Stability Assessment of First Order Statistics Features Computed on ADC Maps in Soft-Tissue Sarcoma

Marco Bologna, Eros Montin, Valentina D.A. Corino and Luca T. Mainardi

Abstract— **Radiomics extracts a large number of features from medical images to perform a quantitative characterization. Aim of this study was to assess radiomic features stability and relevance. Apparent diffusion coefficient (ADC) maps were computed from diffusion-weighted magnetic resonance images (DW-MRI) of 18 patients diagnosed with soft-tissue sarcomas (STSs). Thirty-seven intensity-based features were computed on the regions of interest (ROIs). First, ROIs of the images were subjected to translations and rotations in specific ranges. The 37 features computed on the original and transformed ROIs were compared in terms of percentage of variations. The intra-class correlation coefficient (ICC) was computed. To be accepted, a feature should satisfy the following conditions: the ICC after a minimum entity transformation is > 0.6 and the ICC after a maximum entity translation is < 0.4. In total, 31 features out of 37 were accepted by the algorithm. This stability analysis can be used as a first step in the features selection process.**

I. INTRODUCTION

The advances in medical and computational sciences have led to significant improvements in oncology and cancer treatment in the last years. A goal of modern oncology is "personalized medicine", i.e., the use of a treatment tailored to the characteristics of the single patient. In oncology, information about the tumor tissue are required to perform personalized cure. Tumor characteristics can be obtained through biopsies, which are invasive, expensive and localized to a small portion of the tumor. Therefore, biopsies cannot provide complete information in case of highly heterogeneous tissues [1].

Medical imaging can be a useful tool to provide a non-invasive tissue characterization. Starting from this latter assumption, the field of "radiomics" has emerged. Radiomics focuses on extracting and mining a large number of image characteristics (features). Radiomic features are non-invasively computed from images obtained in clinical routine, such as computed tomography, magnetic resonance imaging (MRI) and positron emission tomography [2,3]. The hypothesis is that image features can quantify information regarding intra-tumor heterogeneity, highlighting tumor phenotype [3,4]. It has been shown that this heterogeneity could have profound implications on tumor prognosis [4].

Radiomics can extract hundreds of features and not all of them should be used: some may be redundant, or may be too sensitive to the parameters of the acquisition. The assessment of features that are informative and stable is a first step in the features selection process [2]. There are several techniques used to test feature stability: test-retest [5,6], comparison of features extracted with different machines, reconstruction methods or parameters [6–8], comparison of features computed on slightly different regions of interest (ROIs) [5,9]. Relevance and non-redundancy of the features is usually defined using specific features extraction algorithms.

Among the medical imaging techniques, diffusion weighted MRI (DW-MRI) can capture changes at the cellular level of the tissue thanks to differences in movement of water protons. The apparent diffusion coefficient (ADC) map, derived from several DW-MRIs has been shown to be very useful for tumor detection and characterization [10], evaluation of treatment response [11] and tumor staging [4,12].

The aim of this study was to develop a method to assess stability and relevance of radiomic features extracted from DW-MRI of soft-tissue sarcomas (STSs), by analyzing how the features vary when geometric transformations are applied on the ROIs. Although studies regarding features stability already exist in literature, to the knowledge of the authors, this is the first work in which geometric transformation of the ROIs are used to analyze features relevance.

II. MATERIALS AND METHODS

A. Study population

The dataset for this study consisted of 18 patients, who had a histological diagnosis of STS of intermediate (4 patients) or high (14 patients) malignancy grade according to the FNCLCC (French Fédération Nationale des Centres de Lutte Contre le Cancer) system. The FNCLCC system is based on tumor differentiation, mitotic rate and amount of tumor necrosis. A score is attributed independently to each parameter and the grade is obtained by adding the three attributed scores. Patient characteristics were homogeneous between the two groups. All patients underwent a DW-MRI acquisition. The study was approved by the ethical committee of Fondazione IRCCS-Istituto Nazionale dei Tumori of Milan, Italy, and all patients gave their written informed consent. All patients' data were anonymized prior to the analysis.

B. Image acquisition

DW-MRI images were acquired using Achieva 1.5 T system (Philips Medical system Achieva, Nederlands) or a Magnetom Avanto 1.5 T system (Siemens Medical Solutions, Erlangen, Germany) - 13 patients -, both with a body-matrix coil and spine array coil for signal reception. The data were

M. Bologna is with the Depatment of Electronics, Informatic and Bioengineering, Politecnico di Milano, Milan, Italy (corresponding author, +39 02 23993322; fax +39 02 23993360; e-mail: marco.bologna@polimi.it).

