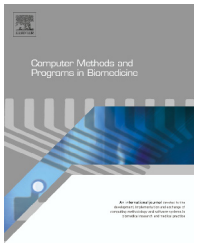
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 111 (2013) 519–524

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set

Tao Sun[a], *Jingjing Wang*[a], *Xia Li*[a,b], *Pingxin Lv*[c], *Fen Liu*[a,b], *Yanxia Luo*[a,b], *Qi Gao*[a,b], *Huiping Zhu*[a,b], *Xiuhua Guo*[a,b,*]

[a] *School of Public Health, Capital Medical University, Beijing 100069, China*
[b] *Beijing Municipal Key Laboratory of Clinical Epidemiology, Beijing 100069, China*
[c] *Department of Radiology, Beijing Chest Hospital, Capital Medical University, Beijing 101149, China*

ARTICLE INFO

Article history:
Received 30 December 2012
Received in revised form
24 April 2013
Accepted 24 April 2013

Keywords:
CT image
Curvelet
Solitary pulmonary nodule
Support vector machine
Texture extraction

ABSTRACT

Lung cancer is one of the most common forms of cancer resulting in over a million deaths per year worldwide. In this paper, the usage of support vector machine (SVM) classification for lung cancer is investigated, presenting a systematic quantitative evaluation against Boosting, Decision trees, *k*-nearest neighbor, LASSO regressions, neural networks and random forests. A large database of 5984 regions of interest (ROIs) and 488 input features (including textural features, patient characteristics, and morphological features) were used to train the classifiers and evaluate for their performance. The evaluation for classifiers' performance was based on a tenfold cross validation framework, receiver operating characteristic curve (ROC), and Matthews correlation coefficient. Area under curve (AUC) of SVM, Boosting, Decision trees, *k*-nearest neighbor, LASSO, neural networks, random forests were 0.94, 0.86, 0.73, 0.72, 0.91, 0.92, and 0.85, respectively. It was proved that SVM classification offered significantly increased classification performance compared to the reference methods. This scheme may be used as an auxiliary tool to differentiate between benign and malignant SPNs of CT images in future

1. Introduction

As the leading cause of cancer-related mortalities, lung cancer is responsible for approximately 1.38 million deaths annually worldwide. Despite recent advances in medicine and technology, the prognosis for lung cancer remains poor, with the 5-year survival rate approaching only 10% in most countries [1]. Yet, if the cancer can be detected and diagnosed in its early stages, the 10-year survival rate could be greatly promoted [2]. However, it is difficult to diagnose lung cancer efficiently. More than 80% of patients are diagnosed with locally advanced or metastatic disease.

Currently, the diagnosis of lung cancer primarily relies on digital computed tomography (CT). In CT images, lung cancer usually appears as solitary pulmonary nodules (SPNs).

* *Corresponding author at*: School of Public Health, Capital Medical University, Beijing 100069, China. Tel.: +86 1083911508; fax: +86 1083911508.
E-mail addresses: hhhtst@126.com (T. Sun), wjj00a@126.com (J. Wang), lixia_new@163.com (X. Li), lpx1209@163.com (P. Lv), liufen05@ccmu.edu.cn (F. Liu), lyx100@ccmu.edu.cn (Y. Luo), gaoqi@ccmu.edu.cn (Q. Gao), zhuhuiping79@163.com (H. Zhu), guoxiuh@ccmu.edu.cn (X. Guo).
0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2013.04.016

By definition, the solitary pulmonary nodule (SPN) is a single, spherical, well-circumscribed, radiographic opacity that measures ≤ 3 cm in diameter and is surrounded completely by aerated lung. There is no associated atelectasis, hilar enlargement, or pleural effusion. However, the solitary pulmonary nodules (SPNs) of lung cancer share similarities with several benign diseases, such as tuberculosis, inflammatory pseudotumor, hamartoma, and aspergillosis [3]. A meta-analysis [4] found that it has a pooled sensitivity of 0.57 (95% confidence interval, 0.49–0.66) and a pooled specificity, 0.82 (95% confidence interval, 0.77–0.86) for lung cancer using CT.

