International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-7, Issue-5S4, February 2019

Lung Cancer Detection of CT Lung Images

Retz Mahima Devarapalli1, Hemantha Kumar Kalluri, Venkatesulu Dondeti

ABSTRACT--- Cancer is one of the deadliest diseases leading to innumerable deaths worldwide. Early detection of lung cancer could increase the survival rate. To detect cancer various image processing techniques have been innovated and applied like median-wiener filter in the preprocessing stage. In the classification Back Propagation model, SVM (Support Vector Machines), Forward Neural Networks, Convolution Neural Networks are used to detect whether the nodule is cancerous or not. Although, there are many such techniques which are available these days but there is still need to further develop early detection to improve accuracy leading to better survival rate.

Keywords: Lung cancer detection, SVM Classifier, Image Processing.

1. INTRODUCTION

Cancer is the most prevalent terminal disease globally, accounting for an estimated 9.6 million deaths in 2018. Of many types of cancers, lung cancer is one of the frequently occurring diseases that causes death and is identified in both genders over an estimated death of 1.76 million in 2018 [1]. The unbounded growth of anomalous cells that are not normal in one or both lungs is Lung cancer. These abnormal cells disrupt the functions of normal lung cells which might lead to an unhealthy tissue in lung. The growth of these abnormal developments leads to the appearance and formation of tumors and disturbs the functionality of lung that supplies oxygen via blood to the body.

Lung cancer is basically of two types [2], Non-Small-Cell Lung Cancer Cell (NSCLC) and Small-Cell-Lung-Cancer (SCLC). 85% of the lung cancer is especially caused by the Non-Small Lung Cancer and SCLC accounts for the remaining 15% and is usually caused by smoking tobacco [2]. CT (Computed Tomography) Scan images of the chest provides the detailed information of abnormal cells in the lung, which are used in detecting the lung cancer, which lowers the chance of dying of lung cancer [3].

The remaining part of the paper described as follows: Section 2 describes the stages involved to detect the lung cancer and brief explanation of each stage along with the techniques applied in each step. Section 3 describes various data sets used by earlier researchers to detect the lung cancer. Section 4 explains the results achieved by earlier researchers. Section 5 describes the conclusion of this survey.

Revised Manuscript Received on February 11 , 2019.

Retz Mahima Devarapalli1, CSE Department, Vignan's Foundation for Science, Technology & Research, Guntur, A.P, India.(dretzmahima56@gmail.com)

Hemantha Kumar Kalluri, CSE Department, Vignan's Foundation for Science, Technology & Research, Guntur, A.P, India.

Venkatesulu Dondeti, CSE Department, Vignan's Foundation for Science, Technology & Research, Guntur, A.P, India.

2. STAGES INVOLVED TO DETECT LUNG CANCER

In the medical field to detect a disease, many of the image pro-cessing strategies can be applied. To improve the detection of lung cancer in the CT images there are four main steps involved. At each step, various techniques are applied which resulted in different accuracies in detecting the lung cancer. Firstly, the lung CT image is pre-processed to remove any noises that exists in the image. Secondly, the image is segmented to get Region of Interest (ROI). Thirdly, feature extraction is applied to extract features like energy, entropy, variance. Finally, different classification algorithm is applied on the extracted features of the lung CT image. All these steps are involved to detect lung cancer are depicted in Fig 1.

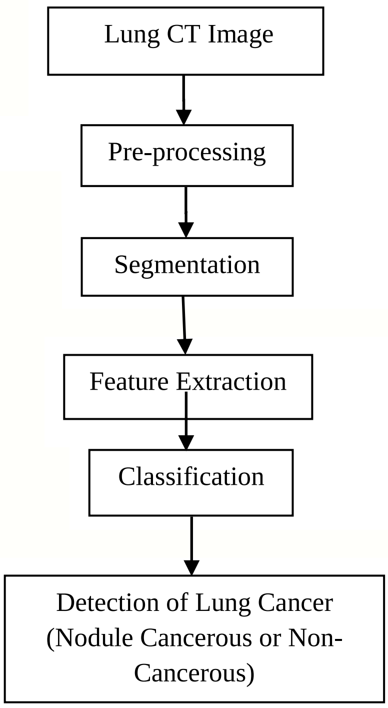

Fig 1: Steps involved to detect Lung Cancer

2.1. Pre-Processing

Image pre-processing is the first and important technique involved in lung cancer detection. Pre-processing technique is needed to improve the detection accuracy and to eliminate some regions of CT Image such as background and surrounding tissues or vessels.According to Sagamithara et al., [4] lung cancer can be detected in the classification stage by applying back-propagation, and in the pre-processing

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

stage, Median filter is applied to remove the noise and Wiener filter is applied to deblur the image. Deep Prakash et al. [5] eliminated the noise of the image and the image is then converted to grayscale image in the processing stage to remove the irregularities present in the image.

