AN EFFECTIVE AND EFFICIENT FEATURE SELECTION METHOD FOR LUNG CANCER DETECTION

Mr. R. Kishore

Assistant Professor, Department of Information Technology
K.L.N College of Engineering,Pottapalayam, Sivagangai Dist

ABSTRACT

Medical image data is growing rapidly. Lung cancer considers to be the most common cause of death among people throughout the world. Early lung cancer detection can increase the chance of people survival. The 5 year survival rate for lung cancer patient increases from 14 to 49% if the disease is detected in time. Computed Tomography can be more efficient than X ray for detecting lung cancer in time. But the problem seemed to merge due to time constraint in detecting the presence of lung cancer. MATLAB have been applied for the study of these techniques. Feature selection is a method to reduce the number of features in medical applications where the image has hundreds or thousands of features. In order to extract the accurate features of an image, an image need to be processed for its effective retreival. Image feature selection is an essential task for recognizing the image and it can be done for overcoming classification problems. However, the quality of the image recognition tasks can be improved with the help of better classification accuracy for enhancing the retrieval performance.

KEYWORDS

Feature selection, image recognition, classification, retrieval.

1. INTRODUCTION

Medical imaging modalities are used to investigate the human body. The resulting images need to be applied with the help of sophisticated image processing methods for extracting the features of an image and image analysis methods are essential for automated or semi-automated lung cancer detection. The result of these methods must be measured and it must be fed into the multiple transformation process of extracting the feature of a particular image. Image feature selection is an essential task which can affect the performance of image classification and recognition. For this purpose, ant colony optimization method has been successfully applied for this activity. The quality of the image recognition tasks can be improved with the help of better classification accuracy for enhancing the retrieval performance.

2. LUNG CANCER DESCRIPTION

Lung cancer is a serious health problem for our body. The mortality rate of lung cancer is highest among other types of cancer. It is one of the serious cancers with the smallest survival rate after the diagnosis. Survival from lung cancer is directly related to growth at its detection time. The earlier the detection, higher the chances of successful treatment. In order to detect the lung cancer in earlier stage, a computer aided diagnosis system is required which can be implemented with image processing techniques.

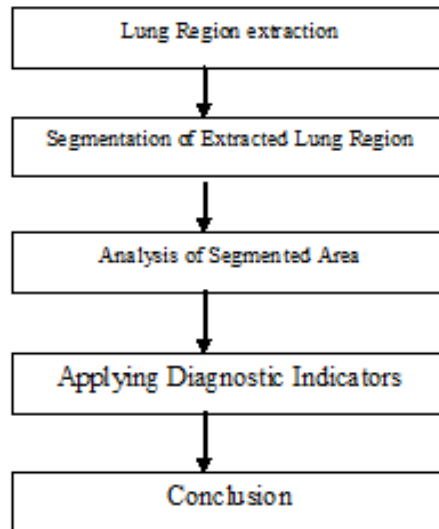## 3. BLOCK DIAGRAM OF LUNG CANCER DETECTION SYSTEM

Fig 1 Lung Cancer Detection System

LCDS is a system which can be used to detect lung nodules. It consists of five main steps such as an extraction of lung region, segmentation of extracted lung region, analysis of segmented area, applying diagnositic indicators and conclusion. For these five steps, image processing techniques need to be applied. Such image processing techniques are listed here: erosion, wiener filter, dilation, lung border extraction, thresholding, region growing, edge detection, ridge detection, morphological operation, fitting of geometrical model and dynamic programming for extraction of lung region. Then segmentation algorithms is to be applied to detect the cancer nodules from the extracted lung region. Rule based technique is applied to classify the cancer nodule. Finally, a set of diagnosis rules are generated from the extracted features.

4. MEDICAL IMAGE PROCESSING

Image processing in medical diagnosis can be done with the help of following stages such as image capture, image enhancement, image segmentation and feature extraction   Figure 1 illustrates a process flow in medical image processing. As illustrated in Figure 1, medical image processing contains different stages. The first stage starts with image capture. The second stage proceeds with image enhancement to get best level of quality and clearness. The third stage proceeds with image segmentation which play an effective role in image processing stages. The fourth stage extracts the general features from enhanced segmented image which is used to give an indication of normality or abnormality of images.