To improve the accuracy and efficiency of CT scans in the diagnosis of lung cancer, a number of research groups are focusing on developing computer-aided diagnoses (CADs) as auxiliary tools, including image segmentation and textural analysis. Murphy et al. [5] extracted textural features, and used *k*-nearest-neighbor classification to detect pulmonary nodule in chest CT. Wang et al. [6] used the gray level co-occurrence matrix and the multi-level model to predict pulmonary nodules. Lee et al. [7] used a two-step approach for feature selection (216 textural features) and classifier (linear discriminant analysis) for evaluation of pulmonary nodule. Sousa et al. [8] explored six stages to extract texture for automatic detection of lung nodules in CT images using support vector machines. Kim et al. [9] used 11 shape features and 13 textural features, with support vector machine and Bayesian classifiers, to improve performance of differentiating obstructive lung diseases, based on high-resolution computerized tomography (HRCT) images. Tan et al. [10] proposed to use textural features and a decision tree (C4.5)-adaboost classifier for classifying normal and tuberculosis in lung computed tomography. Some research groups [11,12] used the support vector machine classifier and other algorithms for classification bases on textural features, and found support vector machines outperformed other classifiers.

In previous studies, there is not a standard way to describe lung nodules (how to extract texture, and other features), the same to prediction models. However, how to describe lung nodules and establish prediction models has a direct effect on the performance of computer-aided diagnoses. In this study, support vector machines, a type of established methodologies used widely in various fields, and other six prediction models were established using a multi-dimensional set of textural features extracted by a Curvelet transformation from regions of interest (ROIs) in CT images, patient demographic characteristics, and morphological features. The aim of this study was to determine which prediction model was more suitable for CT texture analysis and that if this scheme could differentiate between benign and malignant lung cancer of CT images.

2. Materials and methods

2.1. Materials

2.1.1. Image collection

This study was performed with ethics approval (Ethics Committee of Xuanwu Hospital, Capital Medical University, Approval Document No. [2011] 01). This study is cross-sectional and the CT images were collected in 4 hospitals, in 2009–2011. The decision on patient inclusion and exclusion was based on the results of final diagnoses. Malignant cases were confirmed by either surgical removal or biopsy of the lesion, and benign cases by either pathological analysis or a 2 year follow-up. The morphological features were accessed by 8 radiologists, and conflicts in final interpretation of CT images were resolved by consensus discussion.

Table 1 – The description of data set.

	Number of cases	ROIs
Benign cases		
Tuberculosis	29	982
Inflammatory pseudotumor	22	626
Hamartoma	27	696
Pulmonary interstitial edema	1	189
Sclerosing hemangioma	9	93
Clear cell tumor	1	11
Chondroma	1	22
Benign cases diagnosed by 2 year follow-up	4	46
Malignant cases		
Glandular cancer	129	4077
Squamous carcinoma	28	455
Adenosquamous carcinoma	6	208
Malignant carcinoid tumor	2	37

CT scans were obtained using a 64-slice helical CT scanner (GE/Light Speed Ultra System CT99, USA) with a tube voltage of 120 kV and a current of 200 mA The reconstruction thickness and intervals for routine scanning were 0.625 mm The data were reconstructed with a 512 × 512 matrix. All of the SPNs in the CT images were segmented manually to obtain regions of interest (ROIs), and the textural features were extracted ROI by ROI. The reason why nodules were segmented manually was that radiologists could identify accurately the pulmonary nodules which are so difficult to detect by algorithms. The region growing [13] algorithm, a popular tool for image segmentation, was used to remove any background pixels.

A total of 7438 ROIs were acquired from 259 patients, with 2665 benign ROI from 94 patients (50 males, 44 females) and 4773 malignant ROIs from 165 patients (98 males, 67 females). The details are presented in Table 1.

