Classification of pulmonary nodules in lung CT images using shape and texture features

Dhara, Ashis Kumar, Mukhopadhyay, Sudipta, Dutta, Anirvan, Garg, Mandeep, Khandelwal, Niranjan, et al.

Ashis Kumar Dhara, Sudipta Mukhopadhyay, Anirvan Dutta, Mandeep Garg, Niranjan Khandelwal, Prafulla Kumar, "Classification of pulmonary nodules in lung CT images using shape and texture features," Proc. SPIE 9785, Medical Imaging 2016: Computer-Aided Diagnosis, 97852Y (24 March 2016); doi: 10.1117/12.2214466

Event: SPIE Medical Imaging, 2016, San Diego, California, United States

Classification of Pulmonary Nodules in Lung CT Images using Shape and Texture Features

Ashis Kumar Dhara[1], Sudipta Mukhopadhyay[1], Anirvan Dutta [2],
Mandeep Garg[3], Niranjan Khandelwal[3], and Prafulla Kumar[4]
[1]Electronics & Electrical Communication Engineering, Indian Institute of Technology Kharagpur, India 721302.
[2]Electronics & Communication Engineering, Birla Institute of Technology Mesra, India 835215
[3]Radio Diagnosis & Imaging, Postgraduate Institute of Medical Education and Research, India 160023
[4]Department of Electronics and Information Technology, India 110003

ABSTRACT

Differentiation of malignant and benign pulmonary nodules is important for prognosis of lung cancer. In this paper, benign and malignant nodules are classified using support vector machine. Several shape-based and texture-based features are used to represent the pulmonary nodules in the feature space. A semi-automated technique is used for nodule segmentation. Relevant features are selected for efficient representation of nodules in the feature space. The proposed scheme and the competing technique are evaluated on a data set of 542 nodules of Lung Image Database Consortium and Image Database Resource Initiative. The nodules with composite rank of malignancy "1","2" are considered as benign and "4","5" are considered as malignant. Area under the receiver operating characteristics curve is 0.9465 for the proposed method. The proposed method outperforms the competing technique.

Keywords: Lung cancer, Pulmonary nodules, Segmentation of nodules, Feature extraction, Classification of benign and malignant nodules.

1. INTRODUCTION

Lung cancer causes more deaths than the next three common cancer such as breast, colon and prostate.[1] Lung cancer is frequent for smokers as well as non-smokers, and its causes remain unknown. Early diagnosis of lung cancer could improve five-years survival rate from 15% to 80% .[2] Screening program can substantially reduce mortality rate caused by lung cancer.[2] Pulmonary nodules are an early indication of lung cancer.[3] Diameter of pulmonary nodules lie in the range of 3mm to 30mm. A classification scheme is required to assist the radiologists during diagnosis of lung cancer.

Several classification works have been reported using linear discriminant analysis (LDA),[4–6] artificial neural network (ANN),[7,8] and support vector machine (SVM).[9,10] Different shape features[11,12] and texture features[10] have been used for classification of nodules. The combinations of shape and texture features[5,13] are also studied. The present work focuses on the classification

Medical Imaging 2016: Computer-Aided Diagnosis, edited by Georgia D. Tourassi, Samuel G. Armato III,
Proc. of SPIE Vol. 9785, 97852Y · © 2016 SPIE · CCC code: 1605-7422/16/$18 · doi: 10.1117/12.2214466

of benign and malignant nodules using a combination of 3D shape and texture features. The proposed method is compared with the most recent classification work of Han *et al.*,[10] which depends on radiologists for nodule segmentation.

2. CLASSIFICATION OF BENIGN AND MALIGNANT NODULES

End user provides a seed point on the nodule to select a volume of interest (VOI). The technique of Dhara *et al.*[14] is used for nodule segmentation. Block diagram of the proposed scheme is shown in Fig. 1.

