International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

Colorectal Cancer Detection from Colonoscopy Videos: A Comprehensive Survey of Computer-Aided Diagnosis Techniques

Manveer Kaur Saini, Mamta Juneja, Ajay Mittal
University Institute of Engineering and Technology, Panjab University, Chandigarh

ABSTRACT

Colorectal cancer (Colon and rectum carcinoma) is a major health issue across the world and a leadingcause of deaths due to cancer. It is the third most common form of cancer resulting in death in men and the second most common form of deaths in women.It initiates as a non-cancerous tumour but becomes carcinogenic with its growthon the colon wall lining. However, it can be detected at an early stage with regular screening and recent advancements in CAD becomes an effective tool in detection of polyps. Hence there is a need of CAD systems to detect the colorectal polyps at an early stage.In the current scenario research is going on to improve the performance of the CAD system. This survey paper presents a review of CAD and its advancements with focus on the techniques and importance of the CAD systems.

Keywords

CAD; Polyp detection and classification; Feature extraction; CNN; Transfer learning.

INTRODUCTION

Colorectal cancer, a form of gastrointestinal malignancy, is the third major cause of cancer related deathsin men andthe second major cause of cancer deaths in women in US [1].According to an estimate by ACS (American Cancer Society),in the year 2017, 135430 new cases of colorectal cancerwere foundin US and 50,260 peopleapproximatelydied due to this disease. The chance of developing colorectal cancer in men is 4.6% (i.e., 1 in 22) and in women it is 4.2%(i.e., 1 in 24). This disease starts as a non-cancerous growth on the colon's or the rectum's inner lining(which is called a polyp) and takes around 10-20years for growth[2]. The prognosis of colorectal cancer is based upon its early detection to prevent polyps from converting into cancer.Detection of polyps is done by using different invasive as well as non-invasive techniques such as colonoscopy and virtual colonoscopy. However, it is a difficult task to detect polyps in colonoscopy videos as the polyps vary in size, shape, location and texture. The detection and localization of polyps is affected by superimposed anatomical structure which may be missed by an expert without computer-aided diagnostic system. Moreover, there may be significant intra-observer and inter-observer variations.The capability of computers has been well established in the automated diagnosis of various diseases. There are variousCAD systems forlung cancer,breast cancer, colorectal cancer and Tuberculosis(TB). But due to less accuracy of the existing systems there is a need of improvised detection systems for early detection of polyps. Hence this paper analyzes the shortcomings of the existing systems to propose an improvised system in future.

STRUCTURE OF POLYPS

A polyp is an unnatural small vascular tissue growth on the surface of a mucous membrane and exists intwo forms. First is called pedunculated where the polyp is connected to the surface by an elongated narrow stalk. Second is called sessile where stalk is not present. Polyps can be benign(non-cancerous) but can turn malignant over time. Polyps can be discovered during early ages of patients but they usually become cancerous over time after a certain age limit(usually >50 years of age). Various reasons contribute to the conversion of polyp into cancer. Two of them are the size of the polyp and degree of dysplasia. Dysplasia is

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

the abnormal development of organs or cells. The polyps are generally found in stomach, nose, colon, sinus(es), ear, uterus and urinary bladder.

A colorectal polyp is a fleshy growth on the colon or rectum's inner lining. If left untreated, they can turn into colorectal cancer.

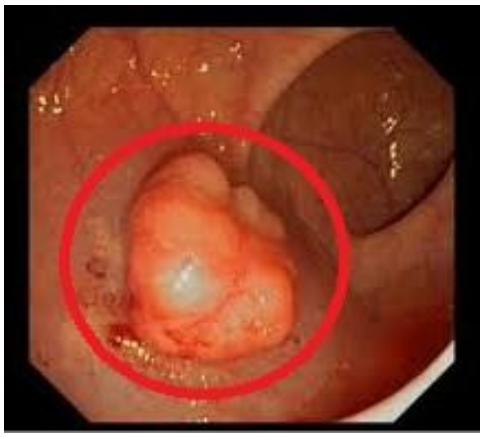

Fig 1. Location of polyp in colon

Types of colorectal polyps

The most common types of colorectal polyps are:

(i) Hyperplastic Polyp: These are mostly found in the distal colon and the rectum;and usually don't turn malignant(cancerous).

