Detection of Lung Nodules in CT Images Using Features fusion and Genetic Algorithm

Hanan M. Amer,Hamdi A. A. Elmikati,Fatma E.Z. Abou-Chadi,Sherif S. Kishk, and Marwa I. Obayya

Abstract— The aim of this study is to increase the accuracy of early detection of pulmonary nodules through the development of a Computer-Aided Detection (CAD) system. A comparative study of performance for the most commonly used techniques in feature extraction and classification was performed to identify the technique that gives the highest accuracy and lowest false positives. This study was conducted on the pulmonary nodule candidates directly without removing the blood vessels to design a precise automatic detection system to detect the pulmonary nodules in early stage. The main significant features of the pulmonary nodule candidates are extracted by using four feature extraction technique: Histogram of Oriented Gradients (HOG) features, Computerized Tomography (CT) Value Histogram (VH) features, texture features of Gray Level CO-Occurrence Matrix (GLCM) based on wavelet coefficients, and statistical features of first and second order. To make use of the extracted features, a feature fusion technique was used to concatenate the extracted features together and select features in a new hybrid feature vector. A Genetic Algorithm (GA) search based on the classification accuracy rate (CAR)of the utilized classifier was also applied to the hybrid feature vector. To get the highest classification accuracy, three classifiers were selected and their performance was compared. These are: Artificial Neural Network (ANN), Radial Basis Function Neural Network (RBF-NN) and Support Vector Machine (SVM). Three parameters were used to compare the classifier performance: the classification accuracy rate (CAR) , the sensitivity (S) and the Specificity (SP). The results have shown that using the selected hybrid features vector and the SVM classifier gives the highest CAR of 99.6% and a 0.008 false positive per scan.

Index Terms— Computer-aided detection, computed-tomography, discrete wavelet transform, principal component analysis, CT value histogram, histogram of oriented gradients, gray level co-occurrence matrix.

---

1 Introduction

As a result of the high rates of air pollution and the spread of smoking in recent years, lung cancer has become one of the most important diseases that pose a great threat to humanity because of the high rates of infection and the difficulty of treatment and rapid spread. Developing new techniques in the treatment and early detection of this disease has become the concern of scientists in medical fields [1], [2].

As early detection of this disease increases the chance of survival of the patient for a period of up to 5 years by up to a percentage of 70%, as well as it increases the chance of success of treatment whenever diagnosed in the early stages, this led to the increasing importance of work on the development of early detection techniques [2], [3].

- *Hanan M. Amer is assistant teacher in department of electronics and communications, faculty of engineering, Mansoura University, Egypt. E-mail: hanan.amer@yahoo.com*
- *Hamdi A. A. Elmikati is professor in department of electronics and communications, faculty of engineering, Mansoura University, Egypt.*
- *Fatma E.Z. Abou-Chadi is professor in department electrical engineering, faculty of engineering, the British University in Egypt.*
- *Sherif S. Kishk is Assistant Professor in department of electronics and communications, faculty of engineering, Mansoura University, Egypt.*
- *Marwa I. Obayya is Assistant Professor in department of electronics and communications, faculty of engineering, Mansoura University, Egypt.*

One of the most important techniques used in the diagnosis of lung cancer is Computerized Tomography (CT) of the patient's chest. It is one of the most accurate methods, because it is a lung imaging on many sections, which results in this examination number large images, enabling radiologists and physicians to examine all parts of the lung [4].But as a result of the increase in the number of images resulting from this examination in addition to the use of low radiation doses during the examination to protect the patient from the risk of exposure to large amounts of radiation, all made the examination of these images by a radiologist difficult and onerous task [5].This motivated scientists to design computerized systems that process and analyze these images and allow automatic determination of the presence of pulmonary nodules. These systems are known as Computer-Aided Detection (CAD) systems [6], [7].

In general, any CAD system for the automatic detection of pulmonary nodules is composed of four main stages: a preprocessing step for image contrast enhancement and noise reduction. Then the automatic segmentation stage that aims to extract the human's lung area followed by a feature extraction procedure of the pulmonary nodule candidates in the digital CT images and the final stage is the classification[8].Fig. 1 demonstrates the basic stages of any CAD system.

The present paper presents a detailed description of the third and fourth stages: feature extraction and classification to detect accurately the pulmonary nodules. It is a continuation of our work which aims the development of a complete automatic CAD system for the detection of lung cancer in its early stages. The first and second stages of the system which are image preprocessing step and automatic segmentation system have been completed and published previously in [9]. The performance of the complete system is evaluated through the computation of the classification accuracy rate (CAR)using a set of images from the standard database available by the Early Lung Cancer Action Project (ELCAP) association [10].

Generally, in CAD systems, the process of extracting the pulmonary nodule candidates, is followed by eliminating the blood vessels and extracting the main features of the extracted candidates. However, the elimination of blood vessels can lead to the loss of small pulmonary nodules that are attached to the vessels (i.e., vascularized) [11]. To overcome the problem of possible loss of small pulmonary nodules, the proposed CAD system in the present work will detect the pulmonary nodules without eliminating the blood vessels and will utilize the extracted the features of the segmented pulmonary nodule candidates. The main focus is to extract the main significant features of the pulmonary nodule candidates to increase the detection accuracy of the pulmonary nodules.

