Distinguishing Adenocarcinomas From Granulomas in the CT Scan of the Chest: Performance Degradation Evaluation in the Automatic Segmentation Framework

Mahsa Bank Tavakoli
University of Tehran
Mahdi Orooji
Tarbiat Modares University
Mehdi Teimouri (mehditeimouri@ut.ac.ir)
University of Tehran https://orcid.org/0000-0003-4662-5535
Ramita Shahabifar
Kerman University of Medical Sciences

Technical advance

Keywords: Lung Cancer, Computed Tomography of the Chest, Computer-Aided Diagnosis, Radiomic Features, Vessel Tortuosity, Pulmonary Nodule Segmentation

Posted Date: July 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-36004/v1

TECHNICAL ADVANCE

Distinguishing Adenocarcinomas from Granulomas in the CT Scan of the Chest: Performance Degradation Evaluation in the Automatic Segmentation Framework

Mahsa Bank Tavakoli[1†], Mahdi Orooji[2^], Mehdi Teimouri[1*] and Ramita Shahabifar[3]

*Correspondence:
mehditeimouri@ut.ac.ir
[1]Faculty of New Sciences and Technologies, University of Tehran, North Kargar, 1439957131 Tehran, Iran
Full list of author information is available at the end of the article
[†]m_banktavakoli@ut.ac.ir
[^]morooji@modares.ac.ir

Abstract

Background: The most common histopathologic malignant and benign nodules are Adenocarcinoma and Granuloma, respectively, which have different standards of care. However, distinguishing Adenocarcinoma from Granuloma in the CT scan of the chest is a challenging task, due to the similar appearance in shape and appearance. Indeed, biopsies are needed for the diagnosis. The radiomic features of pulmonary nodules, along with the torsion of the vessels attached to the nodules are accepted by expert radiologists as the biomarker for discriminating the benign nodules from the malignant ones. In this paper, we propose an automatic framework for the distinction of the Adenocarcinomas and the Granulomas in CTs using the radiomic features of nodules and the attached vessel tortuosity.

Methods: To automatically segment the delineated nodule area and the attached vessels area, we apply a morphological-based approach. For distinguishing the malignancy of the segmented nodule, two texture features of the nodule, the curvature mean and the number of the attached vessels are extracted. Then, we apply a trained SVM classifier to identify the segmented nodule as the Adenocarcinoma or the Granuloma.

Results: The proposed framework is evaluated on a private dataset, including 22 CTs for each nodule type, i.e., adenocarcinoma and granuloma (44 CTs for both nodule types). The dataset contains the CTs of the non-smoker patients who are between 30 and 60 years old. Compared to the state-of-the-art feature selection methods which employ only the shape features of the nodule, the texture features of the nodule, or the torsion features of the attached vessels along with the radiomic features of the nodule for differentiating Adenocarcinomas from Granulomas, the accuracy of our framework is improved by 21.39%, 3.72%, and 8.27%, respectively. The area under the ROC curve of the introduced framework for the manually and automatically segmented nodules is 0.8874 and 0.7583, respectively.

Conclusions: The AUC value for the automatically segmented nodules is lower than that of the manual ones labeled by a radiologist. However, the time run of the introduced framework for the automatically segmented nodules is much lower than that of the manual ones.

Keywords: Lung Cancer; Computed Tomography of the Chest; Computer-Aided Diagnosis; Radiomic Features; Vessel Tortuosity; Pulmonary Nodule Segmentation

Background

Diagnosis of malignant nodules in early stages via Computed Tomography (CT) scans is an important step for reducing lung cancer mortality [1]. In this regard, Computer Aided Diagnosis (CADx) systems are presented that use radiomic features of suspicious nodules in CT images [2]. The most common histopathologic malignant nodules that appear as subsolid in CT images are adenocarcinomas; as a result, characterizing adenocarcinomas in CT images is challenging [3]. Also, granulomas are a broad group of benign nodules that appear similar in size, shape, and appearance to lung cancers on CT [4]. Since in most imaging-based evaluations, the benign nodules (such as granulomas) usually look similar to the cancerous nodules (such as adenocarcinomas), a biopsy is often necessary for a certain diagnosis of the cancerous ones. Unfortunately, the biopsy of the lung is an invasive, painful, and costly procedure.

Recently, CADx systems are introduced to help physicians for diagnosing cancerous nodules from benign confounders. Such systems have a significant impact on the treatment road by increasing the accuracy of the diagnosis and reducing the necessity of repeated biopsy. The CADxs for differentiating cancerous nodules from non-cancerous nodules contain three main stages: suspicious detection/segmentation, feature extraction, and classification [2]. According to the suspicious detection/segmentation stage, the CADxs can be divided into two groups: automatic and manual. In the first stage of the automatic CADxs, the boundary of the delineated nodules is determined by the radiomic features and segmentation methods [4, 5, 6, 7]. On the other hand, in the manual CADxs [8, 9, 10, 11, 12, 13, 14], the boundary of the suspicious nodules are identified completely by radiologists.

