PREDICTION OF LUNG CANCER USING IMAGE PROCESSING TECHNIQUES: A REVIEW

Arvind Kumar Tiwari
GGS College of Modern Technology, SAS Nagar, Punjab, India

ABSTRACT

Prediction of lung cancer is most challenging problem due to structure of cancer cell, where most of the cells are overlapped each other. The image processing techniques are mostly used for prediction of lung cancer and also for early detection and treatment to prevent the lung cancer. To predict the lung cancer various features are extracted from the images therefore, pattern recognition based approaches are useful to predict the lung cancer. Here, a comprehensive review for the prediction of lung cancer by previous researcher using image processing techniques is presented. The summary for the prediction of lung cancer by previous researcher using image processing techniques is also presented.

KEYWORDS

Classification, lung cancer, accuracy, image processing techniques

1. INTRODUCTION

Lung Cancer is the uncontrolled growth of abnormal cells, start off in one or both lungs, usually in the line the air passages. The abnormal cells do not develop into healthy lung tissue; they provide rapidly and form tumours. According to American cancer society the cases of lung cancer increases very rapidly and almost 14% newly diagnosed cancers are a lung cancer and also the main cause of cancer death worldwide. The previous study of diagnosis showed that the most of the lung cancer patients belongs to the age of 60 years.

Lung Cancer is one of the most serious human body problems in the world. The death rate of lung cancer is highest of all other types of cancer. The survival rate of lung cancer is very smallest among all types of cancer. So, there is a need to design a computational intelligence based approaches to detect the lung cancer because the survival from lung cancer is directly related to its growth at its detection time. If we detect lung cancer at early stage, then there are more possibilities to survive the patients. It is also showed from previous study that cigarettes smoking are the main cause of lung cancer. It is observed that an estimated 85% of lung cancer cases in males and 75% lung cancer cases in females where cigarette smoking is the main reason

2. LITERATURE SURVEY

In literature various researchers have been used image processing techniques to predict the lung cancer. Sharma et.al. (2011) used lung CT images extracted from NIH/NCI Lung Database Consortium and proposed an automatic computer aided diagnosing system for detection of lung cancer by analyzing these lung CT images. The authors of the paper have used several steps for the detection of lung cancer. Firstly, they extracted the lung region from the computer tomography image using various image processing techniques such as bit image slicing, erosion and wiener filter. In the first step the bit image slicing technique was used to convert the CT
images into a binary image then after extraction the region growing segmentation algorithm was used for segmenting the extracted lung regions. After segmentation of lung region they used rule based model to classify the cancer nodules. Finally, a set of diagnosis rules were generated from the extracted features and with the help of diagnostics indicator. It was observed that the proposed method achieved the overall accuracy of 80%. Anam Tariq et al. (2013) has developed a computerized system, that was detected the lung nodules with the help of CT scan images. The computerized system consists of two stages, first one is lung segmentation and enhancement and second one is feature extraction and classification. For removing background and extracts the nodules from an image, the threshold segmentation technique was applied. When extraction and segmentation were completed, then a feature vector was used to calculate the abnormal region. After that the regions were classified using neuro fuzzy classifier. This system gives the facility to detect the smallest nodules which lead to early diagnosis of lung cancer.

Sundararajan et al. (2010) have been proposed support vector machine for the detection of pneumoconiosis by using various textural features for disjoint segments of the lung and focused on a subset of lung disorders. Le (2011) has been proposed a complex system for the detection of various lung disorders and image processing techniques were used for the detection of lung nodule but they were not used false positive reduction step. Chaudhary et al. (2012) aim has to get more accurate results using the various enhancement and segmentation techniques. The image processing procedures that is, image pre-processing; segmentation and feature extraction were done. In image enhancement, image are compared with Gabor filter, auto enhancement and fast Fourier transform techniques. In the segmentation stage the Watershed and thresholding segmentation were used and comparison has been made. Hashemi et al. (2013) his aim has to improve the efficiency of the lung cancer diagnosis system, with the help of proposing a region growing segmentation method to segment CT scan lung images. First of all, for noise removing linear-filtering and contrast enhancement technique was used as pre-processing step, to prepare the image for segmentation. After that for differentiating among malignant, benign and advance lung nodules, the cancer recognition was presented by fuzzy inference system. The authors also compare the diagnosis performance of the proposed method with the artificial neural network.

