Computers in Biology and Medicine ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Automated colon cancer detection using hybrid of novel geometric features and some traditional features

Saima Rathore [a,b,*], Mutawarra Hussain [a], Asifullah Khan [a]

[a] DCIS, Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan
[b] DCS&IT, University of Azad Jammu and Kashmir, Muzaffarabad, Azad Kashmir

ARTICLE INFO

Article history:
Received 31 October 2014
Accepted 6 March 2015

Keywords:
Colon cancer
Classification
Elliptic objects
Colon biopsy
Elliptic Fourier descriptors
SIFT
Morphological, geometric, and texture features

ABSTRACT

Automatic classification of colon into normal and malignant classes is complex due to numerous factors including similar colors in different biological constituents of histopathological imagery. Therefore, such techniques, which exploit the textural and geometric properties of constituents of colon tissues, are desired. In this paper, a novel feature extraction strategy that mathematically models the geometric characteristics of constituents of colon tissues is proposed. In this study, we also show that the hybrid feature space encompassing diverse knowledge about the tissues' characteristics is quite promising for classification of colon biopsy images. This paper thus presents a hybrid feature space based colon classification (HFS-CC) technique, which utilizes hybrid features for differentiating normal and malignant colon samples. The hybrid feature space is formed to provide the classifier different types of discriminative features such as features having rich information about geometric structure and image texture. Along with the proposed geometric features, a few conventional features such as morphological, texture, scale invariant feature transform (SIFT), and elliptic Fourier descriptors (EFDs) are also used to develop a hybrid feature set. The SIFT features are reduced using minimum redundancy and maximum relevancy (mRMR). Various kernels of support vector machines (SVM) are employed as classifiers, and their performance is analyzed on 174 colon biopsy images. The proposed geometric features have achieved an accuracy of 92.62%, thereby showing their effectiveness. Moreover, the proposed HFS-CC technique achieves 98.07% testing and 99.18% training accuracy. The better performance of HFS-CC is largely due to the discerning ability of the proposed geometric features and the developed hybrid feature space.

1. Introduction

Medical imaging has gained much importance in the last few decades, especially in analyzing different body parts for predicting certain disorders/diseases. Microscopic imaging is one of the medical imaging techniques, wherein the images of biopsy slides are captured. Biopsy images have well-defined organization of tissues and connected components, depending upon the body part from which they are taken [1]. The same is true for colon biopsy images, which are used in our problem for cancer detection. Biologically different constituents in a colon biopsy image can be identified by looking at the spatial organization of its constituents.

Microscopic analysis is the commonly practiced technique of colon cancer diagnosis, wherein histopathologists visually examine the deformation of tissues under microscope, and decide whether the geometric structure and organizational arrangement of sample tissues belong to the class of malignant or normal colon. Microscopic analysis is time consuming as well as subjective. The main reason behind subjectivity is the fact that quantitative cancer grades are assigned depending upon the observed morphology of tissues by the histopathologists. This process also leads to inter/intra-observer variability as quantitative grades assigned to the same sample by different histopathologists, or even by one histopathologist, may vary at times [2,3]. In order to alleviate such problems in diagnosis, researchers are working since long to find automatic quantitative tools, which could measure the degree of deformation and assign quantitative cancer grades to the colon samples.

The research in the field of colon cancer is in various dimensions. A larger subset of the colon cancer detection techniques has been summarized in a recent survey reported by Rathore et al. [4]. Some authors have performed analysis on hyperspectral data of colon biopsies [5,6]. In these schemes, authors select one spectral band amongst several bands of hyperspectral cube, calculate image features, and then based on these features classify samples

* Corresponding author at: DCIS, Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan. Tel.: +92 51 2207381 × 3102.
E-mail address: saimarathore_2k6@yahoo.com (S. Rathore).

http://dx.doi.org/10.1016/j.compbiomed.2015.03.004
0010-4825/

into multiple classes. Some researchers have studied thousands of human genes in parallel by using two variants of microarrays [7,8]. Their aim was to identify such genetic alterations, which were supposed to be responsible for colon cancer. Like genes, blood serum also deviates from its normal composition in case of colon cancer. Researchers have exploited this variation, and have used laser-induced fluorescence and Raman spectroscopy of blood serum for cancer detection [9,10].

