International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163
Issue 10, Volume 6 (October 2019) www.ijirae.com

A VARIOUS CANCER DETECTION TECHNIQUES : A STUDY

Parvathy Unnikrishnan[1], student
Department of CSE, ASIET, Kalady, Kerala.
Parututtu1@gmail.com

Asst.Prof.M.Gokilavani [2]
Department of CSE, ASIET, Kalady, Kerala.
gokilavani.cs@adishankara.ac.in

Manuscript History
Number: **IJIRAE/RS/Vol.06/Issue10/OCAE10083**
Received: 11, October 2019
Final Correction: 16, October 2019
Final Accepted: 22, October 2019
Published: **October 2019**
Citation: Unnikrishnan & Gokilavani (2019). A VARIOUS CANCER DETECTION TECHNIQUES: A STUDY. IJIRAE::International Journal of Innovative Research in Advanced Engineering, Volume VI, 616-621.
doi://10.26562/IJIRAE.2019.OCAE10083
Editor: Dr.A.Arul L.S, Chief Editor, IJIRAE, AM Publications, India

Abstract—Cancer is strong and abnormal growth of cells within the any parts of the body. Advance research in advanced countries talks that main reason of death of people having tumor is incorrect detection of tumor in various parts like brain, blood and liver. It is one of the most dangerous diseases and therefore it should be detected quickly and accurately. This can be done by using automatic tumor detection techniques on medical images. Generally MRI or CT scan that is directed into intracranial cavity produces the complete image of the tumor. Magnetic Resonance Imaging (MRI), a highly developed technique of medical imaging, is used to visualize internal structure of human body without any surgery. For the accurate detection of brain tumor, segmentation of MRI image is important. Classification of tumor, through segmented MR image, into normal and abnormal MRI brain images, is a difficult task due to complexity and alteration in tumor tissue characteristics like its location, size, gray level intensities and shape. In this paper, review of various techniques of automatic detection of tumor in various parts of body using Magnetic Resonance Image (MRI) is proposed.

Keywords— Tumor; Magnetic Resonance Image; detection; classification; features extraction;

I. INTRODUCTION

Accurate diagnosis for different types of cancer plays an important role in determining and choosing the proper treatment to the doctors to assist them. By using classification techniques, possible errors that might occur due to unskilled doctors can be minimized. Challenge facing medical practitioners makes this study of a much greater significance. Since symptoms appear only in the advanced stages thereby causing the mortality rate of lung cancer to be the highest among all other types of cancer, challenging the detection of cancer in its early stages [1]. The objective of undertaking this project is to facilitate doctors to provide the best possible treatment by providing useful insights with the help of predictive models through analysis and diagnosis of lung cancer treatments. This technique can also examine medical data in a shorter time and more precisely. The critical task is to define and specify a good feature space that means the type of features which will discriminate between malignant and benign. The Support Vector Machine (SVM) was first proposed by Vapnik and has since attracted a high degree of interest in the machine learning research community [2]. Several recent studies have reported that the SVM (support vector machines) generally are capable of delivering higher performance in terms of classification accuracy than the other data classification algorithms. Sims have been employed in a wide range of real world problems such as text categorization, hand-written digit recognition, tone recognition, image classification and object detection, micro-array gene expression data analysis, data classification. It has been shown that Sims is consistently superior to other supervised learning methods. However, for some datasets, the performance of SVM is very sensitive to how the cost parameter and kernel parameters are set.

As a result, the user normally needs to conduct extensive cross validation in order to figure out the optimal parameter setting. This process is commonly referred to as model selection. One practical issue with model selection is that this process is very time consuming. We have experimented with a number of parameters associated with the use of the SVM algorithm that can impact the results. These parameters include choice of kernel functions, the standard deviation of the Gaussian kernel, relative weights associated with slack variables to account for the non-uniform distribution of labelled data, and the number of training examples. For example, we have taken four different applications data set such as diabetes data, heart data and satellite data which all have different features, classes, number of training data and different number of testing data. This paper is organized as follows. In next section, we introduce some related background including some basic concepts of SVM, kernel function selection, and model selection (parameters selection) of SVM.

