RESEARCH ARTICLE

Journal of Medical Imaging and Health Informatics
Vol. 6, 1–7, 2016

Robust Computer-Aided Detection of Pulmonary Nodules from Chest Computed Tomography

Zaid Abduh[1], Manal Abdel Wahed[1,*], and Yasser M. Kadah[2]

[1]*Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, Giza 12613, Egypt*
[2]*Electrical and Computer Engineering Department King Abdul Aziz University, Jeddah 21589, Saudi Arabia*

Detection of pulmonary nodules in chest computed tomography scans play an important role in the early diagnosis of lung cancer. A simple yet effective computer-aided detection system is developed to distinguish pulmonary nodules in chest CT scans. The proposed system includes feature extraction, normalization, selection and classification steps. One hundred forty-nine gray level statistical features are extracted from selected regions of interest. A min–max normalization method is used followed by sequential forward feature selection technique with logistic regression model used as criterion function that selected an optimal set of five features for classification. The classification step was done using nearest neighbor and support vector machine (SVM) classifiers with separate training and testing sets. Several measures to evaluate the system performance were used including the area under ROC curve (AUC), sensitivity, specificity, precision, accuracy, F1 score and Cohen-*k* factor. Excellent performance with high sensitivity and specificity is reported using data from two reference datasets as compared to previous work.

Keywords: Computer Aided Diagnosis, Computed Tomography, Pulmonary Nodules.

1. INTRODUCTION

Lung cancer remains the leading cause of cancer-related deaths in the world.[1] Early diagnosis can improve the effectiveness of treatment hence enhance the chances of survival.[2] The detection of pulmonary nodules in chest CT scans helps in early diagnosis of lung cancer and in risk assessment of malignancy. However, pulmonary nodule detection is time-consuming and oftentimes some nodules are missed due to the large number of axial thin-slices per scan (which usually range between 200–400 slices), and the small size of some pulmonary nodules.[3] Therefore, computer aided detection (CADe) has been being developed to aid the radiologist and serve as a second reviewer. The purpose of CADs is to distinguish pulmonary nodules automatically from other normal lung structures in high-resolution CT chest scans. This helps reduce the chances of missing nodules, improve performance, reduce intra- and inter-observer variability of radiologists, and quantitatively estimate the degree of malignancy risk.

Many studies were conducted to develop CADe systems. Ge et al. extracted forty-four features from volumes of interest, including geometric features, histogram-based features, gradient field and ellipsoid features.[4] Wilks' lambda stepwise feature selection feature selection and linear discriminant analysis (LDA) classification were used for nodules classification. Golosio et al. computed forty-three geometric features from selected region of interests (ROIs).[5] Iso-surface triangulation method was used to select regions of interest. Feed forward neural network with one hidden layer and logistic activation function was used for nodules classification. An Algorithm for automatic pulmonary nodule detection in chest CT was also developed by Murphy et al.[6] One hundred and thirty-five features calculated over segmented voxels and two KNN classifiers applied. Sequential-forward-floating-selection (SFFS) with leave-one-out cross-validation was used for feature selection. Messay et al. developed CADe system with a total of two hundreds forty-five features extracted from segmented nodule.[7] They use three feature types: geometric, intensity and gradient features. A sequential forward selection used to select the optimal subset for classification. Fisher linear discriminant (FLD) and quadratic classifiers were used. Another CADe system was developed by Tan et al.,[8] where nodule and vessel enhancement filters were used for nodule segmentation and divergence features were computed to locate the centers of the nodule clusters. Real nodules were differentiated from false positive ones using forty-five features (including invariant features calculated in a 3-D gauge coordinates, shape and regional descriptors). Feature selective classifier based on a genetic algorithm and artificial neural networks were developed for classification. Cao et al. developed another system with forty-three multiple type features based on intensity, shape and gradient to study the class imbalance issue, which occurs during the training of support vector machine (SVM) classifier.[9]

*Author to whom correspondence should be addressed.

They employed particle swarm optimization method (PSO) to improve the performance of classification by simultaneously optimizing the best pair of misclassification cost parameter, feature subset and intrinsic parameters in the objective function of cost sensitive SVM. Choi et al. developed an automated method to detect pulmonary nodules based on three dimensional shape-based feature descriptors, the surface saliency and surface normal vector are shape-based feature descriptors, which were obtained through eigenvalue decomposition of the Hessian matrix. Multi-scale dot enhancement filters were used to detect nodules in the segmented lung volume. Three-dimensional shape-based feature descriptors were then extracted from the detected nodules and an iterative wall elimination method was used to refine the feature descriptors. SVM classifier was used to classify nodules and non-nodules. Tater et al. also developed a new method for pulmonary nodule detection in CT scans using decision trees based on twelve geometric features. Their system used random forest (RF), logistic model trees (LMT) and J48 decision tree classifiers.

