Iranian Journal of Medical Physics

ijmp.mums.ac.ir

Radiomic Feature Reproducibility: The Impact of Inter-Scanner and Inter-Modality Variations

Maryam Heidari[1], Alireza Amouheidari[2], Simin Hemati[3], Hamid Abdollahi[4], Hossein KhanAhmad[5], Parvaneh Shokrani[1*]

1. Department of Medical Physics, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran
2. Department of Radiation Oncology, Isfahan Milad Hospital, Isfahan, Iran
3. Department of Radiotherapy Oncology, Faculty of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran
4. Department of Radiologic Sciences and Medical Physics, Faculty of Allied Medicine, Kerman University of Medical Sciences, Kerman, Iran
5. Department of Genetics and Molecular Biology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

ARTICLE INFO

Article type:
Original Paper

Article history:
Received: Sep 28, 2020
Accepted: Nov 10, 2020

Keywords:
Imaging Genomics
Reproducibility of Results
X-Ray Computed Tomography
Magnetic Resonance Imaging (MRI)
Computer-assisted Image Analysis

ABSTRACT

Introduction: Radiomic features robustness analysis is a critical issue before clinical decision making. In this study, the reproducibility and robustness of radiomic features in computed tomography (CT) and magnetic resonance (MR) images of glioblastoma cancer patients were analyzed regarding inter-scanner and inter-modality variations.

Material and Methods: CT and MR Images of eighteen glioblastoma cancer patients were used to extract the radiomic features following image segmentation. Coefficient of variation (COV), intraclass correlation coefficient (ICC), and concordance correlation coefficient (CCC) analysis were done to select the most robust features in all paired combinations of CT and MR images include T1-T2, T1-FLAIR, T1-ADC, T1-CT, T2-FLAIR, T2-ADC, T2-CT, FLAIR-ADC, FLAIR-CT, and ADC-CT.

Results: The features with COV ≤ 5% or ICC ≥ 90% or CCC ≥ 90%, considered as the most robust features, include the shape features, Minimum (belong to first-order Features), IMC1, IDN, IDMN (belong to GLCM), and Run Length Non-Uniformity (belongs to Gray Level Run Length Matrix).

Conclusion: In this study we presented a large image feature variation among different imaging modalities including CT and MRI. Our results identified several robust features that could be used for further clinical analysis.

►*Please cite this article as:*
Heidari M, Amouheidari A, Hemati S, Abdollahi H, KhanAhmad H, Shokrani P. Radiomic Feature Reproducibility: The Impact of Inter-Scanner and Inter-Modality Variations. Iran J Med Phys 2021; 18: 397-402. 10.22038/IJMP.2020.52406.1863.

Introduction

Quantitative image analysis of medical images, in terms of "Radiomics," has been utilized as a feasible approach for decoding cancer phenotypes [1]. Radiomic analysis is an advanced image processing technique that extracts mineable quantitative features from medical images [2]. In radiotherapy, treatment planning involves the application of computed tomography (CT), magnetic resonance (MR) or positron emission tomography (PET) images. Individualization of radiotherapy treatment planning can be achieved by using tumor specifications of each patient such as inter tumor and intratumor tissue heterogeneities information. It has been shown that such inhomogeneities can influence the treatment response and patient survival [3,4]. Therefore, the application of radiomic features in treatment planning can optimize treatment planning strategies. Radiomics can extract quantitative radiographic phenotype features to predict the genotype of the tumor quickly using data-characterization algorithms. For instance, imaging phenotypes extracted from tomographic images, such as CT, MRI, and PET, can profile gene expression in cancers like hepatic carcinoma (HCC), glioblastoma multiform (GBM), lung cancer, esophageal cancer, and kidney cancer [2].

Glioblastoma brain tumors are highly heterogeneous solid tumors and therefore are regarded as resistant to radiation and chemotherapy agents. To be aware of such heterogeneity, several radiomic analysis based on the various imaging approaches including CT, MRI and PET have been conducted [5]. Results show that several imaging features are correlated to genomic data or could predict tumor grades, stages, therapy response and survival [2, 4, 6].

