Image Processing and Classification Techniques for Early Detection of Lung Cancer for Preventive Health Care: A Survey

Prashant Naresh¹, Dr. Rajashree Shettar²
Dept. of Computer Science and Engineering, RVCE, Bangalore¹
Dept. of Computer Science and Engineering, RVCE, Bangalore²
prashantaresh1988@gmail.com¹, rajashreeshettar@rvce.edu.in²

Abstract: Prediction and diagnosis of cancer is very important and crucial task in healthcare industry. Cancer is of the leading cause of death globally so it is very essential to detect and predict it at very initial stage. In this paper we presented an overview of all the existing techniques to detect lung cancer at initial stage. Data Mining and Image processing plays very crucial role in healthcare industry specially for disease diagnosis. Data Mining is very beneficial for finding hidden information or pattern form the huge databases, some widely used data mining techniques are classification, prediction, association analysis, pattern matching and clustering. Image Processing plays significant role in cancer detection when input data is in the form of images, some techniques used in Image Processing for information retrieval are Image acquisition, Noise Removal, Segmentation, and Morphological operations etc. In this paper a study is presented in which an analysis is provided for the diagnosis and prediction of lung cancer at initial stage using Image Processing and Data Mining techniques. For detecting lung nodules number of tests should be required from the patient. But automated diagnosis system for prediction of lung cancer by using image processing and data mining techniques, plays an important role in time and performance which decreases mortality rate because of early detection of lung cancer. Different hybrid techniques provide different accuracy and sensitivity. This research paper analyzes various image processing and classification techniques and their efficiency used for predicting lung cancer.

Keywords: Image Acquisition, Noise Removal, Segmentation, Morphological Operation, Lung Cancer.

I. INTRODUCTION

Lung cancer is considered to be the main cause of cancer death worldwide, and in its early stages it is difficult to detect because only in the advanced stage symptoms appear causing the mortality rate to be the highest among all other types of cancer [1]. If lung nodules can be identified accurately at an early stage, the patient’s survival rate can be increased by a significant percentage. The rank order of cancers for both males and females among Jordanians in 2008 indicated that there were 356 cases of lung cancer accounting for (7.7 %) of all newly diagnosed cancer cases in 2008 [2]. Lung cancer affected 297 (13.1 %) males and 59 (2.5%) females with a male to female ratio of 5:1 which Lung cancer ranked second among males and 10th among females [2]. In the modern age of computerized fully automated trend of living, the field of automated diagnostic systems plays an important and vital role.
Automated diagnostic system designs in Medical Image Processing are one such field where numerous systems are proposed and still many more under conceptual design due explosive growth of the technology today [3].

The process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics is known as Image segmentation. Literature has a wide range of segmentation techniques used in lung cancer diagnosis. Image processing has wide scope in medical image processing for diagnosing the Lung cancer [1]. The purpose of this paper is to review related work on automatic diagnosis of lung cancer and summarization various segmentation and classification techniques for detection of lung nodules.

The paper is organized in four sections. Section II covers literature pertaining to image processing and classification algorithms, and applications of these techniques for lung cancer diagnosis. Section III provides summary of lung nodule detection and classification. Section IV covers the conclusion.

II. RELATED WORKS

Disha Sharma et.al [4] develops an automatic CAD system for early detection of lung cancer by analyzing LUNG CT images using several steps. First, extracting the lung regions from the CT image using several image processing techniques, including bit image slicing, erosion, and Weiner filter. Bit plane slicing technique is used in the first step in the extraction process to convert the CT image into a binary image. After extraction, the extracted lung regions are segmented using region growing segmentation algorithm. Then rule based technique is applied to classify the cancer nodules. Finally, a set of diagnosis rules are generated from the extracted features and by help of diagnostics indicator achieves accuracy of 80%.

Samuel Cheng et.al [5] presented a nonlinear anomaly detector called kernel RX-algorithm and apply it to CT images for malignant nodule detection. Simple watershed segmentation and additional speed up tricks have also been adopted to decrease the computational complexity of the algorithm. CT images are obtained from an on LIDC-IDRI [6] annotated files dataset. The proposed approach can be an efficient technique for lung cancer detection at early stage.

Anam Tariq et.al [7] proposed a computerized system for lung nodule detection in CT scan images. The system consists of two stages i.e. lung segmentation and enhancement, feature extraction and classification. Threshold segmentation is applied to remove background and extracts the nodules from an image. A feature vector for possible abnormal regions is calculated and regions are classified using neuro fuzzy classifier. System facilitates the detection of small nodules which lead to early diagnosis of lung cancer.

Anita chaudhary et.al [8] aim to get the more accurate results by using various enhancement and segmentation techniques. MATLAB have been used through every procedures made. In image processing procedures, process such as image pre-processing, segmentation and feature extraction discussed in detail. Compare Gabor filter, auto-enhancement and Fast Fourier transform techniques, used for image enhancement. In the segmentation stage the Watershed and Thresholding Segmentation is used and comparison has been made.

