Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2019 February ; 10953: . doi:10.1117/12.2507978.

Advanced statistical analysis to classify high dimensionality textural probability-distribution matrices

Jack Prothero^a, Jean-Baptiste Vimort^b, Antonio Ruellas^c, J.S. Marron^a, Matt McCormick^b, Pablo Hernandez-Cerdan^b, Lucia Cevidanes^c, Erika Benavides^c, and Beatriz Paniagua^b

^aDept. of Statistics and Operations Research, University of North Carolina at Chapel Hill, Hanes Hall, Campus Box 3260, Chapel Hill, NC, USA 27599

^bKitware, Inc. 101 East Weaver Street, Carrboro, NC, USA 25710

^cDept. of Orthodontics and Pediatric Dentistry, University of Michigan, 1011 N University Ave, Ann Arbor, MI, USA 48109

Abstract

Temporomandibular Joint (TMJ) Osteoarthritis (OA) is associated with significant pain and disability. It is really hard to diagnose TMJ OA during early stages of the disease. Subchondral bone texture has been observed to change in the TMJ early during TMJ OA progression. We believe that raw probability-distribution matrices describing image texture encode important information that might aid diagnosing TMJ OA. In this paper we present novel statistical methods for High Dimensionality Low Sample Size Data (HDLSSD) to test the discriminatory power of probability-distribution matrices in computed from TMJ OA medical scans. Our results, and comparison with previous results obtained from the summary features obtained from them indicate that probability-distribution matrices are an important piece of information provided by texture analysis methods and should not be down sampled for analysis.

Keywords

Texture analysis; Statistical analysis; High Dimensionality Low Sample Size Data; Temporomandibular Joint Osteoarthritis; Subchondral bone remodeling

1. INTRODUCTION

Osteoarthritis (OA), the most prevalent arthritis worldwide, is associated with significant pain and disability and affects 13.9% of adults at any given time[1,2]. OA affects the temporomandibular joint (TMJ), among other joints. To date, there is no single sign, symptom, or test that can clearly diagnose early stages of the disease. It has been observed that often times, changes in the TMJ bone occur in early stages of this disease, involving structural changes both in the texture of the bone marrow and the subchondral cortical plate (see figure 1). We have successfully measured subchondral bone texture in the 3D space, and we know it can potentially provide an objective measure of subchondral bone architecture quality [3] using Gray Level Co-occurrence Matrices[4] (GLCM) and Gray Level Run Length Matrices[5] (GLRLM). However, we believe that raw probability-distribution matrices represented in GLCM and GLRLM encode important information that is lost when

computing summary features. In this paper we present statistical methods for High Dimensionality Low Sample Size Data (HDLSSD) to test the discriminatory power of probability-distribution matrices compared with summary features obtained from them.

2. DATA

We will use 16 condylar bone specimens, obtained from 12 patients with diagnosis of TMJ OA who were treated with surgical resection. From those 16 condylar bone specimens, we have 4 pairs of specimens that come from the same patient (left and right condyles. Imaging of all specimens were acquired using μ CT and hr-CBCT. For further information on this dataset see our previously published study in [3].

3. METHODS

3.1 Image analysis

In order to discriminate healthy/damaged sub-images we computed two types of matrices each one characterizing a different aspect of the global image. GLCM, represent a two-dimensional histogram where both entries are the voxel values. The GLCM encodes the number of occurrences of each voxel value combination in the whole image and this in every spatial direction. GLRLM, on the other hand, represent a two-dimensional histogram where one entry is the size of each run (set of consecutive, collinear voxels having the same grey-level value in one direction) and the other entry the intensity of each run. The GLRLM encodes the number of grey-level runs for each run size and voxel intensity for the whole image, in every spatial direction. In other terms, it describes the distribution of regions with same voxel intensity and different sizes in the image.

3.2 Statistical analysis

Throughout this summary, matrix data will be displayed both as a colored grid with higher intensity colors indicating values of larger magnitude (see figure 1.left.d-f and 1.right.b for CBCT-GLCM, and subsequent respective figures) and as a rasterized curve determined by the entries in each matrix (figure 2 a-c and figure 3a for CBCT-GLCM, and subsequent respective figures for the rest of the modalities). Our goal is to determine whether the probability-distribution matrices have comparable discriminatory power to the textural features extracted from them. We will examine each of the four sets of matrices separately and then compare their overall performance to prior analyses. This problem falls within the realm of high-dimension, low-sample-size classification, with each set of matrices constituting 35 observations of dimension 100 ($d=100$, this is so for the selected gray-levels and run-lengths bins in the computation of GLCM and GLRLM matrices). As such we chose distance-weighted discrimination (DWD) as our classification method and validated the results with direction-projection-permutation (DiProPerm) hypothesis tests. DWD was developed as an alternative binary linear classifier to support vector machines (SVM) [6]. The goal was to take more information into account when determining the separating hyperplane than just a small number of support vectors. The hyperplane, determined by a vector w and scalar β , is found by solving a second-order cone program of the form notated

equation 1 (describing the DWD optimization problem formulation), where each x_i is a d -dimensional observation, y_i is the class label of x_i (+1 or -1).

