CT texture analysis can help differentiate between malignant and benign lymph nodes in the mediastinum in patients suspected for lung cancer

Abstract

Background

In patients with Non-Small-Cell Lung Carcinoma NSCLC the lymph node staging in the mediastinum is important due to impact on management and prognosis. Computed tomography texture analysis (CTTA) is a post processing technique that can evaluate the heterogeneity of marked regions in images.

Purpose

To evaluate if CTTA can differentiate between malignant and benign lymph nodes in a cohort of patients with suspected lung cancer.

Material and Methods

With tissue sampling as reference standard, 46 lymph nodes from 29 patients were analyzed using CTTA. For each lymph node, CTTA was performed using a research software "TexRAD" by drawing region of interest (ROI)on all available axial contrast enhanced computed tomography (CT) slices covering the entire volume of the lymph node. Lymph node CTTA comprised image filtration-histogram analysis undertakes 2 stages: First step comprised an application of a Laplacian of Gaussian filter to highlight fine to coarse textures within the ROI, followed by a quantification of textures via histogram analysis using mean grey-level intensity from the entire volume of the lymph nodes.

Results

CTTA demonstrated a statistically significant difference between the malignant and the benign lymph nodes (p=0,001), and by binary logistic regression we obtained a sensitivity of 53% and specificity of 97% in the test population. The area under the Receiver Operating Curve was 83.4% and reproducibility was excellent.

Conclusion

CTTA may be helpful in differentiating between malignant and benign lymph nodes in the mediastinum in patients suspected for lung cancer, with a low intraobserver variance.

Key words: computed tomography, computer applications-detection/diagnosis, thorax, lymphatic, mediastinum.

Introduction

In patients with suspected lung cancer contrast enhanced computed tomography (CT) and/or Fluorine-18 Fluoro-D-Glucose positron emission tomography/computed tomography (F-18-FDG PET/CT) are standard procedures in the identification, characterization and staging of the disease. CT relies on size and contrast enhancement, and PET/CT on the metabolism of F-18-FDG. Studies have shown, that although CT and F-18-FDG PET/CT both improve the diagnostic accuracy, invasive procedures are still mandatory to confirm/exclude malignancy (1,2). Improving the diagnostic accuracy of imaging modalities to identify malignant involvement of the mediastinal lymph nodes in a patient with a suspected lung nodule/mass is important in the diagnostic work-up when therapy and prognosis is considered (3–5). This will further assist to minimize the number of invasive procedures required to confirm the presence of malignancy within lymph nodes and therefore the complications associated with the invasive procedures.

Contrast enhanced CT texture analysis (CTTA), a new post processing technique, has been successfully applied in different oncological applications, to non-invasively quantify heterogeneity within different tumors and potentially leading to improved prognosis, disease risk-stratification/characterization and treatment-response/prediction in a number of cancer sites e.g. non-small cell lung cancer (NSCLC) (6,7), esophageal (8), colorectal (primary and metastatic liver)(9), head and neck (10) and metastatic renal cell cancer (11). Additional texture features on CT seem to be associated with tumor biology (providing a biological-correlate) such as grade, physiology (perfusion and glucose uptake), hypoxia, angiogenesis and genetic mutation in different tumors such as gliomas, colorectal, NSCLC and esophageal cancer (8,9,12).

The purpose of this study was to evaluate if CTTA can differentiate between malignant and benign lymph nodes in the mediastinum in patients suspected to have lung cancer with tissue sampling as gold standard.

Material and Methods

The study conformed to Danish legal requirements. As all subjects received best patient care, Institutional Review Board approval was waived (Journal number: 1-15-0-72-2-09).

Patients

During a 15- month study period, 840 patients with suspected lung cancer were referred to a tertiary hospital. All patients had a CT performed and subjected to a multidisciplinary team (MDT) decision. When indeterminate lesions or malignancy suspect lesions were present on CT, the patients received an additional F-18-FDG PET/CT. 168 patients received both CT and an F-18-FDG PET/CT during the study period. Among the patients 32 had non- NSCLC and 29 patients were excluded from analysis because the CT scan was performed at other institutions using other CT scanners and contrast media policy. Thus, the study population consisted of 107 patients. A potential number of 1926 lymph node stations were available for analysis.

