Computerized Medical Imaging and Graphics 38 (2014) 151–162

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Automated classification of four types of developmental odontogenic cysts

A. Frydenlund [a], M. Eramian [a,*], T. Daley [b,1]

[a] Department of Computer Science, University of Saskatchewan, 110 Science Place, Saskatoon, SK, Canada S7N 5C9
[b] Department of Pathology, Schulich School of Medicine and Dentistry, Dental Sciences Building, Room 4044, University of Western Ontario, London, ON, Canada N6A 5C1

ARTICLE INFO

Article history:
Received 25 July 2013
Received in revised form 31 October 2013
Accepted 2 December 2013

Keywords:
Odontogenic cysts
Classification
Oral cysts
Dentigerous
Keratocyst
Lateral periodontal
Glandluar
Machine learning
Image processing

ABSTRACT

Odontogenic cysts originate from remnants of the tooth forming epithelium in the jaws and gingiva. There are various kinds of such cysts with different biological behaviours that carry different patient risks and require different treatment plans. Types of odontogenic cysts can be distinguished by the properties of their epithelial layers in H&E stained samples. Herein we detail a set of image features for automatically distinguishing between four types of odontogenic cyst in digital micrographs and evaluate their effectiveness using two statistical classifiers – a support vector machine (SVM) and bagging with logistic regression as the base learner (BLR). Cyst type was correctly predicted from among four classes of odontogenic cysts between 83.8% and 92.3% of the time with an SVM and between $90 \pm 0.92\%$ and $95.4 \pm 1.94\%$ with a BLR. One particular cyst type was associated with the majority of misclassifications. Omission of this cyst type from the data set improved the classification rate for the remaining three cyst types to 96.2% for both SVM and BLR.

1. Introduction

Tooth formation involves a complex interaction of the enamel organ and surrounding connective tissues. Occasionally, the remnants of the enamel organ can give rise to a variety of odontogenic tumours and cysts. The rarity of these pathologies means that proper diagnosis is often the responsibility of subspecialists. Proper diagnosis of odontogenic cysts is crucial because different biological behaviours of the various types of cysts require different treatment plans, and present significantly different risks to patients.

As was argued in [7], there is a need for an accurate computer assisted diagnostic protocols to reduce the workload of oral pathologists and to potentially reduce the costs of multiple expert diagnostic opinions; this was the main motivation for that paper and remains the primary motivation for the current work. Odontogenic cysts are distinguished by histologic examination of characteristics of their epithelial layer.

Most research on classification of odontogenic cysts have been based on histopathological features and other clinical considerations which are observed and judged by humans; Phillipsen and Reichart provide a good review of these works [17]. Some image analyses of odontogenic cysts have been reported. The appearance of odontogenic cysts in other imaging modalities such as MR has been studied [11]. Li et al. used imaging analysis techniques to quantify PCNA+ cells in odonotogenic keratocysts and radicular cysts [14]. Landini used image processing to segment H&E stained histopathological images of odontogenic keratocysts and radicular cysts into theoretical cell regions using a semi-automated algorithm and derived various morphometric features of individual cells which proved to be successful in distinguishing radicular cysts from odontogenic keratocysts (95% correct classification), but not subtypes of keratocysts (60% correct classification) [13]. Han et al. reported another comparison of two types of odontogenic keratocyst and radicular cysts in which a cascade of Haar classifiers is used [9]. They reported classification rates similar to that of Landini's previous experiment, namely 100% for distinguishing between keratocysts and radicular cysts, but 60% for distinguishing between the two types of keratocyst. Our work is novel in that we combine properties of Landini's theoretical cells with morphological, spectral, and textural features of the entire visible epithelial region and that we can distinguish between four different types of developmental odontogenic cyst, three of which have not been considered in previous work.

As a first step towards a computer assisted diagnostic system for classifying odontogenic cysts, we presented in [7] an

* Corresponding author. Tel.: +1 306 966 4028; fax: +1 306 966 4884.
E-mail addresses: aaf955@mail.usask.ca (A. Frydenlund), eramian@cs.usask.ca (M. Eramian), Tom.Daley@schulich.uwo.ca (T. Daley).
[1] Tel.: +1 519 661 2111x86405.

0895-6111/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compmedimag.2013.12.002

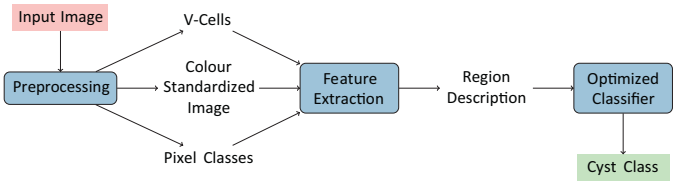

Fig. 1. Cyst classification system overview.

automated epithelial segmentation algorithm for hematoxylin and eosin (H&E) stained digital microscopy images of developmental odontogenic cysts. The algorithm was evaluated using a data set consisting of four types of cysts: dentigerous cysts (DCyst), odontogenic keratocysts (OKC), lateral periodontal cysts (LPC), and glandular odontogenic cysts (GOC). A review of the details of these pathologies, their prevalence, and their radiographic appearance is presented in [7].