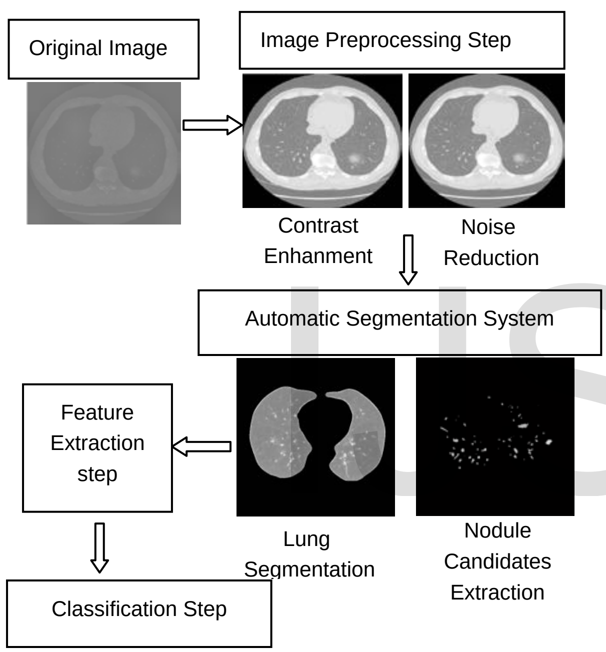


Fig 1: The block diagram of the basic stages of the CAD system.

Existing approaches in the literatures for feature extraction can be roughly categorized into geometric features: radius features, geometric shape features, major and minor axes, boundary and circularity information [4], [5], [12], [17], textural features: co-occurrence features, wavelets features, histogram of oriented gradients (HOG) features[6], [7], [13], [14], [18] and intensity-based features: intensity distributions, CT value histogram (CVH), edge-gradient features and so on [5], [7], [17], [20]. In the field of pulmonary nodules detection, the textural features and intensity-based features are used for potential detection since geometric features need to deal with nodules of fixed geometrical features [7], [8].

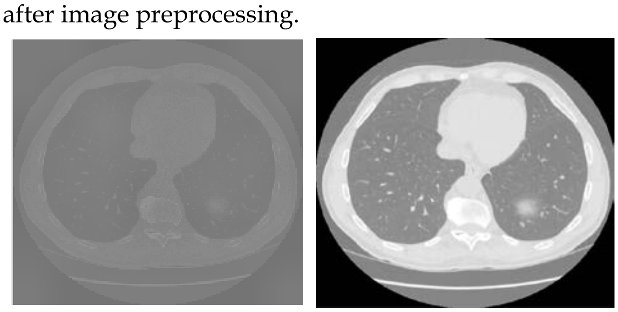
Recently, several studies have been carried out with the aim of increasing the accuracy of classification systems. The most important of these methods were the fusion techniques, which can be classified into three different levels, namely, data fusion, fusion at the level of features, and fusion at the decision level. Data fusion where data is fused from a different data sources to produce a new set of data with higher content of information [53]. decision Fusion where the classification decision of a group of classifiers, the classifiers can be of the same type or of different types using the same set of features or different types of them in order to obtain the results of the classification of accurate and unbiased [54]. feature fusion where different features types are selected and combined with the aim of producing a new set of features with removing correlated and redundant information to increase the classification accuracy of the classifiers. In this work will be applied features fusion[55].

For the classification stage, classifiersare used to detect the pulmonary nodules where they utilize the extracted hybrid features vector of the pulmonary nodule candidates as an input to the classifier and its output is to classify this input pattern as nodules or non-nodules. Several classification approaches have been proposed such as Artificial Neural Networks (ANN) [4], [5], linear discriminate analysis classifier [23], [25], rule-based [24], Bayesian classifier [12], [13], [14], support vector machine (SVM) [5], and k-NN [15], [16].

The performance of the proposed system is compared with that of previous reported classifiers: Artificial Neural Networks (ANN) classifier [4], [5], [6], Radial Basis Function Neural Network (RBF-NN) classifier [31] and Support Vector Machine (SVM) classifier [4]. The organization of the paper is as follows: Section 2 is a description for the proposed system. Section 3 presents the performance evaluation and Section 4 is the final conclusion.

2 THE APPLIED IMAGES

The proposed CAD system has been applied to 40 CT scans containing 320 regions of interest (ROI) available by the Early Lung Cancer Action Project (ELCAP) association [10].The images in this database are available in format of Digital Images and Communication in Medicine (DICOM) and have a resolution of 0.76×0.76×1.25. The size of pulmonary nodules that consider in this work varies from 3 mm to 30 mm with different types including: solid, non-solid, part-solid, juxta-vascular, well-circumscribed, and juxta-pleural pulmonary nodules. Fig.2 shows atypical example of the chest CT images before and after image preprocessing.



(a) (b)

Fig 2: (a) A typical example of CT lung image from ELCAP database and (b) preprocessed image .

3 METHODS

As mentioned previously, any CAD system consists of four main stages; a preprocessing stage followed by an automatic segmentation procedure to extract the lungs and the pulmonary nodule candidates then the third stage is to extract the main features of pulmonary nodule candidates and finally the classification stage to detect the pulmonary nodules. The first two stages of the present work were published in [9]. The segmented pulmonary nodule candidates are used as the input to the third stage of the CAD system which is the feature extraction and selection stage. It aims at extracting the main significant features of the segmented pulmonary nodule candidates using advanced features extraction techniques and then selecting the most significant features. The resultant selected features are then used in classification stage which aims at detecting the pulmonary nodules by classifying the extracted pulmonary nodule candidates into nodules or non-nodules. Detailed description of these two stages is presented in the next subsections.

