

A Secondary Analysis of Panoramic Radiographs Reveals Hotspots in the Maxillofacial Region Associated with Diabetes

Gary D Pack, Ph.D.^{1,2} Mark Craven, Ph.D.¹ Amit Acharya, B.D.S., M.S., Ph.D.²

¹University of Wisconsin-Madison, Madison, WI,

²Center for Oral and Systemic Health, Marshfield Clinic, Marshfield, WI

Abstract

Diabetes mellitus is the putative cause of a number of pathologies occurring in the bony and soft tissues of the maxillofacial region and is known to exacerbate other oral diseases such as periodontitis. We present the first use of clinical panoramic radiographs for a secondary analysis of disease, with a focus on identifying hotspots in the maxillofacial region that are associated with diabetes. We developed a curated data set using Consensus Landmark Points (CLPs) and used that data to develop an analysis pipeline. This pipeline entailed automatic data cleansing, registration, and intensity normalization. The pipeline was used to process 7280 uncurated images that were subsequently analyzed using pixel-wise methods for a case/control study of patients with a history of diabetes. We detected statistically significant clusters of pixels that demarcated anatomical hotspots specific to the diabetic patients.

Introduction

In 2017, the **CDC**¹ estimated 23.0 million people were diagnosed with diabetes and an additional 7.2 million people had undiagnosed diabetes. The **CDC** also estimated that an additional 84.1 million people had prediabetes and only 11.6% are aware of their condition. Many oral and systemic diseases are associated with diabetes,^{2,3} and prediabetes increase the risks for coronary artery disease, stroke and progression to diabetes mellitus. A non-invasive screening test for diabetes that makes use of a huge store of information already available would be highly desirable. It is well established that diabetes causes changes in trabecular bone structure⁴ throughout the body including the maxillofacial region.

The maxillofacial region is one of the most commonly imaged regions of the body. The bony and cartilaginous tissues hold a record of both long and short term effects of pathologies occurring in the maxillofacial region. Increasing access to this information using automated techniques could have a tremendous impact on detecting and treating both local and systemic disease. One imaging modality in particular, panoramic radiographs, offers considerable promise for exploring the maxillofacial region.

Panoramic Radiographs (**PANs**) are acquired as a routine matter for developmental monitoring and general screening.⁵ In fact, many dental practices acquire them regularly at three to five year intervals to screen for chronic infections and cancerous lesions. This means that images are collected independent of acute disease, thereby providing data sets less biased than those that contain only diagnostic images. This property increases their value for use in secondary analysis. However, extracting information from clinical **PANs** for secondary analysis requires careful preprocessing. Clinical images are acquired using different devices, at different locations by many different machine operators, and using a variety of imaging parameters. Additionally, it is common practice to reduce image data from 12 bits at acquisition to 8 bits, (4096 levels of brightness to 256).⁶ The primary reason for this reduction is to replicate the properties of film and match the human limits of the ability to perceive differences in brightness.⁷ As a result, considerable information is lost. Differences in magnification inherent to producing the **PANs** is further compounded by improper positioning of patients in the machines.⁸⁻¹⁰ We have developed a processing pipeline to deal with these issues and to prepare these images for hotspot discovery.

Hotspots are patterns of pixels in images that encode information associated with disease phenotypes. Examples of hotspots are pixels whose statistical properties vary from image to image as a function of differences between cases and controls, “textures” in trabecular bone, metrics based on shapes or sizes of anatomical structures and estimates of density. Our goal is to discover disease phenotypes in the maxillofacial region associated with diabetes for eventual use in detecting established and precursor states of the disease.

The contributions of this work are three-fold. First, this is the first use of panoramic radiographs for secondary, population-based analysis and phenotype discovery. Second, we developed a processing pipeline for data cleansing and image registration enabling pixel-wise analysis of panoramic radiographs. Third, we detected statistically significant hotspots in maxillofacial panoramic radiographs in a case/control study using pixel-wise analysis. This work focuses

on determining if and where unknown phenotypes may be detected and highlights some of the challenges of working with panoramic radiographs.

Our pipeline is laid out in Figure 1. The top part of the figure shows how a set of images is manually annotated in order to identify a template image that serves as a common basis for pixel-wise analysis. In order to identify differences between our case and control groups through pixel-wise analysis, we must first register the images being analyzed to one another. In order to ensure that we have anatomically meaningful registrations, we use human annotators to mark landmark points on a small set of images that we refer to as the curated set. These landmark points then guide the registration process. Once we have computed all pairwise registrations, we identify a template image. The middle part of the figure shows how a larger set of images is processed and assigned to clusters that are anchored by the manually annotated images. This processing involves removing problematic images and determining which annotated image is most similar to each unannotated image in the larger set. Finally, the bottom part of the figure shows how all images are registered with the selected template image and pixel-wise analysis is performed to detect differences between cases and controls.

