

Pulmonary Nodules: Assessing the Imaging Biomarkers of Malignancy in a “Coffee-Break”

A. Talwar, J.M.Y. Willaime, L.C. Pickup, M. Enescu, D. Boukerroui, W. Hickes, N.M. Rahman, M. J. Gooding, T. Kadir, F. V. Gleeson

Abstract

Purpose

To prospectively assess the repeatability of CT texture features and semi-automated volume measurements extracted from pulmonary nodules (PNs) at low-dose multi-detector row computed tomography (CT). The secondary objective was to investigate the influence of two methods of inspiratory control on the repeatability of CTTA and volumetry.

Introduction

Although nodule volumetry is a recognized biomarker of malignancy in pulmonary nodules (PNs), caution is needed in its interpretation because of variables such as cardiac motion, respiratory volume variation and inter-scan variability of up to 25%. The alternative technique of CT texture as a potential independent biomarker of malignancy has been suggested, but interscan variability and respiratory volume variation has not been assessed.

Materials and Methods

40 patients (20 with an indeterminate PN and 20 with pulmonary metastases) underwent two Low Dose CT scans (LDCT) within a 60-minute period (the “Coffee-break”). Patients were randomized to two inspiratory control groups: normal breath hold, and controlled lung volume.

20 texture features were extracted from each automatic contoured region surrounding the PN.

The variability and limits of repeatability of texture measurements within individual nodules was assessed.

Results

The mean difference in volume between the two scans was 6.3%, SD: 29.9%.

The 20 textural features displayed 95% CI below $\pm 17.8\%$, and were less variable than nodule volume (95% CI $\pm 28.9\%$). All features had high repeatability, calculated by the concordance correlation coefficient, ($0.84 \leq CCC \leq 0.99$).

All measurements were more repeatable for the controlled lung volume group than the normal breath-hold group.

Conclusion

CTTA repeatability was comparable to automatic volumetric measurements, and appears to be improved using controlled volume breath holding.

Advances in Knowledge

- CT texture analysis (CTTA) is a potential imaging biomarker of malignancy in many tissues and organs, including pulmonary nodules (PNs).
- This study demonstrates that CTTA is a highly repeatable measure, comparable to nodule volumetry.
- The repeatability of nodule volumetry and CTTA improves with controlled volume breathing.

Implications for Patient Care (*up to 3 bullet points*)

- Alongside nodule volumetry, CTTA has the potential to assist in the diagnosis of PNs.
- The routine use of controlled volume breathing whilst performing LDCT for patients undergoing nodule surveillance and in lung cancer screening programmes should be considered.

Introduction

Current methods of determining if pulmonary nodules (PNs) are benign or malignant are not standardised. The US National Lung Screening Trial (NLST) showed that up to 95% of lung nodules detected on CT scans of the chest were falsely positive for malignancy¹. Detection of non-malignant nodules has the unwanted consequences of unnecessary cost, additional investigations, patient anxiety and increased morbidity. Three studies have reported generic health-related quality of life, anxiety and lung-cancer specific distress data from approximately 2,500 screening participants^{2,3,4}.

At present for small, sub-centimetre indeterminate nodules, follow up scanning to detect growth as a surrogate biomarker for malignancy is the most common recommendation⁵. Prior studies have shown that there is significant variation of up to 25% in these measurements⁶. This has been attributed to patient position, heart pulsation and inspiration levels, which can all influence the assessment of nodule size⁷. Consequently, a change in volume of 25 or 30% (at separate time intervals) has been suggested as a way of avoiding false positives.

Recent studies have shown that incorporating nodule characteristics such as size, growth rate, contrast enhancement and CT texture analysis (CTTA) can improve the accuracy of predicting the risk of malignancy⁸. This allows the stratification of PNs into different investigations and/or follow-up pathways based on the predicted risk of malignancy.

CT Texture Analysis (CTTA) represents a novel imaging approach that can potentially be used as an imaging biomarker in tumour characterization. Texture analysis refers to a variety of mathematical methods that provide information about the spatial arrangement of gray scales or intensities of the pixels within an image, producing measures of tissue heterogeneity, irregularity of

shape and border and surface smoothness as well as many other features that may not be appreciated by the naked eye.

