International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 06 | June -2017 www.irjet.net p-ISSN: 2395-0072

DETECTION OF LUNG CANCER USING SVM CLASSIFICATION

Priyadarshini Jambagi1 and M. S. Shirdhonkar2

[1]M.Tech student, Computer Science and Engineering
BLDEA'S V.P. Dr. P.G. Halakatti College of Engineering and Technology, Vijayapur, Karnataka, India
[2]Professor in Computer Science and Engineering Department
BLDEA'S V.P. Dr. P.G. Halakatti College of Engineering and Technology, Vijayapur, Karnataka, India

---

Abstract - *Lung cancer is one of the most serious diseases in the world with the smallest survival rate after diagnosis and with a progressive increase in the number of deaths each year. Survival from lung cancer is directly related to its growth at its detection time. The early detection and treatment of lung cancer can greatly improve the survival rate of patients. We attempt to develop a computer aided diagnosis (CAD) system for early detection of lung cancer based on the analysis of the sputum color images. We need to resolve several problems to complete the development of such system. Classification is one of the important problems to be considered. In this paper, we present a feature extraction process followed by a SVM classification technique to classify the sputum cell into cancerous or normal cell. We used 100 sputum color images to test the result. The performance criteria such as sensitivity, precision, specificity and accuracy were used to evaluate the proposed techniques.*

Keywords—Lung cancer; Sputum images; Feature extraction; SVM; Classification.

1. INTRODUCTION

Cancer is the second most common disease after heart diseases in the world. Lung cancer remains the leading cause of cancer related death in the world. The overall 5 –year survival rate with lung cancer is 14%. patients with early stage disease who undergo curative resection have a 5-year survival rate of 40% to 70%. The most recent estimate statistics according to the American Cancer Society indicates that in 2014, there were 224,210 new cases, accounting for about 13% of all cancer diagnoses. Lung cancer accounts for more deaths than any other cancer in both men and women. The earlier the detection is, the higher the chances of successful treatment. In the literature, there are many techniques that have been developed for lung cancer segmentation and classification [2]. The detection of lung cancer can be done in several ways, such as computed tomography (CT), magnetic resonance imaging (MRI), and X-ray. All these methods consume a lot of resources in terms of both time and money, in addition to their implementations [3]. Recently, some medical researchers have proved that the analysis of sputum cells can assist for an early diagnosis of lung cancer. Furthermore, image processing techniques provide a reliable tool for improving the manual screening of sputum samples. The segmentation of the sputum cells for lung cancer detections has been addressed by many articles. For example the authors in [4] used the Hopfield Neural Network (HNN) and the mean shift clustering algorithm for segmenting the sputum cells.

2. RELATED WORK

In [3] Authors developed a system Computer Analysis of Computed Tomography Scans Current computed tomography (CT) technology allows for near isotropic, sub millimeter resolution acquisition of the complete chest in a single breath hold. These thin-slice chest scans have become Indispensable in thoracic radiology, but have also the best Substantially increased the data load for radiologists. Automating the analysis of such data is therefore a necessity and this has created a rapidly developing research area in medical imaging. The work presents a review of the literature on computer analysis of the lungs in CT scans and addresses segmentation of various pulmonary structures, registration of chest scans, and applications aimed at detection, classification and quantification of chest abnormalities. In addition, research trends and challenges are identified and directions for future research are discussed.

In [4] Authors Automated clinical image data collection tools and apparatus are becoming increasingly important to the medical industry, and imaging databases are growing at an unprecedented rate. Consequently, grid-based telemedicine efforts require the autonomous classification of patient images from distributed sources for fast and accurate image storage, management, and retrieval. In this paper, they presented a unique algorithm that performs feature discovery to find class-wise isomorphic Association Rules (ARs) among features. By discovering ARs, we are able to find unique and useful knowledge in images. To find knowledge, first uniformly segment every image in a series and extract color and texture features for every segment. Next, Discover ARs for the color and texture features for image segments. Then exploit redundancy in the differentials of rule sets for the autonomous classification of patient image data with significant sensitivity and specificity. They demonstrated the efficacy of our approach with experimental results on a data set of diabetic retinopathy patients.

