DE G DE GRUYTER OPEN

JIM Journal of Interdisciplinary Medicine

CLINICAL UPDATE

INTERNAL MEDICINE // GASTROENTEROLOGY

Computer-Aided Diagnosis in Colorectal Cancer: Current Concepts and Future Prospects

Andrei-Constantin Ioanovici[1], Andrei-Marian Feier[1], Ioan Țilea[1,2], Daniela Dobru[1,3]

[1] University of Medicine and Pharmacy, Tîrgu Mureș, Romania
[2] Department of Clinical Science-Internal Medicine, University of Medicine and Pharmacy, Tîrgu Mureș, Romania
[3] Department of Gastroenterology, County Emergency Clinical Hospital, Tîrgu Mureș, Romania

CORRESPONDENCE

Ioan Țilea
Str. Revoluției nr. 35
540042 Tîrgu Mures, Romania
Tel/Fax: +40 265 211 011
E-mail: ioan.tilea@umftgm.ro

ARTICLE HISTORY

Received: June 14, 2017
Accepted: July 3, 2017

Andrei-Constantin Ioanovici • Str. Gheorghe Marinescu nr. 38, 540139 Tîrgu Mures, Romania. Tel: +40 265 215 551

Andrei-Marian Feier • Str. Gheorghe Marinescu nr. 38, 540139 Tîrgu Mures, Romania. Tel: +40 265 215 551

Daniela Dobru • Str. Gheorghe Marinescu nr. 38, 540139 Tîrgu Mures, Romania. Tel: +40 265 215 551

ABSTRACT

Colorectal cancer is an important health issue, both in terms of the number of people affected and the associated costs. Colonoscopy is an important screening method that has a positive impact on the survival of patients with colorectal cancer. The association of colonoscopy with computer-aided diagnostic tools is currently under researchers' focus, as various methods have already been proposed and show great potential for a better management of this disease. We performed a review of the literature and present a series of aspects, such as the basics of machine learning algorithms, different computational models as well as their benchmarks expressed through measurements such as positive prediction value and accuracy of detection, and the classification of colorectal polyps. Introducing computer-aided diagnostic tools can help clinicians obtain results with a high degree of confidence when performing colonoscopies. The growing field of machine learning in medicine will have a big impact on patient management in the future.

Keywords: colorectal polyps, machine learning, computer-aided diagnosis, colonoscopy

INTRODUCTION

As of 2012, the estimated incidence of colorectal cancer (CRC) is 1.36 million, marginally higher in men than women, making it the third most frequent cancer worldwide. In terms of mortality, in which colorectal cancer ranks fourth, there are approximately 700,000 deaths every year.[1] The estimated costs for CRC, including medical costs, non-medical costs, and income losses, were approximately 32 billion USD in 2010, and these costs are expected to rise by 2030.[2] The introduction of screening for CRC by means of colonoscopy has led to improved survival compared to patients who presented with symptoms initially, since the disease was primarily more advanced in the latter category of patients.[3] Improved diagnosis and medical economic burden are currently being addressed, since researchers have been focusing on developing computer-aided diagnosis

(CAD) tools based on the concept of machine learning.[4] Given that the use of CAD tools in medicine could have a big positive impact in the near future, we aim to present the current level of development of these systems, pointing out several directions for implementation, as well as their drawbacks.

THE BASICS OF MACHINE LEARNING ALGORITHMS

Regarding the management of CRC, a number of methods using machine learning algorithms have been proposed and are currently under research. Among them are the training of algorithms, such as artificial neural networks (ANNs), to predict colorectal cancer survival based on a number of variables, or the analysis of the colon from endoscopic images, distinguishing between neoplastic lesions subtypes based on complementary DNA microarray data.[5–7] Our focus is on machine learning methods that are related to the endoscopic approach of colorectal cancer management.

In order to be of use in the endoscopic approach of colorectal cancer, machine learning algorithms need to be trained on certain datasets of preprocessed images. The video frames and images retrieved from colonoscopy need to go through a process called visual feature extraction and description. Common feature extraction methods make use of color, shape, and texture, and in order to achieve effective recognition, regions of interest need to be extracted from an image. Descriptors that derive from the process are then used for image representation, indexing, and recognition. Classification algorithms such as ANNs and support vector machines (SVMs) are trained on these descriptors and used for the classification and detection of colon lesions (polyps).[8,9]

