Test–Retest Reproducibility Analysis of Bone Mineral Densitometry Radiomics Features

Younes Qasempour
Kerman University of Medical Sciences
Mohammad Rashid Akhash
Kerman University of Medical Sciences
Isaac Shiri
Hopitaux Universitaires de Geneve
Ghasem Hajianfar
Iran University of Medical Sciences School of Behavioral Sciences and Mental Health
Neda Abdalvand
Iran University of Medical Sciences
H Abdollahi (hamid_rbp@yahoo.com)
Kerman University of Medical Sciences https://orcid.org/0000-0003-0761-1309

Research article

Keywords: Bone mineral densitometry, DXA, Test–retest reproducibility, Quantitative radiomics features, ICC

Posted Date: July 27th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-44885/v1

Abstract

Background

Radiomics features reproducibility assessment is a critical issue in imaging biomarker development era. In the present study, we aimed to assess test–retest reproducibility analysis of bone mineral densitometry (BMD) image radiomics features.

Methods

In this prospective research work, eighteen patients were included and were subjected to DXA BMD scans acquired within 10 min of each other under an approved protocol. Seven regions of interest (ROIs) including four lumbar spine regions (L1-L4) and three hip regions (trochanteric, inter trochanteric and neck) in both test and re-test images were segmented and 107 radiomics features from seven different feature sets were extracted. Intra-class correlation coefficient (ICC) were initially used to estimate radiomics features reproducibility.

Results

We showed that there is no radiomics feature with 90% < ICC < 100% in all ROIs, but there are three feature including Strength (from NGTDM feature set), SALGLE (Small Area Low Gray Level Emphasis) (from GLSZM feature set) and Busyness (from NGTDM feature set) with ICC < 70% in all eight ROIs. Shape features has features with ICC < 70%.

Conclusion

Our study on test–retest reproducibility analysis of bone mineral densitometry radiomics features shows radiomics features have several variations against changes of time of image acquisition. The reproducible features may be used as imaging biomarkers in the field of clinical densitometry. The results of this study may be repeated by more radiomics features and more BMD scanners as first line for bone mineral biomarker discovery.

Background

Bone mineral density (BMD) assessment by using dual-energy X-ray absorptiometry (DXA) is the gold standard for diagnosis, monitoring and follow-up of bone mineral loss deficiencies such as osteoporosis and osteopenia. Generally speaking, DXA plays three main roles including diagnosis of osteoporosis, fracture risk assessment, and monitoring of treatment response (1). Studies have indicated that BMD measurement in hip and lumbar regions is the most reliable measurement for predicting hip fracture risk and spinal therapy monitoring respectively. In addition, DXA has several advantages including low scan

times, easy to use, low radiation dose, and good measurement precision. However, although, DXA is a feasible approach for BMD assessment, it suffers from some limitations. For example, it is a two-dimensional (2D) imaging approach and areal density measurement is affected by bone size as well as the true 3D volumetric density of the bone tissue (2).

In addition to DXA, several other approaches are developed for BMD assessment. These methods are including quantitative computed tomography (QCT), peripheral DXA (pDXA), quantitative ultrasound (QUS) and magnetic resonance imaging (MRI) (3–5). Recently, quantitative radiomics texture analysis studies have been applied for a wide range of clinical applications such as diseases detection, diagnosis, prognosis and prediction (6–8). Although radiomics is used comprehensively for cancer studies, it is applicable for bone diseases management such as osteoporosis diagnosis and therapy response prediction and assessment. Previous studies have identified that radiomics features extracted from computed tomography (CT), magnetic resonance imaging (MRI) and radiology images could be used for bone diseases management (9, 10). In our previous study, we showed the feasibility of radiomics features extracted from DXA images to classify osteoporosis and osteopenia patients from normal ones (11).

Radiomics is an advanced image processing approach with several steps including image acquisition, segmentation, feature extraction, feature selection and data modelling (12–14). In radiomics approach, the extracted features could be served as imaging biomarker for diagnosis, prognosis, response prediction and assessment of therapy for several diseases. With this regard, radiomics features have to be reproducible, means that radiomics feature values should stay unchanged or minimally changed when the feature is computed from a repeat scan acquired after a short time interval. A wide range of studies have reported that radiomics features are vulnerable against changes in image acquisition, reconstruction, pre-processing, segmentation and analysis (15–17). In this light and to decrease false positive rate, robust radiomics features have to be found and used for further clinical applications.

Based on the previous radiomics studies, a reproducible/repeatable feature remain the same in different radiomics processes settings or in same radiomics processes, but in different times. To the best of our knowledge, there are no reports on BMD radiomics reproducibility over changes in the times of image accusation. In the present study, we aimed to assess the test-retest reproducibility of radiomics features extracted from DXA images. In this study, for first time, reproducibility of image features was checked in several regions of BMD images in two consecutive times. The reproducibility was calculated based on the statistical tests.

Methods

Patients

This research work was conducted as a prospective study and approved by local ethics committee. In this work, 18 patients that referred to bone densitometry department were included. For each patient, informed consent was taken before examination. All patients underwent two DXA within 10 minutes before any

treatment was delivered. Both scans were obtained with the same DXA scanner by using the same imaging protocol (Lexxos, 4.7 mm Al 75 kVp and 14.2 mm Al 140 kVp). These patients' medical images were divided into two groups: test and re-test.

Contouring

In this work, seven anatomical regions were contoured by a nine years experienced radiologist in the field of bone densitometry. All contours were done by ITK SNAP (18) and then all images and contours were exported for feature extracting and analysis. Differnt contouring was shown in Fig. 1. In this contouring four regions from lumbar spinal including L1, L2, L3 and L4 and three regions from hip including neck (NK), trochanteric (TR) and inter trochanteric (IT) were drawn for radiomics analysis. These regions are analyzed in routine clinical BMD assessment. In addition, all regions were analyzed as a separate unique region, called ALL region (combinations of all regions). These contouring were same in both test and re-tests images.

Radiomics

.. Totally 107 radiomics features were extracted by using Pyradiomics radiomics feature extraction platform (19). These features were listed in supplementary table 1 and were divided into seven groups: 1) first order (FO) (18 features), 2) gray level co-occurrence matrix (GLCM) (24 features), 3) gray-level different matrix (GLDM) (14 features), 4) gray level run-length matrix (GLRLM) (16 features), 5) gray-level size zone length matrix (GLSZM) (16 features), 6) neighborhood gray-tone difference matrix (NGTDM) (5 features) and 7) Shape (14 features).

Data analysis

To estimate the reproducibility radiomics features by using repeat DXA data, the intra-class correlation coefficient (ICC) were initially used. ICC was calculated by the following equation:

$$ICC = \frac{MS_R - MS_W}{MS_R + (k-1)MS_W}$$

Where MSR = mean square for rows, MSW = mean square for residual source of variance, k = number of observers involved and n = number of subjects. In this study, radiomics feature reproducibility were categorized in four categories including 1) ICC < 70%, 2) 70% < ICC < 80%, 3) 80% < ICC < 90% and 4) 90% < ICC < 100%. ICC were shown as Heatmap, box plot, bar plot and density distribution plot.