Janee et al., [6] used colour space transformation, contrast enhancement and image scaling are applied in the pre-processing stage for enhancement of the image. Twinkal et al., [7] in the first stage initiated with the process of enhancing the image, which enhances the percipience of images. One of the image enhancement methods is by equalization of the histogram, which standardizes the contrast in the image. Foreground and background separation method and image enhancement are applied by Mithuna et al. [8] in the Pre-processing. Talebpour et al. [11] firstly, converted all of the images to grayscale images using simple thresholding. Kalaivani et al. [12] converted the image to grayscale and applied Histogram equilization.

2.2. Segmentation

The process of segregating the digital image into numerous parts, so as to use the information retrieved and identify the objects easily from the segmentation process effectively is Image Segmentation. To perform segmentation on an image there are different techniques including, thresholding using Otsu's Method, watershed segmentation, colour based segmentation such as k mean clustering, and texture methods such as texture files.

K-mean clustering is used in the segmentation stage of the pre-processed CT images by Sangamithra et al. [4]. After the prepressing stage, the Region of Image is extracted in the segmentation stage by Deep Prakash Kaucha et al. [5], which selects only the selected region from the lung CT Image. Later, DWT (Discrete Wavelet Technique) is applied on the extracted ROI, to perform compression and multi resolution analysis. By applying DWT, the processed image is divided into four sub-bands LL, LH, HL and HH, again second derivative level DWT is applied on LL band for finding the smooth variation. The maximum content of the information is in the LL Band and edges in other directions apart from horizontal are preserved in the other higher sub bands.

After the pre-processing of the image, Threshold and Watershed-Based-Segmentation are applied by Janee et al. [6]. Twinkal et al. [7] To extirpate the background, without disturbing the other additional pixels, OTSU Segmentation is applied. Adaptive diffusion active contour model is used to perform segmentation by Mithuna et al. [8]. Talebpour et al. [11] used HU units standard to perform segmentation. Later, they have filtered the segmented images using CNEF (Cylindrical Nodule Enhancement filter) which is a 3D filter that improves the enhancement of nodule. Kalaivani et al. [12] segmented the images using thresholding.

2.3. Feature Selection

Feature extraction is a predominant step in the image processing, edges and lines convey the most important information of an image which needs to be extracted. Features that exist are extracted by implementing Gray-Level-Co-Occurrence Matrix (GLCM) where analyzing the texture of an image where spatial relationship of pixels are to be considered. The GLCM functions calculates how often pairs of pixels with specific values and in a specified spatial relationship occur in an image that characterizes the texture of an image. Creating a GLCM, which then resulting in the extraction of measures from the matrix. Some of the statistical features calculated from contrast, correlation, energy, homogeneity.

Sangamithra et al. [4] extracted the features like energy and homogeneity. Entropy and co-relation are also considered for extraction using GLCM after the segmentation stage. Deep Prakash et al, [5] in the feature extraction stage, Using GLC matrix the properties like energy, entropy, variance, dissimilarities and co-occurrence are extracted. GLCM extraction technique used by Janee et al. [6], which arranges massive incorporation of pixel brightness values in a sample. Later, the authors used the properties like mean, entropy, homogeneity, smoothness, fluctuation, standard deviation/n, RMS, IDM, relationship, kurtosis and vitality.

Twinkal et al., [7] extracted three features such as entropy, contrast, energy which improves efficiency and can be easily evaluated for classification. Mithuna et al., [8] extracted features like area, perimeter, circularity. Talebpour et al., [11] extracted both geometric and text features. Geometric features like volume, surface area, spherical density and disproportion, elongation, radius are extracted using binary Mask. Text features like Homogeneity, Moment, Entropy, Kurtosis, Skewness, contrast and energy are extracted using Gray Level. Kalaivani et al. [12] applied regionprops function and extracted features like Perimeter, Major Axis Length, Area, Minor Axis Length, Solidity and Eccentricity.

2.4. Classification

A kind of supervised machine learning where an algorithm classifies new deductions from existing examples of labelled data is nothing but Classification. Different classification algorithms are applied on the extracted feature images like Back Propagation, Support Vector Machines (SVM), Convolution Neural Networks (CNN).

