

Evaluation of Shape and Textural Features from CT as Prognostic Biomarkers in Non-small Cell Lung Cancer

FRANCESCO BIANCONI¹, MARIO LUCA FRAVOLINI¹, RAQUEL BELLO-CEREZO¹,
MATTEO MINISTRINI², MICHELE SCIALPI² and BARBARA PALUMBO²

*Departments of ¹Engineering, and ²Surgical and Biomedical Sciences,
University of Perugia, Perugia, Italy*

Abstract. *Background/Aim.* We retrospectively investigated the prognostic potential (correlation with overall survival) of 9 shape and 21 textural features from non-contrast-enhanced computed tomography (CT) in patients with non-small-cell lung cancer. *Materials and Methods.* We considered a public dataset of 203 individuals with inoperable, histologically- or cytologically-confirmed NSCLC. Three-dimensional shape and textural features from CT were computed using proprietary code and their prognostic potential evaluated through four different statistical protocols. *Results.* Volume and grey-level run length matrix (GLRLM) run length non-uniformity were the only two features to pass all four protocols. Both features correlated negatively with overall survival. The results also showed a strong dependence on the evaluation protocol used. *Conclusion:* Tumour volume and GLRLM run-length non-uniformity from CT were the best predictor of survival in patients with non-small-cell lung cancer. We did not find enough evidence to claim a relationship with survival for the other features.

Much research in recent years has focused on the identification of reliable prognostic factors to enable personalised care for patients with non-small-cell lung cancer (NSCLC) (1-3). Among them, the assessment of tumour heterogeneity through shape and textural features from imaging data has been receiving increasing attention. It is in fact generally believed that heterogeneity is associated with adverse biology, and, ultimately, poor prognosis and worse response to therapy (4-6).

Correspondence to: Francesco Bianconi, Department of Engineering, University of Perugia, Via Goffredo Duranti, 93 – 06125 Perugia (PG), Italy. Tel: +39 0755853706, Fax: +39 0755853703, e-mail: bianco@ieee.org

Key Words: Computed tomography, non-small-cell lung cancer, shape, texture, radiomics.

Computed tomography (CT) is usually the front-line imaging approach in many neoplastic disorders, and, as such, also the primary source of baseline data for most patients with NSCLC. The use of CT-derived features as potential biomarkers to predict survival in patients with NSCLC has, therefore, elicited intense research interest in recent years (7-8).

Ganeshan and Miles were among the first to suggest that textural features from CT could be correlated with tumour metabolism and stage (4). Since then various authors have investigated the subject obtaining different – sometimes diverging – results. Aerts *et al.* for instance analysed 440 radiomic features on a dataset of 1,019 individuals with lung and head-and-neck cancer and found that 238 features yielded a significant survival difference (9). They also noted that features describing heterogeneity correlated with worse survival in all the datasets considered. Fried *et al.* retrospectively investigated 91 patients with stage III NSCLC treated with radio- and chemotherapy and concluded that predictive models incorporating textural features from CT and conventional prognostic factors outperformed those based on the latter alone (10). Hayano *et al.* found strong correlation between histogram intensity features (mean and entropy) and overall survival on a cohort of 35 patients with advanced NSCLC treated with chemotherapy (11). Coroller *et al.* evaluated 635 radiomics features from pre-treatment CT scans in patients with lung adenocarcinoma and determined that 35 features were prognostic for distant metastasis and 12 for survival (12).

On the other hand, however, Sacconi *et al.* did not find any statistically significant correlation between CT textural features and survival in patients with adenocarcinoma (13), whereas only 1 of the 329 texture and non-textural features examined by Balagurunathan *et al.* emerged as statistically significant in separating an independent dataset into low- and high-risk groups of patients with NSCLC (14). It is also not uncommon – and this is quite surprising, if not alarming – for the same feature to be linked both positively and negative outcome in different studies: entropy for instance was significantly associated with favourable outcome (overall

survival) in the study of Win *et al.* (15) and unfavourable in that of Hayano *et al.* (11).

It has been argued that most of these discrepancies can be explained by the different statistical criteria used. Chalkidou *et al.*, for instance, scrutinised 15 research articles and found insufficient evidence to claim a relationship of PET- and CT-derived features with patient survival (16). More recently, McQuaid *et al.* took things further and showed that *p*-values for measurements in studies on texture in CT are very sensitive on factors such as the selection of the optimal cut-off values and the length of the follow-up period (17).

