1}^{-1} z_m' z_m' y_m, \quad (9)$$

where I is the identity matrix of size $k \times k$. Since the algorithm requires W_{m-1} to be invertible, the algorithm must start from

$$\boldsymbol{\lambda}_k \to \boldsymbol{\lambda}_{k+1} \to \cdots \to \boldsymbol{\lambda}_m.$$

At each iteration, we need to store $k \times k$ matrix W_{m-1}. In many applications, k will not be larger than 10 and most likely around 5 or less, which is manageable as far as computer memory is concerned.

5 Online algorithm for F-test

An online algorithm for the F-test is fairly involved but it is based on the online algorithm for linear regression. Let y_i be image intensity values of the i-th image and $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})'$ to be the variables of interest and $\mathbf{z}_i = (z_{i1}, \cdots, z_{ik})'$ to be nuisance variables corresponding to the i-th image. We assume there are $m-1$ images to start with. We are interested in testing the significance of the group variable while accounting for the effect of nuisance covariates. In a general setting, we have a general linear model [9].

$$\mathbf{y}_{m-1} = Z_{m-1}\boldsymbol{\lambda}_{m-1} + X_{m-1}\boldsymbol{\beta}_{m-1},$$

where $Z_{m-1} = (z_{ij})$ is $(m-1) \times k$ design matrix, $X_{m-1} = (x_{ij})$ is $(m-1) \times p$ design matrix. $\boldsymbol{\lambda}_{m-1} = (\lambda_1, \cdots, \lambda_k)'$ and $\boldsymbol{\beta}_{m-1} = (\beta_1, \cdots, \beta_p)'$ are unknown parameter vectors to be estimated at the $(m-1)$-th iteration. The significance of the variable of interests is determined by testing

$$H_0 : \boldsymbol{\beta} = 0 \text{ vs. } H_1 : \boldsymbol{\beta} \neq 0.$$

The fit of the reduced model under H_0, i.e.,

$$\mathbf{y}_{m-1} = Z_{m-1}\boldsymbol{\lambda}_{m-1}^0,$$

is measured by the sum of the squared errors (SSE):

$$\text{SSE}_{m-1}^0 = (\mathbf{y}_{m-1} - Z_{m-1}\boldsymbol{\lambda}_{m-1}^0)'(\mathbf{y}_{m-1} - Z_{m-1}\boldsymbol{\lambda}_{m-1}^0),$$

where the least squares estimate $\boldsymbol{\lambda}_{m-1}^0$ is iteratively estimated using the online algorithm (9). This provide the sequential update of SSE under H_0:

$$\text{SSE}_k^0 \to \text{SSE}_{k+1}^0 \to \cdots \to \text{SSE}_m^0.$$

Similarly the fit of the full model corresponding to $H_1 : \boldsymbol{\beta} \neq 0$ is measured by

$$\text{SSE}_{m-1}^1 = (\mathbf{y}_{m-1} - \mathbb{Z}_{m-1}\boldsymbol{\gamma}_{m-1}^1)'(\mathbf{y}_{m-1} - \mathbb{Z}_{m-1}\boldsymbol{\gamma}_{m-1}^1),$$

where $\mathbb{Z}_{m-1} = [Z_{m-1} X_{m-1}]$ is the combined design matrix of size $(m-1)\times(k+p)$ and

$$\boldsymbol{\gamma}^1_{m-1} = \begin{pmatrix} \boldsymbol{\lambda}^1_{m-1} \\ \boldsymbol{\beta}^1_{m-1} \end{pmatrix}$$

is the combined parameter vector of size $(k+p)\times 1$. Similarly using the online algorithm (9), SSE under H_1 is sequentially computed as

$$\mathrm{SSE}^1_{k+p} \to \mathrm{SSE}^1_{k+1} \to \cdots \to \mathrm{SSE}^1_m.$$

