# Using a radiomics-derived classifier to distinguish between lung adenocarcinomas and granulomas—where are we now?

**David J. Murphy[1], Andrea Bille[2]**

[1]Department of Radiology, St. Vincent's University Hospital, Dublin, Ireland; [2]Department of Thoracic Surgery, Guy's and St. Thomas' NHS Foundation Trust, King's College University, London, UK

*Correspondence to:* David J. Murphy, FRCR. Department of Radiology, St. Vincent's University Hospital, Elm Park, Dublin 4, Ireland. Email: david.murphy@st-vincents.ie.

*Comment on:* Orooji M, Alilou M, Rakshit S, *et al.* Combination of computer extracted shape and texture features enables discrimination of granulomas from adenocarcinoma on chest computed tomography. J Med Imaging (Bellingham) 2018;5:024501.

The ability to accurately distinguish benign from malignant pulmonary nodules based solely on imaging features is the holy grail of chest radiology. Incidentally discovered indeterminate pulmonary nodules are often subject to lengthy imaging follow-up to confirm benign behaviour (1,2). Clinico-radiological scoring systems can also be used to assess the likelihood that a given pulmonary nodule is malignant (3). These validated scoring systems, such as the Brock model, incorporate patient factors such as age, sex, lung cancer family history and smoking status, along with semantic radiological features such as nodule diameter and spiculation to determine a percentage risk of malignancy for a given nodule (4). Despite use of these validated scoring systems, and the addition of advanced imaging techniques such as 18 fluorodeoxyglucose positron emission tomography computed tomography (^{18}F-FDG PET/CT) (5), it is often not possible to confidently distinguish benign from malignant aetiologies. This results in a large number of patients undergoing biopsy (CT guided, bronchoscopic guided or surgical wedge resection) for benign entities, estimated at approximately 26% of all nodule biopsies in the US (6).

The use of machine learning technology in pulmonary nodule assessment represents an exciting development in chest radiology, and has the potential to provide a robust, non-invasive method of distinguishing benign from malignant nodules. The use of computer-aided diagnosis (CAD) algorithms for the detection of pulmonary nodules have been employed since the early 2000s (7). Recent advances in computer technology allows for a more complex assessment, enabling a computerized feature-based analysis of lung nodules that has the potential to aid discrimination of malignant from benign lesions (8). The novel, rapidly changing field of radiomics employs high-throughput computing to extract innumerable quantitative features from diagnostic medical images, converting digital medical images into mineable data (9). The process of radiomics begins with the acquisition of diagnostic quality medical images-in the example of pulmonary nodule assessment, this usually involves acquisition of a non-contrast CT thorax. From these images the region of interest (ROI), for example the nodule under assessment, is then segmented. Nodule segmentation can now often be performed automatically by post-processing software, with minimal operator contour edits (10). Once the lesion is segmented, quantitative features are extracted using high-throughput computing, generating a quantitative report which can then be compared with clinical and genomic data to discover potential relationships. The use of these quantitative features describing the shape and texture of a nodule are amongst the most promising radiomics techniques under investigation in the field of nodule assessment. Texture analysis techniques provides a quantitative description of the internal heterogeneity of a lesion by analysing the distribution and relationship of pixel grey levels in the image (11). Orooji *et al.* (12) explore the ability of a machine learning classifier employing a combination of radiomic shape and texture features to distinguish between

adenocarcinomas and granulomas. The authors chose to compare adenocarcinomas with granulomas, as these are the most common histological malignant and benign diagnoses encountered (13).

