International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-1, October 2018

Enhanced Detecting System for Computer-Aided Diagnosis of CT Lung Cancer

Umar S. Alqasemi, Ahmed A. Qashgari, Mukhtar M. Alansari

Abstract: According to The American Cancer Society, In the US. they estimated that there will be 160,390 deaths from lung cancer and that 213,380 new cases will be diagnosed in 2007. The accurate diagnosis of the lung nodule in case of determine either malignant or benign requires a lot of resources to read the ever increasing volume of detecting nodules on high resolution computed tomography (HRCT). So, the motivation for developing and enhancing the performance and accuracy of computer-aided-diagnosis systems (CADx) for HRCT has been the focus of many research groups to alleviate this burden. In this work, we report the results of a new CADx system that provides enhanced detecting performance. We utilize an optimized set of texture features using feature selection and apply the new system using Datasets provided by National Cancer Institute (NCI)'s and The Cancer Imaging Archive (TCIA). The results compared well to previous work with classification accuracy performance of 89.4% and 86% sensitivity and 93% specificity. The implementation details and analysis of the proposed system are described in this paper. The results of the new system will contribute a significant impact on the accuracy of such systems and hence the enhancement of detecting process outcome.

Keywords: Computer-Aided Diagnosis, Digital Mammogram, Gabor Wavelets Feature, Support Vector Machines, Classification.

I. INTRODUCTION

According to The American Cancer Society, In the US. they estimated that there will be 160, 390 deaths from lung cancer and that 213,380 new cases will be diagnosed in 2007. The accurate diagnosis of the lung nodule in case of determine either malignant or benign requires a lot of resources to read the ever increasing volume of detecting nodules on High Resolution computed tomography (HRCT) [1]. Many research centers focus their effort on developing in Computer Aided Diagnosis system(CADx) [2]. This effort is resulting in many advances in CAD system. The CAD system is improving the detection of cancer by helping the physicians to focusing on the abnormality area and will decreasing the workloads and eliminate waste time also decreasing the human error to increase the quality of care.

Manuscript published on 30 October 2018.
* Correspondence Author (s)
Umar S. Alqasemi, Department of Biomedical Engineering, King Abdul-Aziz University, Jeddah, Saudi Arabia, E-mail: ualqasemi@kau.edu.sa
Ahmed A. Qashgari, Department of Biomedical Engineering, King Abdul-Aziz University, Jeddah, Saudi Arabia, E-mail: ahmed.q@drayaa.com
Mukhtar M. Alansari, Department of Biomedical Engineering, King Abdul-Aziz University, Jeddah, Saudi Arabia, E-mail: mukhtar.rad@gmail.com

Many of the previous research centers collecting huge of data by them self on real patients and volunteers such as the Data used in this research were obtained from The Cancer Imaging Archive (TCIA) sponsored by the SPIE, NCI/NIH, AAPM and The University of Chicago[3]. In this study we use the database provided for applying Machine learning for the 66 abnormal nodules less than 3 cm (33 Benign and 33 malignant). In literature, Several studies have been conducted to evaluate the effects of CADx on radiologists' accuracy for characterization of malignant and benign lung nodules; they are summarized in Table 3. Matsuki et al. [4] training and classifying the 155 benign and malignant cases of high resolution CT by using the power of artificial neural network to get high differentiating performance output model to aid radiologists to increase their accuracy using receiver operating characteristics ROC analysis . Li et al. [5] training and classifying 28 malignant and 28 benign nodules. They used linear discriminant classifier for distinguishing the malignant from the benign nodules with an Area under the curve (AUC) of 0.831 that used as an aid. They noticed that the performance of radiologists increased with CADx. The average AUC improved significantly from 0.785 to 0.853. Shah et al. [6] they training and classifying 15 malignant and 13 benign nodules by 8 radiologists to evaluate the classification accuracy for The CADx system used image features as input to a decision tree classifier. The system achieved a sensitivity of 91% at a specificity of 67% . The area under the curve of each of the 8 readers increased with CADx. The average AUC for all readers increased significantly from 0.75 to 0.81 with the use of CADx output. Awai et al. [7] they using Neural Network ANN for classification of lung nodules and evaluated its outputs on radiologists' performance by using a data set of 18 malignant and 15 benign nodules. The average AUC of 19 reader aided with CADx increased significantly compared to the others who unaided with CADx. Way et al. [8] they developed an automated segmentation and classification CADx system for of lung nodules data set of 124 malignant and 132 benign nodules from 152 patients. The CADx system used to compare radiologists' performance without and with CADx.