Conventionally, CT Chest examinations are performed in full inspiration after appropriate instructions are given to the patients. However this results in inconsistent lung volume acquisition, which can alter nodule volume and position. Its effects on CTTA of PNs are not known. A method of reducing this variability in lung volumes is by asking the patients to breathe in a set volume of air.

The primary objective of this study was to prospectively assess the repeatability of CTTA of PNs at low dose multi-detector row computed tomography (CT) and to compare this to the repeatability of nodule volume. The secondary objective was to investigate the influence of two different methods in inspiratory control on CTTA and nodule volume measurements.

The repeatability of CTTA for each nodule for the risk of malignancy was also analysed.

Materials and Methods

Patient and Nodule Selection

The study was approved by the local Research Ethics Committee and written informed consent was obtained from all patients.

Between March and July 2016, 40 adult patients (24 male, 16 female, age range, 43-88 years, mean age 69 years) with PN(s) between 5-20mm were enrolled into the study.

20 patients had known pulmonary metastases previously shown on chest CT. The majority of these patients (n=17) were referred for chest CT or Positron Emission Tomography (PET) scans as part of cancer surveillance or for baseline chest CT (n=3) before the start of anticancer therapy.

The underlying primary cancers were sarcoma (n= 15) and colorectal cancer (n=5).

The other group of 20 patients was enrolled from the virtual pulmonary nodule clinic. This group had indeterminate PN(s) and were referred for CT surveillance. 3/20 of these patients were subsequently diagnosed with non-small cell lung cancer. 7/20 of the patients had completed two years of nodule surveillance and were diagnosed with a benign nodule based on stability. 10/20 of the patients are still undergoing nodule surveillance to date (March 2017).

Only solid, non-calcified nodules were included in both groups because the software used for nodule volumetry and CTTA was not designed for the analysis of sub-solid nodules. Juxta-pleural, juxta-cardiac and nodules abutting blood vessels were included.

Each patient had one nodule identified that was either under surveillance if indeterminate at the time or the index pulmonary metastasis previously chosen for oncology follow-up.

Image Acquisition

Two low-dose chest CT examinations (LDCT) without contrast material were performed in all cases. Between the two LDCT examinations, patients were asked to get off and on the CT table to simulate the conditions of a repeat examination, with new patient positioning and a new scan localiser performed.

For patients with known pulmonary metastases, a contrast material enhanced standard chest CT examination or PET scan then followed the two LDCT examinations for clinical purposes. For patients with an indeterminate nodule, one additional LDCT examination was performed, with the first scan used for clinical purposes.

All CT scans were acquired using a multi-slice CT scanner GE Lightspeed VCT 64-slice (General Electric Medical Systems, Milwaukee, Wisconsin, USA) at a section thickness of 0.625 mm and reconstructed with a CHEST convolution kernel.

Exposure settings for one additional LDCT examination were 2 millisieverts.

Inspiration

In order to investigate the influence of two different methods of inspiratory control, patients were randomized to two groups. 20 patients underwent both CT scans with a normal breath hold (as per routine clinical practice⁹; Group A) and 20 underwent both scans with controlled lung volumes (Group B). Patients in Group B were asked to breathe from a pre-filled reservoir bag containing 1L of air whilst lying down prior to the scan. Scanning was then performed during a breath hold for approximately 10 seconds.

Nodule contouring and volume measurements

One PN per patient was identified and marked-up on the first LDCT examination (baseline scan) by a single observer (A.T., two years experience in radiology and trained for this specific task) using commercially available software (XD3, Mirada Medical Ltd., UK). Each PN was identified in the second LDCT examination (validation scan) with knowledge of findings from the first study, and matched by using a combination of section number, lung segment and distance to the pleura. All PNs were delineated using a proprietary semi-automatic Otsu based thresholding method that requires a single user click-point inside the nodule to initialize the segmentation (Figure 1).

Texture Analysis

A number of image signatures previously used in combination to predict nodule probability of malignancy¹⁰ were extracted from each contoured region surrounding the PN (see Figure 2). A machine-learning model that maps image signatures onto a probability of malignancy score (ranging from 0 to 100%) was applied to the baseline and validation scans.

