DOI: http://dx.doi.org/10.26483/ijarcs.v9i1.5333

Volume 9, No. 1, January-February 2018

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

FEATURE EXTRACTION OF COLON AND RECTUM CANCER FROM MRI IMAGES

R. Srivaramangai
Department of Information Technology
University of Mumbai
Mumbai, India

Ajay S. Patil
School of Computer Sciences
North Maharashtra University
Jalgaon, India

Nita Patil
School of Computer Sciences
North Maharashtra University
Jalgaon, India

Abstract: Feature extraction, the third major component plays a vital role on the final expected outcome of the application using digital image processing. Feature plays a very significant role in the area of digital image processing. The image feature detection and extraction are based on the colour, shape and texture feature categories. Colour feature refers to the presence of one of the colours either in RGB space (Red, Green, Blue colour model) or in HSxspace (Hue, Saturation, Value/Brightness model) in the particular area of interest. Especially in medical imaging, the internal parts will have a specific colour slightly varying in hue. When the scan modality provides a colour image, colour feature can be used for processing. The structure of the organs and its parts if visible clearly, then the shape parameter can be used for classification of images. The knowledge of the domain experts will enable to find the structure as what and where it is located. Similarity in shape will help in identifying the objects. Irrespective of colour and shape, texture parameter plays an important role in classification problems. In the absence of colour and shape, the only available feature is the textural features of the object. It looks into the properties related with intensity values of each pixel. This paper proposes the textural feature extraction techniques that best suits for retrieving the properties of colon and rectum cancer MRI images which can be further processed for classification.

Keywords: Feature Selection, Feature Extraction, Statistical Parameters, Gabor, GLCM.

INTRODUCTION

Feature extraction is a process of image processing which is used to select and extract those features (properties) which are helpful in identifying the problem of interest. It is a methodology followed not only in digital image processing but also in machine learning, pattern recognition and computer vision [01]. It is also used for reducing the volume of the image by taking only the required features from the available feature set. Broadly the feature image processing can be classified into three main steps: Feature detection, feature selection and feature extraction. Feature detection focused on finding the presence of a feature of certain type (according to the application) in that image. The output of this operation need not be binary but can later be converted into a binary property. The presence of feature takes to the next level of decision making that which are the features that are to be selected so that the application using this is benefitted by the recognition. Generally, the selected features are given as inputs to classifiers which will categorize according to the requirement by matching the selected feature parameters. There are various methods for feature selection and deciding on which feature parameters are to be selected [02]. The processing of the image(s) and to manipulate them to get some findings is dependent on feature extraction. Thus, it can be said as a set of data that are derived from the images which contains certain information. This information needs to be computed or checked to infer the outcomes. When the input data or data set is too large and some of them are redundant, then a subset of those data is extracted to determine the outcome. Feature extraction techniques are helpful in various image processing applications especially in medical imaging. Image pre-processing and feature extraction techniques are required for any image related applications.

LITERATURE SURVEY

Feature extraction plays a major role in the process of diagnosis since the extracted features and their properties and behaviour decide on the type of disorder under study. Like segmentation, there are numerous techniques for extracting the medical data parameters from the images. Finally, feature was extract and compared with these standard metrics. Thus, the proposed method performs better than the existing works. Gladis et al., in their work on brain tumour classification [03], have used linear discriminate analysis for extracting the features of brain tumour taking into consideration intensity and texture features. These are compared with (PCA). It has been found that the number of features extracted by linear discriminant analysis (LDA) selected or features extracted by LDA are more accurate than PCA. Kishore in a research on effective way of feature extraction for lung cancer [04], has developed image clustering technique which is an unsupervised method for feature selection. CBIR and Image mining have been used for retrieval of features. A research on feature extraction of MRI Brain Images [05], has proposed a new technique by combing the topological gradient approach and watershed transformation for segmenting the image. It has reduced the unwanted contours and for obtaining a better result through the topological gradient approach. Kathiravan[06] has used PCA analysis for feature reduction and have used textural features after de noising the image. Christe et al., [07] has used statistical parameters for extracting the features though not related with medical imaging. Priya [08] has used statistical parameters with acoustic shadowing added as one of the feature for cervical cancer classification. Ada and Kaur [09] have also used statistical parameters for lung cancer detection. Dansheng et al., [10] has used six non-redundant features for classification. They have used LDA for feature reduction. Amutha et al., having used active contour model for segmentation of lung cancer nodules have used gray level co-occurrence matrix (GLCM) and statistical parameters [11]. Akif et al., [12] applied gray level run length matrix (GRLM) texture based features for segmentation and found a high accuracy rate. Doyle et. al., [13] used textural and nuclear architectural features for analysis for histo-pathological image of breast cancer. Most of the research related with medical images use texture features than shape and colour features since the malignant cell growth erodes the shape of the organ. Ribo et al., [14] proposed an algorithm that can extract the shape feature considering the adjacent parts is in the tree structure so as to get the best combination of the features. Compared with the traditional extracting algorithms which aim at the whole picture or a single part, the experiments show that the proposed algorithm can obtain a higher diagnostic accuracy. Bhuvaneswariet. al., [15] in their work on lung cancer have done the feature extraction using Gabor filter, Walsh Hadamard transform. The feature selection is done by the correlation based feature selection for lung tumour. A new breast cancer detection algorithm, named as Gabor cancer detection (GCD) algorithm has been developed by Yufeng [16], where Gabor features are used and a Gabor filter bank is formed with five bands by four orientations (horizontal, vertical, 45 and 135 degree) in Fourier frequency domain. For breast cancer detection, in each mammographic image, twenty Gabor-filtered images are extracted to do the classification. Hemanthet. al., [17] has used statistical features which are most commonly used for brain image analysis. These features are specifically found to be superior for image classification and segmentation applications. Han and Chen [18] have proposed an approach for the automatic modality classification of medical images. In their approach, visual feature and textual feature were extracted for modality classification. They also suggested fusion of different extracted visual and textual features. Visual feature was extracted based on global and local features. Statistical features from the local area of the images from the CT images of Chest has extracted by Murphy et al., [19]. LDA is used for brain tumour feature selection and extraction by Gladis et al., [03]. They have constantly used LDA in their research works related with tumour detection. Eltoukhy et al., [20] in their research on breast cancer diagnosis presents a new method for digital mammogram images. Multi-resolution representations, wavelet or curvelet, are used to transform the mammogram images into a long vector of coefficients. A feature extraction method is developed based on the statistical t-test method. The method is ranking the features according to its capability to differentiate the classes. Then, a dynamic threshold is applied to optimize the number of features, which can achieve the maximum classification accuracy rate. The method depends on extracting the features that can maximize the ability to discriminate between different classes. Saraswati et al., [21] has used an efficient feature extraction method using curvelet transform and feature selection using swarm intelligence. In this work, coefficients are extracted from the region of interest (ROI) mammogram images by fast discrete curvelet transform via wrapping method.

