

Upgraded Spatial Gray Level Dependence Matrices for Textural Analysis in Colon Cancer Tissues

B. Saroja^{1*}, A. Selwin Mich Priyadharson²

¹Research Scholar, Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai.

²Associate Professor, Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai.

*Corresponding author E-mail: boda.saroja@gmail.com

Abstract

Colon or Bowel or Colorectal Cancer (CRC) is commonly determined by diagnosing a sample of colon tissue and further analysed by medical imaging. The colon tissue classification method count on specific changes between texture features extracted from benign and malignant regions. The variations in the image acquisition methods effects the colon tissue analysis. In this paper, an Upgraded Spatial Gray Level Dependence Matrices (U-SGLDM) is emphasized to extract textural features. The licensed image set of all applicable types of tissues within colon cancer are used for experimentation. Several texture feature sets are extracted to show the significant differences among the eight colon cancer biopsy images in the image data set. The fractal dimension-Hurst Coefficient is added to U-SGLDM for long range assessment. The Prominence of the analysis evoked in the representation of histopathological image structure over longer periods.

Keywords: Colorectal cancer, textural features, U-SGLDM, fractal features, histopathological images.

1. Introduction

Medical images generally have non homogeneous structures without regularity. The CRC tissue is one such histopathological image which contain several distinct tissue types. To study and measure these distinguished tissues, various statistical and structural approaches are used. However; still several approaches are being in progress. Feature analysis is a vital concern in CRC detection and classification [2, 11]. It comprises of various mathematical practices in processing the CRC image to quantify different gray levels within the image. Basically CRC is analyzed on basis of primary statistical features and derived or secondary statistical feature sets. The primary statistics are derived from the set of gray level averages from the pixel values. This limits the number of feature for discrimination. Many literatures show several statistical approaches in analyzing CRC viz., FOM, SGLDM, GLDM [4], RUNL [5], FPS [6], FDTA, SRDM [7], and HM [8]. The non-deterministic properties of texture in CRC, govern the distribution and associations between the gray levels of the image. Literature further show that some information is lost which are significant in calculating the features; In SGLDM method [7, 8 &9] area of interest (AOI) is selected manually, which is interior or exterior to the mass for calculating textural features. This is due to the operators which are used in the selection of AOI in the image. So, the textural analysis should be perfect-invariant. An Upgraded SGLDM method [1] is suggested in this paper for whole image segmentation of the CRC. This paper is organized as I. Introduction and Literature review

2. Proposed methodology

The approach used in this paper for textural feature analysis is illustrated using figure 1.

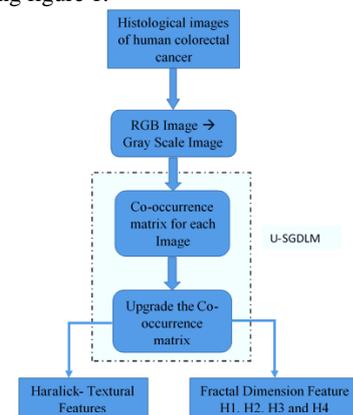


Fig. 1. Approach for Feature Extraction.

Upgraded spatial gray level dependence matrices (U-SGLDM)

The U-SGLDM method is based on whole segmentation of the histopathological image. The U-SGLDM is a second order probabilistic statistical method, which calculates a set of features based on co-occurrence of spatial distribution of gray scale levels in the AOI. The significant idea of the U-SGLDM is as follows: Step 1: Calculate the co-occurrence matrix of the image. The co-occurrence matrix is defined as

$$P_{\theta,d}(i,j) = \sum_j \sum_i \begin{cases} 1, & \text{if } f(x,y) = i \text{ and } f(x+\theta, y+d) = j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

