Reliability of 2D Magnetic Resonance Imaging Texture Analysis in Cerebral Gliomas: Influence of Slice Selection Bias on Reproducibility of Radiomic Features

Serebral Gliomlardaki İki Boyutlu Manyetik Rezonans Görüntüleme Yapısal Analizinin Güvenilirliği: Kesit Seçiminin Radyomik Özelliklerin Yeniden Üretilebilirliğine Etkisi

Burak Koçak

İstanbul Training and Research Hospital, Clinic of Radilogy, İstanbul, Turkey

ABSTRACT

Introduction: In this study, we aimed to investigate the reproducibility of two-dimensional (2D) texture features between adjacent magnetic resonance imaging (MRI) slices in patients with cerebral gliomas.

Methods: For this retrospective methodological study, T2weighted MRI and semi-automatic segmentation data of 25 patients with lower-grade gliomas were obtained from a public database. Only two regions of interests were used in this study: (i), the largest slice and (ii) one of the adjacent slices. Using PyRadiomics, an open source software to extract radiomic features from medical images, a total of 1116 texture features from six different feature classes were extracted from original, Laplacian of Gaussian-filtered, and wavelet-transformed images. Intra-class correlation coefficient (ICC) values with and without 95% confidence interval (CI) were used for reliability analysis. The ICC threshold for excellent reproducibility was 0.9.

Results: In the reliability analysis without considering the 95% CI for the ICC values, 28% of the texture features had excellent reproducibility. On the other hand, considering the 95% CI, only 10% of the texture features had excellent reproducibility. Neither a feature class (range of excellent reproducibility rates without 95% CI, 21.2%-34.4%; with 95% CI, 2.1%-18.3%) nor an image type (range of excellent reproducibility rates without 95% CI, 22.3%-41.9%; with 95% CI, 9.1%-14%) had considerable reliability in two adjacent MRI slices.

Conclusion: 2D MRI texture analysis of gliomas using T2-weighted sequence is substantially sensitive to slice selection bias, which may lead to non-reproducible results in radiomic works.

Keywords: Glioma, MRI, texture analysis, radiomics, reliability

ÖΖ

Amaç: Bu çalışmadaki amacımız serebral gliomlardaki iki boyutlu (2D) manyetik rezonans görüntüleme (MRG) yapısal analizinde kesit seçiminin radyomik özelliklerin yeniden üretilebilirliğine etkisini ardışık kesitler kullanarak araştırmak.

Yöntemler: Bu metodolojik çalışmaya halka açık bir veri bankasından düşük dereceli gliomu bulunan 25 hastanın T2 ağırlıklı MRG görüntüleri ve semi-otomatik segmentasyon verileri dahil edildi. Sadece iki ilgili bölge veya segmentasyon alanı kullanıldı: (i), en büyük kesitten elde edilen ve (ii), hemen komşuluğundaki kesitten elde edilen. Radyomik özellikler açık kaynak kodlu PyRadiomics yazılımı kullanılarak elde edildi. 6 farklı özellik sınıfından 3 farklı görüntü tipi kullanılarak toplamda 1116 yapısal özellik elde edildi. Güvenilirlik analizi %95 güven aralığı (GA) kullanılarak ve kullanılmadan sınıf içi katsayısı (SiK) ile yapıldı. Mükemmel yeniden üretilebilirlik için SiK eşik değeri 0,9 idi.

Bulgular: %95 GA kullanılmadan yapılan güvenilirlik analizinde, yapısal özelliklerin %28'i mükemmel yeniden üretilebilirliğe sahipti. Bunun yanında, %95 GA kullanılarak yapılan güvenilirlik analizinde ise, yapısal özelliklerin sadece %10'u mükemmel yeniden üretilebilirliğe sahipti. Ardışık MRG kesitlerinde ne bir özellik sınıfı (aralık, %95 GA kullanılmadan %21,2-%34,4; kullanılarak, %2,1-%18,3) ne de bir görüntü tipi (aralık, %95 GA kullanılmadan, %22,3-%41,9; kullanılarak, %9,1-%14) önemli düzeyde yeniden üretilebilirliğe sahipti.