This paper by Orooji *et al.* (12) is a retrospective study across two centres examining the ability of a machine learning classifier to discriminate between adenocarcinomas and granulomas using a combination of radiomic texture and shape features derived from non-contrast CT scans. After excluding patients with multiple nodules and CT scans with artefacts, 195 nodules from 195 subjects with histological confirmation from two sites were included. These were divided into two groups, a training set of 139 nodules (70 adenocarcinomas and 69 granulomas) from one institution, and an independent test set of 56 nodules (34 adenocarcinomas and 22 granulomas) from a second institution. The nodules were manually segmented, and a total of 645 two-dimensional (2D) texture and 24 three-dimensional (3D) shape features were extracted from the segmented nodule ROIs. The top computer-extracted discriminating radiomics features were then optimized on the training set to determine the likelihood of a nodule being an adenocarcinoma. The classifier was then independently validated on the test set, and compared with the interpretation of two human readers (one expert thoracic radiologist, one pulmonology fellow with training in CT thorax interpretation). The texture feature examined were extensive, including 1st order histogram features, Haralick features, Laws features, Laplacian pyramids, grey level features, Gabor features, gradient features and local binary patterns. Three-dimensional shape features were extracted in an attempt to quantify irregularities in shape that can result from internal tumour heterogeneity, and included measurements of nodule width, height, depth, area, perimeter, eccentricity, extend, compactness, radial distance, roughness, elongation, convexity, equivalent diameter and sphericity. A comprehensive description of these radiomics texture and shape features is beyond the scope of this article, but are described in detail in the manuscript text (12), and in excellent review articles by Lubner *et al.* (11), Gillies *et al.* (9) and Bashir *et al.* (14). The authors used a feature selection approach to identify the combined top 6 texture and shape features, which were then used to train three machine learning classifiers [linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and support vector machine (SVM)]. These machine learning classifiers were then applied to the validation set to predict the probability

of a nodule being an adenocarcinoma, and the results compared with the two human readers. An attempt was made to determine the robustness of the radiomic features across the two sites, different CT scanners and CT slice thickness by calculating the preparation-induced instability (PI) score for the top shape and texture radiomic features identified in the testing and training cohorts. The PI number is a score between 0–1 which quantifies the stability of radiomic features across two separate cohorts, with a PI score closer to 0 implying stability of the feature (15).

The top texture features identified across the three classifiers were the "skewness of Laws features (L5 × E5) and (L5 × R5)", "skewness of gradient features" and "Gabor texture features". The top performing shape features were "mean of extend", "mean of convexity" and "variance of eccentricity". Texture features outperformed shape features overall; for example, the SVM classifier AUC for the top performing texture feature "skewness of Law L5 × E5" was 81.9%±0.9% and for the top performing "mean of extend" shape feature was 69.3%±0.9%. The most stable and reproducible texture feature was "variance of sum variation", and "mean of extend" was the most stable shape feature, as measured by the PI score. Tumour nodules showed more internal heterogeneity than granulomas, but there was no significant difference in nodule diameter or mean Hounsfield units (Hu). The best AUC on the training set was 92.9%±1.1% for a combination of 4 texture and 2 shape features using the SVM machine-learning classifier, with a resulting AUC of 77.8% for the locked down SVM classifier on the independent test set. The AUCs for the expert thoracic radiologist and pulmonologist readers were 72.4% and 69.7% respectively. Interestingly, there was no significant difference found in the adenocarcinoma nodule prediction results between the locked down SVM classifier and the expert thoracic radiologist on the independent test set.

The major limitation of this study was their decision to limit their analysis to one specific type of benign and malignant pathology. This potentially limits the immediate clinical applicability, as there are a myriad of potential histological diagnoses when faced with a pulmonary nodule. In addition, the human readers were not provided with any clinical details when making their assessment, which may have negatively impacted their performance. Although the authors did make use of datasets from separate institutions with different scanners and scan protocols, the question of how generalizable the results are across multiple different sites remains.

