Improved Marginsharpness Pulmonary Nodules Analysis Type And Treatment Prediction Using Lungs X-Ray Images

K. Priyadharshini[1], S. Sowmiya[2], M. Sri Senbaga Kuzhalvai Mozhi[3], Mrs. K. Kalaivani M.E[4]

[1, 2, 3] *UG Students, Department of Computer Science and Engineering*
[4]*Assistant Professor of Computer Science and Engineering Arasu Engineering College, Kumbakonam, Tamil Nadu, India*

Abstract— *The lung cancer is the reason of lot of deaths on population around the world. An Which acquired for the processing image. Here the X-Ray images with low dose can used for the nodule identification. The low dose Usually the exiting work of pulmonary nodule identification can be done through the image can be reconstructed through the effective algorithm so called Global Dictionary based Statistical Iterative Reconstruction small pulmonary nodules, a Computer-Aided Diagnosis (CAD) tool provides assistance to treatment options. Due to complexity diagnosis of radiologist aiming the improvement in the lungs nodules*

Keywords— *Imageprocessing, Datamining, Support Vector Machine (SVM), Global Dictionary Based Statistical Iterative Reconstruction,Genetic Algorithm.*

Learning about lung cancer is an important part of preparing for the decisions you will be making about your treatment and aftercare. The more you know about lung cancer, the better position you will be in to make informed decisions and advocate for your own interests. This chapter provides an overview of lung cancer topics including: lung cancer risk factors, how lung cancer develops, grows, and spreads, and types of lung cancers. The information will help you understand other chapters in this book, and your health care providers' recommendations for your care. Lung cancer develops when normal lung cells sustain genetic damage that eventually leads to uncontrolled cell proliferation. Like all cancers, lung cancer cells have the ability to invade neighboring tissues and spread or metastasize to distant parts of the body. Left untreated, lung cancer eventually causes death. Lung cancer is sometimes referred to as bronchogenic cancer or bronchogenic carcinoma.

I. INTRODUCTION

Image processing is processing of images using mathematical operations by using any form of signal processing for which the input is an image, a series of images, or a video, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image.[1] Most image-processing techniques involve treating the image as a two dimensional signal and applying standard signal-processing techniques to it. Images are also processed as three-dimensional signals with the third-dimension being time or the z-axis.

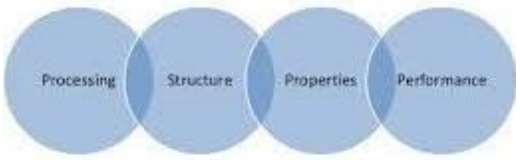

Fig 1.1 Steps For Image processing

The word bronchogenic means originating from the bronchi, the airways of the lungs. Most lung cancers begin in the cells lining the bronchi of the lungs. There are two main types of lung cancer, and they are treated differently. Learning about your particular type of lung cancer can help you talk with your health care providers about your treatment choices.

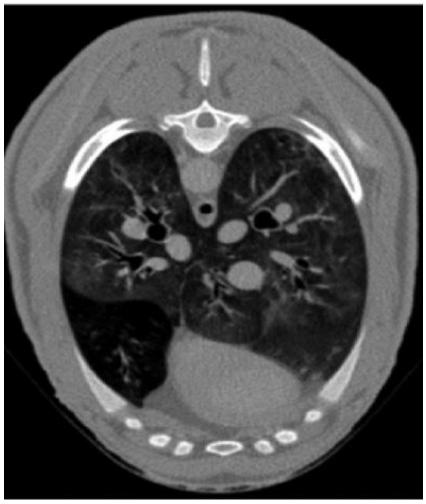

Fig 1.2 Image of Lungs

Data mining is the process of analyzing data from different perspective and summarizing into useful information –information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing the data. It allows user to analyze data from many different dimension or angles, categorized it, and summarize the relationship identified.

Technically data mining is a process of finding co- relations or patterns among dozens of fields in large relational database. Data mining is also called knowledge, discovery in database. It process of discovering interesting and useful patterns, relating large amount of data.

