LUNG DISEASES CLASSIFICATION -A SURVEY

Thamke Latika A., Vaidya Madhav V.,
1M. Tech, 2 Professor Head of department,
1Department of Information Technology,
1SGGSIE&T, Vishnupuri, Nanded India

Abstract— Lung diseases are the disorder, issues that affect the lungs, the organs that permit us to breathe and it is the most frequent medical conditions worldwide especially in India. When the diseases are not identified, they may also lead to serious troubles and may also even lead to transience. As a consequence, to assist the medical community this study helps in classifying some of the lung diseases, especially bronchitis, pleural effusion, emphysema and normal lung images. Lung diseases mainly occur as airway disease, lung tissue disease and lung circulation diseases are studied. In the computerized detection and classification of lung images, four main methodologies are used. Automatic classifications of lung disease in CT (Computed Tomography) images are an important diagnostic tool used for CADS (Computer Aided Diagnosis System). In this paper, we have compared the various segmentation methods for classification of Lung diseases, i.e. the Adaptive Thresholding and Gradient Magnitude and then GLCM feature can be extracted. The K-NN and Multiclass SVM classifier can be used for classification purpose. The total Datasets are 90 of each 30 classes, i.e. Normal Lung, Bronchitis and Emphysema. For training, the 63 CT images are used and for testing the 27 CT images are used. The Multiclass-SVM gives 85.19% accuracy.

Index Terms— CT images, Median Filter, Adaptive Thresholding, Multiclass SVM.

I. INTRODUCTION

Lung diseases refer to more than a few disorders affecting the lungs, such as chronic obstructive pulmonary (COPD) diseases, asthma, infections such as tuberculosis, pneumonia, lung cancer and other breathing problems. Lung disorder signs and symptoms can fluctuate via the type of the affected diseases. Common symptoms are trouble in breathing, shortness of breath, feeling like you are now not getting adequate air, decreased capacity to exercise, a cough that won't go away, coughing up blood or pain, mucus or pain when respiratory in or out. Medical image evaluation and manner requires an environment for data analysis, fast access, and revelation processing and algorithm development. Medical imaging is the method and system used to create images of the human body for clinical motive for analysis and evaluation or clinical science.

1.1 Emphysema

Emphysema is the condition of lung marked by irregular swelling of the alveoli with the defeat of pulmonary elasticity. The following fig 1 shows some CT images of the Lung diseases of Emphysema.

1.2 Pleural Effusion

A pleural effusion is a build-up of liquid (fluid) in the plural gap (space), and region (area) between the layer of tissue that lines the lung and the chest cavity. It may also be referred to as pulmonary effusion or effusion. The following Fig 2 shows some CT images of the Lung diseases of Pleural Effusion.

1.3 Bronchitis

The Bronchitis is the inflammation of the mucous membrane of the bronchus. The Bronchitis is divided into two types, i.e. chronic and acute bronchitis. The chronic bronchitis is a disease in which phlegm and cough for many months, i.e. about three months or for more than two years. The Acute bronchitis is characterized by fever and sputum production, which is commonly caused by the viral organism and also bacterial organisms. The following Fig 3 shows some CT images of Lung diseases of Bronchitis.
II. METHODOLOGY

The most common methodology for classification of lung disease can be shown in the Fig.4.

![Block diagram for Methodology](image)

2.1 Pre-Processing

It is the method for improving image quality previous to computational pre-processing and also used to remove the low frequency noise, reactions and covering portions of these images. Image processing method is useful on the captured images so that the necessary features are removed for further characteristics. Following Fig. 5 shows the Pre-processing images.

![Pre-processing images](image)

2.2 Segmentation

Segmentation is the procedure of separating a picture into several parts which is characteristically used towards recognizes substance otherwise other relevant data in digital pictures. The watershed transform finds a picture by treat it as floors where light pixels are excessive in addition to darkish pixels are low. Segmentation the use of the watershed transform identifies otherwise spot foreground substance as well as background positions.

![Segmentation images](image)

2.3 Feature Extraction

After the process of image pre-processing, various features of the input image are extracted from the infected region. The feature which is used in Lung disease detection is texture and sharp etc.
2.4 Feature Selection

Feature selection is a dimensionality reduction method broadly used for information mining and expertise discovery. It additionally approves getting rid of beside the point features, whilst retaking the underlying discriminatory information. Feature selection implies much less data transmission and efficient statistics mining.

2.5 Classification

It is the most important part in digital analysis. Classification can be executed on spectral feature like density, texture, etc. and then divides the features, space into many classes using different machine learning algorithms.