II. MEDICAL IMAGE PROCESSING

A. Algorithm used in medical image processing projects:

Medical Image Processing Projects are developed based on image processing simulation tool named as Matlab. Using the tool processing more medical images of human organs are (Brain, Lung, Kidney, Skin, Retina, Finger, Tissues and Skull). According to the modality results the physician can easily observe the pathologies directly but sometimes it took more time to analysing. Image analysis process can be automated for producing interesting results about human diseases. In medical image processing projects we have to use more algorithms to identify and classify the diseases in the images. Segmentation and classification methods are used to detect the disease and known the status of the human.

Some of the commonly used classification algorithms are

- Support Vector Machine
- Fuzzy C-Means Clustering.
- K-NN Classification.
- Naive Bayes Classification.
- Decision Trees.
- Genetic Algorithm.
- Neural network Classification.

Before enter into the process of classification we must do the processes of pre-processing, feature extraction and feature reduction. Pre-processing based on some gray scale conversion methods, noise removal concepts. Feature extraction is the process of extracting features in the images with its pixels. Most commonly extracted features are color, shape, texture, geometric features. In medical image processing greatly define the texture features. Then extracted features are to be reduced using feature reduction or selection method. The particular selected features are used to identify the disease in the human organ. Requirements for medical image processing are image enhancement, changing density range of B/W images, manipulating colors, image line profile display, image restoration, image smoothing, Biomedical image area calculation, detection of contour. Applying the algorithms of filtering and reconstruction automatically identify the 3D image data, it must take more number of object slices for processing.

B. Image Processing Technique:

Image processing is a method to convert an image into digital form and perform some operations on it in order to get an enhanced image or to extract some useful information from it.Image processing basically includes the following three steps.

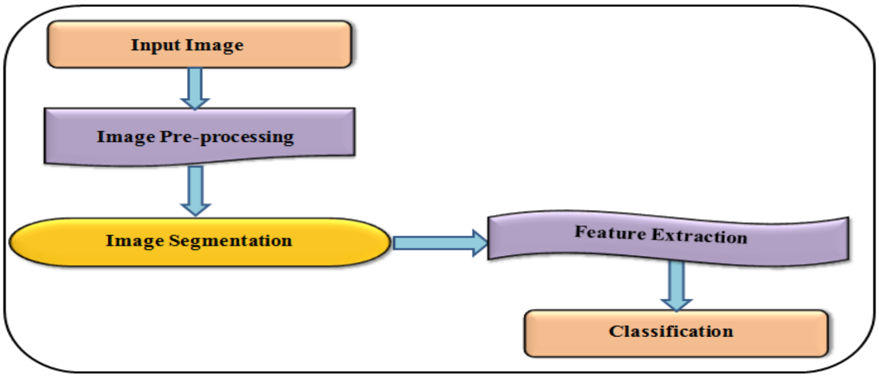

Fig 1: Image processing technique

- Importing the image with optical scanner or by digital photography.
- Analyzing and manipulating the image that includes data compression and image enhancement and spotting patterns.
- The last is the output in which result can be altered image.

III. SVM CLASSIFIER

Next phase in the proposed system is the classification of occurrence and non-occurrence of cancer nodule for the supplied lung image. The classifier used is Support Vector Machine. The aim of classification is to group items that have similar feature values into groups.

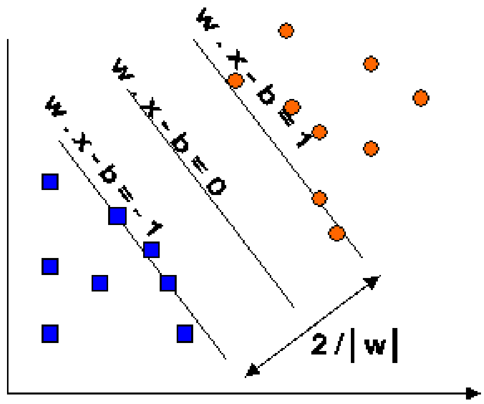

Fig 2: Maximum margin hyper planes for a SVM trained with samples from two classes

SVMs are set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classification. A special property of SVM is, SVM simultaneously minimize the empirical classification error and maximize the geometric margin. So SVM called Maximum Margin Classifiers. SVM is based on the Structural risk Minimization (SRM). SVM map input vector to a higher dimensional space where a maximal separating hyper plane is constructed. Two parallel hyper planes are constructed on each side of the hyper plane that separate the data. The separating hyper plane is the hyper plane that maximizes the distance between the two parallel hyper planes. An assumption is made that the larger the margin or distance between these parallel hyper planes the better the generalization error of the classifier will be [1].Classifier achieves this by making a classification decision based on the value of the linear combination of the features. SVM is a binary classification method that takes as input labeled data from two classes and outputs a model file for classifying new unlabeled/labeled data into one of two classes.

