Histologic classification of colonic polyps based on fractal dimension analysis: comparison of results using support vector machine and logistic regression

Abdelrahim N. Esgiar[1], Mussa Mabrok[2], Abdullah H. Abdullah[1] and Ahmad Almhdie[3]*

[1]Department of Electrical and Electronic Engineering, Faculty of Engineering, Sirte University, Sirte, Libya
[2]Centre for Telecommunication Research and Innovation (CeTRI), Faculty of Electronics and Computer Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia
[3]Department of Electrical and Electronic Engineering, Faculty of Technical and Engineering Sciences, University of Sebha, Libya
[*]Corresponding author: Email:anasser6474@yahoo.co.uk

Abstract

The aim of this study was to evaluate fractal analysis as a tool for differentiating between normal tissue and adenomatous polyp lesions. Images of colon samples from 140 patients were analyzed. There were 70 subjects in each of the normal and polyp groups. Two texture features based on fractal analysis were studied: fractal dimension (FD) and lacunarity (Lac), extracted using the overlapping box-counting method. The proposed classification models based on fractal analysis of normal colon and abnormal polyp images were performed using two classification methods: the support vector machine (SVM) and the logistic regression (LR). Several widely-recalled statistical metrics (accuracy, sensitivity, specificity and precision) were used to evaluate the global model performance. To avoid any overfitting problems, all models were evaluated using a 10-fold cross-validation. The SVM method showed better performance in detecting normal colon images than the LR method. As a result, the SVM method provided results with higher accuracy (ACC) and specificity than the LR method (ACCSVM=0.90 vs. ACCLR=0.75). These results give confidence for developing a practical automated analysis technique for detecting colon polyps.

Keywords: *Histologic classification, colonic polyps, fractal dimension analysis, support vector machine, logistic regression*

1. Introduction

Colorectal cancer is one of the most dangerous cancer types in both men and women in the USA [1]. Adenomatous polyps are precursor lesions with a high risk of progression to CRC [2]; early detection and treatment can provide a major opportunity to save lives. In clinical routine, manual evaluation of histological slides is still indispensable, which is time consuming and requires the visual interpretation of complex images. This interpretation is usually based on subjective assessment techniques and can lead to significant interobserver variation in grading. A great interest has been shown in developing image processing- based methods for quantitative and high-throughput analysis of tumor tissues [3]. The automation of polyp histology analysis could provide a valuable objective assessment, as well as contribute to a reduction in the diagnostic sources of error associated with subjective visual analysis. The automatic classification of colon images has been the interest of limited number of studies [4]–[8], exploiting the textural changes in normal and malignant colon biopsy images. Fractal geometric analysis has been introduced for identifying the diagnosis the different diagnostic categories of colorectal polyps [5], [9]. None of these studied examined lacunarities.

Research in the classification of microscopic images of colonic mucosa has shown that texture features derived from a grey-level co-occurrence matrix (GLCM) are useful when applied to medical image analysis [10], [11]. A wide variety of texture-based techniques have been proposed for the analysis of colonic cancer images, including Haralick's features (entropy, correlation, inverse difference moment and angular second moment) [11]–[14], as well as fractal dimension and lacunarity [9], [15] -[18]. However, only few studies analyzed polyp images.

Table 1 presents a summary of the previous studies on the automated analysis of normal colon and colon polyp tissues using fractal analysis techniques. It can be seen that only the study in [16] used both fractal dimension and lacunarity for colon cancer images, but not for polyp images. Furthermore, the image samples were captured by computed tomography (CT) scans, and not through digital microscopy, as is the case in this study.