In addition, Bland-Altman graphs were drawn for all features and some graphs were presented. This graphical method was used to quantify the agreement between two radiomics features by studying the mean difference within which 95% of the difference between the retest features in comparison to the test features. R package version 3.1.3 IRR was used for all statistical analysis.

Results

The results of this study were shown as clustering heatmap, bar plot, box plot, density distribution and Bland-Altman graphs. In Fig. 2, the ICC for all radiomics features in all ROIs were depicted as a heatmap based on four ICC categories. As was seen, there is no radiomics feature with 90% < ICC < 100% in all eight ROIs, but there are three feature including Strength (from NGTDM feature set), SALGLE (Small Area Low Gray Level Emphasis) (from GLSZM feature set) and Busyness (from NGTDM feature set) with ICC < 70% in all eight ROIs. Shape features has features with ICC < 70%.

In Fig. 3, the bar plot shows the percentage of all radiomics features reproducibility from all regions based on the four ICC categories. As was depicted, 75% of 12 radiomic features have the 90% < ICC < 100%. These features are 90-Percentile, Energy, RMS (Root Mean Squared), Median and TE (Total Energy) from first order feature set and Flatness, Least Axis, Mesh Volume, Minor Axis, Sphericity, Surface Area and Voxel Volume from Shape feature set. There is a feature namely Strength from NGTDM feature set that has 100% ICC < 70% in all regions.

The results for percentage of all radiomic features reproducibility based on the four ICC categories for each ROI and for combination of all ROIs was shown in Fig. 4. It is observed that radiomics features from combination of all ROIs (ALL) and L4 region have the highest percentage of radiomic features with 90% < ICC < 100% followed by TR and NK regions. In addition, L3 and IT regions have the highest percentage of radiomics features with ICC < 70%.

We showed the ICC density distribution in Fig. 5. In this plot, it was seen that there are two peaks of high ICC in L4 region and combination of all ROIs (ALL). For L1 and NK regions, also there are peaks in ICC more than 75%. The lowest ICC peak is for radiomics features that extracted from L2 region. In Fig. 6, the box plot of ICC is shown for all ROIs. As was depicted, the maximum ICC is for combination of all ROIs (ALL), followed by L4, TR and L2.

The Bland-Altman analysis for all feature sets was shown in Figs. 7–13. In these figures, features with ICC more than 0.85% were depicted except NGTDM feature set that didn’t have any reproducible features and features with highest ICC were shown. In radiomics reproducibility analysis based on the Bland-Altman, mean, standard deviation (SD), and upper/lower reproducibility limits (U/LRL) are considered to find reproducible radiomics features.

Discussion

Bone mineral densitometry is a critical medical approach to assess, monitor and predict several bone mineral diseases. It is a widely accepted examination with several advantageous such as, easy to use and data reporting, low cost and low radiation dose. But it has some limitations and challenges and new approaches are being studied for bone density assessment. Radiomics is an advanced quantitative imaging approach that could be used for bone diseases diagnosis, prognosis and prediction. In this approach, radiomics features extracted from medical images are used as biomarkers for clinical decision making. However, before any clinical applications, they have to be assessed in terms of reproducibility and repeatability (20).

In the present study, we assessed test-retest reproducibility of radiomics features extracted from DXA BMD images. We observed that radiomics features based on the region being interested, have several changes over changes in the time of image acquisition.

In our study, we found that most radiomics features extracted from BMD images are vulnerable against changes in the time of acquisition. Based on the results presented in the heatmap figure, we observed that Shape and First Order (FO) have the highest most radioproduicble features (90% < ICC < 100%) and GLSZM and NGTDM have the highest less radioproduicble features ICC < 70%. Our results are in concordance with some previous studies. On the test-retest reproducibility of radiomics feature, several studies have conducted. Balagurunathan et al. investigated the test–retest reproducibility analysis of 219 image features extracted from lung CT acquired within 15 min of each other. They found that 66 features (30.14%) with concordance correlation coefficient ≥ 0.90 across test–retest (21). Hu et al. studied the reproducibility with repeat CT in radiomics study for rectal cancer. 775 image features were extracted from two repeated CT scans within average 8.7 days. The results showed that 496 features have high reproducibility (ICC ≥ 0.8), 225 medium reproducibility (0.8 > ICC ≥ 0.5) and 54 low reproducibility (ICC < 0.5) (22).

The mechanisms of radiomics feature changes are not fully understood. But some issues such as the level of different noises, patient's status changes, operational errors, radiation scattering, electronical changes in the scanner, the nature of radiomics features and changes in the biology may induce radiomics features variations (20, 22). In this study, the time between two image acquisitions was ten minutes, this time may not induce any change in the bone status. In previous studies on the cancer image radiomics reproducibility, more times were used between two imaging and this issue may impacts the biology of the tumors and therefore radiomics feature values. Also, in some cases, features were compared before and after a treatment and changes in the features are assessed and are used as therapy monitoring.

This study suffers from some limitations. First, low sample size. Eighteen patients are low, and more sample sizes will be resulted in more reliable results. Second, this study was conducted by one scanner. Further studies with different scanners from different vendors are needed to found new results. Third, segmentation variation. Our segmentation was done manually. This issue may induce some variations in results. We suggest automatic or semi-automatic segmentation and compare the results. We also, suggest that the results of this study may compared by another imaging modalities such as MRI, CT or radiology.

Conclusions

In summary, our study on test–retest reproducibility analysis of bone mineral densitometry radiomics features shows radiomics features have several variations against changes of time of image acquisition. The reproducible features may be used as imaging biomarkers in the field of clinical densitometry. The

results of this study may be repeated by more radiomics features and more BMD scanners as first line for bone mineral biomarker discovery.

Abbreviations

BMD
Bone mineral density, DXA:X-ray absorptiometry, 2D:Two-dimensional, QCT:Quantitative computed tomography, pDXA:Peripheral DXA, QUS:quantitative ultrasound, MRI:Magnetic Resonance Imaging, CT:computed tomography, NK:neck, TR:trochanteric, IT:Inter trochanteric, FO:First order, GLCM:Gray level co-occurrence matrix, GLRLM:Gray-level different matrix, GLRLM:Gray level run-length matrix, GLSZM:Gray-level size zone length matrix, NGTDM:Neighborhood gray-tone difference matrix, ICC:Intra-class correlation coefficient, MSR:Mean square for rows, MSW:Mean square for residual source

Declarations

Ethics approval and consent to participate

All patients provided written informed consent prior to participation and the study was approved by the local ethics committee of Kerman University of Medical Sciences, Kerman, Iran (Ref NO 13990404), and conducted according to the Declaration of Helsinki.

Consent for publication

No applicable.