$$\min_{r, w, \beta, \xi} \sum_{i=1}^n \left(\frac{1}{r_i} + C\xi_i \right), r_i = y_i(x_i'w + \beta) + \xi_i \geq 0, \quad (1)$$

$$\xi \geq 0, \|w\|^2 \leq 1$$

The r_i are meant to represent the residuals from each data point to the hyperplane, so by minimizing the sum of reciprocals we maximize their distance to the hyperplane with more weight placed on points closer to the boundary. ξ is an error vector, penalized by a constant C in the objective, which allows for some data points to fall on the wrong side of the hyperplane if necessary. DiProPerm [7] 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 typical and atypical behavior in high-dimensional settings. To conduct the test, we randomly shuffle the class data labels and project it onto the DWD direction determined by the two new classes. We record the mean difference and repeat the process 1000 times to create an empirical distribution of mean differences. The p-value of the test is the percent of the empirical distribution larger than the mean difference between the original classes.

4. RESULTS

4.1 Analysis of GLCM computed from CBCT

We present the full set of curves for each data class alongside and matrix of overall mean behavior (see figure 2). The correspondence markings (figure 2.d) are included in the first column for clarity.

The interesting structure of these matrices is concentrated almost entirely along the diagonal; only 36 variables out of 100 have a nonzero entry in at least one observation. The damaged samples tend to display higher values at the edges of the diagonal whereas healthy samples reached higher peaks in the third and fourth diagonal entries. Figure 3 displays the results of a DWD test of the two classes. As expected, high values at the edges of the diagonal indicate a damaged observation and high values in the middle of the diagonal indicate a healthy observation. Also note that further down the diagonal, values in off-diagonal entries also seem to indicate a damaged observation. The DiProPerm hypothesis test for this data obtained a p-value=0.002. By this metric the separation between classes is quite significant and on the order of the significance of separation that we found with summary features obtained from the matrices before[3].

4.2 Analysis of GLRLM computed from CBCT

Rather than being concentrated along the diagonal, the information in the CBCT-GLRLM is concentrated in each row. The observations reach a peak at the beginning of each row and

peter off to the right side of the matrix. Figure 4 shows that the healthy observations seem to remain more intense across the third and fourth rows, while damaged observations have extra peaks in the sixth and seventh rows.

From these plots we can clearly determine which portions of the matrix best correlate with each class (see figure 5). Damaged observation peaks tend to be higher in the middle rows of the matrix and run longer on the top and bottom rows, while healthy observations exhibit the opposite behavior. The DiProPerm hypothesis test for this separation achieved a p -value=0.003; indicating that the class separation observed above is similarly significant to that of the CBCT-GLCM data.

4.3 Analysis of GLCM computed from μ CT

These matrices exhibit a combination of the behaviors observed in the CBCT data, some observations are concentrated along the diagonal, while others are periodic, peaking at the start of each row (see figure 6). The periodic behavior turns up in the trabecular mean plots more prominently than the damaged mean plots.

Plots in figure 7 seem to indicate that a damaged observation means high peak values are located strictly along the diagonal as well as in the bottom-right corner, and that other areas of the matrix tend to indicate a healthy observation. This regional association is weaker than what was observed in the CBCT matrices. The loadings curve is more erratic with lots of low-magnitude entries. The DiProPerm hypothesis test shows a non-significant separation between classes (p -value = 0.305).

4.4 Analysis of GLRLM computed from μ CT

Similarly to the CBCT-GLRLM data, the observations in this set of matrices follow a periodic pattern by row (figure 8). Based on the mean behaviors of each class, healthy observations seem to simply be larger on average across the board except in the last few rows. Figure 9 shows the results of DWD on this set of matrices.

As expected from the summarization, large peak values at the beginning of each row are associated with healthy observations. Less apparent from the summarization is the association of peak width with damaged values, especially in lower rows of the matrix (figure 9). The separation jitter plot does not appear as striking as those from CBCT matrices, but according to a DiProPerm hypothesis test the achieved separation is significant with a p -value = 0.032.