Even though the previous studies achieved excellent results using their respective methods, there is always a room for further enhancement of CADe performance that can be targeted. In this study, we develop a comprehensive CADe system for classification lung tissues in chest CT scans by using different combinations of 2-D features based on statistical texture features. Min–max normalization method (Rescaling) utilized. Sequential forward selection (SFS) with logistic regression model as evaluation function was used to select the significant feature set. Finally, *k*-nearest neighbor (KNN) and support vector machine (SVM) classifiers used for classification. The performance of the developed system was verified on LIDC reference database[12] with quantitative measures such as the area under curve (AUC) of the receiver operating characteristics curves (ROC), sensitivity, specificity, precision, accuracy, F1 score and geometric mean are utilized to assess the performance. Moreover, the agreement of the performance of the new system with the ground truth classification was assessed using the cohen-*k* factor. This paper presents the details of the system design as well as the results of its implementation.

2. METHODOLOGY

The components of the CADe system include feature extraction, feature normalization, feature selection and classification (nodule or non-nodule). A block diagram of the system is shown in Figure 1. Each stage will explained briefly in the following subsections.

2.1. Dataset

Several databases for research in lung cancer have been developed over the last decade. Some databases have been made publicly available, whereas others have remained privately owned by the research groups. The most commonly used databases are from the Cancer Imaging Archive.[12] The set that used in this work is the Lung Image Database Consortium (LIDC-IDRI). The LIDC-IDRI data used here is comprised of 112 CT scan images containing 112 nodules with diameter greater than or equal to 5 mm in effective size. The size of all images is 512×512 pixels in DICOM format. LIDC-IDRI database has its own annotation table.

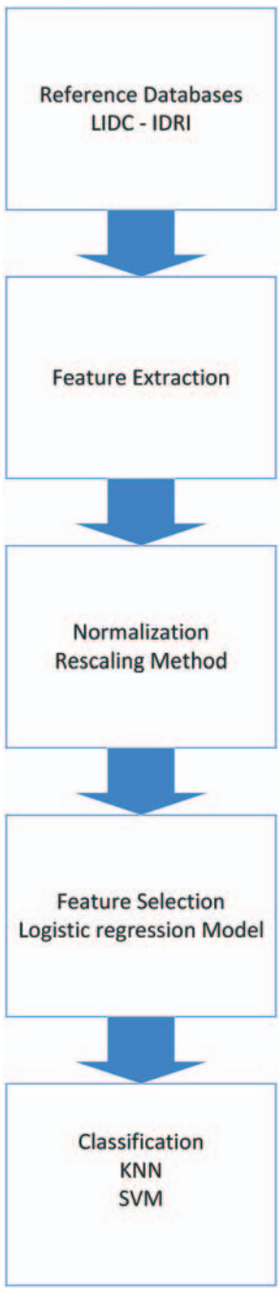

Fig. 1. Block diagram of the proposed system.

2.2. Feature Extraction

Feature extraction in pattern recognition is based on finding mathematical methods for reducing dimensionality of pattern representation.[13] Extraction procedure consists of three steps:

(1) Region of interest (ROI) selection where selection of nodule and non-nodule ROIs in LIDC-IDRI dataset is done manually according to the interpretation of the database and its annotation table.
(2) Extraction of the selected ROIs in 30×30 windows.
(3) Calculation of 149 features from the selected ROIs.

We chose CT-scans including nodules with different sizes and types to measure the sensitivity of the classifiers to most practical types of lung nodules. Total number of ROIs of LIDC-IDRI dataset is 224 (including 112 non-nodules and 112 nodules). Nodule diameters were greater than or equal to 5 mm in effective size and different types are selected including well-circumscribed nodule, juxta-pleural nodule, juxta-vascular nodule and nodule with pleural tail. Figure 2 shows samples of nodule and non-nodule patterns respectively. Features extracted include one hundred forty-nine gray level statistics features summarized as follows:

(1) First order statistical or histogram-based features, which depend only on pixel values, with seventeen features calculated from extracted ROIs (mean, standard deviation, third-order moment, smoothness, skewness, kurtosis, harmonic mean and variance, percentiles).[10, 14]

