IJCERT

International Journal of Computer Engineering in Research Trends

Multidisciplinary, Open Access, Peer-Reviewed and fully refereed

Research Paper | Volume-9, Issue-7 ,2022 Regular Edition | E-ISSN: 2349-7084

Lung Nodule Detection and Classification using Image Processing Techniques

Swathi Velugoti [1], Revuri Harshini Reddy [2], Sadiya Tarannum [3], Sama Tharun Kumar Reddy [4]

[1] Assistant Professor, Department of Computer Science and Engineering, Guru Nanak Institutions Technical Campus Hyderabad – Telangana, India

[2,3,4] UG Students, Department of Computer Science and Engineering, Guru Nanak Institutions Technical Campus, Hyderabad – Telangana , India

Email ID: swathivelugoti@gmail.com , harshinireddyrevuri@gmail.com , sadiyatarannum999@gmail.com , reddytharun2301@gmail.com

**Corresponding Author: Swathi Velugoti*

Available online at: http://www.ijcert.org

Received:26/06/2022, Revised: 16/07/2022, Accepted: 21/07/2022, Published: 25/07/2022

Abstract: - Lung cancer is one of the significant reasons for death among India. Many diagnosis and detection of lungs cancer has been done using various data analysis and classification techniques. Since the cause of lung cancer stay obscure, prevention become impossible, thus early detection of tumor in lungs is the only way to cure lung cancer. Hence, lung cancer detection system using image processing and machine learning is used to classify the presence of lung cancer in a CT- images and blood samples. In spite of CT scan reports are more effective than Mammography; therefore patient CT scan images are categorized in normal and abnormal. The abnormal images are subjected to segmentation to focus on tumor portion. Classification done on features extracted from the images. The efficient method to detect the lung cancer and its stages successfully and also aim to have more accurate results by using SVM and Image Processing techniques.

Keywords: Lung Nodule Detection, CT scans, image processing techniques, Support vector machine

1. Introduction

Cancer is one of the diseases that people are particularly concerned about today. Mortality from lung cancer are expected to continue rising, to become around 17 million worldwide in 2030. Every sixth death in the world is due to cancer, making it the second leading cause of death (second only to cardiovascular diseases), and lung cancer is one of the most deadly diseases. The main cause is the formation of cancerous nodules around the lobe or lung. Therefore, early detection of nodules is very important. Because nodules are small dots in computed tomographic images, clinicians need to examine each image one by one, which is time-consuming and leaves a likely possibility of overlooking a nodule. Early detection of lung cancer can increase the chance of survival among people. There are many techniques to diagnose the lung cancer, such as Chest Radiograph (X-ray), Computed Tomography (CT), Magnetic Resonance Imaging (MRI scan) and Sputum Cytology.

However, most of these techniques are expensive and time consuming. Therefore, there is a great need for a new technology to diagnose the lung cancer in its early stages. Image processing techniques provide a good quality tool for improving the manual analysis. Many diagnosis and detection of lungs cancer has been done using various data analysis and classification techniques. Since the cause of lung cancer stay obscure, prevention become impossible, thus early detection of tumor in lungs is the only way to cure lung cancer. Hence, lung cancer detection system using image processing and machine learning is used to classify the presence of lung cancer in a CT- images and blood samples. In spite of

CT scan reports are more effective than Mammography, therefore patient CT scan images are categorized in normal and abnormal. The abnormal images are subjected to segmentation to focus on tumor portion. Classification done on features extracted from the images. The efficient method to detect the lung cancer and its stages successfully and also aim to have more accurate results by using SVM and Image Processing techniques.

Rest of the paper is organized as follows i.e Section 2 descries Related Work, Section 3 Presents the methodologies, Section 4 presents Results and analysis and finally Section 5 Concludes the summary of research work.

2. Related Work

In this paper[10], some image pre-processing methods such as thresholding, clearing borders, morphological operations (viz., erosion, closing, opening) are discussed to detect lung nodule regions ie, Region of Interest (ROI) in patient lung CT scan images. Also, machine learning techniques such as Support Vector Machine (SVM) and Convolutional Neural Network (CNN) has been discussed for classifying the lung nodules and non-nodules objects in patient lung ct scan images using the sets of lung nodule regions.