A total of 5984 ROIs acquired from 202 patients were randomly selected from the 259 cases as training data, with 1914 benign ROIs from 65 patients (36 males, 29 females) and 4070 malignant ROIs from 136 patients (87 males, 50 females). 1454 ROIs acquired from the remaining 57 patients were used as test data, with 751 benign ROIs (29 benign cases: 14 males, 15 females) and 703 ROIs (28 malignant cases: 12 males, 16 females) of malignant tumors.

2.1.2. Multi-dimensional data set

Texture is a fundamental characteristic of the digital images, and Curvelet transformations were used to extract texture in this article. A Curvelet transformation, a type of second generation Wavelets, overcomes the weaknesses of traditional multi-scale representations [14–16] and yields good results [17,18]. Some studies [19,20] have shown that Curvelet transformations outperform Wavelets. Based on the Curvelet transformation, fourteen CT image textural

features of pulmonary nodules were extracted: Entropy, Mean, Correlation, Energy, Homogeneity, Standard Deviation, Maximum Probability, Inverse Difference Moment, Cluster Tendency, Inertia, Sum-Mean, Difference-Mean, Sum-Entropy, and Difference-Entropy. As a multi-scale transformation, Curvelets decomposed ROI images into 34 sub-bands. Thus, 476 textural features could be extracted from each ROI. Three demographic parameters (age, gender, and smoking habits) of each case were obtained from medical histories. Nine morphological features (including substantial changes, density of SPNs, the presence of spicules, caverns, vacuoles, lobulation, calcification and ground glass in SPNs, and area) were reported by eight experienced radiologists according to CT images.

2.2. Methods

A set of textural features, demographic parameters, and morphological features was used as input data. Support vector machines and other classifiers (including random forests, Boosting, Decision trees, *k*-nearest neighbor, LASSO regressions and neural networks) were used to establish prediction models, and the malignance rate was used as the variable to draw a ROC curve. The malignance rate was defined as:

$$\text{malignance rate} = \frac{\text{the number of malignant images of one case by prediction models}}{\text{the total number of images of one case}} \quad (1)$$

2.2.1. Support vector machines

As suggested by literature to date, support vector machines can be considered good algorithms [21–23] for classification in various research fields. The SVM algorithm developed by Vladimir Vapnik and Alexey Chervonenkis was adopted.

Supposed training data set is as follows: x

$$\{(x_i, y_i), \quad i = 1, 2, \ldots, l\} \quad (2)$$

where x_i is input data, $x_i \in R^N$; y_i is output data, $y_i \in R^N$; l is the number of samples.

The optimal classification surface:

$$w\phi(x) + b = 0, \quad b \in R \quad (3)$$

where w is the dimension of the feature space. $\phi(x)$ is the inner product function.

SVMs can change the issue to be solved from the optimal classification surface into Quadratic Programming. The mathematical term is as follows:

$$\min\left(\frac{1}{2}\|w\|^2 + c\sum_{i=1}^{l}\zeta_i\right) \quad (4)$$

$$s.t.\ y_i(wx_i + b) \geq 1 - \zeta_i \quad (5)$$

$$y_i \in \{-1, 1\}, \zeta_i \geq 0, \quad i = 1, 2, \ldots, l \quad (6)$$

where ζ_i is the slack variable; c is the penalty parameter.

According to the K–T condition, the dual objective function is as follows:

$$\max\left\{\sum_{i=1}^{l}a_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}a_i a_j y_i y_j k(x_i, x_j)\right\}, s.t\ 0 \leq a_i \leq c, \sum_{i=1}^{l}a_i y_i = 0 \quad (7)$$

where a_i is the support value. If $0 < a_i \leq c$, x_i is called support vector (SV). If $0 > a_i < c$, x_i is called normal support vector (NSV).

$$b = \frac{1}{N_{NSV}}\sum_{x_i \in NSV}\left\{\left[y_i - \sum_{x_j \in SV}a_j y_j k(x_i, x_j)\right]\right\} \quad (8)$$

where N_{NSV} is the number of NSV.