II. MATERIALS AND METHODOLOGY

A. Database Source

In this study we use the database provided by The Cancer Imaging Archive (TCIA) [3]. The database contain 70 cases digitized Images. The database Images have size 512x512 in DICOM format divided to 3 classes (Malignant nodules, benign nodules and suspicious Malignant nodules), 10 cases for testing set and 60 cases for training set.

*Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)*

B. Pre-processing of TCIA DICOM Images

The nodules are actually in a 3-D shape but in this study we choice the Nodule Center Image in DICOM system its refers to the "Instance Number" of the image that demonstrates the largest extent of the nodule under consideration, In this data the series of CT images are taken without contrast agent and reconstructed for high resolution with slice thickness 1mm. We processing them by MATLAB software, we have 66 number of images in data set, including 33 benign cases and 33 malignant cases, all the images were cropped by 32 x 32 region of interest (ROI), these ROI including malignant and benign nodules those have size less than 3 cm.

C. Feature Extraction

Feature extractions are based on mathematical or statistical findings methods for distinguishing the details of pathology that correlated:

1. Higher Order Statistical Features

The Gray Level Co-occurrence Matrix (GLCM) method use to redistribute the gray level of the image to the new level to make a relation between a gray level of each pixel according to the distance and the angles of directions, Therefore, we use 8 gray levels and angle 0° with distance 1 that give us high performance. The definitions of the GLCM are given in [9].

2. First Order Statistical Features

We applying first order statistical features on the original images and The Gray Level Co-occurrence Matrix of each images, the features Including Mean, Quantiles (10, 30, 50, 70 and 90), Standard deviation, Variance, skewness, Integration ,mean of Integration, standard deviation of Integration, Derivative, mean of Derivative , Standard deviation of Derivative , Uniformity of histogram with 5 levels , Entropy of histogram with 5 levels, Uniformity of GLCM with 8 levels and Entropy of GLCM with 8 levels .

3. Spatial Domain Features

Two type of transformation were applied in this study, first we applying wavelet 2D decomposition with level 5 and “haar” wavelet [11]. And we apply 2D Fourier Transformation (2D-FFT).

D. Feature Selection

We want to select features who could adjacent between benign and malignant or the features who give us high accuracy, also the features who do not correlate with pathology and Features that are not independent. in this study we apply two methods. First, the sequential forward search (SFS) in this method we remove feature one by one and in each we measure the accuracy, if the accuracy decrease that mean this feature is bad so eliminate it and so on. second The sequential backward search (SBS) in this method we looking for the best feature by running each of them alone and then rank them and choose the best one, second we take the best one and running with it the other features one by one individually and select the best feature who increase the accuracy and so on. Finally the best features were varies according to the type of classifier.

E. Classification

For choosing the classifier that provides best performance, we use three different types of classifiers as follows support vector machine (SVM), Artificial Neural Network Pattern Recognition (ANN) and k-voting nearest neighbour (KNN) classifiers. We used the features of testing set and training set which consisted of 14 malignant ROIs and 14 benign ROIs for testing and 19 malignant ROIs and 19 benign ROIs for the training set, but for ANN classifier we training by the 60% of data and validating by 20% and testing by 20%.

III. PERFORMANCE ASSESSMENTS

To evaluate classification performance, we calculate Error Rate, Confusion Matrix, Specificity, Sensitivity, False Negative Rate, the Area Under Curve (AUC) from the Receiver operator characteristic(ROC) and overall Accuracy [10]. We focus on the accuracy to make a comparison with the literatures and have to make sure that the values of Specificity and Sensitivity are near as much as possible to ensure the reliability of the diagnosis system of the computer-aided diagnosis of computed tomography.

IV. RESULTS AND DISCUSSION

The TCIA database High Resolution Computed Tomograpy Images were binary classified as Benign vs Malignant cases by features extractions were extrated the seven features of the first and higher order statistical as mentioned in feature selection paragraph and the evaluation of the three classifier (support vector machine (SVM), Artificial Neural Network (ANN) and K-voting Nearest Nieghbor (KNN)) were learning the system by 58% of data sets and testing them by 42% all these processing have been carried out in MATLAB software, furthermore the same selected features parameters are applying on each classifiers as shown their results below in Table.1.