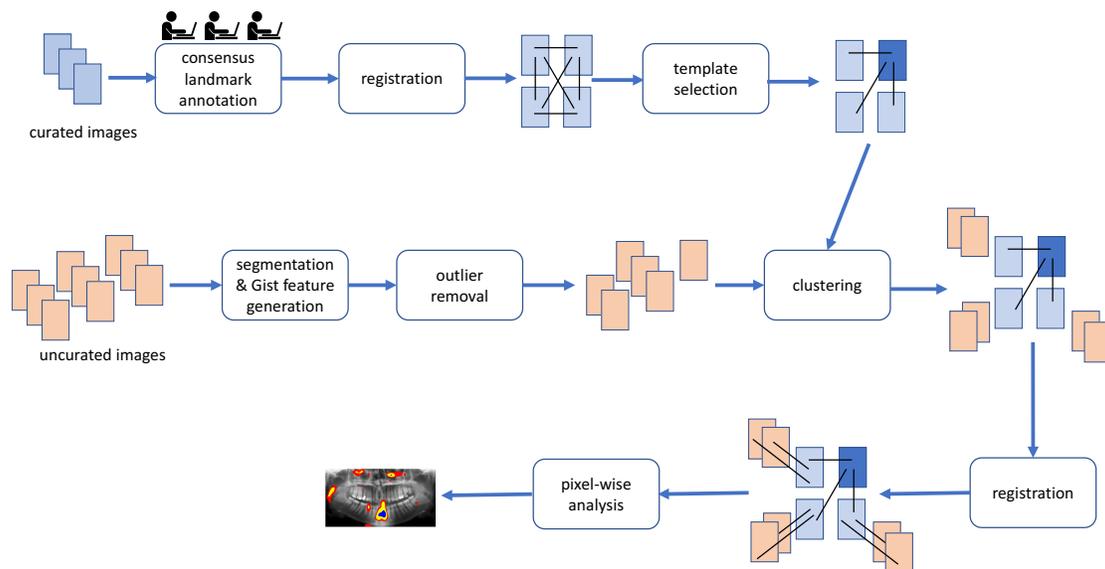


Figure 1: An overview of the processing steps required to produce and analyze the curated and uncurated data sets.

Methods

Curated Panoramic Radiographs

An initial set of 416 PANs were drawn at random from a database containing the dental records of patients at the Marshfield Clinic Dental service area. Each PAN represented a unique patient. We filtered the set of images by requiring at least 6 teeth in each quadrant of the mouth, and a field of view defined by the inclusion of the left and right glenoid fossa, the hyoid bones and a point midway between the vomer and nasal bones on the nasal septum. Only 105 of the 416 images met all the requirements and comprise the curated image set.

Uncurated Panoramic Radiographs

Medical records drawn from the Marshfield Clinic Research Foundation were used to identify¹¹ patients and matching controls for our uncurated case/control study. The patients ranged in age from 13 years old to 93 years old with the number of females approximately equal to the number of males. From these patient records, we selected uncurated panoramic images for 3460 subjects with 3460 matched controls for a total of 7280 images.

Design of the Consensus Landmark Point (CLP) Annotations

A consensus landmark point represents the consensus location, among a set of expert annotators, of a key feature point in the image. The CLP approach is necessary as our preliminary studies showed that standard feature point detectors such as SIFT, MSER, FAST, Hessian-Laplace, Harris-Laplace, HOG, SURF, DoG, and SUSAN¹² failed to identify

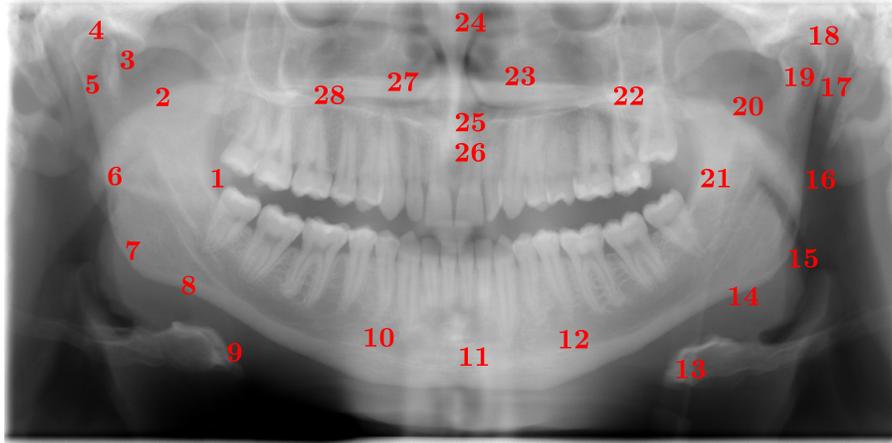


Figure 2: Reference image marked with anatomical reference points.

feature points in these images that could subsequently be used as landmark points.