(ii) Neoplastic Polyp: Neoplasm is an abnormal growth of a tissue whose cells have lost their differentiation nature and can be either benign or malignant. The benign neoplastic polyp is called adenoma.

(iii) Hamartomatous Polyp: These occur as growths on organs due to faulty development.

(iv)Inflammatory Polyp: Polyps associated with inflammations such as Ulcerative Colitis and Crohns disease.

CAD PROCESSING AND PERFORMANCE METRICS

1. CAD Processing

Nowadays polyp detection has become an easy task due to automated machine learning methods.

Automated computer-aided diagnostic systems not only detect polyps but also tell the location of the polyp based upon their curvature (Polyp curvature forms the basis of all the CAD systems).

In most of the cases, the algorithms are compared on the basis of less number of large polyps. Here a brief working of the CAD system is discussed. Firstly, to perform segmentation for finding the polyp, CAD makes use of the curvature method. Then,each polyp is assigned a value which is used to measure them according to their texture, area, shape, energy, intensity and classify them using classifiers.

2. Performance metrics

Two most commonly used performance measures for CAD system areSensitivity and Specificity

(i) Sensitivity

Sensitivity is defined as if a person has a disease then how many times the test will be positive(true positive rate). The CAD algorithms must be highly sensitive to detect even small sized polyps (our main focus is to achieve 100% sensitivity).

Sensitivity= true positives/(true positive + false negative)

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

(ii) Specificity

Specificity is defined as if a person does not have a disease then how many times will the test be negative (true negative rate or number of false positives). CAD schemes have not been able to achieve 100% specificity till now. This is due to the complex structure of the polyps. Folds, residual colonic materials result in false positives. CAD schemes are able to find 100% true positives but only when polyp size is large. All CAD schemes initially apply segmentation on the images to detect polyps. Then for identification it is required to reduce the false positives which could be folds in the colon wall etc. There is always a tradeoff between these two measures. Lower the probability of missing the polyps higher is the sensitivity of the algorithm but higher in the number of false alarms. On the other hand, CAD systems with lower sensitivity have less performance of the CAD algorithm and are not able to detect small sized polyps. The main aim is to design a CAD system with 100% sensitivity and 100% specificity.

Specificity= true negatives/(true negative + false positives)

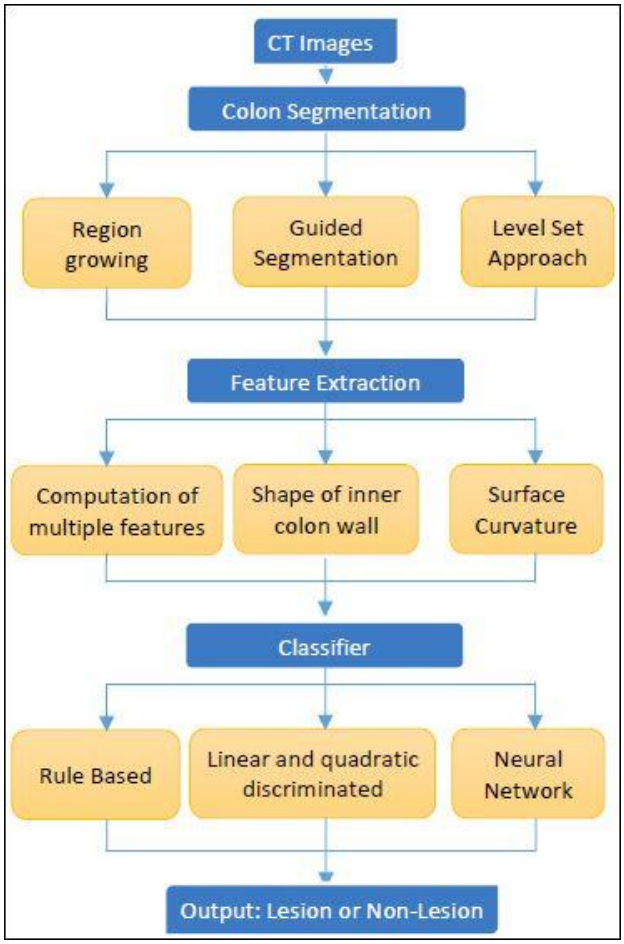

Fig 2. CAD scheme Layout

LITERATURE SURVEY

There is inefficiency of human observations due tohuman visual system limitations, carelessness and fatigue and human observations have a risk of inter-and intra-observer variations. This led to employ CAD to detect the colorectal polyps. The research in CAD to detect colorectal polyps has been going for around a decade. This survey has reviewed various techniques used in CAD for colorectal cancer detection. Some of the features to distinguish malignant polyps are: shape, colour, size, texture, curvature and haustral folds.

