J. Basic. Appl. Sci. Res., 3(2s)243-251, 2013
© 2013, TextRoad Publication
ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Stomach Cancer Diagnosis by Using a Combination of Image Processing Algorithms, Local Binary Pattern Algorithm and Support Vector Machine

Danial Ahmadzadeh [1,*], Mohammad Fiuzy [2,3], Javad Haddadnia [2,3]

[1] Department of Computer Engineering, Science and Research branch, Islamic Azad University, Khorasan Razavi, Neyshabur, Iran
[2] Biomedical Engineering Department, Electrical and Computer Faculty, Hakim Sabzevari University, Sabzevar, Iran
[3] Research Center for Advanced Medical Technologies, Sabzevar University of Medical Sciences, Sabzevar, Iran

Received: June 10 2013
Accepted: July 10 2013

ABSTRACT

Although the amount of stomach cancer has reduced obviously during last decades in western countries, but this illness is still one of the main causes of death in developing countries. In Iran , stomach cancer is one of the most common illness in some areas like northwest and northeast . In this study we aim to suggest a new way for diagnosing this illness. One of the main problems with this illness is that the diagnosis process doesn't take place at the right time. Nowadays doctors try to diagnose this illness on the basis of their experiences, knowledge, and complicated surveys but all humans have error in their works. Data in this study are selected from 55 subjects (selected randomly). By using image processing and artificial intelligence algorithms like primary preprocessors for increasing the quality and statistical characteristics of image, local binary pattern algorithm for elicit characteristics, image histogram algorithm to elicit impaired characteristics and support vector machine has been used for accurate classification among impaired and suspected subjects and also for accurate diagnosis of impairment. The suggested system by using a combination of mentioned methods was succeed to achieve 91.8% accurate diagnosis. Although available methods are accurate but they are really expensive and time consuming, by comparing this method with those mentioned above are will have a better understanding of its accuracy and usefulness.

KEYWORDS: Patient, Stomach Cancer, Artifical Intelligent Algorithms, Local Binary Pattern, Support Vector Machine.

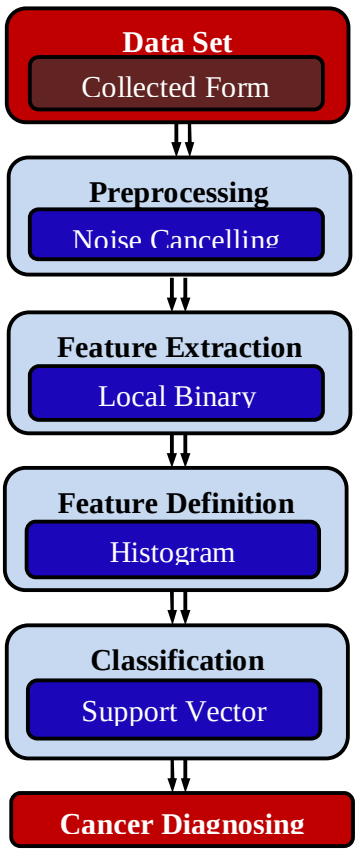
1.INTRODUCTION

Cancer is one of the most important concern in peoples life. As having the highest death rate after heart diseases[1]. In addition to its high death rate, it has economic-psychological and impairment consequences which cause the delay in diagnosis and accurate treat[2]. In accordance to the statistics presented by The World Health Organization, stomach cancer is the second prevalent cancer in the world. As 50% of this cancer are inoperable and with available treatments there is no chance for more than 5 years life for the sufferers. The prevalent of this impairment is unknown, but conducted studies in Japan, China, South America and East Europe shows that this impairment is very common, as in Japan 20% to 30% of cancers are stomach cancer, and 22000 new cases are found in United States every year[1,3].