According to Sagamithara et al., [4] lung cancer can be detected using a back-propagation model, which resulted in the accuracy of 90.65%. Deep Prakash et al., [5] applied Support Vector Machines (SVM) to perform classification, which uses optimum linear separating hyperplanes that classifies the image. To classify into normal or cancerous, linear kernel is applied. This model has achieved an accuracy of 95.16%.

Instead of detecting the cancer, Janee et al., [6] proposed to predict the nodule on the CT Image. In the classification stage they have utilized SVM classifier for the data classification, where they bisected the whole area affected by the disease with the total area of the lung to calculate the affected region of the image which resulted in the accuracy of 97%. Mithuna et al., [8] used SVM for the classification purpose.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

To distinguish nodules from non-nodule objects Talebpour et al., [11] used a three-layered Feed-Forward-Neural-Network and also applied the back-propagation method. Kalaivani et al. [12] also applied back propagation model and implementation is done on Matlab, which acquired an efficiency of 78%.

Though, there are different steps involved, there is no necessity to follow the same approach for detecting the lung cancer. Hongyang Jiang et al. [10] used LIDC Dataset where the input is multigroup 2D Lung CT Images. It involves 3 steps. In the First step Lung Contours are rebuilt via a method which is slope analysis. In the second step, structures that look like vessels in CT Image are eliminated using a filter called Frangi Filter. In the third step, CNN Structure is designed using two groups of images one is original and the group 2 images are binary images generated through complex binarization processing. Initially, the CT images are examined discretely. The results marked by them are distributed among them to re-analyze.

During the next stage, each of them re-analyzed the results marked by them and again made their annotations separately. Two different kinds of nodules were identified, small and large nodules. Small nodules can be detected using image filters and only large nodules are detected using this method. Brightness of the lung image is darker, shape of the lung is highlighted first through local variance between clusters.

Noises are eliminated using morphological methods such as opening and closing operations. Padding operator is used to pack the bubbles in the lung that enhances the lineation. This method also mending the lung contour for juxtra plural nodules near thorax wall. Patch will automatically adjust with the angle of lung contour. Patch will determine the changes of lung contour and lung parenchyma was segmented by corrective lung contours.

Vessels are bright tissues that are tabular in structure where the vessels are eliminated. Although, there are many algorithms like Vessel Enhancing Diffusion Algorithm, Frangi filter, sato Filter, the researchers applied Frangi Filter, which resulted in the vessel Image. This image is subtracted from original image to get the vessel eliminated images. Because of this, there will be a decrease in the no. of suspicious regions and there is an increase in the nodule detection rate. In the final stage, Convolution Neural Network is used to classify whether the nodule is cancerous or not. The sensitivity achieved by the researchers is 94%.

3. DATASETS & RESULTS

These are the different datasets that have been used by various authors:

1. UCI machine learning database – 500 Lung Cancer CT Jpeg Images, Janee et al., [6] utilized this dataset.
2. CT (Computed Tomography) scan images collected from the image archive radio frequency image database utilized by Mithuna et al., [8].
3. From Abderrahmen Mami Hospital of Radiology Department, sub pleural nodules of 40 CT scans – Ariana, this data set is used by Rekha et al., [10].
4. LIDC/IDRI (Lung Image Database Consortium/ Image Database Resource Initiative) dataset Composed of 1018 cases, contains, chest CT images, medically annotated by 4 radiologists stored as XML files. Precisely, 7371 images are labelled as nodules, of them 2669 images are greater than 3mm nodules. This dataset is used by the Deep Prakash et al., [5], Hongyang et al., [10], Talebpour et al., [11].
5. Lung CT database acquired from Cancer Imaging Archive, contains, 4682 images stored as DICOM (Digital Imaging and Communications in Medicine) of 61 Patients. Kalaivani et al., [12] used this data set.

4. RESULTS & DISCUSSIONS

Sangamithra et al., [4] used LIDC/IDRI database which resulted in the accuracy of 90.65%. Prakash et al., [5] also used LIDC Dataset that resulted an accuracy of 95.16% and a sensitivity of 98.21%. Janee et al., [6] used UCI Machine Learning database that resulted un the detection accuracy of 97% and prediction accuracy of 87%. Rekka et al., [9] used the CT images from Image Archive Radio Frequency Database that resulted an accuracy 0f 98.52% and Sensitivity of 76.75%. Hongyang et al., [10] used the LIDC/IDRI dataset in their research that resulted a sensitivity of 94%. Talebpour et al., [11] used the LIDC/IDRI dataset that resulted in a sensitivity of 90%. All the accuracies that have been achieved by earlier researchers so far by applying various techniques at each stage to detect the lung cancer is shown in Table 1.

