REVIEW ON: SEGMENTATION AND CLASSIFICATION OF SPUTUM CELL IN LUNG CANCER DETECTION

Money Sethi¹, Sukhdeep Kaur²

¹, ²Assistant Professor, ¹, ²Adesh Institute of Engg. & Tech. Faridkot, Punjab (151203)

Abstract: Diagnosis of lung cancer in its primal stage is a major issue confronted by the medical world. For that an efficient segmentation technique is required for the detection of Area of interest i.e. cancer cells. However, many common forms of techniques are available in market and all we need is the accuracy of segmentation of the Area of Interest (Cells) and Non-Area of Interest (Back-Ground). In this paper, we will discuss the segmentation and various classification techniques which have helped in lung cancer detection. The lung cancer detection method uses sputum cytology images.

Keywords: Lung Cancer Detection, Segmentation and Classification

1. INTRODUCTION

A. Lung Cancer

Lung cancer is one of the serious and life threatening cancer. Lung cancer is one of the most frequent cancers worldwide. Like breast cancer in female, lung cancer is the leading cancer site in males, comprising 17% of the total new cancer cases and 23% of the total cancer deaths. In medical image processing the detection of various types of cancer affecting different organs is a prime ongoing research area. Cancer detection should be more accurate since its reports will be used for diagnosis. So many researchers have provided different techniques, but most of the methods got delimited to the segmentation step itself. The usage of fine needle or other equipments to the affected area is more painful to the patient. Surgery is needed for the diagnosis of lung cancer, which will be costly and pain full. Another commonly used method for lung cancer detection is the analysis of CT images. But the information obtained from CT has the problem of intensity variation. Hence, there is chance to get fault results and it may lead to wrong diagnosis and other related problems, which harms medical practices. For this analysis, the sputum of the affected person can be used which can be obtained without surgery or any other harmful method. Sputum images can be used for the detection of lung cancer.

B. Image Segmentation

Image segmentation is the procedure of dividing a digital image into multiple segments. The objective of segmentation is to improve and change the representation of an image into something that is more important and easier to investigate. Image segmentation is regularly used to find objects present in the image. There are following types of image segmentation:

Threshold based segmentation: Histogram thresholding and slicing procedures are utilized to segment the image. They may be connected directly to an image but can likewise be consolidated with pre-processing and post-processing systems.

Edge based segmentation: With this system recognized edges in an image are accepted to represent object boundaries and used to recognize these items.

Region based segmentation: Where an edge based system may endeavour to discover the object boundaries and then find the object itself by filling them in a region based system takes the opposite method by starting in the middle of an item and then “growing” outward until it meets the object boundaries.

Clustering techniques: Although clustering is sometimes utilized as an equivalent word for segmentation systems and we utilize it here to signify procedures that are essentially utilized as a part of exploratory information examination of high-dimensional estimation designs. In this connection grouping routines endeavour to gathering together examples that are comparable in some sense. This objective is very much alike to what we are endeavouring to do when we section an image and without a doubt some clustering methods can promptly be sought image segmentation.

II. CLASSIFICATION TECHNIQUES

Classification is mainly used for pattern recognition. Pattern recognition means identification of the ideal objects. To recognize an object we must receive some information or features from the object on basis of these features we assign the object to particular class that represents a pattern. The set of features that distinguishes objects of different classes is key for classification and recognition. Identifying features is important step in the process of recognition. This process is called feature selection. Decision process is developed through learning. There are two types of learning:

Supervised: Supervised learning assumes that a set of training data has been provided, consisting of a set of objects that have been properly labeled by hand with the correct output e.g. support vector machine, artificial neural network.