The procedure for estimating landmark point locations involved the following: An expert annotator carefully marked a reference **PAN** with 28 reference landmark points as shown in Figure 2. Four annotators (one DDS and three non-expert colleagues) were trained to use software to mark points corresponding to these reference points on each of the 105 **PAN**s. Each annotator performed the following procedure. Two monitors were placed side-by-side. The reference image was displayed on the right-hand monitor and the image to be marked was displayed on the left-hand monitor. The cursor location was indicated with a large set of cross-hairs. The cursor was moved to the estimated point location. Clicking the left-hand mouse button marked the point location. The annotator stepped sequentially through 28 features and marked each point that corresponded to reference landmark points on the reference image. Each annotator repeated this procedure for each of the 105 panoramic radiographs. The annotators were required to wait at least one week, before repeating the full procedure. This ensured that the choices made in each session were as independent as possible.

Within-class consistency in estimating landmark point locations was scored for each annotator using Cronbach's α ($C\alpha$).¹³⁻¹⁵ In this context, a class is a set of locations an annotator chooses for a given landmark point. $C\alpha$ is a measure of reliability or internal consistency. $C\alpha$ is derived using a repeated-measures ANOVA with continuous variables. Therefore, $C\alpha$ is a measure of how consistently points were marked on the images on different occasions, with values in the range $(-\infty, 1]$. The higher the $C\alpha$ score the more consistent the choice. $C\alpha = 1$ is perfectly consistent, $C\alpha > 0.6$ is generally considered acceptably consistent, and $C\alpha < 0.6$ indicates inconsistent choices. A **CLP** is the weighted average of all four of the annotators' placement for a given landmark point, where the weight of each annotator's selection is proportional to that annotator's $C\alpha$ score.

Registrations

For the next step, we used the **CLPs** to perform pair-wise image registrations using two types of global geometric transforms to identify a template image for each type of transform. Each transform was applied pixel by pixel to put the original image into the new image space. The first geometric transform we used was the 2D affine transform.^{16,17} The equation for the transform is:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} s_x \cos(\theta) & sh_y \sin(\theta) & t_x \\ -sh_x \sin(\theta) & s_y \cos(\theta) & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} \quad (1)$$

Where x'_i and y'_i are pixel coordinates before applying the transform and x_i and y_i are the new coordinates of the pixels after applying the transform. For the variables in the square matrix, $sh_y = 1/sh_x$ represent the shear, s_x is the scale in the x or horizontal direction and s_y is the scale in the y or vertical direction, θ is the angle of rotation of the image, and t_x and t_y are translations in the x and y directions, respectively.

To compute the transform, x'_i and y'_i are set to the coordinates of the **CLPs** of each **PAN** and the x_i and y_i are set to the coordinates of the **CLPs** in another **PAN**. The solution of the affine equation determines the values of sh_y , $1/sh_x$,

s_x , s_y , θ , t_x , and t_y . This process was repeated for all image pairs.

A more flexible geometric transform is the 2D projective transform.¹⁷ The equation is:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_4 & p_5 & p_6 \\ p_7 & p_8 & 1 \end{bmatrix} \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} \quad (2)$$

Where, x'_i and y'_i are the i^{th} coordinates of the image pixels before applying the transform and x_i and y_i are the i^{th} coordinates of the image pixels after applying the transform. The p_k , $k \in 1, \dots, 8$ are the elements of the projective transform matrix. To compute the transform, x'_i and y'_i are set to the coordinates of the **CLPs** of each **PAN** and the x_i and y_i are set to the coordinates of the **CLPs** in another **PAN**. This process was repeated for all image pairs.