The solution to the classification is given by the decision function.

$$f(x) = sign\left\{\left[\sum_{i=1}^{l}\alpha_i y_i k(x_i, x_j)\right] + b\right\} \quad (9)$$

where α_j is the positive Lagrange multiplier, $k(x_i, x_j)$ is the function for convolution of the kernel of the decision function.

2.2.2. Other prediction models

Recently, for the characteristics of time saving and good classification performance, other algorithms, such as random forests (RF), have received increased attention, with reports [24,25] suggesting that random forests classifiers may outperform support vector machines. Other classifiers, such as Boosting, Decision trees, *k*-nearest neighbor, LASSO regressions and neural networks are also getting more and more attention, and have good performances for predicting cancers. Variation classifiers [26,27] also have received some attention. In this paper we just compare the support vector machines, Boosting, Decision trees, *k*-nearest neighbor, LASSO regressions and neural networks to find to an optimal classifier for analysis.

The parameters of classifiers play an important key on the classification performance and we chose the parameters as follows to enhance the performance based on previous studies: the radial basis function kernel as the kernel of the SVMs; 6 as the number of nearest neighbor; 2 as the number of units in the hidden layer; J48 as Decision trees.

The program for implementing the prediction models was performed by R software version 2.14.0. In this study, we chose the radial basis function kernel as the kernel of the SVMs.

Table 2 – The distribution of three demographic parameters between benign and malignant cases.

	Benign	Malignance	Statistic	P
Smoking habits				
N (missing)	94 (0)	165 (0)	2.23	0.1356
No (%)	58 (61.7)	86 (52.12)		
Yes (%)	36 (38.3)	79 (47.88)		
Sex				
N (missing)	94 (0)	165 (0)	0.9407	0.3321
Female (%)	44 (46.81)	67 (40.61)		
Male (%)	50 (53.19)	98 (59.39)		
Age				
N (missing)	94 (0)	165 (0)	44.19	<0.0001
Mean (std)	51.04 (13.32)	62.9 (11.3)		
Median (Q1, Q3)	50.5(42.60)	64(54.73)		

2.2.3. *Validation of the classification methods*

A tenfold cross validation framework for training data was used to evaluate the ability of the classifiers to differentiate benign from malignant nodules, incorporated into the area under curve (AUC) with establishing the receive operation characteristic (ROC) based on test data. Also, we calculate Matthews correlation coefficient (Mcc) [28] typically used in machine learning as a measure of the quality of binary (two-class) classifications expressed by

$$\mathrm{Mcc} = \frac{(\mathrm{Tp} \times \mathrm{Tn}) - (\mathrm{Fp} \times \mathrm{Fn})}{\sqrt{(\mathrm{Tp} + \mathrm{Fp}) \times (\mathrm{Tp} + \mathrm{Fn}) \times (\mathrm{Tn} + \mathrm{Fp}) \times (\mathrm{Tn} + \mathrm{Fn})}} \quad (10)$$

where Tp is the number of malignant nodules correctly classified as malignant and Fn is the number of benign nodules wrongly classified as malignant; Tn is the number of benign nodules correctly classified as benign and Fp is the number of malignant nodules wrongly classified as benign. Mcc more closely of +1 value represents a perfect prediction, 0 an average random prediction and −1 an inverse prediction.

3. Results

The descriptions of demographic parameters between benign and malignant cases are shown in Table 2. The descriptions of parameters used in the study are provided in Table 3.

These 476 textural features, together with 3 demographic parameters of the patient and 9 morphological measurements of ROIs were used as input data to establish prediction models. Accuracy based on cross-evaluation and Mcc of SVM, Boosting, Decision trees, *k*-nearest neighbor, LASSO regressions, neural networks and random forests prediction models are listed as follows (Table 4). SVM, Boosting, Decision trees, *k*-nearest neighbor, LASSO regressions, neural networks and random forests prediction models were successfully established using the 488 features as input data. The information about the cases was analyzed, and the malignance rate (Formula (1)) was used as the independent variable to draw ROC curves. The performances of these prediction models were evaluated and the results which include sensitivity and specificity are presented in Fig. 1 and Table 4.