5. STAGES IN MEDICAL IMAGE PROCESSING

The lung CT image having low noise when compared to MRI image. So I consider CT scan images for detecting the abnormal function in the lung. The main advantage of the computed tomography image having better clarity, low noise and distortion.
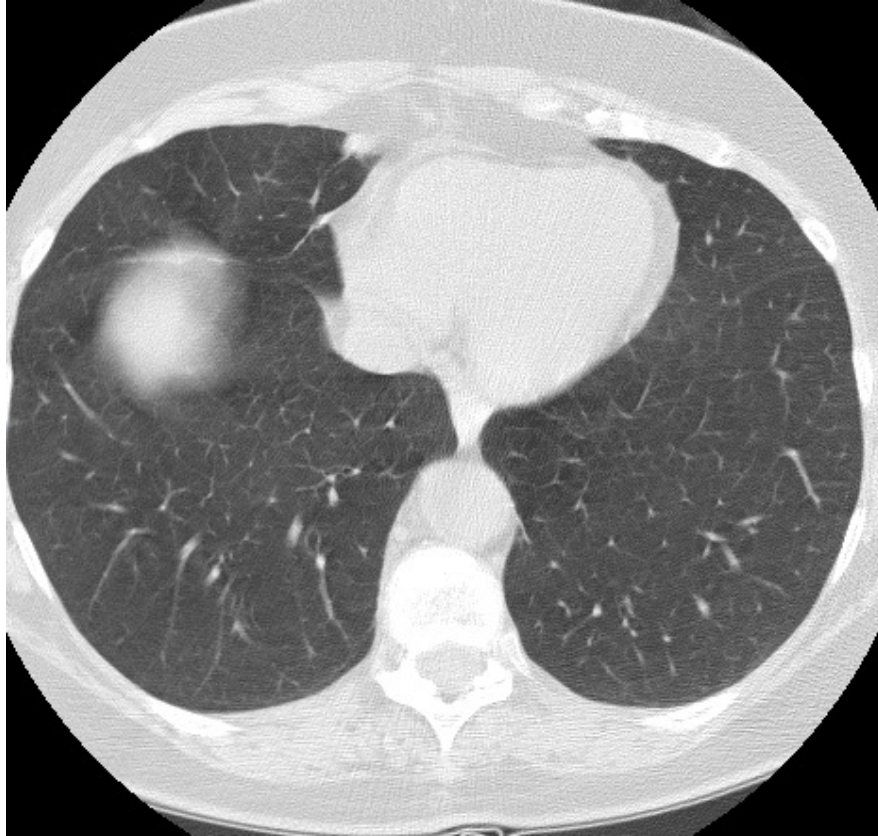
Fig 2 The Lung CT Image

6. FEATURE SELECTION

Feature selection is a image reduction technique which is being widely used for image mining and knowledge discovery and it allows elimination of irrelevant and redundant features while retaining the relevant features. Feature selection implies less image transmission and efficient image mining. Feature selection is one of the research topics used in machine learning and other related fields. It remove the irrelevant and redundant features and the quality of the image recognition system is to be improved for enhancing the performance of learning systems. It can be done with the help of Support Vector Machine. Rapid growth of medical image databases motivated Content Based Image Retrieval requires efficient search schemes. It will be provided with the help of clustering technique. Low level visual features including color, texture and shape, are automatically selected for representing medical images. It is one of the pre-processing step to machine learning for image recognition.
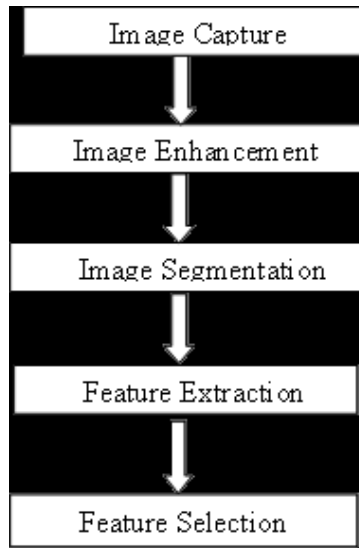
Figure 3 Process Flow in Medical Image Processing

7. FUNDAMENTAL FEATURE SELECTION METHODS

Selection of an image can be done with the help of feature selection technique such as screening, ranking and selecting. Screening is used to remove unimportant and problematic features and records or cases, such as features with too many missing values or features with too much or too little variation to be useful. Ranking is used to sort remaining features and assigns ranks based on importance. Selecting is used to identify the subset of features by preserving only the most important features and filtering all others. The Feature Selection technique screens, ranks and selects the features that are most important.