In [5] Authors have presented Comparison of Hopfield Neural Network and mean shift algorithm in segmenting sputum color images for lung Cancer Diagnosis Lung cancer continues to rank as the leading cause of cancer deaths worldwide. One of the most promising techniques for early detection of cancerous cells relies on sputum cell analysis. For this reason, we attempt to come with a computer aided diagnosis (CAD) system for early detection and diagnosis of lung cancer based on the analysis of the sputum color images. Therefore, the CAD system can play a significant role in the early detection of lung cancer. This paper, presents a comparison between two segmentation methods, a Hopfield Neural Network (HNN) and a mean shift clustering algorithm, for segmenting sputum color images to detect the lung cancer in its early stages. The two methods are designed to classify the image of N pixels among M classes. In this study, they used 100 sputum color images to test both methods. And used some performance criteria such as recall, precision, and accuracy to evaluate the proposed methods and the mean shift algorithm has shown a better segmentation performance compared to the HNN.

In [6] Authors proposed, Lung cancer has been the largest cause of cancer deaths worldwide with an overall 5-year survival rate of only 15%. Its symptoms can be found exclusively in advanced stages where the chances for patients to survive are very low, thus making the mortality rates the highest among all other types of cancer. The present work deals with the attempt to design computer-aided detection or diagnosis (CAD) systems for early detection of lung cancer based on the analysis of sputum color images. The actual goal was to reduce the false negative rate and to increase the true positive rate as much as possible. The early detection of lung cancer from sputum images is a challenging problem, due to both the structure of the cancer cells and the stained method which are employed in the formulation of the sputum cells. They presented here a framework for the extraction and segmentation of sputum cells in sputum images using, respectively, a threshold classifier, a Bayesian classification and mean shift segmentation. Their methods are validated and compared with other competitive techniques via a series of experimentation conducted with a data set of 100 images. The extraction and segmentation results will be used as a base for a CAD system for early detection of lung cancer which will improve the chances of survival for the patient.

3. PROPOSED SYSTEM

Proposed system includes Pre-processing is the initial step in all case of image related diagnosis system and it helps in accurate feature extraction which ultimately results in high classification accuracy. Before segmentation of lung region, we attempt to improve the lung CT images by evaluation of several image enhancement techniques. Image enhancement is most important step in our proposed method since the contrast of cancerous region in the lung images are initially very low. So the image enhancing technique is a best way to detect the lung cancer in introductory stages, more than a value related to some noise can be reducer and facilitating the better accuracy. Images of CT scan are having more knowledge rather than any other images. Since trying with series of images to train the set and the output is stored in a particular file. Sometimes it is difficult to detect in early but using CAD we can detect it easily.

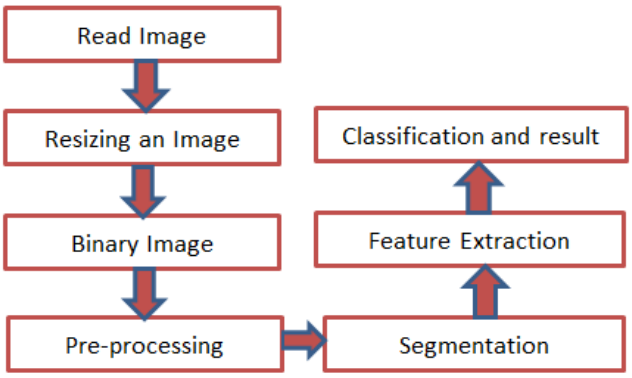

Fig-1: Architecture of proposed system

Thus to remove noise and undesired region the images are subjected to pre-processing steps. First fetch the input image from database. Resize the image according to size acceptable by the system to process. Convert resized image into gray scale image so that it uses only one color channel. Comparison in gray scale involves simple scalar algebraic operators. Gray scale image is sufficient to distinguish intensity peaks. After gray scale is converted into binary image. That Binary image is a digital image with two possible values for each pixel. This removes noise, holes and other impurities in the image one the system architecture is one of the most important steps for analyzing the complete flow of the project. In this architecture, as shown in Fig- 1 initially we are reading the input image from the databases. Databases are publically available in hospitals, collected for real time situations. By using this database we are reading image from the dataset. After acquiring an image next thing is to resizing it. Because, every image is having different size to make it with a same size we are resizing it. After resizing an input image those are converting it to gray scale image. So that features can be accurately fetch. Next step is image enhancement as a collected database images may contain some noise, blur or any impurities, so in order to remove all these we are using image enhancement technique. In our proposed CAD system, we extracted the following features: NC ratio, perimeter, density, curvature, and circularity and Eigen ratio. The first feature is NC ratio, which is computed by dividing the nucleus area (total number of the pixels in the nucleus region) over the cytoplasm area (total number of pixels in the cytoplasmic region)