In accordance to available reports, stomach cancer is more prevalent in northwest and northeast of Iran. Up to dated and standard statistics of this cancer in ratio to the world population is 26.1 in men and 11.1 in women[4]. Most of the sufferers visit doctors in complicated stages but even those who are aware sooner and are under suitable treatments usually die because of recurrence. It is believed biologically that Carcinogen tumor nowadays is exploring pathology from image endoscopy was done by knowledgeable doctors[2]. Although they were not successful in their project this process is really slow and can be affected by bored, and environmental factors, so they can cause error and are regularly mental and qualitative[5]. So there are many calculating techniques for image processing which are able to identify patterns, edges, clusters and…. which can help doctors to diagnose the impairment. In recent years there is a main attention to data processing methods in science. It is proven theoretically that one group of classifieds present more accurate results. [6-12]. We can see that each class of data and method has its own strength and weakness which may do well in one system and not in the other [13]. For this to achieve a better result, a mixture of developing intelligent methods is preferred. Till now many methods in a

***Corresponding Author:** Danial Ahmadzadeh, Department of Computer Engineering, Science and Research branch, Islamic Azad University, Khorasan Razavi, Neyshabur, Iran.
E-Mail: d.ahmadzadeh@srbiau.ac.ir ; Mobile: 0098-9153510373

number of studies has been used by adapting nerve systems, statistical studies and image processing, which are considered in references [14,15].In this study first the overall method will be shown in flowchart (1) then pattern recognition will be considered as a principle after that materials and used methods, then evaluation and result will be present and finally we will present discussion and conclusion. We hope this study help to improve the methods in diagnosis of this impairment.



Flowchart 1. Suggested process to diagnose stomach cancer

2. Characteristics Selection and Pattern Recognition

Characteristics selection is to select those characteristics which have the maximum prediction ability in outcome [16].To identify what is an optimize subset depends on the problem we want to solve [17].Characteristics selection algorithms can be divided to two class in accordance to their evaluation process. If characteristics selection is done independence of all learning algorithm (like an independent preprocessing) that is called filter, if evaluation process is in a relation with classifying algorithm that will be called wrapper. Second class usually causes better results. The pattern recognition system has four parts: eliciting characteristics, characteristics selection, design and teaching classifier and finally experiment. In this study we will use support vector machine for experiment and classifying data, local binary algorithm for eliciting characteristics and histogram statistical technique for identifying and selection of characteristic. The main reason for using histogram is that it shows the amount of characteristics of each sample in accordance to the frequency of that characteristic [18]. To say in another word it is a specific method for the selection of characteristic in each sample.

3. Stomach Cancer

Cancer is one of the most dangerous illnesses in the modern world. In the United States, in each four death one is of cancer [19]. It is estimated that in Europe in 2006 there were 3.2 million people who suffer from cancer and 1.7 million of them died. Until 2004the number of new sufferers reached 300000 people and by aging the Europe population this number will be increased in next decades [20]. As more than 50% of this cancer cannot be operated with available treatments there is no chance for more than 5 years

life for the sufferers. World prevalence of this malignity is unknown, but in the conducted studies in Japan, China, South America and East Europe has more prevalence , as in Japan 20% to 30% of cancers are stomach cancer, and 22000new cases are found in United States every year.[21,22]. . In accordance to available reports, stomach cancer is more prevalent in northwest and northeast of Iran. Up to dated and standard statistics of this cancer in ratio to the world population is 26.1 in men and 11.1 in women in each 100000 person [23]. Today 95% of digestion system cancers are caused by Adenoma. Early Adenoma diagnosis can prevent the development of cancer. In this study we suggest an intelligent system for the diagnosis of Adenoma in stomach by the use of endoscopy images.

4. Database

For collecting data we used endoscopy images kept in digestion subspecialty center by Dr. Shahram Bayat in Neyshabur. The mentioned bank consists of 6 to 8 images from different parts of stomach by the size of 240 in 352 pixels from each person. Below we will consider a number of these samples. For study, presenting, and implementation of the suggested method we use the image of fifth part of 55 suspected subjects. The images of this experiment in figure 1 is belong to the subject number 25, is shown below.

Fig. 1. A selection of 6 images used(subject 25), from the upper line from left,we worked on 5th image

5. Suggested Method

As was mentioned in introduction the purpose of this study is to present an intelligent mean for diagnosing stomach cancer. This method has different stages that we will discuss them. In the first stage we will present the preprocessing operation then local binary pattern for characteristic elicitation after that histogram for selection and introducing the characteristic and finally for classifying data and the result of the previous stages we will present the support vector machine.

5.1. Preprocessing

First the images from data base will be gathered then by the use of image processing techniques (differentiating and median) the noises will be deleted. A search motor in science's images must be sensitive to the changes within the classes. Implementing suitable preprocessing techniques can reduce these changes [24]. The primary images usually include noises which sometimes make the image worse or even delete it (figure2). In this picture we add some salt and pepper noise to the image to delete these noises after the implementation of undesirable signals.