In this paper [11] Lung segmentation in Computed Tomography (CT) images plays a vital role in the diagnosis, detection and three-dimensional visualization of lung nodules. In addition, the stability, accuracy and efficiency of lung segmentation in CT images have a significant impact on the performance of Computer-Aided Detection (CAD) systems. Lung segmentation is usually the first step in lung CT images analysis. In this paper, a fully automated algorithm for recognition and segmentation the lung in 3D X-ray images using the Active Shape Model (ASM) is presented.

In this paper[12], a fully automated model is presented for NSCLC nodule(s) segmentation from CT scan image. The proposed method follows four steps: (1) Preprocessing, (2) Automatic Lung Parenchyma Extraction and Border Repair (ALPE&BR), (3) Automatic lung nodules segmentation using Connected Component Analysis (CCA) and Threshold BasedMathematical Nodule (TBMN) refinement algorithm and (4) Nodules filtering using Hounsfield Unit (HU) value and true cancerous nodule extraction.

This paper [13] proposes an adaptive solution to mitigate the difficulty of thresholding-based method in lung segmentation. Sufficient detection power for nodule candidates is inevitably accompanied by many (obvious) FPs. A rule-based filtering operation is often employed to cheaply and drastically reduce the number of obvious FPs, so that their influence on the computationally more expensive learning process can be eliminated. In general, FP reduction using machine learning has been extensively studied in the literature. Compared with unsupervised learning that aims to find hidden structures in unlabelled data, supervised learning, which aims to infer a function from labelled training data, is more frequently used to design a CADe system. Compared with the existing approaches, the morphology based lung cancer detection can be an alternative with either comparable detection performance and less computational cost, or comparable cost and better detection performance

As there are various preprocessing filters are available and been introduced, the purpose of using Preprocessing is to enhance some features and to remove unwanted features for segmenting the nodule. Segmentation is an important factor in image processing which helps to detect cancer tissue earlier for treatment.

3. Methodologies

The lung CT images having low noise when compared to scan image and MRI image. The main advantage of the computer tomography image having better clarity, low noise and distortion. So we can take the CT images for processing the lung image. Then segmentation is applied to the lung image.

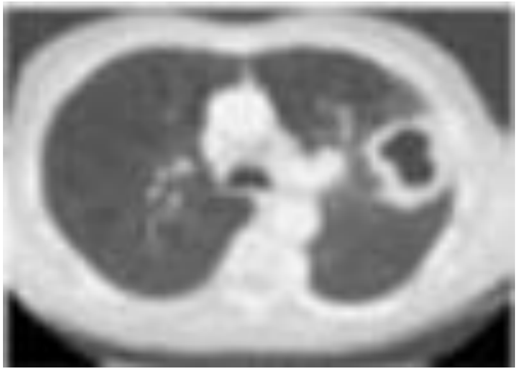

Figure 1 Input lung CT image

The proposed system for lung cancer detection in CT images is shown with the help of architecture in figure 2.The methodology is carried out in five main steps and each step of this system is discussed in detail in section below