This paper is an interesting attempt to identify, and implement, the key shape and textural features derived from a computerized CT image analysis that enables a reliable, non-invasive method of distinguishing between adenocarcinomas and granulomas. Previous studies have examined the potential role of texture analysis in distinguishing benign from malignant nodules, but this is the first attempt to use a combination of texture and shape radiomic features to discern adenocarcinomas from granulomas. Dennie *et al.* (16) utilised Haralick-related textural features to distinguish between granulomas and primary lung cancers (any histological subtype) in 55 nodules, reporting an AUC of 90.2%. This is similar to the AUC of 92.9% for the combination of textural and shape features in the training set, although they did not validate their findings on an independent test set. Kido *et al.* (17) examined the fractal dimension (FD) of nodule edges in 117 patients. FD is a mathematical measurement of an object's intrinsic shape. They found that malignant lesions had a lower 3D FD than two benign pathologies (organizing pneumonia and tuberculoma). The same group also examined the FD of 70 subjects with lung tumours, finding that lepidic-type adenocarcinomas had higher FDs than non-lepidic adenocarcinomas and squamous cell tumours (18). McNitt-Gray *et al.* (19) used second-order grey-level co-occurrence matrix (GLCM) texture features to try and distinguish benign from malignant aetiologies in 32 nodules, finding that four features classified 94% (n=30) of nodules correctly, and that all nodules were correctly classified when 9 features were utilised. Suo *et al.* (20) examined textural heterogeneity differences between the edge and core of 48 nodules (24 malignant, 20 inflammatory), finding a significant difference in malignant lesions. A combination of mean HU difference, entropy difference and lesion volume gave an AUC of 86.4% in detecting malignant nodules. Lee *et al.* (21) employed textural analysis in 77 part-solid nodules (PSNs) to try and distinguish transient from persistent PSNs, finding significant differences in mean HU attention, skewness and mean HU ratio between transient and persistent PSNs, although they did not have histological correlation for the persistent PSNs.

The body of literature published to-date suggests that radiomics could play a role in helping radiologists and clinicians distinguish benign from malignant lung lesions non-invasively, but how and when this technique may come into routine clinical practice is as of yet unclear. The majority of radiomics studies published are single-centre with small cohorts, which potentially limits their generalizability. The innumerable potentially evaluable radiomics features results in heterogeneity in the various shape and textural features examined across individual studies. This, in addition to the lack of validation cohorts in many studies, may also limit immediate clinical applicability. The methodology used by Orooji *et al.* (12) in combining an examination of textural and shape features in a cohort with histological confirmation, and their use of a separate validation test cohort, is a good attempt to develop a robust radiomics model for discriminating benign from malignant nodules. The ideal scenario is to develop an accurate, robust and reproducible classifier based on computer extracted shape and texture features that provides a decision support tool for thoracic radiologists when determining the risk that a given nodule is malignant. However, before any radiomics-derived classifier can be employed into the routine clinical reporting workflow, further work will be needed to evaluate the discriminability of the features identified across the plethora of potential benign and malignant nodule aetiologies. Furthermore, a more rigorous assessment of the impact of the use of different scanners, scanning protocols, iodinated contrast, slice thickness and reconstruction algorithms on any proposed discriminatory radiomics features needs to be performed to determine generalizability (22,23). It would also be interesting to examine any incremental benefit in using a radiomics classifier in addition to one of the existing, validated clinico-radiological nodule malignancy predictor scoring systems (3), particularly given the lack of an overall significant difference between the locked down classifier and the expert thoracic radiologist in discriminating adenocarcinomas from granulomas in the test set (12).

In conclusion, computer-extracted texture and shape features appear to be feasible and reproducible as discriminators of adenocarcinomas from granulomas. These classifiers are not yet ready for routine clinical practice, but following further validation in larger, multi-centre cohorts with a wider array of histologically confirmed benign and malignant pathologies, a classifier based on these computer-extracted features has the potential to provide a valuable decision support tool aiding the non-invasive discrimination of malignant from benign pulmonary nodules.

## Footnote

*Provenance and Peer Review:* This article was commissioned by the editorial office, *Journal of Medical Artificial Intelligence*. The article did not undergo external peer review.