Unsupervised: Unsupervised learning assumes training data that has not been hand-labeled, and attempts to find inherent patterns in the data that can then be used to determine the correct output value for new data instances, e.g. k-means clustering. Other approaches used for decision process
through learning are clustering, graph matching, etc. In medical image processing, there are a number of classification techniques that have been used for detection of lung cancer some of them mentioned in papers are discussed below:

A. Bayesian classifier: This method was proposed by Naïve Bayes in 1950. In machine learning, this method remained popular for text categorization, the problem of judging document as being to one category or the other. It is competitive in its domain and is more advanced method. A Naïve Bayes classifier or Bayesian classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. It is used for judging whether a particular document belongs to one category or the other and we can categorize object belongs to one category or the other. For example a fruit may be considered as an apple if it is red, round and is of 4 inches in diameter. Even if these features depend upon each other or upon the existence of other features the Naïve Bayes classifier will consider all these properties to contribute independently to the probability that this fruit is an apple. Naïve Bayes classifiers can be trained very efficiently in a supervised learning setting. The Naïve Bayes classification has worked well in many complex real-world problems. In 2004, analysis of the Bayesian classification problem has shown that there are some theoretical reasons for unreasonable efficiency of naïve Bayes classifiers. Still, a comprehensive comparison with other classification methods in 2006 showed that Bayes classification is outperformed by more current approaches. It only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification.

B. Support vector machine: The SVM algorithm was invented by Vladimir N. Vapnik in 1963. SVM are supervised learning (machine learning which includes construction and study of algorithms that can learn from and make predictions on data) model with associated learning algorithm that analyse data and recognize pattern used for classification. If we have a lots of training examples and each example belongs to one of the two categories. SVM training algorithm builds a model which will determine whether the given example belongs to which one of the two categories. It represents example as points in spaces so that examples of separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into same space and predicted to belong to a category based on which side of the gap they fall on. They are suitable for binary classification tasks. More formally, SVM constructs hyper plane or a set of hyper-planes which can be used for classification. A good separation is achieved by the hyper-plane that has largest distance to the nearest training data points of any class. They are used to solve the real world problems. They are helpful in text and hypertext categorization. Classification of images can be performed by using SVM. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes

C. Hopfield neural network: Hopfield Neural Network (HNN) is another classifier which is one of the artificial neural networks, which has been proposed for Segmenting both gray-level and color images. If one have the segmentation problem for gray-level images as minimizing a suitable energy function with HNN, it derived the network architecture from the energy function, and classify the sputum cells into nuclei, cytoplasm and background classes.

D. Fuzzy Technique: It is one of the most difficult tasks in image analysis & computer vision. It is to classify the pixel in an image correctly, when there is no crisp boundaries between objects in an image thus in order to address this difficulty, fuzzy technique is used. Fuzzy technique classifies pixel values with great extent of accuracy & it is basically suitable for decision oriented applications like tissue classification & tumor detection etc. Fuzzy clustering divides the input pixels into clusters or groups on the basis of some similarity criterion, such that similar pixels belong to same cluster.

III. COMPARISON OF THE TECHNIQUES

The comparison of different techniques is shown in Table 1.

<table>
<thead>
<tr>
<th>Statistical Parameter</th>
<th>Threshold and Bayesian</th>
<th>Threshold and SVM</th>
<th>Region Based and Hopfield-NN</th>
<th>Edge Based And Fuzzy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specificity</td>
<td>60.86%</td>
<td>91.30%</td>
<td>51.84%</td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>81.82%</td>
<td>81.82%</td>
<td>62.64%</td>
<td></td>
</tr>
<tr>
<td>PLR</td>
<td>2.09</td>
<td>9.41</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td>NLR</td>
<td>0.299</td>
<td>0.1991</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>PPV</td>
<td>66.7%</td>
<td>90.0%</td>
<td>54.87%</td>
<td></td>
</tr>
<tr>
<td>NPV</td>
<td>77.8%</td>
<td>84.0%</td>
<td>67.63%</td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>71.1%</td>
<td>86.7%</td>
<td>70.13%</td>
<td>70.4%</td>
</tr>
</tbody>
</table>

Table I: Results of the above techniques.

IV. CONCLUSION

In this paper, we classify and discuss main image segmentation techniques which have been used for lung cancer detection. These techniques distinguish between the area of interest (cells) and non-area of interest (background). We are planning to create a work which will include a segmentation technique which will recognize some of the cells in the image. This segmentation technique may not determine all the cancer cells in the image but will help in training a classifier which will recognize all of the remaining cancer cells present in the image leading to an efficient cancer detection method. The results will be compared with the results obtained by one of the above image classifier such as Bayesian classifier and Support Vector Machine which showes high accuracy level.
REFERENCES