Table.1 The Performance Results of CADx System

Classifier	Sensitivity	Specificity	AUC	Accuracy
SVM-Polynomial	80%	84.62%	74.5%	82.14%
ANN	86.11%	93.33%	~93.8%	89.39%
KNN -K=5	85.71%	85.71%	84.69%	86%

The confusion matrix for the selected features classified by KNN classifier shown in table.2 that the 28/33 of the Benign nodules cases were diagnosed as a Benign cases and the 31/33 of the Malignant nodules cases were diagnosed as an Malignant cases, which is good, furthermore 2 of Malignant cases were diagnosed as a Benign cases and 5 nodules of Benign cases were diagnosed as an Malignant which is acceptable in case of distinguishing between two abnormalities lung nodules unaided by radiologist.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Table. 2 The Confusion Matrix for the ANN Classifier

Actual	Predicted	
	Benign	Malignant
Benign	(28 nodules)	(5 nodules)
Malignant	(2 nodules)	(31 nodules)

The researchers applying a varieties of classifiers and features extraction to enhance the accuracy of detection in the computer aided diagnosis CADx of a computed tomography scanning a lung nodules, the researcher studies shown in Table.3 We compared our method results shown in Table.1 with literatures as. Henschke et al. they use 28 nodules (14 benign and 14 malignant) with slice thickness 5 mm, The statistical-multiple object detection and location system (S-MODALS) classifying by Neural Network technique was developed for automatic target recognition (ATR) to distinguish between malignant and benign nodules[12].

Table.3 Comparison of Different Performance Measurement for Classifying TCIA Database of Binary Class (Malignant– Benign).

Study	Techniques		Classification Performance %			
	Classifier	Features	Sens	Spec	AUC	Acc
[12]	ANN	S-MODALS	100	79	~79	89
[13]	LDA	correlation and difference entropy	88.2	92.2	-	90.3
[14]	LDA	GLCM	89	92	99.2	90.6
[15]	ANN	MTANN	100	48	88.2	-
Our study	ANN	Multi-features	86	93	94	89.4

McNitt-Gray et al. they use 31 nodules (14 benign and 17malignant) with slice thickness ≤3 mm and nodules size (5-30mm), solitary nodules were identified using semiautomated contouring techniques for using the varieties of nodule size, shape, attenuation and texture as a features and applying linear discriminant analysis[13].

McNitt-Gray et al. they apply GLCM on the 32 nodules (13 benign and 19 malignant) with slice thickness ≤3 mm and nodules size (5-30mm , mean=17mm). and classify them by linear discriminant analysis[14]. Suzuki et al. they use 489 nodules (413 benign and 76 malignant) with slice thickness 10 mm and nodules size <30 mm. to distinguishing between benign and malignant nodules in Low Dose CT scans by use of a massive training artificial neural network (MTANN)[15].

Regarding to the results of literatures contributions shown in table.3 the researchers attempt to enhance the performance of distinction between malignant and benign nodules by manipulating in many features and classifiers, the researchers notice that the preprocessing of the images could play main roles in enhancement, some of them try to use high resolution CT images with slice thickness less than 1mm, further the size and shape of nodules might leads to know which the type of nodules. In this paper we used three types of classifies as shown in table.1 the ANN and KNN classifiers gave high performance regarding to our experiments and compared to literatures in table.3, therefor we have to select the classifier who has the lower false negative rate(FNrate), that means the classifier who could increase the detection of Malignant nodules because if the patient have a malignant nodules and the CADx system diagnosed it as a benign that will allow the cancer rapidly spread and might cause death. So the FNrate for ANN AND KNN are 13.89% and 14.29%, respectively.

The features selected for ANN are (mean, standard deviation, variance, skewness, percentile, mean of Integration, standard deviation of Integration, percentile of Integration, mean and standard deviation of differentiation,(uniformity ,entropy and mean of histogram),(uniformity and entropy of gray level co-occurrence matrix), (uniformity , mean, standard deviation, percentile of 2D wavelet decomposition) and (mean, standard deviation, percentile of FFT), those features gave high performance for our study .

V. CONCLUSION

In this study we designed a computer-aided diagnosed system (CADx) that will enhance the accuracy of distinction between Malignant and benign nodules, so far we selected the best classifier among three different classifiers those built in function in MATLAB and selected features those obtained the better performance and then we classified the selected features by Artificial Neural Network Pattern Recognition method, further all these processing have been carried out in MATLAB software. This CADx system will aided radiologists to avoid invasive procedures that required biopsy and histology test that will effect on the healthcare of patients and waste time and money.

ACKNOWLEDGMENT

Authors acknowledge their sincere thanks to Dr. Yasser kadah, Associate Professor, Department of biomedical engineering, king Abdul-Aziz University for his valuable comments and suggestion.