Some researchers have exploited the variation in the texture of normal and malignant colon biopsy images for cancer detection. In this context, Esgiar et al. analyzed distinctiveness of six texture features (angular second moment, contrast, correlation, entropy, inverse difference moment, and dissimilarity) for classification of colon biopsy images [11]. They found the combination of entropy and correlation to be the most distinctive feature set, providing an overall accuracy of 90.2%. They further extended the idea by introducing fractal dimensions into the classification process, and proved that a combination of entropy, correlation and image fractal dimensions yields classification accuracy of 94.1% [12]. Followed by their work, Masood et al. proposed a few valuable methodologies for classification of colon. In their first method, they calculated morphological features of shape, size and orientation, and gray-level co-occurrence matrix (GLCM) based features of energy, inertia, and local homogeneity from colon biopsy images [13]. They employed polynomial SVM classifier, and achieved classification accuracy of 84% and 90% using morphological and GLCM based features, respectively. Masood et al. further extended the previous work [13], and calculated circular local binary patterns in order to classify colon biopsy images [14]. They obtained an accuracy of 90% by employing Gaussian SVM for classification. Further, Rathore et al. proposed a colon biopsy image based classification technique (CBIC) [15], wherein a hybrid feature set comprising traditional histogram of oriented gradients based features, and novel variants of statistical moments and Haralick texture features has been used for classification of colon biopsy images. A majority voting based ensemble of SVM classifiers has been used for classification, and 98.85% classification accuracy has been observed.

Recently, Altunbay et al. proposed a colon cancer detection technique [16], wherein they constructed a graph on different objects, obtained by using circle fitting algorithm [1] on the white, pink and purple clusters of colon biopsy image. Features of degree, average clustering coefficient, and diameter are computed from the graphs. The features are then used to classify given samples by using linear SVM kernel. In addition, Ozdemir et al. presented a method for automated colon cancer detection [17]. In this work, reference graphs of a few normal images are generated by employing previously proposed method of graph creation [18], and are stored for further referencing. Some query graphs are generated from the test images, and are searched in the reference graphs. Three most similar graphs are found in the reference images. Finally, normal or malignant class is assigned to the test sample based on the degree of similarity of the query graph with the three most similar graphs.

The techniques mentioned in the previous paragraphs have a few limitations. First, graph based techniques [16,17] are computationally expensive. Second, texture features based techniques [11–15] have exploited general texture features for classification, and have not exploited the background knowledge about the morphology of colon tissues for classification. Therefore, a computer-aided diagnostic technique, which could exploit the morphology of normal and malignant colon tissues in a computationally tractable manner, is required.

In this research, a novel hybrid feature space based colon classification (HFS-CC) technique has been proposed for robust and effective classification of colon biopsy images. We propose a novel feature type that mathematically quantifies the geometric structure of constituents of colon tissues. Further, we compute some other feature types such as morphological, SIFT, EFDs, and texture features, and combine those features with geometric features to form a hybrid feature set. HFS-CC differs from its previous counterparts in two aspects. First, it incorporates background knowledge about tissues organization into the classification process by introducing novel geometric features, thus leads to effective and promising results. Second, it caters different categories of features, and exploits positive aspects of each category. There are several abbreviations used in subsequent sections. These abbreviations are given in Table 1.

The remainder of this paper is organized as follows. Section 2 describes the structure of normal and malignant colon tissues. Section 3 presents the proposed scheme in detail. Section 4 describes performance measures. Section 5 demonstrates experimental results, and Section 6 concludes the paper.

Table 1
Abbreviations used in the text.

Acronym	Abbreviations
EFDs	Elliptic Fourier descriptors
FEM	Feature extraction module
H&E	Hematoxylin & Eosin
MCC	Matthews correlation coefficient
mRMR	Minimum redundancy and maximum relevancy
OSDU	Object spatial distribution uniformity
OSU	Object size uniformity
RBF	Radial basis function
SIFT	Scale invariant feature transform
SVM	Support vector machine

2. Cell structure: Normal and malignant colon tissues

Normal colon tissues have well-defined organizational structure. There are three constituents of a normal colon tissue, namely, epithelial cells, non-epithelial cells, and connecting tissues. The detailed structure of a normal colon tissue is shown in Fig. 1(a). Lumen lies in the middle of the tissue and is surrounded by epithelial cells that form glandular structure, whereas, non-epithelial cells lie in between glandular structures and are known as stroma. Cells and connected tissues are organized and coherent in case of normal colon. But, this organizational structure deviates considerably for malignant tissues as shown in Fig. 1(b)–(d).

Histopathologists analyze the samples under microscope and decide whether tissue is normal or not. Furthermore, histopathologists also assign quantitative cancer grades to the malignant colon samples. Grade of colon cancer is the differentiability level of malignant tissues from normal ones. There are three colon cancer grades: well-, moderately-, and poorly differentiable. In a well differentiable grade, tissues are slightly similar to normal ones as shown in Fig. 1(b). In this particular grade, cancer progresses at low speed. In moderately differentiable cancer grade, tissues are different from normal ones as shown in Fig. 1(c), and cancer progresses at moderate speed in this grade. In a poorly differentiable cancer grade, malignant tissues are totally different from normal tissues as shown in Fig. 1(d), and cancer progresses at very high rate in this particular grade.

3. Proposed scheme

The proposed HSF-CC scheme comprises six main stages, namely, preprocessing, feature extraction, feature reduction, feature concatenation,

Please cite this article as: S. Rathore, et al., Automated colon cancer detection using hybrid of novel geometric features and some traditional features, Comput. Biol. Med. (2015), http://dx.doi.org/10.1016/j.compbiomed.2015.03.004