The objective of this work was to evaluate experimentally the potential of 30 CT-derived features (nine shape features and 21 textural features) as prognostic biomarkers in NSCLC. For reproducible research purposes, the analysis was carried out on a publicly available dataset.

Materials and Methods

We considered 203 conventional (non-contrast-enhanced) baseline CT scans from as many individuals with inoperable, histologically or cytologically confirmed NSCLC. This patient series is a subset of the ‘NSCLC-Radiomics’ collection, which is publicly accessible at the Cancer Imaging Archive (18), and all necessary approvals and authorizations were obtained at the institutions where the data were collected (9). Of the 422 patients originally included in the aforementioned dataset, we retained those for which both a pre-treatment (baseline) CT scan and manual segmentation of the lesion were available (n=318) (samples shown in Figure 1). We further discarded 45 cases that either did not allow for correct reconstruction of the 3D volume of the lesion or for which the segmentation provided was patently wrong or dubious. Another 70 cases for which only contrast-enhanced CTs were available were also removed from the study. The tumour and patient characteristics of the series are summarised in Table I; further details about the image acquisition protocol and related settings are available elsewhere (9).

For image analysis, we used nine shape and 21 textural features (Table II). Image pre-processing involved windowing to a central value of 50 HU and width of 300 HU [same settings as in (19)] and linear resampling to 256 levels (1 level≈1.2 HU). No further pre-processing steps such as filtering or contrast enhancement were applied.

The nine shape features were: volume (the total volume of the lesion), sphericity (ratio between the area of the surface of a sphere with the same volume V as the object and the area A of the surface of the object – in our implementation we considered V as the total number of voxels in the lesion and A the number of voxels in the outer shell of the lesion), rectangular fit (ratio between the volume of the lesion and the volume of the minimal rectangular bounding box), three mass shape factors (relative elongation of the lesion along the three principal axes – *i.e.* ratios between the eigenvalues of the inertia matrix of the lesion weighted by the intensity level) and three volume shape factors (same as the mass shape factors, but non-weighted by the intensity level).

The textural features belonged to four groups as detailed below.

- Seven first-order statistics: entropy (9), mean (20), mean of positive values (21), standard deviation (20), skewness (20), kurtosis (20) and uniformity (21);

Table I. Characteristics of the patient series.

Characteristic	Value
Age (range) years	68.2 (42.5-87.1)
Gender, n (%)	
Male	143 (70.4%)
Female	60 (29.6%)
Clinical T-stage*, n (%)	
T0	0 (0.0%)
T1	54 (26.6%)
T2	79 (38.9%)
T3	21 (10.3%)
T4	47 (23.2%)
Tx	2 (1.0%)
Clinical N-stage*, n (%)	
N0	94 (46.3%)
N1	6 (3.0%)
N2	65 (32.0%)
N3	38 (18.7%)
Nx	0 (0.0%)
Histology, n (%)	
Adenocarcinoma	21 (10.4%)
Large cell carcinoma	80 (39.4%)
Squamous cell carcinoma	50 (24.6%)
Not otherwise specified	29 (14.3%)
Unknown/not available	23 (11.3%)

*According to the TNM staging system (27).

- Six features from an isotropic three-dimensional grey-level co-occurrence matrix (GLCM) with displacement of 1 pixel averaged over 26 directions: contrast, correlation, dissimilarity, energy, entropy and homogeneity (9).
- Three features from a neighbourhood grey-tone difference matrix (NGTDM) with the same neighbourhood settings as in GLCM: coarseness, contrast and texture strength (22). With respect to the original formulation, a normalisation factor was introduced to guarantee that the features were independent of the number of voxels in the lesion.
- Four features from a grey-level run-length matrix (GLRLM) with the same directions as GLCM: grey-level non-uniformity, long runs emphasis, run length non-uniformity and short runs emphasis (23).