Llobet et al., [22], presented and compared five different feature extraction methods for breast cancer detection in digitized mammograms. All the methods are based on features extracted from a local window. Li, Shutao, and Liao [23] used kernel-based feature extraction method from gene expression data is proposed for cancer classification. The performances of four kernel algorithms, namely, kernel Fisher discriminant analysis (KFDA), kernel principal component analysis (KPCA), kernel partial least squares (KPLS), and kernel independent component analysis (KICA), are compared on three benchmarked datasets: breast cancer, leukemia and colon cancer. Experimental results show that the proposed kernel-based feature extraction methods work well for three benchmark gene dataset. Overall, the KPLS and KFDA show the best performance, followed by KPCA and KICA. Novel schemes are developed to extract new texture features from the texture spectra in the chromatic and achromatic domains, and colour features for a selected region of interest from each color component histogram of the colonoscopic images. These features are reduced in size using PCA in the work carried out by Tjoa et al., [24]. Huang et al., [25] has investigated the relationship between the B-scan optical coherence tomography (BOCT) image features and histology of malignant human colorectal tissues, also en-face OCT image and the endoscopic image pattern. The in-vitro experiments were performed by a swept-source optical coherence tomography (SSOCT) system. Liu et al., [26] conducted a feature extraction method named Optimal Mean based Block Robust Feature Extraction method (OMBRFE) to identify feature genes associated with advanced colorectal cancer in clinical stage by using the integrated colorectal cancer data.

FEATURE EXTRACTION TECHNIQUES

A general feature extraction and classification based on features is depicted in figure 1. In this, the features extracted from the set of training images are stored in a database. When a testing image is given as input for classification, the same type of features of that testing image are extracted and a pattern matching is done by retrieving those patterns of features in the database. The result of the pattern recognition is checked for correctness. Thus, for any classification problem, feature extraction plays a crucial role.

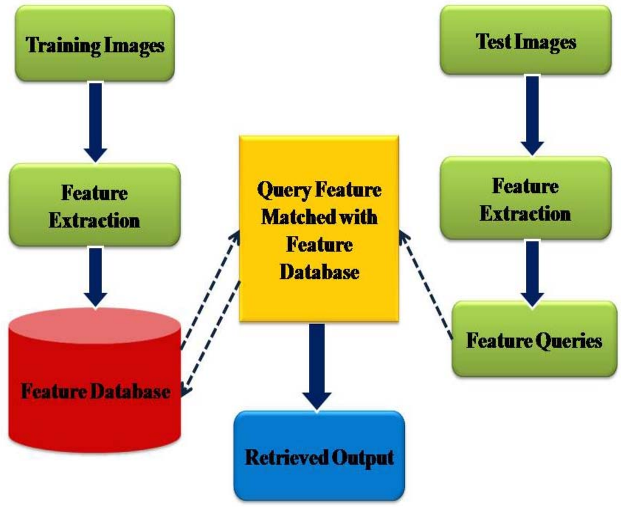

Figure 1. A General Feature Extraction Process

An image has three main characteristic visual features namely shape, colour and texture. Such features are extracted to analyze and classify them as per the properties of those features which is usually done in Content-Based Image Retrieval (CBIR) systems [09][10]. Thresholding is a most popular method of extracting integral parts and shapes of an object when there is enough control on lighting and illumination. It also helps in gathering and stipulating the low-level features. When there is a need of identifying the shape and extracting that shape, template matching is one probable solution. It is a model based approach, where the shape is extracted by finding the best correlation between the template that is used (known model) and the pixels of the image under consideration. Apart from shape, colour plays an important role in identifying the region of interest and extracts the ROI. If the image acquired is a colour image like CT scan, PET scan, digital photographic images etc., colour can also be used as a feature parameter with the values of intensities of the colour present in that region. In case of medical images, especially for cancers, the tumour will result in lump formation which thickens the skin and darkens the colour. Any swelling will have reddishness around the affected area. So, the possible solution in colour images is to identify the original colour of the organ and the suspected region of interest. Texture is one of the most significant characteristics of an image. It is used to describe the local spatial differences in image brightness which is related to image properties such as coarseness, and regularity. This is achieved by performing numerical management of digitized.

PROPOSED FEATURE EXTRACTION TECHNIQUES FOR COLON AND RECTUM CANCER MRI

In colon and rectum MRI image, the colour feature is not required since it is a black and white image which then converted to grayscale image[. Since the objective of this research is to find the abnormality if it exists and to detect the stage of colorectal cancer, the spread of cancer in the colon and rectum hides the shape of rectum. Moreover, the slices of the images obtained through the axial T2 orientation need not have the uniform representation where rectum will be clearly visible as it is in other images. Therefore, textural features become significant for classification of the cancer. There are diverse techniques called as filters available to extract the variety of textural features in an image. Among those, the filters which are widely used for medical imaging especially cancer MRI imaging are Gabor filter, statistical parameter extraction and Gray Level Co variance Matrix parameters. The study on comparison of feature extraction techniques on cancer images has given a better performance result of the classification with these filters mentioned. This chapter gives the description of these three filters and extracts the features for colon and rectum. The extracted features are stored in .csv format for further processing.

A. *Statistical Parameters Features*

These are techniques which are intensity based features basically using the moments. In order to calculate the moment, mean, standard deviation, kurtosis, skewness are used. Here we are extracting these parameters of the image along with entropy. This comes under content based image retrieval system (CBIR)[13].Each image and the corresponding histogram consist of range of pixel values. Table 1 gives the formula for calculation of these features. The mean of these pixel values is calculated which may represent brightness or called as high intensity pixels. Bright image should have high mean while dark image should have low mean, and also mean values characterize individual calcifications. The Standard Deviation (SD) is computed by using this mean value which is usually used to give information on the contrast of the image. SD also characterizes the cluster. Skewness is basically meant for detecting the edges of darker region in an image from the white background. It also implies how the values are directed. The values can be positively or negatively skewed. Skew measures are how asymmetry (unbalance) the distribution of the gray level. If the image shows bimodal Image histogram distribution of the object then it should have high variance and low skew distribution (one peak at each side of mean).