(2) Gray-level co-occurrence matrix (GLCM) high order statistical features, which depend on pixel values and their spatial interrelationships. The GLCM matrix is constructed in four direction $\delta = 0°, 45°, 90°$ and $135°$ with distance $d = 1$ pixel. Based on the computed GLCM, 88 statistical features were extracted including auto-correlation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, sum of squares variance, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation, information of correlation 1, information of correlation 2, inversed difference moment normalized, and inversed difference normalized.[15–18]

(3) Gray-level-run-length matrix (GLRLM) high order statistical features where this matrix is constructed then its properties are calculated as features. Features are calculated in four direction (0°, 45°, 90° and 135°) using zigzag method. A total of 44 features were computed including short run emphasis (SRE), long run emphasis (LRE), gray-level non uniformity (GLN), run-length non-uniformity (RLN), run percentage (RP), low gray-level run emphasis (LGRE), high gray-level run emphasis (HGRE), short run low gray-level emphasis (SRLGE), short run high gray-level emphasis (SRHGE), long run low gray-level emphasis (LRLGE), and long run high gray-level emphasis.[19]

The details of all features including their mathematical definitions are presented in Appendix I.

2.3. Normalization

Normalization of feature values is necessary preprocessing step before training classifiers. The main advantage of normalization (scaling) is to avoid dominating of the features with greater numeric ranges.[20] Another advantage is to avoid numerical difficulties during the calculation. Recommend linearly scaling each feature to the range $[-1, +1]$ or $[0, 1]$.[18] The simplest normalization method used is the min–max rescaling method. Equation (1) describes the min–max rescaling method in which each value in the feature vector subtracted from minimum value and divided by the difference between minimum and maximum values to make all values in range $[0, 1]$.

$$x' = (x - \min(x))/(\max(x) - \min(x)) \quad (1)$$

where x, an original value of a feature, x' is the normalized value. So each value in the features vector will be within range of $[0, 1]$.

2.4. Feature Selection

Feature selection is an important part of any classification scheme. Only a few features may be relevant and hence useful while many may contain irrelevant or redundant information that may result in degradation of classification performance. The success of a classification scheme largely depends on the selected features and the extent of their role in the model.[21] In this study, we utilize forward stepwise feature selection (SFS) algorithm to select the best features for classification by using logistic regression model as evaluation function (feature selection criterion). SFS is a wrapper approach that uses the performance of a classifier as its optimization criterion. It adds features sequentially from a candidate subset while evaluating the criterion, and seeks to minimize over all possible feature subsets. Common criteria used for this purpose include mean squared error (for regression models) and misclassification rate (for classification models). In logistic regression method, a logistic model is fitted to the data and deviation of this fit (mean square error) is computed. Such deviation of the model changes by adding of any other extra feature. The minimum value of this deviation indicates significant subset features and so it will be selected for inclusion. The algorithm continues adding features so long as the change in deviation is more than an amount that has a chi-square distribution with one degree of freedom.[21]

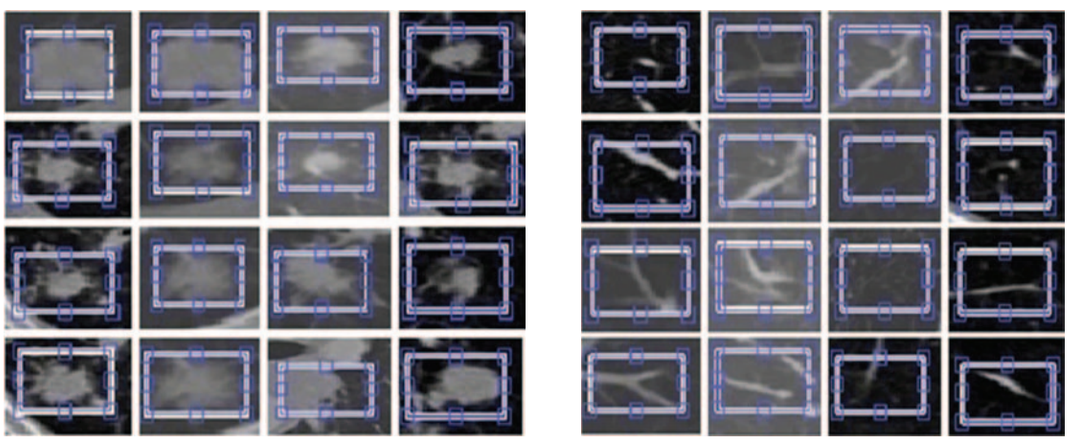

Fig. 2. Illustration of sample images of nodule (left) and non-nodule (right) patterns.