Statistical analysis. To assess the predictive power of the image features described above, we employed two forms of analysis that are commonly used in related studies (11, 15, 17, 19, 21): Cox proportional hazards univariate regression analysis and Kaplan–Meier survival analysis. The results of the two tests were combined using four different evaluation protocols as described below. The endpoint was the overall survival in all the test performed. We considered a Cox’s univariate regression model as follows:

$$h(t)=h_0(t)e^{\beta(X)} \tag{Eq. 1}$$

where *h(t)* is the hazard function (probability of an individual who is under observation experiencing an event at time t), *h₀(t)* the hazard at the baseline, *β* the coefficient and *X* the explanatory variable – which in this case represents the radiomic feature the

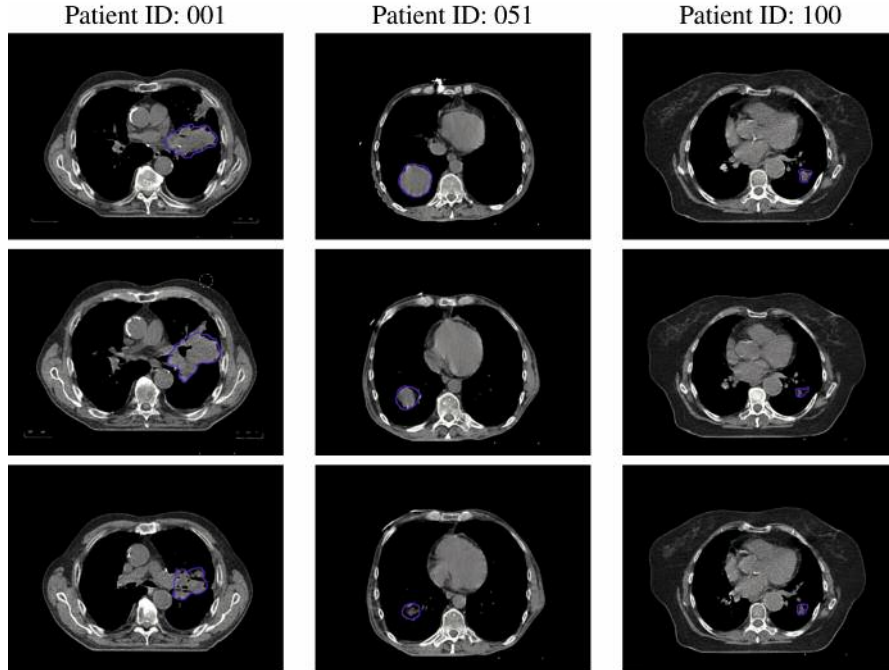


Figure 1. Computed tomographic slices from three different individuals with non-small-cell lung cancer showing the manually delineated lesions.

Table II. Shape and textural computed tomographic features: summary table.

Shape features	Textural features			
	Intensity histogram	GLCM	NGTDM	GLRLM
Mass shape factor I	Entropy	Contrast	Coarseness	Grey-level non-uniformity
Mass shape factor II	Mean	Correlation	Contrast	Long runs emphasis
Mass shape factor III	Mean of positive values	Dissimilarity	Texture strength	Run-length non-uniformity
Rectangular fit	Entropy	Energy		Run percentage
Sphericity	Kurtosis	Entropy		Short runs emphasis
Volume	Skewness	Homogeneity		
Volume shape factor I	Standard deviation			
Volume shape factor II				
Volume shape factor III				

GLCM: Grey-level co-occurrence matrix; NGTDM: neighborhood gray-tone difference matrix; GLRLM: grey-level run-length matrix.

predictive power of which is being investigated. We computed the baseline hazard at the average of X (indicate as ‘ x ’ below), therefore the hazard rate at t can be expressed as follows:

$$h(t) = e^{\beta(X-x)} \tag{Eq. 2}$$

A value for a feature which is higher than the average is associated with a higher hazard ratio (HR) when $\beta > 0$ ($HR > 1$) and a lower hazard when $\beta < 0$ ($HR < 1$).

For Kaplan–Meier survival analysis, we adopted an optimal cut-off approach to dichotomize the population into high- and low-risk group. This was done by testing a set of $K=7$ candidate cut-off

points $C_k, k \in \{1, \dots, K\}$ corresponding to the 20th, 30th, 40th, ..., 80th percentiles of the distribution of each feature considered. For each C_k , we estimated the Kaplan–Meier survival curves of the resulting high- and low-risk group and compared them through log-rank test. For each feature, we finally retained the cut-off value that yielded the highest significance (lowest p -value).

We analysed the results of the two statistical models described above according to the following four evaluation protocols (sorted from the least to the most strict):

- Protocol A: This model only considered the results of the Kaplan–Meier survival analysis. The significance level was $\alpha=0.05$. No correction for multiple tests was applied.