NC ratio = (Area (Nucleus))/ (Area (cytoplasm))*100 (1)

Therefore, based on medical information the morphology, the size, and the growing correlation of the nuclei and its corresponding cytoplasm regions reflect the diagnostic

situation of the cell life cycle. The second feature is the nucleus perimeter defined by:

P (Nucleus) = ∫t√(x2(t) +y2 (t) dt) (2)

Where x (t) and y (t) are the parameterized contour point coordinates The third feature which represents the density and it is based on the darkness of the nucleus area after staining it with a certain dye, thus it is based on the mean value of the nucleus region. The mean value represents the average intensity value of all pixels that belong to the same region, and in our case, each mean value is represented as a vector of RGB components, and it is calculated as follows for a given nucleus.

Mean(Nucleus)=(∑Ni=1(Intensity(i))/(Area(Nucleus))) (3)

Where i is the intensity color value and N is the total number of the pixels in the nucleus area. The curvature at a single point in the boundary can be defined by its adjacent tangent line segments. The difference between slopes of two adjacent straight line segments is a good measure of the curvature at that point of intersection, the slope is defined by:

Φ (t) =tan-1 ((y (t))/(x (t))) (4)

Where y (t) and x (t) denote the derivative of x (t) and y (t). The fifth feature is the circularity, which is a feature that describes the roundness of the nucleus, and it is defined as:

Circularity = (4πNucleus (A))/p2 (5)

Nucleus (A) is the nucleus area; P is the perimeter of the nucleus area. Cells in cleavage are normally round, so their roundness value will be higher. On the other hand, cancerous growing cells are irregular so their roundness value will be lower.

Computer Aided Diagnosis (CAD) is a proven clinical tool that helps the physician for detecting deadly diseases. These systems are said to be second opinion for diagnostic results. CAD is nothing but feature identification system which makes the radiologist to reduce false identification. In order to finalize the diagnostic result, The Radiologist activates CAD system and makes evaluation based on result. CAD system has its enormous impact in the medical field as it provides the hospital interpretation in the detection of the deadly diseases. CAD techniques attracted by the different fields. The main motto of the development of CAD systems is to enhance the performance for most widely using the system in medical Industry.

The performance and accuracy of a computer-aided cancer detection and diagnosis system needs an integration of many independent and self-contained systems to get the best performance for the result. The in-depth knowledge of the subsystems containing the feature extraction and classification systems are necessary to obtain the improved performance of the whole system to improve the efficiency of radiologists in the detection of cancerous nodules, they are not widely used in clinical aspects. As a result, CAD systems have become one of the most attracting area or field of research in medical image processing. Radiology field consisting of visual perspective of an image and description or cognition. The CAD algorithms need a digital data set of the image for analysis and calculations. The medical images are usually in Digital Imaging and Communications in Medicine (DICOM) format compatible and reliable with the open source software.

4. RESULTS AND ANALYSIS

Initially the normal image is given as input then it is converted into binary image after preprocessing. As shown in Fig- 2

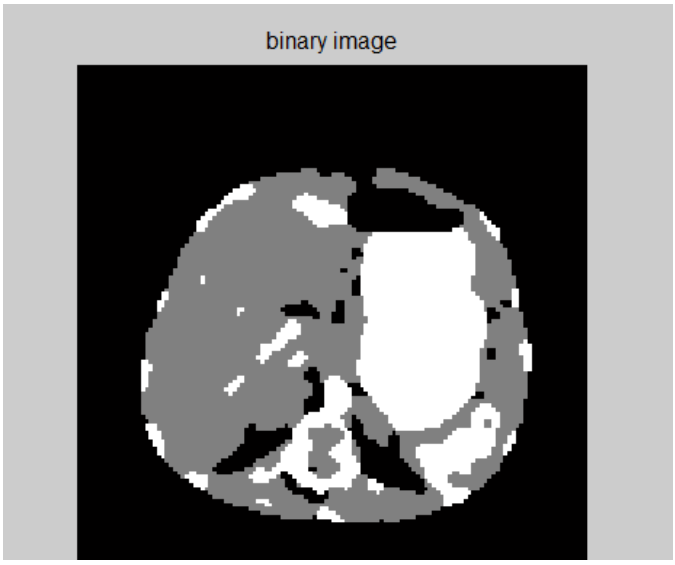

Fig-2: Normal Image is converted to binary Image

The binary image is then segmented to differentiate nucleus and cytoplasm as shown in Fig- 3.