In previously published work, a large number of textural features were computed on an independent dataset⁵, including filter-based Laws features¹¹, Laplacian of Gaussian (LoG)¹², features derived from the grey-level co-occurrence matrix (GLCM)¹³ the Fractal Dimension¹⁴ and First Order Statistics derived from grey-level intensity distributions. The 20 most discriminative features in combination were selected using a greedy-algorithm on this independent dataset. A Support Vector Regressor (SVR) and 5-fold cross-validation technique was used to train a model which maps the selected features onto a probability of malignancy score (ranging from 0 to 100%). The subset of 20 features and the probability of malignancy score was computed for each PN on the baseline and validation scans.

Image analysis tasks were performed using MATLAB 2015a (The Mathworks, Natick, MA, USA).

Statistical Evaluation

Repeatability of volume measurements, and textural features were assessed by computing relative differences (%) between baseline and validation scans:

$$100 \times (\text{feat}_{\text{validation}} - \text{feat}_{\text{baseline}}) / [(\text{feat}_{\text{baseline}} + \text{feat}_{\text{validation}}) \times 0.5].$$

Mean and SD (standard-deviation) differences were estimated from the relative differences. Lower and Upper Limits of Repeatability (LLR & ULR, respectively) which defined the 95% confidence interval (CI) of normal fluctuation in feature measurements were calculated as $\text{mean}_{\text{diff}} \pm 1.96 \times \text{SD}_{\text{diff}}$. The repeatability of the probability of malignancy score (scaled between 0 and 100%) was assessed by computing differences between baseline and validation scans: $(\text{feat}_{\text{validation}} - \text{feat}_{\text{baseline}})$. Limits of repeatability were defined as before. Concordance correlation coefficient (CCC)¹⁵ was used to assess the repeatability of the image features for this group of patients.

Statistical analyses and data visualization were generated using R (version 3.2.1)¹⁶ in RStudio (version 0.99.467) with the ggplot2 library¹⁷.

Results

38 of 40 cases of non-calcified solid intraparenchymal PNs were analysed. Two cases were excluded following quality control of the scans. The first contained a nodule attached to a vessel for which the shared boundary was undefined. The second was a nodule for which the baseline and validation scans were reconstructed with different reconstruction protocols (CHEST and SOFT tissue kernels, respectively). Of the 38 PNs, 16 were located in the lung parenchyma, 9 were juxta-pleural, 12 juxta-vessel and 1 was perifissural.

Repeatability of Volume Measurements

Nodule size ranged from 76 to 2090 mm³ (mean: 439 mm³, SD: 445 mm³; longest diameter in 2D from 5 to 17mm), with the exception of one case, which had a nodule size of 8130 mm³ (longest diameter in 2D: 20mm) due to nodule growth between the time the patient was recruited to the study and scanned.

There was no statistically significant difference in nodule volume ($p = 0.20$, Mann-Whitney U test) between the patients who underwent conventional breath-hold (Group A) and those who inhaled a set volume of air (Group B).

Repeatability of volume measurements measured using Bland-Altman was high with 95% confidence intervals below 30% (mean: -2.1% SD: 13.7%; ULR: 24.7%, LLR: -28.9%) and a concordance correlation coefficient (CCC) of 0.99 across the 38 PNs. Slightly lower variability in volume measurements was observed for nodules less than 500 mm³ (mean: 0.6% SD: 12.4%; ULR: 23.7%, LLR: -24.8%) whereas the CCC was unchanged (0.98).

Volumetric measurements were more repeatable in Group B (ULR: 14.0%, LLR: -13.3%; CCC = 0.99) than in Group A (ULR: 31.3%, LLR: -40.8%; CCC = 0.95).

There was no difference in volumetric repeatability between nodules abutting the pleura or vessels (95% Bland-Altman CI: 28.0%) and those located in the lung parenchyma (95% Bland-Altman CI: 26.3%).

Repeatability of Texture Features

The 20 textural features displayed ULR and LLR below $\pm 17.8\%$ and had high repeatability as measured using the CCC ($0.84 \leq \text{CCC} \leq 0.99$) (Fig. 3, suppl. Fig. 1). The repeatability of individual features was greater in Group B (Bland-Altman 95% CI $\leq \pm 11.8\%$; $0.90 \leq \text{CCC} \leq 0.99$) than in Group A (Bland-Altman 95% CI $\leq \pm 26.6\%$; $0.67 \leq \text{CCC} \leq 0.99$) (suppl. Table 1).