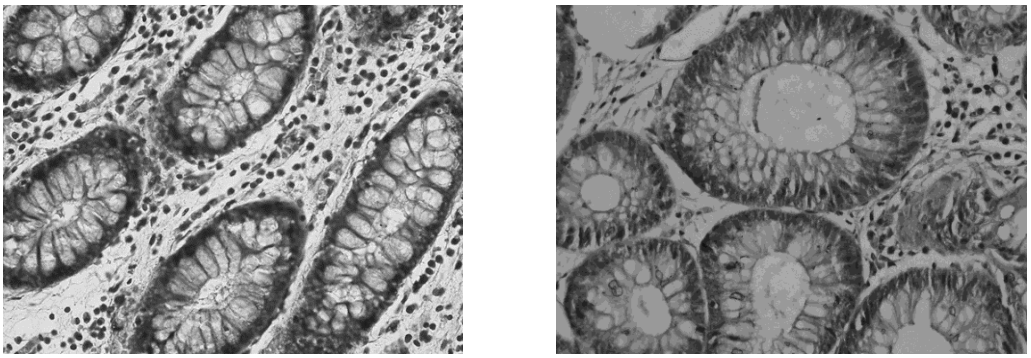
Compared to our previous study [15], we used for the current study a larger dataset, including normal subjects and adenomatous polyp cases, in collaboration with the University of the Philippines. In addition, the classification process was assisted by both fractal dimension and lacunarity features, using two fractal analysis techniques: the non-overlapping box-counting and the sliding-box scan approaches. The aim of this study was therefore to evaluate two well-known classification techniques, the support vector machine (SVM) and the logistic regression (LR), using a completely new clinical dataset. As of our knowledge, we are the first to provide such a classification comparison for the quantification of clinical histological differences between normal colon tissue and adenomatous colon polyps.

2. Patients and methods

2.1 Patients

We used for the analysis of this study 140 colonic images, acquired by digital microscopy. These images consisted of two equally sized classes, namely 'normal' and 'adenomatous polyps'. These images gave a general representation of the infiltrative edge of the adenomatous polyps and of the thickness of the normal mucosa. A zone of interest of 300×400 pixels was considered for each monochromatic image. Figures 1 and 2 represent samples from normal and adenomatous-polyp's patients.

The images used in this study were derived from slides of cases randomly selected from surgical pathology files of the Philippine General Hospital (PGH), Manila, Philippines, in 2007 and 2008. These were previously diagnosed as colonic adenocarcinoma, adenomatous polyps from the colon, as well as tumour-free colonic resection planes to serve as controls. The slides were routinely processed using a Sakura tissue processor and cut at 8μm using a standard microtome. All were stained with haematoxylin and eosin [19].



A B

Figure 1: Example of a normal sample (A) and abnormal colon polyp sample (B).

2.2 Fractal analysis

Fractal analysis can provide quantitative parameters as quantifiers of complexity for the measurement of texture patterns. In this paper, we analysed fractal dimension and lacunarity of the colonic microscopic image samples. The box-counting technique was used to determine the self-similarity dimension [20]. In a self-similar structure, there is a relationship between the box-size scale factor, ε, and the number of boxes N(ε), into which the structure can be divided. This relationship is the fractal dimension, D_f.

$$D_f = \ln(N(\varepsilon))/\ln(1/\varepsilon) \quad (1)$$

Since fractals cannot be completely characterised by their fractal dimension, lacunarity can be used

as a complementary measure to compensate for the lack of quantification of texture variations (in homogeneities), or for deviation of a geometric structure from its translational invariance. In other words, lacunarity provides a tool to reveal 'gap' textures within an otherwise fractal distribution [21]. There is no formal definition of lacunarity and, indeed, there has been some controversy over how it should be measured [22]. Its concept and formulation were derived from differentiating two objects defined by the same fractal dimension, but showing various visual textural patterns [23]. Therefore, lacunarity is considered as a measure of perceived gaps, or holes, in the geometric structure of the image.

The parameter Lac_ε describes lacunarity as a function of the box (window) of size ε from which the data are extracted, according to equation (2) [23], [24]:

$$Lac_\varepsilon=(CV_\varepsilon)^2=\left(\sigma_\varepsilon/\mu_\varepsilon\right)^2 \qquad (2)$$

CV is the coefficient of variation for pixel distribution, σ is the standard deviation, μ is the mean of the data and ε is the scale factor applied to an object 'image'. In this study we applied two box-counting techniques to extract the relevant features, from the ratio of increasing detail to the increasing scale factor (ε). These two techniques are described in the two following sections.