Sonuç: Serebral gliomlardaki 2D T2 ağırlıklı MRG yapısal analizi kesit seçimine duyarlıdır. Bu durum dikkate alınmadığında radyomik çalışmalarda yeniden üretilebilirlik sorunlarına neden olabilir.

Anahtar Kelimeler: Gliom, MRG, yapısal analiz, radyomik, güvenilirlik



Address for Correspondence/Yazışma Adresi: Burak Koçak, İstanbul Training and Research Hospital, Clinic of Radilogy, İstanbul, Turkey

Phone: +90 530 821 08 44 E-mail: drburakkocak@gmail.com ORCID ID: orcid.org/0000-0002-7307-396X Cite this article as/Atuf: Koçak B. Reliability of 2D Magnetic Resonance Imaging Texture Analysis in Cerebral Gliomas: Influence of Slice Selection Bias on Reproducibility of Radiomic Features. istanbul Med J 2019; 20(5): 413-7.

©Copyright 2019 by the İstanbul Training and Research Hospital/İstanbul Medical Journal published by Galenos Publishing House. ©Telif Hakkı 2019 İstanbul Eğitim ve Araştırma Hastanesi/İstanbul Tıp Dergisi, Galenos Yayınevi tarafından basılmıştır. Received/Geliş Tarihi: 27.04.2019 Accepted/Kabul Tarihi: 29.07.2019

Introduction

The most common primary malignant cerebral tumors in adults are gliomas (1). Considering certain prognostic and therapeutic implications, gliomas can be divided into low-grade [World Health Organization (WHO)], grade I and grade II) and high-grade (WHO grade III and grade IV) based on histopathological and clinical criteria (2). Furthermore, WHO grade II and grade III diffuse gliomas are grouped as lower-grade gliomas, forming a heterogeneous group of tumors that have a wide range of malignancy characteristics (3). High-grade gliomas have very poor survival, while low-grade gliomas are associated with a longer life expectancy. Correct histopathological and genomic diagnosis of gliomas is crucial for appropriate treatment of all gliomas. Although biopsy is the gold standard for this purpose, this has been widely challenged by non-invasive conventional and advanced imaging techniques (4).

Texture analysis has been used for quantifying distribution and patterns of pixels or voxels in traditional or advanced medical images (5,6). In contrast to conventional qualitative and subjective clinical assessment, which might lead to significant variability depending on the experience of radiologists, image texture analysis offers an objective and more accurate non-invasive diagnosis that may influence patient management by more personalized management. Recently, texture analysis has been used in predicting histopathological tumor types, prognostic clinicopathological features, genomic characteristics and survival (7). However, the major problem of this field is the reproducibility of texture feature parameters, resulting in a challenge for creating powerful and stable predictive models to be used in clinical practice (8,9).

Although three-dimensional segmentation is the most representative for tumor texture, several studies have been published using a single image slice in the texture analysis of gliomas (10-13). However, this technique is prone to slice selection bias. To the best of our knowledge, the reproducibility of texture features between image slices has not been studied so far. In this study, we hypothesized that texture feature parameters obtained from different and even adjacent slices might not be correlated with each other and have a dependency to the selected slice. Therefore, in this study, we investigated the reproducibility of twodimensional (2D) texture features between adjacent magnetic resonance imaging (MRI) slices in patients with lower-grade gliomas.

Methods

Database Characteristics

No ethical approval was obtained for this retrospective methodological study because all patients included in this study were publicly and freely available for scientific purposes in the cancer imaging archive (TCIA) (14). The imaging and segmentation data of the patients used in this study were obtained from the collection named "LGG-1p19qDeletion" in TCIA (14-16).

One hundred and fifty-nine patients in the collection were reviewed for identifying patients with a uniform image acquisition protocol and no signs of previous surgery or biopsy, which would influence texture feature parameters. Following the initial evaluation of the collection, a randomly selected subset of 25 patients with MRI and tumor segmentation data was included in this reproducibility study.