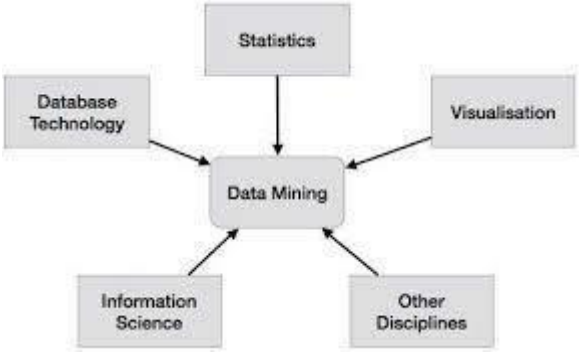

Fig 1.3 Related areas of data mining

II. RELATED WORK

Related research in various field:

Died Erich[2] proposed present "screening data from a nonrandomized trial by using low dose computed tomography (CT)"and a simple algorithm based on size and attenuation of detected nodules to guide diagnostic work-up. Lung cancer screening with low-dose CT demonstrated a prevalence of asymptomatic cancers in 1.3% of a smoking population, including a high proportion of early tumor stages and a 20% (three of 15) rate of invasive procedures for benign lesions.

McLennan G proposed development of computer aided diagnostic (CAD) method for lungs nodules detection, classification, quantitative assessment can be facilitated through a characterized repository of computed tomography (CT) scans. The LIDC/IDRI Database is expected to provide an essential medical imaging research resource to spur CAD development, validation, and dissemination in clinical practice.

Abhishek Kumar Tripathi Two performance metrics which quantify the spread and flatness of the image histogram are proposed. These metrics help us to distinguish between the low and high contrast images. Once we have effective performance metric for contrast, it will be possible to determine whether there is a need of contrast enhancement or not. Simulation results with popular images, natural images, and medical reveal that out of the two proposed metrics HS is more meaningful than HFM. The reason is that HS consider the histogram counts as well as histogram bin locations. HS can effectively discriminate low & high contrast images. Once the quantitative value for the contrast of the image is known, it is easier to say whether there is a requirement of contrast enhancement or not.

McNitt-Gray the variability of lesion size measurements in computed tomography data sets of patients imaged under a "no change" ("coffee break") condition and to determine the impact of two reading paradigms on measurement variability. Independent reading paradigm, the mean percent change for 1D measurement is quite low.

Napel S Method to quantify the margin sharpness of lesions on CT and to evaluate it in simulations and CT scans of liver and lung lesions. New image feature to quantify the margin sharpness of lesions. It has strong correlation with known margin sharpness in simulated images and in clinical CT images containing liver lesions and lung nodules. New image feature to quantify the margin sharpness of lesions.

III. METHODOLOGIES

A. Proposed Architecture

The X Ray Images also used for the utilization of the lung pulmonary nodule detection. Thus the x-Ray Images acquired from the acquisition device with low dose. It need image reconstruction to identify the nodules; the process can perform through global dictionary based statistical iterative reconstruction (GDSIR). Because structures in various subjects

are quite similar in a given application, a training set of patches for construction of a global dictionary can be extracted from images of similar objects.

On the other hand, the training set of patches for construction of an adaptive dictionary can be extracted from an intermediate image and dynamically updated during the reconstruction process. In this paper, the margin sharpness of a nodule is represented using the combination of histogram spread of averaged gradient and acutance.

The proposed method outperforms the competing techniques in terms of precision, MS, and speed. The system computed two attributes of margin sharpness: the intensity difference between a lesion and its surroundings, and the sharpness of the intensity transition across the lesion boundary. These two attributes were extracted from sigmoid curves fitted along lines automatically drawn orthogonal to the lesion margin. It represented the margin characteristics for each lesion by a feature vector containing histograms of these parameters.