Template Image Selection

A template image is an image that is most similar to all of the other images. We used pair-wise Normalized Mutual Information (**NMI**)¹⁸ as our similarity metric. The “most” similar image was determined by:

$$t = \max_i \{ \min_j (\mathbf{NMI}(PAN_i, PAN_j)) \}, \quad (3)$$

where $\mathbf{NMI}(PAN_i, PAN_j)$ is the **NMI** score between PAN_i and PAN_j , with $i, j \in \{1, \dots, N\}, i \neq j$, and N is the number of images. t is the maximum of the minimum of the pairwise **NMI** scores and this value points to the template image. The image whose similarity is t is the image that is most similar to all of the other images. A convenient method for solving this equation is Johnson’s algorithm.¹⁹ Our version of Johnson’s algorithm returns an $N \times N$ matrix representing the pair-wise similarities between all of the images. The minimum of the column-wise values is a vector of the minimum pair-wise similarities. The vector containing the largest pair-wise similarity in this vector is t and the index of that element points to the template image. A template image was determined for each transform type.

The **CLPs** were used to determine pair-wise affine and projective geometric transforms. First, the transforms were computed between all 105 images using Equation 1. This gave us a set of 10920 transforms. Next, all images were registered using the calculated transforms and an **NMI** score was calculated for each pair of images. Then, Johnson’s algorithm and Equation 3 were used to find a template image for the affine image set. Once the template image was found, the remaining 104 images were registered to the template image using the corresponding transforms. The images were intensity normalized using the template image as the reference and **NMI** scores were recorded between the template image and all other images. Additionally, the mean and standard deviation were calculated for the distances between the template image’s **CLPs** and the corresponding **CLPs** on each registered image. Finally, the same procedure was used to determine the projective template image substituting Equation 2 for Equation 1; **NMIs** were calculated using the projective template image.

An additional processing step was required to deal with noise and residual registration errors. Smoothing kernels are non-parametric functions that are applied pixel-by-pixel to each image to “regress out” noise and reduce the effects of registration errors. We chose a Gaussian kernel function whose size is specified by its Full Width Half Maximum (**FWHM**) value. The Gaussian kernel function is defined as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \bar{x})^2}{2\sigma^2}\right) \quad (4)$$

The **FWHM** $\approx 2.355\sigma$ for a Gaussian kernel and σ can be adjusted so that $f(x)$ returns the required **FWHM** filter. During pre-study testing we determined that the results were relatively insensitive to kernel size. Consequently, we chose a kernel size of **FWHM** = 15mm from the middle of the range of tested values.

Pixel-Wise Analysis

Prior to pixel-wise analysis it was necessary to normalize the intensities of the images. The histograms of each image were matched to the histogram of the reference image. The histogram of the reference image was adjusted to reduce shadowing using Contrast Limited Adaptive Histogram Equalization²⁰ implemented in the MATLAB²¹ function *adaphisteq*. Following this step, the histograms of each of the other images were adjusted to match the histogram of the adjusted reference image.

Our task was to determine if pixel-wise differences in maxillofacial region associated with disease could be determined using a case-control approach comparing images from a population of subjects with diabetes and images from matched population of subjects without diabetes.

Pixel-wise image analysis can be thought of as “stacking” images together then sampling corresponding pixels. The difference between cases and controls was assessed using a two sample t-test. Testing the entire image area required in about two million t-tests. Therefore, we adjusted for multiple comparisons by controlling the False Discovery Rate²² (**FDR**). We used the Storey **FDR**²³, to estimate the p-value required to bound the expected number of false positives in a group estimated to be positives. For example, for an **FDR** < .001, only 1 false positive would be expected in a group of 1000 predicted positives. It is possible to have a well controlled **FDR** and still have a small effect size. Consequently, we focus on pixels that survive FDR control and have a large effect size. A well-known metric of effect size is Cohen’s d.²⁴ It is the difference between the mean values of the case and control sets divided by the pooled standard deviation of those sets.

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}}, \quad (5)$$

where \bar{x}_i, s_i^2 , and n_i are the mean, variance and number of pixels in each set. $i = 1$ is the control set and $i = 2$ is the case set. It is common to consider $|d| > 0.8$ a large effect size.