3.1 Feature Extraction

An essential procedure in a classification process is the feature extraction. Feature extraction is the process of defining a set of features, or characteristics, which will most efficiently or meaningfully represent the information that is important for analysis and classification [51]. The goal of feature extraction is to achieve significant data reduction and to determine informative measures. It involves dimensionality reduction in which data in high-dimensional space are transformed to a space of fewer dimensions. Resulting data (features) are intended to be informative, non-redundant, facilitating the subsequent learning and generalization steps.

In this work, four different techniques of feature extraction are used; are the Histogram of Oriented Gradients(HOG) features, the texture features of Gray Level CO-Occurrence Matrix (GLCM) based on wavelet coefficients, the Value Histogram(VH) feature, and the statistical features. These will be explained in details as follows:

1- The Histogram of Oriented Gradients (HOG) features are used to describe the texture of an image by calculating the gradients distribution of image parts in various orientations. This type of features was adopted first by Dalal and Triggs [33] where the input image is subdivided into small areas called "cells". For each cell, the pixels gradient is calculated to compose the histogram of the gradients. These gradients are describing the contours of objects, giving the dark shape and outline of image objects, describing the texture of the image objects and providing an information on the direction in which a maximum change in intensity is occurring around the pixel. Each of neighboring cells are grouping in block. To increase the accuracy, the cells histogram are normalized based on the histograms of the cells that are belong to the same block. This normalization gives better invariance to the shading and lighting changes.

The HOG was calculated for each nodule candidate image. The resulting feature vectors have large dimensions that may reach up to 1806336 for each image. To decrease the feature vectors dimension and eliminate the information redundancy, the Principal Component Analysis (PCA) was employed, so that data is to be more efficient in manipulated and stored[30].

PCA is proposed by Hotelling and it is a mathematical method [31].The PCA is a transformation algorithm. The inputs to the PCA algorithm are a collection of correlated variables and the outputs are collection of linearly uncorrelated variables by applied an orthogonal transformation these uncorrelated variables called principal components. By applying the PCA, There is an extra advantage next to getting on an uncorrelated variables. It is the dimension of output data from PCA is smaller than input data due to getting the main components which are comes from organizing the eigenvalues that calculated from the covariance matrix of the input data in order [32].After applying the PCA, the feature vector size of each ROIs reached to the dimension of 150.

2- Texture features of Gray Level CO-Occurrence Matrix (GLCM) based on wavelet transform coefficients: the wavelet transform is an effective and popular method to get the wavelet descriptors that are used in texture analysis, because of their efficiency in detecting a localized information from spatial and frequency domain in addition to its features scale to get the multiresolution characteristics [37]. In this work, wavelet features have been calculated as follows; the nodule candidates patterns are decomposed in to two levels by applying the 2-D Daubechies (db2) of wavelet transform. Then, each of the horizontal, diagonal, vertical detailed were obtained from wavelet decomposition structure of decomposition level One and Two in addition to the approximate coefficients of level Two.

After get the wavelet coefficients; the texture features of GLCM were calculated from them. The GLCM is a two-dimensional statistical dependence matrix based upon the spatial relationship between adjacent or neighboring pixels to describe the main texture information of an image [52]. First, the GLCM has been determined for each wavelet sub-bands of level one and two to characterize the nodule candidates. The GLCM was used in the many previous works to extract the texture features of the pulmonary nodule candidates because of their efficiency to catch the spatial dependence of image gray level intensity values [34], [35]. The multiresolution analysis of wavelet transform allows to get the characterization features around the pulmonary nodule candidates in various levels and scales, furthermore applied the GLCM for each wavelet sup-bands then calculated the second order statistical texture properties of each GLCM to get amore characterized and un-correlated texture features based on wavelet descriptors of the pulmonary nodule candidates to enhancement the classification accuracy [36].The computed second order statistical features are entropy, contrast, energy, correlation and homogeneity. There are a set of 5 features calculated For each wavelet sub-band, getting a total of 35 features.

3- The Value Histogram(VH) feature meaning the Histogram distribution for the CT image intensity values[26]. The VH features are extracted by constructing the histogram distribution of CT image pixels intensity values. The features vector size is specified by the number of extracted histogram bins which is determined according to the number of bins that gives the best Classifier Accuracy Rate (CAR) of the Support Vector Machine(SVM) classifier. In this way, different VHs were constructed with various bins number [26]. For each VH features vector of specified bins number is represent as an input feature vectors for the SVM classifier to determine the corresponding CAR. Finally, the number of bins which lead to the best CAR of the SVM classifier will be adopted to extract the VH features in this present work. In the present work, the optimum number of bins was found to be 16.

4- Statistical Features: The statistical calculations are described the texture of an object in the image in an indirect manner upon the non-deterministic properties which manage and describe the relationships between the gray levels intensity values of an image. Two categories of statistical features were used: first-order and second-order statistical features. The histogram calculation of the image grey level intensity values has been used to determine the first-order statistic features. The first-order statistic features that are determined from this histogram comprise the following; average gray level, skewness, variance, kurtosis, entropy, and the foreground area pixels [41]. The statistical properties of the Gray Level Co-occurrence Matrix (GLCM) that were calculated for the grey level intensity image are Correlation, homogeneity, energy, contrast and entropy. These are considered as the second-order statistic features and were calculated by applying the Haralick transformation [41]. The resultant statistical features vector consists of 11 features.

