NON-INVASIVE DIFFERENTIAL DIAGNOSIS OF DENTAL PERIAPICAL LESIONS IN CONE-BEAM CT

Arturo Flores[1], *Steven Rysavy*[1], *Reyes Enciso*[2], *and Kazunori Okada*[1]

[1]San Francisco State University, San Francisco, CA 94132 USA
[2]University of Southern California, Los Angeles, CA 90089 USA
{aflores,rysavy,kazokada}@sfsu.edu,renciso@usc.edu

ABSTRACT

This paper proposes a novel application of computer-aided diagnosis to a clinically significant dental problem: non-invasive differential diagnosis of periapical lesions using cone-beam computed tomography (CBCT). The proposed semi-automatic solution combines graph-theoretic random walks segmentation and machine learning-based LDA and AdaBoost classifiers. Our quantitative experiments show the effectiveness of the proposed method by demonstrating 94.1% correct classification rate. Furthermore, we compare classification performances with two independent ground-truth sets from the biopsy and CBCT diagnoses. ROC analysis reveals our method improves accuracy for both cases and behaves more in agreement with the CBCT diagnosis than with biopsy, supporting a hypothesis presented in a recent clinical report.

***Index Terms*—** periapical lesion, CBCT, classification, Adaboost, LDA

1. INTRODUCTION

Dental periapical lesions can be classified as either cysts or granulomas [1]. Granulomas are composed of chronic inflammatory cells. Being the product of a necrotic pulp in a non-vital tooth, a granuloma must be treated by endodontic therapy or by tooth extraction. Apical cysts are lesions in which epithelial rests of Malassez are stimulated to proliferate and form a debris-filled central cavity. Removal of its cause (endodontic therapy) and surgical removal of the cyst is assumed to be curative. However, a granuloma may heal without surgical treatment if given the opportunity [2].

Differentiating between these types of lesions remains an open research problem. Standard treatment for these lesions includes surgical extraction and a *post hoc* biopsy for histological diagnosis. This is the only reliable method of diagnosis currently available. Since this procedure is *invasive*, the chance for the granuloma to heal by itself is lost. As a result the healing rate for granulomas remains largely unknown. Furthermore, the patient is subjected to potentially unnecessary surgery and associated complications, including infection and discomfort. Other methods of diagnosis, such as periapical radiographs [3] and papanicolou smears [4], have been unsuccessful in producing an accurate diagnosis.

A recent study by Simon et al.[5] has proposed a *non-invasive* diagnostic technique of periapical lesions using 3D cone-beam computed tomography (CBCT). This study used the lowest intensity value at the center of the lesion to differentiate between cysts and granulomas. Seventeen CBCT scans were evaluated in this study. These lesions were subsequently surgically extracted and biopsies were performed to ascertain histological diagnoses. In 13 cases, the CBCT and biopsy diagnoses coincide, but in the remaining four cases the CBCT and biopsy diagnoses differ. Inspection of the biopsy notes on these four split cases suggest the CBCT diagnosis may have been more accurate [5].

This study shows 3D CBCT is a good modality toward the differential diagnosis of these lesions. This non-invasive technique not only improves the quality of patient care but also allows for the granuloma's healing process to be studied. However, their method requires the endodontic expert to analyze the whole lesion manually in search of the cystic cavity, which is time consuming and prone to human error. Furthermore, the simple thresholding technique used is unreliable due to the low-dosage nature of the dental CBCT. Moreover, the voxel grey values are patient-specific and do not correspond 1:1 to Hounsfield units since the scanner adapts the dosage for each patient in order to minimize the radiation to the maxillofacial area.

Addressing these issues, this paper proposes a semi-automatic solution for the non-invasive differential diagnosis of these periapical lesions. The goal is to improve the approach of [5] in accuracy, repeatability and usability. We adapt the computer-aided diagnosis (CAD) approach, exploiting advanced computational algorithms. Our solution takes three user-specified click-points, indicating a region of interest (ROI) as an input, and outputs the differential diagnosis between cyst and granuloma.

One of the main contributions of this work is that to our knowledge, this is the first to adapt the CAD approach to the above clinical problem in dentistry using 3D CBCT. CAD has been successful in many high-impact clinical areas, includ-

ing lung nodule detection [6] and breast tumor diagnosis [7], providing medical professionals with a valuable second opinion. However, the field of dentistry has not yet fully benefited from the advancements of medical image analysis, despite how common the dental practice is to our daily life.

We also analyze our system with two independent ground-truth sets from biopsy and CBCT diagnoses. The results show our system improves accuracy for both cases and behaves more in agreement with the CBCT diagnosis. This indicates the CBCT scan may provide a safer, faster, and more accurate method for differential diagnosis of these lesions.