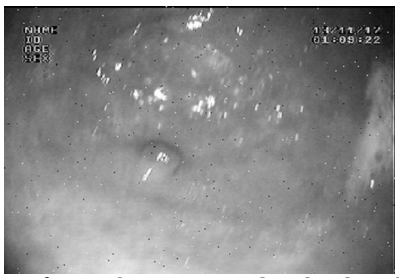


Fig. 2. 5th image from subject 25 mixed with salt and pepper noise

For this we use median nonlinear filter to omit stroke noises[25,26].

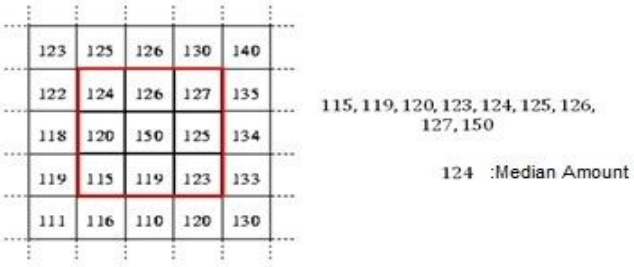


Fig. 3. Median filter

Some of these noises may affect obviously in image processing techniques like edge detection, segmentation, data compression and identification of a specific part of the image [27]. This filter by masking start to calibrate all pixels and the base of this work is to use the neighborhood pixels. In another word all neighborhood pixels will be arranged, and select the middle element of arranged numbers and replace it with the central pixel. We implement this filter on figure 2 and you can see the result in figure 4.

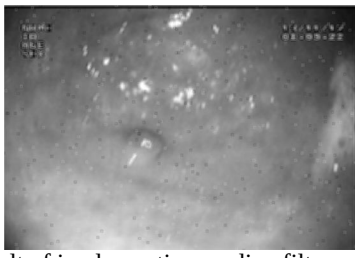


Fig. 4. The result of implementing median filter on input image 2

If you look at figure 4 carefully, you can see that it is more opaque than the original one, and this is because of using the median filter. As the median filter is a function of sum and total family then subtraction is the diverse of sum, and will have the diverse function of median filter. For applying subtraction filter we use derivational operator in discrete space. And by using equation number 1, 2, 3 implement subtraction filter.

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (1)$$

$$\nabla^2 f = [f(x+1,y) + f(x-1,y) + f(x,y-1) + f(x,y-1)] - 4f(x,y) \quad (2)$$

With the equation 2 the bellowing filter will be designed that can be shown in a 3*3 mask. These equations can be implemented by a filter in a picture-result from isotropic in 90 mutations-according to the filters shown, in both cases the total point minus subtraction points equals to $-8f(x)$ equation 3.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (3)$$

This is called sharpening filter. In figure 5 you can see the effect of Laplasin filter very well. If we compare this picture with figure 4 you can see a better clarity and also there are no more opaque points.

Fig. 5. The result of implementing Laplace filter on image 4 with observable mounds

5.2. Eliciting Characteristic by Local Binary Pattern

One of the methods used for classifying image tissue is local binary pattern [28].LBP was presented by Ojala and his assistants [29] as a two leveled version from tissue spectrum method [30]which is used in 3 levels(2,1,0) in local tissue pattern. Local binary pattern uses 28=256 possible tissue units instead of 38=6561 units used in spectrum method, which will cause better representation of tissue and lead the discrimination of the comparable tissue [31]. Local binary pattern in 3*3 neighborhood pixel is shown below, main neighborhood of 3*3 (figure 6) is limited in 2 levels (1, 0) this process uses central pixel. The amount of pixels in the neighborhood of the limited part is multiplied by special weights related to the specific pixels. The amount of eight pixels for calculating a united quantity will be added for the related pattern. LBP characteristics vectors by LBP distribution histogram in a part of a picture make an image.

