Informatics in Medicine Unlocked 8 (2017) 74–79

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

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

Detection of breast cancer on digital histopathology images: Present status and future possibilities

M.A. Aswathy, M. Jagannath*

School of Electronics Engineering, VIT University, Chennai 600127, Tamilnadu, India

ARTICLE INFO

Keywords:
Breast cancer
Biopsy
Histopathological image analysis
Image segmentation
Image classification
Feature extraction

ABSTRACT

Breast cancer is a very common type of cancer in women around the world and more so in India. It affects not only women but also men. In India, we have a very disturbing trend of increase in cancer patients especially in urban areas. Out of 100 cancer patients, 25–32 patients are affected with breast cancer. The modern medical science has several advanced methodologies and techniques for the identification of breast cancer. Digital pathology is one of the emerging trends in modern medicine. Pathological studies are getting more prominence in detection of various types of cancers. This article reviews and summarizes the applications of digital image processing techniques on histopathological images for the detection of breast cancer and discusses its future possibilities.

1. Introduction

Cancer has become a major threat to individuals all around the world. In India, the rate of death due to cancer is huge and surpassing 8 lakhs reported cases by the year 2000 according to Indian populace registration information [1]. It is the second largest chronic infection in India in charge of greatest mortality with around 0.3 million passing for every year. The increase in the number of cancer patients every year from 2005 is shown in Fig. 1 [1]. The increasing trend of cancer patients in the last few decades enables us to predict the count of patients by the end of 2020 in India.

Breast cancer topped the list for females and mouth cancer for males [19]. Breast cancer became a chronic disease affecting worldwide women population. Even though the most of elemental causes and other features are common around the world, every region has its own specific causes for cancer. In India around 1.5 lakh new cases with breast cancer (over 10% of all cancers) have been registered in 2016 [31]. It's not the first time medical researchers are targeting breast cancer. Due to lack of advancement in medical field, this disease is growing as one of the most chronic diseases of the era. Recent trends in image processing shows what an engineer can do in the medical field. Medical image processing became highly acceptable mechanism now days. The vast development in information technology makes this mechanism reliable and efficient. Medical image processing is not only used in cancer detection but also for diagnosis of other diseases [37]. The procedure that takes out a piece of the mass or a sample from human body for testing is called a biopsy. The tissue sample is called the biopsy sample or specimen. The testing process is referred to as pathology. The process of histopathology defined as the detailed analysis of a biopsy sample by a pathologist. First the sample has been processed and then sectioned onto glass slides. On the other hand, histopathology is the examination of biopsy sample taken by either intrusive or minimal intrusive methods by a pathologist under an instrument like microscope for studying the growth of cancer, tumour etc.

Different structures of the tissue are coloured with different stains for the sake of conceptualizing under the microscope. Then, a detailed study on these tissues has been done by a pathologist for the detection of lesions or tumours [15]. Scientists in have been familiar with the significance of quantitative examination of histopathological images. Quantitative examination can be used to support pathologists' decision about the closeness or the nonappearance of a disease, besides to help in infection development evaluation. Additionally, quantitative depiction is key, not only for clinical use (e.g., to expand the demonstrative unwavering quality), additionally for exploration applications (e.g., drug revelation [35]) and organic systems of ailment [13]. As an outcome, the utilization of PC supported analysis in pathology can considerably upgrade the productivity and precision of pathologist's choices, and generally speaking advantage the patient.

On account of new advancement in image processing techniques several methods have been suggested for accurate detection of breast cancer. Among the different studies, robotized segmentation and classification of nuclei/cell is a repeating task, especially troublesome on histopathology images. The segmentation of nuclei (histological

* Corresponding author.
E-mail addresses: aswathym.a2016@vitstudent.ac.in (M.A. Aswathy), jagan.faith@gmail.com (M. Jagannath).

http://dx.doi.org/10.1016/j.imu.2016.11.001
Received 6 October 2016; Accepted 5 November 2016
Available online 11 November 2016

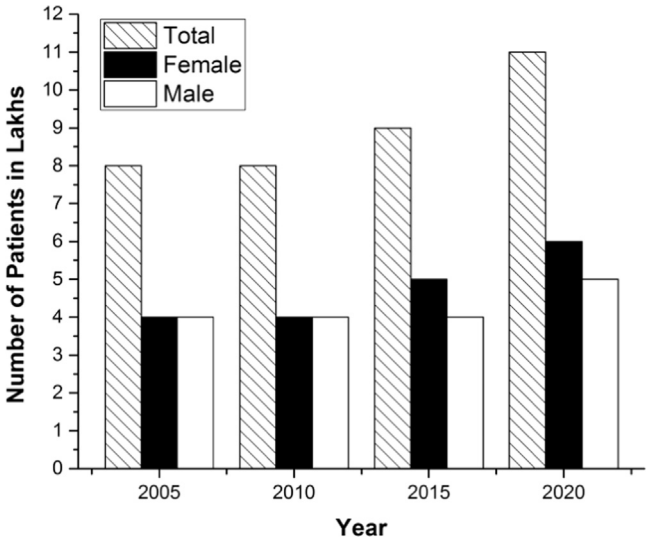

Fig. 1. Total cancer prevalence in India year wise (predicted case for the year 2020).

structures) on histopathological images is more troublesome since the greater parts of the cells are frequently part of unpredictable and sporadic visual angles [15]. This article reviews some of the most accepted computer aided techniques for the analysis of breast cancer from a histopathology image.