The inclusion criteria for a lymph node was: 1) a long axis diameter of > 1 cm as to reliable get some better statistics in terms of more number of voxels available within the lymph node for CTTA calculations, 2) The largest lymph node within each lymph node station was selected, 3) Lymph nodes with large vessels going through was excluded as such vessels may disturb the texture calculations 4) Lymph nodes distorted by artifacts was excluded as such artifacts may also disturb texture calculations and 5) tissue sampling was performed either by surgery or mediastinoscopy.

1788 lymph node stations did not contain a lymph node with a long axis diameter of > 1 cm; 38 lymph node stations were hampered by CT artifacts; 30 lymph node stations were in patients with distant metastases so no histopathology was obtained, 17 lymph node stations were out of the range of mediastinoscopy or surgery and in 7 lymph node stations tissue sampling was not performed for unknown reasons. So 46 lymph nodes in 29 patients were appropriate for analysis (Fig.1).

Contrast enhanced CT

CT examinations included the chest and the upper abdomen. CT was performed with a multiple-row detector CT scanner (Philips Brilliance CT 64-channel scanner; Philips Healthcare, Best, The Netherlands). CT acquisition parameters were 64 x 0.625 mm collimation, kV 120-140, mAs/slice 150-250, rotation time 0.75, reconstruction thickness 2 mm, increment 1 mm, pitch 1.078, FOV 35 cm and matrix 512 x 512. CT examinations included the chest and the upper abdomen.. Iodixanol 270 mg/ml (Visipaque® 270; GE Healthcare, Oslo, Norway), or iohexol 300 mg/ml (Omnipaque® 300; GE Healthcare), was injected intravenously in weight-adjusted doses of 2 ml/kg body weight to compensate for differences in distribution volume. A bolus tracking technique with a ROI in the descending aorta on the level of Carina was used to compensate for differences in cardiac output. CT was performed after a delay of 15 seconds for the chest and upper abdomen (late arterial phase), and 65 seconds for the upper abdomen (portovenous phase), and raw picture data sets were transferred to a Philips Extended Brilliance™ Workspace workstation v4.02 for further analysis (1,2).

Reference standard

Tissue sampling was obtained either by surgery with complete lymph node resection, by mediastinoscopy with sampling from nodal stations 2R/L, 4R/L and 7 or by anterior

mediastinotomy from stations 3A, 5 and 6. Among the 46 lymph nodes included in the study tissue sampling was obtained by surgery in 74% (34/46) and by mediastinoscopy/-tomy in 26% (12/46).

CTTA of lymph nodes

CTTA was performed using TexRAD, a research software (TexRAD Ltd, , Cambridge, UK) The technique comprised an image filtration-histogram approach; in which an initial filtration step using a Laplacian of Gaussian band-pass filter (non-orthogonal Wavelet) was employed to extract and enhance features of different size based on the spatial scale filter (SSF) values varying from 1.4-4.1 mm in diameter (Fig. 2.) Following image filtration, each filtered image texture map was quantified using histogram parameters such as mean image intensity.

Two resident radiologists performed the analysis (over an interval of 3-6 months for the assessment of intra- and inter observer variance). The radiologists were blinded to all patient identifiers and clinical data as well as results from CT, F-18-FDG PET/CT and tissue sampling. For each lymph node a Region of interest (ROI) was drawn on all axial images. For conglomerates the entire structure was enclosed within a ROI. A semi-automated approach further refined the ROI enclosing the lymph node to exclude air and fat a thresholding procedure that removed pixels having value <-50 HU.

Statistical analysis

Mean grey-level intensity for all filter sizes was analyzed with the Shapiro-Wilk's test as well as QQ-plots to determine if they were Gaussian distributed. All but the unfiltered mean image intensity showed a Gaussian distribution. Using the following transformation: $\sqrt[2]{(x_{max}+1)-x}$, where x = mean grey-level intensity and x_{max} = maximum mean grey-level intensity observed across

all lesions. After transformation the unfiltered mean grey level intensity approached Gaussian distribution for both the malignant and benign group.