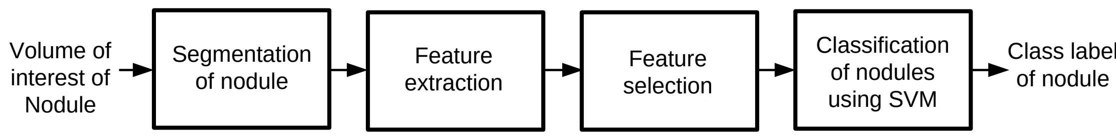

Figure 1: Block diagram of the proposed classification scheme

2.1 Segmentation of Nodules

The technique of Dhara *et al.*[14] is able to segment all types of nodules (viz. solid, part-solid and non-solid) based on their internal texture. In this framework, two separate algorithms were developed for segmentation of solid/part-solid and non-solid nodules, respectively.

2.2 Feature Extraction

Surface roughness and texture are important for discrimination of benign and malignant nodules. The nodules are represented using several 3D shape- and texture-based features. Shape features are volume, surface area, sphericity, equivalent diameter, circularity, convex surface area, major axis length, and minor axis length. Nine gray level co-occurrence matrix (GLCM) were computed considering nine directions[10] to compute Haralick features.[15] The mean and range of each Haralick feature over the nine directions are computed as[10]

$$H^{k}{}_{mean} = \frac{1}{n}\sum_{i=1}^{9} H_{ik}, \tag{1}$$

$$H^{k}{}_{range} = \max_{i=1:9}\{H_{ik}\} - \min_{i=1:9}\{H_{ik}\}, \tag{2}$$

where H_{ik} is the k^{th} Haralick feature considering i^{th} direction. Total number of features computed from the segmented nodule is 34.

2.3 Feature Selection

Relevant features are selected based on the area under the receiver operating characteristics curve (AUC) and p-values. ROCKIT[16] is used for computation AUC and Student's **t**-test is used for **p**-values. AUC is ≥ 0.60 and p-value is ≤ 0.60 for 27 features as shown in Table 1. Hence, the set of 27 features are used in the proposed classification scheme.

Table 1: List of selected features based on **AUC** and **p**-value

Type	**Feature no.**	**Feature Name**	**AUC**	**p values**
3D shape	1	Equivalent diameter	0.88	3.83E-56
	2	Minor axis length	0.88	1.71E-55
	3	Volume of nodule	0.88	7.73E-28
	4	Surface area of nodule	0.87	9.62E-29
	5	Circularity	0.87	2.66E-16
	6	Convex surface area	0.85	1.11E-26
	7	Major axis length	0.81	3.06E-22
	8	Sphericity	0.67	2.90E-07
Haralick 3D	9	Mean information measure of Correlation[1]	0.92	0.2.6E-73
	10	Range of sum entropy	0.92	1.16E-95
	11	Mean inverse difference moment	0.88	7.46E-40
	12	Mean angular second moment	0.87	4.16E-37
	13	Range of angular second moment	0.87	4.99E-38
	14	Mean entropy	0.87	4.61E-57
	15	Range of sum average	0.80	2.42E-31
	16	Range of contrast	0.79	1.65E-24
	17	Mean contrast	0.79	1.21E-24
	18	Mean sum entropy	0.76	1.43E-23
	19	Range of difference entropy	0.74	9.46E-18
	20	Range of difference variance	0.71	5.56E-17
	21	Range of sum squares of variance	0.70	7.11E-14
	22	Range of sum variance	0.68	3.67E-15
	23	Range of inverse difference moment	0.67	5.07E-08
	24	Mean sum variance	0.64	4.57E-19
	25	Mean difference variance	0.64	4.33E-19
	26	Range of entropy	0.60	1.5E-03
	27	Range of mean information measure of correlation[1]	0.60	1.25-4

3. RESULTS AND DISCUSSION

The proposed classification method is compared with the method of Han *et al.*[10] The proposed method and the competing technique are evaluated on a data set of 542 nodules of LIDC/IDRI database.[17] The composite rank of malignancy 1, 2 are considered as benign and 4, 5 are considered as malignant.[10] The data set consists of 279 benign nodules and 263 malignant nodules. The SVM is used for classification of nodules. Five-fold cross-validation approach is used to generate training and testing data sets. Mean AUC is obtained from 100 classification outcomes. Mean AUC for the proposed method and the competing technique are 0.9465 and 0.9449, respectively. The ROC plots are shown in Figure 2. The proposed classification scheme outperforms the competing technique due to the use of 3D shape-based and texture-based features.