III. LITERATURE REVIEW

Raman deep Kaur and Prince Verma (2015) [1] Proposed a method in which lung CT images can be classified by using the improved MLP-NN classifiers. The median filter and morphological smoothing is used for pre-processing the images. The MAD technique with Gabor and WHT is used for feature extraction the images. The GA is used for the elements feature selection of the lung images. The lung can be categories as bronchitis, emphysema, pleural effusion and normal lung. The CT images can be classified using the MLP and Improved MLP-NN classifiers. The result is 88% for MLP and 93% for Improved MLP-NN classifiers. Saher B. Shaker, and Marleen de Bruijine (2010) [2] Proposed a KNN classifier to identify the CT image as normal tissue, Centrilobular emphysema based on the texture feature analysis. Rotation invariant LBP (Local Binary Patterns) and rotation invariant GFB texture features are calculated. Based on the above two features a feature histogram is calculated. In which the sequential forward selection algorithm has been applied after which classified with an accuracy rate of 95.2 % and qualitative measure of emphysema is derived. It has achieved high correlation with Pulmonary Function Test (PFT) than with other standard measure such as relative area (RA) of emphysema.

C. Bhuveswari (2013) [3] Proposed a method. For classify lung nodule into four classes. The CT images can be used the median filter is used for pre-processing the pictures. The feature extraction of images is done by using the Gabor filter. The fuzzy c-means clustering water shaded algorithm is used for clustering and segmentation. The lung can be categories as small cell lung carcinoma, normal, non small cell lung carcinoma as well as large cell lung carcinoma. The CT images can be classified using the naive bays’ and random fret classifiers. Anna N. Karahaliou, Panayiotis D. Korfiatis (2010) [4] Proposed a method for identification and characterization of lung parenchyma. The MD CT images can be used in the 3D. The lung can be categories as normal, ground glass, reticular patterns. The performance are evaluated in the metric of volume overlap true positive, false positive on the scanned image. The Multi Detector Compute Tomography (MDCT) acquires the acquisition of volumetric datasets which helps for enabling visualization, characterization and quantification of lung anatomy. The Computer Aided Detection automatically identifies and characterizes the DPLD patterns in CT image and two main stages are carried out. They are segmentation and classification. Segmentation is done using the gray level methods and classification is for classifying the tissue as normal and abnormal. The k nearest neighbor (KNN) voxel classification is done after the removal vessel tree volume for efficient characterization of lung parenchyma. With the help of volume overlap the performances are evaluated.

C. Bhuveswari (2014) [5] Proposed a method. For classify lung nodule into three classes. The CT images can be used the novel shape bases i.e. multiscale (size) filter is used for pre-processing the images. The genetic algorithm is used for feature extraction the images. The lung can be categories as bronchitis, pneumonia, normal lung. The CT images can be classified using the MLP-NN, KNN and bays net classifiers. The result is 86.75, 85.2, and 83.4 using the MLP-NN, KNN and Bays Net. Marios Anthimopoulos (2016) [6] Proposed a method for classify lung nodule into 7 classes. The CT images can be used in 3D. The lung can be categories as healthy. Ground glass opacity, micro nodules, consolidation, reticulations, honeycombing. The images can be classified using the CNN classifiers. The result is 85.5 % by using the CNN classifiers.

C.Bhuveswari (2014) [7] Proposed a method in which lung CT images can be classify by using the decision tree and naive byes classifier. The median filter is used for preprocessing the images. The moment invariance is used for feature extraction the images. The feature selection of the lung images is done by using the Genetic algorithm. The lung can be categories as normal and pleural effusion. The CT images can be classified using the Naive byes and decision tree classifiers. The location on of lung nodule is found. David S. Paik (2004) [8] proposed a Technique for classification of the lung nodules using the CAD (computer aided detection) algorithm. The CAD algorithm is also called as the surface normal overlap Technique and that Technique can be used to detect the lung nodule in the helical CT image to colonic polyp. The Surface normal overlap, candidate lesion selection and Gradient orientation are also used for detection of the lung nodules. The 8 CT images are used for testing purpose. The results are 5.6 FP/dataset and 90% Sensitivity.

Kuhnigk (2006) [9] proposed Technique for the classification of the lung nodules using the morphological operation. The morphological operation is applied on the pulmonary nodules for classification the lung diseases. The 700 datasets are used for the training and testing purpose. The results give the 95% accuracy by using the morphological operation. P. Reeves (2006) [10] proposed Technique for measuring the changes in the nodule size for classification of the Lung diseases. The Adaptive Thresholding Techniques are used for the segmentation of the lung nodules and the knowledge based shape matching is used for the detection of the lung diseases. The 50 datasets are used for the testing purpose.