Availability of data and material

The datasets of the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

There was no funding support for this study.

Authors' contributions

YQ and MRA collected the data and managed patents. IS and GH performed the data analysis. NA interpreted the data. HA designed, conceptualized the study and drafted the manuscript.

Acknowledgements

The authors would like to acknowledge the Radiology Staffs at Rafsanjan University of Medical Sciences Hospitals who supported patient recruitment and data acquisition.

Conflict of interest statement:

The authors declare that they have no conflict of interest.

References

1. Blake GM, Fogelman I. An update on dual-energy x-ray absorptiometry. Semin Nucl Med. 2010;40(1):62–73.
2. Lewiecki EM. Benefits and limitations of bone mineral density and bone turnover markers to monitor patients treated for osteoporosis. Curr Osteoporos Rep. 2010;8(1):15–22.
3. Engelke K, Adams JE, Armbrecht G, Augat P, Bogado CE, Bouxsein ML, et al. Clinical use of quantitative computed tomography and peripheral quantitative computed tomography in the management of osteoporosis in adults: the 2007 ISCD Official Positions. J Clin Densitom. 2008;11(1):123–62.
4. Nardone V, Tini P, Carbone S, Grassi A, Biondi M, Sebaste L, et al. Bone texture analysis using CT-simulation scans to individuate risk parameters for radiation-induced insufficiency fractures. Osteoporos Int. 2017;28(6):1915–23.
5. Wolski M, Podsiadlo P, Stachowiak G, Englund M, Haugen I. Trabecular bone texture detected by plain radiography is associated with MRI-defined osteophytes in finger joints of women without radiographic osteoarthritis. Osteoarthritis Cartilage. 2018;26(7):924–8.
6. Abdollahi H. Radiotherapy dose painting by circadian rhythm based radiomics. Med Hypotheses. 2019;133:109415.
7. Abdollahi H, Mahdavi SR, Mofid B, Bakhshandeh M, Razzaghdoust A, Saadipoor A, et al. Rectal wall MRI radiomics in prostate cancer patients: prediction of and correlation with early rectal toxicity. Int J Radiat Biol. 2018;94(9):829–37.
8. Abdollahi H, Mofid B, Shiri I, Razzaghdoust A, Saadipoor A, Mahdavi A, et al. Machine learning-based radiomic models to predict intensity-modulated radiation therapy response, Gleason score and stage in prostate cancer. Radiol Med. 2019;124(6):555–67.
9. Abdollahi H, Mahdavi SR, Shiri I, Mofid B, Bakhshandeh M, Rahmani K. Magnetic resonance imaging radiomic feature analysis of radiation-induced femoral head changes in prostate cancer radiotherapy. J Cancer Res Ther. 2019;15(Supplement):11-s9.
10. MacKay JW, Murray PJ, Kasmai B, Johnson G, Donell ST, Toms AP. Subchondral bone in osteoarthritis: association between MRI texture analysis and histomorphometry. Osteoarthritis Cartilage. 2017;25(5):700–7.
11. Rastegar S, Vaziri M, Qasempour Y, Akhash M, Abdalvand N, Shiri I, et al. Radiomics for classification of bone mineral loss: A machine learning study. Diagn Interv Imaging. 2020.

12. Abdollahi H, Mostafaei S, Cheraghi S, Shiri I, Rabi Mahdavi S, Kazemnejad A. Cochlea CT radiomics predicts chemoradiotherapy induced sensorineural hearing loss in head and neck cancer patients: A machine learning and multi-variable modelling study. Phys Med. 2018;45:192–7.
13. Abdollahi H, Shiri I, Heydari M. Medical Imaging Technologists in Radiomics Era: An Alice in Wonderland Problem. Iran J Public Health. 2019;48(1):184–6.
14. Abdollahi H, Tanha K, Mofid B, Razzaghdoust A, Saadipoor A, Khalafi L, et al. MRI Radiomic Analysis of IMRT-Induced Bladder Wall Changes in Prostate Cancer Patients: A Relationship with Radiation Dose and Toxicity. J Med Imaging Radiat Sci. 2019;50(2):252–60.
15. Mostafaei S, Abdollahi H, Kazempour Dehkordi S, Shiri I, Razzaghdoust A, Zoljalali Moghaddam SH, et al. CT imaging markers to improve radiation toxicity prediction in prostate cancer radiotherapy by stacking regression algorithm. La Radiologia medica. 2019.
16. Saeedi E, Dezhkam A, Beigi J, Rastegar S, Yousefi Z, Mehdipour LA, et al. Radiomic Feature Robustness and Reproducibility in Quantitative Bone Radiography: A Study on Radiologic Parameter Changes. J Clin Densitom. 2019;22(2):203–13.
17. 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. Eur Radiol. 2017;27(11):4498–509.
18. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage. 2006;31(3):1116–28.
19. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer research. 2017;77(21):e104-e7.
20. Rastegar S, Beigi J, Saeedi E, Shiri I, Qasempour Y, Rezaei M, et al. Radiographic Image Radiomics Feature Reproducibility: A Preliminary Study on the Impact of Field Size. J Med Imaging Radiat Sci. 2020.
21. Balagurunathan Y, Kumar V, Gu Y, Kim J, Wang H, Liu Y, et al. Test–retest reproducibility analysis of lung CT image features. J Digit Imaging. 2014;27(6):805–23.
22. Hu P, Wang J, Zhong H, Zhou Z, Shen L, Hu W, et al. Reproducibility with repeat CT in radiomics study for rectal cancer. Oncotarget. 2016;7(44):71440.