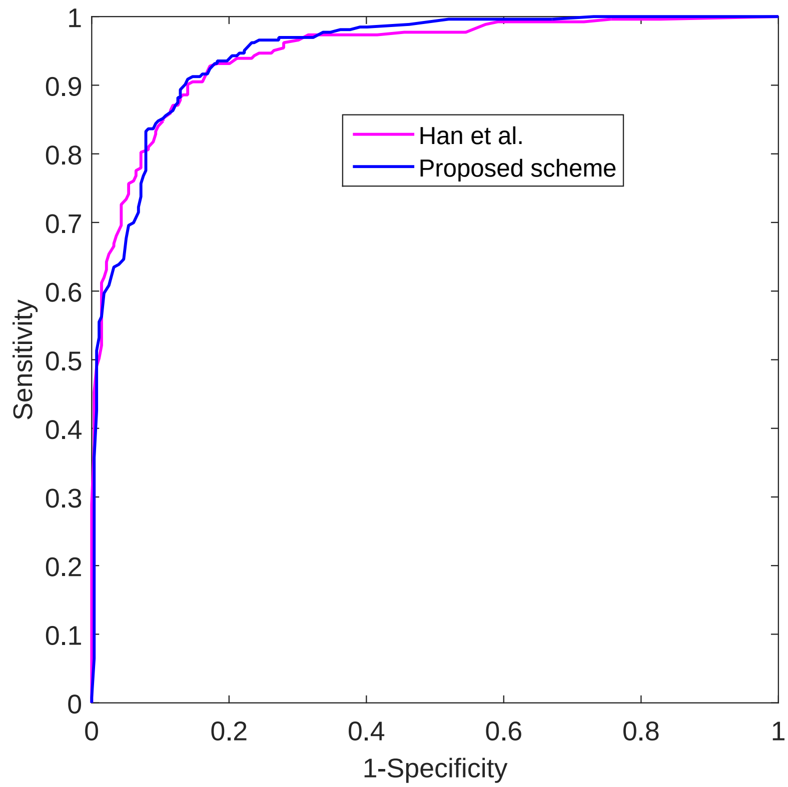

Figure 2: ROC plot of the performance of classification of benign and malignant nodules.

4. CONCLUSION

A classification scheme is developed for discrimination of benign and malignant nodules. The proposed scheme could assists the radiologists in screening of lung cancer. The proposed system is easy to use as only a seed point is expected from radiologists for nodule segmentation. More research is required to improve the accuracy of classification by means of improving segmentation accuracy and introducing new features.

Acknowledgements

This study was funded by Department of Electronics and Information Technology, Govt. of India with Grant number $1(2)/2013 - ME\&TMD/ESDA$.

REFERENCES

[1] *American Cancer Society, Global cancer facts and figure, 2^{nd} Edition, Atlanta, GA* (2012).

[2] Diederich, S., Wormanns, D., Semik, M., Thomas, M., Lenzen, H., Roos, N., and Heindel, W., "Screening for early lung cancer with low-dose spiral CT: Prevalence in 817 asymptomatic smokers," *Radiology* **222**(3), 773–781 (2002).

[3] Austin, J., Müller, N., Friedman, P. J., Hansell, D. M., Naidich, D. P., Remy-Jardin, M., Webb, W. R., and Zerhouni, E. A., "Glossary of terms for CT of the lungs: recommendations of the nomenclature committee of the fleischner society," *Radiology* **200**(2), 327–331 (1996).

[4] McNitt-Gray, M. F., Hart, E. M., Wyckoff, N., Sayre, J. W., Goldin, J. G., and Aberle, D. R., "A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution CT: preliminary results," *Medical Physics* **26**(6), 880–888 (1999).