In the feature extraction stage of the CADxs, the radiomic features of the detected or segmented nodules are extracted. Radiomic features which are widely used in the CADx for lung cancer diagnosis are texture and shape features of the nodule area [2]. In [4], an automatic CADx is proposed for differentiating adenocarcinomas from granulomas, which only uses shape features of the nodules. In [14], a manual CADx is also proposed for differentiating adenocarcinomas from granulomas, which uses texture features of the nodules. Expert radiologists admit that the torsion of the vessels, which are connected to the suspicious nodules, are correlated with the histological types of the nodules, i.e, malignant and benign nodules [13, 15, 16]. This is because the vessels are pulled toward the cancerous nodules for better nourishing and growing. However, the tortuosity features of the attached vessels are considered in very few CADxs for lung cancer diagnosis. In [13], a manual CADx is proposed for differentiating adenocarcinomas from granulomas, in which the tortuosity features of the attached vessels are extracted.

In this paper, we introduce an automatic CADx that uses the radiomic features of nodules plus the tortuosity and the number of the vessels attached to the nodules for distinguishing granulomas from adenocarcinomas in lung CTs. However, in [14], [4] and [13], 12 texture features of the nodules, 3 shape features of the nodules, and 12 tortuosity features of the attached vessels are extracted, respectively. In our system, first, the nodules and the attached vessels are segmented. For segmenting the delineated nodules and the attached vessels, we apply the morphological framework introduced in [17]. The radiomic features of the nodules plus the

tortuosity and the number of the vessels attached to the nodules are then extracted. The tortuosity features are the curvature Mean, the fractional dimension Mean, and the distance metric Mean of the attached vessels. The radiomic features are 855 features, including the shape, the sharpness, and the texture features of the nodules. Next, a subset of four features is selected by the forward selection algorithm. These features are the standard deviation of the correlation feature and the kurtosis of the diagonal gradient images from the nodule area and the curvature Mean and the number of the attached vessels. The selected features are extracted from the manual and automatic annotations of the nodules and the attached vessels in the CT images. Finally, the SVM classifications with 3-fold cross-validation are performed on the selected features.

Generally, the proposed automatic CADx, like other CADxs try to help physicians for diagnosing Adenocarcinomas from Granulomas. It is not going to perform better than the physicians, but it would speed up the diagnosis process and would reduce the necessity of repeated biopsy.

The rest of the paper is as follows. In Section 2, the methods and the materials are discussed. The experimental design, classification accuracy, and the discussion are explained in Section 3. In this section, the accuracy of the proposed schema is compared with the accuracy of the state-of-the-art feature selection methods [4, 13, 14], which employ three shape features of the nodule area, 12 tortuosity features of the vessels attached to the nodule, and 12 texture features of the nodule respectively for this task.

Methods

Nodule and the Attached Vessels Segmentation

We employ the framework introduced in [17] for the segmentation of each nodule and the connected vessels in the Region Of Interest (ROI) (, i.e, a volume around the seed point). The steps are as follow:

At first, a Region Growing (RG) segmentation is applied to determine the mask of the nodule and the connected vessels. We compute the 2D vectors containing the Solidity Radius (SR) and the curvature amounts for all edge pixels of the mask. The SR value is equal to the radius of the solidity circle [17]. The solidity circle for an edge pixel is the biggest in the mask so that the interior and frontier pixels have value 1, as well as its frontier crosses the edge pixel [17]. We then label the 2D vectors as the nodule or the non-nodule by a Support Vector Machine (SVM) classifier. We specify the mask of the nodule using the convex hull image of the edge pixels classified as the nodule. At last, the connected vessels are segmented using Hessian-based multiscale filtering and a RG segmentation in the ROI.

In [17], the vessels incorrectly annotated as the nodule are specified and are removed from the nodule mask in a three-step algorithm. First, the Solidity Convex Hull (SCH) value of the nodule mask in each slice is calculated. The SCH value of the nodule mask is described as the ratio of the mask area to the area of the convex hull image [17]. In [17], the SCH values of a nodule (correctly annotated) and a vessel fragment are described more than 0.8 and less than 0.6, respectively. In addition, if a vessel fragment is connected to the nodule and is incorrectly annotated

as the nodule, this value is between 0.6 and 0.8 [17]. Then, we specify the skeleton of the annotated nodule and all junction points on it. In [17], the median junction point is described as the center point. Moreover in [17], the remote point is described as the farthest junction point from the center point. The mean distance is equal to the mean distances of the center point and the junction points [17]. The connection point is between the center and the remote points, and with a distance equal to the mean distance [17]. At last, we remove the line between the remote point and the connection point on the skeleton. The corrected nodule contains the pixels of the nodule mask related to the remained skeleton.

Feature Extraction

Three sets of 3D features, including 830 texture features, 13 shape features, and 12 sharpness features, are automatically extracted for nodule characterization. We also extract 4 features, including 3 tortuosity features and the number of the attached vessels, from the segmented vessels area.

Features of the Nodule

The texture features are divided into 3 classes. These classes are the first-order statistics, the second-order statistics, and the features of the texture images. First-order statistics characterize the distribution of voxel intensities within the image region defined by the mask [6]. We use the 6 first-order statistics features, including energy, mean, standard deviation, skewness, kurtosis, and uniformity of the nodule segment. The second-order statistics features which are extracted from the texture matrices are mentioned in Table 1. We also calculate 130 texture images from each nodule segment, which are classified into 8 classes. Then, the 6 first-order features are extracted from each texture image. The texture images are listed in Table 2.