Schilham et al. (2006) have been proposed to use a k-nearest neighbor (k-NN) classifier by using the images from the JSRT database for searching the Gaussian scale space maxima in multi-scale to detect different sized lung nodules. It was also observed that the proposed method performed better only at higher false positive rates. Pereira et al. (2007a) has been proposed a sliding band filter based on the convergence of radial gradients to detect the lung nodule by using the JSRT database without a false positive reduction step. S.K. Vijai Anand et al. (2010) have proposed a system which was predicted lung tumor from CT images with the help of image processing techniques coupled with neural network classification, to recognize whether benign or malignant. Then optimal thresholding was applied for noise removal to segregate lung regions. To segment the lung nodules region growing method was used. After that, back propagation neural network was used to classify as cancerous or non-cancerous. S.L.A. Lee et al. (2008) have been proposed a random forest based classifier to detect all the nodules in the images and recorded a low false detection rate. The proposed method includes three stages such as image acquisition, background removal and nodule detection for identification of lung nodules by using images from the LIDC lung databases. The performance of the system was observed on some images containing nodules and some images containing no-nodules that were normally selected from database images.

Vijay A. Gajdhane et al. (2014) have been proposed detection of lung cancer by using various image processing techniques with the help of CT scan images. There were mainly three processes used throughout the report; pre-processing step, feature extraction step, and finally the classification step. In pre-processing step noise was removed with the help of median filter, and Gabor filter was used for image enhancement and segmentation. Area, perimeter, and eccentricity were the extracted features. The extracted ROI (Region of Interest) transformation function was
created an array of ROI. At last, Support Vector Machine (SVM) was used for classification purpose. Fan Zhang et al. (2013) have been proposed Support Vector Machine (SVM) based classifier by using feature based imaging classification method to classify the lung nodules in Low Dose Computed Tomography (LDCT) slides into four categories that is, well circumscribed, vascularized, juxta-pleural and pleural-tail. The proposed support vector machine based method was trained with the help of polynomial kernel by C-SVC and the probability estimates upon the different types were predicted with the obtained SVM model which was used to classify the feature descriptors into four categories. Lo et al. (1995) applied a matched filter and sphere template matching to enhance round-shaped objects, and filtered false results with a convolutional neural network (CNN). They measured good performance on a small set of images. S. Sivakumar et al. (2013) have proposed an efficient lung nodules detection scheme by performing nodule segmentation. For nodule segmentation weighted fuzzy probabilistic based clustering was carried out for lung cancer images. SVM was used for classification purpose. Iiya Levner et al. (2007) have proposed the novel approach that was created the topographical function and object markers used within watershed segmentation. The watershed algorithm was commonly used within the unsupervised setting of segmenting the image into a set of non-overlapping regions. Yoshida et al. (1995) have been proposed a combination of two enhancers such as a difference image algorithm and wavelet transform sensitive to subtle nodules. The proposed method integrated overlapping and lesion models were used in combination. It was observed that the proposed model obtained better result in compared to single algorithm without any incorporation of false positive reduction technique. Sunil Kumar et al. (2014) have proposed Computer aided diagnosis (CAD) that was detected the lung cancer at early stage by using various image processing techniques with the help of CT scan images. In pre-processing step noise was removed and converted the image into gray scale. Histogram thresholding was used for thresholding purpose. Finally, region growing algorithm was used for classification purpose. Xu et al. (1997) proposed another difference image algorithm. For false positive reduction, they created a hybrid system with the help of an artificial neural network (ANN) and a rule-based technique. M. Arfan Jaffar et al. (2008) have presented a method for lung segmentation based on Genetic Algorithm (GA) and morphological image processing techniques. GA was applied on the normalized histogram to determine the threshold to separate out background and object. When background was removed the morphological operation was performed in three operations i.e. to filter noise, to smooth the image and to detect edges.

Giger et al. (1990) have been proposed a difference image technique by using image processing techniques to create a nodule enhanced and a nodule suppressed image. They found that the proposed method eliminates everything remaining the shadow of nodule-like objects. They used less powerful methods for classification by using various geometry features to leave out false findings. Hiram Madero Orzczo et al. (2013) have been proposed a SVM based method by using eight texture feature that were extracted from histogram and the gray level concurrence matrix to classify the image with nodules or without nodules without considering the segmentation stage. Fatma Taher et al. (2012) have been proposed a combination of Bayesian classification and a Hopfield Neural Network (HNN) for extracted and segmented the sputum cells for a diagnosis of lung cancer. It was observed that The HNN segmentation algorithm outperforms the Fuzzy C-Mean clustering, that allow the extraction of nuclei and cytoplasm regions successfully. For improved the performance of HNN algorithm, morphological was processing on the segmented image. Kesav Kancherla et al. (2013) have proposed an early lung cancer detection methodology with the help of Nucleus Based features. To perform nucleus segmentation the seeded region growing segmentation technique was used. An additional criterion like nucleus size to seeded region growing method was used for better accuracy. JIA Tong et al. (2007) have been presented an automatic computer-aided detection(CAD) scheme, that could identify the pulmonary nodule at an every stage from CT images. An important pre-processing step of lung cancer was an improved optimal gray-level Threshold segment lung parenchyma. Region growing algorithm was proposed for Trachea and Main Airway Bronchi Elimination then the filter of nodule
candidate, the detection of nodule candidate, the feature extraction and classification and three-
dimensional visualization was done. The accuracy of the CAD system was 95% which indicate
good performance.