Segmentation, Clustering, and Outlier Removal

Clinical imaging data sets are comprised of images of varying quality. As part of the data cleansing process we needed to detect and remove images that were of insufficient quality. Issues we encountered included images that were improperly labeled as panoramic radiographs, were partially occluded by lead aprons or large pieces of jewelry, had required portions of the anatomy that were out of the field of view, had insufficient contrast, or were highly distorted due to positioning or acquisition errors. The first two conditions on the list were detected and removed using sensor information in the image header to confirm the image size and by detecting large areas of image saturation near the image midline. The problems in the rest of the list involved determining if the regions of the anatomy were present. Our curated images provided examples of what was acceptable and formed the exemplars for what we refer to as supervised clustering. Observation suggested that edges in the image that represented the large scale structure were a combination of both weak and strong edges with many irrelevant examples of both throughout the image. Detection of lines parallel to nearby weak edge were “real” while more isolated lines or lines orthogonal to weak edges were less likely to be associated to large scale structure. The following procedure was used to perform a coarse segmentation. First, the images were over-segmented using mean shift based super-pixels.²⁵ Second, the superpixels were clustered using a procedure similar to Simple Linear Iterative Clustering.²⁶ Third, a line finding algorithm²⁷ as implemented in the Scenes/Objects Recognition toolbox²⁸ was applied to the image. Fourth, line segments approximately parallel to superpixel boundaries were kept and all others discarded. Finally, the coarse segmentation consisting of just the line segments was processed to extract Gist Feature Vector (**GVF**)^{29,30} for each image. This procedure was applied to both the curated images and the uncurated data set so that all images were represented by their **GVF**s.

The clustering stage of the pipeline used the **GVF**s of the curated images as cluster centers. The **GVF**s of uncurated images were then clustered using a cosine distance function. Cosine distances more than one standard deviation from their assigned cluster center were considered to be outliers and the corresponding images were discarded.

Results

Consensus Landmark Points

Figure 3 summarizes the intra-annotator **C α** scores for the most inconsistent annotator. This annotator placed approximately 70% of the points in the images consistently. This was one of the non-expert annotators. However, the other annotators were only slightly more consistent with the DDS annotator in the middle of the group. Each **CLP** is the result of the combination of at least two or more annotators with **C α** scores > 0.6. We met our requirements for **C α** scores for all points, so we conclude that **CLP** locations are reasonably well estimated.

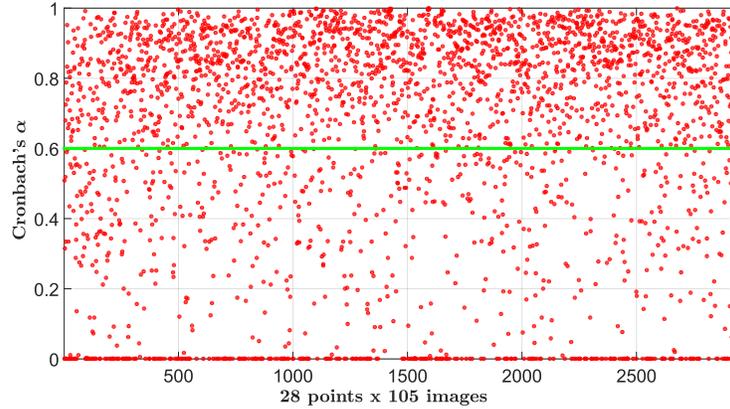


Figure 3: The least consistent annotator's results. Approximately 70% of the points in the images were marked consistently. Annotators tended to mark different points with different levels of consistency, leading us to weight each annotator's contribution to a given CLP in proportion to their $C\alpha$ score for that point. The green line is the $C\alpha = .6$ reference line.

Registration and Template Image Determination

Figure 4(a) shows the **NMI** scores between the affine registered images and the affine template image and the **NMI** scores between the projective registered images and the projective template image. To ease visualization, the images are ordered by the affine **NMI** scores and both affine and projective scores are plotted together. As can be seen, the projective transform images often have much higher **NMI** scores. However, it is important to keep in mind that **NMI** scores are only comparable when looking at relative values on the same image pair. Additionally, the more dissimilar two images are, the less reliable the magnitude of the **NMI** score becomes. It can be seen that the projective transform gave higher **NMI** values in most cases, but in some cases it did worse.

Figure 4(b) shows the mean and standard deviation of the pairwise distances between points for each of the 105×105 image pairings. To ease visualization, the images were ordered by the mean affine distances and both affine and projective were plotted together. Roughly twenty percent of registrations showed little difference in mean distances for the affine or projective transforms. This suggests that some of the images are fairly similar to the template image already. As the figure shows, the projective transform resulted in smaller mean distances than the affine transform for nearly all of the images.