The article composed in seven sections. Section 2 presents the types of dyeing and distinctive image modalities in histopathology. Section 3 highlights various methods used for histopathology image analysis. Different parameters used for performance assessment of a classifier explained in Section 4. Section 5 discusses and compares different algorithms used for nuclei detection, segmentation and classification. Section 6 describes conclusion along with future possibilities explained in Section 7.

2. Staining of histology slides

Histology is the anatomical study of tissues and histology slide preparation involves steps viz: fixing, processing, embedding, sectioning and staining. The type of stain is selected according to the tissue that going to process. The tissue is dyed with one or more stains for clear visualization under the microscope [2].

The most normally utilized recoloring frameworks are Haemotoxylin and Eosin (H & E) staining and Immunohistochemistry (IHC) staining. Eosin is a negatively charged acidic color. It stains basic (or acidophilic) structures red or pink. This is likewise some of the time termed 'eosinophilic'. Along these lines the cytoplasm, stroma, etc. are stained pink by H & E staining. Haematoxylin is a basic colour utilized to dye acidic (or basophilic) structures a purplish blue. Hence the nucleus is stained purple [32]. Immunohistochemistry (IHC) is another technique using for staining. According to the presence and absence of some particular proteins (antigens or antibodies) in the tissue, we can predict stage of cancer. Fig. 2 shows examples of H & E and IHC stained images. H & E image is adapted from UCSB dataset of a benign case (stage 0) [11]. IHC stained image of invasive breast carcinoma (stage 2) shown in Fig. 2 is adapted from Jennifer et al. [18]. The quality of IHC is the natural visual yield that uncovers the presence and limitation of the particular protein with regards to various types of cells, organic states, and/or sub cellular localization inside complex tissues [9].

3. Histopathology image analysis

Histopathology refers to the tiny examination of a biopsy sample that is artificially taking outside and separated onto microscopic slides to study cancer growth, genetic progression furthermore, cell morphology for tumour finding and anticipation. The word histopathology came from Greek words: 'Histos' which means tissue, pathos which means disease and logos which means study [25]. The fundamental utilization of histopathology is in clinical medicine where it commonly includes the examination of a biopsy (i.e. a surgically expelled test or example taken from a patient for the motivations behind point by point study) by an expert doctor called a pathologist. With the late appearance of whole slide computerized scanners, tissue histopathology slides can now be digitized and put away in computerized image i.e. in a digital form.

Histological images can be obtained by using a Charge Coupled Device (CCD) camera with microscope in which an automated computerized technique can be performed [37]. The main objective of most of the digitized techniques is to get quantitative data from images. These quantitative data include size of the cell, abnormalities in the tissue and disproportionate number of cells. The main steps involved in digital image analysis: Preprocessing, segmentation, feature extraction and classification. There are many algorithms which are computer aided, available for histopathology image analysis of breast cancer. Some of the important algorithms used for histopathology image analysis are summarized in Table 1.

3.1. Preprocessing of histopathology image

For an automated detection using image processing techniques, preprocessing is the first step. Image preprocessing generally involves undesirable noise expulsion and corresponding enhancement in the image. Preprocessing can be achieved through different morphological operations like dilation and erosion, low pass filtering (median and averaging) thresholding etc. [17]. Region of interest (ROI) can be obtained in preprocessing stage which will reduces the processing time. Different preprocessing steps are illuminant normalization, color normalization, noise expulsion with smoothening and ROI detection. Illuminant normalization is a technique that clarifies the illuminant variations in the image due to a non standard imaging source. Color normalization is the method for addressing the problem of color variations in the image due to staining. Li and Plataniotis [23] suggested a method, complete normalization scheme, which is suitable for both color and illuminant normalization. Other methods are white shading correction, histogram normalization etc. Noise reduction is accomplished by using thresholding technique after which follows filtering and background correction.

3.2. Detection and segmentation of nuclei

Lymphocyte and epithelial cells are the most important types of nuclei. Nuclei structure may look dissimilar as per various components for example, type of nuclei, severity of disease, and nuclei life cycle. By analyzing histopathological images, an overall idea regarding cell structure, depth of cancer growth etc can be obtain. Information like size and shape of tumour and other cytological data can be avail through the analysis procedure. Thus prediction of breast cancer development is possible. To examine the region of interest histopathology images firstly ought to be segmented. Segmentation is the process of separating region of interest from the background and is a troublesome task in microscopic images [5]. Usually segmentation can be used to detect nuclei, stroma and background [12].

Nuclei detection and classification on histopathology images is difficult because of its complex structure. Basic segmentation techniques like Hidden Markov Model (HMM), Active Contour Model (ACM), watershed algorithm and extended forms of then can be utilized for detection. Fuzzy logic (I and II types), region growing using seeds etc are other new methodologies with same purpose. Some of the important techniques that have been used for nuclei detection and segmentation are discussed below.