Automatic Expert Diagnostic System for CRC Detection

Hiba A. Muhareb, Sara Tedmori, Kais Hasan
Princess Sumaya University for Technology, Jordan
Hiba.m99@gmail.com; S.Tedmori@psut.edu.jo; Q.Hassan@psut.edu.jo

ABSTRACT

Early detection of precancerous polyps that impersonate normal colon cells is very important to enhance the mortality of individuals with high risk of colorectal cancer. Hence, additional technologies for early colorectal cancer detection are required to improve the current detection methods.

Virtual colonoscopy is a new detection method that doesn't require any surgical preparation as opposed to optical colonoscopy. The aim of this paper is to utilize virtual colonoscopy in order to aid in the early detection of colorectal cancer. This will potentially increase the protective effect of early colonoscopy screenings.

In this paper, a fully automated colorectal cancer detection system is proposed. The system utilizes advanced image processing and machine learning techniques to extract polyps from 3D DICOM images. After which, the polyps are analyzed and classified as being either benign or malignant.

Keywords: Medical imaging, Machine learning, Colon cancer, Shape detection.

1 INTRODUCTION

Recent global studies have shown that every year Colorectal Cancer (CRC) causes death of almost half a million of the world's population [1]. According to a study conducted by the Jordan National Cancer Registry (JNCR), CRC is the second most frequent cause of death in Jordan. The study reports that the five most common cancers affecting Jordanians are breast, colorectal, lung, lymphoma and leukemia. Moreover, the study reveals that the crude of incidence rate of all cancer among Jordanians is 67.1 per population of 100,000 (63.9% males and 70.5% females) [2].

The estimates of the growing rate of colon cancer associated with ulcerative colitis are 2.5% at 10 years, 7.6% at 30 years, and 10.8% at 50 years. Patients at higher risk of cancer are those with a family history of colon cancer, a long duration of colitis, extensive colon involvement with colitis, and those with an associated liver disease, sclerosing cholangitis [3].

2 BACKGROUND

Recently, "virtual colonoscopy" (aka colonography) has been utilized as a screening technique for colorectal cancer. Virtual colonoscopy employs a Computerized Tomography (CT) scan using low doses of radiation to create images of the rectum and entire colon, which makes visualizing the colon and finding polyps or masses easier. The CT scan procedure typically involves a bowel preparation with laxatives, introducing air into the colon. CT is a much safer detection method than optical colonoscopy since no sedation nor surgery is required.

In studies comparing virtual colonoscopy and standard colonoscopy, patients seem to prefer virtual colonoscopy as minimal bowel preparation is required.

Similar to standard colonoscopy, virtual colonoscopy can detect irritated and swollen tissue, ulcers, and larger polyps (over 1 centimeter in size). Automating the process will consequently enhance the accuracy of the polyp detection by minimizing false positives via automated systems. Virtual colonoscopy is considered as a semi-automated method in CRC detection.

2.1 Detection methods for CRC

2.1.1 Classical Test

Fecal Occult Blood Test (FOBT) is the first step towards CRC detection. As the cancerous polyps can bleed, the FOBT is used to detect traces of blood that often go unnoticed with the naked eye [4],[5].

It is recommended to perform the FOBT every 1 to 2 years in people aged between 50 to 80 years. This helps reduce the number of deaths associated with colorectal cancer by 15 to 33 percent [4].

2.1.2 Manual Screening

Sigmoidoscopy is a test that allows doctors to examine the rectum and sigmoid colon using a flexible lighted instrument called a sigmoid-scope. The sigmoid-scope is introduced through the anus into the rectum and sigmoid colon [6].

Regular screening with Sigmoidoscopy for people above 50 years reduces the risk of death due to cancer of the rectum and lower colon by 60 to 70 percent [4].

Standard or Optical Colonoscopy is a test that examines the rectum and the entire colon using a flexible lighted instrument called a colonoscope. The colonoscope is inserted through the anus into the rectum and the colon. Colonoscopy reduces deaths from colorectal cancer by about 60 to 70 percent [4],[7].

2.1.3 Semi-automated

Virtual colonoscopy (aka computed tomographic [CT] colonography) is a fairly new method of screening in which a special x-ray equipment (a CT scanner) is used to produce pictures of the colon and the rectum from outside the body. A computer then assembles these pictures into detailed images that can show polyps and other abnormalities. The accuracy of virtual colonoscopy is similar to that of standard colonoscopy and it has a lower risk of complications. However, if polyps or other abnormal growths are found during a virtual colonoscopy, a standard colonoscopy is usually performed to remove them [8].

Double-Contrast Barium Enema (DCBE) is another method of visualizing the colon from outside the body. In DCBE, a series of x-ray images of the entire colon and rectum are taken after the patient is given an enema with a barium solution [4].

2.1.4 Fully Automated

MATCH is a special design for automatic diagnosis system to support colon cancer treatment [9].