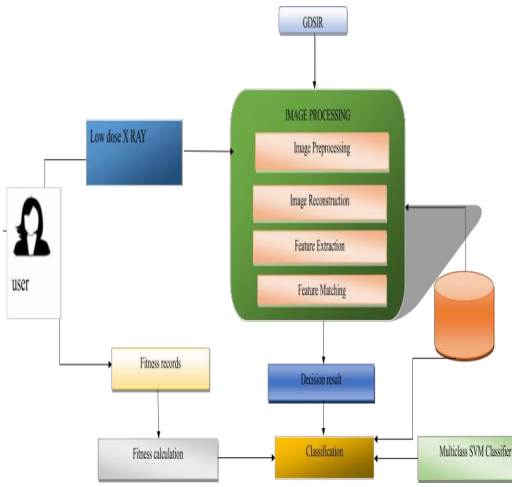

Fig. 3.1 System Architecture of Proposed System

B. IMAGE PROCESSING

a. Image Preprocessing
Preprocessing is a common name of operations with image at the lowest level abstraction both input and output are intensity images.The aim of preprocessing is improvement of image data that suppresses unwanted distortion or enhance some image features important for further processing.

b. Image reconstruction
It is refers to iterative reconstruction used to reconstruct 2D and 3D images is certain imaging techniques.In computed tomography images must be reconstructed from projection of images.

c. Feature segmentation
It is the process of portioning a digital image into multiple segments.The goal of segmentation is simply and changes the representation of images into something that is more meaningfull and easier to analyze.

d. Feature matching means finding the coreesponding features from to similar datasets based on search distance

C. SUPPORT VECTOR MACHINE

SVM is a supervised learning algorithm for classification problems. It is based on the structural risk minimization principle from computational learning theory. The idea is to find a hypothesis for which we can guarantee the lowest true error. The true error of a hypothesis is the probability that this hypothesis will make an error on an unseen and randomly selected test example.

One important property of SVMs is that their ability to learn can be independent of the dimensionality of the feature space. SVM's measure the complexity of hypotheses based on the margin with which they separate the data, and not the number of features. In its simplest, linear form, an SVM is a hyper plane that separates a set of positive examples from a set of negative examples with maximum margin. The bounded region is called the margin, and samples on the margin are called the support vectors. In the linear case, the margin is defined by the distance of the hyper plane to the nearest of the positive and negative examples.

SVM is fundamentally a two-class classifier. In practice we often have problems involving classification to more than two classes. In order to build a multi-class classifier, there have been proposed different approaches. Common approach is to construct a multi-class classifier by combining several binary classifiers. Topic identification with SVM implies a kind of semantic space in the sense that the learned hyper plane separates documents which belong to different topics in the input space.

One way to avoid high dimensional input spaces is to assume that most of the features are irrelevant. Unfortunately, in text categorization there are only few irrelevant features. Results of researches show that even features ranked lowest still contain considerable information and are somewhat relevant.

Since SVMs use over fitting protection, which does not necessarily depend on the number of features, they have the potential to handle large feature spaces. Because of that characteristic, this method is suitable for text

classification, as shown in several researches.

Genetic Operators Genetic algorithm uses the principle of selection to produce solutions at each generation. Matting of parents is represented by cross-over and mutation operations. Selection is used to select the individuals for next generation. The crossover is used to recombine the information. It is used to recombine two input strings that is called as parent string to get better output string usually called as child string. Crossover has different types like one point crossover, two point crossover and uniform crossover. Mutation may be the chromosomes of individuals to be different from their parent individuals.

D. GDSIR and ADSIR

In Section II-B the regularization term in represents prior information on reconstructed images. Various assumptions about the prior information lead to different reconstruction algorithms. For example, the assumption of smooth variation over adjacent pixels suggests a regularization in terms of quadratic differences between adjacent pixels.

The piecewise constant assumption supports a TV regularization. A more general form of regularization is the - generalized Gaussian Markov field (-GGMRF*)* prior, which has two adjustable parameters. The quadratic and TV regularization functions are special cases of - *GGMRF*. Because all these regularization means are based on the relationship of adjacent pixels, it is difficult for them to distinguish weak structures and strong noise.

On the other hand, the dictionary learning and sparse representation techniques perform well in sensing structures and suppressing noise.