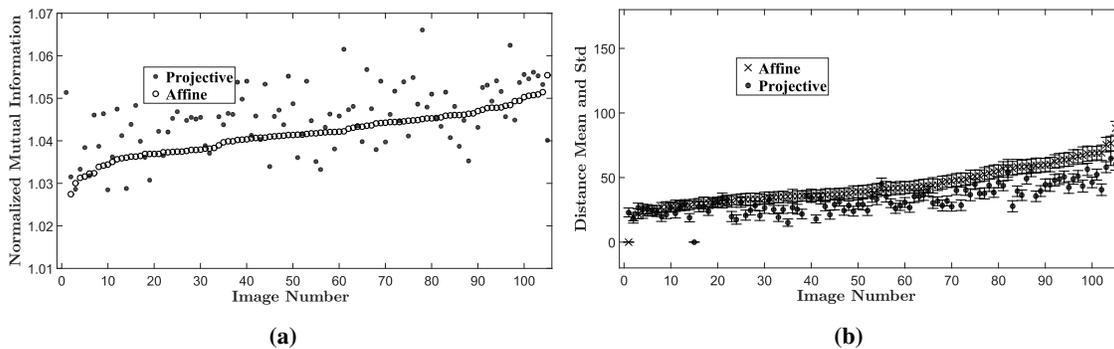


Figure 4: (a) After registration, **NMI** was computed between each image and the corresponding template image. Higher values are better. (b) After registration, the mean and standard deviation distances between **CLPs** and the corresponding template were computed. Lower values are better.

Given the results summarized in Figure 4, the projective template image was chosen as the template image for all subsequent analysis. The template image and the projective transforms between it and the other curated images were passed to the pipeline to be used in processing the uncurated data set as illustrated in the top row of Figure 1.

Segmentation, Clustering, and Outlier Removal

Figure 5 steps through the coarse segmentation process leading to the Gist feature vectors used for clustering images. Low contrast images resulted in very large superpixels and did not capture much structure. The threshold for the maximum superpixel size was set to 1.1 times the maximum superpixel size found in the curated images. The norm of the Gist features vector was also used to further filter images. The threshold for the minimum norm was set to .9 times the minimum norm of the uncurated image set. Among the 7280 uncurated images, 4729 images were filtered out using these criteria. Cosine distance using the Gist feature vectors was used to perform clustering where each uncurated image was assigned to the nearest curated image. Each uncurated image was registered to the curated image at its cluster center using automatic intensity based affine registration. Images sometimes required multiple registration attempts using a sweep of the optimizer initial values before convergence. Even then some registrations failed to converge and those images were discarded. To avoid another degenerate registration case, the size of the uncurated image after registration was required to be greater than 60% of the size of the original image. We discarded 281 images for failing these two tests. The final uncurated study data set contained 2270 images. The patients represented by these images ranged in age from thirteen years old to ninety three years old and consisted of 1587 males and 683 females.

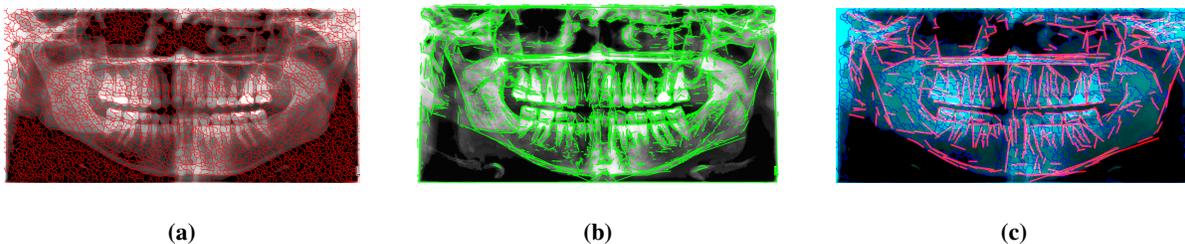


Figure 5: Building coarse segmentation for Gist features: (a) First, mean shift superpixels were created, shown in red. (b) Next, line finding using the superpixels boundaries was used to determine strong lines, shown in green. (c) Finally, the magenta lines show the coarse segmentation.

Pixel-wise Analysis

After filtering and registration, 2270 images remained of which 1020 were cases and 1250 were controls. Figure 6 depicts the regions where there were group differences in pixel values and indicates the effect size in those regions. Statistically significant differences were found in 11 regions, five of which were spatially distinct. Note that there is an asymmetry to the hotspots with more appearing on one side in the images than the other. When we relax the FDR and Cohen's d requirements similar hotspots are found on the opposite side of the mandible.