We also extract the shape features of the segmented nodule. The shape features are volume, surface area, surface area to volume ratio, sphericity, compactness 1, compactness 2, spherical disproportion, Feret diameter, major axis length, minor axis length, least axis length, elongation, and flatness [6].

Furthermore, we employ the sharpness features of the nodule boundaries for differentiating the benign nodules from the malignant nodules. This is because the boundaries of the malignant nodules appear more spiculated and lobulated in CT images; however, the boundaries of the benign nodules are smoother [18]. The sharpness features are calculated by the intensities of the voxels on the nodule margin. These features are the difference of the max and min values, the sum of values, sum of squares, sum of logs, arithmetic mean, geometric mean, population variance, sample variance, standard deviation, kurtosis measure, skewness measure, and Second Central Moment (SCM) [19].

Features of the Attached Vessels

The vessel features, including 3 tortuosity features and the number of the attached vessels, are extracted from the segmented vessel area. The tortuosity features are the curvature Mean, the fractal dimension Mean, and the distance metric Mean.

The vascular distance metric is the ratio of the actual vessel path length to the linear distance between its endpoints [20]. All segments of the vessels attached

to the nodule in the box around the seed point are first identified. A vessel segment is defined as the path between either two branching points or between a branching point and an endpoint. For each segment, the 3D length is divided by the Euclidean distance between the endpoints. The mean value of the distance metric from all vessel segments is the distance metric Mean.

The fractal dimension is computed for each attached vessel by applying a 3D extension of the well-validated box-counting method [21]. In box-counting, the image volume which encloses the attached vessel is divided into a grid of equal cubes, and the number of cubes containing part of the attached vessel centerline is counted. This process is repeated for different cube sizes. The number of cubes containing the centerline is plotted against the cube size in a double logarithmic plot. The fractal dimension is equivalent to the slope of the fitted line. The fractal dimension mean is the average of the fractal dimensions from the attached vessels.

If a curve is defined in polar coordinates by the radius r as a function of the polar angle θ, then its curvature is defined in Eq. (0.1),

$$\kappa(\theta) = \frac{\|r^2 + 2(\frac{\mathrm{d}r}{\mathrm{d}\theta})^2 - r\frac{\mathrm{d}^2 r}{\mathrm{d}\theta^2}\|}{(r^2 + (\frac{\mathrm{d}r}{\mathrm{d}\theta})^2)^{\frac{3}{2}}}, \tag{0.1}$$

where $\frac{\mathrm{d}r}{\mathrm{d}\theta}$ and $\frac{\mathrm{d}^2 r}{\mathrm{d}\theta^2}$ are the first and second derivatives of r with respect to θ [**?**]. We extract the centerlines of the segments in the attached vessels. Then, the curvature values are calculated for all the centerlines points. The curvature mean is the average of all the curvature values.

Feature Selection

A feature set including 859 features is extracted for each CT images. A subset of four features is then selected by the forward selection. Forward selection is an iterative method in which the model has no feature at first. In each iteration, the feature which best improves the model is added. This continues until the addition of a new feature does not improve the performance of the model.

We define the loss function in the forward selection algorithm to be the number of misclassified nodules for the linear SVM model. For each candidate feature subset, the 3-fold cross-validation is performed by repeatedly calling the loss function with different training and test subsets. Hence, the features with the local minimum loss are selected. The selected features are the standard deviation of the correlation feature and the kurtosis of the diagonal gradient images from the nodule area, as well as the curvature Mean and the number of the attached vessels from the vessels area. Figure **??** illustrates the selected features for a granuloma and an adenocarcinoma.

SVM Classification

We employ a SVM classifier in conjunction with four selected features to discriminate between adenocarcinomas and granulomas. The classifier employs the linear kernel. Moreover, we use 3-fold cross-validation for obtaining classification results.

The performance of the classifiers is measured via the Area Under ROC Curve (AUC) of the Receiver Operating Characteristic (ROC) curve.

To compare the classification accuracy of the proposed feature selection method to the state-of-the-art feature selection methods [4, 13, 14], we introduce eight different scenarios that are performed by eight different SVM classifiers. The scenarios correspond to the combination of two different annotations and four different feature sets. Two types of annotation are obtained by two segmentation approaches, manual and the automatic framework. The manual annotations are made by the radiologist. The automatic annotations are resulted by the introduced framework in [17].

We consider the feature sets as follows. In two scenarios, four selected features (FS_{nv}) are employed. The feature set FS_{nv} includes the standard deviation of the correlation feature and the kurtosis of the diagonal gradient images from the nodule area, as well as the curvature Mean and the number of the attached vessels. In the other two scenarios, we consider two texture features of the nodule (FS_n). The feature set FS_n includes the standard deviation of the correlation feature and the kurtosis of the diagonal gradient images. These texture features are also used in [14] for distinguishing between adenocarcinomas and granulomas.