2.5. Feature Classification

To differentiate between nodule and non-nodule regions, a classifier is trained to discriminate between the classes in the feature space based only on the significant selected features. In this work, we used k-nearest neighbor (KNN) and the support vector machine (SVM) classifiers. Both can be used to perform a hard classification, where the output is a class label 0 label for non-nodule and 1 label for nodule. The KNN classifier is one of the simplest and widely used non-parametric machine learning algorithms. An object is classified based on computing the Euclidean distances from its k neighbors. The unknown test case is assigned to the closest cluster based on a majority vote among its labelled k neighbors. In this work, to minimize the errors due to the dimensionality of the feature space and sparsity of samples within that space, we use nearest neighbor classifier ($K = 1$). The SVM classifier is a supervised learning model. It has been applied to classification and regression problems with exceptionally good performance on a range of binary classification tasks. In SVM the original input space is mapped into a high dimensional dot product space called feature space. In this space, the optimal classification hyper-plane is determined to maximize the generalization ability of the classifier.[22] Table I presents the set of parameters associated with the used classifiers. Half-and-Half method used for training and testing of KNN and SVM classifiers.

2.6. Performance Evaluation

Given that we are considering a detection problem, we evaluate the performance of classifiers by considering the confusion matrix for a dichotomous problem where the testing results can be divided into four categories as shown in Table II with columns referring to the true (expected) values and rows referring to the predicted values. P stands for positive cases and N stands for negative cases. True positives (TP) are positive cases correctly classified as positive, true negatives (TN) are negative cases correctly classified as negative; false positives (FP) are negative cases incorrectly classified as positive and false negatives (FN) are positive cases incorrectly classified as negative. Using these distinct notions of correct and incorrect classification, we can utilize different measures to quantitatively evaluate the performance of a classifier as follows:[23]

• The *Sensitivity* measures the ability to identify the presence of the disease, expresses the ratio between the correctly predicted positive cases and the total number of the positive cases:

$$\text{Sensitivity} = TP/(TP+FN) \tag{2}$$

• The *Specificity* measures the ability to identify the absence of the disease, expresses the ratio between the correctly predicted Negative cases and the total number of the negative cases:

$$\text{Specificity} = TN/(TN+FP) \tag{3}$$

Table I. Parameters of the used classifiers.

Classifier	Parameters
KNN	$K = 1$, Euclidean distance metric, and nearest neighbor rule
SVM	Linear kernel function with sequential minimal optimization method (SMO)

Table II. Confusion matrix for a dichotomous problem.

	Expected	
	P	N
Predicted		
P	TP	FP
N	FN	TN

• The *Precision* measures the reliability of the positive result, expresses the ratio between the correctly predicted positive cases and the total number of cases predicted as positive:

$$\text{Precision} = TP/(TP+FP) \tag{4}$$

• The complementary of the Specificity (*1-Specificity*) is the ratio between the cases incorrectly predicted as positive and the total number of negative cases, i.e., it expresses the fraction of the incorrectly classified negative cases with respect to the total number of negative cases:

$$\text{1-Specificity} = 1 - TN/(TN+FP) = FP/(TN+FP) \tag{5}$$

• The *Accuracy* is the ratio between the number of correctly classified cases and the total number of cases given as:

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \tag{6}$$

• The F1 score (F_{Measure}) is a measure of a test's performance when a single value is wanted. It considers both the Precision and the Sensitivity of the test to compute the score. The traditional or balanced F-score is the harmonic mean of Precision and Sensitivity:

$$F_{\text{Measure}} = (2 \times \text{Precision} \times \text{Sensitivity}) / (\text{Precision} + \text{Sensitivity}) \tag{7}$$

• The Cohen-k factor which is a measure of the reproducibility, the methods for estimating intra- and inter variability for categorical variables. The Cohen-k is the ratio between the true agreement ($P_0 - P_a$) and the maximum achievable true agreement ($1 - P_a$). It is the fraction of the observed agreement on its maximum value not due to chance.

$$k = (P_0 - P_a)/(1 - P_a) \tag{8}$$

where

$$P_0 = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

and

$$P_a = \frac{(TP+FP)(TP+FN)+(FP+TN)(FN+TN)}{(TP+TN+FP+FN)^2} \tag{10}$$

The Cohen-k measures concordances and discordances in a dichotomous judgment between two observers, which are in our case the database annotation tables and the output result of the classifiers. If the Cohen-k factor values near one, it indicates near perfect agreement between observers. If it is near zero, it indicate the total absence of agreement between the observers.[23]

Table III. Confusion matrix entries obtained for both classifiers.