3.2 Feature Fusion

Features are the key to the success of the classification process and therefore the selection of the optimal set of features leads to the success of the classification process. In the process of features fusion, a new set of features is created from different sets of features from different domains after removing the insignificant and noisy features [55]. The process of feature fusion is a kind of feature selection, where the process of selecting features is to exclude features that are correlated, insignificant or distorted which in turn reduces the time of computation and increases the classification accuracy of the system. In the present work, the four different features vector Histogram of Oriented Gradients (HOG) features, Value Histogram (VH) features, texture features of Gray Level Co-Occurrence Matrix (GLCM) based on wavelet coefficients, and statistical features were used separately and then fused in a new hybrid feature vector by applying a simple concatenation to improve the classification performance of the system. The performance of each feature set and the new hybrid feature vector was then compared.

Having formed the hybrid feature vector, the next step is to remove any redundant and correlated information which is known as “feature selection”. The removal of the redundant information leads to increase the classification accuracy and reduce the processing data size which in turn improves the speed of classification algorithm by decreasing the burden of calculation [27], [42]. The Genetic Algorithm (GA) was utilized as a feature selection step.

The Genetic Algorithm (GA) was developed originally by Holland [44]. Itis built on a three basic processes: a selection process that is based on the fitness function then it is followed by each of mutation, and crossover. According to the terminology of GA’s, the feature subset is called an individual or chromosome which consists of genes. Also, each of individual quality is measured by a fitness function. The fitness function which was selected in this work is the Classification Accuracy Rate (CAR) of the applied classifiers. The individual size in the GA population equal to the size of the hybrid features vector. The GA starts by first, the initial population is generated randomly which consists of the initial individuals. The individual in the GA's population takes the form of bit string (ones and zeros) where 1 means the corresponding feature in the hybrid features vector is selected and 0 means the corresponding feature is excluded. After the initial population was generated the GA's operates in form of frequent way in order to lead the population to the optimum point in accordance with the fitness function. For each GA's generation, the fittest individuals are chosen through the selection operation to act as the parent individuals to generate the new GA's generation by means of the main operations of the GA's: Reunification of chromosomal genes of each two selected parents (crossover), invert the genes inside an individual randomly from 1 to 0 or vice versa (mutated), and individuals which is taken as it is without change to the next GA generation is specified by the selection operation (elite) [45].

In the present work, the GA algorithm was applied to the hybrid feature vector as a feature selection technique using three classifiers: Artificial Neural Network (ANN), Radial Basis Function Neural Network (RBF-NN) and Support Vector Machine (SVM).

3.3 Nodules Detection

The final stage is to classify the resulted nodule candidates into nodules and non-nodules. Three classifiers were selected and their performance was compared. These are: Artificial Neural Network (ANN), Radial Basis Function Neural Network (RBF-NN) and Support Vector Machine(SVM). The inputs to each classifier are the feature vectors that is resulted from feature extraction and fusion stages. The classifiers are trained and their performance was evaluated. For the training and testing steps of each classifiers, 25% of the available data set size was used for the training phase and they were tested using 75% of the available data set size.

i. The ANN Classifier

The ANN's classifiers consist of three layers [32], [45], [46]. The first layer is the input layer which consists of nodes. The number of nodes depends on the features vector size. The number of neurons in the hidden layer can be specified by trial and error to increase the accuracy of learning process. The third layer is the output layer which consists a number of neurons according to the number of outputs. In the present work, it consists of two neurons according to the two output [0 0] for non-nodules and [1 1] for nodule. The Back-Propagation Algorithm (BPA) was chosen as a supervised learning algorithm [46], with a0.0001 target error, and a 0.9 learning rate.

ii. The Radial Basis Function Neural Network

The structure of the radial basis function neural network (RBF-NN) is made up of only three main layers and utilized a radial basis function as an activation function. The transformation from input to hidden layers is nonlinear. On the other hand, the output is a linear combination of radial basis functions of the inputs and neuron parameters[48].

iii. The SVM classifier

The SVM is a supervised learning methodology. It processed the input date using a kernel function. The results of classification obtained through the final summation with applying an activation function. The hyper-plane is used to classify the input data into two classes (binary classification). The vectors that are much closer to the boundaries is called the support vectors and the distance between hyper-plane and support vectors is called the margin. There are a many types of kernels using in SVM like linear, Radial Basis Function (FBF), and polynomial [45], [49].

Applying the GA technique as a feature selection tool results in a reduction in the size of feature vector as follows: in case of applying the ANN classifier, the number of features was reduced from 182 to 169 features, for the RBF-NN classifier from 182 to 115and in case of the SVM classifier from 182 to 153 features.

4 EXPERIMENTAL RESULTS

To assess the performance of the different proposed classifiers, the classification accuracy, sensitivity, and specificity measures were calculated [45] as follows:

Accuracy=[(TP+TN)/(TP+TN+FP+FN)]*100 (1)

Sensitivity=[TP/(TP+FN)]*100(2)

Specificity=[TN/(TN+FP)]*100 (3)

where FP, FN, TP, TN represent false positive, false negative, true positive, and true negative respectively [45]. The main target of this work is to obtain the highest accuracy, sensitivity, and specificity of the designed CAD system to detect of lung nodules in early stage. To achieve this purpose, applied the feature fusion and feature selection techniques on the extracted features. And to prove that the application of fusing features technology increases the accuracy, sensitivity, and specificity of the applied classifiers, for each classifier, the classification accuracy rate (CAR) , sensitivity (S) and specificity (SP) were calculated based on the extracted four types of features: Histogram of Oriented Gradients (HOG) features, Value Histogram (VH) features, texture features of Gray Level Co-Occurrence Matrix (GLCM) based on wavelet coefficients, and statistical features separately and hybrid feature vector. Tables 1,2,3 show the CAR,S, and SP obtained from the 3 different classifiers.