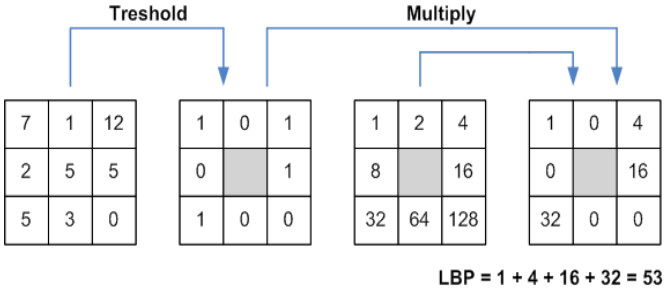


Fig. 6. LBP algorithm process

By applying local binary pattern algorithm on figure 5, figure 7 will be the result.

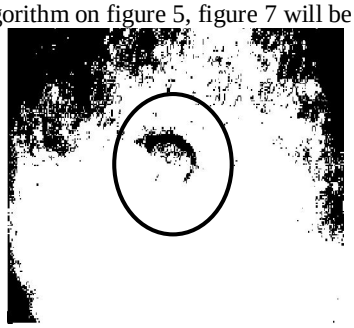


Fig. 7. The result of implementing LBP algorithm on image 5

5.3. Introducing Characteristics by Histogram Algorithm

Histogram is the statistical representation of data in accordance with their frequencies. In this diagram the frequency of data are shown in the term of their numbers. Data in scientific images are light levels of image like figure 6 after applying histogram algorithm on data in figure 5, the result are shown in figure 7. After applying histogram algorithm on an image affected by LBP algorithm the result is figure 8.

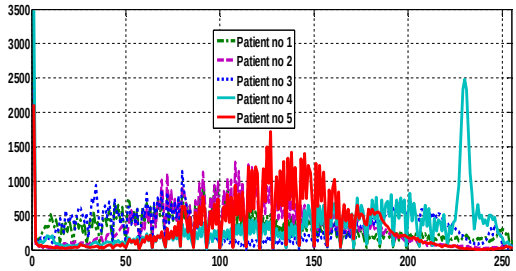


Fig. 8. Histogram of 5 patient

After implementing histogram algorithm on LBP the result is figure 9. It is obvious that data from histogram became less before implementing LBP algorithm.

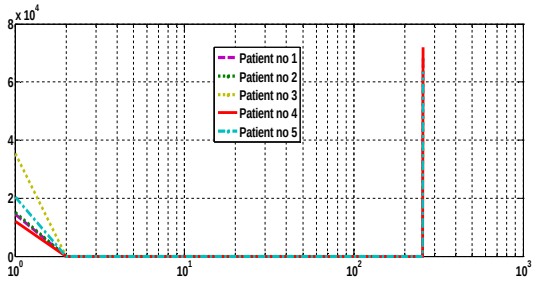


Fig. 9. Implementing histogram on images from LBP

Then in this order all suspected images after extraction will be affected by LBP algorithm and histogram in respect, then with the result of pathology go through support vector machine algorithm to classify impairments into suspect and cancerous samples.

5.4. Support Vector Machine Algorithm

Different classification algorithms are suggested for realization of scientific intelligent projects, like line detachment analysis(LDA)[32, 33], nerve systems [34] and SVM[35]. The last sample is a binary classifier which can create a strong operation even from spread and noisy data. Those will resist against proportionality of taught data, and this classification is not affected by the size rate of the samples [36].In addition to this SVMs are less experimental in comparison with standard nerve systems and proved that they are more accurate than other classifiers like: tissue classification, [37] and CT patterns [38].Support vector machine is a binary classifier which separate 2 classes with a line. In this method by using all bands and an optimization algorithm all samples that shows the classes' border are identified. These samples are support vectors. Some of the taught points which have the least distance to the border can be used as a subset for identifying borders and also as a support vector. In figure 10 two classes and related support vectors are shown. Think that data are made of two classes and classes have xi,i=l,…,L taught points and xi is a vector. These two classes are identified by yi= +-1, for identifying the border between these two classes we use the optimization method [39].In this method the border is calculated in a way that:

1. All samples in class +1 be in one side and class -1 samples in the other side.
2. The border must be in a way that: the distance of the nearest taught samples is the most in both classes in perpendicular line .

The SVM algorithm is shown in figure 10.

