BBE 273 1–12
ARTICLE IN PRESS
BIOCYBERNETICS AND BIOMEDICAL ENGINEERING XXX (2018) XXX–XXX

Available online at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/bbe
ELSEVIER

Original Research Article

Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images

Q1 Pendar Alirezazadeh [a], Behzad Hejrati [a], Alireza Monsef-Esfehani [b,*], Abdolhossein Fathi [a]

[a] Department of Computer Engineering and Information Technology, Razi University, Kermanshah, Iran
[b] Department of Pathology, School of Medicine, Hamedan University of Medical Sciences, Hamedan, Iran

ARTICLE INFO

Article history:
Received 22 December 2017
Received in revised form
27 March 2018
Accepted 28 April 2018
Available online xxx

Keywords:
Whole slide microscopy image analysis
Computer aided diagnosis (CAD) systems
Histopathological image
Breast cancer diagnosis
Domain adaptation
Representation learning

ABSTRACT

Breast cancer has high incidence rate compared to the other cancers among women. This disease leads to die if it does not diagnosis early. Fortunately, by means of modern imaging procedure such as MRI, mammography, thermography, etc., and computer systems, it is possible to diagnose all kind of breast cancers in a short time. One type of BC images is histology images. They are obtained from the entire cut-off texture by use of digital cameras and contain invaluable information to diagnose malignant and benign lesions. Recently by requesting to use the digital workflow in surgical pathology, the diagnosis based on whole slide microscopy image analysis has attracted the attention of many researchers in medical image processing. Computer aided diagnosis (CAD) systems are developed to help pathologist make a better decision. There are some weaknesses in histology images based CAD systems in compared with radiology images based CAD systems. As these images are collected in different laboratory stages and from different samples, they have different distributions leading to mismatch of training (source) domain and test (target) domain. On the other hand, there is the great similarity between images of benign tumors with those of malignant. So if these images are analyzed undiscriminating, this leads to decrease classifier performance and recognition rate. In this research, a new representation learning-based unsupervised domain adaptation method is proposed to overcome these problems. This method attempts to distinguish benign extracted feature vectors from those of malignant ones by learning a domain invariant space as much as possible. This method achieved the average classification rate of 88.5% on BreaKHis dataset and increased 5.1% classification rate compared with basic methods and 1.25% with state-of-art methods.

© 2018 Nalecz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences. Published by Elsevier B.V. All rights reserved.

UNCORRECTED PROOF

* Corresponding author at: Department of Pathology, School of Medicine, Hamedan University of Medical Sciences, Hamedan, Iran.
E-mail addresses: a.pendar@stu.razi.ac.ir (P. Alirezazadeh), behzad.hejrati@gmail.com (B. Hejrati), monsef@umsha.ac.ir (A. Monsef-Esfehani), a.fathi@razi.ac.ir (A. Fathi).

https://doi.org/10.1016/j.bbe.2018.04.008
0208-5216/© 2018 Nalecz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences. Published by Elsevier B.V. All rights reserved.

Please cite this article in press as: Alirezazadeh P, et al. Representation learning-based unsupervised domain adaptation for classification of breast cancer histopathology images. Biocybern Biomed Eng (2018), https://doi.org/10.1016/j.bbe.2018.04.008

1. Introduction

Every year, millions of women in the world are examined by using medical image examination to diagnose breast cancer. Each year, 2.4 million new cases are reported and consequently 523,000 death per year for women and even 10,000 death per year for men are recorded [1]. Research has shown breast cancer (BC) is the second deadly disease in the world. Fortunately, with the early diagnosis, the death rate of this dangerous disease can be decreased.

To define cancer, we need to know the types of tumors. In oncology, neoplasms are divided into two groups: malignant and benign. A tumor is benign when it is microscopic and apparel characteristics are quite harmless. In other words, it remains limited and topical and cannot metastasize to other parts of the body. Therefore, a surgical excisional removal can usually save a patient's life. Malignant tumors are called the cancer. This disease can invade and destroy adjacent structures. It can also metastasize to further body parts and cause death.