Radiomic features are divided into different classes, including morphological (shape), intensity-based, and textural features [7]. Radiomic features are

*Corresponding Author: Tel: (98) 313-7929032; Fax: (98) 311 6688597; E-mail: Shokrani@med.mui.ac.ir.

vulnerable to changes in image acquisition, image reconstruction, region of interest (ROI) segmentation methods, and radiomic analysis methods [7-11]. Therefore, it is critical to identify a set of robust and reproducible features for prognosis and prediction of treatment response [4, 7, 8]. Several statistical analyses criteria have been proposed in order to examine the robustness of radiomic features including coefficient of variation (COV), intraclass correlation coefficient (ICC), Bland-Altman analysis, and concordance correlation coefficient (CCC) [12-14].

Although several studies have analyzed the reproducibility of radiomic features, to our knowledge, no study has identified the multimodality reproducible and robust features. In this study, we aimed to analyze the radiomic features in CT and MR images of glioblastoma cancer patients in terms of multimodality reproducibility and robustness.

Materials and Methods

Patients

This prospective study was conducted between December 2018 and December 2019 according to medical ethics guidelines and criteria approved by Isfahan University of Medical Sciences. In this study, our inclusion criteria were as follows. (a) According to the World Health Organization classification, histologically confirmed high grade glioma by stereotactic biopsy or open tumor resection (b) availability of pretreatment CT scan and MRI data including T1, T2, DWI, and FLAIR sequences. Patients having images of low quality and lack of all MRI sequences and also with brain metastases or recurrence of the disease were excluded. Images of 18 patients including 9 men and 9 women were investigated. 16 patients had grade Ⅳ and two patients had grade Ⅲ glioma cancer. The mean age of patients was 55.28 ± 16.37 years (range, 19-72 years).

MRI acquisition parameters

All patients underwent brain MR imaging before chemo-radiotherapy. MRIs were performed with one of these three 1.5-T units: (Siemens MAGNETOM Aera) (10 patients), (Siemens MAGNETOM Avanto) (4 patients), and (PHILIPS, Ingenia) (4 patients). MRI scans included T1W, T2W, FLAIR and DW imaging sequences. The Aera MRI acquisition parameters (FOV = 230 × 230 (mm^2), NEX = 1, Slice thickness = 5 mm, spacing = 6 mm) were as follows: T1 images (TR/ TE = 62), T2 images (TR/ TE = 33), FLAIR (TR/ TE = 76, and DWI (TR/ TE = 59, spacing = 6.5 mm). The Avento MRI acquisition parameters (slice thickness = 5.5 mm, spacing = 7.15 mm), were as follows: T1 images (TR/ TE = 36), T2 images (TR/ TE = 50), FLAIR (TR/ TE = 82), and DWI (TR/ TE 32, Slice thickness = 5 mm, spacing = 6.5 mm). The Ingenia MRI acquisition parameters, were as follows: T1 images (TR/ TE = 37, FOV= 230 × 230 (mm^2), NEX = 1, slice thickness = 5 mm, spacing = 6.1 mm), T2 images (TR/ TE = 45, slice thickness = 5 mm, spacing = 5.7 mm), FLAIR (TR/ TE = 14, slice thickness = 3 mm, spacing = 5 mm), DWI (TR/ TE = 32, slice thickness = 5 mm, spacing = 5.5 mm).

CT acquisition parameters

All patients underwent CT simulation in treatment position before radiation therapy. CT scans were acquired either on SIEMENS Healthcare, SOMATUM scope, or on PHILIPS Healthcare, Brilliance. Detailed scanning parameters are provided in Table 1.

Table 1. CT scanning parameters

Scanning parameters	SIEMENS, SOMATUM scope	PHILIPS, Brilliance
Tube voltage (kV)	130	120
Tube current (mAs)	170	250
FOV (mm)	265	350
Slice thickness (mm)	2-4	2-4

Image segmentation and feature extraction

For each patient, the gross tumor volume was delineated manually slice by slice in all CT and MRI slices using the open source software, 3D-Slicer version 4.10.2 (http://www.slicer.org). Because in glioma tumors, the margin is indistinct, contouring was performed manually based on T1-post contrast enhancement images and then mapped to the other MRI sequences and modified if necessary. The contours were approved by a radiation oncologist with 13 years of experience. An example of image segmentation is shown in Figure 1.