Note

$$\mathrm{SSE}^{m-1}_1 = \min_{\boldsymbol{\lambda},\boldsymbol{\beta}} \sum_{i=1}^{m-1} (y_i - \mathbf{z}_i\boldsymbol{\lambda} - \mathbf{x}_i\boldsymbol{\beta})^2 \le \min_{\boldsymbol{\lambda}} \sum_{i=1}^{m-1} (y_i - \mathbf{z}_i\boldsymbol{\lambda})^2 = \mathrm{SSE}^{m-1}_0$$

and thus, $\mathrm{SSE}_0 - \mathrm{SSE}_1 \ge 0$. Larger the value of $\mathrm{SSE}_0 - \mathrm{SSE}_1$ is, more significant the contribution of the coefficients $\boldsymbol{\beta}$ is. Under H_0, the test statistic at the m-th iteration F_m is given by

$$F_m = \frac{(\mathrm{SSE}_0 - \mathrm{SSE}_1)/p}{\mathrm{SSE}_0/(m-p-k)} \sim F_{p,m-p-k}. \tag{10}$$

Which is the F-statistic with p and $m-p-k$ degrees of freedom.

6 Random field theory

Since the t-statistic maps $T(x)$ in (5) and (6) are correlated over x, it is necessary to correct multiple comparisons using the random field theory [27,10]. Consider hypotheses,

$$J_0 : \mu(x) = \mu_0 \text{ for all } x \quad \text{vs.} \quad J_1 : \mu(x) > \mu_0 \text{ for some } x. \tag{11}$$

Note that the hypotheses (4) are at each fixed voxel x, while the hypotheses (11) are for over all possible voxels in the image. Therefore, (11) is the multiple comparisons version of (4). We reject J_0, if we can reject $H_0(x)$ for at least one pixel x. Then the overall type-I error for continuously indexed hypotheses (11) is given by

$$\begin{aligned} P(\text{ reject } J_0 | J_0 \text{ is true }) &= P\Big(\bigcup_x \{T(x) > h\}\Big) \\ &= 1 - P\Big(\bigcap_x \{T(x) \le h\}\Big) \\ &= 1 - P\Big(\sup_x T(x) \le h\Big) \\ &= P\Big(\sup_x T(x) > h\Big) \end{aligned}$$

for some h using t random field as a test statistic $T(x)$. The computation for the supremum distribution of a random field $T(x)$ is based on the expected Euler characteristic (EC) approach [1,7,23,27,30]. For sufficiently high threshold h, we have

$$P\Big(\sup_{x\in\mathcal{M}} T(x) > h\Big) = \sum_{d=0}^{N} \mu_d(\mathcal{M})\rho_d(h), \tag{12}$$

where $\mu_d(\mathcal{M})$ is the d-th Minkowski functional or intrinsic volume of $\mathcal{M}$ [21]. The d-th intrinsic volume of $\mathcal{M}$ is a generalization of d-dimensional volume.

$\mu_0(\mathcal{M})$ is the Euler characteristic of $\mathcal{M}$ while $\mu_N(\mathcal{M})$ is the volume of $\mathcal{M}$. For irregular jagged shapes such as the mandible shape $\mathcal{M}$, the intrinsic volume can be estimated using complicated enumeration techniques [10,29]. For t random field with $m-1$ degrees of freedom, EC-density ρ_d are [21,28]

$$\begin{aligned}
\rho_0(y) &= \int_y^\infty \frac{\Gamma(\frac{m}{2})}{((m-1)\pi)^{1/2}\Gamma(\frac{m-1}{2})}\Big(1+\frac{y^2}{m-1}\Big)^{-m/2}\,dy,\\
\rho_1(y) &= \lambda^{1/2}\frac{(4\ln 2)^{1/2}}{2\pi}\Big(1+\frac{y^2}{m-1}\Big)^{-(m-2)/2}\\
\rho_2(y) &= \lambda\frac{4\ln 2}{(2\pi)^{3/2}}\frac{\Gamma(\frac{m}{2})y}{(\frac{m-1}{2})^{1/2}\Gamma(\frac{m-1}{2})}\Big(1+\frac{y^2}{m-1}\Big)^{-(m-2)/2}\\
\rho_3(y) &= \lambda^{3/2}\frac{(4\ln 2)^{3/2}}{(2\pi)^2}\Big(1+\frac{y^2}{m-1}\Big)^{-(m-2)/2}\Big(\frac{m-2}{m-1}y^2-1\Big),
\end{aligned}$$