8. REVIEW OF LITERATURE ON FEATURE SELECTION METHOD

Haleh V. and Kenneth D. J, a image processing community member played a vital role for the development of various techniques for the feature selection technique with the help of a classification problem founded during image recognition[1]. Sérgio *et al.* evaluate the ranking method for a family of feature selection methods based on genetic algorithm, it tends to improve the accuracy of content-based image retrieval systems and it also evaluate the ranking quality for improving retrieval performance[2]. Medical images play a vital role in patient diagnosis, therapy, surgical planning, medical reference, and training. With the recent increase in the availability of filmless radiology equipment, the management of medical images is receiving attention. Picture Archiving and Communication Systems (PACS) is the technology used in many hospitals and specialized clinics for providing quick access to screening exams and the actors are being integrated to involve in the enterprise workflow. The radiological databases was originally built for storing medical images which is being evolved from simple storage servers to active repositories for research and decision support. Jaba and Shanthi reviewed on continuous feature discretization and identified the defined characteristics of the method. Then a new supervised approach was being proposed which is used to combine discretization and feature selection for selecting the relevant features which is being used for classification purpose. Associative Classifier is one of the classification technique used here [3]. Medical images play a fundamental role for medical diagnosis and treatment. These images differ from photographic images because

they reveal the internal anatomy as opposed to an image of surfaces. Huanzhang *et al.,* discussed feature selection as an important technique when training classifiers in machine learning problems and described the complexity of the classifier parameters adjustment during training increases with the number of features. So there is a need of novel embedded feature selection method called ESFS, which was inspired from the wrapper method[4]. Georgia *et al.,* discussed the information theoretic approach to feature selection for computer-aided diagnosis based on mutual information (MI) concept. [5]. Mohamed *et al.,* discussed an approach which was capable to develop a computer-aided diagnosis (CAD) system which is being useful for radiologist in diagnosing the appropriate medical patterns and efficiency of feature selection seen on the CAD system. The system is being implemented in four stages which are [6]:

a) selection of the region of interest for a given image
b) The feature extraction stage based on the region of interest
c) The feature selection stage which select the relevant features to be used in next stage, and
d) The classification stage which is used to recognize image

Cancer[7] that forms in the growth of lung tissue which usually occurs in the cells lining air passages. lung cancer can be categorized into two types (i) Small cell lung cancer is also called as oat cancer. This type of lung cancer tends to spread quickly. (ii) Non small cell lung cancer is the most common type of lung cancer. Lung cancer may found as a mass or a tumor on a chest of a patient.

9. IMAGE CLUSTERING

Clustering is an unsupervised feature selection technique. CBIR and Image Mining have been combined to increase the speed of the image retrieval system. Users often query image based on the feature. It is one of the feature extraction process used in medical image processing.

10. AUTOMATED COMPUTER AIDED DIAGNOSIS SYSTEM

The automated CAD System is being proposed for early lung cancer detection by analyzing lung CT images using several steps. This approach starts by extracting the lung regions from the CT image using image processing techniques such as bit image slicing, erosion, wiener filter. Bit image slicing is used in the extraction process to convert the CT image into a binary image. After the extraction step, the extracted lung regions are segmented using region growing segmentation algorithm. Then, the initial candidate nodules resulting from the region growing segmentation algorithm are analyzed to extract a set of features to be used in the diagnostic rules. After this, lung regions are used to detect cancerous cells.

11. EXTRACTING LUNG REGION

It is very important to separate region of interest from other parts of the image.

12. SEGMENTING EXTRACTED LUNG REGION

Image segmentation is used to locate the objects and boundaries in the region of interest.

13. FEATURE EXTRACTION

The image feature extraction stage is used to extract desired portions of an image. After the segmentation is performed on lung region, the features can be extracted from it and the diagnosis rule can be designed to detect the cancerous nodules in the lungs.