2. PROPOSED DENTAL CAD

The proposed system consists of the following three successive stages: 1) graph-theoretic lesion segmentation, 2) feature extraction of lesion intensity statistics, and 3) machine learning-based lesion classification using linear discriminant analysis (LDA) [8] and AdaBoost [9].We refer to this system as *dental CAD System*.

2.1. Lesion Segmentation

Our pilot study [10] applied three segmentation algorithms to the dental CBCT data. This study showed the random walks algorithm by Grady [11] performs better than other segmentation methods on our data. We adapt this method to our 3D lesion segmentation problem. This method is also refined to improve final classification performance [12].

We also implement a method to place initial seeds in 3D with minimal user-interaction. The method specifies two concentric spheres. It then places foreground and background seed labels along the surface of the inner and outer sphere, respectively. Three parameters are required to generate these seed points: the center of the spheres, an inner radius and an outer radius. The inner radius specifies a sphere small enough to be completely contained within the lesion. Likewise, the outer radius specifies a sphere large enough to entirely enclose the lesion. These parameters are specified via mouse clicks. Figure 1 shows an example segmentation.

2.2. Feature Extraction

The next step is to extract a set of intensity statistics from each segmented lesion as a feature vector. In [5] only the minimum intensity value at the center of the lesion was used for classification. In this study, we consider a set of eight features computed from the lesion's intensity distribution: maximum, minimum, mean, median, standard deviation, skewness, kurtosis, and entropy.

2.3. Lesion Classification

The final step is to classify the lesion using the features extracted in the previous step. In this study, we compare three classification methods: a simple threshold classifier using a single feature, Linear Discriminant Analysis (LDA), and AdaBoost.

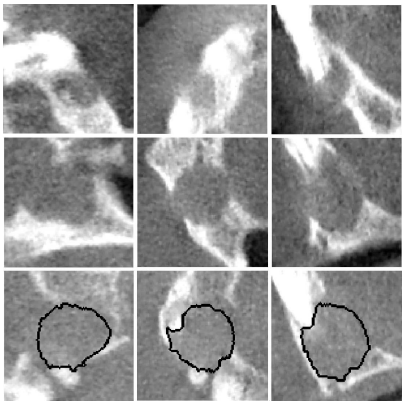

Fig. 1. Example of periapical lesions and 3D segmentation by the random walks algorithm. Sagittal, coronal, and axial view of a granuloma (top) and a cyst (middle) are displayed. Visually, they are hard to differentiate. Bottom row shows an example of 3D segmentation by our method.

LDA seeks a basis that maximizes the ratio of between-class (S_B) and within-class (S_W) scatter [8]. The definitions of the scatter matrices are given by $S_B = \sum_c N_c(\mu_c - \overline{x})(\mu_c - \overline{x})^T$ and $S_W = \sum_c \sum_{i \in C}(x_i - \mu_c)(x_i - \mu_c)^T$, where sample mean and class-mean are defined as $\overline{x} = \frac{1}{N}\sum_i x_i = \frac{1}{N}\sum_c N_c \mu_c$ and $\mu_c = \frac{1}{N_c}\sum_{i \in C} x_i$. N_c is the number of samples in class c.

AdaBoost [9] is an ensemble learner in which the joint decision rule of multiple weak classifiers form a final strong classifier. Adaboost assigns an initial uniform weight to each training sample. After each boosting iteration, weights for correctly classified samples are decreased and weights for incorrectly classified samples are increased. This allows Adaboost to concentrate on the informative and difficult training samples. The resulting classifier of the Adaboost algorithm is given by $H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$, where α_t is the weighted error, h_t is the t-th weak classifier hypothesis, and $H(x)$ is the strong classifier hypothesis.

3. DATA

Seventeen anonymous 3D dental CBCT scans used in [5] were available for our study in the DICOM format. The CBCT images were captured using the NewTom 3G scanner. This study was approved by the institutional review board. Each 0.2mm axial slice is a 512×512 image with a 12 bit intensity range ($4,096$ grayscale). An ROI of 100 cubic voxels is pre-computed for each lesion to run our experiments efficiently. Fig.1 shows an example. Each lesion underwent a histological biopsy and was diagnosed as either a cyst or granuloma by a certified oral pathologist. Note we treat not only these biopsy results but also the CBCT diagnostic results by endodontic experts in [5] as our ground-truth labels. In

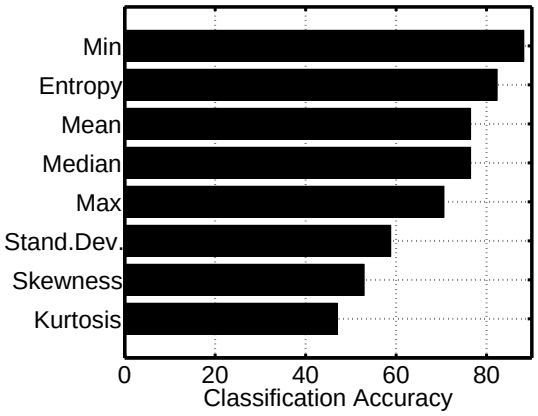

Fig. 2. Classificatuion accuracy results using a simple threshold classifier on each individual feature.