An element at location (i, j) of the image matrix signifies the probability that two different resolution cells which are separated at a distance d in a specific directional angle θ from each other. This will have gray level values i and j respectively. The relative frequencies of co-occurrence P θ , d (i, j) describe how often two pixels with gray levels (i, j) appear in the image 'f' of window size (M x N). The frequencies are represented at $\theta = \{00, 450, 900, \dots, 3600\}$. For different θ and d values, different SGLD matrices results.

$$P_{0^{\circ},d}(i,j) = \left| \left\{ [(k,l), (m,n)] \in D : k-m=0 \right. \right. \\ \left. \left. | |l-n|=d, f(k,l) = i, f(m,n) = j \right\} \right| \quad (2)$$

$$P_{45^{\circ},d}(i,j) = \left| \left\{ [(k,l), (m,n)] \in D : k-m=d \right. \right. \\ \left. \left. | (l-n) = -d, \text{OR } ((k-m) = -d, (l-n) = d) \right. \right. \\ \left. \left. | f(k,l) = i, f(m,n) = j \right\} \right| \quad (3)$$

$$P_{90^{\circ},d}(i,j) = \left| \left\{ [(k,l), (m,n)] \in D : k-m=d \right. \right. \\ \left. \left. | |l-n|=0, f(k,l) = i, f(m,n) = j \right\} \right| \quad (4)$$

and so on, where $|\{\dots\}|$ refers to set cardinality and $D = (M \times N) \times (M \times N)$.

Step 2: Eliminate the rows and columns of the co-occurrence matrix corresponding to all gray level values comprised between number of distinct gray levels and the maximum gray level value present in image AOI.

Step 3: Obtained is the upgraded co-occurrence matrix that corresponds to the whole image. From the upgraded co-occurrence matrix, eight textural features, are calculated [Haralick et al] for colon biopsy as follows, 1. Energy: Energy is the measure of uniformity in an image by the extent of pixel pair recurrences. The energy value will be high for similar pixel values and defined in equation 4.

From the upgraded co-occurrence matrix, eight textural features, are calculated [Haralick et al] for colon biopsy as follows,

1. Energy: Energy is the measure of uniformity in an image by the extent of pixel pair recurrences. The energy value will be high for similar pixel values and defined in equation 4.

$$Energy = \sum_{i,j} [P_{\theta,d}(i,j)]^2 \quad (5)$$

2. Contrast: Intensity measure of pixel of the image is defined as contrast given by Equation (6).

$$Contrast = \sum_{i,j} (i-j)^2 P(i,j) \quad (6)$$

3. Correlation: The correlation feature defined in equation (7) is a measure of gray tone linear dependencies in the image.

$$Correlartion = \frac{\sum_{i,j} (ij)P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (7)$$

Where μ_x , μ_y and σ_x , σ_y are mean and variance of the co-occurrence matrix, P_x and P_y . P_x is the marginal probability matrix obtained from $P_{\theta,d}(i,j)$ by summing the rows and P_y is the marginal probability matrix obtained from $P_{\theta,d}(i,j)$ by summing the columns.

4. Sum of squares: The sum of squares also known as variance, defined in equation 8. Variance increases when the gray level values differ from their mean.

$$variation = \sum_{i,j} (i-\mu)^2 P(i,j) \quad (8)$$

5. Inverse Difference Moment (IDM): This statistic is also called as Inverse Homogeneity is a measure of image homogeneity. It is more sensitive to the presence of near diagonal elements. It has maximum value when all elements in the image are same. The

homogeneity is inversely related to contrast feature increases at constant energy.

$$IDM = \sum_{i,j \& i \neq j} \frac{1}{1 + [i-j]^2} P(i,j) \quad (9)$$

6. Sum Average: This is a secondary textural feature given by equation 10.

$$Sum\ Average = \sum_i i \cdot P_{x+y}(i) \quad (10)$$

Where,

$$P_{x+y}(l) = \sum_{i,j} P(i,j), \quad i+j = l = 2,3, \dots, G \quad (10.a)$$

and G is the number of distinct gray scale levels.