Penedo et.al. (1998) have been proposed a neural network based method for nodule enhancement
by using another ANN to select only true findings and the results were presented on a moderate-
sized, private database extended by simulated nodules. Mao et.al. (1998) has evaluated the
nodule enhancing capabilities of fragmentary window filtering. The authors were presented
preliminary results on a few generated radiographs M. Gomathi & P. Thangaraj et.al. (2010) have
been proposed a support vector machine based (CAD) system for automatically detection of the
lung cancer nodules with reduction in false positive rates. They used various image processing
techniques to obtain the lung regions from CT scan images such as the segmentation was carried
to apply fuzzy possibility C Mean (FPCM) clustering algorithm. Amjed S. Eslam b Jaber et.al.
(2014) have been proposed an automatic computer aided diagnosis (CAD) system that includes
three steps such as thresholding the CT image by segmentation, labeling the founded regions and
regions are extracted for further analysis for the detection of lung cancer using CT scan images.

Ghayoumi H, Siamak J et.al. (2013) have been proposed an image analysis approach for
automated detection, segmentation and classification of abnormal cancer cells from normal cells
was introduced. Gaussian smoothing was used for detection and segmentation. At last Fast
Fourier Transforms (FFT) was used for extracted the position of the objected cells. Patil,
Kuchanur et.al. (2012) have been proposed a system to identify the lung nodules by extracted the
various geometrical and textural features like area, perimeter, entropy etc. using various image
processing techniques. When geometrical and textural features were extracted then these features
were applied for the classification of lung cancer.

Muhm et. al. (1983), Quekel et. al. (1999), Shah et. al. (2003) and Doi (2007) mentioned that
many of the tumors were overlooked by radiologists in the early stage. Prashant Naresh et.al
.(2014) have predicted lung cancer at initial stage by using various image processing and data
mining techniques. He was proposed an automated diagnosis system for the prediction of lung
cancer. Image acquisition, noise removal, segmentation, and morphological operations were used
and De Boo et.al. (2009) studied that the efficiency of lung screening can be improved by using
a computer aided diagnosis system, while Hoop et.al. (2010) warns that further training of the
users may be important, as they often find it hard to differentiate between true and false findings
of the machine. Ada (2013) has been proposed a method that is based on the combination of
feature extraction and Principal Component Analysis (PCA) for lung detection with the help of
CT scan images by using the histogram equalization for per-processing step of the images and
features was extracted with the help of PCA. Sudha. V, Jayashree. P et.al. (2012) have been also
proposed an efficient lung nodule detection system by using CT scan images. The proposed
method performed two steps; first of all lung region segmentation through thresholding and then
thresholding and morphological operations were used segmenting the lug nodules. Weng et.al.
(2009) have been proposed a method based on a difference image technique and SVM
classification by using the statistics of gradients as features but they also used JSRT image
database for testing. So, they obtained the performance of the proposed method on the merged
database. Snoeren et.al. (2010) have been proposed a neural network based on gradient
convergence based method and location features to train a proposed model with simulated
nodules without using a false positive reduction step.
3. **Summary For The Classification Of Lung Nodule**

In Literature, it is observed that the image processing techniques with the computational intelligence based approaches are useful for the prediction and decision making of lung cancer. Table 1 gives the summary of image processing techniques and classification approaches with their performance analysis for the detection of lung nodules. The summary of lung nodule detection and classification is presented in Table-1.