where $\lambda = \frac{1}{2t^2}$ measures the smoothness of the field, which is masked by smoothing with kernel K_t. The exact expression for the EC density ρ_d is available for other random fields such as t, χ^2, F fields [26], Hotelling's T^2 fields [6] and scale-space random fields [22]. In each case, the EC density ρ_d is proportional to $\lambda^{d/2}$, and it changes depending on the smoothness of the field. For instance, for zero mean and unit variance Gaussian field Z, we have

$$\begin{aligned}
\rho_0 &= P(Z > h) = 1 - \Phi(h)\\
\rho_1 &= \lambda^{1/2}\frac{e^{-h^2/2}}{2\pi}\\
\rho_2 &= \lambda h\frac{e^{-h^2/2}}{(2\pi)^{3/2}}\\
\rho_3 &= \lambda^{3/2}(h^2-1)\frac{e^{-h^2/2}}{(2\pi)^2},
\end{aligned}$$

where λ measures the smoothness of the field, defined as the variance of the derivative of the field.

7 Application

The methods were applied to 290 3D CT mandible images from an extant medical imaging database to localize the regions of growth between 0 to 20 years.

7.1 Subjects

Using an IRB approved extant imaging database, the dataset selected for this study consisted of 290 CT images from 290 typically developing individuals ranging in age from birth to 20 years old. Study inclusion criteria for the dataset were subjects who were imaged for medical reasons that do not affect growth and development of the head and neck, and who had a normal bite (Class I). Additionally, only studies with slice thickness 1.25 and 2.5mm slice thickness where the entire mandible could be visualized. Only CT images showing the full mandible without any motion or any other artifacts were selected though minimal dental artifacts were tolerated. Additional detail on scanner parameter and inclusion criteria is provided in [15,24]. The age distribution of the subjects is 9.66 ± 6.34 years. The minimum age was 0.17 years and maximum age was 19.92 years. A total of 160 male and 130 female subjects were divided into 3 groups [11]. Group I (age below 7) contained 130 subjects. , Group II (between 7 and 13) contained 48 subjects. Group III (between 13 and 20) contained 112 subjects.

7.2 Image preprocessing

CT images were visually inspected and determined to capture the whole mandible geometry withminor dental artifacts. As described in [24], these CT images were collected retrospectively, following University of Wisconsin-Madison Institutional Review Board (IRB) approval. The mandibles in CT were semi-automatically segmented using an in-house processing pipeline that involves image intensity thresholding using the Analyze software package (AnalyzeDirect, Inc., Overland Park, KS). Each of the processed mandibles were examined visually and edited manually by raters [11]. The segmented binary images were then affine registered to the mandible labeled as `F226-15-04-002-M` using the Advanced Normalization Tools (ANTS) [4] (Fig. 1). The mandible `F226-15-04-002-M` served as the initial template.

CT images are inherently noisy due to errors associated with image acquisition. Compounding the image acquisition errors, there are errors caused by image registration and semiautomatic segmentation. So it is necessary to smooth out the affine registered segmented images. We smoothed the binary images with Gaussian kernel with bandwidth $\sigma = 20$ voxels (Fig. 2). Gaussian kernel smoothing is often used to negate possible image registration artifacts, and increase the signal-to-noise ratio. The average of all 290 smoothed binary images was computed and used as the final template where the online statistical analyses are performed.