Entry	KNN	SVM
TP	56	56
TN	53	55
FP	3	1
FN	0	0

• The receiver operator characteristic (ROC) analysis graph is used where the area under the curve (AUC) measure is computed. The ROC comprises two performance evaluation measures; sensitivity and specificity where it represents a plot of the true positive rate (sensitivity) against the false positive rate (1-specificity) for the different possible cut points of a diagnostic test. It clearly demonstrates the tradeoff between sensitivity and specificity where any increase in sensitivity will be accompanied by a decrease in specificity. The area under the curve is widely used a measure of classification performance where an area close unity represents a perfect classification and an area near one half represents a poor classification.

3. RESULTS AND DISCUSSION

The total number of ROIs used from the LIDC-IDRI database was 224 (including 112 Non-nodules and 112 nodules). The training and testing of KNN and SVM classifiers was based on independent half-and-half training and testing sets with random selection of each from the database. Min–max normalization method was used to make each feature values range between 0 and 1. The outcome of the feature selection resulted in five optimal features that were skewness, 8th percentile, short run high gray-level emphasis (SRHGE) at direction 0°, run length non-uniformity (RLN) at direction 45°, and maximum probability at direction 0°. Table III summarizes the confusion matrix entries computed for both KNN and SVM classifiers. Table IV summaries the performance measures calculated from the results in Table III. Figure 3 shows the comparison between KNN and SVM in a graph. The ROC analysis of both classifiers with selected features shows that SVM was better than KNN where the AUC measure for SVM was 0.9868 whereas it was 0.9703 for KNN.

Table V shows the comparison of the proposed CAD system with the previously reported CAD systems. From the table, it is clear that the proposed CAD system achieves highest sensitivity and smallest false positives. These particular measures were used because of their importance to the selection of the best detection system and due to the availability of the information on such previous methods in the literature. We expect the conclusions of this comparison to hold for other measures given that both measures displayed affect all of them.

Feature space distributions, the dimensionality, and the available training sample sizes play an important role on feature selection method and classifiers relative performance.[24] Nodule size affects the performance of the classification process as well. Moreover, the large number of nodules that was used for training

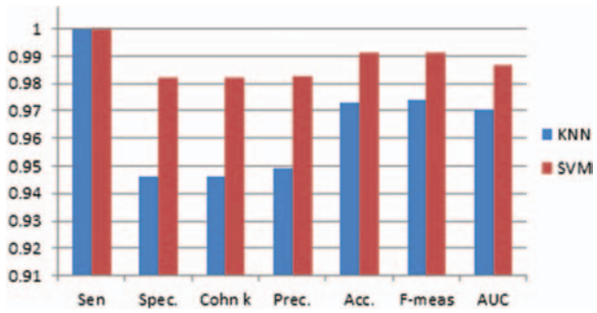

Fig. 3. Comparison of KNN and SVM performance measures.

enhances the performance of classification. Our datasets contained reasonable nodules number 112 nodules and nodules of diameter 15 mm were the dominant.

In using ROC analysis, it is very important to realize that the ROC curve is highly dependent on the dataset being used for evaluation and the nodule size. If the dominant nodule size 5 mm and above, very high sensitivity will be achieved, the curve of ROC will be closer to the left-hand border and then the top border of the ROC space and area under curve (AUC) will reach one. This is explaining why our ROC graphs reach the perfect. SVM classifier got better values than KNN classifier. Area under curve of ROC graph for SVM was 0.9868 whereas Area under curve of ROC graph for KNN was 0.9703.

Normalization and feature selection boost the performance of the classification process. Normalization will avoid numerical instabilities in the process of training the classifiers and allow variations in different features to be well represented equally with no dominating features that happen to have wider numeric ranges. Feature selection reduces the time of classification and will choose the optimal set that yield the best performance. Feature selection chooses optimal set of features. Number of the selected features depends on the datasets used. The training and testing method (half-and-half or cross validation) is chosen according to the available number of samples. If we have large number of samples, it is preferred to use half-and-half.

Since each measure indicates only one variable, we used more than one measure to compare the performance of classifiers in best way. Table IV summarized the different measures that used to evaluate the performance of the used classifiers. SVM classifier got the best values in all measures. This indicates that the feature space is somewhat sparse and therefore correct class assignment may not be possible by simple distance measures. SVM shows better generalization characteristics that play well into the results.