In two other scenarios, three shape features of the nodule area proposed by [4] (FS_s) are used in conjunction with two annotations. The feature set FS_s includes roughness, convexity, and sphericity. Also, to compare the classification accuracy of the proposed feature selection method to the state-of-the-art feature selection method [13], two more scenarios are considered. In [13], 12 tortuosity features of the vessels attached to nodules are introduced for distinguishing granulomas from adenocarcinomas. We select two features from these 12 features by the forward selection algorithm. Two selected features (FS_v) are the Max value of the maximum curvature values of the vessel branches and the 4th bin value from the Histogram of torsion measurements of the branches. These features are important for distinguishing benign nodules from the malignant ones because the malignant one tends to pull the vessels toward itself for better feeding and growing. As a result, the vessels which reach the malignant nodule are more torsion. Hence, in these scenarios, two selected tortuosity features of the attached vessels FS_v are used in conjunction with two annotations.

Results

Data

We employ a database from the Afzalipour Hospital of Kerman. The database includes the CTs of 44 non-smoker patients, who were between 30 and 60 years old. Each case had a dubious nodule of size 11.91 ± 4.36mm. The database consists of 22 CTs for each nodule type, i.e., Adenocarcinoma and Granuloma. All the CT scans were collected as part of an Institutional Review Board-approved, HIPAA-compliant protocol. In addition, these CTs were constructed by the Siemens scanner machine with exposure 120 KVp, slice thickness of 1-5mm, and an X-ray tube current of 41-200mAs. All the CTs have 100 to 400 slices. The resolution of each slice is 512×512 pixels. The type of each nodule is also described using the microscopic analysis of the nodule specimen gathered by biopsy and/or surgical resection. Moreover, the

CTs consist of three popular types of nodules (containing solid, part-solid, and non-solid). The regions of the nodule and the connected vessels in the database are also annotated by a skilled radiologist.

Classification Accuracy

The classification results of manually and automatically segmented nodules are provided in table 3. The classification results of the segmented nodules using the feature set (proposed in this paper) FS_{nv}, three shape features (introduced in [4]) FS_s, two selected texture features (introduced in [14]) FS_n, and two selected tortuosity features of the attached vessels (used in [13]) FS_v are also reported in table 3. As it can be seen in table 3, the AUC values of the classifications by the feature set FS_{nv} improve compared to those of the feature sets FS_n, FS_s, and FS_v for both manual and automatically segmented nodules.

Discussion

In this study, we investigated the role of automatic segmentation of the delineated pulmonary nodule and the attached vessels, as well as computerized image analysis to identify a set of nodule texture features and the attached vessels tortuosity that best distinguish adenocarcinomas from granulomas on CT scans of the chest. Our study revealed that the standard deviation of the correlation feature and the kurtosis of the diagonal gradient images from the nodule area, as well as the curvature Mean and the number of the attached vessels from the vessels area, were the most predictive and discriminating features. The performance of the SVM classifier has resulted in an AUC of 75.83% and 88.74%, on the texture and tortuosity features FS_{nv}, using the automatic framework (proposed in [17]) and the manual segments (labeled by an expert radiologist), respectively. The performance of the SVM classifier using the manual segments is about 13% better than that of the automatic segments. However, the time needed for the annotation of each nodule and the attached vessels by a radiologist is much higher than that of the automatic segmentation. Hence, in the process of lung cancer screening in which the number of cases is high, using the automatic segmentation would be much more cost-effective.

In [14], 12 texture features of the nodule are employed to discriminate granulomas from adenocarcinomas. We selected two features (FS_n) from these texture features by the forward selection algorithm. We then extracted two selected texture features FS_n from our dataset to differentiate the granulomas from the adenocarcinomas. In [4], three shape features of the nodule (FS_s), including roughness, convexity, and sphericity are also employed, to discriminate granulomas from adenocarcinomas. In our study, we also extracted these shape features from our dataset to differentiate the granulomas from the adenocarcinomas. Moreover, In [13], 12 tortuosity features of the attached vessels are employed to discriminate granulomas from adenocarcinomas. We selected two features (FS_v) from these 12 tortuosity features of the attached vessels by the forward selection algorithm. We then extracted two selected tortuosity features FS_v from our dataset.

In the case of applying the segmentation framework in [17], the use of the texture and tortuosity features (FS_{nv}) amplified the performance of the SVM classifier compared to those of two selected tortuosity features (FS_v), two selected

texture features (FS_n), or the shape features (FS_s) alone, with an increase of 8.27%, 3.72%, and 21.39% in the AUC values, respectively. In the case of using manual segments (labeled by the radiologist), an AUC value of 85.95%, 73.33%, and 63.42% was obtained on two selected texture features (FS_n), two selected tortuosity features (FS_v), and the shape features (FS_s) alone, respectively. As a result, the AUC value of the classifier with the texture and tortuosity features (FS_{nv}) improves, in comparison to those of two texture features (FS_n), two tortuosity features (FS_v), and the shape features (FS_s) alone, with a rise of 2.79%, 15.41%, and 25.32%, respectively. This result demonstrates the superiority of the proposed feature selection method compared to the state-of-the-art feature selection methods [4, 13, 14].

Conclusions

Early distinguishing granulomas from adenocarcinomas as the most common benign and malignant nodules, respectively, is very important to increase the five-year survival rate for lung cancer. Granulomas and adenocarcinomas look similar in the CT scans. As a result, it is needed to perform surgical wedge resections or biopsies for precise diagnosis. Many CADx systems have been proposed to diagnose lung cancer early by the radiomic features of nodules. Tortuosity features of the vessels that are attached to nodules are also biomarkers for differentiating cancerous nodules from benign nodules. However, very few CADxs [13] employ the tortuosity features. In this paper, we developed an integrated schema for the automatic segmentation and the radiomic-based and tortuosity-based characterization of nodules and the attached vessels for discriminating adenocarcinomas from granulomas, respectively.