7.3 Statistical analysis

The analysis was done contrasting the probability maps between different ages (groups I to II, II to III, and I to III) by performing two-sample t-tests. Due to the computational load, the online algorithm was used to compute t-statistics

and determine p-values. Since the probability map was sufficiently smooth, the random field theory was used to account for multiple comparisons.

7.4 Results

Age effects. We performed two-sample t-test to assess age effects between the groups. We tested the statistical significance of the differences in mean probability maps while accounting for group variability differences (Fig. 3). The resulting t-statistic maps are displayed in Fig. 4-top and summarized in Table 1. Since it is difficult to visualize the 3D results on paper, we have also superimposed the statistic maps on the initial template `F226-15-04-002-M` (Fig. 5).

Voxels above or below ± 4.41 were considered significant in the t-statistic between Groups I and II at the 0.05 level after the multiple comparisons correction. Similarly for other age group comparisons, voxels above or below ± 4.43 (between II and III) and 4.37 (between I and III) were considered significant at the 0.05 level. These regions are colored dark red or dark blue. The dark red regions show positive growth (bone deposition) and dark blue regions show negative growth (bone resorption).

Between Groups I and II, significant growth in dentition is present (red regions), while the menton region (chin area), the midpoint of the lower border of the human mandible, shows negative growth, which is likely reflective of changes in bone shape/angle and size.the. Between Groups II and III, the condyle show significant growth in depth (red regions), while again the mental portion of the mandible (chin, blue region) show significant negative growth. The latter can only happen if the mandible is changing in shape such as widening or elongating. By comparing Groups I and III, we were able to increase the contrast of the growth and detect large clusters of growth regions in almost every part of the mandible. Statistically significant growth was detected in the condyle and ramus regions, indicating significant vertical growth/depth of the mandible after age 13. At the same time, significant negative growth in the mental (chin) region indicating that the angle of the U-shaped mandible was undergoing changes in shape and size. Even though the random field theory based thresholding gives very conservative results, we were able to detect large wide spread regions of morphometric change. The findings are consistent with our previous study that used a different technique, i.e., 2D surface deformation [11] and landmarks [15]. This is the first study to demonstrate that the proposed probabilistic map of tissue segmentation can be effectively used to characterize growth/morphometric change.

Sex effects. Within each group, we tested the significance of sexual dimorphism by performing the two-sample t-test between males and females. Note this analysis further reduces the sample size in each group and subsequently reduces the effectiveness of the statistical analysis compared to the age effect analysis. The resulting t-statistic maps are displayed in Fig. 4-bottom and results are summarized in Table 2. Since 3D statistical maps are superimposed on top of the initial template `F226-15-04-002-M` (Fig. 5).

Table 1: The results of statistical analysis on age effects. $< 10^{-d}$ indicates p-value is smaller than 10^{-d}, i.e., d decimal places.

	Group II - I	Group III - II	Group III - I
d.f.	176	158	240
Min. t-stat.	-5.22	-7.44	-13.61
Corrected p-value	0.019	$< 10^{-6}$	$< 10^{-25}$
Max. t-stat.	8.86	6.57	13.50
Corrected p-value	$< 10^{-9}$	0.000052	$< 10^{-24}$

Table 2: The results of statistical analysis on sex effects. * indicates p-value is above 0.1 and not significant.

	Group I (M-F)	Group II (M-F)	Group III (M-F)
d.f.	128	46	110
Min. t-stat.	-3.33	-2.42	-5.46
Corrected p-value	*	*	0.016
Max. t-stat.	2.26	2.85	2.99
Corrected p-value	*	*	*

Any regions with voxels above or below $\pm$ 4.37, 4.89 and 4.50 (for groups I, II and III respectively) were considered significant at 0.05 level after the multiple comparisons correction. There were no sex differences in groups I and II. In group III, the statistical significance was localized in the ramus on both sides. Note that this analysis is assessing morphometric differences between the two sexes, it is not quantifying overall size differences. Such global size differences were all removed after the affine registration. Thus, the significant signal is basically detecting relative size differences after removal of global size differences. Such findings are consistent with general findings on sexual dimorphism that become evident during puberty.

