Are the teeth cracked?

-- Evaluating the performance of imaging methods on cracked-teeth identification via functional data analysis

Consultant: Wei Angel Huang

Clients: Dr. Beatriz Paniagua, Dr. Jared Vicory, Kitware Inc, Carrboro, NC

Professors: Dr. Steve Marron, Dr. Perry Haaland

March 4th, 2021

Abstract

Microfractures (cracks) are the third most common cause of tooth loss in industrialized countries. If not detected early, microfractures may continue to progress, causing sharp pain and tooth loss. To improve early detection of cracks in a non-invasive, data-driven way, a new imaging analysis algorithm was developed in combination with high resolution cone beam computed tomography (hr-CBCT). To compare its performance with baseline method micro computed tomography (micro-CT), we analyzed two datasets derived from micro-CT and hr-CBCT, both containing cracked teeth as well as healthy teeth. Using functional data analysis, we found that for micro-CT, cracked teeth and healthy teeth are hardly distinguishable due to the high false positives of cracks detected. However, for hr-CBCT, there are much fewer false positives and the data from cracked teeth and healthy teeth show significant separation. This result supports that the hr-CBCT pipeline outperforms micro-CT in early detection of cracked teeth, which could potentially lead to earlier treatment and longer tooth retention.

Executive Summary

An overview of the major findings and their corresponding subsection number is as follows:

- *The distributions of the number of cracks in cracked and healthy teeth are different in the hr-CBCT dataset but mostly overlapping in the micro-CT dataset (Result 1).*
- *The crack-size curves for cracked vs. healthy teeth are more separable in the hr-CBCT dataset than the micro-CT dataset, and is better visualized in logarithmic scale (Result 2).*
- *The hr-CBCT but not the micro-CT dataset shows good separation of teeth conditions in principal component directions (Result 3).*
- *The hr-CBCT but not the micro-CT dataset shows separation of means of the quantile functions between the two teeth conditions in later quantiles of data (Result 4).*
- *Quantitative testing reveals a significant separation of the two teeth conditions using hr-CBCT but not micro-CT (Result 5).*

Introduction

Cracked teeth are the third most common cause for tooth loss in industrialized countries.[1] Once a crack develops, it is colonized by bacteria, which can cause pulpal and periapical disease[2], both of which cause intense pain and are the most common reasons for emergency dental care[3]. If left undetected, cracks continue to progress and ultimately result in tooth loss.

However, cracked teeth are extremely hard to detect, especially during the early stage. 2D intraoral radiographs and cone beam computed tomography (CBCT) scans are imaging tools used to detect cracks. CBCT performs better than 2D intraoral radiographs by capturing the 3D structure, but still has its limitations. To this end, the client combined high-resolution CBCT (hr-CBCT) with advanced image analysis and machine learning in the hope of devising a better tool for early detection of cracks in a non-invasive, data-driven way. The goal of this report is to evaluate the data derived from this pipeline on its ability to separate out cracked teeth from healthy teeth. Another dataset derived from micro computed tomography (micro-CT) was also evaluated using the same pipeline as a baseline comparison.

Methods

Data

Data was provided by Jared Vicory, Kitware Inc, Carrboro. There are two datasets, one from micro-CT (n=45 teeth, 31 cracked and 14 healthy), another from synthetic hr-CBCT (n=25 teeth, 19 cracked and 6 healthy). Detailed description of data generation methods are here[4].

In short, there are 45 extracted human premolars (first and second molars) in the micro-CT dataset. 31 teeth were randomly selected to undergo simulated stress microfractures while the other 14 teeth were used as healthy controls. All the teeth were then examined by two blinded investigators to mark the presence(=1)/absence(=0) of superficial microcracks in each tooth. It was confirmed that the simulated stress reliably induces cracks which mimics the common clinical presentations of stress induced dental cracks.

Another set of teeth data was generated using synthetic hr-CBCT[5] without the smoothing filters. The smoothing filters were used in normal hr-CBCT for better perception of gross tooth structure but destroys fine features like microfractures. Since an hr-CBCT dataset without filter is not available, we used the data that were synthetically generated for hr-CBCT as in this paper[5].

In both datasets, each tooth is represented by a vector of putative crack sizes, sorted from the largest crack (i.e. largest number of connected voxels) to the smallest (i.e. fewest number of connected voxels). Each tooth can have a different number of cracks and each crack can have different sizes (measured as number of connected voxels).

Functional data analysis

Functional data analysis (FDA) is an extension of multivariate analysis that is suitable to analyze the variation in a population of curves or distributions. Since each of our tooth samples is a vector of crack sizes that are sorted from largest to smallest, it can be viewed as a curve. Thus, we chose FDA as our main analytical method. The FDA package used in this analysis was developed by Dr. Steve Marron[6].

The analysis scripts were written in Matlab 2020a and can be found in https://github.com/angelvv/STOR765_Spring2020_Angel_Huang.