MRI Acquisition Parameters

Only T2-weighted spin-echo MRI images were included in the study. The images were obtained using a 1.5 Tesla MRI unit. Acquisition parameters were uniform, except for time to echo and repetition time. The representative acquisition parameters were as follows: time to echo, 98; repetition time, 4000; slice thickness, 3 mm; pixel spacing, 0.937x0.937 mm²; echo train length, 16; and acquisition matrix, 256x256.

Texture Feature Extraction

Before feature extraction, the low-frequency signals that would corrupt MRI images were corrected in all images using N4 bias field correction algorithm (17). Then, gray-level intensity values were normalized and discretized (18,19). Normalization procedure was performed using the ± 3 sigma technique based on the following mathematical formula:

$$f(x) = \frac{s(x - \mu(x))}{\sigma(x)}$$

where f (x) is normalized gray-level intensity, x is original gray-level intensity, μ (x) is mean gray-level intensity value, σ (x) is the standard deviation of gray-level intensity, and s is the scaling factor, which was 100 in this study.

The discretization was based on the following mathematical formula:

$$X_{b,i} = \lfloor \frac{X_{gl,i}}{W} \rfloor - \lfloor \frac{\min(X_{gl,i})}{W} \rfloor + 1$$

where, $X_{b,i}$ is gray-level intensity following discretization; $X_{gl,i}$ is gray-level intensity prior to discretization; W is the bin-width value, which was five in this study.

The segmentation datasets were obtained from TCIA (14-16), which was used in the study of Akkus et al. (15). The segmentations had been done with semi-automatic fashion and based on normal brain atlas, the posterior probability of the voxels, and geodesic active contour (20-22). Original segmentation data included three image slices. Nonetheless, we only used the largest slice and one of the adjacent slices in the radiomic analysis. Segmentation style and usage in this study are presented in Figure 1.



Figure 1. Segmentation style in a 47-year-old male with a grade III diffuse glioma. (a) T2-weighted axial slice shows the right-sided frontal tumor with extension to the left side. (b) Reconstructed coronal T2-weighted MRI image presents the segmentation data decomposed to three consecutive slices. Two axial slices including middle (green-colored segmentation) and upper (red-colored segmentation) or lower (blue-colored segmentation) ones were included in this study. (c) Axial T2-weighted image shows one of the two selected segmentations

MRI: magnetic resonance imaging

Texture features were extracted from the two adjacent MRI slices using PyRadiomics software (PyRadiomics version 2.0.1; Numpy version 1.13.1; SimpleITK version 1.1.0.dev370; PyWavelet version 0.5.2; Python version 2.7.13) (23). Using the original image, the extracted texture feature groups were as follows: (i), 18 first-order features; (ii), 14 gray-level dependence matrix (GLDM) features; (iii), 24 gray-level co-occurrence matrix (GLCM) features; (iv), 16 gray-level run length matrix (GLRLM) features; (v), 16 gray-level size zone matrix (GLSZM) features; and (vi), 5 neighboring gray-tone difference matrix (NGTDM) features. In addition to the original image, we also used Laplacian of Gaussian (LoG)-filtered and wavelet-transformed images in extracting texture features. The LoG filter was used for image filtration with values of 2 mm, 4 mm, and 6 mm; where, 2 mm, 4 mm, and 6 mm represent fine, medium, and coarse patterns, respectively. Wavelet-based texture features were created using eight different frequency band combinations. The total number of the features extracted was 1116 [93 from the original image; 279 (93x3) from LoG-filtered images; and 744 (93x8) from wavelet-transformed images] per lesion. Detailed definitions and mathematical formulas for these features have been described in the website of PyRadiomics in detail, https://pyradiomics.readthedocs.io/en/latest/.