8 Discussion

The image processing and analysis somewhat resembles the voxel-based morphometry (VBM) widely used in modeling the gray and white matter tissue probability maps in structural brain magnetic resonance imaging studies [10]. In VBM, the posterior probability map is further estimated using the prior map. However, there is no such prior map in mandibles. We propose to distribute our average probability map as a potential prior map for other researchers so that they can obtain a better probability map. The advantage of the VBM framework over the ROI-volumetry approach is that it is completely automated and does not require artificial partitioning of the anatomical shapes, which introduces undesirable bias. Furthermore, it is not restricted to a priori ROI, enabling us to perform the statistical analysis at each voxel level and to pinpoint the exact location of the anatomical differences within ROI, even if there are no ROI size

differences. Although VBM was originally developed for whole brain MRI, we successfully applied it to mandible growth analysis for the first time.

In analyzing mandible growth, we did not pursue region of interest (ROI) volumetry. The main shortcoming of the ROI-volumetry is the artificial partitioning of the regions. It is also a time consuming process to partition mandibles into few ROIs. The ROI-volumetry may dilute the power of detection if the anatomical difference occurs near the partition boundary [10]. We also did not use the landmark based method. The landmark based approaches are based on few selective landmarks and it is difficult to infer in the regions with no landmarks. In this study, we proposed an alternative voxel-wise approach that avoid predefined regions of interests (ROI) or landmarks.

Unlike our deformation-based morphometry [11], which used the amount of displacement between corresponding mesh vertices, VBM uses probability as a feature. Therefore, we cannot give any direct physical interpretation to the result of VBM. Since we are not using any deformation vector field, we cannot infer the direction of growth or visulize any growth directions visually. Deformation-based morphometry [11] is a 2D surface-based morphometry that directly analyzed how the mandible surface grows over time using surface deformation. In contrast, the present study is a 3D volume-based morphometry that indirectly analyzed how the probability density changed over time at each voxel.

The probability map-based approach is based on the mathematical concept of probability density and is the indirect measure of the actual bone density. To directly determine if actual bone deposition or resorption occurs, we need to use the actual bone density measure and correlate with our VBM results. This is worth pursuing in further studies as it is of clinical relevance. To address the issue of computational burden as well as the need for sequential statistical inference procedures, an online algorithms for performing both one-sample and two-sample t-test was developed. We have shown the method can be further extended to linear regressions and the F -test. We expect this study to motivates researchers in imaging to develop online algorithms for more complex statistical inference procedures.

Acknowledgements

This work was supported by NIH Research Grant R01 DC6282 (MRI and CT Studies of the Developing Vocal Tract), from the National Institute of Deafness and other Communicative Disorders (NIDCD); and core grants P-30 HD03352 and U54 HD090256 to the Waisman Center from the National Institute of Child Health and Human Development (NICHHD).

References

1. R.J. Adler. *The Geometry of Random Fields*. John Wiley & Sons, 1981.
2. J. Ashburner and K. Friston. Voxel-based morphometry - the methods. *NeuroImage*, 11:805–821, 2000.