Table III. Results of Cox proportional hazards univariate regression analysis (left) and Kaplan–Meier survival analysis (right) of shape and textural computed tomographic features. Boldface indicates a p-value <0.01.

Radiomic feature	β	p-Value	HR	p-Value
Volume	0.004	0.0004	1.480	0.0011
Sphericity	-0.195	0.7489	1.555	0.1438
Volume shape factor I	0.129	0.7607	1.847	0.1164
Volume shape factor II	0.271	0.6041	1.674	0.3455
Volume shape factor III	0.109	0.9235	1.487	0.4890
Mass shape factor I	0.836	0.0608	1.515	0.0155
Mass shape factor II	1.457	0.0077	1.332	0.0070
Mass shape factor III	2.352	0.0382	1.427	0.0340
Rectangular fit	3.434	0.0013	1.023	0.0052
Mean	0.023	0.0013	0.889	0.0012
Mean of positive values	0.021	0.0025	0.836	0.0020
Entropy	-0.096	0.8274	0.870	0.2160
Standard deviation	-0.016	0.1430	0.782	0.0582
Skewness	-0.081	0.6719	0.753	0.3917
Kurtosis	0.204	0.0019	1.057	0.0387
Uniformity	37.52	0.2472	1.324	0.0749
GLCM contrast	0.000	0.1118	1.058	0.0235
GLCM correlation	0.000	0.2639	0.798	0.1303
GLCM dissimilarity	-0.021	0.0773	0.761	0.0028
GLCM energy	-263.138	0.6200	0.725	0.3893
GLCM entropy	0.156	0.1753	0.912	0.1879
GLCM homogeneity	8.181	0.0609	1.146	0.0237
NGTDM coarseness	10.214	0.1143	1.156	0.0344
NGTDM contrast	-2.273	0.4542	1.010	0.0020
NGTDM texture strength	0	0.0349	0.879	0.0123
GLRLM short run emphasis	1.208	0.9089	0.839	0.1872
GLRLM long run emphasis	-0.034	0.9892	0.839	0.1686
GLRLM grey-level non-uniformity	0.017	0.0006	0.830	0.0039
GLRLM run-length non-uniformity	0.001	0.0006	1.062	0.0012
GLRLM run percentage	3.434	0.0013	1.040	0.0052

GLCM: Grey-level co-occurrence matrix; NGTDM: neighborhood gray-tone difference matrix; GLRLM: grey-level run-length matrix.

- Protocol B: Same as in A, but with correction for multiple tests. The significance level was adjusted using the most conservative (lowest α after correction) between Bonferroni and Benjamini–Hochberg procedure.
- Protocol C: A feature was considered significant when all the following conditions were satisfied: i) significant according to model B; ii) significant according to Cox univariate regression ($\alpha=0.05$), and iii) the feature correlated either positively or negatively with good outcome in both Cox’s regression and Kaplan–Meier analyses.
- Protocol D: Same as in C, but with a significance level $\alpha=0.01$ instead of $\alpha=0.05$ for both tests.

Results

The overall results of Cox univariate regression and Kaplan–Meier survival analysis are summarised in Table III; the performance of the radiomic features in each of the four evaluation protocols is reported in Table I. As can be seen, evaluation protocol A returned 17 statistically significant features: volume; mass shape factor I, II and

III; rectangular fit; mean; mean of positive values; kurtosis; GLCM contrast, dissimilarity and homogeneity; NGTDM coarseness, contrast and texture strength; GLRLM grey-level non-uniformity, run-length non-uniformity and run percentage. Four of them, namely mean, mean of positive values, NGTDM texture strength and GLRLM grey-level non-uniformity were positively correlated with outcome, whereas all the remaining ones were negatively correlated. However, after applying correction for multiple tests (protocol B), the set of significant features were reduced to volume, mean, mean of positive values and GLRLM run-length non-uniformity. Combined application of Cox’s regression and Kaplan–Meier survival analysis (protocol C) further limited the number of significant features from four to two, namely, volume and GLRLM run-length non-uniformity. These two features were still significantly positively associated with outcome after reducing the threshold from $\alpha=0.05$ to $\alpha=0.01$ (protocol D).

Discussion

Prior work has advocated the use of image features from CT as prognostic bio-markers in NSCLC (9-11, 14). Recent articles, however, have reported insufficient evidence to support any such relationships, suggesting – by contrast – that there could be a potentially high number of false discoveries (type-I errors) due to a number of factors such as small sample size and inappropriate data analysis (16-17).