5. DISCUSSION

In this paper we present novel statistical analysis methods to analyze data that falls within the realm of high-dimension, low-sample-size classification. Our data samples consisted on probability-distribution matrices obtained from two well-known methods for texture analysis[4,5]. The results presented build upon previous work of our team[3], and present new and novel information based on this new data and methods. For a full comparison of the effectiveness of the summary matrices as compared to the extracted features, a table of the

DiProPerm p-values of the separation between the two classes for each subset of data is given below.

It appears that extracting features from the CBCT matrices loses some discriminatory information, but extracting features from the μ CT matrices reduces or works around some apparent noise. In conclusion, these results compared with previous results obtained from the summary features indicate that probability-distribution matrices are an important piece of information provided by texture analysis methods that yields higher discrimination power than summary features and should not be ignored for analysis.

REFERENCES

- [1]. Hussain AM, Packota G, Major PW and Flores-Mir C, "Role of different imaging modalities in assessment of temporomandibular joint erosions and osteophytes: a systematic review.," *Dentomaxillofac. Radiol* 37(2), 63–71 (2008). [PubMed: 18239033]
- [2]. Centers for Disease Control., "CDC Data on Osteoarthritis," 2017, <https://www.cdc.gov/arthritis/data_statistics/arthritis-related-stats.htm>.
- [3]. Vimort J-B, Ruellas A, Protrero J, Marron J, Shah H, Cevidanes L, Benavides E and Paniagua B, "Detection of bone loss via subchondral bone analysis," *Accept. to SPIE Med. Imaging* (2018).
- [4]. Haralick RM, Shanmugam K and Dinstein I, "Textural Features for Image Classification," *IEEE Trans. Syst. Man. Cybern* 3(6), 610–621 (1973).
- [5]. Galloway MM, "Texture analysis using gray level run lengths," *Comput. Graph. Image Process* 4(2), 172–179 (1975).
- [6]. Marron JS, Todd MJ and Ahn J, "Distance-Weighted Discrimination," *J. Am. Stat. Assoc* 102(480), 1267–1271 (2007).
- [7]. Wei S, Lee C, Wichers L and Marron JS, "Direction-Projection-Permutation for High-Dimensional Hypothesis Tests," *J. Comput. Graph. Stat* 25(2), 549–569 (2016).

Figure 1.

Three-dimensional representation of a temporomandibular joint condyle on the left and two different planar radiographical representations of subcondral bone from two different patients with and without arthritic changes.

Figure 2.

CBCT-GLCM Data Summary: a) Plot of all 35 curves, mean curves of b) damaged and c) healthy classes, d) overall mean matrix, and Mean matrices of e) damaged and f) healthy classes.

Figure 3.

Summary of DWD on CBCT-GLCM data. a) Raw data curves, with red indicating a damaged observation and blue indicating a healthy observation, b) matrix of DWD direction loadings. Green values in the matrices are positive, magenta values are negative. c) Curve of DWD direction loadings. Positive values point towards the damaged class, negative values point towards the healthy class, d) Jitter plot of the projection of the data onto the DWD direction, with smooth histograms for each population.

Figure 4. CBCT-GLRLM Data Summary: a) Plot of all 35 curves, mean curves of b) damaged and c) healthy classes, d) overall mean matrix, and Mean matrices of e) damaged and f) healthy classes.

Figure 5.

Summary of DWD on CBCT-GLRLM data. a) Raw data, b) matrix of DWD direction loadings, c) curve of DWD direction loadings, d) Jitter plot of DWD projection of the data.

Figure 6.
 μ CT-GLCM Data Summary: a) Plot of all 35 curves, mean curves of b) damaged and c) healthy classes, d) overall mean matrix, and Mean matrices of e) damaged and f) healthy classes.

Figure 7.

Summary of DWD on μ CT -GLCM data. a) Raw data curves, b) matrix of DWD direction loadings, c) curve of DWD direction loadings, d) Jitter plot of the projection of the data onto the DWD direction, with associated probability distributions.

Figure 8.

μ CT -GLRLM Data Summary: a) Plot of all 35 curves, mean curves of b) damaged and c) healthy classes, d) overall mean matrix, and Mean matrices of e) damaged and f) healthy classes.

Figure 9. Summary of DWD on μ CT-GLRLM data. a) Raw curves, DWD direction loadings in b) matrix and c) curve forms, d) Jitter plot of the projection of the data onto the DWD direction.

Table 1.

P-value summaries for matrices and extracted features.

	CBCT GLCM	CBCT GLRLM	μ CT GLCM	μ CT GLRLM
Matrices	0.002	0.003	0.305	0.032
Summary features	0.007	0.006	0.165	0.002