E. Montin, L.T. Mainardi and V.D.A. Corino are with the Depatment of Electronics, Informatic and Bioengineering, Politecnico di Milano, Italy (e-mail: eros.montin@polimi.it, luca.mainardi@polimi.it, valentina.corino@polimi.it)

acquired axially by means of echo planar imaging. DWI were acquired using four b-values (50, 400, 800 and 1000 s/mm^2).

C. Image processing

For each acquisition, the ADC was computed as the slope of the linear regression of the logarithm of the DWI exponential signal decay on the b-values [13]. The calculation was performed pixel-wise using ITK 4.8 [4]. The segmentation of the gross tumor volume was performed by an expert radiologist on the DW-MRI computed with the lowest b-value (50 s/mm^2), on which the anatomical details are maintained and the heterogeneity of the tumor is more visible [15]. The segmentation was performed using 3D Slicer [16].

D. Radiomic features extraction

In this study, 37 intensity-based radiomic features where computed. Those features included statistical information about the signal intensity and histogram distribution of the pixels in the ROI. The histogram was evaluated between 0 and 4000*10^{-6} mm^2/s using N bins. In this study, three different values of N were tested (16, 32 and 64 bins) to evaluate whether the bin number affects the stability of the features. Thus, the total number of computed features was 87 (12 features are bin-independent and the other 25 are bin-dependent and thus they are counted three times).The algorithm used to compute all the features was implemented in ITK 4.8 [4,16].

Features extraction was performed on the ADC maps for each patient first by using the original ROIs as masks and then by using the transformed ROIs (see *II.E*). The results were grouped per type and entity of the transformation.

E. Stability/relevance analysis

To assess the stability and the relevance of the radiomic features, modifications to the ROIs were performed. We assessed the variation of the features after the application of a series of geometric transformations to the original ROI: i) translations along x (medial-lateral direction) in a range of ± 100% of the length of the bounding box in the x direction, with a step ST = 10%, for a total of 21 ROIs (20 transformed plus the original one); ii) same as i) along y (anterior-posterior) direction; iii) rotations around z axis (longitudinal axis of the body) in a range of ±180° with a step SR=18°.

For each ROI, transformation type (translation along x or y, and rotation around z) and entity, a comparison between the features computed for all the patients on the original ROIs and those computed on the transformed ones was made. For each comparison, the percentage variation with respect to the reference was computed as:

$$Diff\% = \frac{F_{Transf} - F_{Original}}{F_{Original}} \cdot 100 \qquad (1)$$

being F_{Transf} and $F_{Original}$ the features computed on the transformed and original ROIs, respectively. Moreover, the intra-class correlation coefficient (ICC) was computed [17,18]. Briefly, the ICC measures the bivariate relation of variables representing different measurement classes, and can be used to assess the agreement of data. In this study, a two-way mixed effect model was used [18]. The maximum value of ICC is 1, when all the elements of each group are the same.

The lower the ICC, the lower the similarity among the elements of the group.

The ICC values were used to assess the stability of the features and their ability to discriminate between healthy and tumor tissue. To perform this task, two hypotheses were made: a) when a transformation (translation of rotation) of minimum entity (translation = ± ST or rotation= ± SR, see Fig.1 A for a schematization and Fig.1 C for an example with a real image) is applied, the features computed on the original and transformed ROIs should be similar; b) when a translation of maximum entity (schematization in Fig.1 B, real example in Fig.1 D) is applied, the features computed on the original and transformed ROIs should be different. These two hypotheses mean that a small movement of the ROI should not change the computed feature as well as the feature should be different when computed on healthy tissue. The two hypotheses are translated into the following two conditions: 1) Let ICC$_{10}$ be an ICC obtained by comparing the features computed on the original ROIs and the ones computed on the ROIs after a minimum entity transformation; hypothesis a) is satisfied if ICC$_{10}$>0.6 [9]. 2) Let ICC$_{100}$ be an ICC obtained comparing the features computed on the original ROIs and the ones computed on the ROIs after a maximum entity translation; hypothesis b) is satisfied if ICC$_{100}$<0.4 [19]. A graphical representation of the range of acceptability of the ICC values for the possible transformations (rotations and translations) is shown in Fig.1 E.
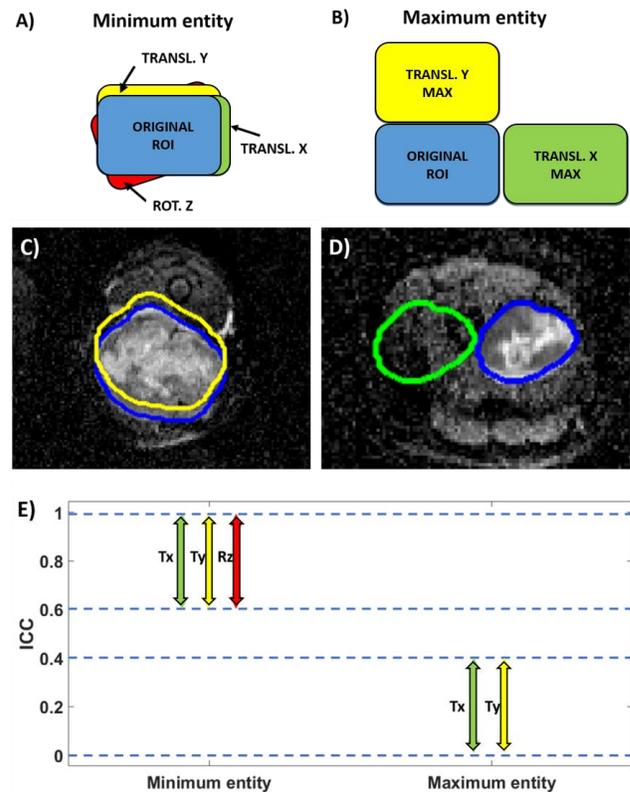