Our schema used the nodule and the attached vessels segmentation framework in [17] that employs morphological methods and extracts the nodule and the attached vessels area. A set of 3D texture and tortuosity features were extracted from the nodule and the attached vessels area, respectively. These features were used to distinguish granulomas and adenocarcinomas using a SVM classifier. Four features were then selected. The selected features are the standard deviation of the correlation feature and the kurtosis of the diagonal gradient images from the nodule area, as well as the curvature Mean and the number of the attached vessels. At last, we compared the proposed feature selection method with the state-of-art feature selection methods [4, 13, 14] for distinguishing adenocarcinomas and granulomas. The performance of the classifier with the proposed texture and tortuosity features improved in comparison to those of two selected texture features of the nodule proposed in [14], two selected tortuosity features of the attached vessels proposed in [13], or three shape features of the nodule proposed in [4]. The major finding of our study is that the tortuosity and number of the attached vessels, along with the radiomic features of the nodule are efficient biomarkers for discriminating granulomas from adenocarcinomas.

Our study did have its limitations which included using datasets consist of one specific type of benign and malignant pathology, i.e. granulomas and adenocarcinomas from only one institution. For a more general conclusion, the introduced frameworks must be evaluated on the independent cohort. As future work, it is valuable to evaluate the discriminability of the features and the classifier in distinguishing other benign conditions such as hamartoma and fibrosis from other types of non-small cell lung cancers like squamous cell carcinomas.

Abbreviations

AUC: Area Under ROC Curve
CADx: Computer Aided Diagnosis
CT: Computed Tomography
RG: Region Growing
ROC: Receiver Operating Characteristic
ROI: Region Of Interest
SCH: Solidity Convex Hull
SCM: Second Central Moment
SR: Solidity Radius
SVM: Support Vector Machine

Declarations

Ethics approval and consent to participate

In the research, All the CT scans in the private Dataset were achieved as the clinical standard care of the patients in accordance with the Declaration of Helsinki supervised by the Ethics Committee of the Kerman University Of Medical Science.

Consent for publication

Not applicable

Availability of data and materials

The data that support the findings of this study are available from the Kerman University Of Medical Science but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the Kerman University Of Medical Science.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable

Author's contributions

MB, MT, and MO wrote the main manuscript. The computational method and algorithms of the paper were developed by MB and MT. All figures and tables are generated by MB and MT. Data collection was mentored by MT and MO, and collected by MB. RS labeled the nodules and the attached vessels in our private dataset. All authors reviewed the manuscript.

Acknowledgements

Not Applicable

Author details

[1]Faculty of New Sciences and Technologies, University of Tehran, North Kargar, 1439957131 Tehran, Iran. [2]Department of Electrical and Computer Engineering, Tarbiat Modares University, Jalal AleAhmad, 14115-111 Tehran, Iran. [3]Kerman University Of Medical Science, Kerman, Iran.