Repeatability of malignancy prediction scores

The malignancy prediction score was highly repeatable (ULR & LLR $\leq \pm 8.7\%$; CCC = 0.96). Limits of repeatability were marginally tighter for Group B (ULR & LLR $\leq \pm 7.6\%$; CCC = 0.94) than for Group A (ULR & LLR $\leq \pm 10.0\%$; CCC = 0.94).

Discussion

CT Texture analysis has previously been shown to be a promising imaging biomarker of malignancy both on its own⁶ and when used in conjunction with other nodule characteristics¹⁸.

To our knowledge, this is the first study to assess the repeatability of CTTA used in combination to predict PN malignancy and to examine the effect of different inspiratory techniques on CTTA.

The results demonstrate that the repeatability of CTTA was comparable to automatic volumetric measurements that are currently recommended for use in clinical practice.

The volumetric repeatability ranged from -28.9% to $+24.7\%$ (Bland-Altman, 95% CI) in 38 PNs (CCC = 0.98). Similar to Gietema et al³; those limits fell below $\pm 25\%$ for nodules less than 500 mm³. Wormanns et al¹⁹ also reported the limits of repeatability on Bland-Altman analysis of -20.4% to 21.9% , although this study computed Volume Measurement Error in such a way that it corresponds to half the Relative Difference. This means that at the 95% CI, the limits of repeatability were equivalent to -40.8% to $+43.8\%$.

Interestingly, we found no difference in volume repeatability between nodules abutting the pleura, vessels or fissures and nodules located in the lung parenchyma, which has been a feature of some prior reports.

The repeatability of textural features was high with a test-retest variability below $\pm 18\%$ (Bland-Altman, 95% CI) and concordance correlation coefficients in agreement with the findings of a recent study¹⁵ (CCC ≥ 0.85 for a large proportion of textural features) for similar CT acquisition and reconstruction protocols. When combined in a model to predict nodule malignancy, texture analysis yielded a highly repeatable score (ULR & LLR $\leq \pm 8.7\%$; CCC = 0.96).

Our results indicate that controlling the volume of air inhaled by patients during CT examination improves the repeatability of both volume and texture analysis measurements. This intervention is readily applicable and may be worth considering for routine use for patients in lung cancer screening programmes.

There are a number of limitations to our study. Firstly, only 40 patients with 40 PNs were recruited, although the volume results are similar to other published volume analysis studies, suggesting that the sample and data generated is representative. Secondly, only one set of reconstruction parameters was used in this study (0.625mm and CHEST kernel). However, Zhao et al¹⁵ found that textural features were repeatable across a wide range of CT imaging parameter settings used in

clinical care (slice thicknesses ranging from 1.25mm to 5mm). Thinner slice thicknesses and smoother reconstruction kernels were associated with higher repeatability. Additionally, as with other studies such as Kostis et al ⁴, SSNs were not included due to the segmentation method used.

Conclusion

CTTA is a highly repeatable measure, and appears at least as reproducible to volume analysis. The use of controlled volume breathing appears to increase reproducibility. Whether CTTA is a more reliable predictor of malignancy than growth, or will be additive to the information provided by growth over time requires further research.