Table 1: Summary of published studies using fractal analysis of colon polyp images acquired by digital microscopy

Materials and Methods	[15] 2002	[16] 2009	[18] 2015	[17] 2020
Colon	Yes	Yes	Yes	No
Polyp	No	No	No	No
Image analysis techniques	Texture features from a grey-level co-occurrence matrix and fractal analysis	CT	Fractal analysis	Fractal analysis
Features	Correlation, Entropy and FD	Fractal analysis	FD	FD
Numerical Results	Correlation (mean± SD): Normal=0.0326± 0.0075, Cancer=0.0510± 0.0097. Entropy (mean ±SD): Normal=1.01 ±0.15, Cancer=1.26± 0.11, FD (mean±SD): Normal=1.757±0.023, Cancerous=1.785±0.037, All p<0.0001.	Fractal dimension (FD), Abundance and Lacunarity (Lac)	FD (mean): no SD Well-differentiated Cancer=1.431, Moderately differentiated=1.516 Weakly differentiated=1.669 Undifferentiated =1.741, p<0.001.	FD (mean± SD): Normal =1.712±0.030 Cancer = 1.741±0.016, p<0.001

SD indicates standard deviation, p is the p-value, and FD denotes the Fractal Dimension.

2.2.1 Fractal dimension using non-overlapping box-counting approach

In this method, a non-overlapping regular square grid with scale factor ε traversed the image to measure the box-counting dimension. The image was superimposed on a regular grid with scale factor ε, and the number of grid boxes (windows) was counted. This gave a number, N(ε), with a value dependent on ε. Then we progressively changed ε to smaller sizes and counted the corresponding N(ε). Next, we plotted the distribution log(N(ε))/log(1/ε) for each image, fitted a straight (regression) line to the points, and measured its slope (equation 1), to give the box-counting dimension. Each part of the image was sampled only once for each box size, and repeated until the whole surface area of the image had been traversed [38], [42].

2.2.2 Fractal dimension using sliding-box scan approach

The following settings were developed according to the recommendations of the FracLac user manual [25]. The size of the series pertaining to the sliding-box technique was set to decrease linearly from a maximum box size of 41% of the entire image or region of interest (ROI) size to a minimum size of 3×3 pixels. The square box of size ε was slid over the entire image so that it

overlapped itself with each movement. The sliding-box scan then counted the number of pixels inside the box and the number of boxes, N(ε), slid the boxes horizontally by a fixed number of pixels (x), then recounted the pixels that fell on the box, and the number of boxes, N(ε). At the end of each row, the box was slid down by a fixed number of pixels (y), and the row was scanned again in the same way until the entire image had been scanned. Then we progressively changed ε to smaller sizes and counted pixels that fell on the box and the corresponding N(ε).This process was repeated for each box size until the entire area of the image had been scanned using each ε. Next, we plotted log(N(ε))/log(1/ε), fitted a straight (regression) line to the points, and measured its slope (equation 1). This technique differs from a regular box-counting scan in which all boxes are of a fixed size and are laid on a non-overlapping grid [23], [25]. The non-overlapping grid can be seen as a special case of the sliding-box algorithm, with horizontal and vertical increments equal to the scale factor, ε.

2.2.3 Lacunarity using non-overlapping and sliding-box scan approaches

For both non-overlapping and sliding-box counting, lacunarity was determined from the probability distribution for pixel 'mass distribution'. Lacunarity at a particular ε was labelled as $Lac(\varepsilon)$, and calculated as the variation in pixel density at different box sizes, using CV for pixel distribution as in equation (2). Lacunarity varied with the size of the sampling unit. Thus, in order to arrive at a single number, the values for $Lac(\varepsilon)$ were summarised as the mean, $\overline{Lac}$, over all scale factors, ε [23], [25].