Atiye Hashemi et.al [9] aim at presenting a method to improve the efficiency of the lung cancer diagnosis system, through proposing a region growing segmentation method to segment CT scan lung images. Linear-filtering and contrast enhancement used as preprocessing step for noise removal, to prepare the image for segmentation. Afterwards, cancer recognition are presenting by Fuzzy Inference System (FIS) for differentiating between malignant, benign and advanced lung nodules. The diagnostic performances of FIS system is comparing by using artificial neural networks (ANNs).

Dansheng Song et.al [10] provides an objective and quantitative assessment of centrosomal numeral and morphological abnormalities and shows the magnitude of these differences. Regions of interest were selected to include one cell and its centrosomes. After segmentation, feature abstraction, and optimization, six non redundant
features were used for statistical analysis and classification. Linear discriminant analysis (LDA) and support vector machines (SVM) with 10-fold cross validation used for classification and gets an accuracy of 85%.

S.K. Vijai Anand et.al [11] proposed a system which efficiently predicts lung tumor from Computed Tomography (CT) images through image processing techniques coupled with neural network classification as either benign or malignant. Optimal thresholding is applied to the denoised image to segregate lung regions. Region growing method is used to segment Lung nodules which are of relatively high density found within the lung regions. A set of textural features extracted from the extracted ROIs is classified as cancerous or non-cancerous by the back propagation neural network.

Yang Liu et.al [12] proposed a two stage scheme for automatic lung nodules detection in Multi-Slice Computed Tomography (MSCT) scans with multiple SVMs. Multi SVM is used to reduce no. of false positive lung nodules. Three SVMs classifiers for different slice directions are used on preprocessed images to categorize the candidates as nodule or non-nodule.

Aparna Kanakatte et.al [13] presented an automated process of tumor delineation and volume detection from each frame of PET lung images. Data is represented by using spatial features (geometric moments) and frequency domain features (discrete cosine transform, wavelets) and comparison has been made. K-nearest neighbor and support vector machines (SVM) classifiers were used to analyze performance of these features. This procedure facilitates the physicians in accurate staging and radiotherapy treatment planning for lung tumors.

S.L.A. Lee et.al [14] presented a method that includes three stages: image acquisition, background removal and nodule detection for identification of lung nodules. Images selected from the LIDC lung databases. The system was tested on few images containing nodules and few containing no nodules which were randomly selected from the database images. Random forest based classifier performs well to detect all the nodules in the images and recorded a low false detection rate.

Yang Liu et.al [15] presented a computer aided lung nodule detection scheme based on analysis of enhanced voxel in three dimensional (3D) CT image and evaluated the performance of the scheme on two CT data sets. SVM classifier is applied to categorize the initial nodule candidates as nodule or non-nodule. Input to SVM classifier is eight features extracted from 3D initial nodule candidate voxels. 93.75% sensitivity with 4.6 FPs per case has been achieved by combining SVM classifier and a decision rule. The performance of a lung CAD system can be improved by analyzing the 3D voxel characteristic.

Fan Zhang et.al [16] presented a feature-based imaging classification method to classify the lung nodules in low dose computed tomography (LDCT) slides into four categories: well circumscribed, vascularized, juxta-pleural and pleural-tail. SVM classifier is used to conduct the classification. Specifically, a four-type SVM is trained with polynomial kernel by C-SVC from [17], and the probability estimates upon the different types are predicted with the obtained SVM model, which is used to classify the feature descriptors into four categories.

S.Sivakumar et.al [18] develop an efficient lung nodule detection scheme by performing nodule segmentation through weighted fuzzy possibilistic based clustering is carried out for lung cancer images. Support Vector Machine (SVM), a machine learning technique is used for classification. The RBF kernel based SVM classifier performs better than linear and polynomial kernel based classifier.

M. Arfan Jaffar et.al [19] describes a method for lung segmentation based on Genetic Algorithm (GA) and morphological image processing techniques. GA is applied on the normalized histogram determine the threshold to separate out background and object. After background removal morphological operation is performed in three operations: to filter noise, to smooth the image, to detect edges. Susan thinning algorithm [20] is used to reduce the borders to the width of one pixel.
JIA Tong et.al [21] an automatic computer-aided detection (CAD) scheme is presented that can identify the pulmonary nodule at an early stage from CT images. An improved optimal gray-level threshold segment lung parenchyma is an important pre-processing step of lung cancer computer-aided detection system. Region-growing algorithm is proposed for Trachea and Main Airway Bronchi Elimination. Then, the filter of nodule candidates, the detection of nodule candidates, the feature extraction and classification, three-dimensional visualization is done. Results of this CAD system having accuracy 95%, which indicate good performance when compared with physician diagnosis.

