Tissue and Cell 48 (2016) 265–273

Contents lists available at ScienceDirect

Tissue and Cell

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

Quantitative microscopic evaluation of mucin areas and its percentage in mucinous carcinoma of the breast using tissue histological images

Monjoy Saha [a], Indu Arun [b], Bijan Basak [b], Sanjit Agarwal [b], Rosina Ahmed [b], Sanjoy Chatterjee [b], Rohit Bhargava [c], Chandan Chakraborty [a,*]

[a] School of Medical Science and Technology, IIT Kharagpur, 721 302, India
[b] Tata Medical Center, New Town, Rajarhat, Kolkata 700 156, India
[c] Department of Bioengineering, University of Illinois at Urbana–Champaign, United States

ARTICLE INFO

Article history:
Received 6 September 2015
Received in revised form 30 January 2016
Accepted 23 February 2016
Available online 26 February 2016

Keywords:
Breast
Mucinous carcinoma
Invasive ductal carcinoma
Mucin
Mucin index
Computer-aided methodology

ABSTRACT

Mucinous carcinoma (MC) of the breast is very rare (~1–7% of all breast cancers), invasive ductal carcinoma. Presence of pools of extracellular mucin is one of the most important histological features for MC. This paper aims at developing a quantitative computer-aided methodology for automated identification of mucin areas and its percentage using tissue histological images. The proposed method includes pre-processing (i.e., colour space transformation and colour normalization), mucin regions segmentation, post-processing, and performance evaluation. The proposed algorithm achieved 97.74% segmentation accuracy in comparison to ground truths. In addition, the percentage of mucin present in the tissue regions is calculated by the mucin index (*MI*) for grading MC (pure, moderately, minimally mucinous).

1. Introduction

Breast cancer (BC) is the 2nd most common cancer among women in developed as well as developing countries (Ferlay et al., 2015). In 2012, BC caused approximately 521,000 deaths worldwide (IAR, 2016; Russo et al., 2016; WCR, 2016). BC incidence rate (3.1% annual increment from the year 1980 to 2010) is geared up due to urbanization, adopting the western lifestyle, physical inactivity, increasing alcohol consumption, heavy smoking (especially before pregnancy), night shift work, food habit, etc. (Davis Tsu et al., 2013; Forouzanfar et al., 2011). BC has a higher risk in all over the world in spite of independent of modernization (NBCF, 2016). In the year 2015, it is predicted that almost 60,290 in situ and 231,840 new invasive BC patients would have been diagnosed in the US (Siegel et al., 2015). Among all BC cases, especially pure mucinous carcinomas (MC) (also known as colloid carcinoma) are rare and accounts for 1–7% (Komaki et al., 1988; Bloom and Richardson, 1957; Lacroix-Triki et al., 2010). This carcinoma usually affects post-menopausal women or older women (age > 60 years) (Memis et al., 2000). Histologically, MC are characterized by nests of small and uniform tumour cells floating in lakes of extracellular mucin separated by delicate fibrovascular septae (Dumitru et al., 2015). MCs generally show low nuclear atypia, however, some cases can show increased atypia and mitosis. MCs can be pure or mixed type. A pure tumour is composed of more than 90% MC. Mixed tumours usually show an admixture of MC and invasive carcinoma of no special type. Pure MCs are associated with lower recurrence rates, lower incidence of lymph node metastasis and have excellent disease free survival when compared to mixed mucinous carcinomas and invasive carcinoma of no special type.

In a study, depending on the quantity of mucin present in the tumour, MCs were subcategorized into pure mucinous (mucin ≥ 90%), moderately mucinous (90% > mucin > 30%) and minimally mucinous (mucin ≤ 30%) (Conant et al., 1994). The World Health Organization reported that mixed MC contains 90% > mucin > 50% quantity of mucin (Lacroix-Triki et al., 2010). In some papers, it is claimed that MC can also be categorized on the basis of mucin content (Mecklin et al., 1986; Lam et al., 2004), as the variation of the quantity of mucin has a great impact on BC diagnosis. A radiologic and pathologic correlation of the cases showed that MC with higher percentages of mucin showed less aggressive, slower growth rates and lower grades while those with small percentages of mucin showed radiologic and histologic characteristics of more aggressive invasive ductal carcinomas. Studies have shown

* Corresponding author.
E-mail address: chandanc@smst.iitkgp.ernet.in (C. Chakraborty).

http://dx.doi.org/10.1016/j.tice.2016.02.005
0040-8166/

that prognosis varies with the amount of mucin substance present in the tumour, so that survival rate is high in pure type carcinoma than moderately and minimally MC (Toikkanen and Kujari, 1989). Grade I tumours generally come under pure mucinous, grade II and III are either moderately or minimally mucinous (Conant et al., 1994). MC of the breast is typically positive for progesterone and estrogen receptors and negative for human epidermal growth factor receptor 2 or HER2/neu (Dumitru et al., 2015; Kashiwagi et al., 2013).

Generally, after the confirmation of BC clinicians recommend for the core biopsy. Amongst various pathological target features, mucin plays an important role for histopathological evaluation of MC. The qualitative assessment of MC histopathological slide is usually done under 4×, 10× and 40× magnifications. At very low magnification, architectural changes are observed. To achieve a better grading procedure, the mucin regions need to be quantitatively evaluated. The purpose of this paper is to develop a quantitative computer-aided methodology for automated identification of mucin areas and its percentage using tissue histological images.