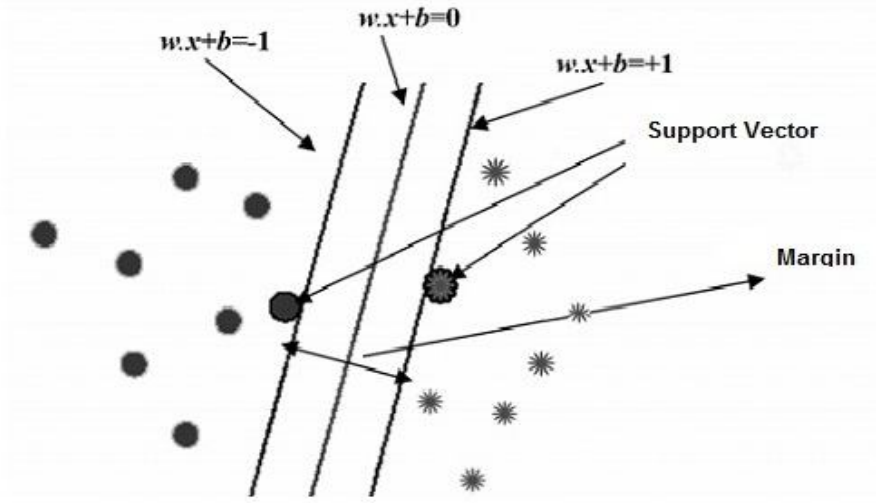


Fig. 10. SVM algorithm's in classifying 2 class of data

6. The Accuracy of the Final Evaluation Result

Overall the process which causes the diagnosis of impairment is: first by 5 class of 50 data in a way that each part has all kind of subjects. In this kind of experiment, 9 parts of data are used for teaching and

for testing. This process of teaching and testing will be done 10 times in a way that in each time one different part is put aside for testing.

7. Result of Evaluation in Suggested Method

For applying the suggested method MATLAB software was used. All suggested processes was in this order: first images with biological result were elicited from impairments then after all preprocessing stages for omitting noise. LBP was administered on all images. Then after applying histogram LBP images with aimed vector went through support vector machine for classification.SVM uses different methods for separating data like: RBF,MLP,Linear,Gausian,…. In this study for classifying impaired ones from healthy persons we used RBF, MLP and Linear. The operation of SVM algorithm in 2 dimensions of data,Linear and RBF, is shown in figure 11,12 .

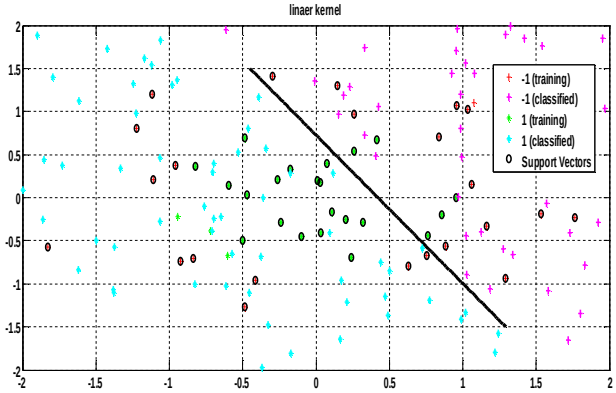


Fig. 11. Separating and classifying data on the linear structure base

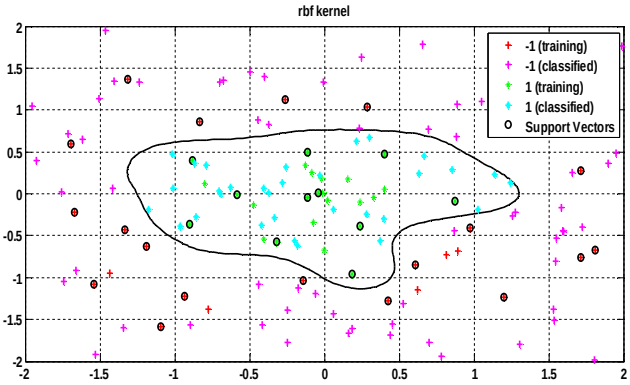


Fig. 12. Separating and classifying data on the base of RBF structure

As was mentioned before 9 parts of data were used for teaching and 1 for testing, this was done 10 times to achieve the best result. Testing data are given to the system to diagnose the impairment. For checking system operation we used accuracy parameter, which is shown in equation 4.

$$Total\ \ Accuracy = \frac{Total\ True}{Total\ input} = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \qquad (4)$$

This relation is a function of 42 parameter which is mentioned below:
Tp: Is the number positive elements that are classified correctly
Tn: Is the number negative elements that are classified correctly
Fp: Is the number positive elements that are classified incorrectly
Fn: Is the number negative elements that are classified incorrectly
Among these 55 subjects who have been as testing and teaching subjects, the system succeed to achieve an accuracy of 91.8% by using RBF.