Statistical Analysis

The statistical analysis was performed using SPSS version 20 (SPSS Inc.). The degree of correlation and agreement of quantitative texture features between MRI slices were assessed using intra-class correlation coefficient (ICC) (24). For the ICC analysis, we used a two-way model, single-rating, and absolute agreement. The strength of reproducibility was defined as follows: (i), ICC<0.9, not excellent reproducibility; and (ii), ICC>0.9, excellent reproducibility (24). The reproducibility was assessed using the ICC values with and without considering 95% confidence interval (CI).

Results

Overall Reproducibility

In the analysis performed without considering 95% CI for the ICC values, approximately one-fourth of the texture features were excellently reproducible (Figure 2a). On the other hand, considering the 95% CI, only one-tenth of the texture features were excellently reproducible (Figure 2b).



Figure 2. Pie charts show the overall reproducibility rates of 2D texture features between adjacent MRI slices without (a) and with (b) considering 95% confidence interval

MRI: magnetic resonance imaging

Reproducibility Based on Image Types

In the analysis without considering the 95% CI for the ICC values, approximately less than half of the texture features extracted from the original and LoG-filtered image types were excellently reproducible. Nonetheless, for the wavelet-transformed images, approximately one-fourth of the features were excellently reproducible (Figure 3a).

In the analysis with considering the 95% CI for the ICC values, approximately only one-tenth of the texture features extracted from the original, LoG-filtered, and wavelet-transformed image types were excellently reproducible (Figure 3b).

Reproducibility Based on Feature Classes

In the analysis without considering the 95% CI for the ICC values, the feature classes with the highest and lowest rates for excellent reproducibility were GLRLM and GLCM, respectively. For the first-order and GLCM feature classes, approximately one-fourth of the texture features were excellently reproducible. Meanwhile, for the other feature classes (GLDM, GLRLM, GLSZM, NGTDM), approximately one-third of the features were excellently reproducible (Figure 4a).

In the analysis with considering the 95% CI for the ICC values, the feature classes with the highest and lowest rates for excellent reproducibility were NGTDM and GLCM, respectively. For the first-order and GLCM features, only less than one-tenth of the texture features were excellently reproducible. Meanwhile, for the other groups (GLDM, GLRLM, GLSZM, NGTDM), approximately less than one-fifth of the texture features were excellently reproducible (Figure 4b).



Figure 3. Bar charts present the reproducibility rates of 2D texture features between adjacent MRI slices based on image types without (a) and with (b) considering 95% confidence interval

MRI: magnetic resonance imaging, LoG: laplacian of gaussian



Figure 4. Bar charts present the reproducibility rates of 2D texture features between adjacent MRI slices based on feature classes without (a) and with (b) considering 95% confidence interval.

GLCM: gray-level co-occurrence matrix features, GLDM: gray-level dependence matrix features, GLRLM: gray-level run length matrix features, GLSZM: gray-level size zone matrix features, NGTDM: neighboring gray-tone difference matrix features

Discussion

In this study, we investigated the reproducibility of 2D texture-based radiomic feature parameter values between two adjacent conventional T2-weighted MRI slices in lower-grade (WHO grade II and III) glioma patients. The vast majority of high-dimensional texture features were not excellently correlated between adjacent T2-weighted MRI slices. Neither a feature class nor an image type had considerable reliability in two adjacent MRI slices.

To obtain reliable values in a quantitative method, the parameter values obtained must be resistant to various factors such as segmentation variability, acquisition differences or use of different scanners from different vendors. Although much work has been done using 2D MRI texture analysis in cerebral gliomas (10-13), there is a scarcity of papers regarding the reliability of the technique. Only few papers draw our attention to the in vivo stability of the texture feature parameters. The most significant of those is a methodological study dealing with volume bias, slice bias, and region of interest bias in glioblastomas (9). Although it has been conducted with a very limited number of features, in their seminal work, the authors suggested that increasing fractal tumor volume and even a minimal change of a region of interest area significantly influence the texture feature parameters, providing evidence regarding susceptible nature of the texture analysis. However, the stability of parameters across different slices has not been studied so far. Therefore, a direct comparison of this study with others is not possible.