In 2000, Summers et al. [3] proposed a technique to detect polyps based on local variations of the curvature of the colonic wall. The parameters used to filter anomalies and to reduceFPs are: dimensionless ratio sphericity, mean curvature and minimum polyp size. Yoshida et al. [4] in 2001 proposed a 3-D CAD technique based on the existing CAD system. The method extracted a thick region covering the entire colonic wall which not only contained the surface but the complete internal structure of the polyp also. It helped in differentiating colonic folds and colonic wall. They used 3-D geometric features of volumetric shape index(topological shape of the volume) and volumetric curvedness(size of the polyp) for characterizing polyps and used fuzzy clustering forclustering and volumetric feature gradient(GC) and directional gradient(DGC) for classification. Huang et al. [5] used cubic B-spline, paraboloid and quadratic polynomials to compute curvature. Pilkinton et al. [6] used non-linear edge-conserving smoothing filters to enhance automated polyp detection in CTC bydecreasing the number of false positives per patient.

Jerebko et al. [7] analyzed symmetry in curvature patterns of colonic lumen using 3D Canny Edge detection for finding boundaries between the tissue and the colon lumen. He then calculated the Mean and Gaussian curvatures for every boundary voxel by using the original intensity values. Further Pickhardt et al. [8] used eigen values of the Hessian matrixrather than calculating curvature on the voxel grid. Initially CAD and its different techniques focused on the polyp features. But Dundar et al. [9] used a technique called optimized cascadeclassifierwhich minimized the objective function by using the AND-OR framework. Suzuki et al. [10] developed a MTANN (Massive Training Artificial Neural Network) for increasing the CAD performance on FN cases in a substantial clinical test. To differentiate between polyps and non-polyps,3D MTANNs were developed and had achieved 71.4% sensitivity on polyps that were not detected by the radiologists. This method suffered from a drawback that still 5 false-positives per patient were generated and the method categorized them into easy, moderate and difficult. To overcome this limitation, Aman et al. [11] used Content Based Image Retrieval(CBIR) to minimize the FP rate. In this, the detections were ranked based on their likelihood of being false-positives. Low ranked features were eliminated and this almost halved theFP detections from the CADe system and maintained high sensitivity.

The existing CAD systems suffered from three problems. Firstly, scatter effect on neighbouring tissues in fecal-tagging CTCdue to high density contrast agents. Secondly, for surface curvature computation the kernel approach yielded erroneous results for thin structures(6-8mm). Thirdly, over 150 features are selected from all polyp candidates and this high dimensionality of features doesn't allow good feature boundaries and reduces accuracy of predictions. To overcome these problems, Liu et al. [12] introduced three new techniques. First, to reduce pseudo enhancement effects scale-based scatter correction algorithm was developed. Second, to estimate curvatures accurately a cubic spline interpolation method was used. Third, DMLLE (Diffusion Maps and Local Linear Embedding) was used forFP reduction and classification. Tu et al. [13] gave a method for 3D detection of polyps from CT images. He adopted Probabilistic Boosting Tree(PBT) for detecting polyps probabilistically; gave a denotative analysis of convergence rate for AdaBoost algorithm without the need for colon segmentation and achieved 92% sensitivity on polyps of size larger than 6mm.

Ju et al. [15] used curvature tensor smoothing and spectral compression to reduce the number of false-positives(FP), caused by minute bumpy structures and disconnected curvature fields of a discretized surface or volume, and preserve true positives. The method gave a sensitivity of 96% for polyp size >10mm and reduced FP by 92%. Kilic et al. [16] usedgenetic algorithm, cellular neural network(CNN), and 3-D template matching along with fuzzy rule-based thresholding to detect colon polyps in CT images. The method gave high sensitivity and satisfactoryFP rate for images with area broken into templates but did not give high accuracy and high sensitivity when whole area is considered as one. Fiori et al. [17] introduced a CADe system to flag candidate polyp based upon texture and geometric features. The method also considered surrounding area for each candidate polyp(by computing differential features instead of absolute features) and tested regions of multiple sizes. But there was a limitation that the FP rate did not decrease much in comparison to previous CAD scheme.