Figures

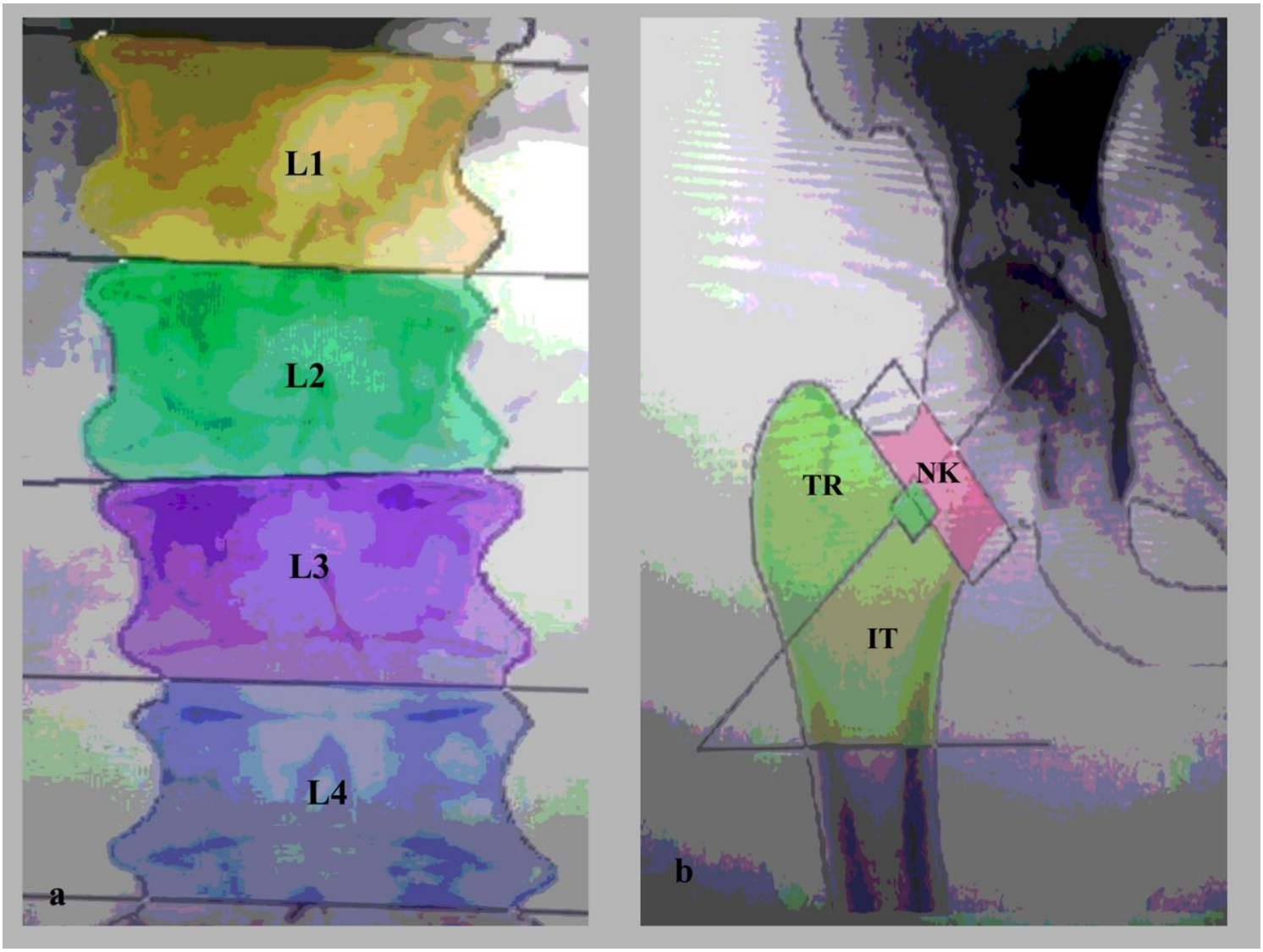

Figure 1

Regions of interest (ROI) identification for radiomics study. L; Lumbar; NK: Neck; TR: Trochanteric and IT: Inter trochanteric.

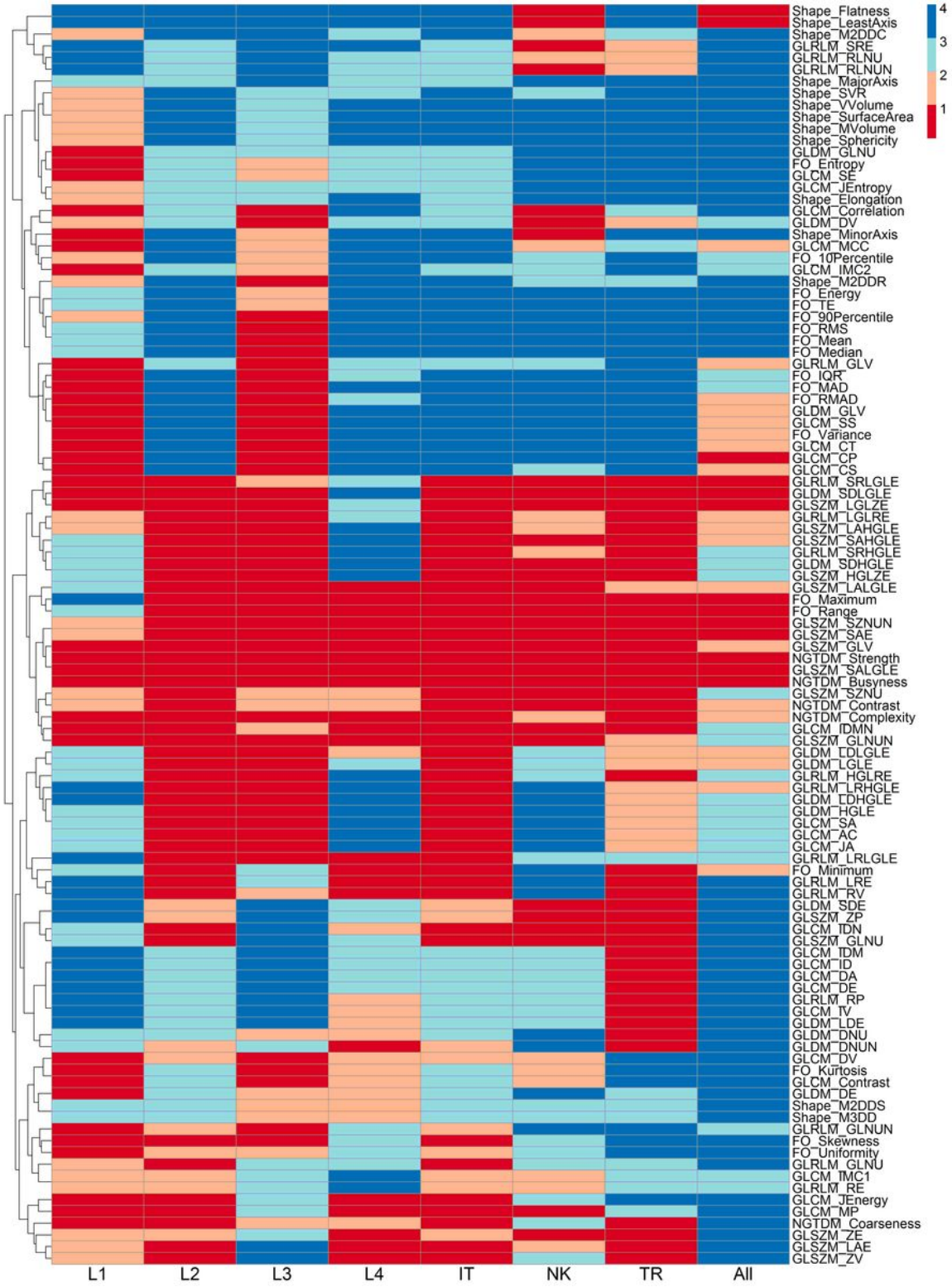

Figure 2

The clustering Heatmap showing radiomics features reproducibility based on the ICC for all regions. The number 1 to 4 are: 1) ICC < 70%, 2) 70% < ICC < 80%, 3) 80% < ICC < 90% and 4) 90% < ICC < 100%. L; Lumbar; NK: Neck; TR: Trochanteric and IT: Inter trochanteric.