Results

Due to the fact that the two datasets were generated in different settings and in different sets of teeth, no direct quantitative comparison was made between the two datasets. Instead, the focus of this study is to evaluate the performance of the hr-CBCT pipeline, with Micro-CT serving as a qualitative baseline comparison. The exact same analysis pipeline was applied to both datasets. In fact, to reduce the potential bias in analysis, the consultant was kept blinded to the conditions of the datasets during the analysis and was not aware of the comparison until the end of analysis.

1. *The hr-CBCT shows distinct distributions of the number of cracks between teeth conditions, whereas micro-CT does not.*

Since each tooth has a different number of putative cracks (i.e. features), we first visualized the distribution of number of cracks (Fig. 1). For micro-CT dataset, cracked teeth have a larger spread of the number of cracks (between 315 and 1000) than healthy teeth (between 408 and 751). But there is a big overlap of distribution between cracked and healthy teeth, indicating that many cracks are probably false-positives in the healthy teeth. For hr-CBCT, cracked teeth have generally many more cracks (between 47 and 948) than healthy teeth (between 6 and 148).

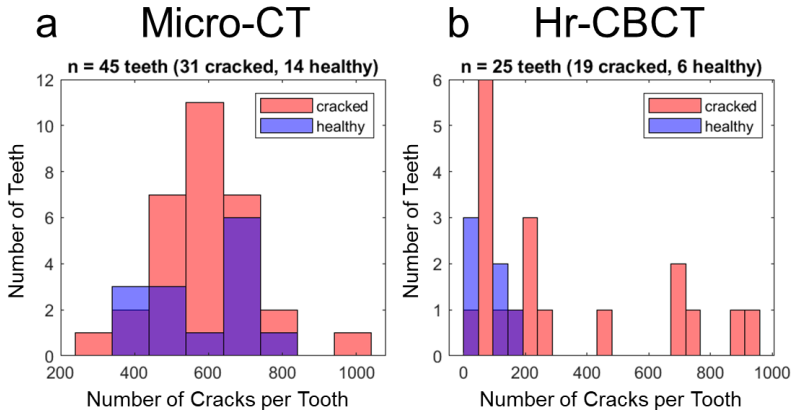


Figure 1. Distribution of number of cracks per tooth for micro-CT (a) and hr-CBCT (b). Hr-CBCT shows bigger separation of distribution between cracked teeth (red) and healthy teeth (blue).

2. *The crack-size curves for cracked vs. healthy teeth are more separable in the hr-CBCT dataset than in the micro-CT dataset*

To further visualize how crack sizes are distributed in each tooth, we plotted the crack-size curves for both micro-CT and hr-CBCT datasets (Fig. 2). Each curve represents one tooth and each tooth has many cracks. We sorted the cracks by their sizes (measured by number of connected voxels) in decreasing order and displayed the largest 50 cracks for each tooth. This shows the distribution of crack sizes for each tooth. For each crack, Since the largest crack size is normally in the hundreds or thousands and the smaller crack sizes are in the single or double digits, a logarithmic scale of data provides better visual separation across this wide range of scales (Fig. 2 bottom). As a result, we will use a logarithmic scale of data for further analysis.

In addition, we compared the largest crack in each dataset (Fig. 2 bottom, Crack ID=1) since the larger the crack, the more likely it resembles a real crack than noise or artifacts. For micro-CT, the range of the largest crack in cracked teeth (range from 125 to 1336 voxels) is a little broader than that of control teeth (range from 149 to 1210 voxels), suggesting a higher variability in cracked teeth samples. For hr-CBCT, a more distinct difference is seen -- the range of the largest crack in cracked teeth (range from 10,403 to 1,151,856) is more right-shifted than healthy teeth (range from 14 to 66,927). This shows that hr-CBCT is more likely to identify larger cracks (even in healthy teeth) but is also better at identifying real cracks. This can also be seen in the separation of the first feature of two teeth conditions (Fig. 2 bottom).

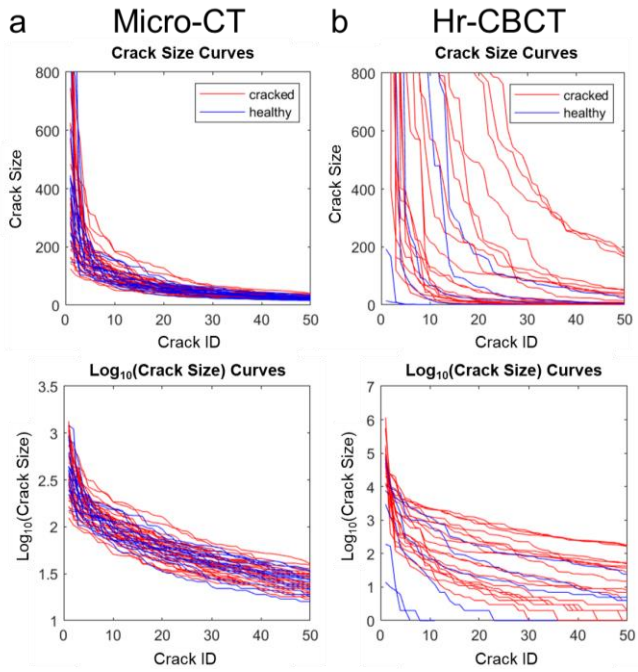


Figure 2. Crack-size curves of the largest 50 cracks in each tooth for micro-CT (a) and hr-CBCT (b) in their original unit (top) and logarithmic scale (bottom). Each curve represents one tooth, with its crack sizes displayed in descending order. For both datasets, logarithmic scale provides better visualization of the distribution. The hr-CBCT dataset shows qualitatively better separation of crack-size distributions between cracked (red) and healthy (blue) teeth.

