COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 121 (2015) 92–108

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Novel structural descriptors for automated colon cancer detection and grading

Saima Rathore[a,b,*], *Mutawarra Hussain*[a], *Muhammad Aksam Iftikhar*[a,c], *Abdul Jalil*[a]

[a] *DCIS, Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan*
[b] *DCS&IT, University of Azad Jammu and Kashmir, Muzaffarabad, Azad Kashmir, Pakistan*
[c] *Comsats Institute of Information Technology, Lahore, Pakistan*

ARTICLE INFO

Article history:
Received 26 December 2014
Received in revised form
25 May 2015
Accepted 27 May 2015

Keywords:
Colon cancer
Colon cancer detection
Colon cancer grading
Colon classification

ABSTRACT

The histopathological examination of tissue specimens is necessary for the diagnosis and grading of colon cancer. However, the process is subjective and leads to significant inter/intra observer variation in diagnosis as it mainly relies on the visual assessment of histopathologists. Therefore, a reliable computer-aided technique, which can automatically classify normal and malignant colon samples, and determine grades of malignant samples, is required. In this paper, we propose a novel colon cancer diagnostic (CCD) system, which initially classifies colon biopsy images into normal and malignant classes, and then automatically determines the grades of colon cancer for malignant images. To this end, various novel structural descriptors, which mathematically model and quantify the variation among the structure of normal colon tissues and malignant tissues of various cancer grades, have been employed. Radial basis function (RBF) kernel of support vector machines (SVM) has been employed as classifier in order to classify/grade colon samples based on these descriptors. The proposed system has been tested on 92 malignant and 82 normal colon biopsy images. The classification performance has been measured in terms of various performance measures, and quite promising performance has been observed. Compared with previous techniques, the proposed system has demonstrated better cancer detection (classification accuracy = 95.40%) and grading (classification accuracy = 93.47%) capability. Therefore, the proposed CCD system can provide a reliable second opinion to the histopathologists.

1. Introduction

Colon cancer is one of the leading causes of cancer related deaths in modern and industrialized world. About half a million people die every year worldwide due to colon cancer [1]. Primary reason of colon cancer is chain smoking, but there are some other reasons of colon cancer such as family history of colon cancer, increasing age, and unbalanced diet such as diets with low consumption of fruits/vegetables and heavy consumption of meat [2].

The conventional method of colon cancer diagnosis is microscopic analysis of colon biopsy samples. In such an examination, histopathologists analyze the biopsy samples under microscope, and diagnose the tissue as normal/malignant based on the morphology of tissues. Further,

* *Corresponding author at*: DCIS, Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan. Tel.: +92 51 2207381x3102.
E-mail address: saimarathore_2k6@yahoo.com (S. Rathore).

http://dx.doi.org/10.1016/j.cmpb.2015.05.008
0169-2607/

histopathologists assign quantitative cancer grades depending upon the morphology of malignant tissues they observe under microscope. However, the manual process of colon cancer diagnosis has a few limitations. First, it consumes valuable time of the histopathologists as they have to examine many images per day. Second, the process is subjective, and may lead to inter- and intra-observer variability in diagnosis due to workload and experience of histopathologists [3,4]. Therefore, a computer-aided diagnostic system, which could accurately identify normal and malignant samples, and quantify the cancer grade as well, is required.

In the past decade, a few computer-aided diagnostic systems have been proposed for automatic detection of normal and malignant colon biopsy images. These techniques exploit the textural changes in normal and malignant colon biopsy images, and have been summarized in a recent survey [5]. The texture analysis of colon biopsy images is characterized by the extraction of discerning features from the images. The extracted features are then used as input to different classifiers for identifying normal and malignant images. For instance, Esgiar et al. calculated texture features of contrast, entropy, angular second moment, dissimilarity, inverse difference moment and correlation from gray-level co-occurrence matrix (GLCM) of the input colon biopsy images [6]. They achieved 90.20% classification accuracy by employing linear discriminate analysis (LDA) and K-nearest neighbor (KNN) classifiers. In another work, Esgiar et al. combined features of entropy and correlation with image fractal dimensions [7], and obtained 94.10% classification accuracy with the same set of classifiers i.e. KNN and LDA. Later, Masood et al. employed GLCM based texture features of energy, inertia and local homogeneity, and morphological features of shape, size and orientation in order to classify normal and malignant colon biopsy images. They employed third degree polynomial kernel of support vector machines (SVM) for classification, and achieved an accuracy of 84% and 90%, respectively, by using morphological and texture features [8]. Masood et al. further extended their work, and employed circular local binary patterns in order to classify colon biopsy images [9]. They employed Gaussian kernel of SVM for classification, and obtained classification success rate of 90%.