FIGURES

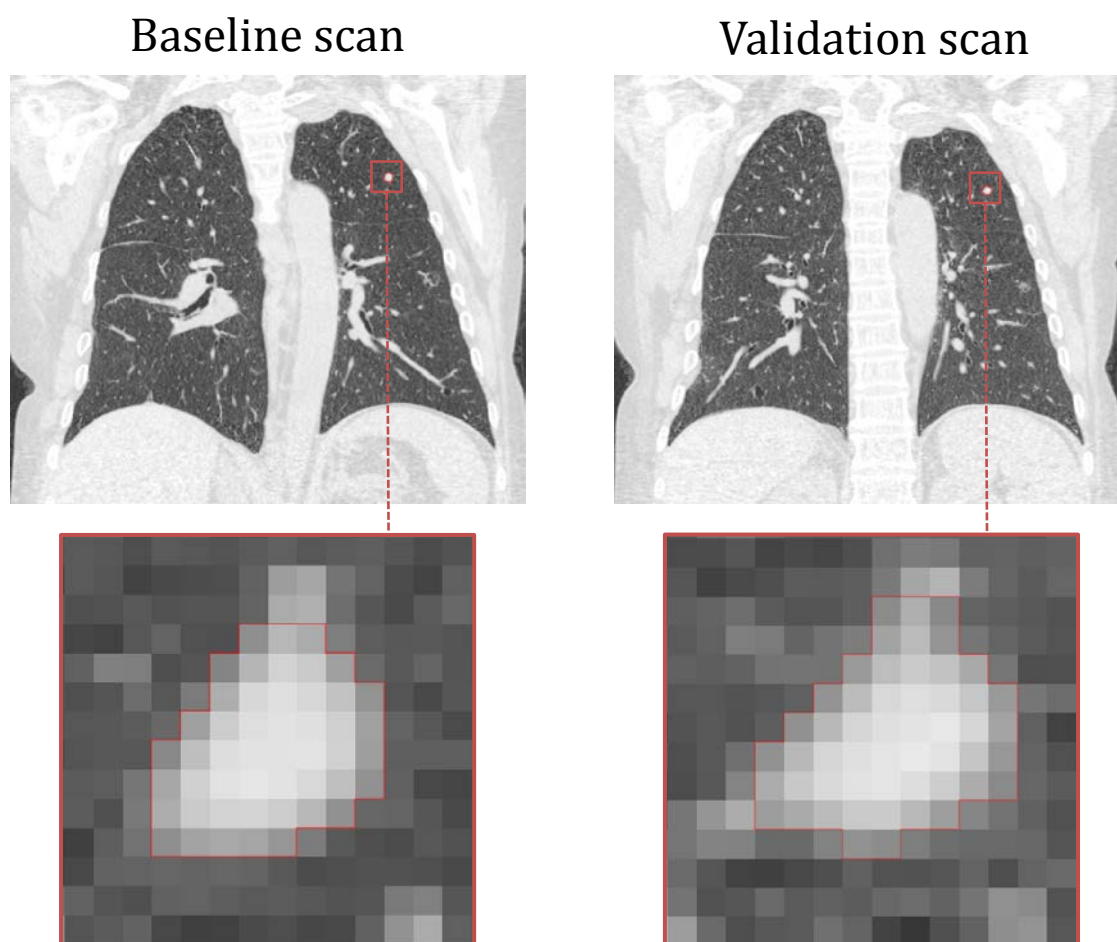


Figure 1: Example of pulmonary nodule on baseline and validation scans. The PN was delineated using a semi-automatic Otsu based thresholding method.