Herein we present our second step towards a fully automated algorithm for classifying developmental odontogenic cysts. The assumption is made that the epithelial regions of digital images of odontogenic cysts can be identified with perfect accuracy using an automated algorithm such as that presented in [7]. We propose a set of image features that can be computed from such epithelial regions to form descriptions of the regions and show that the proposed region descriptions can be used to distinguish between and accurately classify samples of the four types of developmental odontogenic cysts mentioned above using standard classification algorithms.

This study serves as a proof of principle that there exist quantitative features of odontogenic cyst epithelia which can be computed from digital images that can be used to distinguish between types of odontogenic cyst. There are additional types of odontogenic cysts beyond those studied herein that would need to be distinguishable by our system in order for it to have practical value as a diagnostic tool, but those cysts present additional challenges that are beyond the scope of this initial, proof-of-principle study; these challenges and future research directions are discussed in the conclusions.

1.1. System overview

The high-level architecture of our system is depicted in Fig. 1. The system architecture is fairly standard in that it follows the stages of input preprocessing, feature extraction, and classification. Each of the stages depicted in Fig. 1 are explained in detail in Section 2 but we summarize the operation of the system here.

Input images are preprocessed to reduce luminance variation caused by uneven lighting, and stain colour hue variations caused by differences in sample thickness, sample age, amount of stain, etc. which can adversely affect features based on colour; the process is the same as that used in [7]. We also extract other information about virtual cells (V-cells, [12], explained in more detail in Section 2.5) and whether each pixel is predominantly hematoxylin-stained, eosin-stained, or unstained, corresponding very roughly to nuclei, other tissue, and non-tissue, respectively; we call these the *pixel classes*; the idea of using purple, pink and white hues to characterize primitive cytological components of a histologic image is from [19] and is described in Section 2.4. This information is used to extract features which form a region description (feature vector) that is presented to the trained cyst classifier which outputs its prediction of the input image's cyst type.

The classifier was trained from a set of training images. For each training image a region description was produced in the same manner as depicted in Fig. 1. The set of all such region descriptions was used to train two standard classifiers. The parameters of the classifiers were then optimized using epithelial region descriptions extracted from a validation set of images, disjoint from the training set; this is the "optimized classifier" block in Fig. 1. The optimized classifier was then evaluated using a third, disjoint set of test images.

Table 1
Summary of training data set composition (73 images total) denoting counts of samples for dentigerous cysts (DCyst), odontogenic keratocysts (OKC), lateral periodontal cysts (LPC) and glandular odontogenic cysts (GOC). Two groups of training data were gathered at different times with the same equipment, but slightly different resolution.

	Subset 1	Subset 2	Class total
DCyst	10	10	20
LPC	10	10	20
OKC	10	10	20
GOC	8	5	13
Total	38	35	73
Resolution (px)	1300×1030	1080×734	
Scale (px/μm)	3	3	

2. Materials and methods

2.1. Image data sets

Three different sets of images which were acquired at different times with slightly different acquisition parameters were used in our work.

2.1.1. Training set

Epithelial region descriptions (feature vectors) from the images in the training set were used to train classifiers with examples of descriptions from each cyst class and to construct colour histogram models of epithelial and stromal regions in the colour standardization process (Section 2.3).

A total of 73 images were obtained using a Leitz Dialux 20 microscope (Leitz Wetzlar, Germany) with an NPL 25X objective lens. Of these, 38 images were obtained at a resolution of 1300×1030 pixels with an inter-pixel distance of 1/3 μm (3 pixels per micron), consisting of 10 dentigerous cysts (DCyst), 10 lateral periodontal cysts (LPC), 10 odontogenic keratocysts (OKC), and 8 glandular odontogenic cysts (GOC). The remaining 35 images were obtained at a resolution of 1080×734 pixels with an inter pixel distance of 1/3 μm (3 pixels per micron), consisting of 10 DCyst, 10 LPC, 10 OKC, and 5 GOC. Composition of the training set is summarized in Table 1. Each sample in the training set is from a unique patient.

The inter-pixel distance (scale) for the images was determined by measuring the diameter in pixels of red blood cells (RBC) incidentally occurring in the images which are known to be 7 μm in diameter. The mean diameter in pixels of 60 RBC sampled from image subsets sharing the same acquisition parameters was measured and divided by 7 to determine the image scale in pixels per micron (px/μm). Although the actual size of the RBC has been slightly reduced due to shrinkage caused by the tissue fixation process, the diameter of the RBC in pixels and the assumed actual RBC diameter is used only for relative calibration of quantitative image features involving distances, and any assumed RBC diameter would produce an equivalent calibration.

2.1.2. Validation and test sets

The validation and test sets were collected at the same time using a Zeiss Axio microscope (Carl Zeiss MicroImaging GmbH, Göttingen, Germany) with N-Achroplan 20X objective lens. These images have size of 1300×1030 pixels and an inter-pixel distance of 0.385 μm (2.6 pixels per micron). Scale was determined in the same manner as that for the training set, using 92 sample blood cells. The number of images of each type of cyst contained in each