Tissue and Cell 53 (2018) 111–119

Contents lists available at ScienceDirect

Tissue and Cell

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

Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis

Dev Kumar Das[a,*], Surajit Bose[b], Asok Kumar Maiti[c], Bhaskar Mitra[d], Gopeswar Mukherjee[b], Pranab Kumar Dutta[e]

[a] School of Medical Science and Technology, Indian Institute of Technology Kharagpur, West Bengal, 721302, India
[b] Barasat Cancer Research and Welfare Centre, Barasat, West Bengal, 700124, India
[c] BC Roy Technology Hospital, IIT Kharagpur, West Bengal, 721302, India
[d] Drs.Tribedi and Roy, Kolkata, West Bengal, 7000016, India
[e] Dept. of Electrical Engineering, IIT Kharagpur, West Bengal, 721302, India

ARTICLE INFO

Keywords:
Oral cancer
Epithelial layer
Keratin pearl
Convolution neural network

ABSTRACT

Identification of various constituent layers such as epithelial, subepithelial, and keratin of oral mucosa and characterization of keratin pearls within keratin region as well, are the important and mandatory tasks for clinicians during the diagnosis of different stages in oral cancer (such as precancerous and cancerous). The architectural variations of epithelial layers and the presence of keratin pearls, which can be observed in microscopic images, are the key visual features in oral cancer diagnosis. The computer aided tool doing the same identification task would certainly provide crucial aid to clinicians for evaluation of histological images during diagnosis. In this paper, a two-stage approach is proposed for computing oral histology images, where 12-layered ($7 \times 7 \times 3$ channel patches) deep convolution neural network (CNN) are used for segmentation of constituent layers in the first stage and in the second stage the keratin pearls are detected from the segmented keratin regions using texture-based feature (Gabor filter) trained random forests. The performance of the proposed computing algorithm is tested in our developed oral cancer microscopic image database. The proposed texture-based random forest classifier has achieved 96.88% detection accuracy for detection of keratin pearls.

1. Introduction

The incidence rate of oral squamous cell carcinoma (OSCC) is increasing in many Asian countries due to the frequent consumption of alcohol and excessive tobacco chewing. OSCC is an important subtype of oral cancer that represents above 90% of total oral cancer cases (França et al., 2012). Like all other cancer diagnosis, the histological evaluation, i.e. the study of tissue samples of affected region under the microscope, is the clinical practice to diagnose OSCC and its different grades. This method of microscopic investigation is referred as gold standard in cancer diagnosis. According to WHO, OSCC has been categorized into two groups; low grade and high grade depending on different histological parameters (Pindborg et al., 1997). In conventional practice, pathologists used to investigate histopathological images of an oral mucosa under lower magnification ($2\times$ or $4\times$ or $5\times$ objective lens with effective magnification of $20\times$, $40\times$, and $50\times$) to find any abnormality associated with oral submucous fibrosis (OSF), oral epithelial dysplasia, or OSCC. Oral mucosa is a tiny mucous membrane inside the oral cavity. Histologically it is categorized into three subtypes such as, keratinized stratified squamous epithelium (found in hard palate, gingiva, and dorsum area of the tongue), non-keratinized stratified squamous epithelium (present in most of the oral region), and specialized mucosa (found in taste buds of lingual papillae) (Nanci, 2007). Oral mucosa consists of stratified squamous epithelium and subepithelium connective tissue, which is also known as lamina propria. It protects underlying connective tissue from various injurious effects. Apart from OSCC, the visual appearance of oral mucosa layer is also a key factor to assess the threats due to the diabetes, vitamin deficiency, or some chronic symptoms for tobacco users. The determination of precancerous and cancerous states is a necessary task for oral cancer treatment and prognosis. The histological variations and changes in squamous epithelium and connective tissue region help clinicians to differentiate between these two states. Deformation of epithelial architectural and cellular morphology is observed in precancerous condition. However, in cancerous condition, specifically for OSCC, a keratinized area encloses keratin pearl, cellular polymorphism, mitotic population, etc. are common well established visual features for identification. Among them, the presence of keratin pearl within

* Corresponding author.
E-mail addresses: dev.biomedical@gmal.com, dev.biomedical@smst.iitkgp.ernet.in (D.K. Das).

https://doi.org/10.1016/j.tice.2018.06.004
Received 16 March 2018; Received in revised form 29 May 2018; Accepted 18 June 2018
Available online 28 June 2018
0040-8166/

keratinized area and also the percentage of keratinization are the key clinical factors in the initial assessment of OSCC grades. Keratin is an abundant cellular protein constitutes in epithelial cytoskeleton uniformly distributed over different epithelial layers (Mohanta and Mohanty, 2016; Ramulu et al., 2013). Chronic physical trauma and heat due to smoking are the main causes for higher keratin formation. Over 20% keratin is visualized in low-grade OSCC; however, < 20% keratin present in high-grade OSCC (Pindborg et al., 1997). Keratin pearl is a type of keratinized structure frequently visible in low-grade carcinomas; however, it is rarely found in high-grade carcinoma too. In conventional diagnosis process, pathologists visually examine and quantify the keratinized area under lower magnification. The process of manual assessment fully depends upon the pathologist's expertise and experience, but still it leads to inter-rater variations which in turn affect diagnostic accuracy of OSCC grading. There is indeed, a need for automated quantification of the keratinized area from oral histological images for OSCC grading to reduce diagnostic error introduced during the manual evaluation process. The early onset detection of OSCC can certainly improve the survival rate. In view of this, the computer assisted tool could be a good option for the clinicians to validate their intervention leading to a fast and error optimized diagnosis.