Figure 1 A) Schematization of minimum entity transformations: translation along x (green); translation along y (yellow); rotation around z (red). B) Schematization of maximum entity translation along x (green) and y (yellow). C) ROI in an ADC map (blue contour) and its minimum translation along y (yellow contour). D) ROI in an ADC map (blue contour) and its maximum translation along x (green contour). E) Ranges of acceptability of ICC for the transformations of maximum and minimum entity; Tx, Ty are translation along x and y respectively and Rz is rotation around z.

III. RESULTS AND DISCUSSION

According to two conditions for stability, 31 out of 37 radiomic features were accepted. All the features that were accepted were accepted independently on the number of bins used for histogram discretization. In Table I, all the features analyzed are listed, divided by result obtained for the stability analysis. It can be noted that 31 over 37 features are stable. Fig.3 shows a heat map of the ICC values measured for all the features and bin numbers.

Fig.2 A and B show the variation of two features (signal mean and quantile 0.01) when a translation along x-axis is applied. In particular, in Fig.2 A the average percentage variation shows a gradual but clear decreasing trend when the ROI is translated from its original position. The average percentage variation of the feature in Fig.2 B shows a decreasing trend but with a high variation even for small transformations. Their corresponding ICCs are shown in Fig. 2 C and D, respectively. ICC in Fig. 2 C satisfies the two conditions for stability: ICC_{10} is large for small transformations, and it is very small for the maximum translations. ICC in Fig. 2 D rapidly decreases below 0.6 for minimum entity translations and thus the feature is considered unstable. Fig. 3 shows a heat map with all the ICC values for all the features (computed with 32 bins discretization) when a translation along x is applied to the ROIs. As it can be noted, all the accepted features have a similar trend: the ICC is above 0.6 when the translation is between ±20% of the bounding box and usually goes below the value 0.4 for translation greater than ±50% of the bounding box. The unstable and the irrelevant features do not have this trend. As a matter of fact the rows related to unstable features (signal min, quantiles 0.01 and 0.1) appear as low intensity rows with an high intensity value in the middle. Irrelevant features (histogram total frequency, mean and min) present rows with high intensity values for each entity of translation, because the values measured for those features are always very similar.
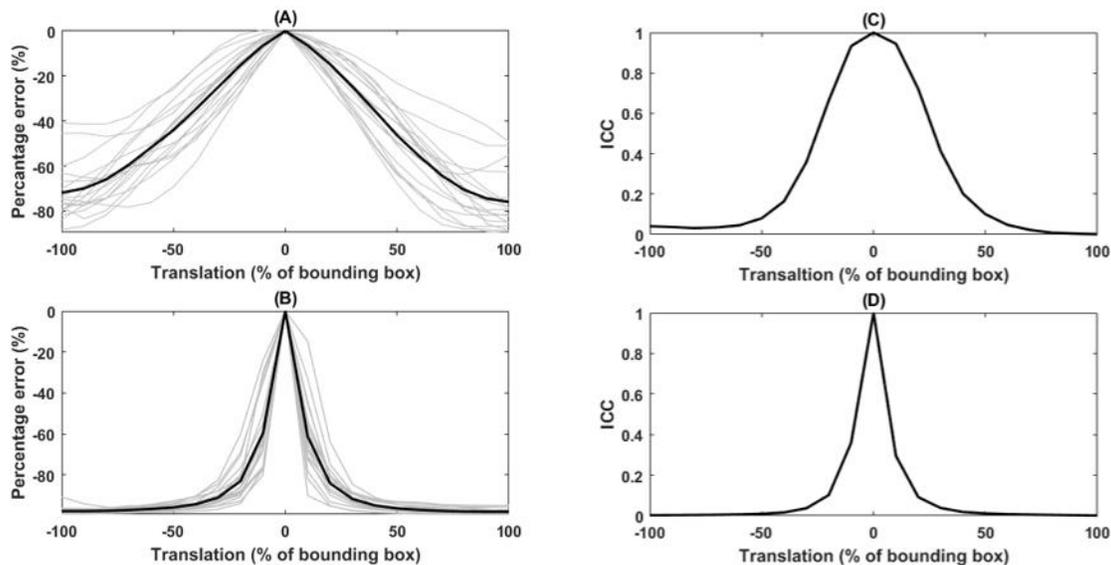
Figure 2 A) Percentage variation of an accepted feature (signal mean) as function of the applied translation along x. B) Percentage variation of a not accepted feature (quantile 0.01) as function of the applied translation along x. C-D) ICC plots corresponding to the transformations reported in A) and B). In A) and B) the thin grey lines represent the errors for each patient, while the black thick lines represents mean value.