[5] Armato III, S. G., Altman, M. B., Wilkie, J., Sone, S., Li, F., Doi, K., and Roy, A. S., "Automated lung nodule classification following automated nodule detection on CT: A serial approach," *Medical Physics* **30**(6), 1188–1197 (2003).

[6] Li, F., Aoyama, M., Shiraishi, J., Abe, H., Li, Q., Suzuki, K., Engelmann, R., Sone, S., MacMahon, H., and Doi, K., "Radiologists' performance for differentiating benign from malignant lung nodules on high-resolution CT using computer-estimated likelihood of malignancy," *American Journal of Roentgenology* **183**(5), 1209–1215 (2004).

[7] Matsuki, Y., Nakamura, K., Watanabe, H., Aoki, T., Nakata, H., Katsuragawa, S., and Doi, K., "Usefulness of an artificial neural network for differentiating benign from malignant pulmonary nodules on high-resolution CT: evaluation with receiver operating characteristic analysis," *American Journal of Roentgenology* **178**(3), 657–663 (2002).

[8] Suzuki, K., Li, F., Sone, S., and Kunio, D., "Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network," *IEEE Transactions on Medical Imaging* **24**(9), 1138–1150 (2005).

[9] Zhu, Y., Tan, Y., Hua, Y., Wang, M., Zhang, G., and Zhang, J., "Feature selection and performance evaluation of support vector machine (svm)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography," *Journal of digital imaging* **23**(1), 51–65 (2010).

[10] Han, F., Wang, H., Zhang, G., Han, H., Song, B., Li, L., Moore, W., Lu, H., Zhao, H., and Liang, Z., "Texture feature analysis for computer-aided diagnosis on pulmonary nodules," *Journal of Digital Imaging* **28**(1), 99–115 (2014).

[11] Iwano, S., Nakamura, T., Kamioka, Y., and Ishigaki, T., "Computer-aided diagnosis: a shape classification of pulmonary nodules imaged by high-resolution CT," *Computerized Medical Imaging and Graphics* **29**(7), 565–570 (2005).

[12] El-Baz, A., Nitzken, M., Vanbogaert, E., Gimel'farb, G., Falk, R., and El-Ghar, M. A., "A novel shape-based diagnostic approach for early diagnosis of lung nodules," in [*Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro*], 137–140 (March 2011).

[13] Elizabeth, D., Nehemiah, H., Raj, C. R., and Kannan, A., "Computer-aided diagnosis of lung cancer based on analysis of the significant slice of chest computed tomography image," *IET Image Processing* **6**(6), 697–705 (2012).

[14] Dhara, A. K., Mukhopadhyay, S., Das Gupta, R., Garg, M., and Khandelwal, N., "A segmentation framework of pulmonary nodules in lung CT images," *Journal of Digital Imaging* **29**(1), 148–148 (2015).

[15] Haralick, R. M., Shanmugam, K., and Dinstein, I. H., "Textural features for image classification," *IEEE Transactions on Systems, Man and Cybernetics* (6), 610–621 (1973).

[16] Noessner, J., Niepert, M., and Stuckenschmidt, H., "ROCKIT: Exploiting parallelism and symmetry for map inference in statistical relational models," *arXiv preprint arXiv:1304.4379* (2013).

[17] Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., Max, D., Pais, R. C., Qing, D. P. Y., Roberts, R. Y., Smith, A. R., Starkey, A., Batra, P., Caligiuri, P., Farooqi, A., Gladish, G. W., Jude, C. M., Munden, R. F., Petkovska, I., Quint, L. E., Schwartz, L. H., Sundaram, B., Dodd, L. E., Fenimore, C., Gur, D., Petrick, N., Freymann, J., Kirby, J., Hughes, B., Casteele, A. V., Gupte, S., Sallam, M., Heath, M. D., Kuhn, M. H., Dharaiya, E., Burns, R., Fryd, D. S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B. Y., and Clarke, L. P., "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans," *Medical Physics* **38**(2), 915–931 (2011).