Several quantitative approaches have been reported for cancer diagnosis using histopathological images (Gurcan et al., 2009; He et al., 2012; Komura and Ishikawa, 2017; Veta et al., 2014). The computational approach is applied for identification of different immunohistochemical marker K19 (Safadi et al., 2010), *β-catenin* (Albasri et al., 2015) to detect OSF and OSCC. But a limited research reported for automated identification of histological parameter to detect oral cancer using computer vision or machine learning approaches. It is suggested in one report that automated quantification is needed for cancer or precancerous (viz., OSF) diagnosis to reduce grading dilemma between clinicians (Eid and Landini, 2003). Epithelial, subepithelial, and connective tissue junction for normal, OSF, and squamous cell carcinoma quantified using fractal dimension measure (Landini and Rippin, 1996). Graph theoretical approach was employed for analyzing architectural changes over epithelial layer (Landini and Othman, 2004). Different image analysis approaches have been reported to identify histological parameters namely epithelial (Krishnan et al., 2011; Pal et al., 2008; Ray et al., 2008), subepithelial for OSF diagnosis (Jadhav et al., 2006; Krishnan et al., 2009; Patra et al., 2012). After an extensive literature review, it was observed that the reported works mainly focuses on OSF diagnosis. Limited research reported for keratin region extraction from oral histological images (Das et al., 2015). Keratin pearl is one of the initial screening criteria for OSCC grading but till date (as per our knowledge) no attempted has been seen to detect keratin pearl structure. In this paper, a pixel-based CNN architecture is designed for segmentation of oral constituent layers. Furthermore, keratin pearl structure is detected using random forest tree classifier by incorporating Gabor features.

2. Materials and methodology

The step wise description of proposed methodology for computing microscopic images of the oral mucosa is illustrated in the Fig. 1. Our method follows a two-stage framework; performing segmentation of constituent layers in the first stage and second step does keratin pearl detection. To train CNN with known data, the training patches were created from ground truth or labelled images of oral mucosa histology. The trained CNN architecture would then be used for epithelial, subepithelial, and keratin region delineation. Once keratinized area was determined, hand crafted feature-based machine learning would be accomplished to detect the accommodated pearl region inside keratin.

2.1. Microscopic imaging of oral mucosa histology

The clinically valued oral tissue histopathological slides (25 low grades, 15 high grades, and two healthy subjects) were collected from Barasat Cancer Research & Welfare Centre, Barasat, WB, India. Three images from each slide (total 126) were acquired using the in house set up of Leica DM2700 M microscope under 50× effective magnifications. The resolution of the images is 2048 × 1536 pixels (1 pixel is equivalent to 1.163 μm). Fig. 2 shows the acquired microscopic images from oral tissue histology. The data labelling and grading of cancer were manually accomplished by clinical expert. On a different note, the institutional ethical approval was obtained to conduct this study.

2.2. Data labelling and training patch generation

In this section, the strategy for selection and extraction of patch region from the label data is described. After image acquisition, 100 s of sub images from 80 images were manually labeled by clinical expert, and later those labeled images were utilized for generating training patches (Fig. 3). 20 images manually labeled by expert clinician (10 for epithelial and 10 for keratin region) are taken for testing (did not participate in training process) the proposed method to evaluate its performance. Each training image contains four regions such as keratin, epithelial, subepithelial and background. 7 × 7 RGB patches are extracted (in the patch directory, ~25% are overlapping and rest ~75% are nonoverlapping patches), essentially the patch size is 7 × 7 × 3 for performing two-dimensional CNN. In this study, total one million patches (0.25 million patches from each regions) were taken for trainings of CNN.

2.3. Convolution neural network architecture and training model design

Deep CNN is a parametric neural network containing several layers to perform sequential operations through involving numerous hyperparameters, which could be used to derive a robust classification model. The motivation and key concept of developing deep neural network initiated from the animal visual cortex model. It seems from outside; the conventional deep CNN follows the general approach of doing machine learning like first feature extraction and learning and then pattern classification. But the key difference is, it learns high level features from the set of image patches through performing sequential operations (convolution, maximum pooling etc.) at different layers (may be two, three, or more) and then testing them into classifier, instead of computing hand crafted features. To make this network stable and optimized, the error at the classification layer can be back propagated and iterated for n^{th} times. In this work, 12 layers are used to construct the deep architecture of CNN as shown in Fig. 4.

2.3.1. Feature extraction layer

Feature extraction layer comprises of a set of multiple convolutions, rectified linear unit (ReLU), and pooling layers. During propagation of CNN model, the feature map after sequential operations (convolution-ReLU-Maxpool) reduces sequence by sequence. Thereby, the network extracts the coarse features from the input patches.

2.3.1.1. Convolution layer. This is the main block for CNN, which extracts local region description from image patches. The convolution layer consists of multiple numbers of learnable filters. Each filter is convolved across the whole patches toproduce different feature maps (Chen et al., 2016). The output of a neuron at the i^{th} layer of j^{th} feature map at some position *(x,y)* in the patch can be defined as:

$$v_{ij}^{xy} = b_{ij} + \sum_{p=1}^{h}\sum_{q=1}^{w}\sum_{k=1}^{d} w_{ijk}^{pq} v_{(i-1)k}^{(x+p)(y+q)} \quad (1)$$

Where w_{ijk}^{pq} represents weight at the position *(p,q)*for the k^{th} channel, $v_{(i-1)k}^{(x+p)(y+q)}$ defines feature map of *(i-1)th* layer, and b_{ij} is the overall bias added in this convolution layer.