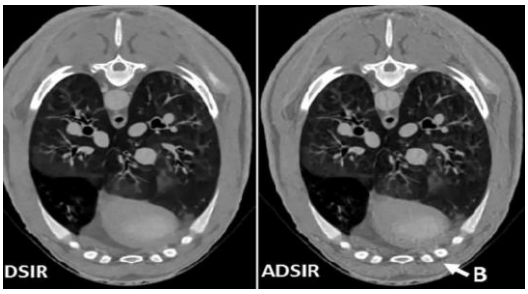

Fig. 4.1 Example for GDSIR – ADSIR

IV. PERFORMANCE ANALYSIS

We implemented the proposed model based on two techniques, Global Dictionary Based Statically Iterative Reconstruction and Support Vector Machine (SVM). In the SVM is a supervised learning algorithm for classification problems. It is based on the structural risk minimization principle from computational learning theory .The Global Dictionary Based Statically Iterative Reconstruction can be used to reconstructing the image features and segmenting their images. Genetic Algorithm is used matching the images.

V. SYSTEM SCREEN SHOTS

Fig. 5.1 Screen Shot For Image Preprocessing

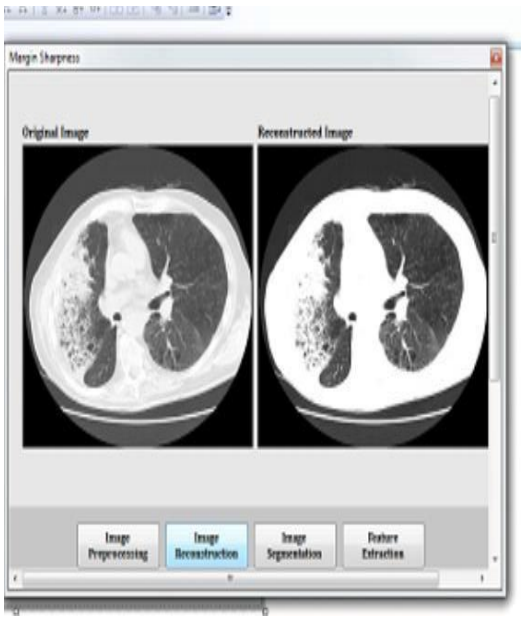

Fig. 5.1 Screen Shot For Image Reconstruction

VI. CONCLUSION AND FUTURE WORK

In this paper study the significance of image processing and data mining to the lung cancer. Multiclass SVM classifier classifies the collected fitness data. The margin sharpness of a nodule is represent using the combination of

histogram spread of averaged gradient acutance. Hence the proposed work can identified lungs cancer that will be produce from the result. In future proposed approach is not only confined to lungs cancer disease, it may also extent to infer information pertaining to other cancer diseases data such as breast cancer, brain tumor. In future,

References

[1]. Armato III, S.G., McLennan, G., Bidaut, L., et al.: 'The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans', Med.

[2]. Phys., 2011, 38, (2), pp. 915–931

[3]. C. Kamphuis and F. Beekman, "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm," *IEEE Trans. Med. Image.*, vol. 17, no. 6, pp. 1101–1105, Dec. 1998.CT," *Med.*

[4]. *Phys.*, vol. 32, p. 1676, 2005.

[5]. Diederich, S., Wormanns, D., Semik, M., et al.: 'Screening for early lung cancer with low-dose spiral CT: Prevalence in 817 asymptomatic smokers', Radiology, 2002, 222, (3), pp. 773–781

[6]. Horsthemke, W.H., Raicu, D.S., Furst, J.D.: 'Characterizing pulmonary nodule shape using a boundary-region approach' .Proc .of SPIE Medical Imaging 2009, San Diego, United States, February 2009, vol. 7260, pp. 72602Y–72602Y–9

[7]. J.Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sonogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE Trans. Med. Image.*, vol. 25, no. 10, pp. 1272–1283, Oct. 2006.

[8]. Ko, J.P., Naidich, D.P.: 'Computer-aided diagnosis and the evaluation of lung disease', J. Thorac. Imaging, 2004, 19, (3), pp. 136–155

[9]. P. La Rivière, "Penalized-likelihood sonogram smoothing for low-dose

[10]. Siegel, R., Naishadham, D., Jemal, A.: 'Cancer statistics', CA Cancer J. Clin., 2013, 63, (1), pp. 11–30