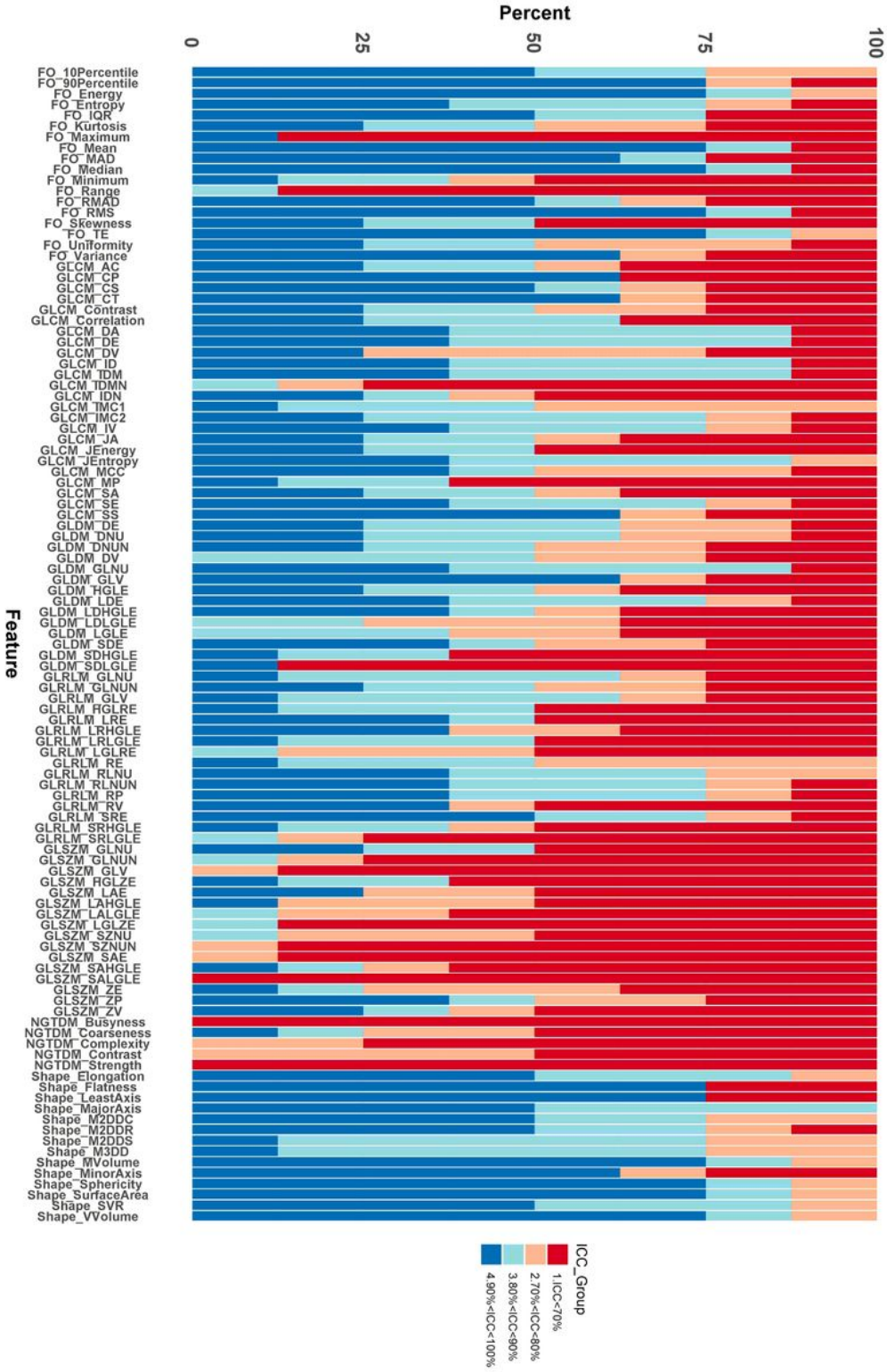

Figure 3

Bar plot showing the percentage of all radiomics features reproducibility from all regions based on the ICC. The number 1 to 4 are: 1) ICC < 70%, 2) 70% < ICC < 80%, 3) 80% < ICC < 90% and 4) 90% < ICC < 100%.