References

1. Midthun, D.E.: Early detection of lung cancer. F1000Res **5** (2016). doi:10.12688/f1000research.7313.1
2. Tharcis, P., Kezi Selva Vijila, C.: Computer-aided diagnosis of lung cancer in computed tomography scans: A review **14**(3), 374–388 (2018)
3. Ko, J.P., Suh, J., Ibidapo, O., Escalon, J.G., Li, J., Pass, H., Naidich, D.P., Crawford, B., Tsai, E.B., Koo, C.W., Mikheev, A., Rusinek, H.: Lung adenocarcinoma: Correlation of quantitative ct findings with pathologic findings. Radiology **280**(3), 931–939 (2016). doi:10.1148/radiol.2016142975. PMID: 27097236. https://doi.org/10.1148/radiol.2016142975
4. Alilou, M., Beig, N., Orooji, M., Rajiah, P., Velcheti, V., Rakshit, S., Reddy, N., Yang, M., Jacono, F., Gilkeson, R.C., Linden, P., Madabhushi, A.: An integrated segmentation and shape-based classification scheme for distinguishing adenocarcinomas from granulomas on lung ct. Med Phys **44**(7), 3556–3569 (2017). doi:10.1002/mp.12208
5. Mukherjee, J., Chakrabarti, A., Shaikh, S.H., Kar, M.: Automatic detection and classification of solitary pulmonary nodules from lung ct images. In: 2014 Fourth International Conference of Emerging Applications of Information Technology, pp. 294–299 (2014). doi:10.1109/EAIT.2014.64
6. Namin, S.T., Moghaddam, H.A., Jafari, R., Esmaeil-Zadeh, M., Gity, M.: Automated detection and classification of pulmonary nodules in 3d thoracic ct images. In: 2010 IEEE International Conference on Systems, Man and Cybernetics, pp. 3774–3779 (2010). doi:10.1109/ICSMC.2010.5641820
7. Gong, J., Liu, J.Y., Sun, X.W., Zheng, B., Nie, S.D.: Computer-aided diagnosis of lung cancer: the effect of training data sets on classification accuracy of lung nodules. Phys Med Biol **63**(3), 035036 (2018). doi:10.1088/1361-6560/aaa610
8. de Carvalho Filho, A.O., Silva, A.C., Cardoso de Paiva, A., Nunes, R.A., Gattass, M.: Computer-aided diagnosis of lung nodules in computed tomography by using phylogenetic diversity, genetic algorithm, and svm. Journal of Digital Imaging **30**(6), 812–822 (2017). doi:10.1007/s10278-017-9973-6
9. de Carvalho Filho, A.O., Silva, A.C., Cardoso de Paiva, A., Nunes, R.A., Gattass, M.: Computer-aided diagnosis system for lung nodules based on computed tomography using shape analysis, a genetic algorithm, and svm. Med Biol Eng Comput **55**(8), 1129–1146 (2017). doi:10.1007/s11517-016-1577-7
10. Huang, P., Park, S., Yan, R., Lee, J., Chu, L.C., Lin, C.T., Hussien, A., Rathmell, J., Thomas, B., Chen, C., Hales, R., Ettinger, D.S., Brock, M., Hu, P., Fishman, E.K., Gabrielson, E., Lam, S.: Added value of computer-aided ct image features for early lung cancer diagnosis with small pulmonary nodules: A matched case-control study. Radiology **286**(1), 286–295 (2018). doi:10.1148/radiol.2017162725
11. Li, Z., Hu, Z., Xu, J., Tan, T., Chen, H., Duan, Z., Liu, P., Tang, J., Cai, G., Ouyang, Q., Tang, Y., Litjens, G., Li, Q.: Computer-aided diagnosis of lung carcinoma using deep learning - a pilot study (2018). 1803.05471
12. Nishio, M., Nishizawa, M., Sugiyama, O., Kojima, R., Yakami, M., Kuroda, T., Togashi, K.: Computer-aided diagnosis of lung nodule using gradient tree boosting and bayesian optimization. PLOS ONE **13**(4), 0195875 (2018). doi:10.1371/journal.pone.0195875
13. Alilou, M., Orooji, M., Beig, N., Prasanna, P., Rajiah, P., Donatelli, C., Velcheti, V., Rakshit, S., Yang, M., Jacono, F., Gilkeson, R., Linden, P., Madabhushi, A.: Quantitative vessel tortuosity: A potential ct imaging biomarker for distinguishing lung granulomas from adenocarcinomas. Scientific Reports **8**(1), 15290 (2018). doi:10.1038/s41598-018-33473-0
14. Beig, N., Khorrami, H., Alilou, M., Prasanna, P., Braman, N., Orooji, M., Rakshit, S., Bera, K., Rajiah, P., Ginsberg, J., Donatelli, C., Thawani, R., Yang, M., Jacono, F., Tiwari, P., Velcheti, V., Gilkeson, R., Linden, P., Madabhushi, A.: Perinodular and intranodular radiomic features on lung ct images distinguish adenocarcinomas from granulomas. Radiology **290**, 180910 (2018). doi:10.1148/radiol.2018180910
15. Bullitt, E., Wolthusen, P.A., Brubaker, L., Lin, W., Zeng, D., Van Dyke, T.: Malignancy-associated vessel tortuosity: a computer-assisted, mr angiographic study of choroid plexus carcinoma in genetically engineered mice. AJNR Am J Neuroradiol **27**(3), 612–9 (2006)
16. Khunger, M., Alilou, M., Thawani, R., Madabhushi, A., Velcheti, V.: Computer extracted measurements of vessel tortuosity on baseline ct scans to predict response to nivolumab immunotherapy for non-small cell lung cancer. Journal of Clinical Oncology **35**(15_suppl), 11566–11566 (2017). doi:10.1200/JCO.2017.35.15_suppl.11566. https://doi.org/10.1200/JCO.2017.35.15_suppl.11566
17. Bank Tavakoli, M., Orooji, M., Teimouri, M., Shahabifar, R.: Segmentation of the pulmonary nodule and the attached vessels in the ct scan of the chest using morphological features and topological skeleton of the nodule. Accepted for puplication in IET Image Processing **14**(8), 1520–1528. doi:10.1049/iet-ipr.2019.1054
18. Snoeckx, A., Reyntiens, P., Desbuquoit, D., Spinhoven, M.J., Van Schil, P.E., van Meerbeeck, J.P., Parizel, P.M.: Evaluation of the solitary pulmonary nodule: size matters, but do not ignore the power of morphology. Insights Imaging **9**(1), 73–86 (2018). doi:10.1007/s13244-017-0581-2
19. Ferreira, J. J. R., Oliveira, M.C., de Azevedo-Marques, P.M.: Characterization of pulmonary nodules based on features of margin sharpness and texture. J Digit Imaging **31**(4), 451–463 (2018). doi:10.1007/s10278-017-0029-8
20. Helmberger, M., Pienn, M., Urschler, M., Kullnig, P., Stollberger, R., Kovacs, G., Olschewski, A., Olschewski, H., Bálint, Z.: Quantification of tortuosity and fractal dimension of the lung vessels in pulmonary hypertension patients. PLOS ONE **9**(1), 1–9 (2014). doi:10.1371/journal.pone.0087515
21. Ge, M., Lin, Q., Lu, W.: Realizing the box-counting method for calculating fractal dimension of urban form based on remote sensing image. In: 2006 IEEE International Symposium on Geoscience and Remote Sensing, pp. 1423–1426 (2006). doi:10.1109/IGARSS.2006.367