The independent samples t-test was used to compare means between malignant and benign groups. In case of unequal variances determined by Levener's test an unequal variance t-test would be performed instead (Welch's t-test)

The variables that proved a statistically significantly difference between malignant and benign groups was used in a binary logistic regression model to obtain the 2x2 table used for calculating sensitivity, specificity and accuracy. Furthermore, an overlay plot of sensitivity/specificity was performed to obtain the cutoff value giving the highest overall accuracy as well as to obtain the cutoff providing the highest specificity without losing sensitivity in order to avoid potential future patients to be deprived of possibly curative surgery. In case the two cutoff values were not identical the latter was used for analysis.

To assess the method a Receiver Operating Curve (ROC) was created using the mentioned cutoff value.

Comparison of inter- and intraobserver variance was calculated using limits of agreement by a Bland-Altman plot (13).

Results

Lymph node morphology

Among the 46 lymph nodes available for analysis 15 was malignant and 31 benign evidenced by tissue sampling. The 15 malignant lymph nodes had a mean short axis diameter of 9.9 mm (7.0-17.9 mm) and a mean long axis diameter of 16.3 mm (10.1-21.6 mm) while the 31 benign lymph nodes

had a mean short axis diameter of 9.8 mm (6.5-14.4 mm) and a mean long axis diameter of 16.7 mm (10.1-38.6 mm)

After transformation using the natural logarithm both the short axis diameter and long axis diameter were normally distributed for both malignant and benign lymph nodes as assessed by Shapiro-Wilk's test (p>0,5) as well as by QQ-plots.

There was no statistical significant difference for short axis (p=0.882) and long axis diameter (p=0.916) between the malignant and the benign groups as assessed by the independent t-test.

A short axis of 10 mm as a compromise is often considered as the threshold for pathologic lymph nodes in the mediastinum (14). In this study a correct identification of only 7 of the 15 malignant lymph nodes and 19 of the 31 benign lymph nodes were obtained when a short axis larger than 10 mm was considered giving a sensitivity of 46.7% and specificity of 61.3%.

Malignant vs. benign lymph node textures

Mean grey-level intensity for all filter levels were normally distributed for both malignant and benign lymph nodes as assessed by Shapiro-Wilk's test (p>0.5) as well as visually by QQ-plots. This was obtained after transformation of the unfiltered mean image grey-level data as described in the statistical analysis section.

Only unfiltered mean image intensity (Mean HU) showed a statistically significant difference between the malignant and the benign group. The lymph nodes in the malignant group showed significantly higher mean image intensity than the benign group. Mean difference was 1.88(95%CI, 0.88 to 2.88), as Levener's test showed unequal variances (p=0.029) the results presented are from the unequal variance t-test (Welch's t-test) performed on the transformed data: $t(19.556)=3.939$, p=0.001. Mean image intensity at the coarse filtered texture scale (SSF = 3.4 mm) showed a trend

to differentiate between the malignant and benign group but did not reach statistical significance. Mean difference was -6.59(95%CI, -13.62 to 0.45) and the result of the independent t-test was t(44)=-1.888, p = 0.066 for filter size 3.4 mm. No other texture parameters showed significant differences between the two groups.

In Table 1 the distribution characteristics and p-values for the texture mean image intensity is shown and Fig. 3 shows the box and whisker plot for the unfiltered mean image intensity. Three outliers were found, however, when performing both a Mann-Whitney's U test and the unequal variance t-test they both gave roughly the same result, hence the outliers were included in the analysis.

Binary logistic regression analysis and ROC curve

The binary logistic regression analysis on the unfiltered mean image intensity provided a classification table for the CTTA as well as a way to obtain cut off values for future studies. In order to determine the optimal cut off value for both sensitivity and specificity we performed an overlay plot of sensitivity and specificity at various probability cut offs (Fig 4).