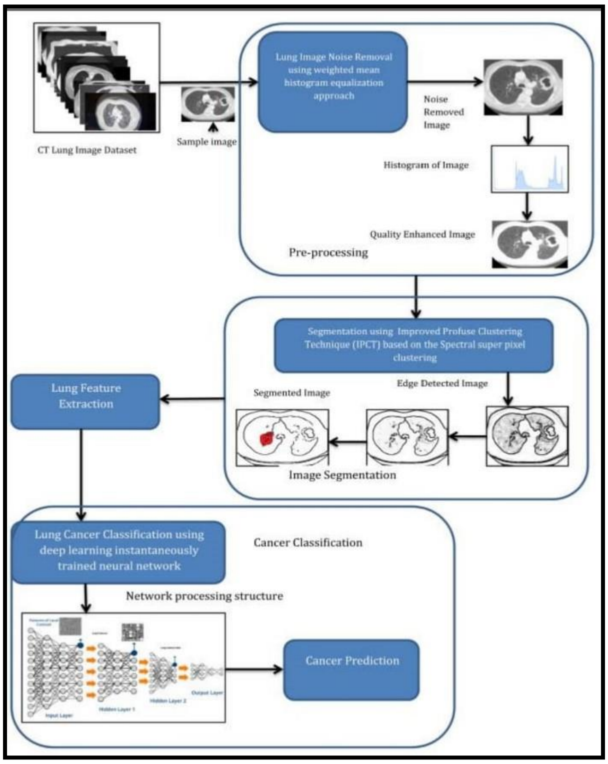

Figure 2. Block fiagram of the Proposed model

3.1 Data Set

For the proposed work the dataset used in the Lung Nodule Analysis (LUNA) which is derived from Lung image Database Consortium and Image Database Research Initiative (LIDC/IDRI) database[14]. Out of 1024 patient data available, 888 patient data is used to excluding the slices of thickness greater than 2.5mm. The dataset consists of the DICOM format CT scan images where they are showed in MHD and RAW image format. MHD file format is the Meta image that is based tagged file format for the medical images and RAW file contains the processed data from the image sensor of either camera, scanner. The dataset includes the annotations of the CT scans provided in the dataset. These annotations are obtained from the expertized radiologists including all the information (diameter, X, Y, Z co-ordinates of the images) about the CT scans images of lung. The dataset LUNA was the challenge as the part of 2016 IEEE International Symposium on Biomedical Imaging.

3.2 Nodule Extraction and Pre-Processing

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

3.3 Image Segmentation

The process of separating out required region of interest from the image is known as segmentation. Mathematical morphological operations are powerful tools in acquiring lung region from binary images. In our methodology, first the preprocessed gray scale images were converted to binary images. Morphological opening operation was performed to the binary image with disk structuring element for removal of unwanted components from the image. The opened image was then complemented and clear border operation was performed to it. The lung masks were obtained by filling the holes and gaps present in the lungs. Finally exclusive OR operation was performed to lung mask output and clear border output to give us the segmented tumor region.

3.4 Feature Extraction

Feature extraction is the most essential step that transforms input data into required features. This stage extracts out significant features of segmented region of interest and these features serve as input for classification of CT scan images. The size and shape of tumor present in the lungs is estimated by extracting three geometrical features. The features are area, perimeter and eccentricity of cancerous lung nodule. 1. Area: This is a scalar quantity which gives total number of pixels acquired by cancerous lung nodule. The area is evaluated from the binary image by taking summation of pixel areas in the image that are registered with value 1. 2. Perimeter: This is a scalar quantity that gives the total pixels present at the border of the lung tumor. The perimeter is evaluated from the binary image by summing the pixels registered with value 1, at the outline of lung nodule. 3. Eccentricity: This metric value is also referred to irregularity index (I) or circularity or roundness. For a circular shape eccentricity value is equal 1 and the value is less than 1 for any other shape. Eccentricity =length of major axis /length of minor axis

3.5 Classification

The Classification stage involves labeling the CT scan images as normal and abnormal. In our method SVM algorithm will be used for detection of lung cancer in CT images. SVM classifiers are supervised learning models that analyze input data and classify them according to pattern. The SVM classifier builds a model by using training dataset and categorizes it into two classes. The SVM algorithm then assigns new examples of testing dataset to one of the two classes. SVM classifier thus finds the best hyper plane that separates the two groups and thus classifies the lung CT images. For the best hyper plane data points of one class are separated from the other by largest margin between the two classes.

4. Results

The proposed model is then developed in MATLAB R2016a. MATLAB is one of the tools for research development and analysis [15]. Both detection and features extraction are implemented in MATLAB and classification is implemented using machine learning toolbox. Classification learner toolbox aids in developing the trained prediction model from the features extracted easily and very fast. 5 folds cross validation was used to prevent from overfitting during the training process. Different 16 DICOM images from LIDC are used for training the classifier and result is validated using 5 images with total 15 nodules. Database For This Study Was Obtained From The Luna Dataset[14] .Figure 2 Shows The Ct Scan Image Of Patient Affected By Lung Cancer.