Figure Legends

Table Legends

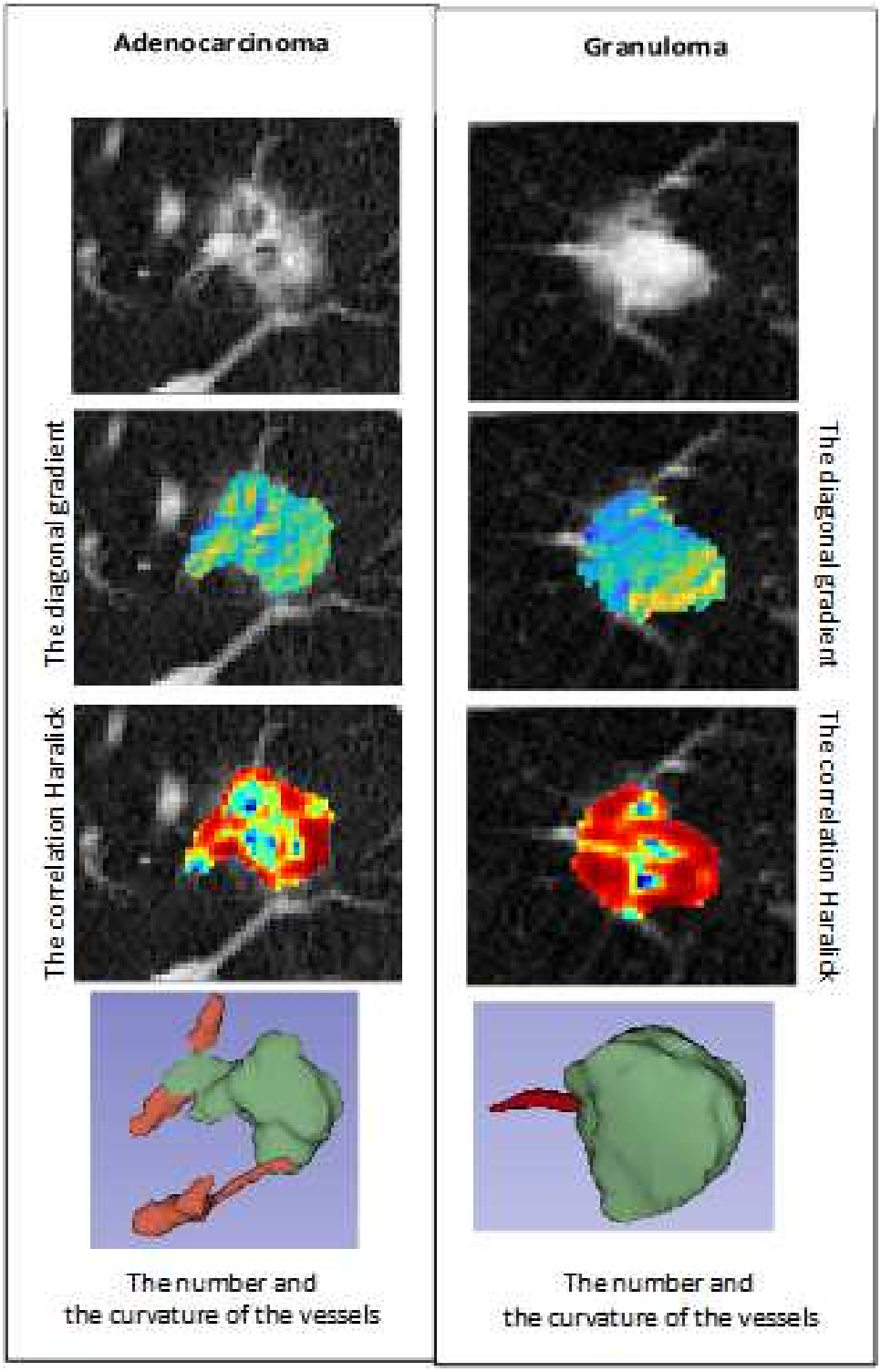

Figure 1 The illustration of the 4 selected features for a granuloma and an adenocarcinoma.

Table 1 The second-order features which are extracted from the texture matrices.