Table 1Theclassification accuracy rate (CAR) of the three classifiers using the four feature extraction techniques and the hybrid features.

Features \ Classifiers	ANN	RBF-NN	SVM
Wavelet Features	85.8%	87.5%	94%
VH	89.6%	95.4%	87%
HOG	63%	65.8%	71.6%
Statistical Features	83.7%	78.3%	95.8%
Hybrid Features	96.3%	97%	95%

Table 2 The sensitivity (S) of the three classifiers using the four feature extraction techniques and the hybrid features.

Features \ Classifiers	ANN	RBF-NN	SVM
Wavelet Features	**83.6%**	**95.9%**	**100%**
VH	**88.6%**	**96.6%**	**81.6%**
HOG	**62.8%**	**69%**	**67.8%**
Statistical Features	**98.8%**	**96%**	**99.1%**
Hybrid Features	**99.1%**	**100%**	**100%**

Table 3 The specificity (SP) of the three classifiers using the four feature extraction techniques and the hybrid features

Features \ Classifiers	ANN	RBF-NN	SVM
Wavelet Features	**88.4%**	**81.7%**	**89.6%**
VH	**90.6%**	**94.3%**	**95%**
HOG	**63%**	**63.6%**	**77.7%**
Statistical Features	**75.8%**	**70.5%**	**93%**
Hybrid Features	**94%**	**91%**	**95%**

Comparing the results of classification shown in the three tables, it is clear that the use ofthe feature fusion technique led to obtaining the highest classification results for the three classifiers. The CAR reached 96.3%, 97% and 95%, the S reached to 99.1%, 100% and100% and the SP reached to 94%, 91% and 95%with False Positive (FP) of .06, .1, and .058and the False Negative (FN) equal .008, 0 and 0 in the case of Artificial Neural Network (ANN), Radial Basis Function Neural Network (RBF-NN) and Support Vector Machine (SVM), respectively. Therefore, the adoption of feature fusion gives a significant impact on the improvement of the classification performance of the CAD system.

Applying the GA technique as a feature selection technique to the hybrid feature vector using the three classifiers results in reduction in the feature vector size. Table 4 depicts the resulted number of features before and after reduction and the CAR, S and SP corresponding to each classifier.

ISSN 2229-5518

Table 4 The classification accuracy rate (CAR), the sensitivity (S) and the specificity (SP) of the three classifiers and the size of feature vector after using the genetic algorithm (GA) technique.

Classifier	Hybrid feature size	Selected feature size	CAR	S	SP
ANN	182	169	99.6%	100%	99.2%
RB_NN	182	115	99.2%	100%	98.4%
SVM	182	153	99.6%	100%	99.2%

As clear from Table 4, the application of GA feature selection technique has increased the CAR, S and SPin addition to reducing the feature vector size. This has led to an improvement in the system performance and a reduction in the computational time. There is a significant decrease in the number of features resulted from using the RB_NN classifier but the CAR and SP are relatively lower than those of the other two classifiers ANN and SVM. While the result of the classification of both the ANN and SVM equal, but the number of features in the case of the SVM of less than in the case of ANN.

5 DISCUSSION

The main objective of the present work is the development of a Computer-Aided Detection system(CAD) for accurate detection of lung nodules in CT scans images. The main task was to obtain the highest detection accuracy of the pulmonary nodules. The system consists of four main stages. These are; image preprocessing stage to enhance the quality of the CT images, an automatic segmentation stage to automatically extract the human's lung and the pulmonary nodule candidates, a feature extraction and selection stage to get the most significant features and finally a classification stage to detect the pulmonary nodules.

Forty CT scans with 320regions of interest (ROI) were made available through the early lung cancer action project (ELCAP) association and used to train and test the performance of the proposed system.

In this proposed system, it were applied a novel Image Size Dependent Normalization Technique (ISDNT) [56] to enhancement the CT image contrast and the wiener filter [9] as a noise reduction technique to ameliorate the CT image quality in the image preprocessing stage.

For the automatic segmentation stage it was utilized the bi-level thresholding technique based on the optimal gray level threshold has been calculated by the proposed method and then it were employed each of median filter and the mathematical morphological operation to suppress any unwanted pixels [9].

In the third stage of the proposed system, four feature extraction techniques were utilize. These are: are the Histogram of Oriented Gradients (HOG) features, the texture features of Gray Level CO-Occurrence Matrix (GLCM) based on wavelet coefficients, the Value Histogram (VH) feature, and the statistical features of first and second order. A feature fusion step was employed on the four different set of extracted features to produce the hybrid features vector. The five feature vectors were used as an input to three types of classifiers and their performance was evaluated. The classifiers are: Artificial Neural Network (ANN), Radial Basis Function Neural Network (RB_NN), and Support Vector Machine (SVM). Each classifier was trained using 25% of the dataset and tested using the remained 75% of available data input.

The Classification Accuracy Rate (CAR), the Sensitivity (S), and the Specificity (SP)were calculated for each classifier using each of the five feature vectors. Comparing the CAR, S, and SP resulted from each classifier has showed that the hybrid features gave the highest CAR, S, and SP. This leads to conclude that the feature fusion technique increased the detection accuracy of pulmonary nodules and improves the system performance .