2.3 Statistical analysis

Two well-known classification methods were evaluated; the support vector machine (SVM) [26], [27] and the logistic regression (LR) [28], [29]. Several statistical metrics were used to evaluate the global model performance. The first metric is the accuracy which is defined as the ratio of the number of correct predictions to the total number of predictions. It is expressed by:

$$\mathrm{ACC} = \mathrm{TP} + \mathrm{TN}/\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN} \quad (3)$$

where TP, FP, TN, and FN mean True Positive, False Positive, True Negative, and False Negative, respectively. The second metric is the precision, or the positive predictive value, which is the percentage of correctly classified cases to the number of actual cases.

$$\mathrm{Precision} = \mathrm{TP}/\mathrm{TP} + \mathrm{FP} \quad (4)$$

The third metric is the recall, or the sensitivity, which provides the ability of a model to classify all cases and refers to the true positive rate.

$$\mathrm{Recall} = \mathrm{TP}/\mathrm{TP} + \mathrm{FN} \quad (5)$$

The fourth metric is the specificity, which provides the ability of a model to classify all controls and refers to the true negative rate.

$$\mathrm{Specificity} = \mathrm{TN}/\mathrm{TN} + \mathrm{FP} \quad (6)$$

To avoid overfitting problems, all models were evaluated using a 10-fold Leave-One-Out cross-validation [27].

3. Results and discussion

The results show that the classification evaluations based on SVM outperformed those based on LR. Both classification methods provided a good performance in terms of accuracy. However, the SVM provided more accurate classification (ACC= 0.90) than that obtained by the LR (ACC= 0.75). Comparative results of the classification performance of both SVM and LR methods using the other previously mentioned metrics are reported in Table 2. classification and discrimination of normal and cancerous achieves significantly higher correct classification rates.

Table 2: Summary of published studies using fractal analysis of colon polyp images

Method	Accuracy	Precision	Recall	Specificity
SVM	0.90	0.95	0.84	0.96
LR	0.75	0.72	0.81	0.78

Conclusions

This study has demonstrated that fractal analysis of simple digital microscopy images can significantly help in differentiating between normal tissues and adenomatous polyps. This study therefore provides evidence for the value of fractal dimension and lacunarity features for effectively differentiating between these two clinical groups. This study demonstrates that fractal dimension and lacunarity features extracted from images can result in highly significant correct colonic mucosa. We conclude that the use of accurate and robust texture features can significantly contribute to a more automated reliable diagnostic.
a robust technique for PM of pump problems and can determine optimum solutions with a wide range of variations in GA parameters. The application of KBS, the formulation of the evaluation function and the design of the GA operators can help in optimizing the GA technique to solve genuine large-scale PM of a pump.

Acknowledgment

We are very grateful to Dr. Jose Maria C. Avila who provided the research team with the pathology and control images used in this study, with an ethical permission granted by Philippine General Hospital (PGH), Manila, Philippines in 2007. Unfortunately, Dr Avila has passed away in April 2021, otherwise, he would have been one of our co-authors.

References

[1] R. L. Siegel et al., « Colorectal cancer statistics, 2020 », CA Cancer J Clin, vol. 70, no. 3, p. 145-164, mai 2020, doi: 10.3322/caac.21601.

[2] Y. Tang, A. D. Polydorides, S. Anandasabapathy, et R. R. Richards-Kortum, « Quantitative analysis of in vivo high-resolution microendoscopic images for the detection of neoplastic colorectal polyps », JBO, vol. 23, no. 11, p. 116003, nov. 2018, doi: 10.1117/1.JBO.23.11.116003.

[3] K. N. Manjunath, P. C. Siddalingaswamy, et G. K. Prabhu, « Domain-Based Analysis of Colon Polyp in CT Colonography Using Image-Processing Techniques », Asian Pac J Cancer Prev, vol. 20, no. 2, p. 629-637, févr. 2019, doi: 10.31557/APJCP.2019.20.2.629.