8. Conclusion

In this study we were looking for a way to decrease the calculating process and increase the accuracy of the diagnosis process of stomach cancer, for these we used SVM algorithm. In table 1 there is a comparison among this method and those that presented before.

Table 1. Accuracy comparison among intelligent algorithms

Best accuracy percent in experiment data	Method
76.89%	Multi Layer Perceptron (MLP)
79.69%	Adaptive Neuro Fuzzy (ANFIS) Inference System
85.14%	Linear SVM algorithm(in this study)
89.25%	SVM algorithm by MLP structure(in this study)
91.8%	SVM algorithm by RBF structure(in this study)

9. Acknowledgement

Here I have to give my special thanks to dear Dr. Shahram Bayat for his support and helpful tips.

REFERENCES

[1] Fuchs CS, Mayer RJ. Gastric carcinoma. N Engl J Med 1995; 333(1): 32-41.
[2] Breaux JR, Bringaze W, Chappuis C, Cohn L. Jadenocarcinoma of the stomach: A review of 35 years and 1,710 cases. World J Surg 1990; 14: 580-586.
[3] Abeloff MD. Clinical Oncology. 2nd ed. New York, Churchill Livingstone, 2000; PP: 1545-1579.
[4] Sadjadi A, Nouraie M, Mohagheghi MA, Mousavi-Jarrahi A, Malekezadeh R, et al. Cancer Occurrence in Iran in 2002, an International Perspective. Asian Pacific J Cancer Prev 2005; 6: 359-363.
[5] Rangayyan, R. M. (2005). Biomedical Image Analysis. Canada, CRC Press.
[6] L.I. Kuncheva; J.C. Bezdek; R.P.W. Duin, "*Decision templates for multiple classifier fusion: an experimental comparison*", Pattern Recognition, 34(2): 299–314, 2001.
[7] Bi. Yaxin; Bell. David; Wang. Hui; Guo. Gongde; Guan. Jiwen, "*Combining Multiple Classifiers Using Dempster's rule for text Caregorization*", Applied Artificial Intelligence, 21:3, 211- 239, 2007.
[8] L.I. Kuncheva, *Combining Pattern Classifiers, Methods and Algorithms*,New York, NY: Wiley Interscience, 2005.
[9] S. Tulyakov; S. Jaeger; V. Govindaraju; D. Doermann, "*Review of Classifier combination Methods*", Studies in Computational Inteligence(SCI) 90,361-386 , 2008.
[10] M. Sugeno, "Industrial Applications of Fuzzy Control", Elsevier,Book, New York, 1985.
[11] F. Bergh and A. Engelbrecht, "A new locallyconvergent particle swarm optimizer", IEEE, 2002,Fundamenta Informatic Journal
[12] J.Kittler; M.Hatef; R.Duin P. W; J.Matas, "*On Combining Classifiers*", IEEE Transactions on Pattern Analysis and Machine Intel ligence, 20(3):226–239, 1998.
[13] S. Tulyakov; S. Jaeger; V. Govindaraju; D. Doermann, "*Review of Classifier combination Methods*", Studies in Computational Inteligence(SCI) 90,361-386 , 2008.
[14] Baopu Li, Max Q.-H, Meng, " Automatic polyp detection for wireless capsule endoscopy images", Expert Systems with Applications 39 (2012) 10952–10958.
[15] Dimitris K. Iakovidis, Dimitris E. Maroulis, Stavros A. Karkanis, " An intelligent system for automatic detection of gastrointestinal adenomas in video endoscopy", Computers in Biology and Medicine, 38 (2011), 150-164
[16] Zhai L, Dong S, and Ma H," *Recent Methods and Applications on Image Edge Detection*", In Proceedings of the 2008 international Workshop on Education Technology and Training & 2008 international Workshop on Geoscience and Remote Sensing - Volume 01 (December 21–22, 2008). ETTANDGRS. IEEE Computer Society, Washington,332335.