ANNs were inspired by the design of the human brain and are models for computation consisting in processing elements (neurons), organized into successively connected layers. Usually there are input layers, where different types of data are presented, and output layers that give the response of the network. Between them are hidden layers that are used for processing data. This is the basic design of a feedforward network, one of the main architectures of neural networks that need to be trained in order to be able to solve complex problems. Training is achieved by the use of a backpropagation algorithm, a feedback process in which the output of a network is compared to a certain desired output, and the difference between the two outputs is then used to adjust the network. With repetition, the difference gets smaller, and the output of the network can be identical to the desired output, thus being able to give a solution to a problem. Training can be done in multiple ways, for instance by supervised learning, in which the desired output is known by the network and adapts from it, and by unsupervised learning, in which the network learns based on observation of responses to inputs. The former method is the most frequently used in ANNs for cancer research.[10,11] SVMs are supervised machine learning algorithms associated with learning algorithms that analyze data and recognize specific patterns and are mainly used for classification, as well as for regression analysis.[12]

CLINICAL APPLICATIONS OF ARTIFICIAL NEURAL NETWORKS IN COLORECTAL DISEASE

Maroulis *et al.* (2003) developed an ANN for the recognition of premalignant lesions (polyps) from video frame sequences and single images. To describe the images, texture-based features were extracted. The study reported a success rate of over 95%, without retraining required for matching types of lesions, even when applied to different patients, and stated that the incorporation of distinct classifiers and feature extraction tools could supply even higher detection success rates.[13] Oliva *et al.* (2016) created a prototype system with the purpose of aiding in the study and analysis of colon tissue images. Texture-based feature extraction was employed, along with the building of different types of classification algorithms to evaluate image fragments. One classifier, namely the naïve Bayes classifier, obtained the best results in terms of positive predictive value (PPV) – 90.2% and specificity – 92.54%. The study concluded that the system can assist medical professionals in carrying out a better evaluation of medical images, by pointing out patterns related to diseases such as CRC.[14] Tjoa *et al.* (2003) proposed a hybrid method to classify the status of the colon by combining texture- and color-based feature extraction that relies on colonoscopy images. The assessment was made using a backpropagation neural network associated with principal component analysis (PCA) and reportedly had a classification accuracy of 97.72 % of the images retrieved from colonoscopy. PCA is an important step, since it reduces the dimensions of images while retaining all the valuable information about them.[15] In a study by Biswas *et al.* (2016), 1,185 images of normal colon aspect, as well as colon polyps, Crohn's disease, colon ulcer, and sigmoid colon cancer images were used to train and test the classifier, a multiclass SVM, whereas feature extraction was realized by the application of cross-wavelet transform (CWT), and image reduction was done by PCA.

The results were promising, as classification accuracy was 98.46% for all diseases and 98.83% for sigmoid colon cancer alone, ultimately being able to save time for the colonoscopy procedure.[16]

HYBRID METHODS IN DIAGNOSING COLORECTAL CANCER

Ribeiro *et al.* (2016) made use of a database consisting of two classes that contained 25 healthy images from 18 patients and 75 abnormal images from 56 patients for the training of a convolutional neural network (CNN).[17] Data augmentation was performed, thus 800 images resulted, while only 75 subimages from each healthy image and 25 subimages from the abnormal images were used to train the classifier. The classification accuracy of colon polyps was 90.96% and the sensitivity was 95.16%, thus reporting a good score for false negatives, while the false positive rate was high, the specificity being only 74.19 %.[17]

Throughout various endoscopy sessions, the distance, zoom, and perspective towards the walls are not always the same. A study by Uhl el al. (2014) emphasized the importance of texture descriptors that are invariant to scale and viewpoint conditions and proposed a novel descriptor that had 84.6% accuracy in classifying images into two classes compared to other methods. The first class was comprised of normal mucosa and hyperplastic polyps, while in the second class there were neoplastic, adenomatous or carcinomatous structures. The images were obtained with the aid of a high-definition (HD) endoscope without magnification in combination with virtual chromo-endoscopy technology and conventional chromoscopy.[18] The method proposed by Kominami for the CAD of colorectal polyp histology was a real-time image recognition system combined with narrow-band imaging (NBI) magnifying colonoscopy. Images containing 188 lesions were obtained from 41 patients who underwent endoscopic submucosal resection and biopsy, and a descriptor that was invariant to image translation, scaling, and rotation, and partially invariant to illumination changes was used. The accuracy of evaluating concordance between diagnosis by the image recognition system and the histologic findings was 94.9%, with the sensitivity being 95.9% and the specificity 93.3%.[19] A similar approach was researched in a 2011 study which developed a recognition system for classifying NBI images of colorectal tumors into three types of structures of microvessels on the colorectal surface, according to the Hiroshima classification system.[20] The 907 images used were collected by physicians during NBI colonoscopy examinations with similar lighting conditions and image zooming, and the maximum recognition rate was 94.1%, while the recall rate specific for C3-type polyps was only 73.6%.[21]