Author	Technique Applied	Datasets	Performance Measure
Sangamithra et al., [4]	Median Filter, EK Mean Clustering, GLCM, Back Propagation neural network Algorithm	LIDC/IDRI Dataset	Accuracy - 90.65%
Deep Prakash et al., [5]	Gray Scale Image, ROI and DWT is applied, GLCM, SVM Classifier Algorithm	LIDC/IDRI Dataset	Accuracy - 95.16% Sensitivity – 98.21%
Janee et al., [6]	Watershed segmentation, GLCM, SVM Classifier algorithm	UCI machine learning dataset	Detection Accuracy- 97% Prediction Accuracy – 87%
Rekka et al., [9]	Otsu thresholding Method, Morphological closing, mathematical subtraction, Clear Border Operation	40 CT scans from Abderrahmen Mami Hospital	Accuracy - 98.52% Sensitivity – 76.75

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

Hongyang et al., [10]	Morphological methods – closing and opening, Frangi Filter, Convolution Neural Networks	LIDC/IDRI Database	Sensitivity – 94%
Talebpour et al., [11]	Thresholding method, Cylindrical Nodule Enhancement Filter, Binary mask, Gray Level, Feed Forward Neural Network and Back Propagation Model	LIDC/IDRI Dataset	Sensitivity – 90%

5. CONCLUSION

The work which has been carried out through this paper mainly focuses on the literature survey to detect lung cancer. Median filter, Histogram Equalization are applied and used in preprocessing stage. Watershed and ROI also are segmented in the next stage which is a part of segmentation. Techniques like Back Propagation Model, Support Vector Machine classifier, Feed Forward Neural Networks and Convolution Neural Networks are applied in the classification stage. By applying these techniques at various stages resulted in different accuracies detection and confirmation of early detection of cancer. When compared with other machines of cancer detection only Support Vector Machine has achieved the highest accuracy.

REFERENCES

1. http://www.who.int/news-room/fact-sheets/detail/cancer
2. https://www.lungcancer.org/find_information/publications/163-lung_cancer_101/268-types_and_staging
3. https://www.cancer.org/cancer/non-small-cell-lung-cancer/detection-diagnosis-staging/detection.html
4. P.B. Sangamithra, & S. Govindaraju, "Lung Tumor Detection and Classififcation using EK- Mean Clustering", International Conference on Wireless Communications, Signal Processing and Networking, IEEE WiSPNET conference, pp. 2201-2206,2016.
5. Deep Prakash K, Prasad P W C, Alsadoon A, Sreedharan S, "Early Detection of Lung Cancer using SVM Classifier in Biomedical Image Processing", IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, pp. 3143-3148,2017.
6. Janee Alam, Sabrina Alam, Alamgir Hossan, "Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifier", International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018
7. Ms. Twinkal Patel, Vimal Nayak, "Hybrid Approach for Feature Extraction of Lung Cancer Detection", 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018), pp. 1431-1433, 2018.
8. Mithuna B.N, Dr. Pushpa Ravikumar, Arpitha C.N, "A Quantitative Approach for Determining Lung Cancer Using CT scan Images", Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology ICECA 2018pp. 1786-1790,2018.
9. Rekka Mastouri, Henda Neji, Saoussen Hantous-Zannad, Nawres Khlifa, "A morphological operation-based approach for Subpleural lung nodule detection fom CT images", 2018 IEEE 4th Middle East Conference on Biomedical Engineering, MECBME, pp.84-89,2018.
10. Hongyang Jiang, He Ma, Wei Qian, Mengdi Gao & Yan Li, An Automatic Detection System of Lung Nodule Based on Multigroup PatchBased Deep Learning Network, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 22, NO. 4, pp. 1227-1237, JULY 2018
11. A.R. Talebpour, H.R. Hemmati, M. Zarif Hosseinian, "Automatic Lung Nodules Detection In Computed Tomography Images Using Nodule Filtering and Neural Networks", The 22nd Iranian Conference on Electrical Engineering (ICEE 2014), pp. 1883-1887,2014.
12. S. Kalaivani, Pramit Chatterjee, Shikhar Juyal, Rishi Gupta, "Lung Cancer Detection Using Digital Image Processing and Artificial Neural Networks", International Conference on Electronics, Communication and Aerospace Technology, ICECA, 100-103, 2017.