The logistic regression was performed to assess the effects of unfiltered mean values on the likelihood of a lymph node to be malignant. We used a cut off of 0,6 instead of the intersection value to obtain the highest value of specificity without losing too much sensitivity to avoid patients to be deprived of possible curative surgery (Table 2). The logistic regression model was statistically significant, $\chi^2(1) = 18.014$, $p < 0.0005$. The model demonstrated 45.2% (Nagelkerke R^2) of the variance in the malignant lymph nodes and classified 82.6% of the cases correctly. Sensitivity = 53.3%, Specificity = 96.8%, Positive predictive value = 88.9% and Negative predictive value = 81.1%

Furthermore, a ROC curve was derived from the data provided by the logistic regression analysis which gave us an area under the curve of 83.4% with a 95% confidence interval of 70.9% to 96.0% (Fig 5).

Intra- and interobserver variance

The median ratio of reproducibility for intraobserver variance was essentially zero. Thus, no systematic bias was observed between the two readings. Furthermore, the limits of agreement were narrow. The mean difference between malignant and benign lymph nodes was 19.97 HU units and the limits of agreement were between -3.94 and 3.72 HU units. The variability was roughly consistent across the graph (Fig. 6.).

However, with regard to inter-observer variance a systematic bias was observed, as the mean of the difference was 7.19 HU units. Furthermore, the limits of agreement were wide (Fig. 7.).

Discussion

In a previous study (2,15), both CT as well as F-18-FDG PET/CT performed equally poor when mediastinal disease in lung cancer patients were considered Furthermore, in the presented study, the mean size of the malignant lymph nodes were not significantly larger than the benign. Of 46 lymph nodes 27 were not pathologically enlarged on CT (larger than 10 mm in the short axis) (14) . In RECIST 1.1 (16) the short axis needs to be larger than 15 mm before a lymph node is considered pathologic. F-18-FDG PET/CT has in several studies demonstrated difficulties when assessing small lesions (1,2,17), so most likely many of the lymph nodes in our study would be below the detection limit for F-18-FDG PET/CT

Consequently, at most institutions even non-enlarged lymph nodes and nodes without significant uptake of FDG go for Endobronchial Ultrasound/Endoscopic Ultrasound (EBUS/EUS).

Unfortunately, this study was performed at a time where mediastinoscopy was standard procedure at our institution. Although EBUS/EUS can add to the diagnostic accuracy in these small lesions the accuracy is not 100% (18,19). In a recent report (20) the authors also found that, texture analysis may be complementary in the CT evaluation of mediastinal lymph nodes in lung cancer patients. In that study, non-enhanced CT scans were considered. So although from a theoretical point of view differences in heterogeneity is expected to be augmented in a contrast enhanced phase it also seems that a non-enhanced CT can display important heterogeneity features. Further studies comparing the impact of texture features in different contrast media phases is warranted.

In the present study, we found 8 malignant non-enlarged lymph nodes and 12 enlarged benign lymph nodes in accordance to the gold standard. Additional texture analysis found 5 malignant non-enlarged lymph nodes and benignancy in 11 enlarged lymph nodes. Thus in this study with 33% prevalence of malignancy we had a sensitivity of 53.3% and a specificity of 96.8%. Thus, on a per lesion basis CTTA and CT is supplementary when considering both malignancy and benignancy.

The unfiltered mean image grey level showed a significant difference between malignant and benign lesions. However, when applying different filter sizes it showed a strong tendency for a filter size of 3.4 mm ($p = 0.066$) although non-significant to distinguish between malignant and benign lymph nodes. Probably the low overall number of lymph nodes analyzed as well as the small size of lymph nodes is the reason for the non-significant results. CTTA has turned out to be a useful post-processing modality and have been used for prognosis, disease risk-stratification/characterization and treatment-response/prediction, as well as been correlated with physiology (perfusion and glucose uptake), hypoxia, angiogenesis and genetic mutation in different tumors (6–9,11,12).

CTTA is highly reproducible when intra-observer variance is considered. However, a systematic bias was introduced when looking at inter-observer variance. The second reader concistently had

higher values than the first reader. This may reflect inclusion of nearby calcifications or vessel walls. Despite this bias, it had no impact on the differtion between malignant and benign lymph nodes.