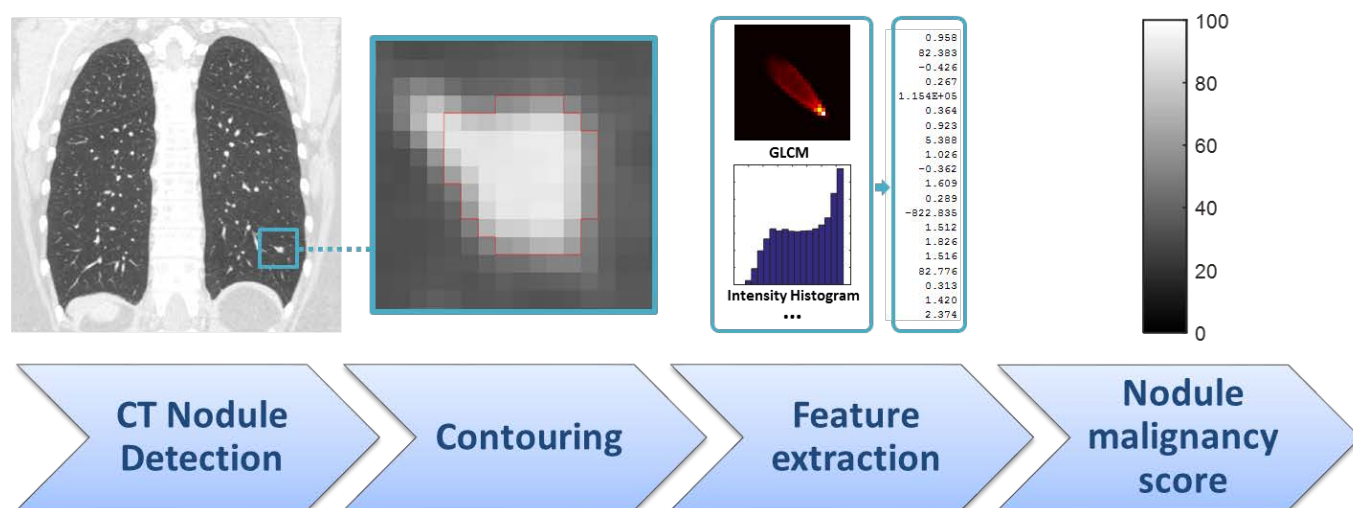


Figure 2: Quantitative Texture Analysis. Computer Aided Diagnosis (CADx) pipeline

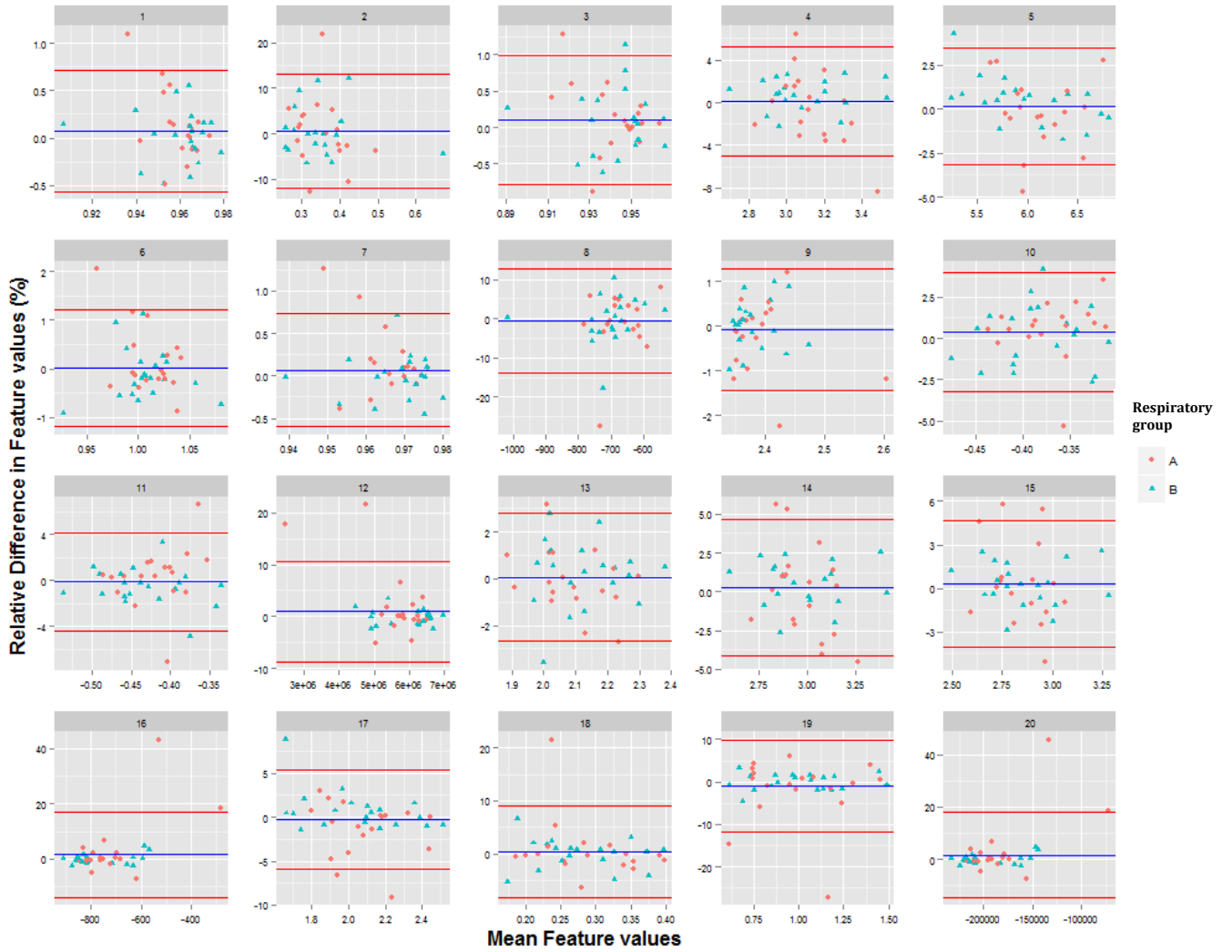
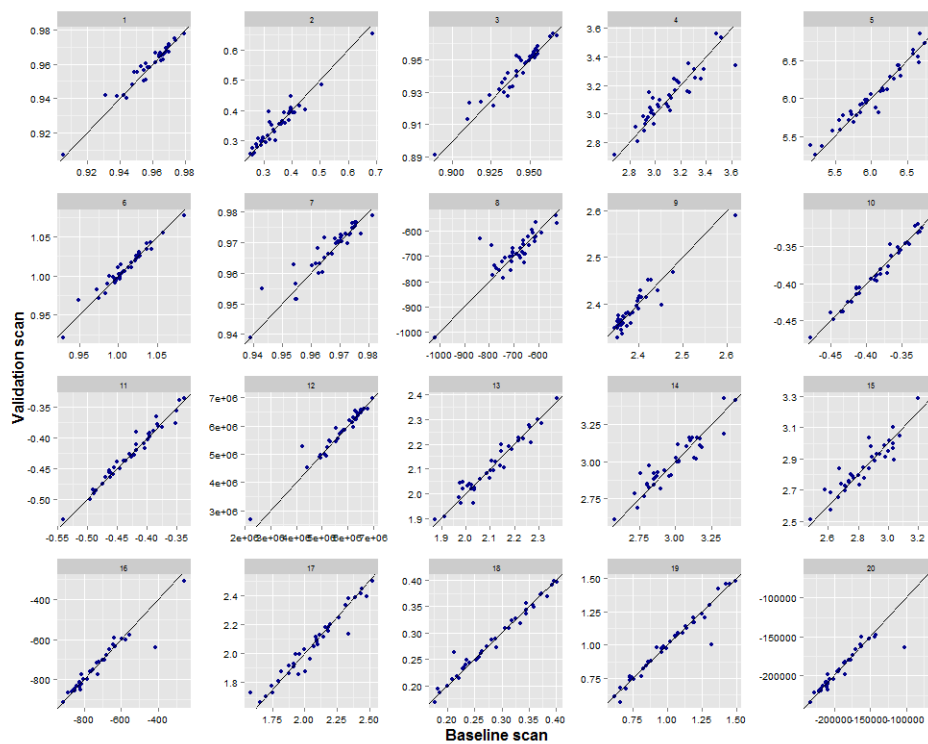