Our calculations actually seem to confirm that the results can be strongly dependent on the statistical analysis performed. By using four different evaluation protocols, we found that the outcome was in fact quite different, with the number of statistically significant features ranging from two to 17. The two features that reached statistical significance in all the tests were volume and GLRLM run-length non-uniformity. These results are in agreement with other findings available in the current literature – see for instance articles (24, 25) on volume and (14) on GLRLM run-length non-uniformity. It should, however, be noted that these two features were strongly correlated ($R^2=0.89$) in our study [see also Fave *et al.* on this point (19)]. This casts doubts on whether GLRLM run-length non-uniformity can actually provide additional information compared to volume alone.

It is also worth commenting on the magnitude of the effects and potential clinical value of volume and GLRLM run-length non-uniformity. As for the volume, our study indicated an increased relative risk (Table III – right) of negative outcome of about 48% per 1 cm³ increment of lesion volume. This is a value not far from those reported by Zhang *et al.* (24). The effect is, therefore, measurable suggesting that tumour size can actually provide additional predictive information for survival besides other parameters, as other authors have also claimed (24, 25). By contrast, the potential clinical value of GLRLM run-length non-uniformity is more difficult to assess: the increase in the relative risk was in fact lesser in this case (*approx.* 6% per unit), and – differently from the volume – the unit of measure of this feature has no direct physical interpretation. Further studies are therefore needed to validate the use of this parameter in clinical practice.

In conclusion, previous studies investigating textures features from CT as potential markers for survival and response to treatment in NSCLC suggested that patients with heterogeneous (non-uniform) tumours had poorer survival (21) and lower response to treatment (22). Yet we found insufficient evidence to claim a relationship between heterogeneity and overall survival in this study. Classic textural features that are indicative of heterogeneity (*e.g.* entropy, standard deviation, GLCM contrast and GLCM correlation) failed to reach statistical significance in our calculations. Likewise, the average tissue density, which was considered significant in previous studies (11), gave divergent results in this work.

Another suggestive finding is the strong dependence of the outcome on the evaluation protocol used. In our study, the number of statistically significant features varied from 2 to 17 when switching from the tightest to the loosest protocol. We observed that the simple adjustment of the significance level through correction for multiple tests reduced the number of significant features approximately fourfold. This results seem to confirm – on a quantitative basis – the concerns recently raised by Chalkidou *et al.* (16), namely the risk of type-I error inflation and, consequently, of increased false-discovery rates in studies with radiomic features.

References

- 1 Kley K and Oehr P: Usefulness of combined FDG-PET with CT or tumour markers in lung cancer diagnosis. *Anticancer Res* 30: 1741-1745, 2010.
- 2 Lin Y, Lin W-Y, Kao C-H, Yen K-Y, Chen S-W and Yeh J-J: Prognostic value of preoperative metabolic tumor volumes on PET-CT in predicting disease-free survival of patients with stage I non-small cell lung cancer. *Anticancer Res* 32: 5087-5092, 2012.
- 3 Cuyún Carter G, Barrett AM, Kaye JA, Liepa AM, Winfree KB and John WJ: A comprehensive review of nongenetic prognostic and predictive factors influencing the heterogeneity of outcomes in advanced non-small-cell lung cancer. *Cancer Manag Res* 6: 437-449, 2014.
- 4 Ganeshan B and Miles KA: Quantifying tumour heterogeneity with CT. *Cancer Imaging* 13: 140-149, 2013.
- 5 Bashir U, Siddique MM, McLean E, Goh V and Cook GJ: Imaging heterogeneity in lung cancer: Techniques, applications, and challenges. *Am J Roentgenol* 207: 534-543, September 2016.
- 6 Koh YW, Park SY, Hyun SH and Lee SJ: Associations between PET textural features and GLUT1 expression, and the prognostic significance of textural features in lung adenocarcinoma. *Anticancer Res* 38: 1067-1071, 2018.
- 7 Chen B, Zhang R, Gan Y, Yang L and Li W: Development and clinical application of radiomics in lung cancer. *Radiat Oncol* 12: 154, 2017.
- 8 Lee G, Lee HY, Park H, Schiebler ML, van Beek EJR, Ohno Y, Seo JB and Leung A: Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. *Eur J Radiol* 86: 297-307, 2017.
- 9 Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ and Lambin P: Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 5: 4006, 2014.
- 10 Fried DV, Tucker SL, Zhou S, Liao Z, Mawlawi O, Ibbott G and Court LE: Prognostic value and reproducibility of pretreatment CT textural features in stage III non-small cell lung cancer. *Int J Radiat Oncol Biol Phys* 90: 834-842, 2014.
- 11 Hayano K, Kulkarni NM, Duda DG, Heist RS and Sahani DV: Exploration of imaging biomarkers for predicting survival of patients with advanced non-small cell lung cancer treated with antiangiogenic chemotherapy. *AJR Am J Roentgenol* 206: 987-993, 2016.