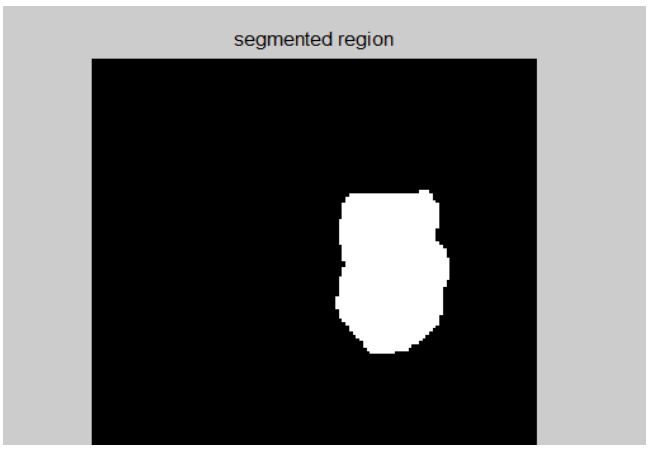

Fig-3 Image segmentation to extract nucleus and cytoplasm.

When nucleus is differentiated the features are calculated based on the circularity, mean and standard deviation so as to evaluate the cancer detection parts, finally the results are displayed with red line indicating the cancer affected region as shown in Fig-4.

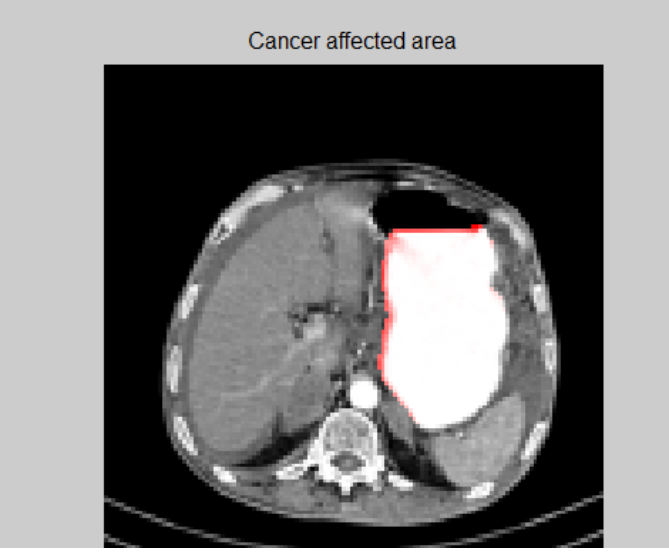

Fig-4 Final output displaying red colour as cancer region

5. CONCLUSION

In this paper, we presented a technique for early detection of lung cancer. The technique is based on a feature extraction followed by SVM classification. A set of different features such as NC ratio, perimeter, density, curvature, circularity and Eigen ratio were extracted from the nucleus region to be used in the diagnostic process. The experimentation shows the potential of the rule based method after applying it to the extracted features The preliminary results of the CAD system have yielded promising results that would supplement the use of current technologies for diagnosing lung cancer, where it can be used to monitor treatment effectiveness.

REFERENCES

[1] "Cancer Facts & Figures 2015." [Online]. Available: http://www.cancer.org/research/cancerfacts&figures/cancerfactsfigures/cancer-facts-figures-2015.

[2] F. B. J. M. Thunnissen, "Sputum examination for early Detection of lung cancer," J. Clin. Pathol. vol. 56, no. 11, pp. 805–810, Nov. 2003.

[3] I. Sluimer, A. Schilham, M. Prokop, and B. Van Ginneken, "Computer analysis of computed tomography scans of the lung: a survey," IEEE Trans. Med. Imaging, vol. 25, no.4, pp. 385–405, 2006.

[4] S. Dua, V. Jain, and H. W. Thompson, "Patient classification using association mining of clinical images," in 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2008, 2008, pp.253–256.2008

[5] F. Taher, N. Werghi, and H. Al-Ahmad, "Comparison of Hopfield Neural Network and Mean Shift algorithm in Segmenting Sputum Color Images for Lung Cancer Diagnosis," in 20th IEEE International Conference on Electronics, Circuits, and Systems (ICECS), Abu Dhabi, UAE, pp. 649–652.2013

[6] F. Taher, N. Werghi, H. Al-Ahmad, and C. Donner, "Extraction and Segmentation of Sputum Cells for Lung Cancer Early Diagnosis," Algorithms, vol. 6, no. 3, pp. 512–531, Aug. 2013.