Training an SVM involves feeding known data to the SVM along with previously known decision values, thus forming a finite training set. It is from the training set that an SVM gets its intelligence to classify unknown data. In SVM, for two class classification problem, input data is mapped into higher dimensional space using RBF kernel. Then a hyper plane linear classifier is applied in this transformed space utilizing those patterns vectors that are closest to the decision boundary, shown in fig-2.

Consider the pattern classifier, which uses a hyper plane to separate two classes of patterns based on given examples {x (i), y (i)}, i=....n .Where (i) is a vector in the input space I=Rk and y (i) denotes the class index taking value 1 or 0. A support vector machine is a machine learning method that classifies binary classes by finding and using a class boundary the hyper plane maximizing the margin in the given training data. The training data samples along the hyper planes near the class boundary are called support vectors, and the margin is the distance between the support vectors and the class boundary hyper planes. The SVM are based on the concept of decision planes that define decision boundaries.

A. Training the classifier

In the training phase, known data is given and the classifier is trained. Given training data (xi, yi) for i = 1 ...N, with xi ∈Rd and yi∈ {−1, 1} .The training points satisfy the following conditions.

$$F(x) = W^T x_i + b \geq +1 \text{ for } y_i = +1$$

$$F(x) = W^T x_i + b \leq -1 \text{ for } y_i = -1$$

B. Testing the data

In testing phase, unknown data are given and the classification is performed using trained classifier. Classification is done by using following decision function.

$$F\ (x, \{w, b\}) = \text{sign}\ (w.x+b)$$

Every input x is initially mapped into a higher dimension feature space F, by z=φ(x) through a nonlinear mapping φ: Rn→F. W is the normal to the line, x is the feature vector and b the bias. W is known as the weight vector and b is bias.

IV. RELATED WORK

Tumor is an uncontrollable and abnormal growth of cells in the any parts of body. Tumors are of two types-primary or benign tumors and metastatic or malignant tumors.

A primary tumor starts and spread only in the body. Metastatic tumors can initiate somewhere in the body as cancer and extended. Various methods, which are available in diagnosis, are expert opinion, human inspection, biopsy, and etc. These methods have some drawbacks like time consumption, incorrect inspection etc.

So image processing techniques can be helpful to detect tumor. There are various medical imaging techniques like x-ray, computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), are available for tumor detection. The MRI is the most commonly used modality for tumor growth imaging and location detection due to its higher resolution. Magnetic Resonance Imaging (MRI) is an imaging technique which non-invasively provides high contrast images of different anatomical structures. It provides better differentiation of tissues than other medical imaging techniques. Evaluation and analysis of MRI images by radiologists is error-prone and time consuming. Hence radiologists can use an algorithmic image processing in tumor diagnosis in MR images, especially due to large alterations in shape and size of structures needs to be considered for tumor detection and segmentation. Therefore automatic analysis and classification of such medical images is essential. In recent years, various methods have been proposed for image segmentation and classification techniques of tumors.

A. SVM based Lung Cancer diagnosis using multiple image features in PET/CT:

The details of patient characterization are summarized in table below:

TABLE I. PATIENT CHARACTERISTICS

Variable	Level	Frequency	Percentage (%)
Gender	Male	19	54
	Female	16	46
Tumor Subtype	Adenocarcinoma	18	51
	Pleomorphic carcinoma	1	4
	Bronchioloalveolar	1	4
	SCC	3	8
	NSCLC	2	5
Benign		10	28