Based on the extensive literature survey, it is observed that quantitative evaluation for automated identification of mucin area from breast histological images has not been attempted. Few researchers reported mucin carcinoma diagnosis methodology using magnetic resonance images (Molinari et al., 2015; Zhang et al., 2015). No attempt has been made towards automatic identification, segmentation and grading using MC histological images (Dumitru et al., 2015). Pathologists determine the proportion of mucin by microscopic examination and by a rough estimation of the percentage that is subjective and less accurate. Computer-aided image analysis is a pixel by pixel and cluster dotted point technique which provides more accurate results, reduce false positivity, decreases the inter-observer variability and also helps in the accurate identification of pure, moderate/mixed and minimal MC. In view of this, the aim is to develop a quantitative microscopic evaluation of mucin areas and its percentage in mucinous carcinoma of the breast using tissue histological images at 4× magnification.

2. Materials and methods

A graphical block-diagram of the proposed method is presented in Fig. B1. We formulated a sequence of processing steps which for helping us to determine the specific characteristic feature of MC. Steps included in this methodology are: (i) microscopic imaging, (ii) ground truth segmentation, (iii) pre-processing, (iv) mucin region extraction and (v) post-processing.

2.1. Microscopic imaging from breast MC histopathological slides

For automatic mucin region segmentation, a total of 45 slides (15 grade I, 15 grade II, and 15 grade III) from 20 (7 grade I patients, 6 grade II patients, 7 grade III patients) confirmed breast MC patients were used. All the slides were prepared with formalin fixation and paraffin embedding, followed by cutting 4 μm thick sections and staining with haematoxylin and eosin on Leica ST 5020 multistainer automated platform. The images were grabbed from all the slides using a Leica microscope (Model DM750 and camera model ICC50) at a magnification of 4× (objective lens). Image dimensions were 2048 × 1536 pixels (1 pixel = 1.6 μm). From each slide, mucin region was selected. Then MCs classified histologically on the basis of mucin content (Conant et al., 1994) and marked by the two experienced pathologists from Tata Medical Center, Kolkata, India.

2.2. Ground truth segmentation

To create more accurate and detailed ground truth, a semi-automated ground truth tool has been developed. This tool provides freehand drawing methods with full flexibility to the pathologists to mark the mucin regions. MATLAB "freehand" function has been used to develop this algorithm. Ground truth was performed in the marked regions on all the slides. Fig. B2 shows ground truths and automated segmented images of breast MC at 4× magnification.

2.3. Pre-processing

The aim of the pre-processing is to enhance and normalize the image colour. The pre-processing starts with colour transformation followed by colour normalization.

2.3.1. Colour transformation

First step is the transformation of RGB images into another colour space for contrast enhancement. In view of this, different colour channels are extracted from various colour transformation techniques such as *Lab*, *YCbCr*, *HSV*, etc. In "*Lab*" colour space, '*L*' denotes lightness and '*a*' and '*b*' denotes the colour opponent dimensions. In *YCbCr*, *Y* represents luminance component, *Cb* and *Cr* represents blue and red difference chrominance components respectively. *HSV* denotes hue, saturation and value. Colour transformation step is widely used in pre-processing step to diminish the effect of varying staining and illuminance. In the proposed method, *RGB* to *YCbCr* colour transformation showed optimum result and high accuracy. The RGB to *YCbCr* colour transformation has been done by Eq. (A1) (Zhang et al., 2009). In Eq. (A1), R, G and B represent red, green, and blue components. *Y*, *Cb* and *Cr* channels were next processed for colour normalization.

2.3.2. Colour normalization

After colour transformation, all colour channels were normalized to remove the variation of luminance and to improve the efficacy. Mean and standard deviation were calculated to normalize *Y*, *Cb* and *Cr* channels. Fig. B3 shows normalized images at various colour spaces and the corresponding output images. The mean (M) and standard deviation (S_d) of each channel have been calculated by Eq. (A2) (Gonzalez et al., 2004). In Eq. (A2), 'n' is the sample size, 'x_i' represents the observed values. Using the values of mean and standard deviation we calculated normalized image ($N_{a\times b}$), defined by Eq. (A3).

2.4. Mucin region segmentation

This part deals with the challenge of tracing the exact boundary of the mucin region. It has no proper shape and it varies from image to image. More specifically, it is a problem of dynamic shape recognition. In this view, Fuzzy c-means (FCM) model has been taken into consideration to achieve expected optimal regions. Only FCM could not solve the issues hence, we used FCM along with gap statistics and intensity variation. The mucin regions were determined by the fuzzy membership function of each class for all pixels. This function is limited in the closed interval [0,1]. The partition of the given data set is carried out through the iterative optimization technique and minimization of the objective function (Pc) for a known number of clusters. The estimation of the number of clusters is calculated by gap statistic. The gap value (Gp_n) has been calculated by Eq. (A4) (Tibshirani et al., 2001).

In Eq. (A4), 'sz' is the sample size, 'x' is the number of clusters being evaluated, and 'W_x' is the pooled within-cluster dispersion measurement, and is defined by Eq. (A5).

In Eq. (A5), 'sz_i' is the number of data points in cluster i and 'S_i' is the sum of the pairwise distances for all points in cluster i. $E_{sz} \times \{\log(W_x)\}$ is the expected value and determined by reference distribution. $\log(W_x)$ is determined from the sample data. The FCM objective function (Pc) has been calculated by Eq. (A6) (Chen and Zhang, 2004).