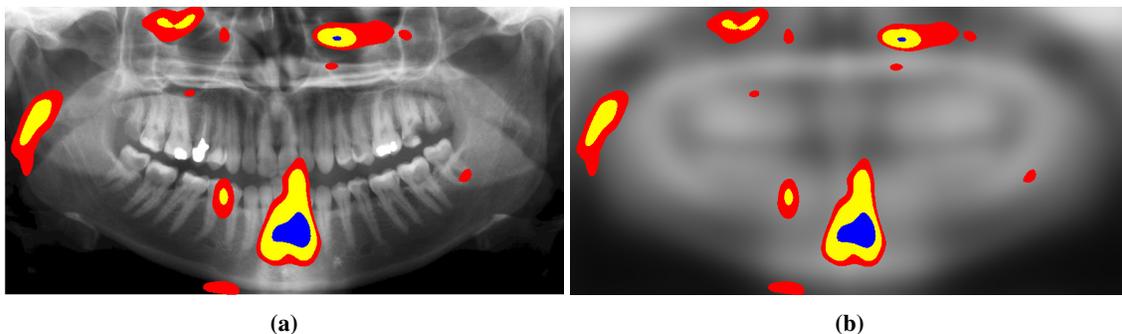


Figure 6: Pixel-wise analysis of 1020 cases and 1250 controls. Highlighted pixels indicate those with significant differences between cases and controls with $FDR < .001$. The color of the highlighted pixels indicate the effect size and represent nested contours. Red pixels have Cohen's $d \geq .15$ (2335 pixels). Yellow pixels have Cohen's $d \geq 0.2$. Blue pixels have Cohen's $d \geq .25$. a) Significant pixels projected onto the projective template image. b) Significant pixels projected onto an image that is the mean of the smoothed control group images

Discussion

Our study represents the first use of panoramic radiographs for secondary, population-based analysis and phenotype discovery. We were successful in using images to identify regions where there are significant differences between patients with a history of diabetes and controls. We developed a pipeline for large-scale unsupervised registration, discovery, and analysis of hotspots that can be applied to millions of PAN images for a wide range of applications.

Our study confirmed the common observation that expertise is not necessarily a prerequisite for good annotation. Experts and non-experts were similar in their consistency in point selection. It may be that experts are biased in choosing landmarks because they try to interpret the anatomy in the image. On the other hand, non-expert annotators may make their decisions based on image contrast without trying to understand the anatomy.

The fact that we filtered out 69 percent of the uncurated images was not unexpected. There have been many studies evaluating errors in panoramic radiographs.^{9,31,32} The quality of the images depends greatly on where and why they were collected. Image data sets with higher error rates were collected from general practice dental offices. Conversely, image data sets collected for prospective studies in academic institutions had the fewest errors. In either case, approximately 25 percent of images reviewed in these other studies had a sufficient number of errors that it was necessary to retake them. We have even more stringent criteria since we require that the uncurated images be similar to the curated images.

Some of our hotspots overlay the region where a cancer associated with diabetes occurs so the diagnostic value of the images may well extend to other diseases. For example diabetes is a suspected risk factor for nasopharyngeal carcinoma³³. The large red and yellow hotspot on the left side of Figure 6 (a) overlays the nasopharyngeal region and warrants further study. The large hotspot centered on the chin or the mental region overlays a collection of muscle attachment points called the mental splines. It is also proximal to the mental foramen where the mental nerve exits the mandible. The precise role that diabetes would play in the pathologies of this region are uncertain but, given the strength of our this hotspot, warrants further study. The other hotspots appear to be associated with points where nerves and blood vessels pierce bone. For example the large blue, yellow, and red area on the upper right side of Figure 6 (a) is proximal to the infraorbital foramen. The reason for the left/right asymmetry is unknown. The images are shown in radiological orientation so the left side of the image is the right side of the face and vice versa. We speculate that this asymmetry might associated with handedness. This will be further explored in future work.

We used a comparatively small set of curated images to build the models in our pipeline. A much larger curated image set would likely improve our results. One further limitation of our study is that we searched the entire maxillofacial region using a single class, i.e., presence or absence of a diagnosis of diabetes. It is unlikely that all patients would show hotspots in the same regions since they range in age from thirteen years old to ninety three years old. They are also likely to be in significantly different stages of disease progression. Taking these factors into account would greatly improve the sensitivity of our model.

Acknowledgements

Gary D Pack, PhD acknowledges support of the NLM training grant: 5T15LM007359 and the support of the Marshfield Clinic BIRC - Dental Informatics Research and Training Program. Mark Craven, PhD acknowledges NIH BD2K program grant U54 AI117924.

Thanks to the annotator team: Neel Shimpi BDS, MM, PhD, Harshad Hegde MS, Kelsey Schwei PhD, and Cathy Schneider.

References

- [1] Centers for Disease Control and Prevention. National diabetes statistics report, 2017.
- [2] B. L. Mealey. Periodontal disease and diabetes: A two-way street. *The Journal of the American Dental Association*, 137:S26–S31, 2006.
- [3] M. J. Nesbitt, M. A. Reynolds, H. Shiao, K. Choe, E. M Simonsick, and L. Ferrucci. Association of periodontitis and metabolic syndrome in the Baltimore longitudinal study of aging. *Aging Clinical and Experimental Research*, 22(3):238, 2010.