In a study by Tajbakhsh *et al.* (2016), another hybrid technique was proposed for the detection of colon polyps.[22] A combination of both context- and shape-based approach was considered due to the fact that a shape-based approach alone may falsely identify other polyp-like structures such as fecal content, and that a context-based approach may not be able to extract the discriminative geometric information of the polyps. A powerful descriptor that is invariant to rotation and robust against light setting changes was used along with a two-stage classification scheme, which greatly improved the detection performance in colonoscopy videos. The CAD was evaluated both on a public polyp database containing 300 colonoscopy images with 300 polyp instances and on a personal database containing 19,400 frames and a total of 5,200 polyp instances and reported different results: at 0.1 false positives per frame, the sensitivity was 88.0% for the public database, while a sensitivity of only 48% was reported for the personal database. The difference was due to the fact that there were fewer images in the public database, with no images that had no polyps inside.[22]

MACHINE LEARNING ALGORITHMS

Recently, a 2016 study proposed a number of novel approaches in terms of automated detection of polyps from colonoscopy videos, suggesting that temporal information also plays a big role in detection methods. An offline method that evaluates both spatial and temporal features was implemented, combined with an online learning scheme that deals with limited training data. This novel integrated framework was evaluated on another public database, namely the Asu-Mayo Clinic Polyp Database and had a precision of 88.1%, while the offline method alone scored 78.5% on the same database. The combined method managed to detect fewer false positives, but had a longer processing time due to online method's need to update.[23]

A different approach was investigated in a study in which a machine learning algorithm was constructed to diagnose colorectal cancer by transferring features learned from two online databases that are non-medical. Then, an algorithm was used to detect images with polyps, and another was used for their classification. The images were taken under either white light (WL) or NBI endoscopy and were taken under random lighting conditions, zooming, and optical magnification. The results obtained with this method were compared with the ones provided by endoscopists, and although the precision for the classification of the polyps was similar (87.3% vs. 86.4% by human), the proposed method showed a higher recall rate (87.6% vs. 77.0%) and higher accuracy (85.9% vs. 74.3%). The method utilized the background of an image for polyp classification, being consonant with the NICE recommendation which states that comparing colors between the polyp and its surrounding area should be one of the criteria in polyp type classification.[24] Another study published in 2014 reported superior results compared to endoscopists when classifying polyps. A SVM was used to classify colorectal polyps by type and reached 96% accuracy, while human visual inspection was of only 85%, although the proposed method was only able to classify polyps that were either adenomatous or hyperplastic.[25]

LIMITATIONS OF THE PROPOSED METHODS

A commonly reported problem by the research community is the presence of artifacts such as air bubbles, reflections, or fecal content, which make the detection system to classify them as polyps, thus increasing the false positive rate and decreasing the overall performance of the system. The process of creating algorithms that will correctly deal with artifacts or removing them from the images imposes additional difficulties.[24,26,27] Another issue arises in the classification phase of the medical images. Yu *et al.* (2016) and Zhang *et al.* (2016) have noted the fact that there is a high inter-class similarity and intra-class variation regarding colon polyps, rendering it difficult for machine learning algorithms to correctly classify the polyps.[23,24] Finally, small datasets used to train algorithms were considered impediments in obtaining better results, while trial and error is still currently used to determine the best classification method for a certain dataset.[28]

CONCLUSION

Machine learning in medical imaging is a very promising and quickly growing field. By training a CAD system based on a certain set of images, the achievement of results with a high degree of confidence for the clinician is possible. Although in terms of accuracy and detection rates some methods have been close to 100% and performing better than humans, currently these methods have not been implemented in the daily clinical practice. Given the fact that this field has been on an ascending trend, it is likely that computer-aided diagnosis will have a big impact on medical imaging, possibly more than a reliable second observer.

