Artificial Intelligence in Medicine xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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

MuDeRN: Multi-category classification of breast histopathological image using deep residual networks

Ziba Gandomkar[a,*], Patrick C. Brennan[a], Claudia Mello-Thoms[a,b]

[a] Image Optimisation and Perception, Discipline of Medical Imaging and Radiation Sciences, Faculty of Health Sciences, University of Sydney, Sydney, NSW, Australia
[b] Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA

ARTICLE INFO

Article history:
Received 28 September 2017
Received in revised form 20 February 2018
Accepted 13 April 2018

Keywords:
Benign breast lesion
Breast cancer
Breast cancer subtypes
Deep learning
Deep residual networks

ABSTRACT

Motivation: Identifying carcinoma subtype can help to select appropriate treatment options and determining the subtype of benign lesions can be beneficial to estimate the patients' risk of developing cancer in the future. Pathologists' assessment of lesion subtypes is considered as the gold standard, however, sometimes strong disagreements among pathologists for distinction among lesion subtypes have been previously reported in the literature.
Objective: To propose a framework for classifying hematoxylin-eosin stained breast digital slides either as benign or cancer, and then categorizing cancer and benign cases into four different subtypes each.
Materials and methods: We used data from a publicly available database (BreakHis) of 81 patients where each patient had images at four magnification factors (×40, ×100, ×200, and ×400) available, for a total of 7786 images. The proposed framework, called MuDeRN (MUlti-category classification of breast histopathological image using DEep Residual Networks) consisted of two stages. In the first stage, for each magnification factor, a deep residual network (ResNet) with 152 layers has been trained for classifying patches from the images as benign or malignant. In the next stage, the images classified as malignant were subdivided into four cancer subcategories and those categorized as benign were classified into four subtypes. Finally, the diagnosis for each patient was made by combining outputs of ResNets' processed images in different magnification factors using a meta-decision tree.
Results: For the malignant/benign classification of images, MuDeRN's first stage achieved correct classification rates (CCR) of 98.52%, 97.90%, 98.33%, and 97.66% in ×40, ×100, ×200, and ×400 magnification factors respectively. For eight-class categorization of images based on the output of MuDeRN's both stages, CCRs in four magnification factors were 95.40%, 94.90%, 95.70%, and 94.60%. Finally, for making patient-level diagnosis, MuDeRN achieved a CCR of 96.25% for eight-class categorization.
Conclusions: MuDeRN can be helpful in the categorization of breast lesions.

1. Introduction

Breast cancer (BCa) is the most common non-skin cancer among women worldwide. Despite the increase in the incidence rate of BCa over last few decades, the mortality rate from BCa in the developed countries has been decreased due to improvements in treatment options and early detection through screening mammography [1]. For every 1000 women who have participated in screening mammography, 15.6 to 17.5 need a needle biopsy [2] but only one in four are diagnosed with BCa [3]. Therefore, each year, pathologists evaluate a large number of breast histopathological slides, from which only about 25% contains malignancy, and benign lesions are far more prevalent.

The diagnoses made by pathologists on the cases are usually considered as the gold standard for further treatment of the patients. However, recent studies have shown that the pathologists might disagree with an expert consensus-derived reference diagnoses in distinguishing benign cases from cancer [4–6]. In [5], 6900 individual case diagnoses made by 115 pathologists were compared with an expert consensus-derived ground truth and 17% of benign cases with atypia and 3% of benign cases without atypia were misdiagnosed as ductal carcinoma in situ or invasive carcinoma, while 10% of invasive carcinoma or ductal carcinoma in situ were misdiagnosed as benign cases with or without atypia. Also, it was shown that pathologists who inter-

* Corresponding author. Postal address: Room 726, Level 7, Brain and Mind Centre, 94 Mallett Street, Camperdown, NSW, 2050, Australia.
E-mail address: ziba.gandomkar@sydney.edu.au (Z. Gandomkar).

https://doi.org/10.1016/j.artmed.2018.04.005
0933-3657/

pret a smaller number of cases per week and those working as general pathologists make more diagnostic errors than experts [3–5]. Allison et al. [4] divided underlying reasons for disagreement among the pathologists into three categories, which were pathologist-related, diagnostic coding/study methodology-related, and specimen-related. Among pathologist-related factors, "professional differences of opinion on features meeting diagnostic criteria" was ranked first. Computer-assisted analysis can be helpful in reducing the discrepancies in benign/malignant classification by providing an objective classification.

Recently, with the advent of whole slide imaging and production of digital histopathology slides, many researchers have started developing computer-aided detection tools for classification of breast slides as benign or malignant [7]. For example, Weyn et al. [8] used wavelet-based, Haralick, intensity-based, and morphological features extracted from segmented nuclei and their surrounding for classification of breast histopathological slides as benign or malignant, and achieved a correct classification rate (CCR) of 79% for case-based classification. In [9], 84 features (morphological, intensity-based, and textural) extracted from isolated nuclei were utilized for classifying images as either benign or malignant. A sensitivity of 97% and a specificity of 94% has been achieved. However, both methods are computationally expensive as the epithelial nuclei were segmented first. Unlike these methods, Yang et al. [10] extracted textural features using a texton-based approach without segmenting the structures in slides. Using this method, 89% of images were classified correctly.