Table 1 [14]gives a comparative view of different CAD techniques by different authors on the patients based on sensitivity, FP rate and the runtime of the algorithm.

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

Table 1. CAD results in previous studies

Algorithm	Data Centre	Training	Testing	FP/ volume	Polyp Sensitivity	Runtime per vol.
Jerebko[7]	N/A	96	72	4.3	94%(5mm+)	
Tu[13]	N/A	117	35	5.8	92%(6mm+)	
Pickhardt[8]	N/A	15	35	3.1	77.1%(6mm+)	
Dundar[9]	2	169	201	5	86%	4min 43sec
Suzuki[10]	1	10	73	0.55	96.4%(5mm+)	
Liu[12]	3	395	791	2.5	79%(6-9mm)	
Ravesteijn[19]	4	86	307	4-6	85-100%	

Additionally Simona et al. [18] performed detection and segmentation of significantly large lesions in CTC which improved scheme's performance. The classical CADe scheme gave 78 percent sensitivity and this method gave 83 percent sensitivity. But it did not increase accuracy. Ravesteijn et al. [19] used linear logistic classifier(logistic regression) to order candidate polyps based upon three features: colon wall protrusion, feature to cast away detections on the rectal clyster tube and mean internal intensity.The classifier gave better results even in the following cases:dataset for training should have less polyps, a significant imbalance must be thereamong polyp and non-polyp candidates, a small feature space, imbalanced and unidentified misclassification costs, and an exponential distribution in feature spacewith respect to size of the candidate. Jian et al. [20] used MTSVM (Massive Training Support Vector Machine) and MTGPR(Massive Training Gaussian Process Regression), instead of MTANN(Massive Training Artificial Neural Network) which has long training time, for the reduction of FPs in CADe of polyps. For classification of polyps, a 3-dimensional scoring method was applied to the outputs of MTSVM and MTGPR. Ong et al. [21] classified the polyp surface as either, ridge-like fold,bulbous or semi-planar structure using geodesic-ring neighbourhood on which orientation and curvature methods were applied. The method gave 100percent sensitivity as well as specificity for polyps>10mm in size. However, the sensitivity and accuracy decreased for polyps<5mm and it also missed some flat polyps and non-polyp lesions.

Fan et al. [22] introduced graph embedding semi-supervised feature extraction with adaptive kernel support vector machine(AK-SVM) which used prior labeling information as constraints to feature selection framework based on graph. The method increased the accuracy of classification to 93.7%. The major drawback of this scheme is that it needs prior labeling information to work otherwise it is equivalent to other techniques for selecting features. Ismail et al. [23] introduced a segmentation framework which was fully automated and used convex formulation with some anatomical featuresof the active contour model. It used geometrical descriptors to classify colon and non-colon parts from the output of the segmentation. The method gave high sensitivity and an accuracy of 99%. However the method is suitable for only large polyps and doesn't maintain accuracy for polyp size 5-9mm. XU JW et al. [24] proposed a method feature selection based upon SFFS (Sequential Forward Floating Selection) which maximizes the AUC (Area Under the receiver operating characteristic Curve) combined with non-linear SVM (Support Vector Machine) classifier. For testingits performance, the proposed method was compared withmethod based on Wilki's lambda for stepwise feature selection. However the scheme is not suitable for imbalanced datasets.

Bae et al. [25]came up with a method to learn with unbalanced dataset using boosting framework with up- and down- sampling and used partial least square(PLS) analysis for compact and discriminative feature learning. However, there is a challenge due to small amount of polyp samples despite polyp diversity and difficulty in polyp and non-polyp classification due to similar colour and texture features. Yang et al. [26] proposed graph inference methods to deal with adjacent air-filled organs' existence like lungs, small intestine andstomach and the break down of colon by poor insufflation to achieve high quality segmentation. However, the method did not provide information about the region of the polyp and was limited to small polyps.