3. *Hr-CBCT shows better separation of teeth conditions in principal component directions*

For data that contains a large number of features like in our datasets, Principal Component Analysis (PCA) is an effective way to extract the key features that correspond to lower dimensional representations of the data. In functional data analysis, PCA helps to visualize the difference between distributions. In the top row of Figure 3, we plotted the data, its mean, mean residual, and the R2 of the residual. It is interesting to see that for micro-CT, the shape of the data was mostly determined by the mean (98% of the total sum of square of data) whereas for hr-CBCT, the shape of the data is a feature of the mean (57% of the total sum of square) as well as the variability about the mean (43% of the total sum of square). We then extracted first four principal components for each dataset (Fig. 3 row 2-5). For micro-CT, none of the PC direction showed good separation between the two teeth conditions as shown by the overlapping blue and red colors. For hr-CBCT, PC1 (explains 91% of the variance) seemed to extract the vertical shift of data, and other PCs extracted more of the dynamic of the distribution. In addition, in both PC1 and PC4 dimensions, healthy teeth corresponded to higher PC scores than cracked teeth in general.

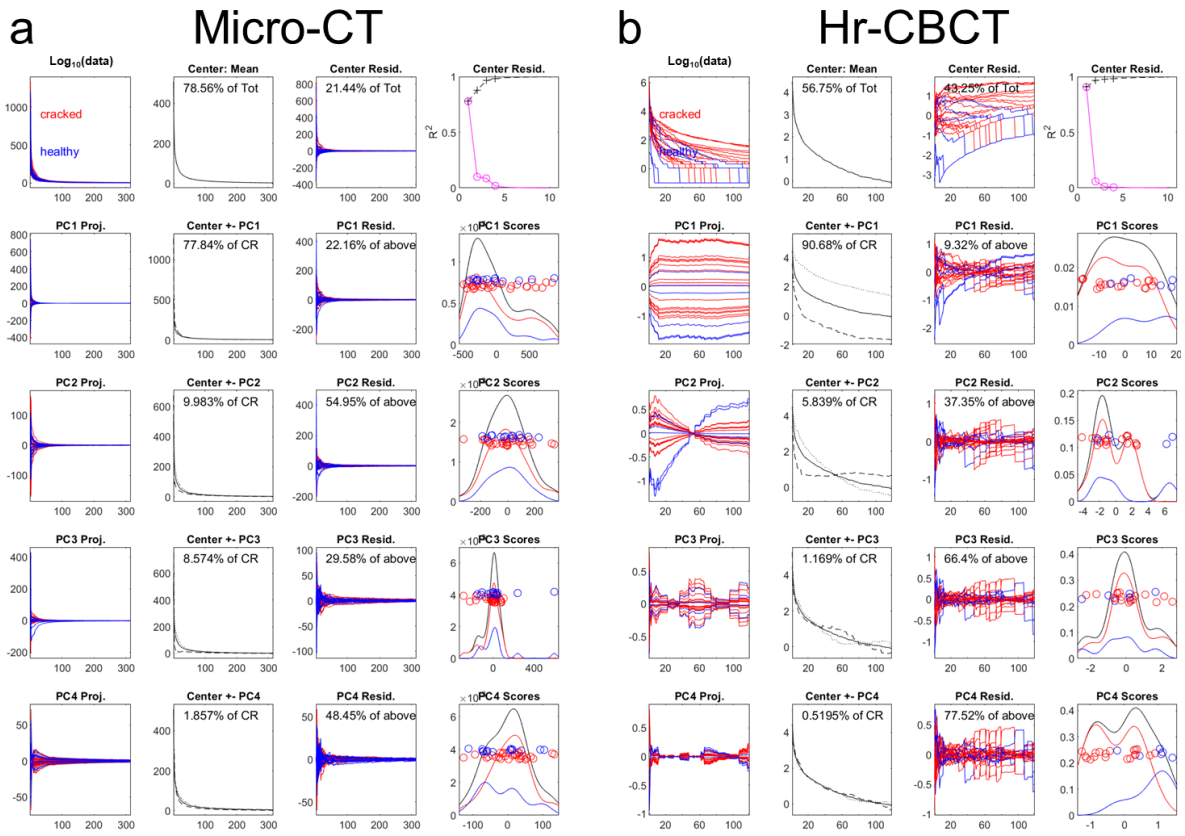


Figure 3. Visualization of modes of variation using principal component analysis for micro-CT (a) and hr-CBCT (b). Each line in column 1 and 3, and each dot in column 4 represents a tooth sample, either cracked (red) or healthy (blue). In hr-CBCT but not micro-CT, there is some separation between cracked vs. healthy teeth in the principal component directions.