We think that our study has very significant pre-clinical and clinical implications. In general, a texture-based high-dimensional radiomic workflow includes a few crucial steps as follows: (i), preprocessing of the images; (ii), segmentation of the tumors or lesions; (iii), radiomic feature extraction; (iv), dimension reduction to avoid redundant features, which is optional: and (v), statistical model development using conventional or advanced methods (25). The segmentation step is known to be the most critical and challenging one in radiomic works (6). Therefore, our focus in this work was on the segmentation step with a different perspective, that is, slice selection bias. The most important implication of our work was that 2D MRI texture analysis would lead to non-reproducible feature parameter values due to the high susceptibility of the texture analysis to the slice of interest or slice selection bias. Therefore, the 2D MRI texture analysis using a single slice must be used cautiously in radiomic workflows. If this technique is used in gliomas, a reliability analysis regarding the slice selection bias should be included in the radiomic workflow to exclude the features with poor reproducibility.

A few limitations to this methodological study need to be acknowledged. First, the nature of the study was retrospective, which was disadvantageous due to dependency on limited data. Second, although the image acquisition protocol is fairly uniform, we had to perform a few preprocessing steps to minimize small differences like bias field, the number of gray levels, and relative gray-level intensity range (18,19). It is worth to emphasize that the texture analysis has a dependency on these preprocessing steps to obtain comparable parameters (18,19). For this reason, all of the MRI images in our study underwent N4 bias field correction, gray-level normalization, and gray-level discretization (17-19). We did not consider pixel rescaling because it was homogeneous in all patients. Third, we included only T2-weighted MRI images, because they are widely used in radiomic works (26,27). This study can be expanded using other sequences in future studies. Fourth, we only included lowergrade tumors (WHO grade II and III) to represent gliomas. Nonetheless, whether our findings might be extrapolated to other gliomas should be further studied. Fifth, a Bland-Altman analysis could have been included as a statistical method to reveal the degree of agreement between the slices. Instead, we used the ICC in this study, which can serve as a single strong metric not only for the degree of correlation but also for the agreement between quantitative measurements (24).

Conclusion

2D MRI texture analysis of gliomas was substantially susceptible to selected slices, which may lead to non-reproducible results in radiomic works. The vast majority of high-dimensional texture features were not excellently correlated between adjacent T2-weighted MRI slices. Neither a feature class nor an image type had considerable reliability in two adjacent MRI slices. Therefore, a reliability analysis with considering different slices must be incorporated into every scientific research using this technique. Otherwise, the unstable feature parameters might cause non-reproducible outcomes in terms of selected texture features and statistical predictive models.

Ethics Committee Approval: Not required due to use of public data.

Informed Consent: Not required due to use of public data.

Peer-review: Internally peer-reviewed.

Financial Disclosure: The authors declared that this study received no financial support.