An attempt was made to increase the classification accuracy, enhance the system performance and to reduce the computational time using the Genetic Algorithm (GA)as a features selection algorithm on the hybrid features vector. The CAR, S and SP results of the three learned classifiers; ANN, RB_NN, and SVM showed an increase in the values of CAR, S and SP of the three classifiers. The CAR reached 96.3%, 97% and 95%, for the ANN, RB_NN, and SVM, respectively. Sensitivity S reached 100 for the three classifiers. As for Specificity (SP), it reached 99.2%, 98.4% and 99.2% for the ANN, RB_NN, and SVM, respectively. A reduction of the False Positive (FP) of 0.008, 0.017 and 0.008 have been obtained while the reduction in the False Negative (FN) reached0for the three classifiers: ANN, RB_NN, and SVM. Based on these results, it can be concluded that applying the (GA) as a feature selection technique to the hybrid feature vector composed of HOG features, GLMC texture features, the VH feature, and the statistical features of first and second order increases the classification performance of the system significantly. Both the ANN classifier and the SVM gave the same results of classification but SVM utilizes a smaller number of features and therefore, the computational time needed is less than that of the ANN.

Table 5 shows a comparison of the performance of the suggested system and five systems developed in previous researches to detect pulmonary nodules. It has been found that the suggested system achieves the best classification rate and the lowest false positives.

Table 5 The accuracy and false positive of the proposed system and other work in the pulmonary nodules detection

Work	Accuracy	False Positives
Choi et al. , 2013 [57]	**97.61%**	**2.27**
Kuruvilla et al. , 2.014 [58]	**93.30%**	**2.00**
Demir et al. , 2015 [59]	**90.12%**	**2.45**
Manikandan et al. , 2016 [60]	**94.00%**	**0.38**
Sweetlin et al. , 2017 [61]	**94.00%**	
Proposed system	**99.60%**	**0.008**

Still much work is needed for discriminating benign and malignant tumors of the lung nodules. This is the aim of the next stage of the work.

REFERENCES

[1] A. Jemal, T. Murray, E. Ward, A. Samuels, R. C. Tiwari, A. Ghafoor, E. J. Feuer, M.J. Thun, Cancer statistics, CA Cancer J. Clin. Vol. 55, pp. 10–30, 2005.

[2] A. Micheli, P. Baili, M. Quinn, E. Mugno, R. Capocaccia, P. Grosclaude, Life expectancy and Cancer Survival, The EUROCARE Working Group, in the EUROCARE-3 cancer registry areas, Ann. Oncol. V28–V40, 2003.

[3] G.K. Singh, B.A. Miller, B.F. Hankey, Changing area socioeconomic patterns in U.S. cancer mortality, 1950–1988. Part II: lung and colorectal cancers, J. Natl. Cancer Inst. 94 916–925, 2002.

[4] X. Ye, X. Lin, G. Beddoe, J. Dehmeshki, "Efficient computer-aided detection of ground-glass opacity nodules in thoracic CT images", in Proc.29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 4449–4452, 2007.

[5] T.W. Way, B. Sahiner, H.P. Chan, L. Hadjiisk, P.N .Cascade, A. Chughtai, N. Bogot, E. Kazerooni, "Computer-aided diagnosis of pulmonary nodules on CT scans: Improvement of classification performance with nodule surface features", Med. Phys., vol. 36, no. 7, pp. 3086–3098, 2009.

[6] S.C. Park, J. Tan, X. Wang, V. Lederman, J.K. Leader, S.H. Kim, B. Zheng, "Computer-aided detection of early interstitial lung diseases using low-dose CT images", Phys. Med. Biol., vol. 56, no. 4, pp. 1139–1153, 2011.

[7] L. Song, X. Liu, A. Yang, K. Pang, C. Zhou, X. Zhao, Y. Zhao, "A novel approach of computer-aided detection of focal ground-glass opacity in 2D lung CT images", Proc. SPIE, vol. 8670, pp. 86702W, 1–6, 2013.

[8] T. Manikandan, "Challenges in lung cancer detection using computer-aided diagnosis (CAD) systems – a key for survival of patients", Arch Gen Intern Med Volume 1 Issue 2, 2017.

[9] H.A. Elmikati, F.E. Abou-Chadi, Sh.S. Kishk, M.I. Obayya, H.M. Amer, "Automatic Detection of Human's Lung Nodules Using an Optimal Gray Level Threshold", International Journal of Scientific & Engineering Research Volume 8, Issue 5, ISSN 2229-5518, May-2017.

[10] Early Lung Cancer Action Program(ELCAP), available from: http://www.via.cornell.edu/lungdb.html.[Last cited on 2011 Dec 05].

[11] W.J. Kostis, A.P. Reeves, D.F. Yankelevitz, and C.I. Henschke, "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images," IEEE Transactions on Medical Imaging, vol. 22, no. 10, pp. 1259–1274, 2003.

[12] R. Shen, I. Cheng, A. Basu, "A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs", IEEE Trans. Biomed. Eng., vol. 57, no. 11, pp. 2646–56, Nov. 2010.

[13] R. Uppaluri, T. Mitsa, M. Sonka, E.A. Hoffman, G. McLennan, "Quantification of pulmonary emphysema from lung computed tomography images", Amer. J. Respir. Crit. Care Med., vol. 156, pp. 248–254, Jul.1997.

[14] F. Chabat, G.Z. Yang, D.M. Hansell, "Obstructive lung diseases: Texture classification for differentiation at CT", Radiology, vol. 228, pp. 871–877, Jul. 2003.