Since multiple PC directions showed promising separation of teeth conditions, we further plotted hr-CBCT data on combinations of PC directions to see if it would improve the visual separation (Fig. 4). As expected, data projected onto a combination of PC1 and PC4 directions (Fig. 4, first row, last column) showed improved visual separation between teeth conditions compared to that onto a single PC direction, although no strong clustering was observed probably due to small sample sizes (Fig. 3).

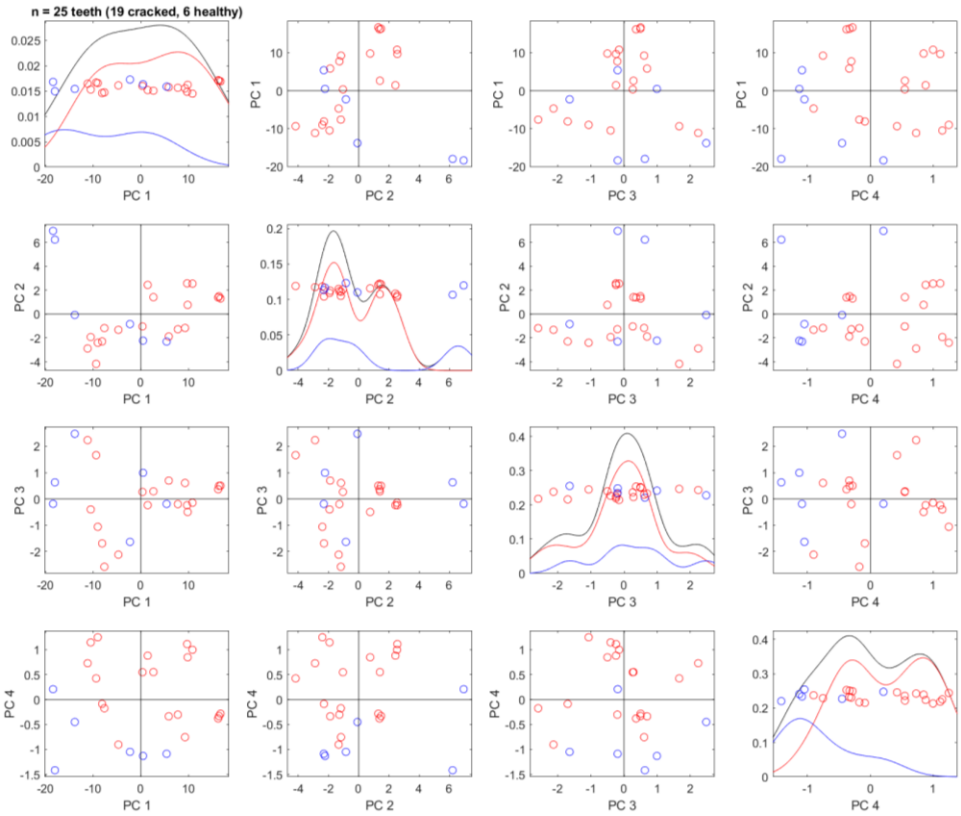


Figure 4. For hr-CBCT, projection of data onto combinations of PC directions. Sub-density estimates in the 1-d distributions for cracked teeth (red), healthy teeth (blue), and all teeth (black) are on the diagonal. PC1 and PC4 combination shows improved separation of cracked (red) and healthy (blue) teeth.

It should be noted that PCA requires each sample to have the same number of features. Since this is not the case for our dataset, we had been using the median number of features (i.e. cracks) in each dataset as a cutoff. To be more specific, we only used the first 119 features (i.e. the median number of features) in the hr-CBCT dataset for samples with more features, and zero-padded the samples with fewer features. An alternative way to solve this problem is to use a number of quantiles, which captures the trend of data as well as ensures the same number of features (i.e. quantiles) across samples. Here we used 100 quantiles (i.e. percentile) to represent the data (Supp. Fig.1). Using PCA, the modes of variation of quantiles (Supp. Fig. 2) showed a very similar pattern to what we saw in Fig. 3. This demonstrates the robustness of our analysis and supports our finding that hr-CBCT performs better in separating teeth conditions than micro-CT.

4. *Hr-CBCT but not micro-CT shows separation of means between the two teeth conditions in later quantiles of data.*

To visualize a large number of variables and identify which ones contribute more to the structure of data and the separation of the groups, a marginal distribution plot was used to visualize the property of quantiles (Fig. 5). Since we have 100 quantiles in total, we selected a representative subset of these quantiles to visualize (e.g. the 1st, 8th, …93th, 100th quantiles). Here we focused on the mean of quantiles, since no additional information was found for this analysis applied on other statistics (e.g. skewness, max, min, and standard deviation).