[17] A.Baraldi, and P.Blonda, "A Survey of Fuzzy Clustering Algorithms for Pattern Recognition—Part I and II", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 29, NO. 6, 1999.
[18] M.R.NaemAbadi, N.Amir Ahmadi Chamachar, E.Tahami, H.Rabani, "Diabete diagnosis with SVM algorithm",14th Iran electronic engineering conference, 2011.
[19] Jemal, A., R. Siegel, et al. (2008). "Cancer Statistics, 2008." CA Cancer J Clin 58(2): 71-96.
[20] Gouveia, J., M. P. Coleman, et al. (2008). "Improving cancer control in the European Union: Conclusions from the Lisbon round -table under the Portuguese EU Presidency, 2007." European Journal of Cancer 44(10): 1457-1462.
[21] Fuchs CS, Mayer RJ. Gastric carcinoma. N Engl J Med 1995; 333(1): 32-41.
[22] Abeloff MD. Clinical Oncology. 2nd ed. New York, Churchill Livingstone, 2000; PP: 1545-1579.
[23] Abeloff MD. Clinical Oncology. 3nd ed. New York, Churchill Livingstone, 2000; PP: 1545-1579.
[24] Linda G. Shapiro and George C. Stockman (2001): "Computer Vision", pp 279-325, New Jersey, Prentice-Hall,ISBN,0-13-030796-3
[25] "*Computer Vision*", pp 279-325, New Jersey, Prentice- Hall,ISBN,0-13-030796-3
[26] Zhai L, Dong S, and Ma H," *Recent Methods and Applications on Image Edge Detection*", In Proceedings of the 2008 international Workshop on Education Technology and Training & 2008 international Workshop on Geoscience and Remote Sensing - Volume 01 (December 21–22, 2008). ETTANDGRS. IEEE Computer Society, Washington,332335.
[27] A.Baraldi, and P.Blonda, "A Survey of Fuzzy Clustering Algorithms for Pattern Recognition—Part I and II", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 29, NO. 6, 1999.
[28] T. Ojala, M. Pietik¨ainen, and T. M¨aenp¨a¨a, "Multiresolution gray scale and rotation invariant texture analysis with local binary patterns," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, Jul. 2002.
[29] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with lassification based on feature distributions, Pattern Recognition 29 (1996) 51–99.
[30] D.C. He, L. Wang, Texture unit, texture spectrum, and texture analysis, IEEE Trans. Geosci. Remote Sensing 28 (1990), 509–512.
[31] M. Pietikäinen, T. Ojala, Nonparametric texture analysis with complementary spatial operator, in: M. Pietikäinen (Ed.), Texture Analysis in Machine Vision, Series in Machine Perception and artificial Intelligence, vol. 40, World Scientific, Singapore, 2000, pp. 3–18.
[32] S.A. Karkanis, D.K. Iakovidis, D.E. Maroulis, D.A. Karras, M. Tzivras, Computer aided tumor detection in endoscopic video using color wavelet features, IEEE Trans. Inform. Technol. Biomed. 7 (2003) 141–152.
[33] D. West, V. West, Model selection for a medical diagnostic decision support system: a breast cancer detection case, Artif. Intell. Med. 20 (2000) 183–204.
[34] K. Papik, B. Molnal, R. Schaefer, Z. Dombovari, Z. Tulassay, J. Feher, Application of neural networks in medicine—a review, Diagnostics Med. Technol. 4 (3) (1998) 538–546.
[35] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[36] I. El-Naqa, Y. Yongyi,M.N.Wernick, N.P. Galatsanos, R.M. Nishikawa, A support vector machine approach for detection of microcalcifications, IEEE Trans. Med. Imaging 21 (12) (2002) 1552–1563.
[37] S. Li, J.T. Kwok, H. Zhu, Y.Wand, Texture classification using the support vector machines, Pattern Recognition 36 (2003), 2883–2893.
[38] S.B. Gokturk, C. Tomasi, B. Acar, C.F. Beaulieu, D.S. Paik, R.B. Jeffrey Jr., J. Yee, S. Napel, A statistical 3-D pattern processing method for computer-aided detection of polyps in CT colonography, IEEE Trans.Med. Imaging 20 (12) (2001), 1251–1260.
[39] V. Vapnik and A. Chervonekis,"The necessary and sufficient conditions for consistency in the empirical risk minimization method," Pattern Recognition and Image Analysis, vol. 1, no. 3, pp. 283-305, 1991.