REFERENCES

1. American Cancer Society. Cancer facts & figures. 2007.
2. H.-P. Chan, L. Hadjiiski, C. Zhou, and B. Sahiner, "Computer-Aided Diagnosis of Lung Cancer and Pulmonary Embolism in Computed Tomography - A Review," Academic radiology, vol. 15, pp. 535-555, 2008.
3. Armato III, Samuel G.; Hadjiiski, Lubomir; Tourassi, Georgia D.; Drukker, Karen; Giger, Maryellen L.; Li, Feng; Redmond, George; Farahani, Keyvan; Kirby, Justin S.; Clarke, Laurence P. (2015). SPIE-AAPM-NCI Lung Nodule Classification Challenge Dataset. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2015.UZLSU3FL.
4. Y. Matsuki, K. Nakamura, H. Watanabe, T. Aoki, H. Nakata, S. Katsuragawa, and K. Doi, "Usefulness of an Artificial Neural Network for Differentiating Benign from Malignant Pulmonary Nodules on High-Resolution CT," American Journal of Roentgenology, vol. 178, pp. 657-663, 2002/03/01 2002.
5. F. Li, M. Aoyama, J. Shiraishi, H. Abe, Q. Li, K. Suzuki, R. Engelmann, S. Sone, H. MacMahon, and K. Doi, "Radiologists' Performance for Differentiating Benign from Malignant Lung Nodules on High-Resolution CT Using Computer-Estimated Likelihood of Malignancy," American Journal of Roentgenology, vol. 183, pp. 1209-1215, 2004/11/01 2004.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

6. S. K. Shah, M. F. McNitt-Gray, K. R. De Zoysa, J. W. Sayre, H. J. Kim, P. Batra, A. Behrashi, K. Brown, L. E. Greaser, J. M. Park, D. K. Roback, C. Wu, E. Zaragoza, J. G. Goldin, R. D. Suh, M. S. Brown, and D. R. Aberle, "Solitary pulmonary nodule diagnosis on CT," Academic Radiology, vol. 12, pp. 496-501.
7. K. Awai, K. Murao, A. Ozawa, Y. Nakayama, T. Nakaura, D. Liu, K. Kawanaka, Y. Funama, S. Morishita, and Y. Yamashita, "Pulmonary Nodules: Estimation of Malignancy at Thin-Section Helical CT—Effect of Computer-aided Diagnosis on Performance of Radiologists," Radiology, vol. 239, pp. 276-284, 2006.
8. Way, TW.; Chan, HP.; Stojanovska-Nojkova, J., et al. RSNA Program Book. Vol. 2007. 2007. Effect of computer-aided diagnosis (CAD) on radiologists' characterization of lung nodules on CT: An ROC study; p. 267.
9. R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. Syst., Man, Cyb., vol. 3, pp. 610–621, Nov. 1973.
10. http://www.saedsayad.com/model_evaluation_c.htm.
11. S. Bouyahia, J. Mbainaibeye, N. Ellouze, Wavelet based microcalcifications in digitized mammograms, Icgst-Gvip J. 8 (2009) 23-31.
12. C. I. Henschke, D. F. Yankelevitz, I. Mateescu, D. W. Brettle, T. G. Rainey, and F. S. Weingard, "Neural networks for the analysis of small pulmonary nodules," Clinical imaging, vol. 21, pp. 390-399, 1997 Nov-Dec 1997.
13. M. F. McNitt-Gray, E. M. Hart, N. Wyckoff, J. W. Sayre, J. G. Goldin, and D. R. Aberle, "A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: Preliminary results," Medical Physics, vol. 26, pp. 880-888, 1999.
14. M. F. McNitt-Gray, N. Wyckoff, J. W. Sayre, J. G. Goldin, and D. R. Aberle, "The effects of co-occurrence matrix based texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography," Computerized Medical Imaging and Graphics, vol. 23, pp. 339-348.
15. K. Suzuki, L. Feng, S. Sone, and K. Doi, "Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network," IEEE Transactions on Medical Imaging, vol. 24, pp. 1138-1150, 2005.

AUTHOR PROFILE

Umar S. Alqasemi, Assistant Professor of Biomedical Engineering at King Abdul-Aziz University, research on biomedical engineering application, having more than twenty published papers, more than fifteen papers review and auditor, interested on photo acoustic imaging, CAD imaging, signal processing and more, guider on real life of biomedical business sector at KAU.

Ahmed A. Qashgari MSc. Electrical and Biomedical Engineering, two scientific paper and ten articles, worked on several research on test equipment calibration of biomedical engineering and CAD system of medical imaging, board members on Saudi scientific society of biomedical engineering, two applied research executed in real market with GE healthcare and Saudi government hospitals, interested on biomedical engineering application sectors.

Mukhtar M. Alansari MSc. Electrical and Biomedical Engineering, three published papers, research on biomedical signals and imaging processing, brain computer interface and tissue engineering, co-founder of nanoparticle manufacturing

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)