- 12 Coroller TP, Grossmann P, Hou Y, Rios Velazquez E, Leijenaar RTH, Hermann G, Lambin P, Haibe-Kains B, Mak RH and Aerts HJWL: CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol* 114: 345-350, 2015.
- 13 Sacconi B, Anzidei M, Leonardi A, Boni F, Saba L, Scipione R, Anile M, Rengo M, Longo F, Bezzi M, Venuta F, Napoli A, Laghi A and Catalano C: Analysis of CT features and quantitative texture analysis in patients with lung adenocarcinoma: a correlation with *EGFR* mutations and survival rates. *Clin Radiol* 72: 443-450, 2017.
- 14 Balagurunathan Y, Gu Y, Wang H, Kumar V, Grove O, Hawkins S, Kim J, Goldgof DB, Hall LO, Gatenby RA and Gillies RJ: Reproducibility and prognosis of quantitative features extracted from CT images. *Transl Oncol* 7: 72-87, 2014.
- 15 Win T, Miles KA, Janes SM, Ganeshan B, Shastry M, Endozo R, Meagher M, Shortman RI, Wan S, Kayani I, Ell PJ and Groves AM: Tumor heterogeneity and permeability as measured on the CT component of PET/CT predict survival in patients with non-small cell lung cancer. *Clin Cancer Res* 19: 3591-3599, 2013.
- 16 Chalkidou A, O'Doherty MJ and Marsden PK: False-discovery rates in PET and CT studies with textural features: A systematic review. *PLoS ONE* 10: e0124165, 2015.
- 17 McQuaid S, Scuffham J, Alobaidli S, Prakash V, Ezhil V, Nisbet A, South C and Evans P: Factors influencing the robustness of p-value measurements in CT texture prognosis studies. *Phys Med Biol* 62: 5403-5416, 2017.
- 18 The cancer imaging archive. Available online at: <http://www.cancerimagingarchive.net/> [Last accessed on Jul 25, 2017].
- 19 Fave X, Zhang L, Yang J, Mackin D, Balter P, Gomez D, Followill D, Jones AK, Stingo F and Court LE: Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung cancer. *Transl Cancer Res* 5: 349-363, 2016.
- 20 Newbold P, Carlson WL and Thorne B: *Statistics for business and economics*. Pearson Education International, 2007.
- 21 Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S and Miles K: Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: A potential marker of survival. *Eur Radiol* 22: 796-802, 2012.
- 22 Adamasun M and King R: Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern Syst* 19: 1264-1274, 1989.
- 23 Galloway MM: Texture analysis using gray level run lengths. *Comput. Gr. Image Process* 4: 172-179, 1975.
- 24 Zhang J, Gold KA, Lin HY, Swisher SG, Xing Y, Lee JJ, Kim ES and William WN Jr: Relationship between tumor size and survival in non-small-cell lung cancer (NSCLC): An analysis of the surveillance, epidemiology, and end results (SEER) registry. *J Thorac Oncol* 10: 682-690, 2015.
- 25 Huynh E, Coroller TP, Narayan V, Agrawal V, Hou Y, Romano J, Franco I, Mak RH and Aerts HJWL: CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. *Radiother Oncol* 120: 258-266, 2016.
- 26 Ravanelli M, Farina D, Morassi M, Roca E, Cavalleri G, Tassi G and Maroldi R: Texture analysis of advanced non-small cell lung cancer (NSCLC) on contrast-enhanced computed tomography: Prediction of the response to the first-line chemotherapy. *Eur Radiol* 23: 3450-3455, 2013.
- 27 Edge S, Byrd DR, Compton CC, Fritz AG, Greene F and Trotti A (eds.): *AJCC Cancer Staging Handbook*, Springer, 2010.

Received January 24, 2018
Revised February 15, 2018
Accepted February 26, 2018