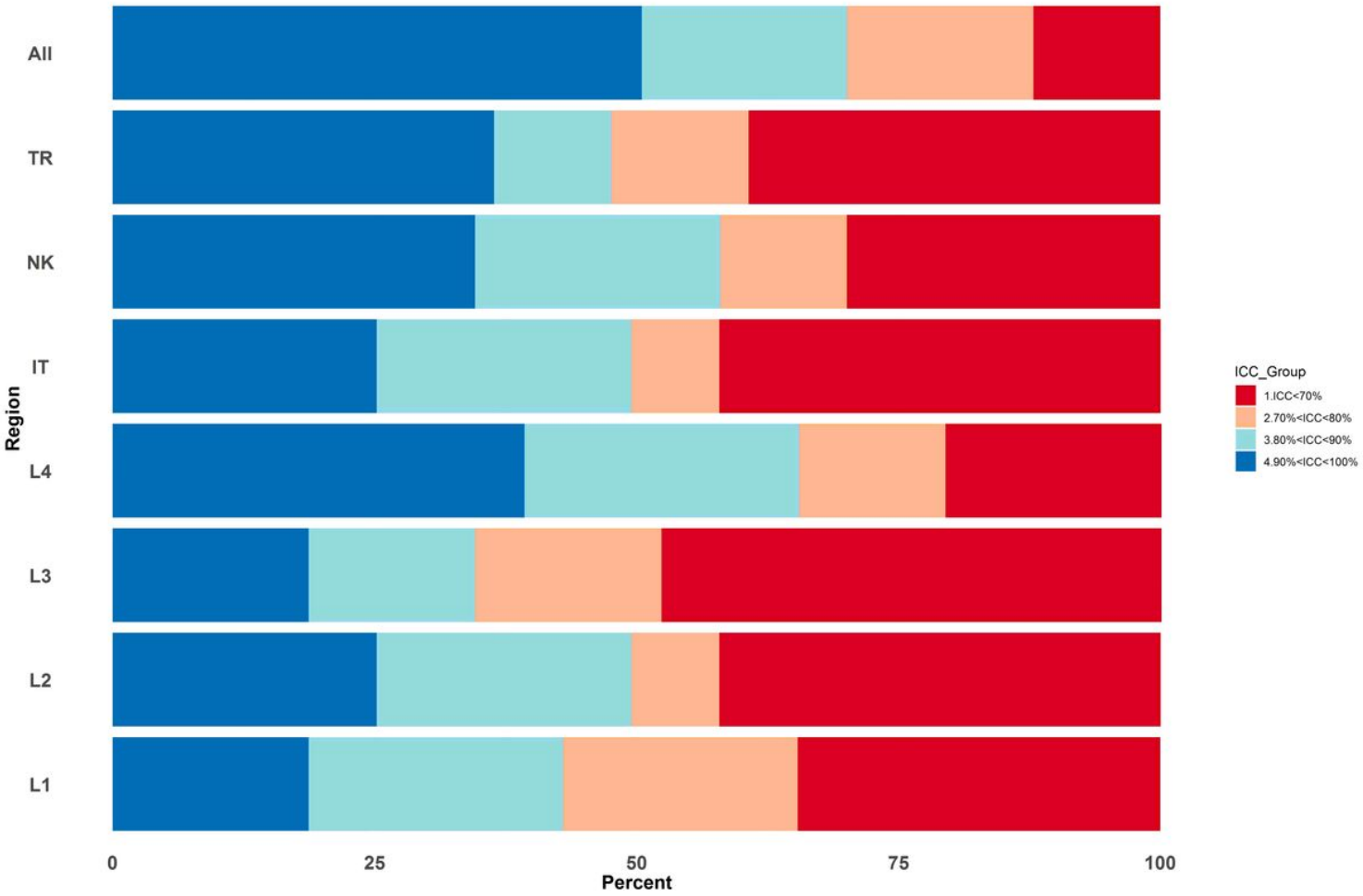

Figure 4

Bar plot showing the percentage of all radiomics features reproducibility based on the ICC for eight regions. The number 1 to 4 are: 1) ICC < 70%, 2) 70% < ICC < 80%, 3) 80% < ICC < 90% and 4) 90% < ICC < 100%. L; Lumbar; NK: Neck; TR: Trochanteric and IT: Inter trochanteric.

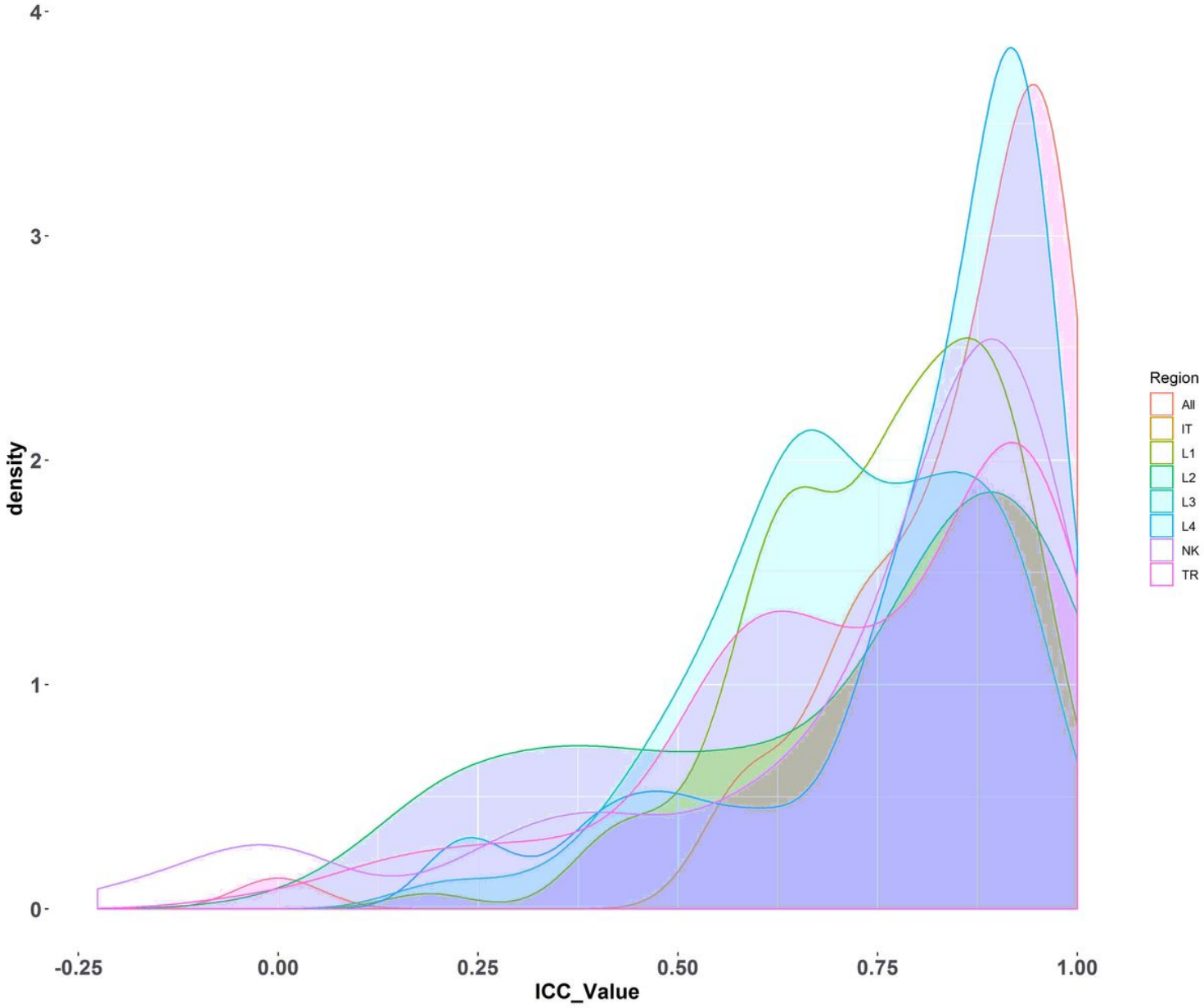

Figure 5

ICC density distribution plot for all radiomics features for all regions. L; Lumbar; NK: Neck; TR: Trochanteric and IT: Inter trochanteric.