2.2 Related work

The applications of artificial neural networks, medical image processing, pattern recognition and classification have improved the accuracy of cancer detection and survival prediction when compared to other statistical or clinicopathological methods [10]. Lashari & Ibrahim (2013) present a new framework for medical imaging classification based on soft set theory to achieve better performance in terms of accuracy, precision and computational speed. Furthermore, the proposed classification algorithm helps improve the physician’s ability to detect and analyze pathologies leading to more reliable diagnosis and treatment of diseases in semi-automatic diagnostic systems [11].

MATCH is a special design for automatic diagnosis system to support colon cancer treatment. MATCH has been proposed in order to overcome the difficulty of decision making in CRC selection of personalized treatment [9]. However, in order to emphasize the semantic modeling, researchers have developed the MATCH system with web services to ensure the interoperability between hospitals. Thus, a larger data set is clustered and similar patients can be automatically categorized. Unfortunately, more generalized diagnostics based on the patient grouping problem will arise.

Another utilization of the MATCH platform has been introduced by Goletsis et al [12]. The researchers performed data integration between medicine and molecular biology. The core of the proposed platform is based on clustering techniques. The system groups patients based on similar clinical features and genetic predisposition to cancer. Accordingly, the clinical and genetic data of the patient can be integrated and used in concluding a real time CRC diagnostics and the effective treatment will be based on the integrated information collected from similar cases.

CRC cell detection has been improved by Chaddad et. al, who used the Haralick’s features in order to enhance the classification of colon cells. In addition, the authors derived a detection approach from the ‘snake’ method using a progressive division of the dimensions of the image to achieve a faster segmentation decreased by 50% [13]. The author’s experimental results show that the application of Haralick’s coefficients enhances cancerous cell detection by obtaining multi spectral images. Results show that the method is efficient in the classification of the following types of cancer cells: Carcinoma (CA), Intraepithelial Neoplasia (IN) and Benign Hyperplasia BH.

The first stage is to judge if a cell is normal or cancerous. The second stage is to deal with the cells that are judged as cancerous by the first ensemble. These predictions are then combined by a prevailing method. Their experiment shows that the neural network ensemble can achieve a high rate of identification and a low rate of false negatives. However, when comparing any neural network model used in image processing with traditional image processing methods, it have been reported by several researches that although the process of training is not an easy task, it can solve medical image processing issues and the time for applying a trained neural network to solve a medical image processing problem is not perceptible [14] [15].

3 RESEARCH PROBLEM

Software to recognize anatomical anomalies in CT scans is required in order to increase the accuracy and enhance the diagnostic results.

Earlier CRC diagnostic systems work by matching the CT scans against a database of stored images. Due to the high variance in the polyps’ shapes, these systems usually have low accuracy.

Although, CRC diagnostic software can recognize various cancer stages, the pre-cancer detection of polyps that impersonate normal colon cells until they develop cancer is still problematic. Different approaches that employ artificial intelligence techniques have been proposed in order to improve cancer cell recognition in colonoscopy. One possible approach for early pre-cancerous polyp detection is the use of neural networks.

4 PROPOSED SYSTEM

In this paper, an Automatic Expert Diagnostic System (AEDS) in CRC is proposed. The AEDS software automatically processes CT scans from the DICOM image dataset, applying on them selected image processing techniques. The system uses a hybrid classification system to obtain a single report of the diagnosis.

The used DICOM images will go through a number of stages before the diagnosis is reported, as follows:

(a) Pre-processing
(b) Image segmentation
(c) Shape matching
(d) Feature extraction
(e) Classification
(f) Report generation

The software will be developed using MATLAB (R2008a, The MathWorks, Image processing toolbox Network, Neural toolbox). The dataset is provided by the National Electrical Manufacturers Association, 2003. Digital Imaging and Communications in Medicine (DICOM), Virginia, USA: National Electrical Manufacturers Association [16].

The next section provides more details regarding the proposed system.

5 METHODOLOGY

The main challenge that faces AEDS is to solve the defined polyp detection. The system has two main processes:

(a) Image processing and segmentation
(b) Classification

In (a), the image is firstly acquired and then processed to enhance the polyp distinguishability using required functions, such as noise removal.

In order to provide a suitable input for the classification, a segmentation process takes place ensuring the following steps:

• Edge detection: an edge of the connected pixels that have the same intensity level between two adjacent pixels will be distinguished by applying the proper operator.

• Boundary detection: the sets of pixels that have been determined using edge detection algorithm barely characterize a boundary completely, due to the effects of noise and non-uniform of the image brightness. Hence, edge detection algorithm is followed by boundary detection procedure in order to assemble edge pixels into meaningful regions.

• Feature extraction: as a final step of image segmentation process, the local region is extracted of the image in (m×n pixels) as an input for the classification process.