Energy measurement is closely related to skew. Highly skew distribution usually gives high-energy measurement.Kurtosis is used to find the resolution level of the image. It finds the measures of noise and resolution, which means a high kurtosis value infer a low noise and low resolution images and vice versa. It can otherwise be called as the measure of sparse.Image with good contrast should have high variance. Entropy measures the average number of bits to code each gray level. It has inverse relationship with skew and energy measurement. Highly skewed distribution tends to yield low Entropy

Table 1. General Statistical Parameters based on Moments

Mean	$\mu = \frac{\Sigma X}{N}$
Standard Deviation	$\sigma = \sqrt{\frac{\sum(Xi-\mu)^2}{N}}$
Kurtosis	$K = n \frac{\sum_{i=1}^{n}(x_i - x_{avg})^4}{\left(\sum_{i=1}^{n}(x_i - x_{avg})^2\right)^2}$
Skewness	$S = \sqrt{n}\frac{\sum_{i=1}^{n}(x_i - x_{avg})^3}{\left(\sum_{i=1}^{n}(x_i - x_{avg})^2\right)^{3/2}}$
Energy	$E = \sum_{i,j=0}^{N-1}(P_{ij})^2$

B. *Gray Level Co-occurrence Matrix*

Gray Level Co-occurrence Matrix (GLCM) is a statistical method which evaluates the texture based on the spatial relationship of pixels which is also known as the gray-level spatial dependence matrix[09]. The GLCM uses its functions and creates a matrix which portrays the texture of an image. It calculates the total number of times each pair of with a particular value and definite spatial relationship occur in an image. From the GLCM created, the statistical measures are extracted. GLCM refers the pair of pixels of comparison as reference pixel and neighbouring pixel when it calculates the spatial relationship between them. The representation (1,0) relation means that 1 pixel is in the x direction, 0 pixels is in the y direction. The GLCM method is a way of extracting second order statistical texture features [27].The matrix is constructed by modelling the spatial relationship between pixels of that region.

Table 2. GLCM – Features based on Spatial Relationships

Energy	$E = \sum_{i,j=0}^{N-1}(P_{ij})^2$
Entropy	$E = \sum_{i,j=0}^{N-1} -\ln(P_{ij})P_{ij}$
Contrast	$\sum_{i,j=0}^{N-1} P_{ij}(i-j)^2$

Homogeneity	$\sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2}$
Correlation	$\sum_{i,j=0}^{N-1} P_{ij} \frac{(i-\mu)(j-\mu)}{\sigma^2}$
Shade	sgn(A) $\lvert A \rvert^{1/3}$
Prominence	sgn(B) $\lvert B \rvert^{1/4}$

where

P_{ij} = Element i,j of the normalized symmetrical GLCM

N = Number of gray levels in the image

μ = the GLCM mean $\mu = \sum_{i,j=0}^{N-1} iP_{ij}$

σ2 = the variance of the intensities of all the pixels that are referenced in the relationships that were used in GLCM and the formula is:

$$\sigma^2 = \sum_{i,j=0}^{N-1} P_{ij}(i-\mu)^2,$$

$$A = \sum_{i,j=0}^{N-1} \frac{(i+j-2\mu)^3 P_{ij}}{\sigma^3\left(2\sqrt{(1+C)}\right)^3},$$

C = the correlation feature,

sgn(x) =Sign of a real number, x = -1 for x < 0, x = 0 for x = 0, x = 1 for x > 0 and

$$B = \sum_{i,j=0}^{N-1} \frac{(i+j-2\mu)^4 P_{ij}}{4\sigma^4(1+C)^2}$$

The GLCM follows a second-order joint conditional probability density functions stated as P(i, j | d, θ) for θ= 0, 45, 90, 135°, etc., are the various directions and different distances are denoted by d = 1, 2, 3, 4, and 5. The probability that two pixels, at an inter-sample distance d indirection θ, have a gray level i and j which is represented as a function P (i, j | d, θ). The spatial relationship is denoted in terms of distance d and angle θ. If the texture is smutty at a very small distance d, then the pair will have similar gray values. For smooth and fine texture, the pairs of pixels at distance d should often be quite different, so that the values in the GLCM are relatively uniformly distributed. Similarly, if the texture smuttier in one direction than another, then the degree of distribution of the values over the main diagonal in the GLCM will vary with the value of θ.

C. *Gabor Filter*

Gabor filter, named after Dennis Gabor, is a linear filter in digital image processing which is especially used for edge detection. It works with finding the magnitude and direction values of the pixel elements with respect to frequency and orientation of the image at each pixel point. Gabor filters are found to be more suitable for texture representation and discrimination since resembles the human visual system[28]. The different set of Gabor filters with different orientations and frequencies will produce a host of features that can be extracted from the image. A Gabor filter is essentially a sinusoidal signal with a given frequency and orientation, modulated by a Gaussian. The general form of a circular 2-D Gabor filter in the spatial domain is represented as:

$$G(x,y,\theta,u,\sigma) = \frac{1}{2\pi\sigma^2} exp2\left\{-\frac{x^2+y^2}{2\sigma^2}\right\} exp\{2\pi i(uxcos\theta + uysin\theta)\}$$

where i=$\sqrt{-1}$, uis the frequency of the sinusoidal wave, controls the orientation of thefunction and $\sigma$is the standard deviation of the Gaussian envelope.The set of frequencies and orientations is designed to localize different, roughly orthogonal, subsets of frequency and orientation information in the input image[17].

EXPERIMENTAL RESULTS

Statistical parameters which are extracted for abnormal colon images are given in table 3 for a sample of 10 images. There are 250 images of colon abnormal images for which the features were extracted. The table 4 shows the statistical features of normal colon images for a sample of 10 images. Here also 250 normal images were taken for input and their features were extracted. These features were taken as input parameters for the next module for classifying the colon images into normal and /or abnormal. Similarly, the statistical parameters of 250 normal rectum images and 250 abnormal rectum images were extracted after segmenting them. The table 5 and 6 gives the sample of 10 abnormal rectum images and 10 normal rectum images.