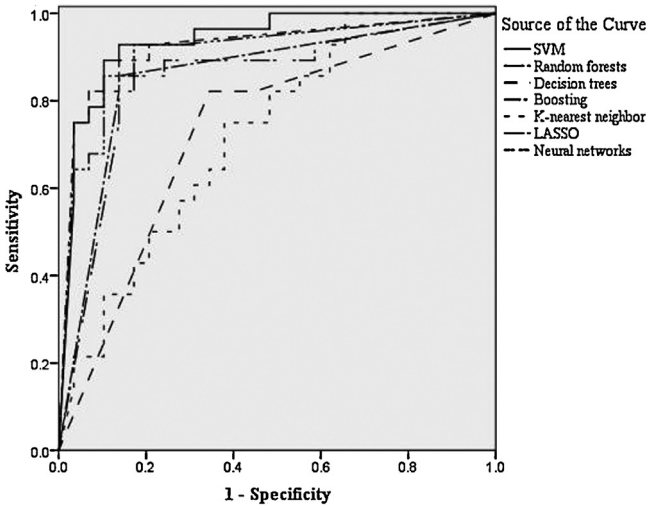


Fig. 1 – ROC curves created by classifiers.

4. Discussions

Prediction models have been used effective in other research fields, such as gene classification and comparison results have been reported in many fields. However, there is no agreement

Table 3 – Descriptions of patient characteristics and morphological features.

Variable	Description
Gender	1 male; 2 female
Age	Continuous
Smoking habits	1 yes; 0 no
Substantial changes	1 yes; 0 no
Uniform density of SPNs	1 yes; 0 no
Presence of spicules in SPNs	1 yes; 0 no
Caverns in SPNs	1 yes; 0 no
Vacuoles in SPNs	1 yes; 0 no
Lobulation in SPNs	1 yes; 0 no
Calcification in SPNs	1 yes; 0 no
Ground glass in SPNs	1 yes; 0 no
Area	Continuous

Table 4 – The performance of classifiers.

	Accuracy of cross-evaluation	AUC	Sensitivity	Specification	Mcc
SVM	0.93	0.94	0.93	0.86	0.79
Boosting	0.88	0.86	0.86	0.86	0.72
Decision trees	0.75	0.73	0.82	0.66	0.48
k-nearest neighbor	0.73	0.72	0.75	0.62	0.37
LASSO	0.90	0.91	0.93	0.83	0.75
Neural networks	0.89	0.92	0.89	0.86	0.75
Random forests	0.90	0.85	0.86	0.83	0.68

on which model has a better performance in the analysis of SPNs in CT images, and different research paper chose different classifiers [5–8,29,30]. In the present study, we have observed that SVMs outperform other classifiers, suggesting that SVMs may be a more appropriate prediction model in the analysis of SPNs in CT images.

Textural features of digital images reflect the microcosmic structure of the pictured objects, overlooking the macroscopic characteristics of the cases. In this study, the textural features extracted by Curvelet transformations, combined with 3 patient demographic characteristics and 9 morphological features, were used as input variables to establish prediction models. Compared to traditional models (such as Mayo Clinic model and VA model), our results appear to increase the rate and accuracy of lung cancer diagnoses, and it proves that using textural features combined with demographic characteristics and morphological features is a better scheme to predict lung cancer than only using textural features or morphological features.

In recent years, lung cancer has emerged as one of the most prevalent cancer in many countries. Due to the difficulties in its early diagnosis, lung cancer has a poor prognosis. The use of CT scans is common in clinical practices to distinguish between benign SPNs and malignant tumors. However, it has poor sensitivity and specificity for lung cancer using CT. As an auxiliary tool, texture extraction method is widely used in medical fields. Compared to the prior studies [5–8,29,30], our results appear to increase the sensitivity for CT scans to accurately predict SPN malignancy thus demonstrating that our methodology is suitable for CT texture analysis and the scheme has potential to be as an auxiliary tool in the future.

There are some limitations involved in this study. In order to ensure enough data for analysis, the images from the same patient were used as training data.