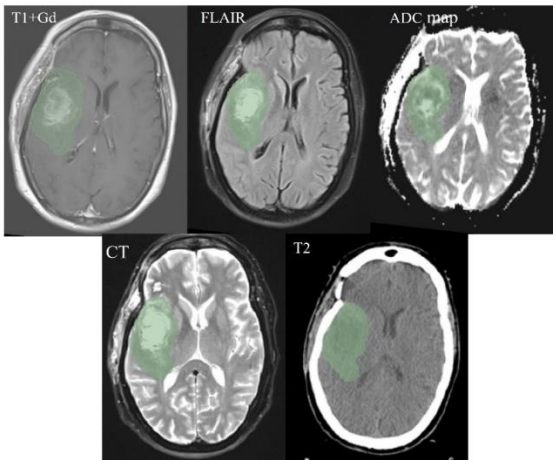

Figure 1. Tumor segmentation on different images (FLAIR, ADC map, CT, and T2) of a GBM patient based on T1-post contrast enhancement image (T1+ Gad) using 3D-Slicer software.

3D slicer can extract radiomic features from DICOM images after the following steps: 1) loading DICOM images, 2) segmentation of ROI, and 3) calculating different radiomic features. In this study 107 radiomic features including 14 shape features, 18 first-order intensity statistics features, and 75 texture features (gray level co-occurrence matrix (GLCM), gray level size zone matrix (GLSZM), gray level run length matrix

(GLRLM), neighboring gray-tone difference matrix (NGTDM), and gray level dependence matrix (GLDM)) were extracted from the segmented tumors.

Statistical analysis

For feature robustness and reproducibility assessment, ICC, CCC, and COV were obtained for each radiomic feature between every paired combination of CT and MR images, including T1 & T2, T1& CT, T1 & ADC map, T1 & FLAIR, T2 & CT, T2 & ADC map, T2 & FLAIR, FLAIR & ADC map, FLAIR & CT, and ADC map & CT. COV was calculated according to equation 1.

$$COV = SD / Mean \times 100 \quad (1)$$

Where SD and mean are standard deviation and mean value of each feature among all patients.

Reproducibility was categorized according to the COV values as (1) high (COV ≤ 5%), (2) intermediate (5% < COV ≤ 10%), (3) low (10% < COV ≤ 20%), and (4) very low (COV > 20%).

ICC is used as a measure of the reliability of the measurements or ratings. Initial definition of ICC was originally recommended by Ronald Fisher according following equation, with considering a data set comprising of N paired data values ($x_{n,1}$, $x_{n,2}$), for n = 1, ..., N [12, 13].

$$r = \frac{1}{NS^2}\sum_{n=1}^{N}(x_{n,1} - x`)(x_{n,2} - x`) \quad (2)$$

where

$$x` = \frac{1}{2N}\sum_{n=1}^{N}(x_{n,1} + x_{n,2})$$

$$S^2 = \frac{1}{2N}\left\{\sum_{n=1}^{N}(x_{n,1} - x`)^2 - \sum_{n=1}^{N}(x_{n,2} - x`)^2\right\}$$

CCC is a desirable reproducibility index to measure the agreement between two variables [14].

