CT based framework for Early detection of Lung Cancer

Sufia Ali
Computer Systems Department
Mehran Univeristy of Engineering and Technology
Jamshoro, Pakistan
sufiaali66@gmail.com

M. Moazzam Jawaid
Department of Computer Systems Engineering.Mehran Univeristy of Engg and Technology
Jamshoro, Pakistan
moazzam.jawaid@faculty.muet.edu.pk

Atiya Jokhio
Computer Science Department
Fast-National University of Computer & Emerging Sciences
Karachi, Pakistan
atiya.jokhio@nu.edu.pk

***Abstract*— Lung cancer in Pakistan is at number three for the cause of death [1].It has affected everyone irrespective of smokers or nonsmokers worldwide. If cancer is detected at an early stage there is a high probability for patient's life. CT scan plays a vital role for inspection and detection of lung cancer. For interpretation and evaluation of such images, there exits Computer aided Diagnostic (CAD) systems. CAD systems have emerged to reduce the overhead work for the radiologists. In CT interpretation, there occur a lot of difficulties. For example, the limitation of the human visual system, insufficient training and experience. A high number of cancers are missed due to the factors like distraction or fatigue. Hence, a second opinion is usually required to reduce erroneous rates. CT is one of the most used and effective method for diagnosing lung cancer.**

***Keywords*— CAD; ROI, Computed Tomography, Lung Nodule.**

I. INTRODUCTION (*HEADING 1*)

Lung cancer in Pakistan is at number three for the cause of death [1]. It has affected everyone irrespective of smokers or nonsmokers worldwide. There is a low survival rate for cancer patients in our region. If cancer is detected at an early stage there is a high probability for patient's life.

Computed Tomography (CT) is able to scan inside of the body and provide detailed information in the form of a pictures by usage of advance X-rays equipment's. CT scan plays a vital role for inspection and detection of lung cancer.

For interpretation and evaluation of such images, there exits Computer aided Diagnostic (CAD) systems. CAD can be further categorized as:

- CADe (computer-aided detection system)
- CADx (computer-aided diagnosis system)

CAD systems have emerged to reduce the overhead work for the radiologists. In addition, it offers a second opinion to them for better decision making.

II. LITERATURE REVIEW

In paper [3], Qing Wu and Wenbing Zhao have proposed a detection method for small cell lung cancer (SCLC) with machine learning algorithm that is based on neural network referred as entropy degradation method (EDM). SCLC detection was treated on the basis of binomial problem. Vectorized histogram of scans from both groups was used to train the algorithm. Their system showed accuracy of 77.8% which indicates it still requires improvement.

In paper [4], a feature subset selection method upgraded correctness of false positive reduction (FPR) for computer-aided detection (CAD) of nodule. The algorithm defined, automatically selected feature set of optimal size and determined the most significant feature. Different downscaling method was applied based on Tomek links After testing, the system showed 56.4% accuracy for detection of false positive reduction rate.

In paper [5], Rachid Sammouda have proposed a method based on Hopfield Artificial Neural Network Classifier that could identify early stage lung cancer. First, the lung portion is drawn out by bit plane slicing and further preprocessing. Segmentation process is done by a artificial neural network of Hopfield model (HNN). His method for segmentation with Unsupervised Hopfield Neural Network Classifier (UHNNC) gave satisfactory results.

In paper [6], incorporates a classification method relying multivariate multinomial Bayesian. Preprocessing of images was done by Histogram Equalization. To examine early stage cancer of a patient feature selection and neutral network classifier were used. For extraction of features gray level co-occurrence matrix (GLCM) was applied. From pool of features, distinct features are necessary so SFS (Sequential Forward Selection) was incorporated. Kernelised Bayesian classified test image lies in normal or abnormal category.