<table>
<thead>
<tr>
<th>AUTHOR</th>
<th>IMAGES</th>
<th>CLASSIFICATION TECHNIQUE</th>
<th>ACCURACY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disha Sharma</td>
<td>CT</td>
<td>Diagnostic Indicators</td>
<td>80%</td>
</tr>
<tr>
<td>Anam Tariq</td>
<td>CT</td>
<td>Neuro-Fuzzy</td>
<td>95%</td>
</tr>
<tr>
<td>F. Paulin, Dr. A. Santhakumaran</td>
<td>CT</td>
<td>Back propagation algorithm used for training Multilayer Perceptron(MLP)</td>
<td>99.28%</td>
</tr>
<tr>
<td>Yang Liu</td>
<td>CT</td>
<td>SVM(GrBF kernel type)</td>
<td>87.82%</td>
</tr>
<tr>
<td>Yao ying huang, Wang sen Li, Xiaojiao ye</td>
<td>CT</td>
<td>Genetic algorithm, feature selection</td>
<td>99.1%</td>
</tr>
<tr>
<td>Dr. K. Usha rani</td>
<td>CT</td>
<td>Feed forward, Back propagation</td>
<td>92%</td>
</tr>
<tr>
<td>Afzan Adam</td>
<td>CT</td>
<td>Genetic algorithm and Back propagation neural network</td>
<td>83.36%</td>
</tr>
<tr>
<td>S.K Vijai Anand</td>
<td>CT</td>
<td>Back propagation network classification</td>
<td>86.30%</td>
</tr>
<tr>
<td>David B. Fogel</td>
<td>CT</td>
<td>Back propagation network classification</td>
<td>98.2%</td>
</tr>
<tr>
<td>JR Marsilin</td>
<td>CT</td>
<td>SVM</td>
<td>78.00%</td>
</tr>
<tr>
<td>Li Rong, Sunyuan</td>
<td>CT</td>
<td>SVM-KNN classifier</td>
<td>98.06%</td>
</tr>
<tr>
<td>F. Eddaoudi</td>
<td>CT</td>
<td>SVM</td>
<td>95%</td>
</tr>
<tr>
<td>Aparna Kanakatte</td>
<td>PET</td>
<td>k-NN, SVM</td>
<td>97%</td>
</tr>
<tr>
<td>S. Aruna, Dr. S.P. Rajagopalan</td>
<td>CT</td>
<td>SVM</td>
<td>98.24%</td>
</tr>
<tr>
<td>S. Sivakumar</td>
<td>CT</td>
<td>SVM(RBF kernel type)</td>
<td>80.36%</td>
</tr>
<tr>
<td>Hiram Madero Orozco</td>
<td>CT</td>
<td>SVM</td>
<td>84%</td>
</tr>
<tr>
<td>Fatma Taher</td>
<td>Sputum</td>
<td>Bayesian</td>
<td>88.62%</td>
</tr>
<tr>
<td>Kesav Kancherla</td>
<td>Sputum</td>
<td>Random forest(bagging)</td>
<td>87%</td>
</tr>
<tr>
<td>Tuba Kiyano</td>
<td>CT</td>
<td>Radial basis function</td>
<td>96.81%</td>
</tr>
</tbody>
</table>
4. **Observation and Recommendation**

- Using CT images the SVM classification technique achieved accuracies between 78% to 98.24%.
- Using CT images the Back Propagation Network classification technique achieved accuracies between 86.30% to 99.28%.
- Using CT images the Neuro Fuzzy classification technique achieved accuracy 95%.
- Using Sputum the Bayesian classification technique achieved accuracy 88.62%.
- Using CT images the Genetic Algorithm classification technique achieved accuracies between 83.36% to 99.10%.

5. **Conclusion**

Prediction of lung cancer is most challenging problem due to structure of cancer cell, where most of the cells are overlapped each other. The image processing techniques are mostly used for prediction of lung cancer and also for early detection and treatment to prevent the lung cancer. To predict the lung cancer various features are extracted from the images therefore, pattern recognition based approaches are useful to predict the lung cancer. Here, a comprehensive review for the prediction of lung cancer by previous researcher using image processing techniques was presented. The summary for the prediction of lung cancer by previous researcher using image processing techniques was also presented.

**References**

7. Machines. In Artificial Intelligence (MICAI), 2013 12th Mexican International Conference on (pp. 277-283). IEEE.


18. Fatma Taher, Naoufel Werghi and Hussain Al-Ahmad (2012), “Bayesian Classification and Artificial Neural Network Methods for Lung Cancer Early Diagnosis”, IEEE.


47. Afzan Adami, Khairuddin Omar2 “Computerized Breast Cancer Diagnosis with Genetic Algorithms AndNeuralNetworkfitt.mmu.edu.my/caic/papers/afzaniCAIET
56. Fatma Taher, Naoufel Werghi and Hussain Al-Ahmad (2012), “Bayesian Classification and Artificial Neural Network Methods for Lung Cancer Early Diagnosis”, IEEE.

Author

Dr. Arvind Kumar Tiwari received the BE degree in Computer Science and Engineering from CCS university, Meerut, INDIA in 2003, and M. Tech. in Computer Science and Engineering from Uttar Pradesh Technical University, Lucknow and Ph.D. in Computer Science and Engineering from IIT (BHU), Varanasi, INDIA. He has worked as and Professor and Vice Principal in GGS College of Modern Technology, Kharar, Punjab, INDIA. His research interests include computational biology and pattern recognition.