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

A new trend in CADe of colorectal polyps is the use of deep learning. Tajbakhsh et al. [27] fine-tuned a convolutional neural network(CNN) which had been trained from a large set of labeled natural images. He used Canny algorithm to get edge-map, removed non-polyp edges from itwith the help of feature extraction and then edge classification. He then used voting scheme to localize polyp candidates in refined edge maps and to each generated candidate assigned a probabilistic confidence value. The method used context information for removimg non-polyps and shape information for localizing polyps having structures with curved boundaries. Ribeiro et al. [28] proposed the use of CNN for automated polyp classification in colon cancer screening. The proposed method used small patches(subimages) to increase the database size and also to classify different regions in the same image. Zhang et al. [29] proposed CNN (Convolutional Neural Network) for detectingpolyps by the transfer of low-level CNN characteristics from the non-medical domain. Endoscopic images were input to the CNN. Two CNNs were used; first one was used to detect polyps and the second one to classify the polyp(SVM was used here). The method was suitable where data available is less and interclass similarties are subtle. Ribeiro et al [30]came up with a method for colonic polyp classification with CNN which used transfer learning from texture database. CNN was used as feature extractor and the feature vectors from the last fully connected layers became input to SVM for performing the final classification.

Table 2[14] shows various polyp features along with the techniques and classifiers used and various performance parameters.

Table 2. Various CAD methods with different features and their observed efficiency

Method	Feature set	Accuracy	Specificity	Sensitivity	Dataset	Classifier
Geometric and texture feature based on flagging[17]	Polyp size of 4-10mm, texture and geometric feature	94%	88%	95%	1300 images	Computer-aided diagnosis (CAD)
Graph embedding semi-supervised[22]	Eliminate redundant and noisy feature; discriminate and geometric structure of an image,	93.7%	74.7%	95.4%	786	Adaptive kernel-support vector machine (AK-SVM)
Max-AUC feature with SFFS[24]	Curvedness shape index, and haustral folds	91%	90%	95%	206	SVM
Data sampling boosting framework[25]	Imbalanced images, shape, texture	89%	81%	95%	2365	Adaboost

CONCLUSION

Early detection of colorectal cancer is required for proper treatment and curbing of disease. Due to different colours, shapes and textures polyps are complex and thus difficult to differentiate. So, automated polyp detection techniques are needed. This review paper presents different colorectal cancer detection and

classification techniques using CAD. After studying and analyzing features of polyps in colon and rectum, it can be concluded that polyps can be detected with their well-defined and round shapes. There can be further improvement in the reviewed CAD schemes. There could be separate study for tuning parameters of these techniques and secondly, to achieve optimization and prevent over-fitting of large databases, cross-validation methods could be used.