3. J. Ashburner and K. Friston. Why voxel-based morphometry should be used. *NeuroImage*, 14:1238–1243, 2001.
4. B.B. Avants, C.L. Epstein, M. Grossman, and J.C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12:26–41, 2008.
5. A. Blum. On-line algorithms in machine learning. In *Online algorithms*, pages 306–325. Springer, 1998.
6. J. Cao and K.J. Worsley. The detection of local shape changes via the geometry of Hotelling's T2 fields. *Annals of Statistics*, 27:925–942, 1999.
7. J. Cao and KJ Worsley. Applications of random fields in human brain mapping. *Spatial Statistics: Methodological Aspects and Applications*, 159:170–182, 2001.
8. T.F. Chan, G.H. Golub, and R.J. LeVeque. Algorithms for computing the sample variance: Analysis and recommendations. *The American Statistician*, 37:242–247, 1983.
9. M.K. Chung. *Statistical and Computational Methods in Brain Image Analysis*. CRC Press, 2013.
10. M.K. Chung, K.M. Dalton, A.L. Alexander, and R.J. Davidson. Less white matter concentration in autism: 2D voxel-based morphometry. *NeuroImage*, 23:242–251, 2004.
11. M.K. Chung, A. Qiu, S. Seo, and H.K. Vorperian. Unified heat kernel regression for diffusion, kernel smoothing and wavelets on manifolds and its application to mandible growth modeling in CT images. *Medical Image Analysis*, 22:63–76, 2015.
12. C.Y. Deng. A generalization of the Sherman–Morrison–Woodbury formula. *Applied Mathematics Letters*, 24:1561–1564, 2011.
13. T. Finch. Incremental calculation of weighted mean and variance. *University of Cambridge*, 4:11–5, 2009.
14. R.M. Karp. On-line algorithms versus off-line algorithms: How much is it worth to know the future? In *IFIP Congress*, volume 12, pages 416–429, 1992.
15. M.P. Kelly, H.K. Vorperian, Y. Wang, K.K. Tillman, H.M. Werner, M.K. Chung, and L.R. Gentry. Characterizing mandibular growth using three-dimensional imaging techniques and anatomic landmarks. *Archives of Oral Biology*, 2017.
16. D. Knuth. The art of computing, vol. ii: Seminumerical algorithms, 1981.
17. T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publisher, 1994.
18. J-B Poline and B.M. Mazoyer. Analysis of individual brain activation maps using hierarchical description and multiscale detection. *IEEE Transactions on Medical Imaging*, 13:702–710, 1994.
19. J.-B. Poline, K.J. Worsley, A.P. Holmes, R.S.J. Frackowiak, and K.J. Friston. Estimating smoothness in statistical parametric maps: Variability of P values. *Journal of Computer Assisted Tomography*, 19:788–796, 1995.
20. CH Salmond, J. Ashburner, F. Vargha-Khadem, A. Connelly, DG Gadian, and KJ Friston. Distributional assumptions in voxel-based morphometry. *NeuroImage*, 17:1027–1030, 2002.
21. V. Schmidt and E. Spodarev. Joint estimators for the specific intrinsic volumes of stationary random sets. *Stochastic Processes and their Applications*, 115:959–981, 2005.
22. D.O. Siegmund and K.J. Worsley. Testing for a signal with unknown location and scale in a stationary gaussian random field. *Annals of Statistics*, 23:608–639, 1996.
23. J.E. Taylor and K.J. Worsley. Detecting sparse signals in random fields, with an application to brain mapping. *Journal of the American Statistical Association*, 102:913–928, 2007.

24. H.K. Vorperian, S. Wang, M.K. Chung, E.M. Schimek, R.B. Durtschi, R.D. Kent, A.J. Ziegert, and L.R. Gentry. Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study a. *The Journal of the Acoustical Society of America*, 125:1666–1678, 2009.
25. A. Witkin. Scale-space filtering. In *Int. Joint Conference on Artificial Intelligence*, pages 1019–1021, 1983.
26. K.J. Worsley. Local maxima and the expected Euler characteristic of excursion sets of χ^2, f and t fields. *Advances in Applied Probability*, 26:13–42, 1994.
27. K.J. Worsley. Detecting activation in fMRI data. *Statistical Methods in Medical Research.*, 12:401–418, 2003.
28. K.J. Worsley, J. Cao, T. Paus, M. Petrides, and A.C. Evans. Applications of random field theory to functional connectivity. *Human Brain Mapping*, 6:364–7, 1998.
29. K.J. Worsley, S. Marrett, P. Neelin, and A.C. Evans. Searching scale space for activation in pet images. *Human Brain Mapping*, 4:74–90, 1996.
30. K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston, and A.C. Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4:58–73, 1996.