13 cases the CBCT and biopsy diagnoses coincide, but in the remaining four cases the CBCT and biopsy diagnoses differ. Section 4.2 discusses how this uncertainty in labels influences the classification performance.

Recall that intensity values are inconsistent between CBCT scan instances. This is due to the adaptive shifting of radiation to minimize radiation dosage to patients. Simple threshold segmentation and classification is therefore not a viable option. Furthermore, the morphology of the periapical lesions can vary. As can be seen in Fig.1, it is non-trivial to segment or classify these lesions.

4. EXPERIMENTS

Due to the small training data set, leave-one-out cross-validation (LOOCV) [13] is used to validate the classification experiments. We use 16 samples to train each classifier then test the classifier on the final sample. This process is repeated until each sample is used as the test sample.

4.1. Classifier comparison

For each single feature, we first evaluate a simple threshold classifier to study the nature of our feature space. The CBCT diagnoses in [5] are used as ground-truth. This test reveals for a simple threshold classifier, minimum intensity yields the best classification accuracy, which agrees with Simon's initial study. See Fig.2 for complete results.

We compare the classification accuracy of simple threshold, LDA, and AdaBoost classifiers using LOOCV. All permutations of two (2nd column) and three (third column) features, and finally all features (fourth column) are considered. The LOOCV results are averaged and summarized in Table 1. The individual classifier with highest accuracy is AdaBoost using the feature pair of median and minimum intensity, resulting in 94.1% LOOCV success rate.

This result suggests a non-linear classifier such as Adaboost is a better choice for this data. Adaboost consistently performed better than LDA in the three different feature sets, although the difference in performance between Adaboost and LDA is not statistically significant. A more interesting result is the comparison of classifier performance using CBCT and biopsy diagnoses as ground truth.

4.2. Ground Truth Analysis

In our previous experiments, we use the CBCT diagnosis by endodontic experts as our ground-truth because [5] suggests the histological findings for the periapical lesions can be misleading. In order to test this hypothesis, we compare the performance of our dental CAD system trained with the two ground-truth sets. We perform the ROC analyses with the LDA classifiers. The LOOCV results of the best performing feature combinations are averaged at varying threshold values, Fig.3 shows the results. Visual inspection of the ROC curves clearly indicate the classifiers with the CBCT ground-truth perform better than those with the biopsy ground-truth. Using biopsy as ground-truth, the best performance of 88.2% success rate was achieved by AdaBoost, using entropy, kurtosis, and skewness as the feature vector.

Fig.4 also compares the two ground-truths according to the correct classification rates during LOOCV for the four split cases, and all cases, over all classifiers tested in Table 1. The results further suggest our dental CAD system agrees more with the CBCT diagnosis than with the biopsy diagnosis, lending further support to the claim that the CBCT diagnosis is the correct diagnosis. Suppose the CBCT diagnosis is our gold standard, then the biopsy diagnosis is correct on 13 out of 17 cases, or 76.5%, and vice versa. In this case, our method with 94.1% success rate can be interpreted as an almost 20% improvement from 76.5% accuracy over the biopsy diagnosis. On the other hand, if we take the biopsy as our gold standard, our approach with 88.2% success rate still achieves a performance increase of 10%. These results demonstrate the advantage of our approach for improving accuracy while not requiring intensive manual labor by the endodontic experts.

5. CONCLUSION

This paper presents our dental CAD system which integrates advanced segmentation and classification algorithms. Our experimental studies show the proposed method effectively per-

Table 1. Comparison of classification accuracy of simple threshold (ST), LDA, and AdaBoost classifiers using LOOCV.