Pathologists are responsible for not only identifying whether a lesion is malignant or benign but also determining the benign or cancer subtypes, as both benign and malignant breast lesions encompass different subcategories. Different treatment options are available for BCa patients and determining the BCa subtype could be helpful in predicting the patient's response to therapy; for example, invasive lobular cancer gains a clear benefit from systemic therapy when compared to invasive ductal cancer [11]. The correct recognition of benign lesion type is also important because the patient's risk of developing subsequent BCa varies among different types of benign lesions [12].

Cserni et al. [13] showed that there are discrepancies among pathologists for determining benign lesion subtypes. The study asked six pathologists to classify benign lesions into three categories, namely fibroadenoma, phyllodes tumor, and anything other these subtypes. The overall Cohen's kappa for categorizing was 0.48, which suggests a moderate agreement [14]. Lawton et al. [15] investigated the agreement among ten pathologists for distinguishing fibroadenomas from phyllodes tumors and found that there was 100% agreement only in 53% of cases. In [16], the inter-observer agreement for classification of invasive breast carcinoma was studied and the highest agreement rates among 13 pathologists were achieved for mucinous, lobular, and tubular subtypes, with agreement rates of 96.0%, 78.7%, and 78.0% respectively.

Similar to the malignant/benign classification, computer-assisted analysis could help pathologists to increase diagnostic agreement in multi-class categorization of lesions. Despite the importance of determining the lesion subtype, only a few previous studies aimed at automatic classification of breast lesions into different subtypes. In [10], six subtypes of BCa were divided into two subgroups: cancer class I, which contains ductal carcinoma in situ and lobular carcinoma in situ, and cancer class II, containing invasive ductal carcinoma, invasive lobular carcinoma, lymph-node-negative metastasis, and soft tissue metastasis. Using a texton-based approach, images were classified into three classes, i.e. benign, cancer type I, and cancer II and a CCR of 80% was achieved.

Recently there has been growing research on the application of Deep Learning (DL) in medical image segmentation and classification [17]. DL revolutionized machine learning in the past few years. Conventionally, in supervised machine learning techniques, the discriminant features were selected based on domain-specific knowledge and computer algorithms determined the optimal decision boundary in the feature space. However, in DL, computers learn the optimal feature representations from the data. Although deep neural networks originated from previously existing artificial neural networks[18], training the deep architectures have recently become practical due to emergence of high-performance GPU computing, which makes it feasible to train networks with many hidden layers in a reasonable time. AlexNet, one the earliest convolution neural networks, won the ImageNet challenge in December 2012 [19]. Since then, further progress in the architecture of networks and learning algorithms has been made [17]. Similar to other machine learning fields, DL algorithms have been widely used for analyzing medical images, especially in segmentation tasks such as anatomical structures segmentation in retinal images [20],or organ segmentation in abdominal CT images [21,22]. DL has also been utilized for classification tasks such as classification of breast lesions as benign or malignant on mammograms [23] or tomosynthesis [24]. In musculoskeletal image analysis, DL was adopted it for different application such as age assessment on x-ray and MRI images or vertebrae localization and identification on CT and MRI images. The promising results are also achieved in other areas such as brain [25] or cardiac [26] images.

In the field of digital pathology, DL has been used for nuclei detection, segmentation, and classification [27]. For example, Xu and Huang [28] utilized a distributed deep neural network architecture to detect cells in whole-slide high-resolution histopathological images. In [29], a multiscale convolutional network was used for accurate segmentation of cervical cytoplasm and nuclei. DL has been also adopted for organ segmentation in large histopathological images such as segmentation of colon glands [30] or neuronal structures [31]. Furthermore, it has been utilized for classification purposes; for example, colon cancer classification [32], thyroid cytopathology classification [33], or Gleason grading of prostate cancer images [34].

DL algorithms were also used for analyzing breast histopathology slides. In [35], it was used to detect mitotic figures within breast slides. In the tutorial on DL provided in [27], AlexNet network schema was used to address the segmentation of nuclei, epithelium, and tubule as well as detection of invasive ductal carcinomas, lymphocyte, and mitosis. Spanhol et al. [36] used AlexNet for classifying breast histopathological images as benign or malignant. In [34], breast cancer areas were detected by incorporating shearlet features inside a convolutional neural networks. Cruz-Roa et al. [37] used a DL approach for automatic detection of invasive ductal carcinoma tissue. In [38], a context-aware stacked convolutional neural network architecture was used for classifying whole slide images as benign, ductal carcinoma in situ, or invasive ductal carcinoma.

This paper focuses on three tasks which are: (i) classification of breast histopathological images as benign or malignant, (ii) categorization of malignant images as ductal carcinoma, lobular carcinoma, mucinous carcinoma, or papillary carcinoma; and (iii) classification of benign images as adenosis, fibroadenoma, phyllodes tumour, or tubular adenoma. Previously Han et al. [39] used GoogLeNet [40] for classifying breast histopathological images into similar eight categories and used majority voting for patient classification. Although this aimed at addressing an almost similar problem, we improved both the image-level (i.e. considering each image individually without incorporating the patient information for decision making) and the patient-level (i.e. appointing a single label to each patient by aggregating the class labels assigned to all images of that patient) classification CCRs. This was achieved by first carrying out stain normalization as a pre-processing step. Secondly, we used a deeper network and a two-stage classifier and

Please cite this article in press as: Gandomkar Z, et al. MuDeRN: Multi-category classification of breast histopathological image using deep residual networks. Artif Intell Med (2018), https://doi.org/10.1016/j.artmed.2018.04.005