7. Entropy: The complexity or disorder of an image as the statistic measures is known as entropy defined in equation (11). The entropy is high for textural non uniformity. Complex textures tend to have high entropy.

$$Entropy = - \sum_{i,j} P(i,j) \log_2 [P(i,j)] \quad (11)$$

8. Directional Moment: The angle measure in image for alignment is known as Directional moment (DM) and it is defined as in equation 12.

$$DM = \sum_{i,j} P(i,j) |i-j| \quad (12)$$

Fractal Dimension Feature -Hurst Coefficient:

The Hurst coefficient [17] is the index of independence which is directly related to fractal dimension D.

$$H = 2 - D, \quad \forall 1 < D < 2 \quad (13)$$

The fractal dimension, D is expressed using scaling rules as in equation 14.

$$N \propto \epsilon^D \quad (14)$$

Where N is the number of lines, ϵ stands for the scaling factor.

The Hurst exponent, H is defined in terms of the asymptotic behavior of fractal dimension as a function of time.

$$E\left(\frac{\bar{G}}{\sigma^2}\right) = C_G^H, \text{ as } G \rightarrow \text{maximum gray level} \quad (15)$$

\bar{G} is the range of the pixel values, and σ^2 is their standard deviation, E is the expected value, G is number of distinct pixel values and C is a constant.

The values of the Hurst exponent vary between 0 and 1 as shown in table 1, with higher values indicating a smoother trend, less volatility, and less roughness. The Hurst exponent represents structure over asymptotically longer periods. The generalized form of the Hurst coefficient or exponent is replaced by 'q', H (q) as discussed by R. M Bryce et al.

Table 1. Range of Hurst Exponent

S.No	H- Range	Indication
1.	0 - 0.5	Long-term changing among high and low values in neighboring pairs.
2.	0.5 - 1	Long-term positive autocorrelation.
3.	0.5	Completely uncorrelated.

3. Data set description image set

The image data is accessed via the following DOI: 10.5281/zenodo.53169, which is a collection of textures in histological images of human colorectal cancer [3]. The images contain more

than one tissue type which are categorized into eight types of colon tissues: Tumor epithelium, Simple stroma, Complex stroma, Immune cells, Debris, Normal mucosal glands, Adipose tissue, no tissue.

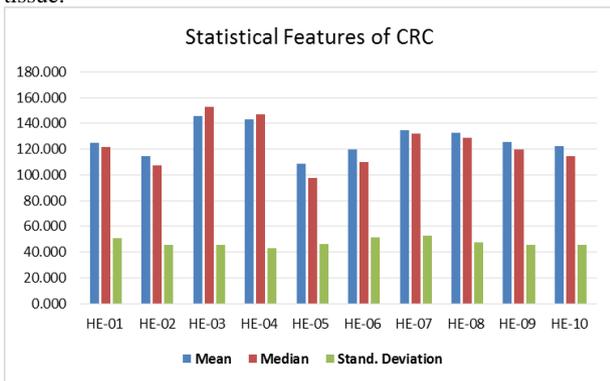


Figure 11. Statistical Feature set for CRC biopsy.

All images are in RGB image format, 0.495 μm per pixel, digitized with an Aperio Scan Scope (Aperio/Leica biosystems). Histological samples are fully anonymized images of formalin-fixed paraffin-embedded human colorectal adenocarcinomas (primary tumors) from pathology archive; Institute of Pathology, University Medical Center Mannheim, Heidelberg University, Mannheim, Germany.

4. Results and discussion

Types of Colon Tissue	Tumor epithelium	Simple stroma	Complex stroma	Immune cells	Debris	Normal mucosal glands	Adipose tissue	No Background
Colon histological Image								
Binary Image								

The images used for experimentation are primarily converted to gray scale images and then to binary images. The variety of histological images used in the experimentation are visualized below

To describe the texture of histological images we consider three distinct sets of features, which are examined in this section.