In this work, we considered 2D textural features to build the CAD system even though the CT data is available for the whole volume. This can be considered a limitation of the current system since the performance may vary with the location and orientation of the lesions with respect to the available slices. Therefore, we expect the performance to be enhanced if 3D features are considered and in particular those that are affine transformation invariant. The development and experimental validation of the

Table IV. Performance evaluation of KNN and SVM classifiers.

Classifier	Sensitivity	Specificity	Cohen-*k*	Precision	Accuracy	F-measure	AUC
KNN	1	0.9464	0.9464	0.9491	0.9732	0.9739	0.9703
SVM	1	0.9821	0.9821	0.9825	0.9911	0.9912	0.9868

Table V. Comparison of proposed method (shown in bold) to reported CAD systems.

CAD system	Nodule size (mm)	Sensitivity (%)	False positives (%)
Ge et al.[4]	3–30.6 mm	80	2.15
Golosio et al.[5]	3–30 mm	79	4
Murphy et al.[6]	≥4.5 mm	80	4.20
Messay et al.[7]	3–30 mm	82.66	3
Tan et al.[8]	3–30 mm	87.5	4
Choi et al.[10]	3–30 mm	97.5	6.76
Tater et al.[11]	2–20 mm	90.5	6
Proposed method	**≥5 mm**	**99.83**	**2**

utility of such 3D features are natural future extensions of the present work.

4. CONCLUSIONS

A simple yet effective system for lung nodule detection is developed. The system implementation of feature extraction, normalization, selection, and classification stages were described. The system was verified on real data from a reference dataset and compared to previous systems. The results obtained show a better performance in both the sensitivity and false positive rate. This indicates its potential for clinical use as a second reviewer to boost the diagnostic process.

APPENDIX I: STATISTICAL TEXTURAL FEATURE DEFINITIONS

A. First-Order Features

These features are computed from the pixel values (intensities) of the ROI and are independent of spatial distribution of pixels. That is, they can be completely defined in terms of the gray level histogram function $P(x_i)$ of the ROI. The definitions for these histogram-based features used are as follows:[10, 14]

Feature	Definition
Mean	$\mu = \sum_{i=1}^{\mathrm{MN}} x_i P(x_i)$
Standard deviation	$\sigma = \sqrt{\sum_{i=1}^{\mathrm{MN}} P(x_i)(x_i-\mu)^2}$
Third moment	$\sum_{i=1}^{\mathrm{MN}} (x_i-\mu)^3$
Smoothness	$1-\frac{1}{1+\sigma^2}$
Skewness	$\frac{\sum_{i=1}^{\mathrm{MN}} P(x_i)(x_i-\mu)^3}{(\sum_{i=1}^{\mathrm{MN}} P(x_i)(x_i-\mu)^2)^{1.5}}$
Kurtosis	$\frac{\sum_{i=1}^{\mathrm{MN}} P(x_i)(x_i-\mu)^4}{(\sum_{i=1}^{\mathrm{MN}} P(x_i)(x_i-\mu)^2)^2}$
Harmonic mean	$\frac{\mathrm{MN}}{\sum_{i=1}^{\mathrm{MN}} 1/x_i}$
Variance	σ^2
Percentiles	$P_n \ni \text{Cardinality}\{x_i : x_i \le P_n\} = \frac{n}{10}MN \quad n=1,2,\ldots,9$

B. Gray Level Co-Occurrence Matrix (GLCM) Features

These features are computed from the GLCM, which takes into account the pixel values as well as their spatial interrelationships and hence can better describe texture variations. The GLCM entry (i, j) is computed as count of pairs of pixels with intensities i and j and separated by a given fixed spatial vector. The resultant GLCM is of dimensions $N_g \times N_g$, where N_g is the number of gray levels in the ROI. The GLCM-based features are defined as follows:[15–18]