14. MATLAB DESCRIPTION

The name MATLAB stands for MATrix LABoratory. It was founded originally to provide easy access to matrix software developed by the LINPACK and EISPACK projects. It is a high performance language for technical computing. MATLAB as an excellent tool for research which will able to solve technical problems. This software package has been commercially available since 1984 and it is now considered as the standard tool for universities and industries worldwide. It has powerful built in routines that enable a wide variety of computations.

Matlab offers many predefined mathematical functions for technical computing which contains a large set of mathematical functions. It has an excellent set of graphic tools. Plotting a given data set is possible with very few commands. The MATLAB command to plot a graph is plot(x,y).

>> x=[1 2 3 4 5 6]
>> y=[3 -1 2 4 5 1]

15. PATTERN

A pattern is a vaguely defined entity that has a name. Pattern recognition is a scientific discipline whose aim is the classify objects into a lot of categories. It is an integral part of machine intelligence system built for decision making. It is an act of taking in raw data and making an action based on the category of the pattern. It is a study of how machines can do the lot of activities. Such activities are listed below:

→ Observe the environment
→ Learn to distinguish patterns of interest
→ Make sound and reasonable decisions about the categories of the patterns.

16. APPLICATION OF PATTERN RECOGNITION

Other than medical diagnosis, there are many pattern recognition applications. These applications are listed below:

Handwritten digit
Biometrics
Speech Recognition
Smell Recognition
Defect detection in chip manufacturing
Interpreting DNA sequences
Fruit recognition
Terrorist Detection
Credit Fraud Detection
Credit Applications

17. CONCLUSION AND DISCUSSION

The performance of three different methods that is being considered are: screens, ranks, and selects. In order to evaluate the performance, image analysis is being done for its recognition. Classification accuracy can be predicted for enhancing the performance of image retrieval. For experimentation of this technique, the CT images are obtained from a NIH/NCI Lung Image

Database Consortium dataset that provides the chance to do this research. The experimentation data consists of 1000 lung images. These 1000 lung images are passed to the system. The diagnostic rules are then generated from those images and these rules will provide input to the classifier for the learning process. After learning, lung images are supplied to the system. Then the system will proceed through its processing steps and finally it will detect whether the supplied lung image is with cancer or not.

REFERENCES

[1] Haleh V. and Kenneth D. J, (1992). Genetic Algorithms as a Tool for Feature Selection in Machine Learning. Artificial Intelligence, 102-109.
[2] Sérgio Francisco da Silva , Marcela Xavier Ribeiro , João do E.S. Batista Neto, Caetano Traina-Jr. ,Agma J.M. Traina, (2011). Improving the ranking quality of medical image retrieval using a genetic feature selection method. Medical Imaging, 1012-1017.
[3] Jaba S. L. and Shanthi V., (2009), International Journal of Computer Theory and Engineering, 1(2), 154-158.
[4] Huanzhang F, Zhongzhe X, Emmanuel D, Weibei D, and Liming C, (2009). Image Categorization Using ESFS: A New Embedded Feature Selection Method Based on SFS. 288–299.
[5] "Georgia D. T.", "Erik D. Frederick", "Mia K. Markey", and "Carey E. F"., (2001). Department of Biomedical Engineering, Duke University, Durham.
[6] "Mohamed A. A", "Wael A. M", "Abo-Bakr M. Y", "Yasser M. K", and "Ahmed S. M", (2009). Feature Selection in Computer Aided Diagnosis System for Microcalcification Detection in Digital Mammograms. Proc of 26th National Radio Science Conferenc (NRC2009).1-9.
[7] American Cancer Society http://www.cancer.org/cancer/lungcancer/

AUTHOR

Mr. R. Kishore, working as a assistant professor in K.L.N College of Engineering, Pottapalayam. He studied B.E Electronics and Instrumentation and M.E Computer Science and Engineering from Thiagarajar College of Engineering, Madurai. He has been worked as a testing engineer in Cognizant Technology Solutions India Private Limited and he has been completed cognizant certification in software testing, quick test professional, test partner, general insurance, personal insurance, life insurance and life science. He is having more than 5 years of teaching experience in engineering colleges.