Table 3. Statistical Parameters of sample of 10 Abnormal Colon Images

Images	*Skewness*	*Kurtosis*	*Mean*	*Standard Deviation*	*Entropy*
1	0.50	2.07	77.77	78.85	5.67
2	1.42	4.88	50.31	59.59	5.84
3	1.34	3.74	47.73	68.40	4.82
4	1.41	4.39	42.65	60.59	4.68
5	1.34	4.16	43.59	60.90	4.63
6	1.34	4.16	44.41	61.89	4.67
7	1.36	3.92	46.10	66.00	4.71
8	1.34	3.79	46.60	66.63	4.78
9	1.34	3.84	47.65	67.66	4.82
10	1.35	3.80	47.88	68.56	4.84

Table 4. Statistical Parameters of sample of 10 Normal Colon Images

Images	*Skewness*	*Kurtosis*	*Mean*	*Standard Deviation*	*Entropy*
1	0.23	2.17	96.76	72.65	7.21
2	0.24	2.18	96.66	72.55	7.23
3	0.24	2.18	96.61	72.49	7.23
4	0.27	2.25	96.48	71.06	7.20
5	0.27	2.25	96.46	71.04	7.20
6	0.27	2.25	96.49	71.10	7.20
7	0.27	2.25	96.47	71.00	7.20
8	0.27	2.25	96.49	71.05	7.20
9	1.03	3.13	61.27	64.22	6.96
10	1.03	3.14	61.24	64.22	6.96

Table 5. Statistical Parameters of sample of 10 Abnormal Rectum Images

Images	*Skewness*	*Kurtosis*	*Mean*	*Standard Deviation*	*Entropy*
1	0.57	3.74	90.16	51.20	7.45
2	-0.04	1.84	107.19	75.45	7.10
3	-0.09	1.84	107.13	73.50	7.15
4	-0.06	1.78	103.64	73.89	7.03
5	0.29	2.83	101.24	64.79	7.28
6	0.24	2.98	108.24	63.27	7.32
7	0.79	2.38	83.62	72.69	7.50
8	0.88	2.52	78.74	72.28	7.42
9	0.36	2.83	103.75	65.69	7.39
10	0.36	2.89	102.52	64.75	7.35

The 20 GLCM parameters have been extracted avoiding the redundant properties. The redundant properties have been removed. The experimental results of GLCM and Gabor filter are listed in appendix from Table 7 to Table 16. Table 7 shows the first 10 values of colon whereas Table 8 shows the last 10 values of colon extracted from the GLCM.Similarly,Table 9and 10 shows the GLCM properties ranging from 1-10 and 11-20 of rectum images. In tables 11 to 13, the Gabor magnitude values of colon images are shown when frequencies 1.0, 2.0 and 3.0. Tables from 14 to 16 shows the Gabor magnitude values of rectum images with frequencies with 1.0, 2.0 and 3.0

Table 6. Statistical Parameters of sample of 10 Normal Rectum Images

Images	*Skewness*	*Kurtosis*	*Mean*	*Standard Deviation*	*Entropy*
1	-0.06	1.71	124.91	81.08	7.82
2	0.14	1.63	119.02	80.77	7.87
3	0.22	2.57	122.40	60.26	7.74
4	0.21	2.59	121.29	60.22	7.74
5	0.09	1.63	122.80	80.25	7.87
6	0.04	1.64	126.58	79.37	7.88
7	0.22	2.56	121.47	60.48	7.75
8	0.23	2.57	120.58	60.35	7.75
9	0.21	1.66	114.95	81.02	7.85
10	0.18	1.62	115.59	81.85	7.84

CONCLUSION

This research has shown the ability of three different filters for extraction of textural features to discriminate the classes of colon and rectum cancer as normal and abnormal and further for investigation of stages. A Gabor filter encodes the textured images into multiple narrow frequency and orientation channels. The obtained preliminary results are interesting. Texture analysis can be used in the process of object detection. Since the number of magnitude values of each orientation for a particular frequency of one image differs from that of each image of the data set, it is very difficult to train a classifier without prior knowledge of the number of parameters to be used. Statistical parameters and GLCM give a standard method of evaluating the images by its feature parameters. Gabor has proven to give better results in some medical image processing. So all these three prominent techniques are used for extracting features. This outcome is for MRI Images of Colon and Rectum and so for different modalities and for different medical images the evaluation needs to be done by experimenting with those images.

REFERENCES

[1] Pergad, Shri ND, S. T. Hamde; "Review on Feature Extraction and Classification of WBC in Bone Marrow for Disease Diagnosis." Innov Res Elec Electron Instrum Control Eng4, No. 1, 2016, pp. 68-73.

[2]Eltoukhy, Mohamed Meselhy, Ibrahima Faye, and BrahimBelhaouari Samir; "A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation.",Computers in biology and medicine 42, No. 1, 2012, pp. 123-128.

[3] GladisRathi, V. P., S. Palani; "Brain tumor MRI image classification with feature selection and extraction using linear discriminant analysis." arXiv preprint arXiv:1208.2128, 2012.

[4] R. Kishore; "An Effective and Efficient Feature Selection Method for Lung Cancer Detection"; International Journal of Computer Science &Information Technology (IJCSIT); Vol. 7; No. 4; 2015; pp. 135-141.

[5] Sivaperumal, S., M. Sundhararajan, T. Nadu; "Advance Feature Extraction of MRI Brain Image and Detection using Local Segmentation Method with Watershed.",Vol.3, 2013, pp. 87-94.

[6] S.Kathiravan; "An Investigation on Image Denoising Technique using Pixel-Component-Analysis"; Innovative Systems Design and Engineering; Vol. 2, No. 4, 2011, pp. 233-241.

[7] S. AllinChriste, M.Vignesh, A.Kandaswamy; "An Efficient FPGA Implementation of MRI Image Filtering and Tumour Characterization using Xilinx System Generator", International Journal of VLSI design & Communication Systems (VLSICS), Vol.2, No.4, 2011, pp.95-109.

[8] SakthiPriya; "Cervical Cancer Screening and Classification Using Acoustic Shadowing; International Journal of Innovative Research in Computer and Communication Engineering; Vol. 1, No. 8, 2013, pp. 1676-1679.