TABLE I.   ACCEPTED AND NON-ACCEPTED FEATURES

Accepted Features	Signal Energy, Signal Kurtosis, Signal Mad, Signal Max, Signal Mean, Signal Median, Signal Range, Signal RMS, Signal Skewness, Signal SD, Signal Variance, Quantile 0.2, Quantile 0.3, Quantile 0.4, Quantile 0.5, Quantile 0.6, Quantile 0.7, Quantile 0.8, Quantile 0.9, Quantile 0.99, Histogram Entropy, Histogram Kurtosis, Histogram Mad, Histogram Max, Histogram Median, Histogram Range, Histogram RMS, Histogram Skewness, Histogram SD, Histogram Variance, Histogram Uniformity
Non-accepted Features	Signal Min, Quantile 0.01, Quantile 0.1, Histogram Mean, Histogram Min, Histogram Total Frequency
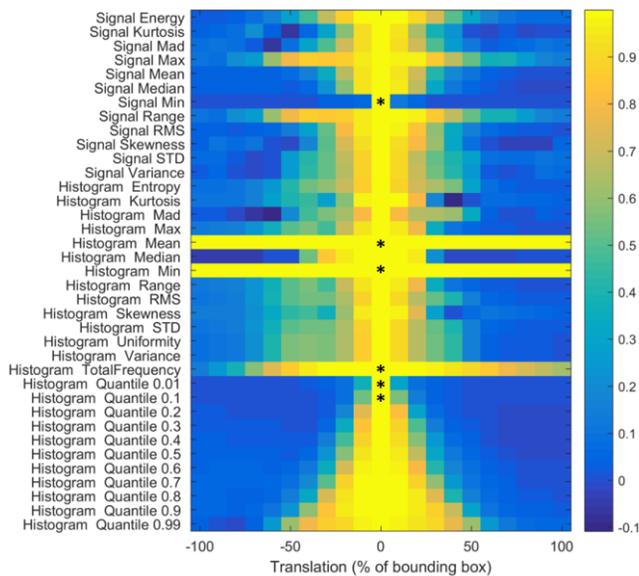
Figure 3 Heat map representing the variation of the ICC of the radiomic features with respect to the translation along x. The non-accepted features (unstable and not relevant) are marked with an asterisk. The features presented in the heat map are computed with a 32 bins discretization.

IV. CONCLUSIONS

A new method to evaluate radiomic features stability computed on ADC maps of soft tissues sarcomas has been proposed. The stability of the features was assessed through translations of the ROIs along the two main axes of the image and rotations around an axis perpendicular to the image plane.

This analysis allows to retain only the features that will not be affected by small differences in manual tumor segmentation (which are a combination of the analyzed geometric transformations). In addition, this kind of analysis can provide a first selection criterion for the radiomic features to be used for tumor characterization, by removing the features that do not give useful information about the ROI.

REFERENCES

[1] H.J. Aerts, E.R. Velazquez, R.T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M.M. Rietbergen, C.R. Leemans, a Dekker, J. Quackenbush, R.J. Gillies, P. Lambin, Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat Commun. 5 (2014) 4006.