Figure 3: Bland-Altman plots for the 20 textural features. Relative differences (%) (y-axis) vs. average feature values (x-axis) across baseline and validation scans. Mean diff. (blue line), ULR and LLR (red lines) computed across all 38 cases were displayed. Data-points for the two different methods of inspiratory control, group A and group B, are shown in red and green respectively.

SUPPLEMENTARY MATERIAL



Suppl. Fig. 1: Scatterplots for the 20 textural features measured on baseline (x-axis) and validation (y-axis) scans. Black line is the line of unity.

Suppl. Table 1: Limits of repeatability ($95\%CI = 1.96 \times SD_{diff}$) and CCC for the 20 textural features.

Feature index	BA 95% CI			CCC		
	All	Group A	Group B	All	Group A	Group B
1	0.6	0.7	0.5	0.97	0.93	0.99
2	12.6	14.7	10.7	0.96	0.91	0.98
3	0.9	0.9	0.9	0.96	0.95	0.97
4	5.1	6.7	2.8	0.91	0.78	0.97
5	3.3	3.9	2.6	0.97	0.93	0.99
6	1.2	1.4	1.0	0.98	0.96	0.99
7	0.7	0.8	0.5	0.93	0.85	0.97
8	13.2	15.3	11.5	0.84	0.67	0.90
9	1.4	1.6	1.1	0.94	0.94	0.94
10	3.6	3.5	3.7	0.99	0.99	0.99
11	4.3	5.2	3.2	0.98	0.96	0.99
12	9.8	13.6	3.0	0.97	0.95	0.99
13	2.7	2.7	2.9	0.97	0.97	0.97
14	4.4	5.7	2.8	0.93	0.83	0.98
15	4.3	5.6	3.0	0.93	0.83	0.98
16	15.5	22.0	3.6	0.95	0.91	0.99
17	5.7	6.2	4.4	0.97	0.93	0.99
18	8.6	11.2	5.6	0.99	0.98	0.99
19	10.9	15.3	3.8	0.97	0.94	1.00
20	16.3	23.1	3.6	0.94	0.90	0.99

References :

-
- ¹ The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. *New England Journal of Medicine*, 2011; 365(5):395–409. doi:10.1056/NEJMoa1102873
- ² van den Bergh KA, Essink-Bot ML, Borsboom GJ, et al. Long-term effects of lung cancer computed tomography screening on health-related quality of life: the NELSON trial. *Eur Respir J*, 2011; 38:154-61
- ³ Vierikko T, Kivisto S, Jarvenpaa R, et al. Psychological impact of computed tomography screening for lung cancer and occupational pulmonary disease among asbestos-exposed workers. *Eur J Cancer Prev*, 2009;18:203-6
- ⁴ Byrne MM, Weissfeld J, Roberts MS. Anxiety, fear of cancer, and perceived risk of cancer following lung cancer screening. *Med Decis Making* 2008;28:917-25
- ⁵ Callister ME, Baldwin DR, Akram AR, et al. British Thoracic Society guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015;70(Suppl 2):ii1–ii54. [Published correction appears in *Thorax* 2015;70(12):1188.]
- ⁶ Gietema HA, Schaefer-Prokop CM, Mali WPTM et al. Pulmonary Nodules: Interscan Variability of Semiautomated Volume Measurements with Multisection CT— Influence of Inspiration Level, Nodule Size, and Segmentation Performance. *Radiology* 2007; 245:888-894
- ⁷ Kostis WJ, Yankelevitz DF, Reeves AP et al. Small pulmonary nodules: reproducibility of three dimensional volumetric measurement and estimation of time to follow-up CT. *Radiology* 2004 231 (2):446-452
- ⁸ Ganeshan B, Miles A. Quantifying tumour heterogeneity with CT. *Cancer Imaging* 2013 13 (1):140-149
- ⁹ American College of Radiology Practice Guideline for the Performance of Thoracic Computed Tomography. 2008
- ¹⁰ Willaime JMY, Pickup LC, Boukerroui D et al. Impact of segmentation techniques on the performance of a CT texture-based lung nodule classification system, poster no. B-0436, ECR 2016, <http://dx.doi.org/10.1594/ecr2016/B-0436>
- ¹¹ Laws, K.I. Rapid Texture Identification. in *Image Processing for Missile Guidance*. 1980. San Diego: SPIE.
- ¹² Ganeshan B, Skogen K, Pressney I et al., Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: preliminary evidence of an association with tumour metabolism, stage, and survival. *Clin Radiol*, 2012. 67(2): p. 157-64.
- ¹³ Haralick R M, Shanmugam K and Dinstein I H. Textural features for image classification *IEEE Trans. Syst. Man Cybern* 1973 3: 610–21
- ¹⁴ B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, Oxford, 1982)
- ¹⁵ Lin, L. I. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989 45, 255–268
- ¹⁶ R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2017 URL <http://www.R-project.org/>.
- ¹⁷ H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009
- ¹⁸ Sun T, Zhang R, Wang J et al. Computer-Aided Diagnosis for Early-Stage Lung Cancer Based on Longitudinal and Balanced Data. *PLoS ONE* 2013 8(5): e63559. doi:10.1371/journal.pone.0063559
- ¹⁹ Wormanns D, Beyer F, Diederich S et al. Diagnostic performance of a commercially available computer-aided diagnosis system for automatic detection of pulmonary nodules: comparison with single and double reading. *Rofo* 2004;176:953–958.