[4] M. Bibbo et al., « Karyometric marker features in normal-appearing glands adjacent to human colonic adenocarcinoma », Cancer Res, vol. 50, no. 1, p. 147-151, janv. 1990.

[5] P. W. Hamilton, P. H. Bartels, D. Thompson, N. H. Anderson, R. Montironi, et J. M. Sloan, « Automated location of dysplastic fields in colorectal histology using image texture analysis », J Pathol, vol. 182, no. 1, p. 68-75, mai 1997, doi: 10.1002/(SICI)1096-9896(199705)182:1<68::AID-PATH811>3.0.CO;2-N.

[6] P. W. Hamilton, D. C. Allen, P. C. Watt, C. C. Patterson, et J. D. Biggart, « Classification of normal colorectal mucosa and adenocarcinoma by morphometry », Histopathology, vol. 11, no. 9, p. 901-911, sept. 1987, doi: 10.1111/j.1365-2559.1987.tb01897.x.

[7] M. F. McNitt-Gray, H. K. Huang, et J. W. Sayre, « Feature selection in the pattern classification problem of digital chest radiograph segmentation », IEEE Trans Med Imaging, vol. 14, no. 3, p. 537-547, 1995, doi: 10.1109/42.414619.

[8] D. Thompson, P. H. Bartels, H. G. Bartels, P. W. Hamilton, et J. M. Sloan, « Knowledge-guided segmentation of colorectal histopathologic imagery », Anal Quant Cytol Histol, vol. 15, no. 4, p. 236-246, août 1993.

[9] S. S. Cross, J. P. Bury, P. B. Silcocks, T. J. Stephenson, et D. W. Cotton, « Fractal geometric analysis of colorectal polyps », J Pathol, vol. 172, no. 4, p. 317-323, avr. 1994, doi: 10.1002/path.1711720406.

[10] R. Cicchi et al., « Scoring of collagen organization in healthy and diseased human dermis by multiphoton microscopy », J Biophotonics, vol. 3, no. 1-2, p. 34-43, janv. 2010, doi: 10.1002/jbio.200910062.

[11] A. N. Esgiar, R. N. Naguib, B. S. Sharif, M. K. Bennett, et A. Murray, « Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa », IEEE Trans Inf Technol Biomed, vol. 2, no. 3, p. 197-203, sept. 1998, doi: 10.1109/4233.735785.

[12] *L. Jiao, Q. Chen, S. Li, et Y. Xu, « Colon Cancer Detection Using Whole Slide Histopathological Images », in World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China, Berlin, Heidelberg, 2013, p. 1283-1286. doi: 10.1007/978-3-642-29305-4_336.*

[13] *K. Masood et N. Rajpoot, « Texture based classification of hyperspectral colon biopsy samples using CLBP », in 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, juin 2009, p. 1011-1014. doi: 10.1109/ISBI.2009.5193226.*

[14] *J. K. Shuttleworth, A. G. Todman, R. N. G. Naguib, B. M. Newman, et M. K. Bennett, « Colour texture analysis using co-occurrence matrices for classification of colon cancer images », in IEEE CCECE2002. Canadian Conference on Electrical and Computer Engineering. Conference Proceedings (Cat. No.02CH37373), mai 2002, vol. 2, p. 1134-1139 vol.2. doi: 10.1109/CCECE.2002.1013107.*

[15] *A. N. Esgiar, R. N. G. Naguib, B. S. Sharif, M. K. Bennett, et A. Murray, « Fractal analysis in the detection of colonic cancer images », IEEE Trans Inf Technol Biomed, vol. 6, no. 1, p. 54-58, mars 2002, doi: 10.1109/4233.992163.*