When there is a set of N paired data values (x_n , y_n), for n= 1,… ,N, the concordance correlation coefficient (ρ_c) is obtained with following equation,

$$\rho_c = \frac{2\,s_{xy}}{s_x^2 + s_y^2 + (x` - y`)^2} \quad (3)$$

Where x`and y`are means for the variables, s_x and s_{xy} are variance and covariance, respectively (14). These variables are computed as follows,

$$x` = \frac{1}{N}\sum_{n=1}^{N} x_n \quad (4)$$

$$S_x^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - x`)^2 \quad (5)$$

$$S_{xy} = \frac{1}{N}\sum_{n=1}^{N}(x_n - x`)(y_n - y`) \quad (6)$$

ICC and CCC were calculated via MedCalc software version 19.1.3. Reproducibility was categorized as follows: (1) high (CCC or ICC ≥ 90%), (2) intermediate (70% < CCC or ICC ≤ 90%), (3) low (CCC or ICC ≤ 70%), and (4) very low (CCC or ICC with very low and negative values). The reproducibility categorization method was the same for CCC and ICC.

Results

COV

The COV results based on the feature sets are shown in Figure 2 (A-G). It demonstrates a heat map in which the brighter color shows higher reproducibility.

Among the shape features, the COV for Sphericity (between T1-T2 and T2-ADC sequences) and the Major Axis Length (between T1-T2 and T2-ADC sequences) ranged between 5% and 10%. For T1-FLAIR, T2-FLAIR, FLAIR-ADC and FLAIR-CT images, the COV value for all shape features was more than 20%.

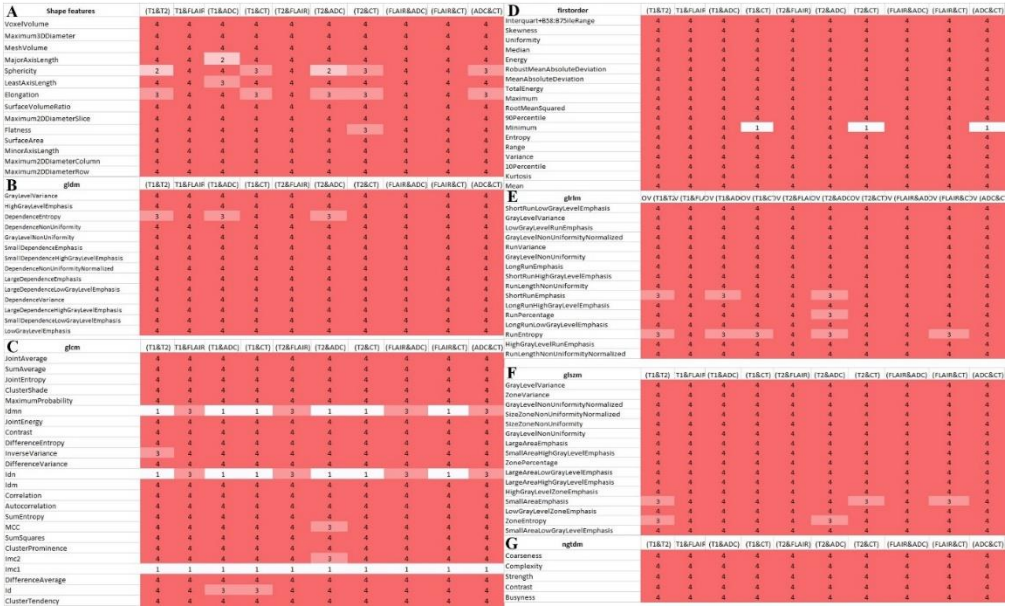

Figure 2. A heat map showing COV results on quantitative feature sets including: A) shape features, B) gldm, C) glcm, D) first-order, E) glrlm, F) glszm, and G) ngtdm among different imaging modalities and sequences.