[15] P. Korfiatis, A. Karahaliou, C. Kalogeropoulou, A. Lazamtzo, L. Costaridou, "Texture based identification and characterization of interstitial pneumonia patterns in lung multi-detector CT", IEEE Trans. Inform. Technol. Biomed., vol. 14, no. 3, pp. 675–680, May. 2010.

[16] L. Sørensen, P. Lo, H. Ashraf, J. Sporring, M. Nielsen, M. Bruijne, "Learning COPD sensitive filters in pulmonary CT", in Proc. Med. Image Comput. Comput.-Assited Intervention Conf., vol. 5762, pp. 699–706, 2009.

[17] I. Sluimer, P.F. Waes, M.A. Viergever, B. Ginneken, "Computer-aided diagnosis in high resolution CT of the lungs", Med. Phys., vol. 30, no. 12, pp. 3081–3090, Dec. 2003.

[18] T. Nuzhnaya, V. Megalooi, H. Ling, M. Kohn, R. Steine, "Classification of texture patterns in CT lung imaging", Proc. SPIE, vol.7963, pp. 796336, 1–7, 2011.

[19] A. Depeursinge, J. Iavindrasana, A. Hidki, "A classification framework for lung tissue categorization", Proc. SPIE, vol. 6919, pp. 69190C,1–12, 2008.

[20] J. Shiraishi, F. Li, K. Doi, "Computer-aided diagnosis for improved detection of lung nodules by use of PA and lateral chest radiographs", Radiology, vol. 237, no. 2, pp. 657–661, 2005.

[21] M.B. Huber, M. Nagarajan, G. Leinsinger, L.A Ray, A. Wismuller, "Classification of interstitial lung disease patterns with topological texture feature", Proc. SPIE, vol. 7624, pp. 762410, 1–8, 2010.

[22] M.J. Gangeh, L. Sørensen, S. Shaker, M.S. Kamel, M.D. Bruijne, "Multiple classifier system in texton-based approach for the classification of CT images of lung", Med. Comput. Vis., vol. 6533, pp. 153–163, Feb.2011.

[23] M.R. Melamed, B.J. Flehinger, M.B. Zaman, R.T. Heelan, W.A. Perchick, N. Martini, "Screening for lung cancer: Results of thememorialsloan-kttering study in New York", Chest, vol. 86, no. 1, pp. 44–53, 1984.

[24] S.F. Huang. R.F. Chang, D.R. Chen, W.K. Moon, "Characterization of speculation on ultrasound lesions", IEEE Trans. Med. Imag., vol. 23, no. 1, pp. 111–121, Jan. 2004.

[25] M. Noguchi, Y. Shimosato, "The development and progression of adenocarcinoma of the lung", Cancer Treatment Res., vol. 72, pp. 131–142, 1995.

[26] X. Liu, L. Ma, L. Song, Y. Zhao, X. Zhao, C. Zhou, "Recognizing Common CT Imaging Signs of Lung Diseases Through a New Feature Selection Method Based on Fisher Criterion and Genetic Optimization", IEEE Trans. Biomedical and Health Informatics, vol. 19, no. 2, March 2015.

[27] Y. Saeys, I. Inza, P. Larra˜naga, "A review of feature selection techniques in bioinformatics", Bioinformatics, vol. 23, no. 19, pp. 2507–2517, 2007.

[28] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection", J. Mach. Learn. Res., vol. 3, pp. 1157–1182, 2003.

[29] M.A. Zuluaga, E.J. Leyton, M.H. Hoyos, M. Orkisz, "Feature selection for SVM-based vascular anomaly detection", in Proc. Int. MICCAI Conf. Med. Comput. Vis., Recognit. Tech. Appl. Med. Imag., pp. 141–152, 2010.

[30] S.C. Park, B.E. Chapman, B. Zheng, "A multistage approach to improve performance of computer-aided detection of pulmonary embolism depicted on CT images: preliminary investigation", IEEE Trans. Biomed. Eng., vol. 58, no. 6, pp. 1519–1527, Jun. 2011.

[31] I.J.S. Jeya, S.N. Deepa, "Lung Cancer Classification Employing Proposed Real Coded Genetic Algorithm Based Radial Basis Function Neural Network Classifier", Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine Volume 2016, Article ID 7493535, 2016.

[32] F.E. Abou-Chadi, M.I. Obayya, H.M. Amer " A Computer-Aided System for Classifying Computed Tomographic (CT) Lung Images Using Artificial Neural Network and Data Fusion", International Journal of Computer Science and Network Security, vol.11 No.10, October 2011.

[33] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", in Proc. Comput. Vis. Pattern Recognit., pp. 886–893, 2005.

[34] M. Firmino, G. Angelo, H. Morais, M.R. Dantas, R. Valentim, "Computer aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy", Firmino et al. Bio. Med. Eng. OnLine vol. 15, Issue 2, 2016.

[35] H. Hotelling, "Analysis of a complex of statistical variables into principal components", Journal of Educational Psychology, vol. 24, Issue 6, pp. 417-441, 2016.

[36] I. Buciu, A. Gacsadi, "Directional features for automatic tumor classification of mammogram images", Biomed Signal Process Control. Volume 6, Issue 4, pp. 370-378, October 2011.

[37] Y. Tao, L. Lu, M. Dewan, A.Y. Chen., J. Corso, J. Xuan, M. Salganicoff, A. Krishnan, "Multi-level ground glass nodule detection and segmentation in CT lung images", in Proc. Med. Image Computer-Assisted Surg., Part II, LNCS 5762, pp. 715–723, 2009.