46 lymph nodes were biopsied and found suitable for analysis, of these 67% were benign. As the prevalence of lung cancer was 80% in patients that had both FDG-PET/CT and CT, and as metastatic disease in patients with malignant lymph nodes often was present in this cohort, tissue sampling was only obtained in 46 lymph nodes that also met the other inclusion criteria. Thus, a selection bias was introduced, as mediastinal staging was not necessary in many of the patients. Even though the lymph nodes examined by CTTA are subjected to a selection bias, the study population reflects the challenges experienced in an MDT setting of a tertiary lung cancer center. Despite the selection bias, CTTA seems to add complementary diagnostic information as 16 of the 46 lymph nodes changed status from malignant to benign or vice versa in accordance with the gold standard, when compared to normal CT.

The results presented in this paper could potentially open up for characterizing lymph nodes in other cancer forms as well as it would be plausible to believe similar changes would be seen in those settings as well.

The study has limitations. First, a relative small number of patients were included. Secondly, our selection bias also limited our study. The reason for this bias was to ensure that different CT techniques with different acquisition parameters and different contrast media policies was not reflected in the heterogeneity measures of texture analysis. Certainly if the study was performed today, more lymph nodes could have been included due to the use of EBUS/EUS guided biopsies in a routine setting. However, the strict selection of lymph nodes evaluated is also a strength when the impact of texture analysis of lymph nodes is considered. In future studies the introduction of

EBUS/EUS and CT guided biopsies from lymph nodes in all lymph node stations may ensure a "full" N-staging so the N-staging by texture can be evaluated on a per patient basis instead of a per lesion basis. Whenever an imaging technique is compared with histopathology of a surgical specimen there will always be the risk that it is not the same lesion that is compared. The risk for such mistakes in the present study is reduced as only the largest lymph node in a given lymph node station on CT was compared with the largest lymph node in the corresponding surgical specimen.

In conclusion CTTA in mediastinal lymph nodes may be helpful and add information to CT in the differentiations between metastatic and benign lymph nodes in patients suspected for lung cancer. However, the technique does not yet have the diagnostic impact necessary to replace current diagnostic measures. Further studies comparing CT + CTTA to EBUS/EUS and F-18-FDG PET/CT on a patient-by-patient basis is warranted.

Reference list:

1. Harders SW, Madsen HH, Hjorthaug K, et al. Characterization of pulmonary lesions in patients with suspected lung cancer: computed tomography versus [^{18}F] fluorodeoxyglucose-positron emission tomography/computed tomography. Cancer Imaging 2012;12:437–446.

2. Harders SW, Madsen HH, Hjorthaug K, et al. Mediastinal staging in Non-Small-Cell Lung Carcinoma: computed tomography versus F-18-fluorodeoxyglucose positron-emission tomography and computed tomography. Cancer Imaging 2014;14:23.

3. Silvestri GA, Gonzalez AV, Jantz MA, et al. Methods for staging non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. Chest 2013;143:e211S –250S.

4. Darling GE, Maziak DE, Inculet RI, et al. Positron emission tomography-computed tomography compared with invasive mediastinal staging in non-small cell lung cancer: results of mediastinal staging in the early lung positron emission tomography trial. J Thorac Oncol 2011;6:1367–1372.

5. Lv Y-L, Yuan D-M, Wang K, et al. Diagnostic performance of integrated positron emission tomography/computed tomography for mediastinal lymph node staging in non-small cell lung cancer: a bivariate systematic review and meta-analysis. J Thorac Oncol 2011;6:1350–1358.

6. Ganeshan B, Abaleke S, Young RCD, et al. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. Cancer Imaging 2010;10:137–143.

7. Ganeshan B, Panayiotou E, Burnand K,et al. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. Eur Radiol 2012;22:796–802.

8. 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;67:157–164.

9. Ng F, Ganeshan B, Kozarski R, et al. Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival. Radiology 2013;266:177–184.

10. Zhang H, Graham CM, Elci O, et al. Locally advanced squamous cell carcinoma of the head and neck: CT texture and histogram analysis allow independent prediction of overall survival in patients treated with induction chemotherapy. Radiology 2013;269:801–809.