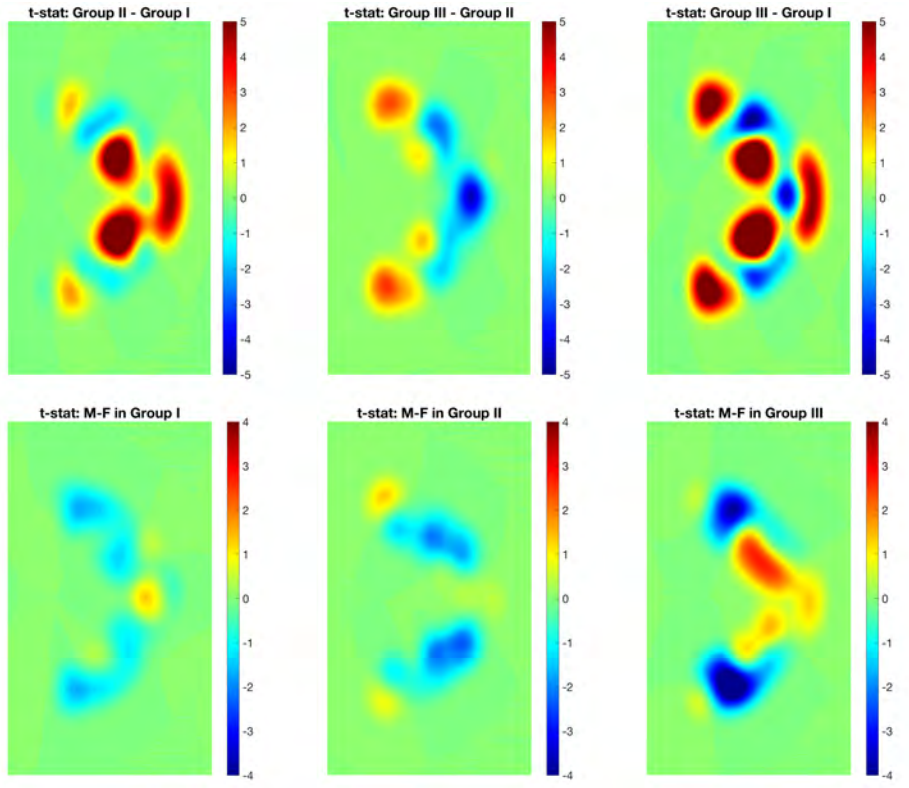

Fig. 4: Top: *t*-statistics maps showing mandibular growth. The morphometric changes of the mandible is shown between Groups I and II, II and III, and I and III. The condyle region show considerable growth in the Group III-I comparison. At the same time, the morphometric changes shown as negative growth in the mental region (chin area) are likely to be reflective of shape and size changes. Bottom: *t*-stat. maps showing sex differences in each age group. The sex difference is not pronounced until age 13.

