Short Communication

Analysis of digital histopathology images for breast cancer diagnosis

Samir Kumar Bandyopadhyay

JIS University, West Bengal, India

Received: 04 April 2019 / Accepted: 27 April 2019

Abstract

Significant development of mammogram-based computer-aided diagnosis using image processing techniques has been reported by different research articles. Huge opportunity still persists to develop a fully automatic system to detect abnormalities and predict future risk factors. Digital histopathology analyzes breast biopsy slide image for diagnosis of breast tumors.

Key words: breast cancer, diagnosis, pathology

Introduction

Cancer is one of the most commonly occurring diseases in the modern world. It remains a challenge to cure the disease. Breast cancer affecting women is known as high mortality unless diagnosed in time. It requires a simple procedure of mammography followed by biopsy of the tumor or lesions present in the breast tissue. Computer-aided diagnosis (CAD) systems involve the development of computational algorithms on medical images. Breast cancer begins in breast tissue. It is made up of glands for milk production, called lobules and ducts. It connects lobules to the nipple. Breast cancer is the most common type of cancer and can cause death if it is not diagnosed at the premature stage. Breast cancer develops from cells lining the milk ducts and slowly grows into a lump or a tumor. Clinicians assumed that it takes a significant amount of time for a tumor to grow 1 cm in size starting from a single cell. A malignant tumor can spread beyond the breast to other parts of the body via the lymphatics or the bloodstream [1-2]. Breast cancer is either invasive or non-invasive. Invasive cancer spreads from the milk duct or lobule to other tissues in the breast. Non-invasive ones are unable to invade other breast tissues. It is called "in situ" and may remain inactive for the entire lifetime.

Breast cancer in India is the second most common cancer in Indian women. The incidence is more in urban in the higher socio-economic groups than rural women. Only early detection and diagnosis is a way of control breast cancer.

Several screening techniques such as ultrasound (USG) imaging use a band of high frequency sound waves to probe the breast, magnetic resonance imaging (MRI) probes the breast using powerful magnetic fields, and mammography. Digital mammography proves the most

Address for Correspondence: Samir Kumar Bandyopadhyay, JIS University, West Bengal, India.
E-mail: 1954samir@gmail.com

DOI: 10.5455/im.41386
This is an Open Access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (https://creativecommons.org/licenses/by-nc/4.0/)
effective screening method for tumor detection at an early stage. Confirmation of breast malignancy is done initially by biopsy. CAD is an interdisciplinary technology which combines digital image processing with medical image processing. CAD based MRI diagnosis helps the radiologist to evaluate the disease in a shorter time. Now computers will be used more often for image interpretation. The digital mammogram image pair is first pre-processed for orientation of the right oriented mammogram. Then, any unwanted artifacts are removed from the image. Then noise removal is done using the Gaussian kernel. Image registration and homogeneity enhancement are made next. The pectoral region is masked out leaving behind the breast region of interest (ROI). Breast contour detection is processed for separation of the ROI from the background. The anatomical segmentation of the breast image is made by tracing each edge path in a clockwise direction to isolate all closed objects within the breast ROI. Statistical 3o is used to identify the suspected region and any region exhibiting intensity values greater are abnormal regions.

Sometimes it is required to identify and monitor markers for high risk patients to prevent a fatality. Asymmetry analysis is one such marker, uses a geometric triangle to represent each breast. Density estimation is done using statistical methods to suppress low intensity regions. [3-4]. The resulting high intensity region is calculated for finding risk factor. Breast volume calculation performed after surgical removal of the breast called mastectomy. The geometric model is used to model the breast delivering highly accurate breast volume measures. This paper proposed the basic method comprising of digital mammography and digitized histopathology slide image specifically focusing towards the breast cancer diagnosis. Histopathological slides are studied for determination of malignancy after the biopsy is performed. Digital images are required to be registered and enhanced before application of any deterministic algorithm.

**Literature Reviews**

Breast is referred to as mammary glands. Different technologies are used to monitor or treat the medical conditions of the human body are used nowadays like X-ray, USG, computed tomography (CT), MRI, etc. In breast cancer, mammography is one of popular technique to identify breast cancer. Mammography can detect breast cancer at an early stage and treatment is more effective and a cure is more likely [5]. Research has shown that early detection with mammography may save lives and increases treatment options. One of the main advantages of mammography is comparative cost-effectiveness and availability.

The breast is a dynamic organ in human-like other organs. The breast extends from the lateral thoracic artery at the top to the inframammary fold at the bottom. The mammary gland has a small projection, the nipple, surrounded by a portion of pigmented skin called the mammary areola.