11. Goh V, Ganeshan B, Nathan P, et al. Assessment of response to tyrosine kinase inhibitors in metastatic renal cell cancer: CT texture as a predictive biomarker. Radiology 2011 ;261:165–171.

12. Skogen K, Ganeshan B, Good C, et al. Measurements of heterogeneity in gliomas on computed tomography relationship to tumour grade. J Neurooncol 2013;111:213–219.

13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 8;1:307–310.

14. Prenzel KL, Mönig SP, Sinning JM, et al. Lymph node size and metastatic infiltration in non-small cell lung cancer. Chest 2003;123:463–467.

15. Gómez-Caro A, Boada M, Cabañas M, et al. False-negative rate after positron emission tomography/computer tomography scan for mediastinal staging in cI stage non-small-cell lung cancer. Eur J Cardio-Thorac Surg 2012;42:93–100; discussion 100.

16. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228–247.

17. Carrillo SA, Daniel VC, Hall N, et al. Fusion positron emission/computed tomography underestimates the presence of hilar nodal metastases in patients with resected non-small cell lung cancer. Ann Thorac Surg 2012;93:1621–1624.

18. Fernández-Esparrach G, Sendino O, et al. [New endoscopic ultrasound (EUS) applications in lung cancer: evaluation of patients with negative mediastinal CT and re-staging after neoadjuvant treatment]. Arch Bronconeumol 2011;47:410–414.

19. Herth FJF. Nonsurgical staging of the mediastinum: EBUS and EUS. Semin Respir Crit Care Med 2011;32:62–68.

20. Bayanati H, E Thornhill R, Souza CA, et al. Quantitative CT texture and shape analysis: can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer? Eur Radiol 2015;25:480–487.

Table 1.: Distribution characteristics for CTTA (mean image intensity) stratified for benign and malignant lymph nodes. P-values for the independent t-test with a null hypothesis that there is no difference between the two groups are listed as well.

Parameter	**Mean**	**95% confidence interval**	**p-values**
	CTTA benign (mean image intensity)		
Without filtration	55.66	50.36 to 60.98	0.001
Filter = 1.4 mm	3.22	0.63 to 5.82	0.566
Filter = 2.0 mm	-8.63	-13.12 to -4.14	0.479
Filter = 2.7 mm	-14.03	-18.15 to -9.90	0.152
Filter = 3.4 mm	-15.06	-18.77 to -11.35	0.066
Filter = 4.1 mm	-11.73	-15.71 to -7.75	0.366
	CTTA malignant (Mean image intensity		

Without filtration	75.32	70.23 to 80.42
Filter = 1.4 mm	2.4	-0.09 to 4.90
Filter = 2.0 mm	-5.17	-8.90 to -1.44
Filter = 2.7 mm	-7.97	-11.67 to -4.27
Filter = 3.4 mm	-8.26	-12.60 to -3.92
Filter = 4.1 mm	-8.24	-12.94 to -3.54

Table 2.: 2x2 classification table derived from the binary logistic regression analysis

Cutoff =0.6	**Benign**	**Malign**
Malign	1	8
Benign	30	7

Figures legends

Fig 1.: Flow chart for obtaining final study group.

Fig 2.: The CTTA window with a drawn RoI and highlighted features of fine 1.4 mm -> coarse textures 4.1 mm. Fine textures in red, medium textures in green and coarse textures in blue colors.

Fig 3.: Box-and-whisker plot for unfiltered mean divided between benign (1) and malignant (2) group. Notice the 3 outliers in the benign group.

Fig 4.: Sensitivity/Specificity plotted against cutoff values.

Fig 5.: Receiver operating curve for the CE-CTTA technique. Area under the curve = 83.4% and $p<0,0001$

Fig 6.: Agreement plots for unfiltered mean intra-observer variance. Punctuated line show the 95% limits of agreement while the solid line show the mean of differences.

Fig 7: Agreement plots for unfiltered mean inter-observer variance. Punctuated line show the 95% limits of agreement while the solid line show the mean of differences. Notice the mean of differences suggest a systematic bias.