Dehmeshki (2008) [11] proposed Technique for segmenting the different type of pulmonary lobes using the efficient algorithm. The Fuzzy connectivity map is used for the contrast the interest of object for region growing algorithm. The volumetric mask is operated using the region growing algorithm and which is created by 1st applying the local adaptive segmentation algorithm. The local adaptive segmentation algorithm is used for identified background and foreground region inside a certain window size. The 815 datasets are used for testing purpose. The results are 84% accuracy using the proposed Techniques. Stefano Diciotti (2010) [12] proposed Technique for estimation of the size of lung nodules through a scale space representation and high intra and inter operator reproducibility. The scale space signature by LoG (Laplacian of Gaussian) kernel is used for to analyse the blob like pattern of the lung nodules. The 50 phantom nodules and 215 nodules are used for the performance analysis. The results are Phantom are 9/146, 6/198 and 10/290 and for No failures are Failures for LDSD, Failures for LDMD and Failures for LiDC Dataset.
Lauge Sorensen (2010) [13] proposed Technique for texture analysis of the CT images. The texture feature can be extracted using the LBP (local binary patterns). The joint local binary patterns (LBP) and the intensity histograms are used to characterize ROIs (regions of interest). The classification is done by using the KNN classifier. The 168 datasets is used for the analysis of the lung images. The performance is the 97.3% sensitivity and 93.2% specificity. Henry Bouma (2009) [14] proposed Technique for automatic classification of the CT images of the pulmonary emphysema. The candidate detection is used for detection of the parenchyma and tissue disease. The shape and location of the pulmonary emphysema can be extracted using the feature computation Techniques. The various classifiers are used for classification of the pulmonary emphysema of the CT images. The 38 datasets are used for the detection of the lung images. The performance is the 63% sensitivity and 4.9 FP.

Yongbume Lee (2001) [15] proposed Technique for automatic classification of the CT images of the pulmonary nodules. The GATM (genetic algorithm template matching) Technique is used to detect the nodules. The genetic algorithm is used for detecting the target location of the lung nodules. Genetic algorithm also used for the selecting the patterns for matching template. The total 13 features are extracted using the template matching models. The 557 Sectional Images and 20 Clinical Cases are used for training and testing purpose. The results are FP 1.1 / Sectional Image with 72% Detection Rate. Kenji Suzuki (2005) [16] Proposed a Technique for classification of the lung nodules of the CT scans images of the lung diseases. The detection of the lung nodules is done by using the artificial neural network for Massive training. The 76 malignant nodules and 413 benign nodules are used for the detection of the lung nodules. The results are 48% accuracy for benign nodules and 100% accuracy for malignant nodule.

Matthew Brown (2001) [17] proposed Technique for automatically classify the lung lobes using the patient specific models of the CT images. The patient specific models are used for the extracting the feature like the shape, location and volume. To detect the candidate nodules, the priori model is used. The 17 test subjects are used for the testing purpose. The accuracy is the 86%. Fracois chabet (2001) [18] Proposed Technique for the classification of the lung nodules using the ERS (edge radius symmetry) transform. The pixels are arranged according to the uniformity, local strength, local symmetry and radial. The intensity of the patterns can be reduced by using the ERS transform. The template matching is used for the adjacent pulmonary vessels for identification. The nine datasets are used for the testing purpose. The performance is Kappa Statistics K >0.5 Ingrid Sliumer (2005) [19] Proposed Technique for classification of lung nodules. The scan images can be segmented using the registration model. The 15-registered normal scan is used for segmentation of the lung. The 10 nodules are used for the testing purpose.

Disha Sharma, [21] proposed method for segmentation and detection of the CT images of lung. The Lung border extraction, median filter, outlining, erosion and dilation are used for detecting the lung regions. The sobel methods are used for segmenting the lung nodules. The Diagnostic Indicators are used for the classification of the lung nodules. The 80% accuracy is obtained by using the proposed methods. Biradar, et al [7] proposed a method to classify lung nodule into two classes the MD CT images can be used in 2D/3D. The lung can be categories as bronchitis, emphysema, pleural effusion and normal lung. The images can be classified using the Support Vector Machines (SVM) classifiers. The Support Vector Machines (SVM) polynomial gives 96.6% accuracy and Support Vector Machines (SVM) quadratic function gives the 92% accuracy. The following Table 1 shows the Classification accuracy of lung diseases using various Method.

Anita, [22] Proposed method for segmentation and classification of the CT images of lung. The Threshold Segmentation is used for segmenting the lung nodules. The Neuro Fuzzy are used for the classification of the lung nodules. The 95% accuracy is obtained by using the proposed methods. Dansheng Song, et al [24] Proposed method for segmentation and classification of the CT images of lung. The Region Growing is used for segmenting the lung nodules. The Fast Artificial Neural Network is used for the classification of the lung nodules. The 95% accuracy is obtained by using the proposed methods.