Classifier	Pair	Triple	All
ST			69.12 ± 14.65
LDA	70.17 ± 16.01	75.72 ± 15.70	73.08 ± 15.76
Adaboost	71.85 ± 12.23	76.58 ± 8.22	74.05 ± 9.76

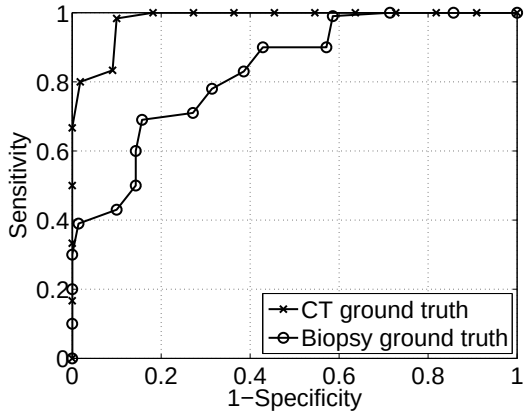

Fig. 3. Ground Truth Analysis. ROC curves generated using CBCT and Biopsy as ground truth.

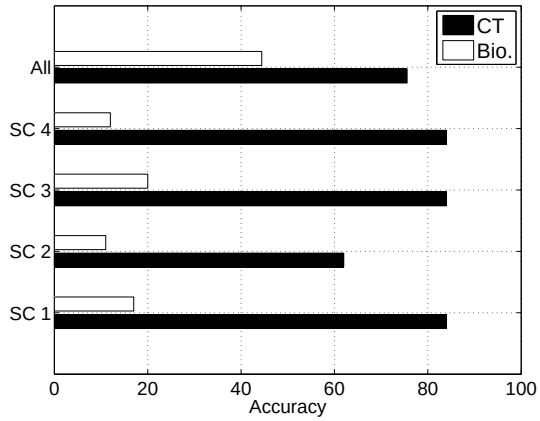

Fig. 4. Classification accuracy comparison using CBCT and biopsy as ground truth. Note the classification accuracy is significantly lower when using biopsy as ground truth.

forms a non-invasive differential diagnosis of enlarged periapical lesions with 94.1% and 88.2% accuracy when we use the CBCT and biopsy diagnosis as the gold standard, respectively. The semi-automatic approach presented in this paper may prove to be more accurate, reproducible, less prone to human error. Furthermore, the ground truth analysis reveals our dental CAD system agrees more with the CBCT diagnosis, supporting the hypothesis made in Simon's initial study.

As future work, we plan to conduct quantitative experiments with more data and extend our ground-truth analysis to correlate more than two ground-truth sets. Such data is currently being collected. Designing a problem specific feature can also boost our classifier's performance. To this end, we plan to incorporate distance-based features to better analyze the voxels surrounding the tip of the periapical root.

6. REFERENCES

[1] Mervyn Shear and Paul Speight, *Cysts of the Oral and Maxillofacial Region*, Oxford: Wright, 3rd edition, 1992.

[2] James Simon, "Incidence of periapical cysts in relation to the root canal," *Journal of Endodontics*, vol. 6, no. 11, pp. 845–848, 1980.

[3] J McCall and Wald S, *Clinical Dental Radiology, 4th Ed.*, pp. 234–251, Philadelphia: Saunders, 1954.

[4] F Howell and V De la Rosa, "Cytologic evaluation of cystic lesions of the jaws: a new diagnostic technique," *Calif Dent Assoc*, vol. 36, pp. 161–166, 1968.

[5] James Simon, Reyes Enciso, Jose-Maria Malfaz, Michelle Bailey-Perry, and Anish Patel, "Differential diagnosis of large periapical lesions using cone-beam computed tomography measurements and biopsy," *Journal of endodontics*, vol. 32, no. 9, pp. 833–837, 2006.

[6] K. Kanazawa, M. Kubo, N. Niki, H. Satoh, H. Ohmatsu, K. Eguchi, and N. Moriyama, "Computer aided diagnosis system for lung cancer based on helical CT images," in *Int. Conf. Pattern Recognition (ICPR '96) Volume III*, 1996, p. 381.

[7] Yu-Len Huang and Kao-Lun Wang, "Diagnosis of breast tumors with ultrasonic texture analysis using support vector machines," *Neural Computing and Applications*, vol. 15, no. 2, pp. 164–169, 3 2006.

[8] R. O. Duda, Hart P. E., and D. G. Stork, *Pattern Classification*, Wiley-Interscience Publication, 2000.

[9] Yoav Freund and Robert E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *European Conference on Computational Learning Theory*, 1995, pp. 23–37.

[10] Steven Rysavy, Arturo Flores, Reyes Enciso, and Kazunori Okada, "Segmentation of large periapical lesions toward dental computer-aided diagnosis in cone-beam ct scans," in *SPIE Medical Imaging*, San Diego, 2008, pp. 6914–153.

[11] Leo Grady, "Random walks for image segmentation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.

[12] Steven Rysavy, Arturo Flores, Reyes Enciso, and Kazunori Okada, "Classifiability criteria for refining of random walks segmentation," in *ICPR 2008*.

[13] Ron Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *IJCAI*, 1995, pp. 1137–1145.