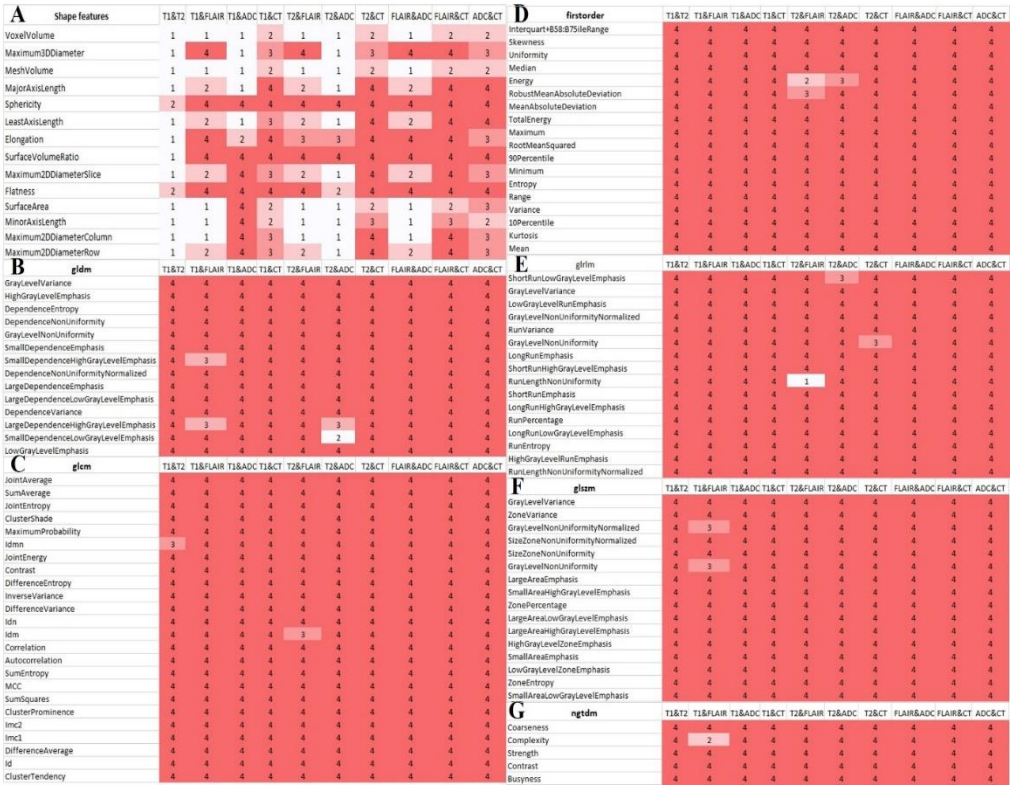

Figure 3. ICC results on feature sets including: A) shape features, B) gldm, C) glcm, D) first-order, E) glrlm, F) glszm, and G) ngtdm among different imaging modalities and sequences.