[9] Ada, Rajneet Kaur, "Feature Extraction and Principal Component Analysis for Lung Cancer Detection in CT scan Images", IJARCSSE, Vol. 3, No. 3, 2013, pp: 187-190.

[10] Dansheng Song, Tatyana A. Zhukov, Olga Markov, Wei Qian, Melvyn S. Tockman; "Prognosis of Stage ILung Cancer Patients through Quantitative Analysis Of Centrosomal Features", IEEE, 2012,pp: 1607-1610.

[11] A.Amutha, R.S.D Wahidabanu, "A Novel Method for Lung Tumor Diagnosis and Segmentation using Level Set-Active Contour Modelling", European Journal of Scientific Research, Vol.90, No.2, 2012, pp.175-187.

[12] AkifBurakTosun, CigdemGunduz-demir, "Graph Run-length Matrices for Histopathological Image Segmentation", IEEE Transactions On Medical Imaging, Vol. 30, No. 3, 2011.

[13] Scott Doyle, Shannon Agner, AnantMadabhushi, Michael Feldman, John Tomaszewski; "Automated Grading of Breast Cancer Histopathology using Spectral Clustering with Textural and Architectural Image Features", IEEE. 2008

[14] Ribo Ge, Hailong Shen; "Researching on Feature Extraction of Brain CT Image"; International Journal of

Signal Processing, Image Processing and Pattern Recognition", Vol. 6, No. 5, 2013, pp. 39-48.

[15] Bhuvaneswari C, Aruna P, Loganathan D; "Performance Evaluation of Feature Extraction and Feature Selection for Classification of Lung Diseases"; Journal of Theoretical and Applied Information Technology, Vol. 61, No. 2, 2014, pp. 389-394.

[16] Yufeng Zheng, "Breast Cancer Detection with Gabor Features from Digital Mammograms", Algorithms, Vol. 3, 2010, pp. 44-62.

[17] Hemanth, D. Jude, and J. Anitha. "Image Pre-processing and Feature Extraction Techniques for Magnetic Resonance Brain Image Analysis.", In Computer Applications for Communication, Networking and Digital Contents, pp. 349-356. Springer Berlin Heidelberg, 2012.

[18] Han, X., Chen, Y;"Modality Classification in Medical Image Retrieval: Multiple Feature Fusion with Normalized Kernel Function", Image CLEF2010 Workshop, (November 2010).

[19] K. Murphy, B.van Ginneken, A. M. R. Schilham, B. J. de Hoop,H. A. Gietema,, M. Prokop; "A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and knearest-neighbour classification", Medical Image Analysis, vol. 13, No.5, pp. 757–770, 2009.

[20] Eltoukhy, Mohamed Meselhy, Ibrahima Faye, and BrahimBelhaouari Samir; "A Statistical based Feature Extraction Method for Breast Cancer Diagnosis in Digital Mammogram using Multiresolution Representation." Computers in Biology and Medicine 42, no. 1, 2012, pp.123-128.

[21] Saraswathi, D., D. Dharani, E. Srinivasan; "An efficient Feature Extraction Technique for Breast Cancer Diagnosis using Curvelet Transform and Swarm Intelligence.",International Conference on, Wireless Communications,Signal Processing and Networking (WiSPNET),IEEE, 2016, pp. 441-445.

[22] Llobet, Rafael, Roberto Paredes, Juan C. Pérez-Cortés; "Comparison of Feature Extraction Methods for Breast Cancer Detection." Iberian Conference on Pattern Recognition and Image Analysis, Springer, Berlin, Heidelberg, 2005, pp. 495-502.

[23] Li, Shutao, Chen Liao; "Feature extraction for Cancer Classification using Kernel-Based Methods." Life System Modeling and Simulation, 2007,pp. 162-171.

[24] Tjoa, Marta P., Shankar M. Krishnan; "Feature extraction for the Analysis of Colon Status from the Endoscopic Images." BioMedical Engineering OnLine 2, No. 1, 2003.

[25] Huang, Shu-Wei, Shan-Yi Yang, Wei-Cheng Huang, Han-Mo Chiu, Chih-Wei Lu; "Study on Image Feature Extraction and Classification for Human Colorectal Cancer using Optical Coherence Tomography.", European Conference on Biomedical Optics, Optical Society of America, 2011.

[26] Liu, Jian, Yuhu Cheng, Xuesong Wang, Lin Zhang, Hui Liu; "An Optimal Mean Based Block Robust Feature Extraction Method to Identify Colorectal Cancer Genes with Integrated Data.", Scientific Reports 7, 2017.

[27] Varalakshmi.K;"Classification of Lung Cancer Nodules using a Hybrid Approach"; Journal of Emerging Trends in Computing and Information Sciences, Vol. 4, No. 1, 2013, pp. 63-68.

[28] Bhuvaneswari C, Aruna P, Loganathan D; "Performance Evaluation of Feature Extraction and Feature Selection for Classification of Lung Diseases"; Journal of Theoretical and Applied Information Technology, Vol. 61, No. 2, 2014, pp. 389-394.

APPENDIX

Table 7: GLCM Parameters of Abnormal Colon Images listing 1-10 parameters

Images	*Autocor relation*	*Contrast*	*Correlation*	*Cluster Prominence*	*Cluster Shade*	*Dissimilarity*	*Energy*	*Entropy*	*Homogeneity*	*Maximum Probability*
1	14.67	0.32	0.97	822.02	47.92	0.24	0.23	2.23	0.89	0.44
2	7.80	0.34	0.94	658.81	58.79	0.24	0.26	2.07	0.89	0.49
3	8.83	0.12	0.98	904.97	84.09	0.11	0.40	1.68	0.95	0.62
4	7.26	0.10	0.98	628.94	59.99	0.10	0.41	1.56	0.95	0.63
5	7.42	0.10	0.98	607.17	57.65	0.09	0.41	1.56	0.95	0.62
6	7.62	0.11	0.98	642.65	60.44	0.10	0.40	1.59	0.95	0.62
7	8.34	0.11	0.98	815.85	75.98	0.10	0.41	1.62	0.95	0.63
8	8.50	0.11	0.98	831.40	77.88	0.10	0.41	1.63	0.95	0.63
9	8.71	0.11	0.98	877.20	80.86	0.10	0.40	1.66	0.95	0.62
10	8.88	0.13	0.98	934.68	85.86	0.11	0.39	1.70	0.94	0.62