REFERENCES

1. Dongqing Chen, Lihong Li, M. Wax, B. Li, A. Kaufman and Zhengrong Liang, "Electronic colon cleansing by colonic material tagging and image segmentation for polyp detection: detection model and method evaluation," *2000 IEEE Nuclear Science Symposium. Conference Record (Cat. No.00CH37149)*, Lyon, 2000, pp. 18/131-18/135 vol.3.
2. Z. Kerekes, Z. Tóth, S. Szénási, Z. Tóth and S. Sergyán, "Colon cancer diagnosis on digital tissue images," *2013 IEEE 9th International Conference on Computational Cybernetics (ICCC)*, Tihany, 2013, pp. 159-163.
3. Summers, Ronald M., et al. "Automated polyp detection at CT colonography: feasibility assessment in a human population." Radiology 219.1 (2001): 51-59.
4. Yoshida, Hiroyuki, and Janne Nappi. "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps." IEEE transactions on medical imaging 20.12 (2001): 1261-1274.
5. Huang, Adam, Ronald M. Summers, and Amy K. Hara. "Surface curvature estimation for automatic colonic polyp detection." Medical Imaging 2005: Physiology, Function, and Structure from Medical Images. Vol. 5746. International Society for Optics and Photonics, 2005.
6. Pilkinton, David, et al. "The effect of edge-preserving image smoothing on automatic colonic polyp detection for CT colonography." Medical Imaging 2006: Physiology, Function, and Structure from Medical Images. Vol. 6143. International Society for Optics and Photonics, 2006.
7. Jerebko, Anna, et al. "Symmetric curvature patterns for colonic polyp detection." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Berlin, Heidelberg, 2006.
8. Pickhardt, Perry J., et al. "Low rates of cancer or high-grade dysplasia in colorectal polyps collected from computed tomography colonography screening." Clinical Gastroenterology and Hepatology 8.7 (2010): 610-615.
9. Dundar, M. Murat, and Jinbo Bi. "Joint optimization of cascaded classifiers for computer aided detection." Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007.
10. Suzuki, Kenji, et al. "Massive-training artificial neural networks for CAD for detection of polyps in CT colonography: false-negative cases in a large multicenter clinical trial." Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on. IEEE, 2008.
11. Aman, Javed, Jianhua Yao, and Ronald M. Summers. "Reducing the false positive rate of computer aided detection for CT colonography using content based image retrieval." Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on. IEEE, 2009.
12. Liu, Jiamin, et al. "High-performance computer aided detection system for polyp detection in CT colonography with fluid and fecal tagging." Medical Imaging 2009: Computer-Aided Diagnosis. Vol. 7260. International Society for Optics and Photonics, 2009.
13. Tu, Zhuowen, et al. "Probabilistic 3D polyp detection in CT images: The role of sample alignment." Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 2. IEEE, 2006.
14. Mittal, Ajay, and Manvjeet Kaur. "Computer-aided-diagnosis in colorectal cancer: A survey of state of the art techniques." Inventive Computation Technologies (ICICT), International Conference on. Vol. 1. IEEE, 2016.
15. Ong, Ju Lynn, and Abd-Krim Seghouane. "False positive reduction in CT colonography using spectral compression and curvature tensor smoothing of surface geometry." Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on. IEEE, 2009.
16. Kilic, Niyazi, Osman N. Ucan, and Onur Osman. "Colonic polyp detection in CT colonography with fuzzy rule based 3D template matching." Journal of medical systems 33.1 (2009): 9.
17. Fiori, Marcelo, et al. "Automatic colon polyp flagging via geometric and texture features." Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE. IEEE, 2010.
18. Grigorescu, Simona E., et al. "Automated detection and segmentation of large lesions in CT Colonography." IEEE Transactions on Biomedical Engineering 57.3 (2010): 675-684.
19. Van Ravesteijn, Vincent F., et al. "Computer-aided detection of polyps in CT colonography using logistic regression." IEEE transactions on medical imaging 29.1 (2010): 120-131.
20. Xu, Jian-Wu, and Kenji Suzuki. "Massive-training support vector regression and Gaussian process for false-positive reduction in computer-aided detection of polyps in CT colonography." Medical Physics 38.4 (2011): 1888-1902.

21. Ong, Ju Lynn, and Abd-Krim Seghouane. "From point to local neighborhood: polyp detection in CT colonography using geodesic ring neighborhoods." IEEE Transactions on Image Processing 20.4 (2011): 1000-1010.
22. Fan, Lei, et al. "Semi-supervised graph embedding-based feature extraction and adaptive kernel-based classification for computer-aided detection in CT colonography." Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE. IEEE, 2012.
23. Ismail, Marwa, et al. "3D automated colon segmentation for efficient polyp detection." Biomedical Engineering Conference (CIBEC), 2012 Cairo International. IEEE, 2012.
24. Xu, Jian-Wu, and Kenji Suzuki. "Max-AUC feature selection in computer-aided detection of polyps in CT colonography." IEEE journal of biomedical and health informatics 18.2 (2014): 585-593.
25. Bae, Seung-Hwan, and Kuk-Jin Yoon. "Polyp detection via imbalanced learning and discriminative feature learning." IEEE transactions on medical imaging 34.11 (2015): 2379-2393.
26. Yang, Xiaoyun, Xujiong Ye, and Greg Slabaugh. "Multilabel region classification and semantic linking for colon segmentation in CT colonography." IEEE Transactions on Biomedical Engineering 62.3 (2015): 948-959.
27. Tajbakhsh, Nima, et al. "Convolutional neural networks for medical image analysis: Full training or fine tuning?." IEEE transactions on medical imaging 35.5 (2016): 1299-1312.
28. Ribeiro, Eduardo, Andreas Uhl, and Michael Häfner. "Colonic polyp classification with convolutional neural networks." Computer-Based Medical Systems (CBMS), 2016 IEEE 29th International Symposium on. IEEE, 2016.
29. Zhang, Ruikai, et al. "Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain." IEEE journal of biomedical and health informatics 21.1 (2017): 41-47.
30. Ribeiro, Eduardo, et al. "Exploring texture transfer learning for colonic polyp classification via convolutional neural networks." Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on. IEEE, 2017.