Both micro-CT and hr-CBCT showed quantile means sorted in increasing order (Fig. 5 upper-left plot). As a result, the variable number on the x-axis corresponded to the quantile number. It should be noted that the first half of these averages were much smaller than the second half and the value increased dramatically towards later variables. This is consistent with what we saw in the data (Fig. 2 and Supp. Fig. 1) where later quantiles had significantly larger values than early quantiles. For the first three rows of variables, the means for cracked (red dots) and healthy (blue dots) teeth seemed to overlap for both micro-CT and hr-CBCT (Fig. 5). However, for the last row of variables, hr-CBCT showed some separation between the two teeth conditions, indicating that the 80th to 100th percentile of the data might be more important in differentiating the teeth conditions. This is consistent with our intuition that cracks with larger sizes are more likely to be real cracks and to indicate the tooth condition correctly.

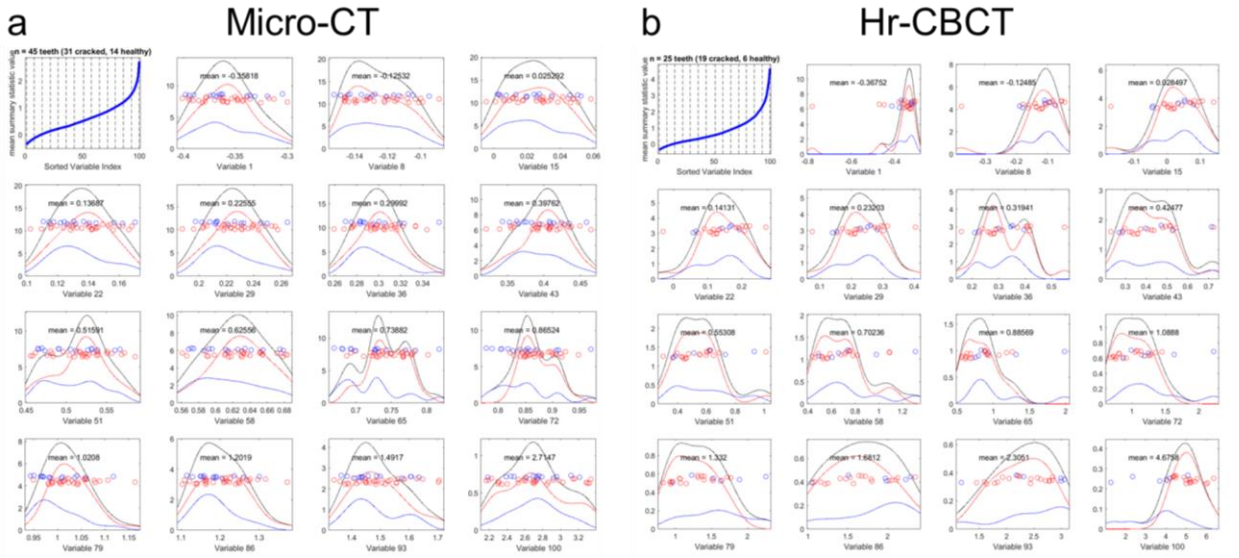


Figure 5. Marginal distribution plot of quantiles for micro-CT (a) and hr-CBCT (b) for cracked (red), healthy (blue), and all (black) teeth. The means of 1-100th quantiles are shown in the upper left panel. Marginal distributions of a representative (equally spaced) set of quantiles, indicated as vertical dashed lines, appear in the remaining panels. Micro-CT did not show separation of teeth conditions in all quantiles shown. Hr-CBCT showed that later quantiles are better at separating the two teeth conditions.

5. *Quantitative testing revealed a significant separation of the two teeth conditions using hr-CBCT but not micro-CT*

In order to quantitatively test if the two teeth conditions are separable, we chose distance-weighted discrimination (DWD) as our classification method to separate the healthy and cracked teeth using the crack-size curves. Statistical significance of this difference was evaluated using the direction-projection-permutation (DiProPerm) hypothesis tests. DWD was developed as an improved version of support vector machines (SVM) for linear classification.[7] DiProPerm[8] is a permutation-based hypothesis test that assesses the chance that the observed degree of separation happened as a result of expected random variation. It was developed with DWD in mind as an area of application, but it represents a general framework of nonparametric hypothesis testing built to discern visually discovered typical and atypical behavior in high-dimensional settings.

Figure 6.a and b showed the results of classifying cracked (red dots) and healthy (blue dots) in the micro-CT dataset, illustrated by projecting each tooth in the DWD and its first orthogonal principal component directions (OPC1). Kernel density estimates of the data projected onto the DWD separating hyperplane showed no clear separation between healthy and cracked teeth (Fig. 6.b, left). Further, DiProPerm test confirmed no significance separation between the two groups (Fig. 6.b, right, p = 0.74). Figure 6.c and d show the results of classification in the hr-CBCT dataset. Unlike in the micro-CT dataset, it shows visual separation of cracked versus healthy teeth using the same plane projection (Fig. 6.d, left). DiProPerm also showed that this separation is statistically significant (Fig. 6.d, right, p = 0.027) even for the small sample size. This confirms quantitatively that the crack-size curves derived from hr-CBCT, but not micro-CT, is able to separate teeth conditions.