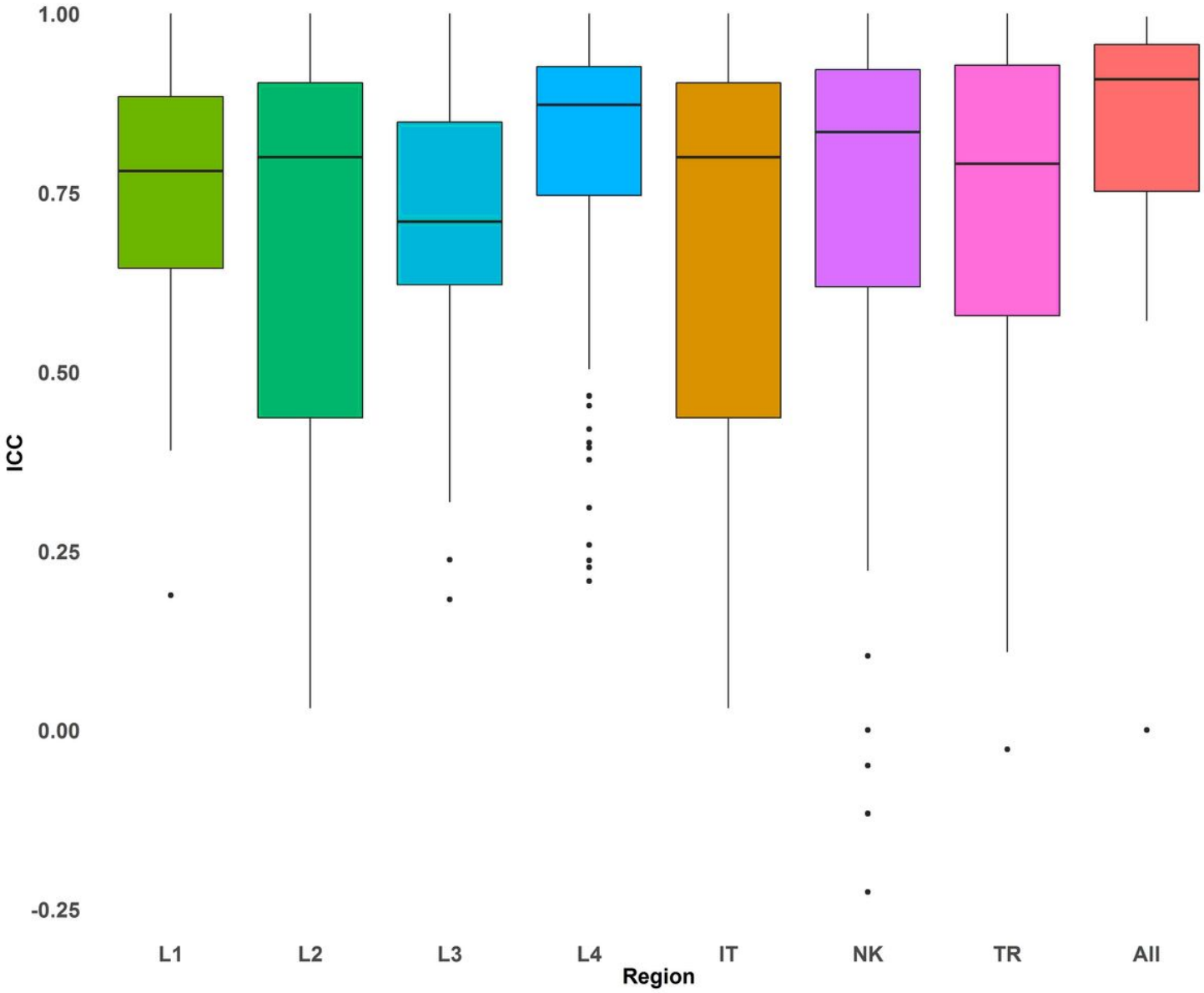

Figure 6

The ICC box plot for all radiomics features for all ROIs. L; Lumbar; NK: Neck; TR: Trochanteric and IT: Inter trochanteric.

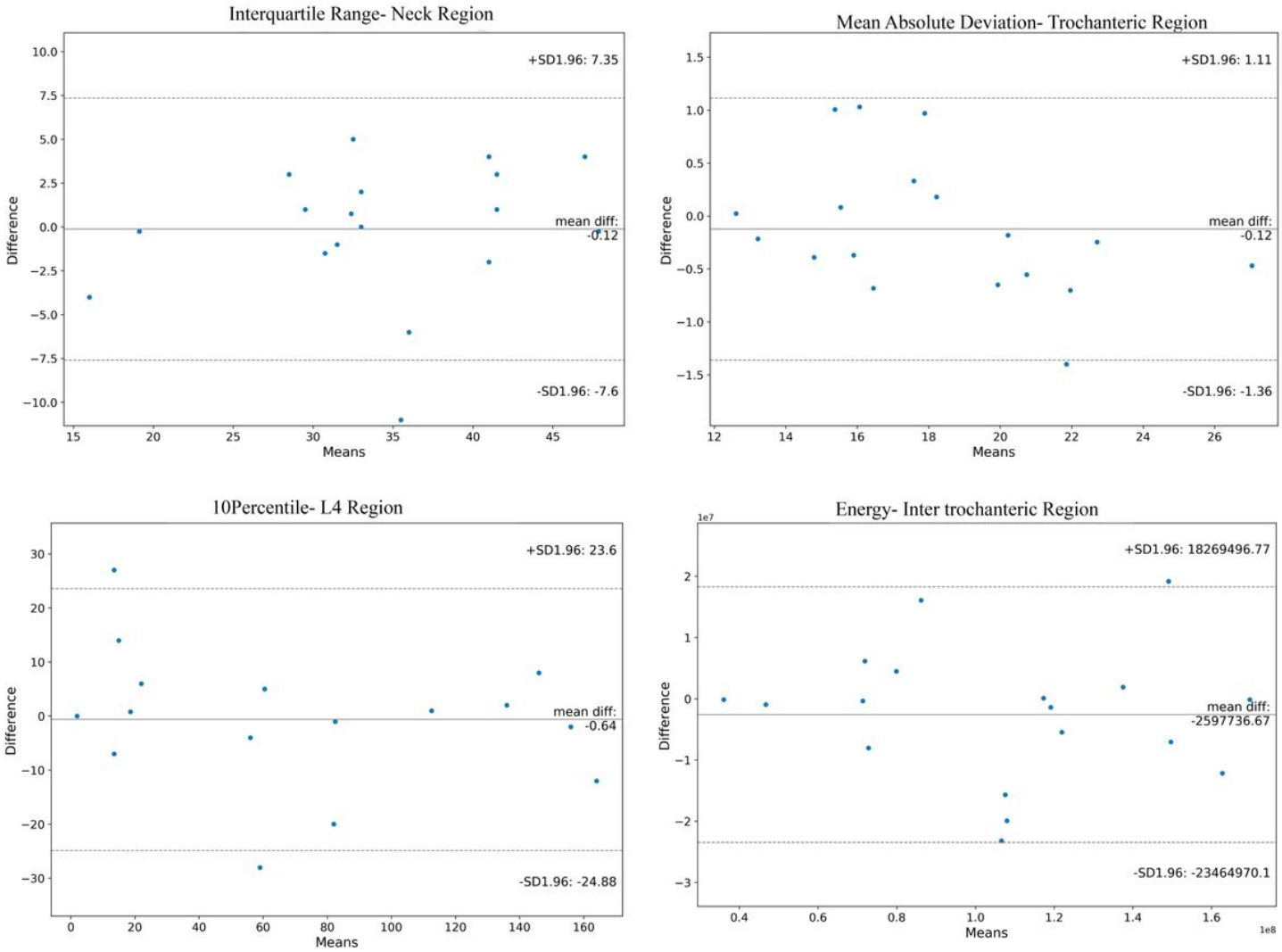

Figure 7

Bland-Altman graph for some first order (FO) radiomics features in some ROIs. These features have ICC ≥ 85%.