TABLE 8: GLCM PARAMETERS OF ABNORMAL COLON IMAGES LISTING 11-20 PARAMETERS

Sample Images	*Sum of Squares of Variance*	*Sum of Average*	*Sum of Variance*	*Sum of Entropy*	*Difference Variance*	*Difference Entropy*	*Information Measure of Correlation*	*Inverse Difference (INV) Homogeneity*	*Inverse Difference Normalized (INN)*	*Inverse Difference Moment Normalized*
1	14.73	6.26	38.00	2.00	0.32	0.60	-0.63	0.89	0.97	1.00
2	7.90	4.49	18.40	1.84	0.34	0.60	-0.57	0.89	0.97	1.00
3	8.82	4.51	23.69	1.58	0.12	0.35	-0.75	0.95	0.99	1.00
4	7.25	4.19	18.92	1.48	0.10	0.32	-0.76	0.95	0.99	1.00
5	7.41	4.26	19.40	1.48	0.10	0.32	-0.76	0.95	0.99	1.00
6	7.61	4.30	19.91	1.50	0.11	0.33	-0.76	0.95	0.99	1.00
7	8.33	4.41	22.30	1.53	0.11	0.33	-0.76	0.95	0.99	1.00
8	8.48	4.44	22.77	1.54	0.11	0.34	-0.76	0.95	0.99	1.00
9	8.70	4.50	23.31	1.57	0.11	0.34	-0.76	0.95	0.99	1.00
10	8.87	4.52	23.79	1.59	0.13	0.36	-0.74	0.95	0.99	1.00

TABLE 9: GLCM PARAMETERS OF ABNORMAL RECTUM IMAGES LISTING 1-10 PARAMETERS

Sample Images	*Autocorrelation*	*Contrast*	*Correlation*	*Cluster Prominence*	*Cluster Shade*	*Dissimilarity*	*Energy*	*Entropy*	*Homogeneity*	*Maximum Probability*
1	13.62	0.12	0.98	359.94	18.33	0.12	0.15	2.20	0.94	0.24
2	20.45	0.15	0.99	676.07	3.20	0.14	0.12	2.46	0.93	0.23
3	20.21	0.14	0.98	618.79	0.13	0.14	0.12	2.45	0.93	0.22
4	19.47	0.15	0.98	609.04	2.43	0.14	0.13	2.43	0.93	0.24
5	17.40	0.19	0.97	565.26	18.34	0.18	0.12	2.42	0.91	0.19
6	18.81	0.19	0.97	541.39	13.81	0.18	0.12	2.43	0.91	0.21
7	14.54	0.60	0.94	876.44	65.46	0.37	0.14	2.73	0.83	0.33
8	13.55	0.60	0.94	909.50	72.86	0.36	0.16	2.63	0.84	0.37
9	18.17	0.15	0.98	613.56	22.71	0.14	0.13	2.35	0.93	0.23
10	17.73	0.15	0.98	576.25	20.53	0.14	0.13	2.34	0.93	0.23

TABLE 10: GLCM PARAMETERS OF ABNORMAL RECTUM IMAGES LISTING 11-20 PARAMETERS

Sample Images	*Sum of Squares of Variance*	*Sum of Average*	*Sum of Variance*	*Sum of Entropy*	*Difference Variance*	*Difference Entropy*	*Information Measure of Correlation*	*Inverse Difference (INV) Homogeneity*	*Inverse Difference Normalized (INN)*	*Inverse Difference Moment Normalized*
1	13.57	6.67	30.88	2.11	0.12	0.37	-0.75	0.94	0.99	1.00
2	20.40	7.90	50.39	2.35	0.15	0.41	-0.76	0.93	0.98	1.00
3	20.15	7.89	49.50	2.34	0.14	0.41	-0.76	0.93	0.98	1.00
4	19.43	7.70	47.75	2.31	0.15	0.41	-0.76	0.93	0.98	1.00
5	17.38	7.43	41.17	2.27	0.19	0.48	-0.70	0.91	0.98	1.00
6	18.79	7.84	44.90	2.28	0.19	0.49	-0.70	0.91	0.98	1.00
7	14.74	6.24	35.04	2.34	0.60	0.80	-0.52	0.84	0.96	0.99
8	13.75	5.94	33.11	2.25	0.60	0.78	-0.52	0.85	0.96	0.99
9	18.13	7.59	43.86	2.24	0.15	0.41	-0.75	0.93	0.98	1.00
10	17.69	7.51	42.56	2.23	0.15	0.41	-0.75	0.93	0.98	1.00

TABLE 11: SAMPLE OF 10 GABOR FEATURES (MAGNITUDE) OF ONE COLON IMAGE WITH FREQUENCY = 1.0

θ = 10	θ = 20	θ = 30	θ = 40	θ = 50	θ = 60	θ= 70	θ = 80	θ = 90	θ = 100	θ = 110	θ = 120	θ = 130	θ = 140	θ=150	θ = 160	θ = 170	θ = 180
0.37	4.26	-0.29	0.26	-0.64	-0.35	-0.70	-0.41	1.08	3.21	0.40	-0.36	-0.50	0.06	-0.13	-0.41	-0.12	-0.08
0.96	2.46	-0.09	0.17	-0.50	-0.09	-0.67	-0.51	1.15	2.85	0.56	-0.52	-0.37	0.09	-0.43	-0.63	-0.11	-0.03
0.98	-0.10	0.06	0.39	-0.19	0.06	-0.66	-0.56	0.30	1.88	-0.17	-0.62	-0.48	-0.33	-0.72	-0.62	-0.38	-0.20
0.75	-0.06	0.30	0.10	-0.24	0.36	-0.81	-0.20	-0.07	1.23	-0.61	-0.45	-0.43	-0.80	-0.49	-0.48	-0.31	-0.53
0.82	0.36	0.53	-0.56	-0.08	0.65	-0.76	-0.01	1.71	1.01	0.53	-0.39	0.12	-0.66	-0.15	-0.36	0.04	-0.83
1.59	1.45	0.77	-0.74	-0.06	0.50	-0.67	-0.24	3.78	0.32	1.89	-0.39	0.35	-0.47	0.23	-0.38	0.23	-0.65
2.49	2.97	1.06	-0.70	-0.33	-0.11	-0.45	-0.59	0.23	2.10	2.31	-0.61	0.15	-0.63	0.48	-0.38	0.31	-0.68
2.48	3.63	0.88	-0.10	-0.76	-0.21	-0.41	-0.02	0.89	1.93	1.19	-0.71	0.02	-0.87	0.35	-0.68	0.14	-0.84
1.51	3.62	0.37	0.13	-0.76	-0.25	-0.57	0.47	1.69	1.46	-0.50	-0.50	-0.02	-0.70	-0.25	-0.66	0.23	-0.45
0.83	3.33	0.12	-0.24	-0.66	-0.23	-0.54	0.13	1.91	0.87	-0.25	-0.44	-0.14	-0.68	-0.38	-0.39	0.09	-0.39