1. Statistical Feature set.
2. U- SGDLM Feature set.
3. Fractal Dimension Feature set. Among the image data set, ten larger images from the tissue region are considered as an application set and were used to test the different combinations of texture features, figure 10.

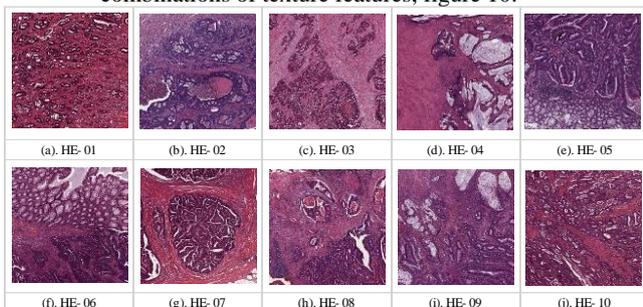


Figure 10. Application set of histology images used for the extracting textural features.

Statistical feature set

The primary statistical features extracted from CRC biopsy are tabulated in table 2; mean, median, standard deviation, skewness and kurtosis.

Table 2. Complete Statistical feature Set

Feature set	Mean	Median	Standard deviation	Skewness	Kurtosis
CRC BIOPSY					
HE-01	125.235	121.867	50.925	0.454	2.501
HE-02	114.431	107.666	45.573	0.724	3.166
HE-03	145.722	152.66	45.558	-0.209	2.419
HE-04	143.319	147.144	42.865	-0.056	2.695
HE-05	108.502	97.798	45.962	0.988	3.476
HE-06	119.542	109.7	51.428	0.629	2.506
HE-07	134.517	132.104	52.427	0.351	2.405
HE-08	132.546	128.983	47.808	0.414	2.598
HE-09	125.821	119.918	45.803	0.513	2.625
HE-10	122.474	114.456	45.33	0.849	3.34

The graphical representation for the mean, median and Standard deviation features are shown in figure 11. Analyzing the features, we can observe much more closeness in statistical values of the features for all CRC biopsies and it is difficult to catalogue into colon tissue type.

Upgraded-SGLDM feature set

The co-occurrence matrices constructed is upgraded using the proposed methodology. The secondary statistical features, known as U-SGDLM feature set is derived and are tabulated in table 3. For fine analysis of extracted feature set, graphical analysis followed.

Table 3. Complete U-SGDLM feature Set

Feature set	Correlation	Sum of squares: variance	Inverse difference moment	Sum average	Entropy	Contrast	Energy
CRC BIOPSY							
HE-01	0.542	34.62	0.042	1.289	10.13	0.542	34.6
HE-02	0.515	10.48	0.039	0.327	10.02	0.515	10.4
HE-03	0.485	10.26	0.035	0.78	10.11	0.485	10.2
HE-04	0.506	3.192	0.055	0.189	9.794	0.506	3.19
HE-05	0.507	9.438	0.041	0.247	9.959	0.507	9.43
HE-06	0.543	5.512	0.04	0.321	10.05	0.543	5.51
HE-07	0.516	27.75	0.041	1.195	10.12	0.516	27.7
HE-08	0.506	12.84	0.042	0.537	10.09	0.506	12.8
HE-09	0.511	11.16	0.036	0.408	10.12	0.511	11.1
HE-10	0.493	30.72	0.041	1.096	9.968	0.493	30.7