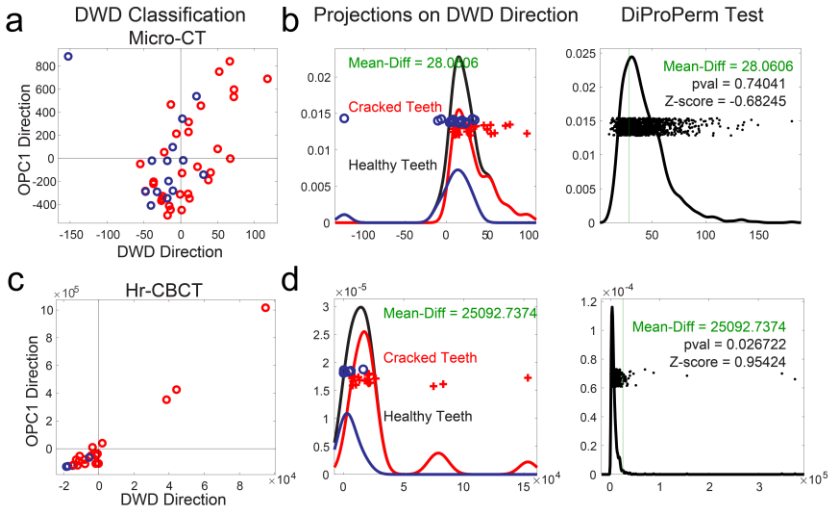


Figure 6. DWD classification of teeth conditions for micro-CT (a) and hr-CBCT (c). Each data sample is projected onto the DWD direction as well as its orthogonal principal component 1 (OPC1) direction. Kernel density estimates of the data projections onto the DWD separating hyperplane and permutation tests using DiProPerm for micro-CT (b) and hr-CBCT (d). Demonstrates a significant separation of the two teeth conditions using hr-CBCT but not micro-CT.

The same analysis was applied to the percentile representation of this data, but neither micro-CT nor hr-CBCT showed significant separation of the two teeth conditions using DiProPerm test (Fig. 7). This might suggest that for healthy teeth, it is important to take into account the fact that they have fewer cracks (using feature number cutoff as in Fig. 6) which is not reflected in the quantile representation of the data (Fig. 7).

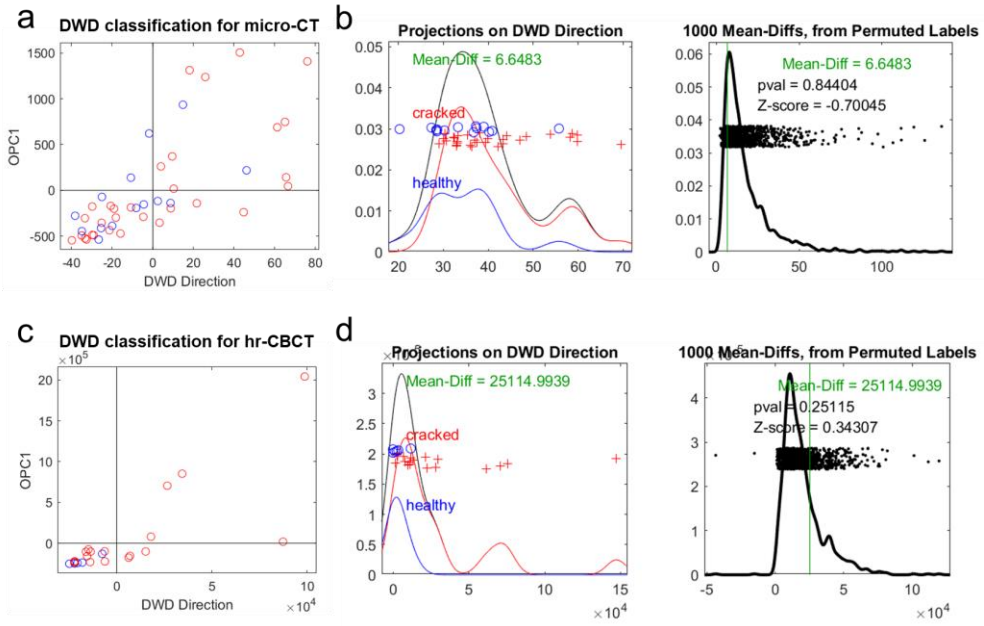


Figure 7. DWD classification for micro-CT (a) and hr-CBCT (c) using quantile data. Each data sample is projected onto the DWD direction as well as its orthogonal principal component 1 (OPC1) direction. Kernel density estimates of the data projections onto the DWD separating hyperplane and permutation tests using DiProPerm for micro-CT (b) and hr-CBCT (d). No significant separation of the two groups using either hr-CBCT or micro-CT.