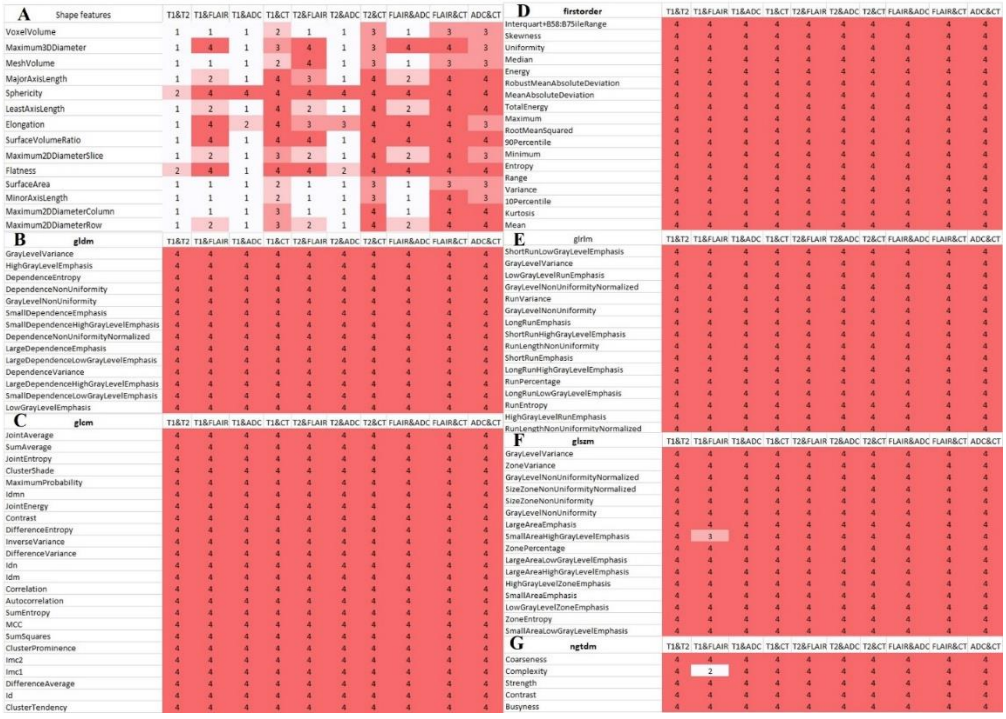

Figure 4. CCC results on the feature sets including: A) shape features, B) gldm, C) glcm, D) first-order, E) glrlm, F) glszm, and G) ngtdm among different imaging modalities and sequences.

For GLDM features, the Dependence Entropy was found as the feature with the lowest COV (10% < COV ≤ 20%) in T1-T2, T1-ADC and T2-ADC images. The COV value for all other GLDM features was more than 20%.

For GLCM features, IMC1 was the most reproducible feature (COV ≤ 5%) among all images. In addition, for features including IDN and IDMN, the COV value was less than 5%, in T1-T2, T1-ADC, T1-CT, T2-ADC, T2-CT and FLAIR-CT image sets. The COV results for T1-FLAIR, T2-FLAIR, FLAIR-ADC, and ADC-CT image sets ranged between 10% and 20%.

For the Minimum feature (belongs to first-order feature class), COV value was less than 5%, in T1-CT, T2-CT, and ADC-CT image sets and for other images, it was more than 20%.

In GLRLM feature class, the Run Entropy feature, for T1-T2, T1-ADC, T1-CT, T2-ADC, and FLAIR-CT image sets, the Run Percentage feature for T2-ADC image set and the Short Run Emphasis feature, for T1-T2, T1-ADC, and T2-ADC image sets, showed COV values in the range of 10% to 20%.

In GLSZM feature class, for T1-T2, T2-CT, and FLAIR-CT image sets, the Small Area Emphasis feature and for T2-ADC image set, the Zone Entropy feature, the COV values ranged between 10% and 20%. For all NGTDM features, the COV values were more than 20%.

ICC

The ICC results based on the feature sets are shown in Figure 3 (A-G). Among the shape features, the most reproducible features (90% < ICC < 100%) belonged to T1-T2 and T2-ADC. In most of the other feature sets, the ICC values were less than 70%.

CCC

The CCC results based on the feature sets are shown in Figure 4 (A-G).

For shape features, the most reproducible features (90% < CCC < 100%) belonged to T1-T2, T1-ADC, and T2-ADC. For GLCM, GLDM, GLRLM, and first-order feature sets, the CCC results in all features were less than 20%. The Small Area High Gray Level Emphasis feature (from GLSZM feature set) and complexity feature (from NGTDM feature set) demonstrated intermediate reproducibility (70 < CCC < 90).

Discussion

Radiomics is now an accepted approach for cancer diagnosis, prognosis and treatment personalization. It is based on quantitative image parameters extracted from high-quality medical images and statistical analyses. However, before introducing to clinical settings, radiomic features have to be assessed in terms of reproducibility and repeatability due to their susceptibility to biasing issues. One of the main biasing issues is variation in the image acquisition parameters. In the present study, we showed that radiomic features extracted from MRI and CT images of glioblastoma patients are highly vulnerable against changes in image acquisition parameters, image sequences, and scanner models.

In the present work, we used COV, ICC, and CCC tests for reproducibility analysis. We defined the most reproducible features as COV ≤ 5%, ICC ≥ 90% and CCC ≥ 90%, based on previous studies [10, 15].

We examined how radiomic features change with different imaging modalities. Finding robust features among inter-modality variations is of interest. Since MRI acts as both anatomical/ functional imaging and CT images provide anatomical information, these features could be used as functional/anatomical image markers. Our results showed that shape features were the most reproducible features in terms of CCC and ICC among all features. Among other features, some textures Minimum (belong to first-order Features), IMC1, IDN, and IDMN (belong to GLCM) in terms of COV, and Run Length Non-Uniformity (belongs to Gray Level Run Length Matrix) in terms of ICC were found as the most robust features among changes in CT-MRI or MRI-MRI variations. Despite the fact that imaging acquisition in MRI and CT scan was different, some features in the shape feature category, including VoxelVolume, MeshVolume, SurfaceArea, and MinorAxisLength showed intermediate reproducibility between CT and MR-T1 images in terms of CCC and ICC. Some other studies have revealed that the shape features in GBM patients' MR images are very reproducible and stable against preprocessing [16].