[16] *V. Goh, B. Sanghera, D. M. Wellsted, J. Sundin, et S. Halligan, « Assessment of the spatial pattern of colorectal tumour perfusion estimated at perfusion CT using two-dimensional fractal analysis », Eur Radiol, vol. 19, no. 6, p. 1358-1365, juin 2009, doi: 10.1007/s00330-009-1304-y.*

[17] *J. Qin, L. Puckett, et X. Qian, « Image Based Fractal Analysis for Detection of Cancer Cells », déc. 2020, p. 1482-1486. doi: 10.1109/BIBM49941.2020.9313176.*

[18] *L. Streba et al., « A pilot study on the role of fractal analysis in the microscopic evaluation of colorectal cancers », Rom J Morphol Embryol, vol. 56, no. 1, p. 191-196, 2015.*

[19] *L. A. Gan Lim, R. Naguib, E. P. Dadios, et J. M. C. Avila, « Image classification of microscopic colonic images using textural properties and KSOM », International Journal of Biomedical Engineering and Technology, vol. 3, no. 3-4, p. 308-318, 2010, doi: 10.1504/IJBET.2010.032698.*

[20] *K. A. Marghani, S. S. Dlay, B. S. Sharif, et A. J. Sims, « Morphological and texture features for cancer tissues microscopic images », in Medical Imaging 2003: Image Processing, mai 2003, vol. 5032, p. 1757-1764. doi: 10.1117/12.481322.*

[21] *M. N. B. Filho et F. J. A. Sobreira, « Accuracy of Lacunarity Algorithms in Texture Classification of High Spatial Resolution Images from Urban Areas », International archives of photogrammetry, remote sensing and spatial information sciences, p. 417-422, 2008.*

[22] *J. Theiler, « Lacunarity in a best estimator of fractal dimension », Physics Letters A, vol. 133, no. 4, p. 195-200, nov. 1988, doi: 10.1016/0375-9601(88)91016-X.*

[23] *A. Karperien, H. Ahammer, et H. F. Jelinek, « Quantitating the subtleties of microglial morphology with fractal analysis », Front Cell Neurosci, vol. 7, p. 3, 2013, doi: 10.3389/fncel.2013.00003.*

[24] *null Allain et null Cloitre, « Characterizing the lacunarity of random and deterministic fractal sets », Phys Rev A, vol. 44, no. 6, p. 3552-3558, sept. 1991, doi: 10.1103/physreva.44.3552.*

[25] A. Karperien, « FracLac for imageJ – FracLac advanced user's manual, Charles Sturt University Australia », 2004. https://imagej.nih.gov/ij/plugins/fraclac/fraclac-manual.pdf

[26] A. A. Miranda, T. A. Chowdhury, O. Ghita, et P. Whelan, « Shape Classification of Colorectal Polyps at CT Colonography using Support Vector Machines », undefined, 2006, Consulté le: déc. 06, 2021. [En ligne]. Disponible sur: https://www.semanticscholar.org/paper/Shape-Classification-of-Colorectal-Polyps-at-CT-Miranda-Chowdhury/96b0ed70d5238ee8017e3329ccc2e1cca5890d73

[27] J. Weston, « Leave-one-out support vector machines », in Proceedings of the 16th international joint conference on Artificial intelligence - Volume 2, San Francisco, CA, USA, juill. 1999, p. 727-731.

[28] M. Kumar et S. K. Rath, « Chapter 15 - Feature Selection and Classification of Microarray Data Using Machine Learning Techniques », in Emerging Trends in Applications and Infrastructures for Computational Biology, Bioinformatics, and Systems Biology, Q. N. Tran et H. R. Arabnia, Éd. Boston: Morgan Kaufmann, 2016, p. 213-242. doi: 10.1016/B978-0-12-804203-8.00015-8.

[29] V. F. van Ravesteijn et al., « Computer-Aided Detection of Polyps in CT Colonography Using Logistic Regression », IEEE Transactions on Medical Imaging, vol. 29, no. 1, p. 120-131, janv. 2010, doi: 10.1109/TMI.2009.2028576.