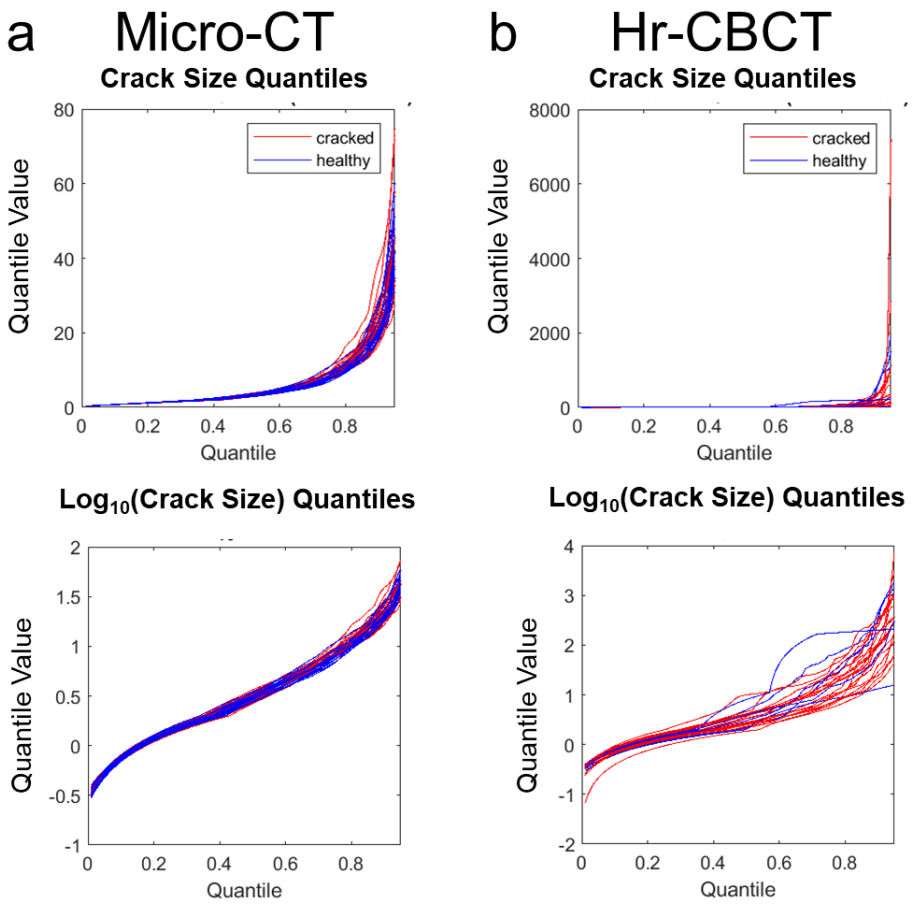
Discussion

Results with this data suggest that combining hr-CBCT and an advanced analysis pipeline can achieve good detection of cracked teeth. It can be applied to hr-CBCT clinically when the images are not over-processed (i.e. without a smoothing filter). In future work, it appears that further thresholding for the presence of large connected voxels can help flag potential cracks. This will provide the clinician additional visual information and confidence to determine if the highlighted area is a crack or not.