CONFLICT OF INTEREST

Nothing to declare.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136:E359-E386.
2. Bloom DE, Cafiero ET, Jané-Llopis E, et al. The global economic burden of non-communicable diseases. Geneva: World Economic Forum, 2011. Available at: http://www.weforum.org/reports/global-economicburden-non-communicable-diseases
3. Wiegering A, Ackermann S, Riegel J, et al. Improved survival of patients with colon cancer detected by screening colonoscopy. *Int J Colorectal Dis*. 2016;31:1039-1045.
4. Regge D, Hassan C, Pickhardt PJ, et al. Impact of computer-aided detection on the cost-effectiveness of CT colonography. *Radiology*. 2009;250:488-497.
5. Dolgobrodov SD, Marshall R, Moore P, Bittern R, Steele RJC, Cuschieri A. e-Science and artificial neural networks in cancer management. *Concurrency Computat Pract Exper*. 2007;19:251-263.
6. Tjoa MP, Krishnan SM. Feature extraction for the analysis of colon status from the endoscopic images. *Biomed Eng Online*. 2003;8;2:9.
7. Selaru FM, Xu Y, Yin J, et al. Artificial neural networks distinguish among subtypes of neoplastic colorectal lesions. *Gastroenterology*. 2002;122:606-613.
8. Djemal K, Cocquerez JP, Precioso F. Visual feature extraction and description. In: Benois-Pineau J, Precioso F, Cord M, eds. Visual Indexing and Retrieval. 1st ed. New York: Springer, 2012; p. 5-20.
9. Igelnik B. Computational Modeling and Simulation of Intellect: Current State and Future Perspectives. 1st ed. Hershey: IGI Global, 2011.
10. Naguib RNG, Sherbet GV. Introduction to artificial neural networks and their use in cancer diagnosis, prognosis, and patient management. In: Naguib RNG, Sherbet GV, eds. Artificial Neural Networks in Cancer Diagnosis, Prognosis, and Patient Management. Boca Raton: CRC Press, 2001; p. 1-8.
11. da Silva IN, Hernane Spatti D, Andrade Flauzino R, Liboni LHB, dos Reis Alves SF. Introduction. In: Artificial Neural Networks: A Practical Course. Cham: Springer International Publishing, 2017; p. 3-19.
12. Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995;20:273-297.
13. Maroulis DE, Iakovidis DK, Karkanis SA, Karras DA. CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames. *Comput Methods Programs Biomed*. 2003;70:151-166.
14. Oliva JT, Lee HT, Spolaôr N, Coy CSR, Wu FC. Prototype system for feature extraction, classification and study of medical images. *Expert Syst Appl*. 2016;63:267-283.
15. Tjoa MP, Krishnan SM. Feature extraction for the analysis of colon status from the endoscopic images. *Biomed Eng Online*. 2003;2:9.
16. Biswas M, Bhattacharya A, Dey D. Classification of various colon diseases in Colonoscopy video using Cross-Wavelet features. 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2016; p. 2141-2145.
17. Ribeiro E, Uhl A, Häfner M. Colonic Polyp Classification with Convolutional Neural Networks. 2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS), Dublin, 2016; p. 253-258.
18. Uhl A, Wimmer G, Hafner M. Shape and size adapted local fractal dimension for the classification of polyps in HD colonoscopy. 2014 IEEE International Conference on Image Processing (ICIP), Paris, 2014; p. 2299-2303.
19. Kominami Y, Yoshida S, Tanaka S, et al. Computer-aided diagnosis of colorectal polyp histology by using a real-time image recognition system and narrow-band imaging magnifying colonoscopy. *Gastrointest Endosc*. 2016;83:643-649.
20. Tanaka S, Kaltenbach T, Chayama K, Soetikno R. High-magnification colonoscopy (with videos). *Gastrointest Endosc*. 2006;64:604-613.
21. Tamaki T, Yoshimuta J, Takeda T, et al. A System for Colorectal Tumor Classification in Magnifying Endoscopic NBI Images. In: Kimmel R, Klette R, Sugimoto A, eds. Computer Vision – ACCV 2010. ACCV 2010. Lecture Notes in Computer Science, vol 6493. Berlin/Heidelberg: Springer, 2011.

22. Tajbakhsh N, Gurudu SR, Liang J. Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information. *IEEE Transactions on Medical Imaging*. 2016;35:630-644.
23. Yu L, Chen H, Dou Q, Qin J, Heng PA. Integrating Online and Offline Three-Dimensional Deep Learning for Automated Polyp Detection in Colonoscopy Videos. *IEEE Journal of Biomedical and Health Informatics*. 2017;21:65-75.
24. Zhang R, Zheng Y, Wing Chuk Mak T, 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*. 2017;21:41-47.
25. Fu JJ, Yu YW, Lin HM, Chai JW, Chen CC. Feature extraction and pattern classification of colorectal polyps in colonoscopic imaging. *Comput Med Imaging Graph*. 2014;38:267-275.
26. Streba CT, Ionescu M, Vere CC, Rogoveanu I. Artificial Intelligence and Automatic Image Interpretation in Modern Medicine. In Wei DQ, Ma Y, Cho WCS, Xu Q, Zhou F, eds. Translational Bioinformatics and Its Application. 1st edition. Springer Netherlands, 2017; p. 371-407.
27. Ameling S, Wirth S, Paulus, D, Lacey, G, Vilarino F. Texture-Based Polyp Detection in Colonoscopy. In: Meinzer HP, Deserno TM, Handels H, Tolxdorff T, eds. Bildverarbeitung für die Medizin 2009: Algorithmen – Systeme – Anwendungen. Berlin/Heidelberg: Springer, 2009; p. 346-350.
28. Kim J, Kim BS, Savarese S. Comparing Image Classification Methods: K-nearest-neighbor and Support-vector-machines. In: Proceedings of the 6th WSEAS International Conference on Computer Engineering and Applications, and Proceedings of the 2012 American Conference on Applied Mathematics Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2012; p. 133-138.