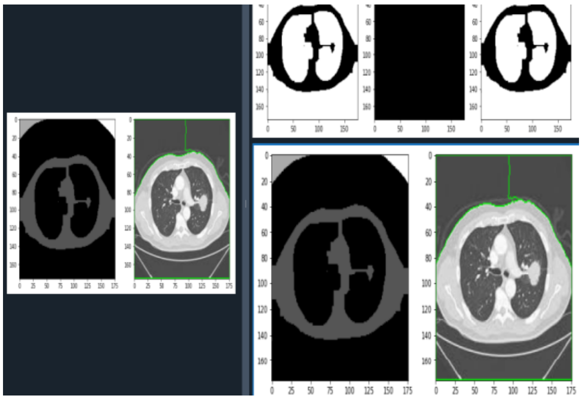

Figure 3. Experimental Illustration

further in the preprocessing stage image enhancement was done using contrast adjustment. in contrast adjustment image intensity values are mapped to full display range of the input data and the contrast of the image is enhanced. figure 2 depicts the contrast enhanced ct scan image fig.2 contrast enhanced image segmentation technique divides input image into various parts and thus gives us the region of interest for further processing. to exact the region of interest, morphological operation based on structuring element were used. the lung masks and tumor region were obtained from ct image .

Figure 3: Bordered corrected output

These features serve as inputs to the svm classifier for categorizing the images into normal (non-cancerous) or abnormal (cancerous).for the given sample ct scan image, the extracted lung tumor as svm classifier is used to determine whether the ct scan images are normal or abnormalusing the templateare normal or abnormal using the template

Figure 4. Detected lung nodule

Comparing the accuracy of proposed model with current model it can be seen that there is progressive increase in accuracy from 88.4% to 93.52%. Sensitivity remained same. Specificity increased from 40% to 50% From the detected cancer nodes, features like Area, Perimeter, Centroid, Diameter, Eccentricity and Mean Intensity of the Pixels were extracted. Extracted features were used to Train Support vector machine and trained model was developed. Training time for classification learner app was 5.93 seconds. Classification learner app evaluates the prediction time for the developed trained model to be 310 observations per second. Scatter plot of trained model are as below.

Table 1. Confussion matrix data

	Existing model	Proposed Model (CNN)
Total number of nodes detected	23	24
Number of True Positive (TP)	21	21
Number of True Negative (TN)	2	2
Number of False Positive (FP)	2	3
Number of False Negative (FN)	0	0

Performance Metrics : To measure the performance of the proposed model using following metrics

Accuracy= (TP+TN)/(TP+TN+FP+FN)
Sensitivity=TP/(TP+FN)
Specificity=TN/(TN+FP)

Therefore, from above result we can say that our proposed model classifies as benign or malignant with accuracy of 93.52%. The classification of nodule as malignant or benign which was not performed in the best model has been successfully implemented.

5. Conclusion

Lung cancer is the most dangerous and widespread in the world according to stage the discovery of the cancer cells in the lungs, this gives us the indication that the process of detection this disease plays a very important and essential role to avoid serious stages. Based on the work done, a lung cancer can be detected and classified using the neural network. This helps for the doctor to improve treatment in the early stage of cancer and avoid many deaths of patients with the effect of lung cancer detection in early stage. The average percentage accuracy for the proposed system is reached 93.52% for detection and classification of lung cancer using CNN.