[2] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G.P.M. Van Stiphout, P. Granton, C.M.L. Zegers, R. Gillies, R. Boellard, A. Dekker, H.J.W.L. Aerts, Radiomics: Extracting more information from medical images using advanced feature analysis, Eur. J. Cancer. 48 (2012) 441–446.

[3] C. Parmar, R.T.H. Leijenaar, P. Grossmann, E. Rios Velazquez, J. Bussink, D. Rietveld, M.M. Rietbergen, B. Haibe-Kains, P. Lambin, H.J.W.L. Aerts, Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer, Sci. Rep. 5 (2015) 1–10.

[4] R. Fisher, L. Pusztai, C. Swanton, Cancer heterogeneity: implications for targeted therapeutics, Br. J. Cancer. 108 (2013) 479–485.

[5] R.T.H. Leijenaar, Stability of FDG-PET Radiomic features: An integrated analysis of test-retest and inter-observer variability, Acta Oncol. 8 (2013) 583–592.

[6] F.H.P. van Velden, G.M. Kramer, V. Frings, I. a. Nissen, E.R. Mulder, A.J. de Langen, O.S. Hoekstra, E.F. Smit, R. Boellaard, Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [18F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation, Mol. Imaging Biol. (2016) 1–8.

[7] L. a Hunter, S. Krafft, F. Stingo, H. Choi, M.K. Martel, S.F. Kry, L.E. Court, High quality machine-robust image features: identification in nonsmall cell lung cancer computed tomography images., Med. Phys. 40 (2013) 121916.

[8] L. He, Y. Huang, Z. Ma, C. Liang, C. Liang, Z. Liu, Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule, Sci. Rep. 6 (2016) 34921.

[9] O. Gevaert, L. a Mitchell, a S. Achrol, J. Xu, S. Echegaray, G.K. Steinberg, S.H. Cheshier, S. Napel, G. Zaharchuk, S.K. Plevritis, Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features, Radiology. 273 (2014) 168–174.

[10] M. Maeda, H. Kato, H. Sakuma, S.E. Maier, K. Takeda, Usefulness of the apparent diffusion coefficient in line scan diffusion-weighted imaging for distinguishing between squamous cell carcinomas and malignant lymphomas of the head and neck, Am. J. Neuroradiol. 26 (2005) 1186–1192.

[11] Y.-S. Sun, X.-P. Zhang, L. Tang, J.-F. Ji, J. Gu, Y. Cai, X.-Y. Zhang, Locally advanced rectal carcinoma treated with preoperative chemotherapy and radiation therapy: preliminary analysis of diffusion-weighted MR imaging for early detection of tumor histopathologic downstaging., Radiology. 254 (2010) 170–178.

[12] V. Vandecaveye, F. De Keyzer, V. Vander Poorten, P. Dirix, E. Verbeken, S. Nuyts, R. Hermans, Head and neck squamous cell carcinoma: value of diffusion-weighted MR imaging for nodal staging., Radiology. 251 (2009) 134–46.

[13] A.R. Padhani, G. Liu, D. Mu-Koh, T.L. Chenevert, H.C. Thoeny, T. Takahara, A. Dzik-Jurasz, B.D. Ross, M. Van Cauteren, D. Collins, D.A. Hammoud, G.J. Rustin, B. Taouli, P.L. Choyke, Diffusion-Weighted Magnetic Resonance Imaging as a Cancer Biomarker: Consensus and Recommendations, Neoplasia. 11 (2009) 102–125.

[14] P.H. Chawla S, Kim S, Wang S, S. Chawla, S. Kim, S. Wang, H. Poptani, Diffusion-weighted imaging in head and neck cancers, Futur. Oncol. 5 (2009) 959–975.

[15] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. V. Miller, S. Pieper, R. Kikinis, 3D Slicer as an image computing platform for the Quantitative Imaging Network, Magn. Reson. Imaging. 30 (2012) 1323–1341.

[16] T.S. Yoo, Insight into images: principles and practice for segmentation, registration, and image analysis, 2004.

[17] P.E. Shrout, J.L. Fleiss, Intraclass correlations: uses in assessing rater reliability. Psychol. Bull. 86 (1979) 420–8.

[18] K.O. Mcgraw, Forming Inferences About Some Intraclass Correlation Coefficients, Psychol. Methods. 1 (1996) 30–46.

[19] J.L. Fleiss, Design and analysis of clinical experiments, New York: John Wiley & Sons, 1986.