Once the image features are extracted and computed, train the support vector machine, a statistical machine tool, to merge the features and learn the properties of the data structure. The trained SVM [3] is then applied for classifying and identifying the benign and early malignant as well as early and advanced malignant in the testing datasets. The **leave-one-out crosses validation** to evaluate the performance of the trained SVM. Individual image features are applied to the classification tasks and then evaluate their performances using **statistical analysis**, the **results** are shown in figure below. It demonstrated that the heterogeneity derived from F-FLT images significantly differentiated benign (0.24 ± 0.09, N=9) from early stage malignancy (0.40 ± 0.09, N=10; P = 0.002), as well as early stage from advanced stage malignancy (0.50 f 0.07, N=13, P = 0.005). Other image features, SUVmean and CT texture, didn't demonstrated similar capability. The method uses the SVM as classification tool and the receiver operation characteristics (ROC) as a figure of merit to characterize the performance. The receiver operation characteristics **(ROC) analysis results** were shown in the below figure. The detection rate of single feature (heterogeneity), two features (heterogeneity with SUVmean), and all three features were 0.304, 0.17 and 0.044, respectively. It clearly demonstrated although heterogeneity is the best individual image feature for the classification, additional information from other image features can help to improve the performance, particularly when these image features provide structure or physiological information of different aspects of tumor. It also shows the best performance was achieved when all three features are combined in the SVM training. In the future, increase the size of test population and evaluate the performance of the proposed method using statistically more accurate methods, such as 10 folder cross validation. They would also like to investigate whether including more image features, such as SUVmax and tumor volume, would further improve the performance of lung cancer staging. SVM analysis and classification with combination of tumor heterogeneity and other effective features has great potential to augment diagnostic accuracy and improve tumor staging in oncological practice.

B. Chest DR Image Classification based on support vector machine:

In order to get the decision-making function, SVM classifier [4] was applied to study on training set of chest DR images. The goal of two classification SVM algorithm is to build hyper-plane to separate the two types of data and ensure the classification of two types having the largest gaps. Through the Lagrange method and the method of quadratic solver of nonlinear optimization find a hyper plane. Key of SVM is the kernel function. As long as the choices of appropriate kernel function, we can get the classification function of high-dimensional space. Support vector machine is constituted to the training sample set and the kernel function. Commonly used kernel functions are mainly polynomial kernel function, radial basis function, and sigmoid function etc.

Support Vector Machine using structural risk minimization to replace the traditional empirical risk minimization, has a very good classification ability and promotional in the small sample, nonlinear and high-dimensional space. So support vector machine is a good selection in the domain of medical image classification. In the SVM classifier, the judgment ability is determined by the values of the penalty factor C and the parameter values of the SVM kernel function to a large extent.

Tab.1 Classification accuracy of different parameters

C / σ	1	10	100	1000	10000
0.25	0.90	0.89	0.89	0.90	0.89
0.5	0.90	0.91	0.92	0.93	0.92
1	0.90	0.90	0.91	0.92	0.91
10	0.90	0.89	0.90	0.92	0.91

C is the penalty factor.SVM is based on structural risk minimization principle, it has high recognition rate and robustness and good generalization ability in solving two-classification and recognition of small samples. SVM method can be used as a tool for early diagnosis and help doctors improving the recognition accuracy rate under conditions of limited medical clinical cases, and it has a good application value.

C. The anisotropic Gaussian Kernel for SVM classification of HRCT images of the lungs:

SVM algorithm [5] is used here for learning lung pattern classification. One of the main issues to resolve when applying the SVM is model selection. This includes tuning the parameters of the algorithm. In this paper tuning is performed by optimizing the radius/margin bound, which is an analytical bound of the accuracy of the SVM algorithm. The method of statistical learning based on the SVM algorithm is chosen to tackle the problem of pixel classification. The SVM algorithm has a geometrical interpretation by which the training instances are implicitly mapped into a high dimensional space using a representation of a dot product in this space by a kernel. The SVM constructs a decision hyper plane in the high dimensional space that separates training instances of two classes in such a way that the distance from the closest training instance to the hyper plane is maximum. This distance is called a margin and, consequently, the SVM finds a decision hyper plane with the largest margin. Gaussian kernel is used here.