References

[1] Prajwal Rao, Nishal Ancelette Pereira, And Raghuram Srinivasan, "Convolutional Neural Networks For Lung Cancer Screening In Computed Tomography (CT) Scans", 978-1-5090-5256-1/16/$31.00_C 2016 IEEE
[2] Sheila Ramaswamy, Karen Truong, "Pulmonary Nodule Classification With Convolutional Neural Networks", Stanford University@2016.
[3] Ayman El-Baz, Garth M. Beache, Georgy Gimel'farb, Kenji Suzuki, Kazunori Okada,Ahmed Elnakib, Ahmed Soliman, And Behnoush Abdollahi, "Computer-Aided Diagnosis Systems For Lung Cancer: Challenges And Methodologies" International Journal Of Biomedical Imaging, Article ID 942353, 46 Pages, 2013
[4] Rotem Golan, Christian Jacob, J¨Org Denzinger, "Lung Nodule Detection In CT Images Using Deep Convolutional Neural Networks", 978- 1-5090-0620-5/16/$31.00_C 2016 IEEE
[5] Xulei Yang, Si Yong Yeo, Jia Mei Hong, Sum Thai Wong, Wai Teng Tang, Zhen Zhou Wu, Gary Lee, Sulin Chen, Vanessa Ding, Brendan Pang, Andre Choo, Yi Su, "A Deep Learning Approach For Tumor Tissue Image Classification", Researchgate, DOI: 10.2316/P.2016.832-025, February 2016.
[6] Allison M Rossetto And Wenjin Zhou, "Deep Learning For Categorization Of Lung Cancer CT Images", 2017 IEEE/ACM International Conference On Connected Health: Applications, Systems And Engineering Technologies (CHASE), 978-1-5090-4722-2/17 $31.00, DOI 10.1109/CHASE.2017.59, © 2017 IEEE.
[7] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Li Wang, Gang Wang, Jianfei Cai, Tsuhan Chen, " Recent Advances In Convolutional Neural Networks", Sciencedirect , Pattern Recognition, Vol 77, Pages 354- 377, 2015.
[8] Md. Badrul Alam Miah, Mohammad Abu Yousuf, "Detection Of Lung Cancer From CT Image Using Image Processing And Neural Network", 2nd Int'l Conf On Electrical Engineering And Information & Communication Technology (ICEEICT) 2015 Jahangirnagar University, Dhaka-1342, Bangladesh, 21-23 May 2015, 978-1-4673-6676-2/15/$31.00 ©2015 IEEE.
[9] Xin-Yu Jin, Yu-Chen Zhang, Qi-Liang Jin, "Pulmonary Nodule Detection Based On CT Images Using Convolution Neural Network2016 9th International Symposium On Computational Intelligence And Design, 2473-3547/16 $31.00, DOI 10.1109/ISCID.2016.52, © 2016 IEEE.
[10] Tekade, Ruchita. "Lung Nodule Detection And Classification Using Machine Learning Techniques." (2018).

[11] Shariaty, Faridoddin Et Al. "Automatic Lung Segmentation In Computed Tomography Images Using Active Shape Model." *2020 IEEE International Conference On Electrical Engineering And Photonics (Eexpolytech)* (2020): 156-159.

[12] Samundeeswari, P. And Ramalingam Gunasundari. "A Novel Multilevel Hybrid Segmentation And Refinement Method For Automatic Heterogeneous True NSCLC Nodules Extraction." *2020 5th International Conference On Devices, Circuits And Systems (ICDCS)* (2020): 226-235.

[13] Karthikeyan, M. V. Et Al. "Lung Cancer Detection Using CT SCAN Images." *International Journal Of Advanced Research In Science, Communication And Technology* (2021): N. Pag.

[14] Fenimore, C. , Armato, S. , Aberle, D. , Brown, M. , Henschke, C. , Mcnitt-Gray, M. , Macmahon, H. , Mclennan, G. , Meyer, C. , Reeves, A. And Yankelevitz, D. (2011), The Lung Image Database Consortium (LIDC) And Image Database Resource Initiative (IDRI): A Completed Reference Database Of Lung Nodules On CT Scans, Medical Physics

[15] Miah, M.B.A., & Yousuf, M.A. (2015) "Detection Of Lung Cancer From CT Image Using Image Processing And Neural Network." 2015 International Conference On Electrical Engineering And Information Communication Technology (ICEEICT): 1-6.