Hiram Madero Orozco et.al [22] presented a very simple but efficient methodology for lung nodule classification without the stage of segmentation. Eight texture features were extracted from the histogram and the gray level co-occurrence matrix (with four different angles) after image acquisition for each CT image. Support vector machine (SVM), used to classify lung tissues into two classes: with lung nodules and without lung nodules. The better reliability results were obtained with 90° and 135° of the GLCM.

Fatma Taher et.al [23] presented a Bayesian classification and a Hopfield Neural Network algorithm for extracting and segmenting the sputum cells for the purpose of lung cancer early diagnosis. The HNN segmentation algorithm outperforms the Fuzzy C-Mean clustering, it allows the extraction of nuclei and cytoplasm regions successfully. Morphological processing on the segmented image improved the performance of HNN algorithm.

Kesav Kancherla et.al [24] proposed an early lung cancer detection methodology using nucleus based features. Seeded region growing segmentation method is used to perform nucleus segmentation. An additional criterion like nucleus size to seeded region growing method is used for better accuracy.

Negar Memarian et.al [25] proposed a novel classification method called iterative linear discriminant analysis and use this method in addition to fuzzy c-means clustering for successful false positive reduction. Estimation of number of false positives in the fuzzy c-means clustering module for every patient is followed by classification of potential nodules in the novel iterative LDA module. Effective FP reduction is achieved. The novel scheme is superior over rule-based FP reduction of detected candidates.

III. SUMMARY OF LUNG NODULE DETECTION AND CLASSIFICATION WORK

Image processing and data mining together have shown significant improvement in medical industry in terms of prediction and decision making of lung cancer. Table 1 gives the summary of image processing and classification work, accuracy and sensitivity of various techniques.
Table 1. Summary of Lung Nodule Detection and Classification Works

<table>
<thead>
<tr>
<th>Author</th>
<th>Images</th>
<th>Technique</th>
<th>Year</th>
<th>Accuracy</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disha Sharma [4]</td>
<td>CT</td>
<td>Edge detection(Sobel)</td>
<td>2011</td>
<td>80%</td>
<td>NA</td>
</tr>
<tr>
<td>Anam Tariq [7]</td>
<td>CT</td>
<td>Threshold Segmentation</td>
<td>2013</td>
<td>95%</td>
<td>NA</td>
</tr>
<tr>
<td>Atiyeh Hashemi [9]</td>
<td>CT</td>
<td>Region Growing</td>
<td>2013</td>
<td>NA</td>
<td>95%</td>
</tr>
<tr>
<td>Dansheng Song [10]</td>
<td>CT</td>
<td>Entropy Threshold</td>
<td>2012</td>
<td>85%</td>
<td>NA</td>
</tr>
<tr>
<td>Yang Liu [12]</td>
<td>CT</td>
<td>Bounding box + Threshold Segmentation</td>
<td>2009</td>
<td>87.82%</td>
<td>93.75%</td>
</tr>
<tr>
<td>S.Sivakumar [18]</td>
<td>CT</td>
<td>Weighted fuzzy probabilistic based clustering</td>
<td>2013</td>
<td>80.36%</td>
<td>82.05%</td>
</tr>
<tr>
<td>JIA Tong [21]</td>
<td>CT</td>
<td>Optimal gray level threshold</td>
<td>2007</td>
<td>NA</td>
<td>95%</td>
</tr>
<tr>
<td>Hiram Madero Orozco [22]</td>
<td>CT</td>
<td>NA</td>
<td>2013</td>
<td>84%</td>
<td>NA</td>
</tr>
<tr>
<td>Fatma Taher [23]</td>
<td>Sputum</td>
<td>Hopfield Neural Network(HNN)</td>
<td>2012</td>
<td>88.62%</td>
<td>NA</td>
</tr>
<tr>
<td>Kesav Kancherla [24]</td>
<td>Sputum</td>
<td>Seeded region growing</td>
<td>2013</td>
<td>87%</td>
<td>NA</td>
</tr>
<tr>
<td>Negar Memarian [25]</td>
<td>CT</td>
<td>NA</td>
<td>2006</td>
<td>NA</td>
<td>80.80%</td>
</tr>
</tbody>
</table>

CONCLUSION

In this paper an analysis is presented for different data mining and Image processing techniques which can be employed in automated lung cancer prediction system. Many existing techniques have been employed in recent years for the prediction of lung cancer at initial stage. In this paper an overview of different algorithm for classification and Image processing used in the field of lung cancer prediction is presented. The main focus is on using different classifiers combined with different segmentation algorithms for nodule detection in lung using image processing. The summary of various segmentation and classification techniques with their classification accuracy and sensitivity of nodule detection has been presented. From the study it has been found that for CT scan the most accurate result i.e. 95% is achieved by use of threshold segmentation with Neuro fuzzy classifier, and with PET images higher accuracy of 97% is achieved by use of k-NN and SVM classifier.

REFERENCES