Author contributions

XG conceptualized and designed the study. Data analysis was done by TS and XG. They are also the guarantors of integrity of the entire study. Clinical study was done by PL. The manuscript was prepared by all authors.

Acknowledgements

Supported by the Natural Science Fund of China (Serial Number: 81172772); the Natural Science Fund of Beijing (Serial Number: 4112015, 7131002); and National S&T Major Project (Serial Number: 2012ZX10005009-003)

REFERENCES

[1] G. Mountzios, M.A. Dimopoulos, J.C. Soria, D. Sanoudoud, C.A. Papadimitriou, Histopathologic and genetic alterations as predictors of response to treatment and survival in lung cancer: a review of published data, Critical Reviews in Oncology/Hematology 75 (2010) 94–109.

[2] N. Seki, K. Eguchi, K. Kaneko, H. Ohmatsu, R. Kakinuma, E. Matsui, M. Kusumoto, T. Tsuchida, H. Nishiyama, N. Moriyama, The adenocarcinoma-specific stage shift in the Anti-lung Cancer Association project: significance of repeated screening for lung cancer for more than 5 years with low-dose helical computed tomography in a high-risk cohort, Lung Cancer 67 (2010) 318–324.

[3] J.W. Chang, C.A. Yi, D.S. Son, N. Choi, J. Lee, H.K. Kim, Y.S. Choi, K.S. Lee, J. Kim, Prediction of lymph node metastasis using the combined criteria of helical CT and mRNA expression profiling for non-small cell lung cancer, Lung Cancer 60 (2008) 264–270.

[4] M.T. Eric, H. Linda, C.M. Douglas, Noninvasive staging of non-small cell lung cancer: a review of the current evidence, Chest 123 (2003) 137S–146S.

[5] K. Murphy, B. Ginneken, A.M.R. Schilham, K. Murphy, B. Ginneken, A.M.R. Schilham, B.J. Hoop, H.A. Gietema, M. Prokop, A large-scale evaluation of automatic pulmonary nodule detection in chest CT using local image features and *k*-nearest-neighbour classification, Medical Image Analysis 13 (2009) 757–770.

[6] H. Wang, X.H. Guo, Z.W. Jia, H.K. Li, Z.G. Liang, K.C. Li, Q. He, Multilevel binomial logistic prediction model for malignant pulmonary nodules based on texture features of CT image, European Journal of Radiology 74 (2010) 124–129.

[7] M.C. Lee, L. Boroczky, K. Sungur-Stasik, A.D. Cann, A.C. Borczuk, S.M. Kawut, C.A. Powell, Computer-aided diagnosis of pulmonary nodules using a two-step approach for feature selection and classifier ensemble construction, Artificial Intelligence in Medicine 1 (2010) 43–53.

[8] J.R. Sousa, A.C. Silva, A.C. Paiva, R.A. Nunes, Methodology for automatic detection of lung nodules in computerized tomography images, Computer Methods and Programs in Biomedicine 98 (2010) 1–14.

[9] N. Kim, J.B. Seo, Y. Lee, J. Goo Lee, S.S. Kim, S.H. Kang, Development of an automatic classification system for differentiation of obstructive lung disease using HRCT, Journal of Digital Imaging 22 (2009) 136–148.

[10] J.H. Tan, U.R. Acharya, C. Tan, K.T. Abraham, C.M. Lim, Computer-assisted diagnosis of tuberculosis: a first order statistical approach to chest radiograph, Journal of Medical Systems 36 (2012) 2751–2759.

[11] M.B. Huber, M.B. Nagarajan, G. Leinsinger, R. Eibel, L.A. Ray, A. Wismüller, Performance of topological texture features to classify fibrotic interstitial lung disease patterns, Medical Physics 38 (2011) 2035–2044.