D. Lung tumor area recognition and classification using EK-mean clustering and SVM:

After GLCM feature extraction, SVM classification is implemented over the extracted features to classify benign or malignant. SVM classifiers are used for getting the hyper plane that partition to wall data points of classes with each other. That resulted hyper planes of SVM classifies [6] the margin of the two classes each other. The vector value for the neighbour boundaries is known as supporting vectors and that calculated distances from that vector and hyper-plane was named as the margins. SVM found the input feature of that vector into high dimensional feature characteristics and then sent to classify that features.

Table 1. Concordance correlation coefficient (CCC) and dynamic range (DR) for CT image and average for tumor volume.

Features	Metric to Test Repeatability		Absolute % difference (|Test-retest|) statistic (μ , σ)
	CCC	**DR**	
convexity	0.897	0.934	(3.87%, 4.25)
entropy of tumor core	0.904	0.954	(1.84%, 2.16)
entropy of tumor boundary	0.826	0.937	(2.02%,2.76)
entropy ratio	0.366	0.924	(1.82%,2.64)

Table 2. Analysis of CT Image use matching center slice

Features	**CCC**	**DR**
Convexity	0.993	0.968
entropy of tumor	0.997	0.973
entropy of tumor boundary	0.996	0.971
entropy ratio	0.997	0.983

Features like entropy, correlation, convexity can be used for SVM classifier. It uses entropy and convexity of concordance correlation coefficient (CCC) and dynamic range for finding the average tumor volume.

E. Early detection of lung cancer using SVM classifier in biomedical image processing:

Plot data items in n dimensional space where n is the number of features with the value of the feature being equal to the value of the coordinate and then perform classification by finding the hyper plane. SVM use optimum linear separating hyper planes which can be used for classification and regression. An optimum hyper plane is used to separate two sets of data in feature space and the optimum hyper plane is produced by distinguishing margins between the two sets. This means, the hyper plane will depend on border training patterns called support vectors. Here, the linear kernel SVM is used to classify the image into normal or cancerous images. The accuracy, sensitivity and specificity obtained from the proposed model for lung cancer type and its different stages are shown in table below. Discrete waveform Transform is applied for image compression and features are extracted using a GLCM. The results are fed into an SVM classifier [7] to determine if the lung image is cancerous or not. The SVM classifier is evaluated based on an LIDC dataset. The classifier achieves an accuracy of 95.16%, sensitivity of 98.21% and specificity of 78.69%. In future work, sensitivity and accuracy could be improved further by improving the candidate nodule pruning algorithm.

V. CONCLUSIONS

This paper attempts to study and provides a brief knowledge about the different image classification approaches and different classification methods. Most common approaches for image classification can be categories as supervised and unsupervised, or parametric and nonparametric or object-oriented, subpixel, per-pixel and perfield or spectral classifiers, contextual classifiers and spectral-contextual classifiers or hard and soft classification. This survey gives theoretical knowledge about different classification methods and provides the advantages and disadvantages of various classification methods.

REFERENCES

1. V. Vapnik. The Nature of Statistical Learning Theory. NY: Springer-Verlag. 1995.
2. Chang, C.-C. and C. J. Lin (2001). LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm
3. Guo, N., Yen, R.-F., Fakhri, G. E., & Li, Q. (2015),"SVM based lung cancer diagnosis using multiple image features in PET/CT" 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). doi:10.1109/nssmic.2015.7582234".
4. Hong, S., Tian-yu, N., Yan, K., & Hong, Z (2010)," Chest DR Image Classification Based on Support Vector Machine", 2010 Second International Workshop on Education Technology and Computer Science. doi:10.1109/etcs.2010.123".
5. Alena Shamsheyeva, & Arcot Sowmya. (n.d.)," The Anisotropic Gaussian Kernel for SVM Classification of HRCT Images of the Lung", Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004. http://doi://10.1109/issnip.2004.1417501".
6. Gopi, K., & Selvakumar, J. (2017),"Lung tumor area recognition and classification using EK-mean clustering and SVM", 2017 International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2). doi:10.1109/icnets2.2017.8067906".
7. Kaucha, D. P., Prasad, P. W. C., Alsadoon, A., Elchouemi, A., & Sreedharan, S. (2017)" Early detection of lung cancer using SVM classifier in biomedical image processing", 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). http://doi://10.1109/icpcsi.2017.8392305