References

- Cha S. Update on brain tumor imaging. Curr Neurol Neurosci Rep 2005; 5: 169-77.
- 2. Perry A, Ellison DW, Reifenberger G, Kleihues P, von Deimling A, Figarella-Branger D, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol 2016; 131: 803-20.
- Cuccarini V, Erbetta A, Farinotti M, Cuppini L, Ghielmetti F, Pollo B, et al. Advanced MRI may complement histological diagnosis of lower grade gliomas and help in predicting survival. J Neurooncol 2016; 126: 279-88.
- Pope WB, Brandal G. Conventional and advanced magnetic resonance imaging in patients with high-grade glioma. Q J Nucl Med Mol Imaging 2018; 62: 239-53.
- 5. Lubner MG, Smith AD, Sandrasegaran K, Sahani D V, Pickhardt PJ. CT Texture Analysis: Definitions, Applications, Biologic Correlates, and Challenges. RadioGraphics 2017; 37: 1483-503.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, They Are Data. Radiology 2016; 278: 563-77.
- 7. Leng Y, Wang X, Liao W, Cao Y. Radiomics in gliomas: A promising assistance for glioma clinical research. J Cent South Univ (Medical Sci 2018; 43: 354-9.
- Mansilla Legorburo F, Pastor-Juan M del R, Sabater S, Canales-Vázquez J, Villas MV, Berenguer R, et al. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. Radiology 2018; 288: 407-15.
- 9. Hainc N, Stippich C, Stieltjes B, Leu S, Bink A. Experimental texture analysis in glioblastoma. Invest Radiol 2017; 52: 367-73.
- Eliat PA, Olivié D, Saïkali S, Carsin B, Saint-Jalmes H, De Certaines JD. Can dynamic contrast-enhanced magnetic resonance imaging combined with texture analysis differentiate malignant glioneuronal tumors from other glioblastoma? Neurol Res Int 2012; 2012: 195176.
- Yang D, Rao G, Martinez J, Veeraraghavan A, Rao A. Evaluation of tumorderived MRI-texture features for discrimination of molecular subtypes and prediction of 12-month survival status in glioblastoma. Med Phys 2015; 42: 6725-35.
- 12. Nakagawa M, Nakaura T, Namimoto T, Kitajima M, Uetani H, Tateishi M, et al. Machine learning based on multi-parametric magnetic resonance imaging to differentiate glioblastoma multiforme from primary cerebral nervous system lymphoma. Eur J Radiol 2018; 108: 147-54.

- Dormagen JB, Ganeshan B, Server A, Schulz A, Skogen K, Helseth E. Texture analysis on diffusion tensor imaging: discriminating glioblastoma from single brain metastasis. Acta radiol 2018; 60: 028418511878088
- Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013; 26: 1045-57.
- Akkus Z, Ali I, Sedlár J, Agrawal JP, Parney IF, Giannini C, et al. Predicting Deletion of Chromosomal Arms 1p/19q in Low-Grade Gliomas from MR Images Using Machine Intelligence. J Digit Imaging 2017; 30: 469-76.
- 16. Erickson B, Akkus Z, Sedlar J, Korfiatis P. Data From LGG-1p19qDeletion. The Cancer Imaging Archive. Epub ahead of print 2017.
- 17. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: Improved N3 bias correction. IEEE Trans Med Imaging 2010; 29: 1310-20.
- Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn Reson Imaging 2004; 22: 81-91.
- 19. Shafiq-ul-Hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 2017; 44: 1050-62.
- Agrawal J, Coufalova L, Warner JD, Korfiatis P, Sedlar J, Erickson BJ, et al. Semiautomated segmentation of pre-operative low grade gliomas in magnetic resonance imaging. Cancer Imaging 2015; 15: 12.
- 21. Rohlfing T, Zahr NM, Sullivan E V., Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp 2010; 31: 798-819.
- Marquez-Neila P, Baumela L, Alvarez L. A morphological approach to curvature-based evolution of curves and surfaces. IEEE Trans Pattern Anal Mach Intell 2014; 36: 2-17.
- 23. Hosny A, van Griethuysen JJM, Parmar C, Aerts HJWL, Fedorov A, Beets-Tan RGH, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 2017; 77: 104-7.
- 24. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016; 15: 155-63.
- Kocak B, Ates E, Durmaz ES, Ulusan MB, Kilickesmez O. Influence of segmentation margin on machine learning–based high-dimensional quantitative CT texture analysis: a reproducibility study on renal clear cell carcinomas. Eur Radiol 2019; 1-11.
- Kinoshita M, Sakai M, Arita H, Shofuda T, Chiba Y, Kagawa N, et al. Introduction of high throughput magnetic resonance T2-weighted image texture analysis for WHO grade 2 and 3 gliomas. PLoS One 2016; 11: e0164268
- 27. Li Y, Liu X, Qian Z, Sun Z, Xu K, Wang K, et al. Genotype prediction of ATRX mutation in lower-grade gliomas using an MRI radiomics signature. Eur Radiol 2018; 28: 2960-8.