The changes in the breast may be seen in females in different stages throughout life. Breast development promotes the sprouting, growth, and development of its nature. The rate for the development of female breast differs for each young female. The structure of the female breast varies significantly at different times during a woman’s life.

Around 5000 years ago, Ayurvedic doctors dealt with the abnormal growth of tumors [6-9]. In Ayurvedic diagnosis, cancer is described as inflammatory or non-inflammatory swelling and named them as either Granthi (minor neoplasm) or Arbuda (major neoplasm) [10,11]. Malignant tumors (Tridosaja) are very harmful, because they can affect all three major body systems, i.e. the nervous system (Vata or air), the venous system (Pitta or fire), and the arterial system (Kapha or water) lose mutual coordination and thus cannot prevent tissue damage, resulting in a deadly morbid condition [12,13]. In modern medical point of view, the cancer is defined as abnormal cell division due to a number of conditions. The multiplications of cells are progressive, uncontrolled, and the cells of cancer infiltrate and destroy surrounding healthy tissue. They have the ability to metastasize and spread throughout the body [10]. Several researchers observed the seriousness of invasive breast cancer due to strongly influenced by the stage of the disease when it is first diagnosed [4]. Breast cancer is the common problem of not only in developed nations but also in developing countries [11]. It has been found that breast cancer is common to a large part of women [12]. In India, the death toll due to breast cancer is increasing at a rapid pace [13]. It is becoming a major challenge in India due to lack of awareness and lethargy of the Indian women towards health care and regular check-up. But the major obstacle is the expensive health care system and facility for treatment of breast cancer is only available in the major cities of India.

Growing age increases the risk factor almost in exponential order in females. The definite cause of breast cancer is still not identified, so, prevention becomes impossible. Early stage detection is more effective for affecting treatment and increases the probability to be cured rapidly. Several studies have shown that early detection saves lives and increases treatment opportunities. It is also evident that early detection and improvements in treatment can decrease the rate of mortality among women considerably.

Over the past 20 years, continuous advancement in the field of image processing algorithms, computer technology in terms of computational power and storage, and digital medical imaging technologies have allowed the development of powerful computer-assisted analytical tools towards the diagnosis of critical diseases like cancer. Last few years, several states of the art CAD systems have been
developed to deal with different fields of disease analysis. The breast cancer diagnosis is one of the most well-accepted implementations of CAD to assist medical practitioners.

The identification process of cancer disease is detection followed by diagnosis by clinicians for the final conclusion. In general, clinicians analyze the externally accessible characteristics such as shape, size, topology, texture, and results of a different medical test conducted like x-ray, mammogram, MRI, etc. Still, the conclusive inference can only be drawn by analyzing the hidden information embedded deep inside the tissues. Breast cancer identification also follows the same path.

Recent studies reveal that the majority of clinical tests including digital mammography and biopsies performed are benign, resulting in wastage of valuable time of medical practitioners, and at the same time increasing the possibility of false detection. The efficient CAD system can screen the benign cases and assist the experts in terms of qualitative and quantitative precision.

The widespread application of CAD can be found after the emergence of digital mammography in the early 1990s [14]. Recently, CAD has become a part of routine clinical detection of breast cancer on mammograms at many mass screening programs and hospitals [12] in the United States. Now the recent introduction of slide digital scanners, histopathology biopsy slides can also be digitized and kept in digital image format. So, image processing algorithms can also be applied to histopathological biopsy slides to analyze the image for decisive inferences. The digital histopathological biopsy slide analysis is new incorporation with high potentiality although currently, it requires complete human intervention. Researchers in the image analysis and histopathological fields have recognized the importance of quantitative analysis of histopathological images [15]. Since most current histopathological diagnosis is based on the subjective opinion of pathologists, there is obviously a requirement for quantitative image-based evaluation of digital histopathological slides. This quantitative analysis of digital histopathology is important not only from a diagnostic viewpoint but also in order to understand the underlying reasons for a specific diagnosis being rendered. In addition, quantitative characterization of histopathology image is important not only for clinical applications but also for advance research.

Proposed Method

Image processing techniques can play an important role in the interpretation of histopathological slides like breast biopsies. Many studies have been started worldwide for automatic interpretation of histopathological slide. Most of them are very much successful to assist the expert to interpret the abnormalities. Due to the growth in computational speed and advanced image processing algorithm digital pathology is now a reality especially for identification of abnormalities. In the proposed work, free Tissue Blocks downloaded from OriGene Technologies [14] are used as the dataset. In the experiments, breast cancer tissues from different patients and non-cancerous breast tissues from different normal females are considered. The sample images are 24-bit bitmap image with the size of 640X480 Pixels. The sample images provided by the OriGene technologies are colored images. The colored image is more informative but at the same time, it will increase the complexity of the method towards the analysis of the image. The 24-bit bitmap can produce almost 16,777,216 number distinct colors shades which are virtually impossible to handle. So, conventionally almost all the medical image processing algorithms use the grey shade image to reduce the complexity of the method and preserve maximum information within the image to investigate.