[12] J. Juntu, J. Sijbers, S. Backer, J. Rajan, D.D. Van, Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images, Journal of Magnetic Resonance Imaging 31 (2010) 680–689.
[13] D. Stroppiana, G. Bordogna, P. Carrara, M. Boschetti, L. Boschetti, P.A. Brivio, A method for extracting burned areas from Landsat TM/ETM+ images by soft aggregation of multiple Spectral Indices and a region growing algorithm, ISPRS Journal of Photogrammetry and Remote Sensing 69 (2012) 88–102.
[14] T. Mandal, Q.M.J. Wu, Y. Yuan, Curvelet based face recognition via dimension reduction, Signal Processing 89 (2009) 2345–2353.
[15] Y.C. Li, Q. Yang, R.H. Jiao, Image compression scheme based on curvelet transform and support vector machine, Expert Systems with Applications 37 (2010) 3063–3069.
[16] B.P. Li, M.Q.H. Meng, Texture analysis for ulcer detection in capsule endoscopy images, Image and Vision Computing 27 (2009) 1336–1342.
[17] K. Vishal, O. Jounada, M.C. Ron, V. Kumar, J. Oueity, R.M. Clowes, F. Herrmann, Enhancing crustal reflection data through curvelet denoising, Tectonophysics 508 (2011) 106–116.
[18] M.M. Eltoukhy, I. Faye, B.B. Samir, Breast cancer diagnosis in digital mammogram using multiscale curvelet transform, Computerized Medical Imaging and Graphics 34 (2010) 269–327.
[19] L. Dettori, L. Semler, A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography, Computers in Biology and Medicine 37 (2007) 486–498.
[20] E.J. Candes, L. Demanet, D.L. Donoho, L.X. Ying, Fast discrete curvelet transforms, Multiscale Modelling and Simulation 5 (2006) 861–899.
[21] Y. Lee, J.B. Seo, J.G. Lee, S.S. Kim, N. Kim, S.H. Kang, Performance testing of several classifiers for differentiating obstructive lung diseases based on texture analysis at high-resolution computerized tomography (HRCT), Computer Methods and Programs in Biomedicine 93 (2009) 206–215.
[22] E.M. Michael, V.G. Harris, D. Nikos, C. Dionisis, T. Sergios, Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers, Artificial Intelligence in Medicine 37 (2006) 145–162.
[23] E.J.R Justino, F. Bortolozzi, R. Sabourin, A comparison of SVM and HMM classifiers in the off-line signature verification, Pattern Recognition Letters 26 (2005) 1377–1385.
[24] D.U. Ramón, A.A. Sara, Gene selection and classification of microarray data using random forest, BMC Bioinformatics 7 (2006) 3.
[25] C.H. Hsieh, R.H. Lu, N.H. Lee, W.T. Chiu, M.H. Hsu, Y.H. Li, Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks, Surgery 149 (2011) 87–93.
[26] M. Liu, L. Lu, J. Bi, V. Raykar, M. Wolf, M. Salganicoff, Robust large scale prone-supine polyp matching using local features: a metric learning approach, Medical Image Computing and Computer-Assisted Intervention 14 (2011) 75–82.
[27] M. Liu, L. Lu, X. Ye, S. Yu, M. Salganicoff, Sparse classification for computer aided diagnosis using learned dictionaries, Medical Image Computing and Computer-Assisted Intervention 14 (2011) 41–48.
[28] B. Matthews, Comparison of the predicted and observed secondary structure of T4 phage lysozyme, Biochimica et Biophysica Acta 405 (1975) 442.
[29] H. Saito, Y. Minamiya, H. Kawai, T. Nakagawa, M. Ito, Y. Hosono, S. Motoyama, M. Hashimoto, K. Ishiyama, J. Ogawa, Usefulness of circumference difference for estimating the likelihood of malignancy in small solitary pulmonary nodules on CT, Lung Cancer 58 (2007) 348–354.
[30] H. Wu, T. Sun, J. Wang, X. Li, W. Wang, D. Huo, P. Lv, W. He, K. Wang, X. Guo, Combination of radiological and gray level co-occurrence matrix textural features used to distinguish solitary pulmonary nodules by computed tomography, Journal of Digital Imaging (2013), http://dx.doi.org/10.1007/s10278-012-9547-6.