Feature	Definition
Energy	$\sum_i \sum_j \mathrm{GLCM}(i,j)^2$
Entropy	$-\sum_i \sum_j \mathrm{GLCM}(i,j)\log \mathrm{GLCM}(i,j)$
Contrast	$\sum_i \sum_j (i-j)^2 \mathrm{GLCM}(i,j)$
Correlation	$\frac{\sum_i \sum_j ij\,\mathrm{GLCM}(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}$
Autocorrelation	$\sum_i \sum_j ij\,\mathrm{GLCM}(i,j)$
Sum of squares	$\sum_i \sum_j (i-\mu)^2 \mathrm{GLCM}(i,j)$
Homogeneity	$\sum_i \sum_j \frac{\mathrm{GLCM}(i,j)}{1+(i-j)^2}$
Cluster prominence	$\sum_i \sum_j (i+j-\mu_x-\mu_y)^4 \mathrm{GLCM}(i,j)$
Cluster shade	$\sum_i \sum_j (i+j-\mu_x-\mu_y)^3 \mathrm{GLCM}(i,j)$
Dissimilarity	$\sum_i \sum_j \lvert i-j\rvert \mathrm{GLCM}(i,j)$
Maximum probability	$\mathrm{GLCM}(i,j)$
Sum average	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$
Sum entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i)\log p_{x+y}(i)$
Sum variance	$\sum_{i=2}^{2N_g} (i-\text{Sum Entropy})^2 p_{x+y}(i)$
Difference variance	$\text{Variance}\{p_{x-y}\}$
Difference entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i)\log p_{x-y}(i)$
Inverse difference moment normalized	$\sum_i \sum_j \frac{\mathrm{GLCM}(i,j)}{1+(i-j/N_g)^2}$
Inverse difference normalized	$\sum_i \sum_j \frac{\mathrm{GLCM}(i,j)}{1+(\lvert i-j\rvert/N_g)}$
Information of correlation 1	$\frac{H_{XY}-H_{XY1}}{\max(H_X,H_Y)}$
Information of correlation 2	$\sqrt{(1-\exp[-2(H_{XY2}-H_{XY})])}$

Notes: $p_x(i) = \sum_{j=1}^{N_g} \mathrm{GLCM}(i,j)$, $\mu_x = \sum_{i=1}^{N_g} i p_x(i)$, $\sigma_x = \sqrt{\sum_{i=1}^{N_g}(p_x(i)-\mu_x(i))^2}$. $p_y(j) = \sum_{i=1}^{N_g} \mathrm{GLCM}(i,j)$, $\mu_y = \sum_{i=0}^{G-1} j p_y(j)$, $\sigma_y = \sqrt{\sum_{i=1}^{N_g}(p_y(j)\mu_y(j))^2}$, $p_{x+y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \mathrm{GLCM}(i,j)$, $p_{x-y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \mathrm{GLCM}(i,j)$, $H_X = -\sum_i p_x(i)\log p_x(i)$, $H_y = \sum_j p_y(j)\log p_y(j)$, $H_{\mathrm{XY}} = -\sum_i \sum_j \mathrm{GLCM}(i,j)\log \mathrm{GLCM}(i,j)$,

$H_{XY1} = -\sum_i \sum_j \mathrm{GLCM}(i,j) \log_2 p_x(i) p_y(j)$, $H_{XY2} = -\sum_i \sum_j p_x(i) \times p_y(j) \log p_x(i) p_y(j)$.

C. Gray Level Runlength Matrix (GLRM) Features

GLRM is a way of describing the image in terms of lengths of consecutive pixels having the same gray level value in a given direction. GLRM element (i, j) represents the count of occurrences of a string of pixels of length j of value i in a specified direction within the ROI. The GLRM-based features are defined as follows:[19]

Feature	Definition
Short run emphasis	$\frac{1}{n_r}\sum_{j=1}^{N}\frac{p_r(j)}{j^2}$
Long run emphasis	$\frac{1}{n_r}\sum_{j=1}^{N} p_r(j)*j^2$
Gray-level nonuniformity	$\frac{1}{n_r}\sum_{i=1}^{M} p_g(i)^2$
Run length nonuniformity	$\frac{1}{n_r}\sum_{j=1}^{N} p_r(j)^2$
Run percentage	n_r/n_p
Low gray-level run emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\frac{p_g(i)}{i^2}$
High gray-level run emphasis	$\frac{1}{n_r}\sum_{i=1}^{M} p_g(i)*i^2$
Short run low gray-level emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)}{i^2*j^2}$
Short run high gray-level emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)*i^2}{j^2}$
Long run low gray-level emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)*j^2}{i^2}$
Long run high gray-level emphasis	$\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N} p(i,j)*j^2*i^2$

Notes: $p_r(j) = \sum_{i=1}^{M} \mathrm{GLRM}(i,j)$, $p_g(i) = \sum_{j=1}^{N} \mathrm{GLRM}(i,j)$, n_r is total number of runs, and n_p is the number of pixels in ROI.