The grey image contains 256 numbers of grey shades to represent an image. The proposed method analyzed different alternative methods to convert the color image to grey scale image. The most important factor here is to preserve the relative color distance within the image color domain. The Euclidean distance is the perfect method to convert the color image to the respective grey shade image. The relative color distance will be perfectly preserved in the color palette. Initially a grey color palette i.e. GP is generated. It is known that the value of red, green and blue of the grey color palette is the same for a particular instance of the intensity of color. The algorithm iteratively read the color pixel (P) from the sample image (I) and split the color information in terms of red, green, and blue.

The grey shade with the lowest distance is selected and value of the grey shade is propagated to red, green and blue component of the pixel, ensuring the conversion from color to grey scale image. The iteration will terminate when all the pixels of the color image are converted to grey scale.

Algorithm: Grey Scale Conversion

```
GREY-SCALE-CONVERSION (I, Height, Width)
I ← Sample Image P ← Single Pixel
Loop i ← 1 to Height
  Do Loop j ← 1 to Width
    Do P.R ← I_i,j.R
    P.G ← I_i,j,G
```

92
The proposed method is implemented using three loops. The outer loop is for height, the inner loop is for width, and the innermost loop is to traverse the grey scale palette. Assuming that height = width = n and grey palette is k. So, the running time of the algorithm is k.n^2. But the grey palette i.e. k is always 256, so, it is a constant. Hence, the running time of the algorithm is the order of n^2.

The initial objective of the process is to eliminate the irrelevant object from the slide to make it more enhanced and clearer. The final objective is to merge the similar objects in a slide by color polarization technique. The initial objective will be achieved by increasing the contrast level of the slide by using a constant (Δ) with the intensity shades. The final objective will be accomplished by color polarization using a threshold (k) determined by the central tendency of image. For simplicity, the average of the highest (MaxV) and lowest (MinV) intensity of the color can be used for thresholding which is already derived in the previous algorithm.

**Algorithm: Color Polarization**

**COLOUR-POLARISATION (I, Height, Width, MaxV, MinV)**

I ← Sample Image

P ← Single Pixel

k ← (MaxV + MinV) / 2

Loop i ← 1 to Height

Do Loop j ← 1 to Width

Do Val ← I_{i,j}

// Enhancing the contrast

If Val ≥ 128

Then Val ← Val * Δ

If Val > 255

Then Val ← 255

Else Val ← Val / Δ

If Val < 0

Then Val ← 0

// Merging similar pixels by color polarization

If Val ≥ k

Then Val ← 255 – MaxV

Else Val ← 255 – MinV

I_{i,j} ← Val
Return I

The proposed method traverses the entire image. The outer loop is used for the height whereas the inner loop is for width. Assuming that, height = width = n. Then the running time of the algorithm is the order of $n^2$.

Experimental Results

The primary objective of the proposed method is to remove the huge amount of fat, connective tissue and gland tissue from the cancerous cells within the histopathological biopsy samples. The stage, intensity, type, future development, and treatment of cancer can only be detected on the basis of orientation of malignant cell, the shape of the cell and duct, density carcinogenic cells in comparison with normal cells. The outputs of aforesaid algorithms are depicted in the following figures for cancerous cells within the biopsy slide and normal slide along with the histogram of the images.

Figure 1. The original histopathological slide shows malignant cells along with the histogram.

Figure 2. The grey shade histopathological slide shows malignant cells along with the histogram.

Figure 3. Finally, the inverse color polarized histopathological slide shows enhanced malignant portion along with the histogram.
Figure 4. The original histopathological slide shows normal cells along with the histogram.

Figure 5. The grey shade histopathological slide shows normal cells along with the histogram.

Figure 6. Finally, the inverse color polarized histopathological slide shows no abnormalities along with the histogram.

Conclusion
Most of the CAD systems of the breast are concentrated in detecting the abnormalities in the screening technologies. But it is also important to analyze the confirmation part of the detection process; to make it a true diagnosis system. The histopathological slide image analysis or digital microscopy has immense potentiality for the future.

Conflict of interest
All authors declare that they have no conflict of interest.

Funding
There was no funding received for this article.
References


2. Griffin J, Treanor D. Digital pathology in clinical use: where are we now and what is holding us back? Histopathology 2017;70:134-145.