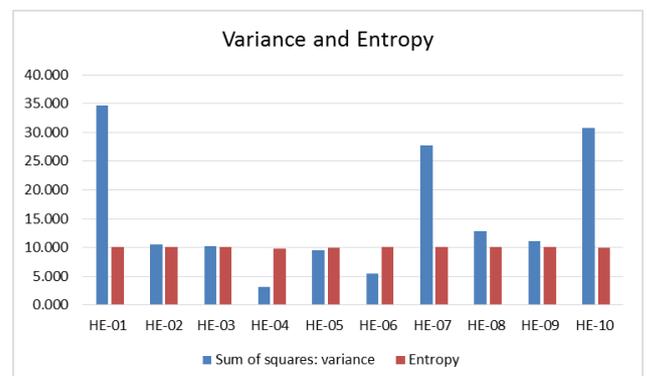


Figure 12. Variance and Entropy features from U-SGDLM feature set. Figure 12. shows the variance and entropy features extracted from U-SGDLM method. From the figure, HE-01, HE-07 and HE-10 show high variance and entropy features. Figure 13. Shows the correlation, inverse homogeneity & sum average features and the images HE-01, HE-07 and HE-10 show degree of variation relative to other images. The images HE-01, HE-07 and HE-10 show relative high variations and these images are categorized as tumor

epithelium tissue. These images contain high tumor content. However, for the said categorized images the contrast and energy features, figure 14. are much similar to other images.

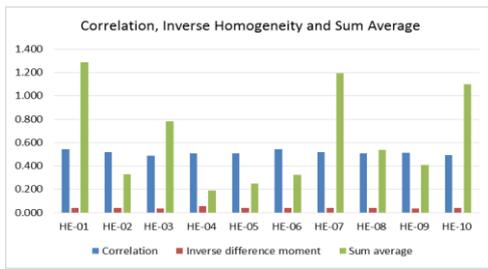


Figure 13. Correlation, inverse homogeneity & sum average features from U-SGDLM feature set.

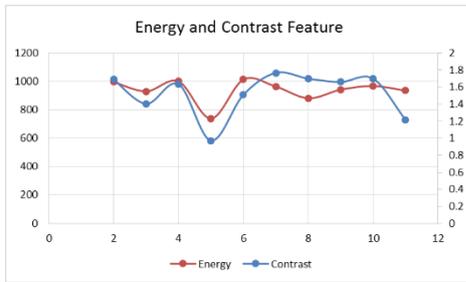


Figure 14. Energy and Contrast features from U-SGDLM feature set.

Fractal dimension feature set

The Hurst exponent $H(q)$ [17], for q values $\{1, 2, 3, \text{ and } 4\}$ are used to extract fractal dimensions in terms of Hurst component features. In addition to it, a derived feature perimeter2 / area. The index of independence of the crc images are evaluated in terms of Hurst components, $H(q) = H1, H2, H3$ and $H4$. Also, as the index $q > 2$, the independence index is variant. For this reason, only first two Hurst components are considered for analysis. From figure 15, it is observed that an high values of $H1$ and $H2$ among the CRC image set.

Table 4. FD - Hurst Component feature Set

Feature set	H1	H2	H3	H4	(perimeter) ² / area
CRC BIOPSY					
HE-01	0.199	0.197	0.132	0.09	6470.356
HE-02	0.119	0.154	0.197	0.17	957.291
HE-03	0.125	0.115	0.131	0.20	4025.59
HE-04	0.181	0.201	0.215	0.21	396.551
HE-05	0.137	0.162	0.191	0.11	503.09
HE-06	0.175	0.192	0.216	0.17	683.442
HE-07	0.169	0.178	0.167	0.13	8751.861
HE-08	0.139	0.164	0.168	0.17	1552.391
HE-09	0.112	0.133	0.18	0.18	656.626
HE-10	0.179	0.179	0.136	0.115	3386.132

The other derived feature in FD is perimeter2/ area, in figure 16 show the image HE-01 and HE-07 are having similar values and cross validate the categorization of HE-01, 07, 10 into tumor epithelium tissue

5. Conclusion

The upgraded SGLDM method used in the paper is based on whole segmentation of the image. Experiments are conducted on the CRC image set, which are released by Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). The biological theory of CRC image set is analyzed by considering, the improved the textural features

obtained from U-SDGLM method. The CRC image set is validated as one of the eight colon tissues. Further the analyzation is validated for longer periods by extracting Hurst component in fractional dimension feature set. As well as, this method can be integrated in the process of the classification of CRC biopsy[10].