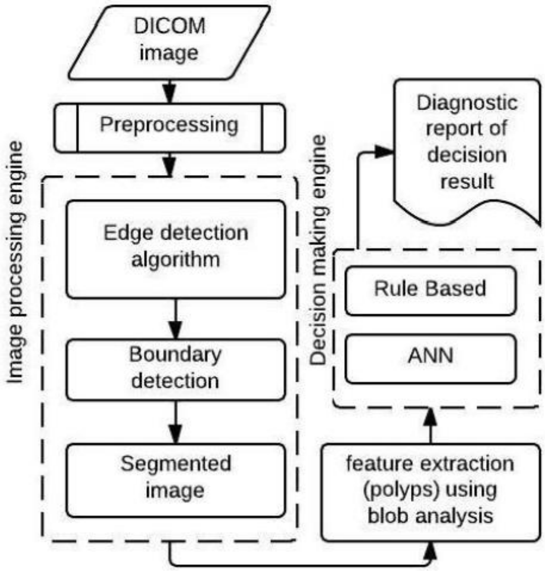

Figure 1: shows the AEDS main processes and the different stages of the presented expert diagnostic system. The used DICOM images go through a number of processes before diagnosis is reported.

Firstly; the dataset is loaded and scans that are in use are considered as an input. The selected scans are then preprocessed (if needed) to enhance the quality of its features. The image processing then takes place in order to extract targeted features. Edge detection algorithm is applied initially, followed by boundary detection to define the colon area. Extracted features will have values of brightness, intensity and depth information that will mainly be used to outline boundaries, which plays a predominant role in the classification process. With these features a segmented image will be used in the neural network to conduct results and generate final output as a diagnostic report.

6 CONCLUSION

Colorectal cancer is a significant cause of mortality and morbidity among world's populations. Hereafter, automated inspection systems of cancerous cells are highly required to enhance the manual methods in CRC detection.

Early detection of CRC relatively, improves the subject's quality of life and increase the likelihood of survival.

The proposed Automatic Expert Diagnostic

System (AEDS) serves the purpose of early CRC detection, with high rate of accuracy, by utilizing Artificial Neural Network (ANN) with image processing algorithms, using virtual colonoscopy scans as the input.

REFERNCES

[1] L. M. Cabal, "Colorectal cancer epidemiology. Assessing," in *The Challenge of Colorectal Cancer: A Review Book*, E. U. Cidón, Ed., India, Research Signpost, 2011.

[2] Jordan Center Registry, "King Hussein Cancer Foundation," 2008. [Online]. Available: http://www.khcc.jo/National_cancer_statistics.aspx. [Accessed 2014].

[3] F. W. Nugent, "Medicine net," 2014. [Online]. Available: http://www.medicinenet.com/colon_cancer/page2.h t m. [Accessed 2014].

[4] National Cancer Institute, “Tests to Detect Colorectal Cancer and Polyps”, 2013 Available: http://www.cancer.gov/cancertopics/factsheet/detection/colorectal-screening. [Accessed 2014].

[5] Web MD, “Stool Tests for Colorectal Cancer”, 2012. [Online]. Available: http://www.webmd.com/colorectal-cancer/fecal-occult-blood-test-fobt. [Accessed 2014].

[6] Web MD, “Sigmoidoscopy (Anoscopy, Proctoscopy)”, 2011. [Online]. Available: http://www.webmd.com/digestivedisorders/sigmoidoscopy-anoscopy-proctoscopy. [Accessed 2012].

[7] Web MD, “The Basics of Colonoscopy”, 2005. [Online]. Available: http://www.webmd.com/digestivedisorders/colonoscopy-overview. [Accessed 2014].

[8] Web MD, “Virtual Colonoscopy”, 2011. [Online]. Available: http://www.webmd.com/digestivedisorders/tc/virtual-colonoscopy-topic-overview. [Accessed 2014].

[9] J. Siddiqi, B. Akhgar, A. Gruzdz, G. Zaefarian, A. Ihnatowicz, "Automated Diagnosis System to Support Colon Cancer Treatment: Match ," *Fifth International Conference on Information Technology: New Generations,* pp. 201-205, 2008.

[10] K-J. Kim and S-B. Cho, "Prediction of colon cancer using an evolutionary neural network", Artificial Life Volume 12, Number 1 181, 2004.

[11] S. A. Lashari and R. Ibrahim, " A Framework for Medical Image Classification Using Soft Set," in *4th International Conference on Electrical Engineering and Informatics, vol 11*, pp. 548-556, 2013.

[12] Y. Goletsis, T.P. Exarchos, N. Giannakeas, D.I. Fotiadis, "Intelligent patient profiling for diagnosis, staging and treatment selection in colon cancer," 8th IEEE International Conference on BioInformatics and BioEngineering, pp.1-6, 2008.

[13] A. Chaddad, C. Tanougast, A. Dandache, A. Al Houseini, A. Bouridane, "Improving of colon cancer cells detection based on Haralick's features on segmented histopathological images," IEEE International Conference on Computer Applications and Industrial Electronics, pp.87-90, 2011.

[14] K. Suzuki, I. Horiba, N. Sugie, M. Nanki, "Neural filter with Selection of Input Features and its Application to Image Quality Improvement of Medical Sequences," *IEICE transaction on information and systems,* vol. E85-D (10), pp. 1710-1718, 2002.

[15] Z. Shi, L. He, "Application of neural networks in medical image processing," in Proceedings of the *Second international symposium on networking and network security,* pp. 23-26, 2010.

[16] The cancer imaging archive, National Cancer Institute, Washington, USA, 2011.