Similarly, a few cancer grading techniques have also been proposed in the past. For example, Altunbay et al. proposed a textural features based technique for classifying colon samples into normal and malignant (low grade and high grade) categories [10]. They constructed a graph on different objects, obtained by using circle fit algorithm [11] on the white, pink and purple clusters of the image. A few structural features such as degree, average clustering coefficient, and diameter are computed from the graphs, and are used to classify given images by using linear SVM.

Ozdemir and Demir also presented an automatic colon cancer detection and grading technique [12]. In this work, a few normal colon tissues are manually selected among large sized colon tissue images, and query graph is generated separately on each selected tissue. Later, reference graph is generated on each test colon biopsy image. The query and the reference graphs are generated using the same way as done in [11]. Further, multiple key regions are located in the reference image that are most similar to a normal biological structure by searching each query graph over the entire reference graph of test image. The similarity is measured in terms of graph edit distance. The graph edit distance actually quantifies the dissimilarity in source and the destination graph by calculating the minimum cost of edit operations that should be applied on source to transform it into destination graph. The basic idea behind this approach is that query graphs of normal glands will be similar to the graphs of selected key regions of normal colon biopsy images compared to those of malignant colon biopsy images. Later, graph edit distance, and some other statistical features computed from these key-regions are given as input to SVM classifier for classification of test images into normal, and malignant tissues of low and high grade cancer. They achieved 92.21% classification accuracy on a colon biopsy image based dataset. But, this technique is computationally expensive due to heavy processing involved in matching query graphs and key regions of reference graphs.

The techniques mentioned in the previous paragraphs have a few limitations. First, the graph based techniques [10,12] are computationally expensive. Generating graphs for extracting these features is not computationally expensive. Rather, these techniques consume large time in fitting circles in the three clusters of colon biopsy images for determining graph nodes, which is a pre-requisite step of graph generation. Further, feature extraction from graphs is also expensive in some of these techniques such as Ozdemir and Demir [12], wherein comparison of query graphs and key regions of reference graphs is expensive. Second major set of techniques for feature extraction [6–9] are those that exploit the texture variation in normal and malignant images. These texture features are general in nature and do not consider the specific pathological variation between normal and malignant colon tissues. For example, colon cancer causes the lumen tissue in a malignant image to turn from near-elliptic to irregular shape. Such pathological variation in structures of tissues is not encoded in the features proposed in literature for colon cancer classification/grading. Therefore, a computationally tractable computer-aided diagnostic technique is required, which could exploit the background knowledge specifically about pathological structure of normal and malignant colon tissues into the classification process.

In this paper, we propose a novel colon cancer diagnosis (CCD) system, which is capable of automatic colon cancer detection, and its classification into various cancer grades. Unlike previously proposed methods, which capture information about the general texture present in colon biopsy images, the proposed method incorporates the background pathological information about the morphology of normal and malignant tissues into the classification and grading process. To this end, some novel structural features, which exploit the shape, convexity, concavity, circularity and area based characteristics of lumen for detection and grading of colon cancer, have been proposed. A data set comprising 192 images has been used for validating the proposed CCD system. Training and testing data has been formulated using Jack-knife 10-fold cross-validation, and radial basis function (RBF) kernel of SVM classifier has been employed for both cancer detection and cancer grading. The proposed features have been proved to yield better results compared to general texture based features in the experimental section.