*Conflicts of Interest:* Both authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/jmai.2018.08.02). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

## References

1. Callister MEJ, Baldwin DR, Akram AR, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. Thorax 2015;70 Suppl 2:ii1-54.
2. MacMahon H, Naidich DP, Goo JM, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. Radiology 2017;284:228-43.
3. Al-Ameri A, Malhotra P, Thygesen H, et al. Risk of malignancy in pulmonary nodules: A validation study of four prediction models. Lung Cancer 2015;89:27-30.
4. McWilliams A, Tammemagi MC, Mayo JR, et al. Probability of cancer in pulmonary nodules detected on first screening CT. N Engl J Med 2013;369:910-9.
5. Herder GJ, van Tinteren H, Golding RP, et al. Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. Chest 2005;128:2490-6.
6. Wood DE, Eapen GA, Ettinger DS, et al. Lung cancer screening. J Natl Compr Canc Netw 2012;10:240-65.
7. Gurcan MN, Sahiner B, Petrick N, et al. Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. Med Phys 2002;29:2552-8.
8. Lee G, Lee HY, Park H, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. Eur J Radiol 2017;86:297-307.
9. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology 2016;278:563-77.
10. Devaraj A, van Ginneken B, Nair A, et al. Use of Volumetry for Lung Nodule Management: Theory and Practice. Radiology 2017;284:630-44.
11. Lubner MG, Smith AD, Sandrasegaran K, et al. CT Texture Analysis: Definitions, Applications, Biologic Correlates, and Challenges. Radiographics 2017;37:1483-503.
12. Orooji M, Alilou M, Rakshit S, et al. Combination of computer extracted shape and texture features enables discrimination of granulomas from adenocarcinoma on chest computed tomography. J Med Imaging (Bellingham) 2018;5:024501.
13. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. J Thorac Oncol 2015;10:1243-60.
14. Bashir U, Siddique MM, Mclean E, et al. Imaging Heterogeneity in Lung Cancer: Techniques, Applications, and Challenges. AJR Am J Roentgenol 2016;207:534-43.
15. Leo P, Lee G, Shih NN, et al. Evaluating stability of histomorphometric features across scanner and staining variations: prostate cancer diagnosis from whole slide images. J Med Imaging (Bellingham) 2016;3:047502.
16. Dennie C, Thornhill R, Sethi-Virmani V, et al. Role of quantitative computed tomography texture analysis in the differentiation of primary lung cancer and granulomatous nodules. Quant Imaging Med Surg 2016;6:6-15.
17. Kido S, Kuriyama K, Higashiyama M, et al. Fractal analysis of small peripheral pulmonary nodules in thin-section CT: evaluation of the lung-nodule interfaces. J Comput Assist Tomogr 2002;26:573-8.
18. Kido S, Kuriyama K, Higashiyama M, et al. Fractal analysis of internal and peripheral textures of small peripheral

bronchogenic carcinomas in thin-section computed tomography: comparison of bronchioloalveolar cell carcinomas with nonbronchioloalveolar cell carcinomas. J Comput Assist Tomogr 2003;27:56-61.

19. McNitt-Gray MF, Wyckoff N, Sayre JW, et al. The effects of co-occurrence matrix based texture parameters on the classification of solitary pulmonary nodules imaged on computed tomography. Comput Med Imaging Graph 1999;23:339-48.

20. Suo S, Cheng J, Cao M, et al. Assessment of Heterogeneity Difference Between Edge and Core by Using Texture Analysis: Differentiation of Malignant From Inflammatory Pulmonary Nodules and Masses. Acad Radiol 2016;23:1115-22.

21. Lee SH, Lee SM, Goo JM, et al. Usefulness of texture analysis in differentiating transient from persistent part-solid nodules(PSNs): a retrospective study. PLoS One 2014;9:e85167.

22. Kim H, Park CM, Lee M, et al. Impact of Reconstruction Algorithms on CT Radiomic Features of Pulmonary Tumors: Analysis of Intra- and Inter-Reader Variability and Inter-Reconstruction Algorithm Variability. PLoS One 2016;11:e0164924.

23. He L, Huang Y, Ma Z, et al. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. Sci Rep 2016;6:34921.