S. Sivakumar, [25] Proposed method for segmentation and classification of the CT images of lung. The Weighted fuzzy Possibilistic based clustering are used for segmenting the lung nodules. The Support vector machine i.e. RBF kernel type classifier are 15 used for the classification of the lung nodules. The 80.36 % accuracy is obtained by using the proposed methods. Fathma Taher, [26] Proposed method for segmentation and classification of the Sputum images of lung. The Hopfield Neural Network (HNN) is used for segmenting the lung nodules. The Bayesian classifiers are used for the classification of the lung nodules. The 88.62 % accuracy is obtained by using the proposed methods. The following Table 1 shows the Classification accuracy of lung diseases using various Method.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Methodology</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bhuvaneswari, [3]</td>
<td>Moment Invariants</td>
<td>90%</td>
</tr>
<tr>
<td>Bhuvaneswari, [5]</td>
<td>Multiscale Filter</td>
<td>86.75%</td>
</tr>
<tr>
<td>Marios, [6]</td>
<td>CNN Classifiers</td>
<td>85.50%</td>
</tr>
<tr>
<td>Biradar, [20]</td>
<td>SVM with Polynomial</td>
<td>96.6%</td>
</tr>
<tr>
<td>Disha Sharma, [21]</td>
<td>Sobel Method</td>
<td>80%</td>
</tr>
<tr>
<td>Anita, [22]</td>
<td>Threshold Segmentation</td>
<td>95%</td>
</tr>
<tr>
<td>Dansheng Song, [23]</td>
<td>Region Growing</td>
<td>95%</td>
</tr>
<tr>
<td>S. Sivakumar, [24]</td>
<td>The Fuzzy Possibilistic Based Clustering</td>
<td>80.36%</td>
</tr>
<tr>
<td>Fathma Taher, [25]</td>
<td>Hopfield Neural Network</td>
<td>88.62%</td>
</tr>
</tbody>
</table>

Table 1. Comparison of different Methods for Classification of Lung Nodules.
IV. PROPOSED WORK

In this paper, we have compared the various segmentation methods for classification of Lung diseases, i.e. the Adaptive Thresholding and Gradient Magnitude and then GLCM feature can be extracted. We have extracted the feature like the Contrast, Correlation, Energy and Homogeneity. Following Fig. 5 shows the Pre-processing images for removing the noise. The noise can be removed using the various methods, i.e. the Median filter, Wiener filter and ordered filter. From the following Table 2, we have observed that the median filter gives the best results as compared to the other filter for removing the salt and pepper noise.

<table>
<thead>
<tr>
<th>Value</th>
<th>Median Filter</th>
<th>Wiener Filter</th>
<th>Ordered Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>29.6607</td>
<td>25.0394</td>
<td>7.9951</td>
</tr>
<tr>
<td>SNR</td>
<td>22.2240</td>
<td>17.6027</td>
<td>0.5584</td>
</tr>
</tbody>
</table>

The following Fig 6. Show the Some segmented images of Lung diseases using the Adaptive Thresholding and Gradient Magnitude.

The following Fig 7. Segmented images (a) Gradient Magnitude (b) Adaptive Thresholding

Fig 8 shows the Performance measure by K-NN and Multiclass-SVM classifier for Gradient Magnitude and Adaptive Thresholding. Table 3 shows the Performance Measure by K-NN Multiclass-SVM classifiers for Gradient Magnitude and Adaptive Thresholding.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient Magnitude and K-NN</td>
<td>0.6667</td>
<td>0.5000</td>
<td>0.3333</td>
<td>0.4000</td>
</tr>
<tr>
<td>Adaptive Thresholding and K-NN</td>
<td>0.8148</td>
<td>0.8333</td>
<td>0.5556</td>
<td>0.6667</td>
</tr>
<tr>
<td>Gradient Magnitude and Multiclass-SVM</td>
<td>0.6667</td>
<td>0.5000</td>
<td>0.1111</td>
<td>0.1818</td>
</tr>
<tr>
<td>Adaptive Thresholding and Multiclass-SVM</td>
<td>0.8519</td>
<td>1.0000</td>
<td>0.5556</td>
<td>0.7143</td>
</tr>
</tbody>
</table>
From the Table 3, we have observed that the Multiclass-SVM with Adaptive Thresholding is better for the Classification of Lung Diseases.

![Fig 8: Performance measure by K-NN and Multiclass-SVM classifier](image)

V. CONCLUSION

This paper presents the survey on different diseases classification techniques used for lung disease. Improving Image quality and accuracy is a core factor of this study and showed the comparison. In this work, In pre-processing, for removing the salt and pepper noise the median filter is better as compared to the other filter. For segmentation, the Adaptive Thresholding is better as compared to the Gradient Magnitude for Classification of the Lung CT images. The Classification Accuracy 81.48 % gives K-NN classifiers with Adaptive Thresholding and 85.19% gives Multiclass-SVM classifiers. To improve recognition rates in the classification process Bays classifiers, Artificial Neural Network, Hybrid algorithm etc. can also be used.

REFERENCES