III. DESIGNED CAD SYSTEM

A. Maintaining the Integrity of the Specifications

Computed Tomography (CT) is able to scan inside of the body and provides detailed information in the form of pictures by usage of advance X-rays equipment's. CT scan plays a vital role for inspection and detection of lung cancer. Mainly, there are four stages for the detection of lung nodule in CADe systems. preprocessing, segmentation of lungs, nodule detection and the reduction of false positives (FP) [2]. The data is acquired from Cancer Imaging Archive (TCIA) Public Access website [8]. It was provided for SPIE-AAPM Lung CT Challenge. We have used 10 different CT scan dataset for our study. In the proposed approach, following are the principal operations:

ISBN-978-969-23372-1-2

A. Preprocessing

Vessels, surrounding tissues and background must be excluded as to improvise the correct detection of the suspected region. (Fig. 1).

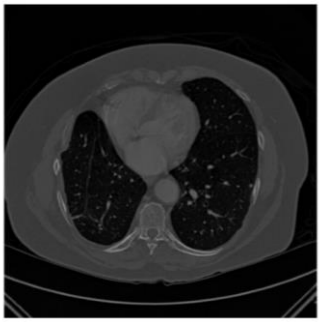

Fig.1. Original CT image

CT scan uses X-ray attention that have HU units. Background and other unnecessary data is being removed so that we can separate lung parenchyma. So for this purpose -500 HU has been used as initial threshold. This value was selected on the basis of histogram of the CT scan image. As a result, lung parenchyma tissue is separated.

Subsequently, for the elimination of holes in the lung tissue dilation and then erosion morphological operations are applied. Hence, lung parenchyma is acquired. (Fig. 3)

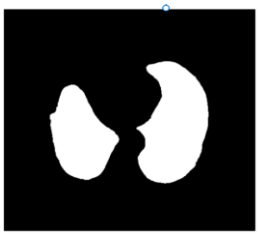

Fig. 3. Lung Parenchyma mask

Accordingly,lung mask is applied with the original CT image and using the initial HU threshold (Fig. 4). we begin processing only with the tissue within lung parenchymal tissue obtained in Fig.4.

Fig.4. Binarized lung module candidates

For the determination of False positive and Lung_RADS system, the information provided in SPIE-AAPM Lung CT Challenge database was used. Later, the binary mask of nodule candidates is obtained (Fig. 5).

Fig. 5. FP reduction step

For obtaining the ROI candidates of lungs, candidates of mask are applied to the original CT scan image. (Fig. 6).

Fig. 6. Segmented ROIs

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this study, SPIE-AAPM Lung CT Challenge dataset have been used for evaluation. The dataset is accessible through the National Cancer Institute's (NCI) The Cancer Image Archive (TCIA). Total of 70 thoracic CT scans (10 for a calibration and 60 for test) are present, which includes DICOM format images, spatial coordinates of the nodule locations and the diagnosis for each nodule in the calibration and test datasets. The dimension of this dataset is 512 x 512 pixels, having a bit depth of 12 bits.

Table 1 presents the information about the dataset and the results obtained after the application of CADs system.

Here are some pictorial views of the results that have been found.

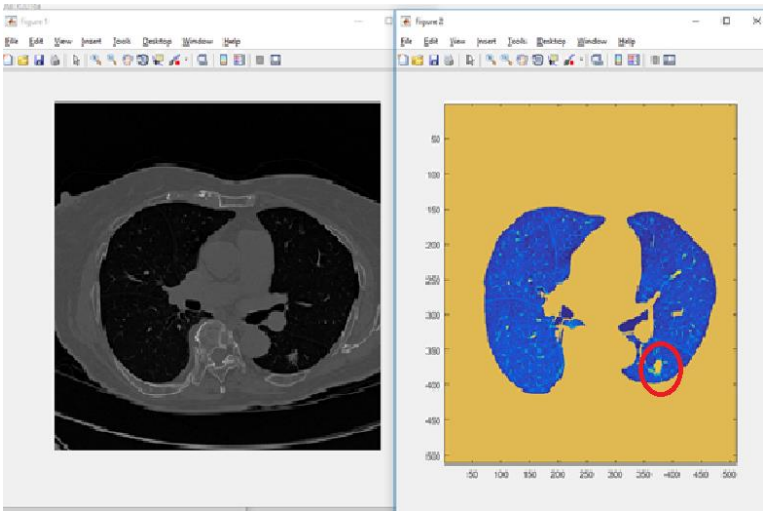

Fig. 7. Patient 1's dataset Result.