texture matrix	features
Gray Level Run Length Matrix (GLRLM)	Short Run Emphasis (SRE), Long Run Emphasis (LRE), Gray-Level Nonuniformity (GLN), Run-Length Nonuniformity (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), Long Run High Gray-Level Emphasis (LRHGE), Gray-Level Variance (GLV), and Run-Length Variance (RLV)
Gray Level Size Zone Matrix (GLSZM)	Small Zone Emphasis (SZE), Large Zone Emphasis (LZE), Gray-Level Nonuniformity (GLN), Zone-Size Nonuniformity (ZSN), Zone Percentage (ZP), Low Gray-Level Zone Emphasis (LGZE), High Gray-Level Zone Emphasis (HGZE), Small Zone Low Gray-Level Emphasis (SZLGE), Small Zone High Gray-Level Emphasis (SZHGE), Large Zone Low Gray-Level Emphasis (LZLGE), Large Zone High Gray-Level Emphasis (LZHGE), Gray-Level Variance (GLV), and Zone-Size Variance (ZSV)
Neigboring Gray Tone Difference Matrix (NGTDM)	coarseness, contrast, busyness, complexity, and strength
Gray Level Dependence Matrix (GLDM)	Small Dependence Emphasis (SDE), Large Dependence Emphasis (LDE), Gray Level Non-Uniformity (GLN), Gray Level Non-Uniformity Normalized (GLNN), Dependence Non-Uniformity (DN), Dependence Non-Uniformity Normalized (DNN), Gray Level Variance (GLV), Dependence Variance (DV), Dependence Entropy (DE), High Gray Level Emphasis (HGLE), Small Dependence Low Gray Level Emphasis (SDLGLE), Large Dependence High Gray Level Emphasis (LDHGLE), and Large Dependence Low Gray Level Emphasis (LDLGLE)

Table 2 The texture images which are extracted from the nodule area.

texture image	imaged features
Haralick	entropy, energy, inertia, inverse difference moment, correlation, information measure of correlation I, information measure of correlation II, sum average, sum variance, sum entropy, difference average, difference variance, difference entropy
Local Binary Pattern (LBP)	one feature with a 3×3 kernel
gray filters	mean, median, standard deviation, and range
gradient filters	x Sobel Edge, y Sobel Edge, diagonal xy Sobel Edge, diagonal yx Sobel Edge, directional gradient, magnitude gradient, sum square of the directional and the magnitude gradients, Kirsch 1, Kirsch 2, Kirsch 3, dx, dy, and diagonal derivative
Laws filters	25 features by 25 5×5 kernels, such as: $L5L5$, $L5E5$, $L5W5$, etc. L, E, and W are one-dimensional filters respectively for Level detection, Edge detection, and Wave detection.
Laws laplace filters	25 features by 25 5×5 kernels, such as: $L5L5$, $L5E5$, $L5W5$, etc.
autocorrelation	one feature
Gabor filters	48 features by a 48th Gabor filter bank

Table 3 The classification results of manually and automatically segmented nodules.

Annotation	Feature Set	AUC
Manual	4 Selected Features (FS_{nv})	0.8874
Manual	2 Selected Texture Features (FS_n) by [14]	0.8595
Manual	2 Selected tortuosity Features by [13] (FS_v)	0.7333
Manual	3 Selected Shape Features by [4] (FS_s)	0.6342
the Automatic Framework [17]	FS_{nv}	0.7583
the Automatic Framework [17]	FS_n	0.7211
the Automatic Framework [17]	FS_v	0.6756
the Automatic Framework [17]	FS_s	0.5444

Figures

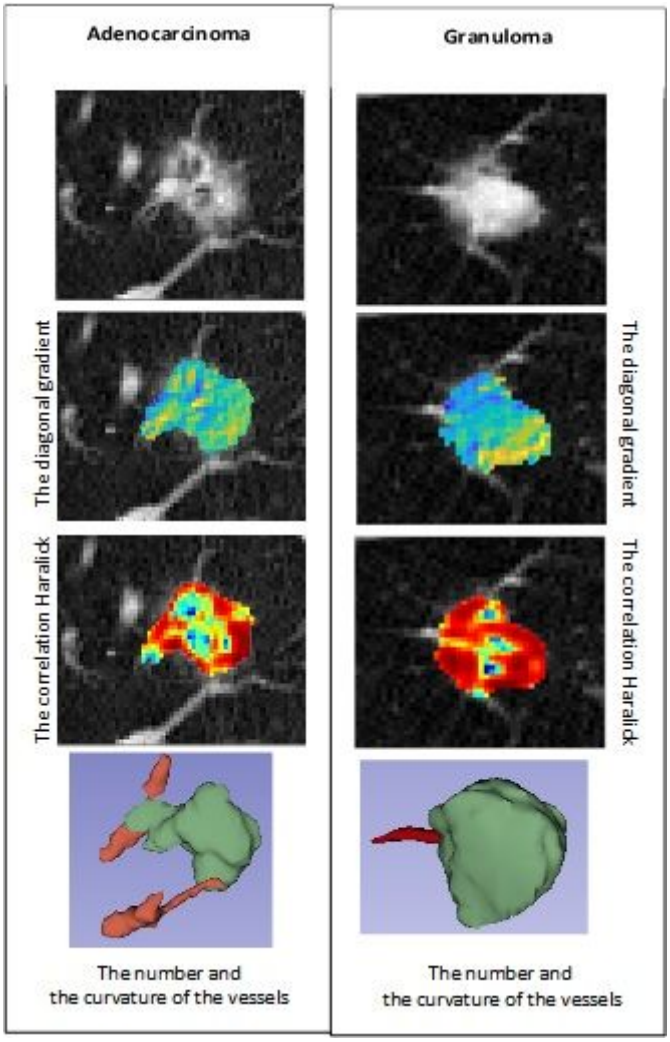

Figure 1

The illustration of the 4 selected features for a granuloma and an adenocarcinoma

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- refs.bib
- BMIMD2000279R2.pdf