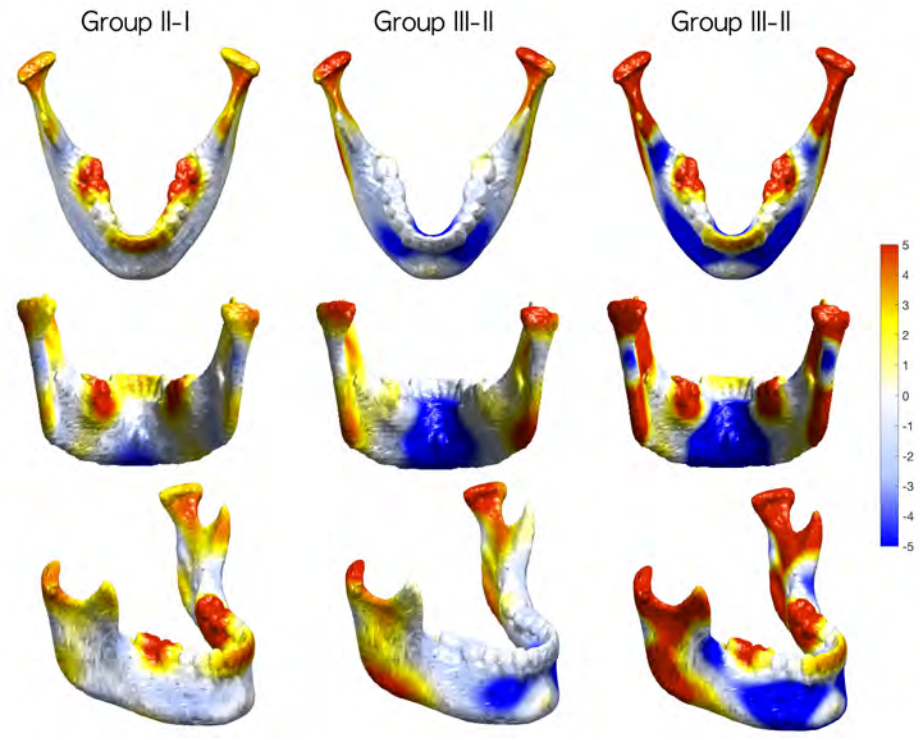

Fig. 5: t-stat. maps showing mandible growth. The elongation of mandible is shown between Groups II and III, and I and III. The condyle regions show prominent growth in Group III- I comparison. At the same time, the elongation is shown as negative growth (dark blue) at the back of the front teeth regions.

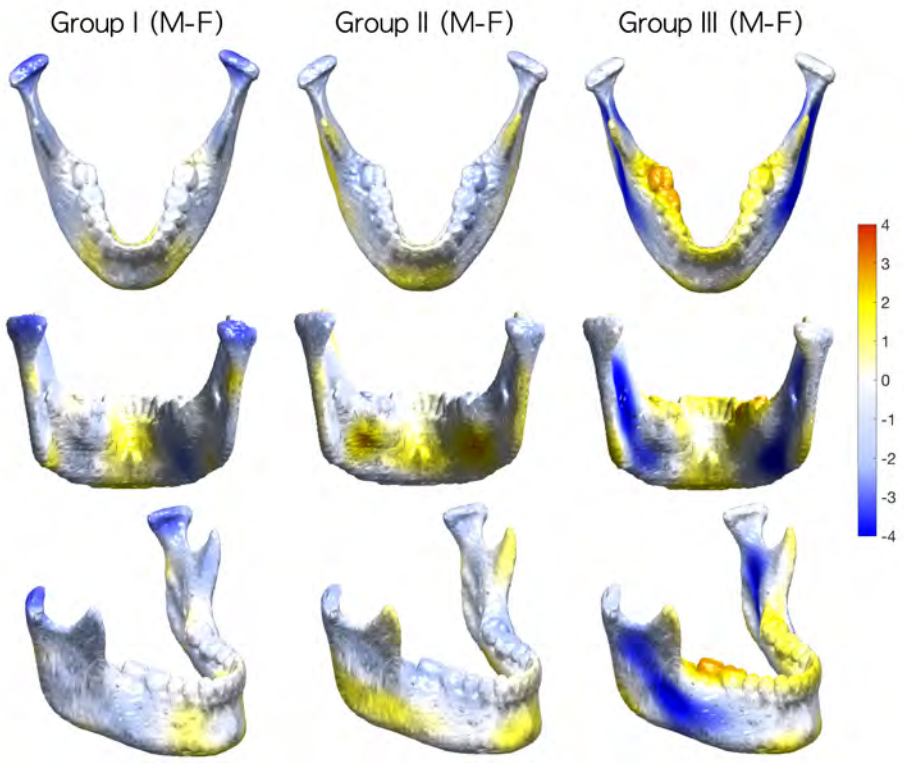

Fig. 6: t-stat. maps (male - female) showing sex differences in each age group. There were no significant sex differences in groups I and II. However, pubertal and post-pubertal sex difference are evident in group III.