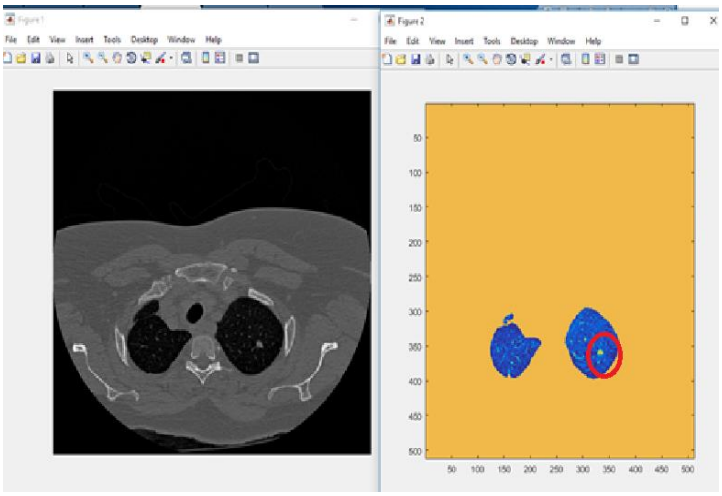

Fig. 7. Patient 2's dataset Result.

V. CONCLUSION

We present a system that can be used for the segmentation and detecting the lung nodules in CT scan images with an average of 80% accuracy. Our system is able to contribute to:

A. It can automatically choose the lung nodule parenchymal tissues exactly.
B. It is also able to detect the suspected region in the lungs effectively.
C. Our system shows enough good performance.

Image Dataset	Resolution	Ground Truth	Found at	Difference
Patient 1	512*512	(374,374,146)	(374,375,149)	-3
Patient 2	512*512	(197,290,41)	(198,290,42)	-1
Patient 3	512*512	(333,356,284)	(334,356,282)	+2
Patient 4	512*512	(128,362,131)	(129,362,130)	+1
Patient 5	512*512	(134,300,214)	(135,300, 212)	+2
Patient 6	512*512	(172,208,80)	(173,208,78)	+2
Patient 7	512*512	(328,242,54)	(327,242,55)	+1
Patient 8	512*512	(123,234,212)	(123,233,214)	+2
Patient 9	512*512	(405,296,163)	(405,296,164)	-1
Patient 10	512*512	(121,98,191)	(122,98,193)	-2

In future, we aim to reduce the FPs with the application of new segmentation techniques.

REFERENCES

[1] Pliro,Lung Cancer in Pakistan and its Treatments March 29,2019. Accessed on: November 20,2020 https://healthblog.pliro.com/lung-cancer-pakistan-treatments/

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," CA: a cancer journal for clinicians, vol. 66, no. 1, pp. 7–30, 2016.

[3] Q. Wu and W. Zhao, "Small-Cell Lung Cancer Detection Using a Supervised Machine Learning Algorithm," 2017 International Symposium on Computer Science and Intelligent Controls (ISCSIC), Budapest, 2017, pp. 88-91

[4] L. Boroczky, L. Zhao and K. P. Lee, "Feature subset selection for improving the performance of false positive reduction in lung nodule CAD," 18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05), Dublin, 2005, pp. 85-90.

[5] R. Sammouda, "Segmentation and Analysis of CT Chest Images for Early Lung Cancer Detection," 2016 Global Summit on Computer & Information Technology (GSCIT), Sousse, 2016, pp. 120-126.

[6] Mr. Sandeep A. Dwivedi , Mr. R. P. Borse , Mr. Anil M. Yametkar, "Lung Cancer detection and Classification by using Machine Learning & Multinomial Bayesian", IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 9, Issue 1, Ver. III (Jan. 2014), pp. 69-75

[7] N. Guo, R. Yen, G. E. Fakhri and Q. Li, "SVM based lung cancer diagnosis using multiple image features in PET/CT," 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), San Diego, CA, 2015, pp. 1-4

[8] http://www.cancerimagingarchive.net/