TABLE 12: SAMPLE OF 10 GABOR FEATURES (MAGNITUDE) OF ONE COLON IMAGE WITH FREQUENCY = 2.0

θ = 10	θ = 20	θ = 30	θ = 40	θ = 50	θ = 60	θ= 70	θ = 80	θ = 90	θ = 100	θ = 110	θ = 120	θ = 130	θ = 140	θ =150	θ = 160	θ = 170	θ = 180
0.96	3.63	0.88	-0.10	-0.76	-0.21	-0.41	-0.02	0.89	1.93	1.19	-0.71	0.02	-0.87	0.35	0.96	3.63	0.88
0.98	2.46	-0.09	0.17	-0.50	-0.09	-0.67	-0.51	1.15	2.85	0.56	-0.52	-0.37	0.09	-0.43	0.98	2.46	-0.09
0.75	-0.10	0.06	0.39	-0.19	0.06	-0.66	-0.56	0.30	1.88	-0.17	-0.62	-0.48	-0.33	-0.72	0.75	-0.10	0.06
0.82	-0.06	0.30	0.10	-0.24	0.36	-0.81	-0.20	-0.07	1.23	-0.61	-0.45	-0.43	-0.80	-0.49	0.82	-0.06	0.30
1.59	0.36	0.53	-0.56	-0.08	0.65	-0.76	-0.01	1.71	1.01	0.53	-0.39	0.12	-0.66	-0.15	1.59	0.36	0.53
2.49	1.45	0.77	-0.74	-0.06	0.50	-0.67	-0.24	3.78	0.32	1.89	-0.39	0.35	-0.47	0.23	2.49	1.45	0.77
4.26	2.97	1.06	-0.70	-0.33	-0.11	-0.45	-0.59	0.23	2.10	2.31	-0.61	0.15	-0.63	0.48	4.26	2.97	1.06
2.46	3.63	0.88	-0.10	-0.76	-0.21	-0.41	-0.02	0.89	1.93	1.19	-0.71	0.02	-0.87	0.35	2.46	3.63	0.88
-0.10	0.00	0.37	0.13	-0.76	-0.25	-0.57	0.47	1.69	1.46	-0.50	-0.50	-0.02	-0.70	-0.25	-0.10	0.00	0.37
-0.06	3.33	0.12	-0.24	-0.66	-0.23	-0.54	0.13	1.91	0.87	-0.25	-0.44	-0.14	-0.68	-0.38	-0.06	3.33	0.12

TABLE 13: SAMPLE OF 10 GABOR FEATURES (MAGNITUDE) OF ONE COLON IMAGE WITH FREQUENCY = 3.0

θ = 10	θ = 20	θ = 30	θ = 40	θ = 50	θ = 60	θ =70	θ = 80	θ = 90	θ = 100	θ = 110	θ = 120	θ = 130	θ = 140	θ =150	θ = 160	θ = 170	θ = 180
-0.21	-0.84	0.96	-0.10	-0.76	0.88	0.89	-0.02	0.35	1.93	-0.41	-0.71	0.02	1.19	0.35	-0.68	0.14	-0.21
-0.09	-0.03	0.98	0.17	-0.50	-0.09	1.15	-0.51	-0.43	2.85	-0.67	-0.52	-0.37	0.56	-0.43	-0.63	-0.11	-0.09
0.06	-0.20	0.75	0.39	-0.19	0.06	0.30	-0.56	-0.72	1.88	-0.66	-0.62	-0.48	-0.17	-0.72	-0.62	-0.38	0.06
0.36	-0.53	0.82	0.10	-0.24	0.30	-0.07	-0.20	-0.49	1.23	-0.81	-0.45	-0.43	-0.61	-0.49	-0.48	-0.31	0.36
0.65	-0.83	1.59	-0.56	-0.08	0.53	1.71	-0.01	-0.15	1.01	-0.76	-0.39	0.12	0.53	-0.15	-0.36	0.04	0.65
0.50	-0.65	2.49	-0.74	-0.06	0.77	3.78	-0.24	0.23	0.32	-0.67	-0.39	0.35	1.89	0.23	-0.38	0.23	0.50
-0.11	-0.68	4.26	-0.70	-0.33	1.06	0.23	-0.59	0.48	2.10	-0.45	-0.61	0.15	2.31	0.48	-0.38	0.31	-0.11
-0.21	-0.84	2.46	-0.10	-0.76	0.88	0.89	-0.02	0.35	1.93	-0.41	-0.71	0.02	1.19	0.35	-0.68	0.14	-0.21
-0.25	-0.45	-0.10	0.13	-0.76	0.37	1.69	0.47	-0.25	1.46	-0.57	-0.50	-0.02	-0.50	-0.25	-0.66	0.23	-0.25
-0.23	-0.39	-0.06	-0.24	-0.66	0.12	1.91	0.13	-0.38	0.87	-0.54	-0.44	-0.14	-0.25	-0.38	-0.39	0.09	-0.23