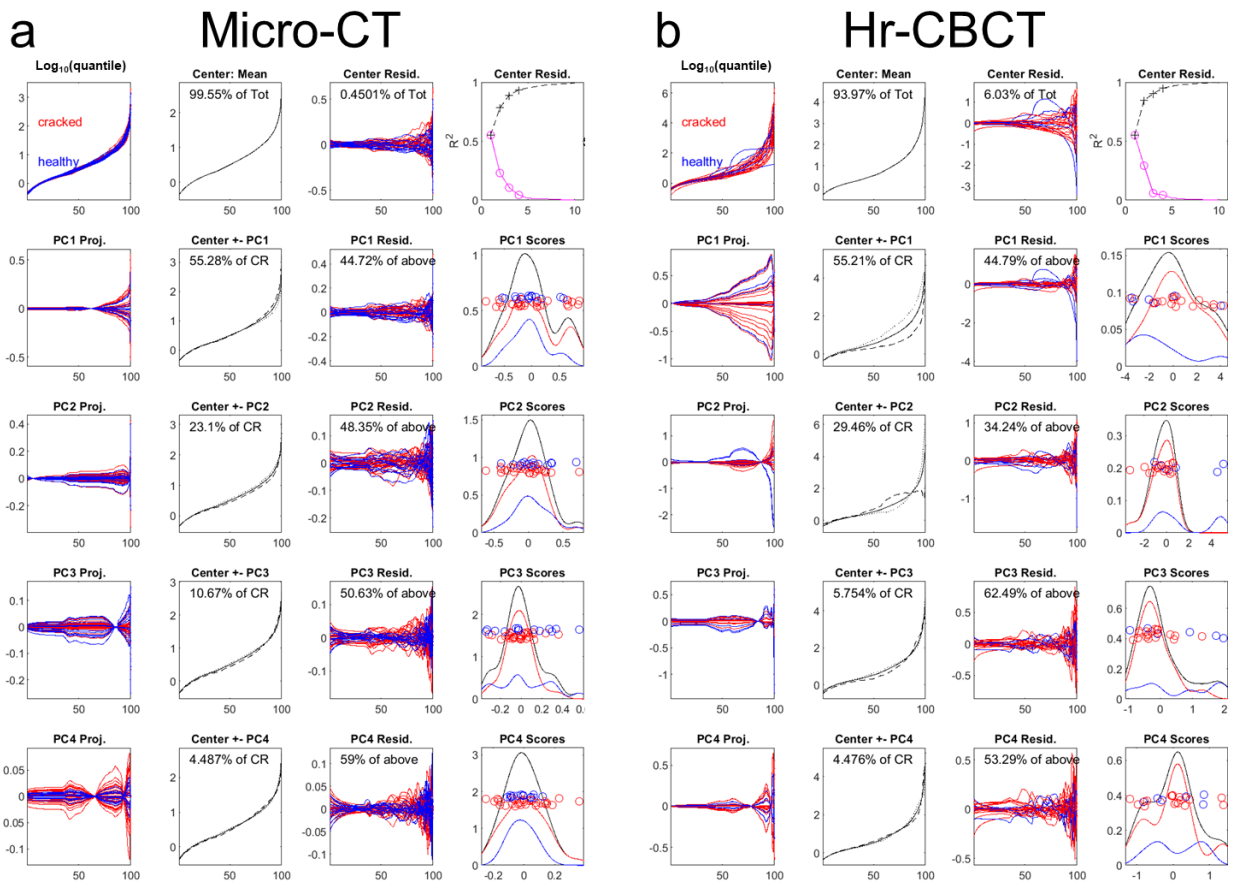
It needs to be noted that the hr-CBCT data was synthetic since the hr-CBCT dataset without filter was not available for this project. In the future, it would be ideal to acquire original images without additional filters and run them through the same pipeline to compare the results. In addition, since different settings and datasets were used, no direct comparison was made between the two imaging methods. To quantitively compare the two methods, it would be important to control for other factors (eg. the setting for teeth

presentation) and ideally using the same teeth samples. At last, the sample size for hr-CBCT is relatively low. There are only 6 samples of healthy control teeth, but there are a greater number of features. A larger sample size would be better for identifying key features in distributions and would be more robust to outliers.

Appendix



Supplemental Figure 1. Distribution of 100 quantiles of crack sizes for micro-CT (a) and hr-CBCT (b). Each line represents one tooth, either cracked (red) or healthy (blue), with its quantiles in ascending order. For each dataset, the top plot displays the raw quantile data and the bottom plot displays data in logarithmic scale. For both datasets, logarithmic scale provides better visualization of the distribution. Hr-CBCT shows qualitatively better separation of quantile distributions between cracked and healthy teeth. (Related to Fig. 2)



Supplemental Figure 2. Visualization of modes of variation using principal component analysis for micro-CT (a) and hr-CBCT (b) for 100 quantiles. Each line in column 1 and 3, and each dot in column 4 represents a tooth sample, either cracked (red) or healthy (blue). In hr-CBCT but not micro-CT, there is some separation between cracked vs. healthy teeth in the principal component directions. (Related to Fig. 3)

References

1 Lynch, C. D. & McConnell, R. J. The cracked tooth syndrome. *Journal (Canadian Dental Association)* **68**, 470-475 (2002).

2 Ricucci, D., Siqueira Jr, J. F., Loghin, S. & Berman, L. H. The cracked tooth: histopathologic and histobacteriologic aspects. *Journal of endodontics* **41**, 343-352 (2015).

3 Portman-Lewis, S. An analysis of the out-of-hours demand and treatment provided by a general dental practice rota over a five-year period. *Primary Dental Care* **14**, 98-104 (2007).

4 Vicory, J. *et al. Dental microfracture detection using wavelet features and machine learning*. Vol. 11596 MI (SPIE, 2021).

5 Shah, H. *et al.* Automatic quantification framework to detect cracks in teeth. *Proc SPIE Int Soc Opt Eng* **10578**, 105781K, doi:10.1117/12.2293603 (2018).

6 Marron, J. S. Object Oriented Data Analysis. doi:https://marron.web.unc.edu/stor-881-object-oriented-data-analysis-fall-2019/ (2019).

7 Marron, J. S., Todd, M. J. & Ahn, J. Distance-Weighted Discrimination. *Journal of the American Statistical Association* **102**, 1267-1271, doi:10.1198/016214507000001120 (2007).

8 Wei, S., Lee, C., Wichers, L. & Marron, J. S. Direction-Projection-Permutation for High-Dimensional Hypothesis Tests. *Journal of Computational and Graphical Statistics* **25**, 549-569, doi:10.1080/10618600.2015.1027773 (2016).