References and Notes

1. American Cancer Society, Cancer Facts and Figures 2015. Available at: http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-44552.pdf.
2. A. El-Baz and J. Suri, Lung Imaging and Computer Aided Diagnosis, Taylor and Francis, London **(2011)**.
3. http://www.diagnijmegen.nl/index.php/Lung_Cancer.
4. Z. Ge, B. Sahiner, H.-P. Chan, L. M. Hadjiiski, P. N. Cascade, N. Bogot, E. A. Kazerooni, J. Wei, and C. Zhou, Computer-aided detection of lung nodules: false positive reduction using a 3D gradient field method and 3D ellipsoid fitting. *Med. Phys.* 32, 2443 **(2005)**.
5. B. Golosio, G. L. Masala, A. Piccioli, P. Oliva, M. Carpinelli, R. Cataldo, P. Cerello, F. De Carlo, F. Falaschi, M. E. Fantacci, G. Gargano, P. Kasae, and M. Torsello, A novel multi threshold method for nodule detection in lung CT. *Med. Phys.* 36, 3607 **(2009)**.
6. K. Murphy, B. van Ginneken, A. M. R. Schilham, B. J. de Hoop, H. A. Gietema, and M. Prokop, A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and *k*-nearest-neighbor classification. *Med. Imag. Anal.* 13, 757 **(2009)**.
7. T. Messay, R. C. Hardie, and S. K. Rogers, A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Med. Imag. Anal.* 14, 390 **(2010)**.
8. M. Tan, R. Deklerck, B. Jansen, M. Bister, and J. Cornelis, A novel computer-aided lung nodule detection system for CT images. *Med. Phys.* 38, 5630 **(2011)**.
9. P. Cao, D. Zhao, and O. Zaiane, Measure Oriented Cost-Sensitive SVM for 3D nodule detection, *Proceedings 35th Annual International Conference IEEE EMBS*, Osaka **(2013)**, pp. 3981–3984.
10. W.-J. Choi and T.-S. Choi, Automated pulmonary nodule detection based on three-dimensional shape-based feature descriptor. *Comput. Meth. Prog. Biomed.* 113, 37 **(2014)**.
11. A. Tater, N. Kilic, and A. Akan, A new method for pulmonary nodule detection in CT scans using decision tree, *Proceedings 35th Annual International Conference IEEE EMBS*, Osaka **(2013)**, pp. 7355–7359.
12. https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI.
13. A. Meyer-Base, Pattern Recognition for Medical Imaging, Elsevier Academic Press, Amsterdam **(2004)**.
14. V. Sudha and P. Jayashree, Lung nodule detection in CT images using thresholding and morphological operations. *Int. J. Emerg. Sci. Eng. (IJESE)* 1, 17 **(2012)**.
15. R. M. Haralick, K. Shanmugam, and I. Dinstein, Textural features of image classification. *IEEE Trans. Sys. Man. Cybern.* SMC-3, 610 **(1973)**.
16. L.-K. Soh and C. Tsatsoulis, Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Trans. Geosci. Remote Sens.* 37, 780 **(1999)**.
17. D. A. Clausi, An analysis of co-occurrence texture statistics as a function of grey level quantization. *Can. J. Remote Sens.* 28, 45 **(2002)**.
18. http://murphylab.web.cmu.edu/publications/boland/boland_node26.html.
19. M. Radhakrishnan and T. Kuttiannan, Comparative analysis of feature extraction methods for the classification of prostate cancer from TRUS medical images. *IJCSI Int. J. Comput. Sci. Issues* 9, 171 **(2012)**.
20. C.-W. Hsu, C.-C. Chang, and C.-J. Lin, A practical guide to support vector classification, Available at: https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf.
21. I. C. Sluimer, P. F. van Waes, M. A. Viergever, and R. van Ginneken, Computer-aided diagnosis in high resolution CT of the lungs. *Med. Phys.* 30, 3081 **(2003)**.
22. L. R. Sudha and R. Bhavani, Performance comparison of SVM and kNN in automatic classification of human gait patterns. *Int. J. Comput.* 6, 19 **(2012)**.
23. F. Sardanelli and G. Di Leo, Biostatistics for Radiologists Planning, Performing, and Writing a Radiologic Study, Springer, Berlin **(2008)**.
24. T. W. Way, B. Sahiner, L. M. Hadjiiski, and H.-P. Chan, Effect of finite sample size on feature selection and classification: A simulation study. *Med. Phys.* 37, 907 **(2010)**.

Received: 10 June 2015. Revised/Accepted: 2 November 2015.