Figure 15. Hurst Components for the CRC Image set.

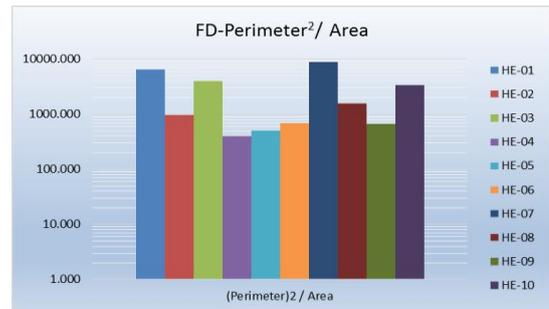
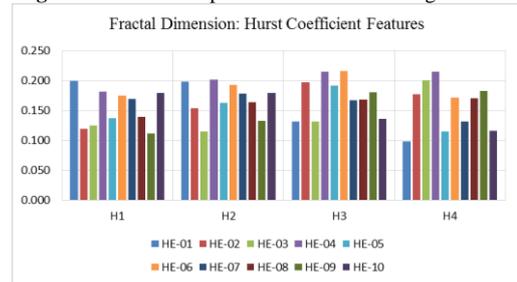


Figure 16. FD-Perimeter2/Area for the CRC Image set.

References

- [1] Sassi OB, Sellami L, Slima MB, Chtourou K & Hamida AB, "Improved spatial gray level dependence matrices for texture analysis", *International Journal of Computer Science & Information Technology*, Vol.4, No.6, (2012).
- [2] Saroja B & SelwinMichpriyadharson A, "Colon cancer Detecion Methods–A Review", *Annual Research & Review in Biology*, Vol. 24, No.2, (2018), pp.1-16.
- [3] Kather JN, Weis CA, Bianconi F, Melchers SM, Schad LR, Gaiser T, Marx A & Zollner F, "Multi-class texture analysis in colorectal cancer histology", *Scientific Reports*, (2016).
- [4] Weszka JS, Dyer CR & Rosenfeld A, "A comparative study of texture measures for terrain classification", *IEEE transactions on Systems, Man, and Cybernetics*, (1976), pp.269-285.
- [5] Galloway MM, "Texture Analysis Using Gray Level Run Lengths", *Computer Graphics and Image Processing*, Vol.4, No.2, (1975), pp.172-179.
- [6] Foster K, Dewbury KC, Griffith AH & Wright AHR, "The Accuracy of Ultrasound in the Detection of Fatty Infiltration of the Liver", *British Journal of Radiology*, Vol.53, (1980), pp.440-442.
- [7] Kim JK, Park JM, Songt KS & Park HW, "Detection of Clustered Microcalcifications on Mammograms Using Surrounding Region Dependence Method and Artificial Neural Network", *Journal of VLSI Signal Processing*, Vol.18, No.3, (1997), pp.251-262.
- [8] Kakkos SK, Stevens JM, Nicolaidis AN, Kyriacou E, Pattichis CS, Geroulakos G & Thomas D, "Texture Analysis of Ultrasonic Images of Symptomatic Carotid Plaques can Identify those Plaques Associated with Ipsilateral Embolic Brain Infarction", *European Journal of Vascular and Endovascular Surgery*, Vol.33, (2007), pp 422-429.
- [9] Chen SJ, Cheng KS, Dai YC, Sun YN, Chen YT, Chang KY, Hsu WC & Chang TW, "The Representations of Sonographic Image Texture for Breast Cancer Using Co-occurrence Matrix", *Journal of Medical and Biological Engineering*, Vol.25, No.4, (2005), pp 193-199.
- [10] Saroja B & SelwinMichpriyadharson A, "Adaptive Pillar K-means Clustering – based Colon Cancer Detection from Biopsy Samples with Outliers", *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, (2017).
- [11] Bryce RM & Sprague KB, "Revisiting detrended fluctuation analysis", *Scientific Reports*, (2012).