Nowadays the in-time detection and diagnosis of tumors with the help of digital image processing and machine learning algorithms can be great help to increase the accuracy of breast cancer diagnosis. Today, there are variety type of imaging producers for breast cancer such as magnetic resonance imaging (MRI) with different type and resolution, ultrasound or sonography, digital mammogram (DM), microscopic (histological) images, and infrared thermography (IRT) [2]. More than four decades, these imaging producers have been used for cancer detection [3]. Hence numerous number of image processing and machine learning algorithms are proposed by many researcher and also many CAD systems are used in most clinics to help experts by reliable detection and diagnosis in a relatively digital work follow. Using these systems decreased mortality rate by 30–70% in recent year [2].

Undoubtedly present designed CAD systems for dealing with radiology images such as MRI, DM and IRT are very strong and common in most clinics but it seems radiology images are not sufficient to have a precise diagnosing and detecting subtype of cancer [4]. Using of biopsy method is a reliable method to detect cancer by more confidence [3]. The biopsy is a medical procedure involving extraction of sample cells or tissues for fixing a part of them in formalin and paraffin on a glass microscope slide which is achieved by surgery from a breast tissue. This sample will be stained by combination of hematoxylin and eosin (H&E). This staining standard has been used for more than a century and it is first routine in pathology clinics to diagnose cancers. If the experts need to know more information about exact type of cancer they will use different biomarker such as immunohistochemistry (IHC) images or the other specific biomarker such as in situ hybridization (ISH) [4]. These complementary staining are usually used along with H&E to achieve more accurate diagnosis.

Analysis of H&E and IHC stained slides by using microscope for pathologist experts are a very time consuming task. During this boring task, some serious error in interpretation may occur due to fatigue of experts, decrease of attention and human sensory limitation such as sight error and wrong discerning color. As mentioned by Gurcan et al. [5], there is an undeniable pressure for computer-assisted diagnosis to lessen the workload on pathologists by helping them to focus more on the difficult-to-diagnose cases instead of simple cases. It is clear CAD cannot make an absolute diagnosis of cancerous images; however, it only provides experts with necessary guidance to confirm a diagnosis by decreasing number of samples [5]. Due to this advantage, histopathology CAD systems play important role in recent year and they have been attracted attention of many image processing and machine learning researchers. CAD systems can be divided into two different categories by considering type of input images. One category is to group these systems according to the type of staining input images, the first group of this category is the systems that use H&E stained images [7–21] and second group uses IHC image [15,22–25]. Since H&E stained images contain more information for the diagnosis of benign and malignant neoplasms, the analysis of H&E histology images requires powerful image processing and machine learning techniques due to the diversity of tissue and tumor structures, different tissue-preparation conditions, stain diffusion, inappropriate crops and shrinkage during installation. On the other side, IHC images are stained by specific biomarker to show specific cells or regions. Therefore it is easy to detect these cells and regions by using simplest algorithms and techniques than H&E stained images [6]. IHC images are most useful when the expert has identified cancer cases with H&E images.

Another category for grouping CAD systems is based on whole slide of image (WSI) or regions of interest image named as (ROI). Either H&E or IHC stain images can be as input data for these CAD systems. In computer aided diagnosis for WSI, entire part of image is used without any separation and segmentation to map images to appropriate categories (cancerous and non-cancerous). High recognition rate is the main purpose of these systems. Instead in ROI approach, some important parts of image detect or segment to focus on them. The aim of CAD for ROI is determining tumor region, scoring of immunostaining, cancer staging, mitosis detection, gland segmentation, and detection and quantification of vascular invasion [7]. Extracting these segments are easier than extracting features from entire part of image which contains different information. In continue, WSI and ROI methods are described as follow:

A. Region of interest (ROI)

In this approach only regions of interest are segmented and detected that contain essential information for diagnosing such as nuclei of cells, mitosis cells, gland, specific patches in tissue, etc. there are two main approach to detect and segment these regions by using machine learning algorithms. The first is the methods that fully detect the various components of the tissue. All steps from the beginning to the end are automatically and experts do not interfere in determining or marking these regions [7]. The algorithms used in this approach include unsupervised learning algorithms such as k-means, autoencoders and principal component analysis (PCA). The main purpose of unsupervised learning is to find out appropriated partition

Download English Version:

https://daneshyari.com/en/article/6484148

Download Persian Version:

https://daneshyari.com/article/6484148

Daneshyari.com