One of the limitations of this study was its small patient sample size. Also, we did not develop a clinical model based on the imaging data. In a future study, the data from a larger patient population and with different grades will be used to model the treatment response and grading using the identified reproducible radiomic features.

Conclusion

In this study, we examined radiomic features variation among CT and MR images to identify the robust features. We concluded that the shape features, in terms of CCC and ICC, and some texture features were the most robust features that can be used for further clinical analysis.

Acknowledgment

The authors would like to acknowledge Dr. Reza Moghareabed, Dr. Iraj Abedi, and Fariba Davanian for their contributions to this study.

References

1. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. European journal of cancer. 2012 Mar 1;48(4):441-6.
2. Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, Schwartz LH. Reproducibility of radiomics for deciphering tumor phenotype with imaging. Scientific reports. 2016 Mar 24;6(1):1-7.
3. Meng Y, Sun J, Qu N, Zhang G, Yu T, Piao H. Application of radiomics for personalized treatment of cancer patients. Cancer management and research. 2019;11:10851.
4. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RG, Fillion-Robin JC, Pieper S, Aerts HJ. Computational radiomics system to decode the radiographic phenotype. Cancer research. 2017 Nov 1;77(21):e104-7.

5. Narang S, Lehrer M, Yang D, Lee J, Rao A. Radiomics in glioblastoma: current status, challenges and potential opportunities. Translational Cancer Research. 2016 Aug 1;5(4):383-97.
6. Peeken JC, Bernhofer M, Wiestler B, Goldberg T, Cremers D, Rost B, Wilkens JJ, Combs SE, Nüsslin F. Radiomics in radiooncology‐challenging the medical physicist. Physica medica. 2018 Apr 1;48:27-36.
7. Baeßler B, Weiss K, Dos Santos DP. Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. Investigative radiology. 2019 Apr 1;54(4):221-8.
8. Cattell R, Chen S, Huang C. Robustness of radiomic features in magnetic resonance imaging: review and a phantom study. Visual computing for industry, biomedicine, and art. 2019 Dec;2(1):1-6.
9. Park JE, Kim HS. Radiomics as a quantitative imaging biomarker: practical considerations and the current standpoint in neuro-oncologic studies. Nuclear medicine and molecular imaging. 2018 Apr;52(2):99-108.
10. Saeedi E, Dezhkam A, Beigi J, Rastegar S, Yousefi Z, Mehdipour LA, Abdollahi H, Tanha K. Radiomic feature robustness and reproducibility in quantitative bone radiography: a study on radiologic parameter changes. Journal of Clinical Densitometry. 2019 Apr 1;22(2):203-13.
11. Shiri I, Rahmim A, Ghaffarian P, Geramifar P, Abdollahi H, Bitarafan-Rajabi A. The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies. European radiology. 2017 Nov;27(11):4498-509.
12. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. Psychological reports. 1966 Aug;19(1):3-11.
13. Fisher RA. Statistical methods for research workers. InBreakthroughs in statistics 1992 (pp. 66-70). Springer, New York, NY.
14. Lawrence I, Lin K. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989 Mar 1:255-68.
15. Rastegar S, Beigi J, Saeedi E, Shiri I, Qasempour Y, Rezaei M, Abdollahi H. Radiographic Image Radiomics Feature Reproducibility: A Preliminary Study on the Impact of Field Size. Journal of medical imaging and radiation sciences. 2020 Mar 1;51(1):128-36.
16. Moradmand H, Aghamiri SM, Ghaderi R. Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma. Journal of applied clinical medical physics. 2020 Jan;21(1):179-90.