[38] H. Wu, T. Sun, J. Wang, X. Li, W. Wang, D. Huo, et al., "Combination of radiological and gray level co-occurrence matrix textural features used to distin-

guish solitary pulmonary nodules by computed tomography", J. Digit Imaging, vol. 26, pp. 797–802, 2013.

[39] A. Kumar, S. Mukhopadhyay, N. Khandelwal, "3d texture analysis of solitary pulmonary nodules using co-occurrence matrix from volumetric lung CT images", In: Edited by Sidky E, Proceedings of the SPIE Medical Imaging Conference:; Lake Buena Vista, Florida, USA. Bellingham, WA: SPIE; vol. 8670. pp. 1–4, 28 February 2013.

[40] H.M Orozco, O.V Villegas, V.G. Sánchez, H.O. Domínguez, M.N. Alfaro, " Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine", Madero Orozco et al. Bio-Medical Engineering OnLine, vol. 14, no. 9, 2015.

[41] P.V. Ramaraju, S. Praveen," Classification of lung tumor Using Geometrical and Texture Features of Chest X-ray Images", International Journal for Research in Applied Science & Engineering Technology (IJRASET), Special Issue-3,ISSN: 2321-9653, November 2014.

[42] P. Arora, M. Singh," New Intensity Based Features for Classification of Mammograms", Conference Paper in Canadian Conference on Electrical and Computer Engineering, 978-1-4799-3010-9, 2014.

[43] Z. Guo, D. Mei," Feature Selection Using Principal Component Analysis", System Science, Engineering Design and Manufacturing Informatization (ICSEM), International Conference on vol. 1, Nov. 2010.

[44] J.H. Holland, "Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence", U Michigan Press; 1975.

[45] Sh. Aalaei, H. Shahraki, A. Rowhanimanesh, S. Eslami, "Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets", Iranian Journal of Basic Medical Sciences, vol. 19, no. 5, pp. 476–482, MAR 2016.

[46] S. Haykin, "Neural Networks: A Comprehensive Foundition," Prentice Education, Inc, 2002.

[47] R. Duda, P. Hat, "Patter Classification and Scene Analysis", New York, Wiley, 1973.

[48] M.V. Rao, N.V. Murty, "Early Lung Cancer Detection using Radial Basis Function Neural Networks (RBFNN)", International Journal of Modern Trends in Engineering and Research, issn(online): 2349-9745, vol. 2, no. 8, pp. 44-48, 2015.

[49] S. Sivakumar, C. Chandrasekar, "Lung Nodule Detection Using Fuzzy Clustering and Support Vector Machines", International Journal of Engineering and Technology (IJET), ISSN : 0975-4024, vol.5, no.1, Feb-Mar 2013.

[50] T. Ahmet, K. Niyazi, A. Aydin, " Classification of Pulmonary Nodules by Using Hybrid Features", Hindawi Publishing Corporation, Computational and Mathematical Methods in Medicine, vol. 2013, Article ID 148363, 2013.

[51] V. Kuppusamy, R.C. Gopalakrishnan, "Feature Extraction Based Lung Nodule Detection in CT Images", International Journal of Applied Engineering Research, ISSN 0973-4562, vol. 11, no. 4, pp 2697-2700, 2016.

[52] V.S. Thakare, N.N. Patil, "Classification of Texture Using Gray Level Co-Occurrence Matrix and Self-Organizing Map", International Conference on Electronic Systems, Signal Processing and Computing Technologies, IEEE, 978-1-4799-2102-7/14, 2014.

[53] B. V. Dasarathy, "Decision Fusion", Computer Society Press, 1994.

[54] L.I. Kuncheva, J.C. Bezdek, R.P.W. Duin, “Decision templates for multiple classifier fusion: an experimental comparison,” Pattern Recognition, vol. 34, pp. 299- 314, February 2001.

[55] U. G. Mangai, S. Samanta, S. Das, P.R. Chowdhury, " A Survey of Decision Fusion and Feature Fusion Strategies for Pattern Classification ", IETE technical review, vol 27, ISSUE 4,JUL-AUG, 2010.

[56] Z. Al-Ameen, G. Sulong., M. Gapar., M. Johar, "Enhancing the Contrast of CT Medical Images by Employing a Novel Image Size Dependent Normalization Technique", International Journal of Bio-Science and Bio-Technology vol. 4, no. 3, September, 2012.

[57] W.-J. Choi, T.-S. Choi," Automated Pulmonary Nodule Detection System in Computed Tomography Images: A Hierarchical Block Classification Approach", Entropy 2013, vol. 15, pp. 507-523, ISSN 1099-4300, 2013.

[58] J. Kuruvilla, K. Gunavathi, "Lung cancer classification using neural networks for CT images", computer methods and programs in biomedicine, vol. 113, pp. 202–209, 2014.

[59] O. Demira, A.Y. Çamurcub, "Computer-aided detection of lung nodules using outer surface features",Bio-Medical Materials and Engineering vol. 26, pp. S1213–S1222, 2015.

[60] T. Manikandan, N. Bharathi, " Lung Cancer Detection Using Fuzzy Auto-Seed Cluster Means Morphological Segmentation and SVM Classifier", SYSTEMS-LEVEL QUALITY IMPROVEMENT, J Med Syst., vol. 40, no. 181, 2016.

[61] J. D. Sweetlin, H. Kh. Nehemiah , A. Kannan, "Computer aided diagnosis of pulmonary hamartoma from CT scan images using ant colony optimization based feature selection", Alexandria University, Alexandria Engineering Journal, ISSN:111.-0168, 2017.