- [4] Kannikar Wongdee and Narattaphol Charoenphandhu. Update on type 2 diabetes-related osteoporosis. *World Journal of Diabetes*, 6(5):673, 2015.
- [5] American Dental Association Council on Scientific Affairs. Dental radiographic examinations: Recommendations for patient selection and limiting radiation exposure. *U.S. Department of Health and Human Services, Public Health Service, Food and Drug Administration*, 2012.
- [6] W. Geelen, A. Wenzel, E. Gotfredsen, M. Kruger, and L. G. Hansson. Reproducibility of cephalometric landmarks on conventional film, hardcopy, and monitor-displayed images obtained by the storage phosphor technique. *The European Journal of Orthodontics*, 20(3):331–340, 1998.
- [7] P. L. Rheingans. Task-based color scale design. In *28th AIPR Workshop: 3D Visualization for Data Exploration and Decision Making*, pages 35–43. International Society for Optics and Photonics, 2000.
- [8] W. Al-Faleh. Common positioning errors in panoramic radiography. *Egyptian Dental Journal*, 51:1813–1817, 2005.
- [9] R. H. N. Rondon, Y. C. L. Pereira, and G. C. do Nascimento. Common positioning errors in panoramic radiography: A review. *Imaging Science in Dentistry*, 44(1):1–6, 2014.
- [10] C. M. Granlund, A. Lith, B. Molander, K. Gröndahl, K. Hansen, and A. Ekestubbe. Frequency of errors and pathology in panoramic images of young orthodontic patients. *The European Journal of Orthodontics*, 34(4):452–457, 2012.
- [11] J. A. Pacheco and W. Thompson. Type 2 diabetes mellitus electronic medical record case and control selection algorithms. *Hemoglobin*, 1:6–5, 2011.
- [12] T. Tuytelaars, K. Mikolajczyk, et al. Local invariant feature detectors: a survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3):177–280, 2008.
- [13] L. J. Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334, 1951.
- [14] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417, 1933.
- [15] G. H. Golub and C. Reinsch. Singular value decomposition and least squares solutions. *Numerische Mathematik*, 14(5):403–420, 1970.
- [16] H. Späth. Fitting affine and orthogonal transformations between two sets of points. *Mathematical Communications*, 9(1):27–34, 2004.
- [17] A. W. Fitzgibbon. Robust registration of 2D and 3D point sets. *Image and Vision Computing*, 21(13):1145–1153, 2003.
- [18] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.
- [19] D. B. Johnson. Efficient algorithms for shortest paths in sparse networks. *Journal of the ACM (JACM)*, 24(1):1–13, 1977.
- [20] K. Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics Gems IV*, pages 474–485. Academic Press Professional, Inc., 1994.
- [21] MathWorks. MATLAB and statistics toolbox release, 2017a.
- [22] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995.
- [23] John D Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.

- [24] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edn. Erlbaum, 1988.
- [25] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM International Conference on Multimedia*, pages 1469–1472. ACM, 2010.
- [26] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.
- [27] R. G. Von Gioi, J. Jakubowicz, J. Morel, and G. Randall. LSD: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):722–732, 2010.
- [28] S Paris. Scenes objects classification toolbox. <https://www.mathworks.com/matlabcentral/fileexchange/29800-scenes-objects-classification-toolbox>. *MATLAB Central File Exchange*, 2019.
- [29] A. Oliva and A. Torralba. Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research*, 155:23–36, 2006.
- [30] A. Chauhan, D. Chauhan, and C. Rout. Role of gist and phog features in computer-aided diagnosis of tuberculosis without segmentation. *PloS ONE*, 9(11):e112980, 2014.
- [31] N. A. Brezden and S. L. Brooks. Evaluation of panoramic dental radiographs taken in private practice. *Oral Surgery, Oral Medicine, Oral Pathology*, 63(5):617–621, 1987.
- [32] B. Peretz, M. Gotler, and I. Kaffe. Common errors in digital panoramic radiographs of patients with mixed dentition and patients with permanent dentition. *International Journal of Dentistry*, 2012, 2012.
- [33] Xing-Si Peng, Guo-Feng Xie, Wen-Ze Qiu, Yun-Hong Tian, Wei-Jun Zhang, and Ka-Jia Cao. Type 2 diabetic mellitus is a risk factor for nasopharyngeal carcinoma: a 1: 2 matched case–control study. *PloS ONE*, 11(10):e0165131, 2016.