TABLE 14: SAMPLE OF 10 GABOR FEATURES (MAGNITUDE) OF ONE RECTUM IMAGE WITH FREQUENCY = 1.0

θ = 10	θ = 20	θ = 30	θ = 40	θ = 50	θ = 60	θ =70	θ = 80	θ = 90	θ = 100	θ = 110	θ = 120	θ = 130	θ = 140	θ =150	θ = 160	θ = 170	θ = 180
0.39	0.32	-0.63	-0.49	-0.35	-0.50	-0.30	0.10	-0.17	0.89	0.12	-0.72	0.50	-0.73	0.54	1.37	-0.72	0.54
0.86	-0.37	0.21	-0.63	-0.40	-0.40	-0.23	0.38	-0.43	-0.36	0.20	-0.84	0.40	-0.54	-0.25	1.35	-0.84	-0.55
0.36	-0.56	1.10	-0.53	-0.41	-0.21	0.29	0.78	-0.55	0.02	0.68	-0.83	-0.27	-0.11	-0.57	1.24	-0.83	-0.57
-0.14	-0.56	0.63	-0.38	-0.46	-0.15	1.36	1.39	-0.61	0.18	0.40	-0.78	-0.34	0.11	-0.40	1.37	-0.78	-0.80
-0.16	0.17	0.19	-0.31	-0.52	-0.36	1.29	1.50	-0.62	0.01	0.19	-0.87	0.77	0.50	-0.48	1.07	-0.87	-0.48
-0.39	2.75	-0.05	-0.40	-0.49	-0.09	-0.50	0.82	-0.50	-0.58	0.37	-0.46	0.20	0.29	-0.45	2.51	-0.46	-0.45
-0.48	4.83	-0.40	-0.38	-0.47	0.15	0.46	1.11	-0.47	0.71	0.16	0.53	-0.28	-0.45	-0.41	2.44	0.53	-0.41
-0.19	4.93	-0.13	-0.41	-0.64	-0.35	0.47	0.14	-0.53	1.55	-0.29	1.05	0.75	-0.69	-0.91	1.42	1.05	-0.91
-0.16	3.26	0.12	-0.53	-0.30	-0.72	0.44	0.75	-0.43	1.50	-0.46	0.26	0.73	-0.36	0.44	0.60	0.26	0.44

TABLE 15: SAMPLE OF 10 GABOR FEATURES (MAGNITUDE) OF ONE RECTUM IMAGE WITH FREQUENCY = 2.0

θ = 10	θ = 20	θ = 30	θ = 40	θ = 50	θ = 60	θ =70	θ = 80	θ = 90	θ = 100	θ = 110	θ = 120	θ = 130	θ = 140	θ =150	θ = 160	θ = 170	θ = 180
-0.17	0.89	0.12	-0.72	0.50	-0.73	0.54	1.37	-0.45	1.23	-0.54	0.11	-0.13	0.01	2.02	0.32	0.11	-0.17
-0.43	-0.36	0.20	-0.84	0.40	-0.54	-0.25	1.35	-0.83	1.24	-0.91	-0.56	1.31	0.52	1.65	-0.12	-0.56	-0.43
-0.55	0.02	0.68	-0.83	-0.27	-0.11	-0.57	1.24	-0.75	0.52	-0.56	-0.53	0.85	1.62	0.35	0.99	-0.53	-0.55
-0.61	0.18	0.40	-0.78	-0.34	0.11	-0.40	1.37	-0.74	-0.09	-0.40	-0.08	0.89	1.92	-0.22	0.88	-0.08	-0.61
-0.62	0.01	0.19	-0.87	0.77	0.50	-0.48	1.07	-0.79	0.78	-0.58	0.12	0.55	-0.72	-0.07	-0.52	0.12	-0.62
-0.50	-0.58	0.37	-0.46	0.20	0.29	-0.45	2.51	-0.80	-0.16	0.52	0.51	-0.24	-0.39	0.34	-0.44	0.51	-0.50
-0.47	0.71	0.16	0.53	-0.28	-0.45	-0.41	2.44	0.66	-0.45	0.34	1.67	-0.46	-0.24	0.73	1.23	1.67	-0.47
-0.53	1.55	-0.29	1.05	0.75	-0.69	-0.91	1.42	-0.46	-0.45	-0.06	1.45	-0.42	-0.24	0.45	0.89	1.45	-0.53
-0.43	1.50	-0.46	0.26	0.73	-0.36	0.44	0.60	-0.81	-0.51	-0.67	1.06	-0.39	-0.27	-0.22	0.27	1.06	-0.43

TABLE 16: SAMPLE OF 10 GABOR FEATURES (MAGNITUDE) OF ONE RECTUM IMAGE WITH FREQUENCY = 3.0

θ = 10	θ = 20	θ = 30	θ = 40	θ = 50	θ = 60	θ =70	θ = 80	θ = 90	θ = 100	θ = 110	θ = 120	θ = 130	θ = 140	θ =150	θ = 160	θ = 170	θ = 180
-0.67	1.48	-0.81	0.39	1.54	0.22	2.55	-0.27	-0.50	-0.53	-0.60	1.08	1.24	-0.20	-0.43	-0.19	1.08	-0.67
-0.83	1.24	-0.91	-0.56	1.31	0.52	1.65	-0.12	-0.32	-0.66	-0.64	0.62	0.95	0.37	-0.53	-0.43	0.62	-0.83
-0.75	0.52	-0.56	-0.53	0.85	1.62	0.35	0.99	-0.66	-0.18	-0.40	0.04	-0.45	-0.33	-0.66	-0.80	0.04	-0.75
-0.74	-0.09	-0.40	-0.08	0.89	1.92	-0.22	0.88	0.99	-0.24	-0.16	-0.12	-0.66	-0.24	-0.60	-0.88	-0.12	-0.74
-0.79	0.78	-0.58	0.12	0.55	-0.72	-0.07	-0.52	0.85	-0.30	0.23	-0.15	-0.40	-0.27	-0.35	0.10	-0.15	-0.79
-0.80	-0.16	0.52	0.51	-0.24	-0.39	0.34	-0.44	0.00	0.00	1.71	-0.54	0.00	-0.25	0.49	0.35	-0.54	-0.80
0.66	-0.45	0.34	1.67	-0.46	-0.24	0.73	1.23	-0.33	-0.32	2.77	-0.59	-0.19	-0.51	-0.64	0.40	-0.59	0.66
-0.46	-0.45	-0.06	1.45	-0.42	-0.24	0.45	0.89	-0.60	-0.66	2.20	0.00	-0.15	-0.66	-0.40	0.31	0.00	-0.46
-0.81	-0.51	-0.67	1.06	-0.39	-0.27	-0.22	0.27	-0.67	0.99	0.86	-0.33	0.21	-0.18	-0.16	-0.19	-0.33	-0.81