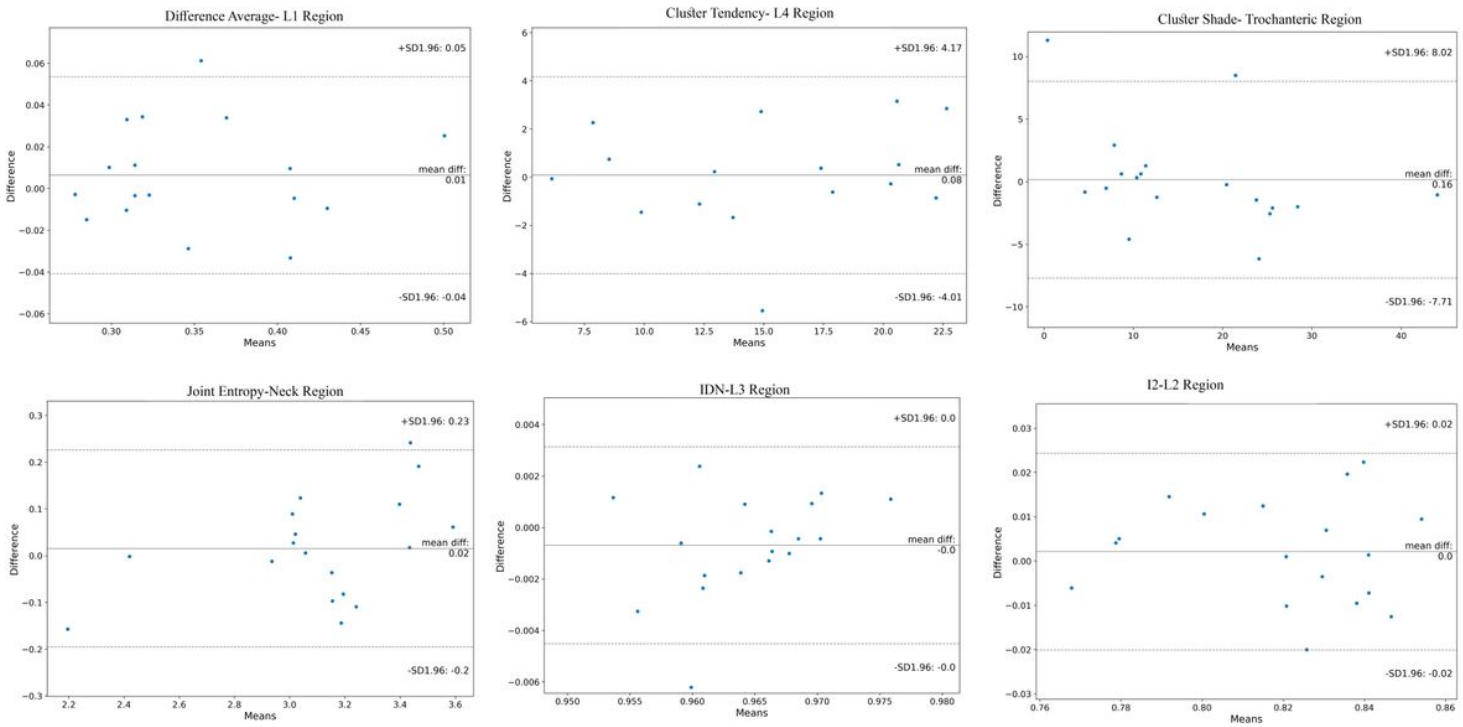

Figure 8

Bland-Altman graph for some GLCM radiomics features in some ROIs. These features have ICC ≥ 85%.

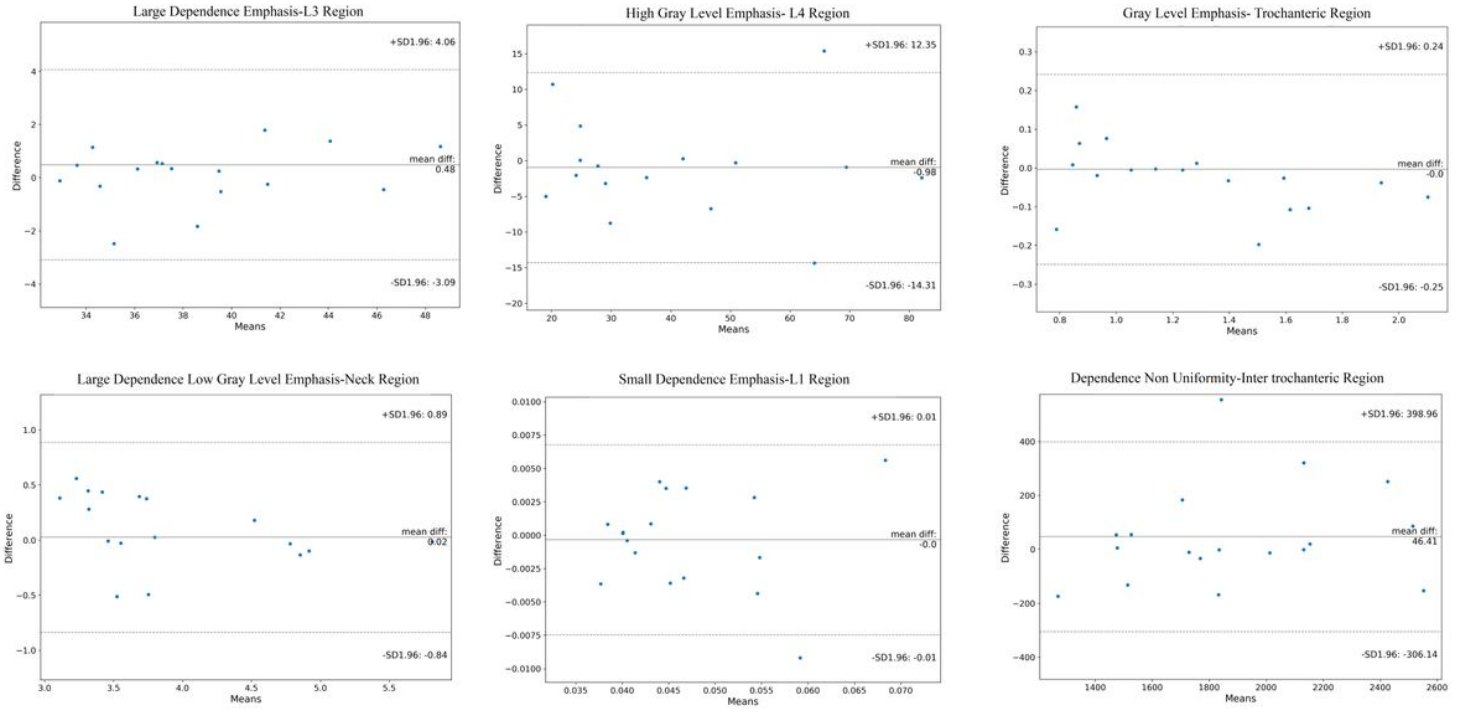

Figure 9

Bland-Altman graph for some first order GLDM radiomics features in some ROIs. These features have ICC ≥ 85%.

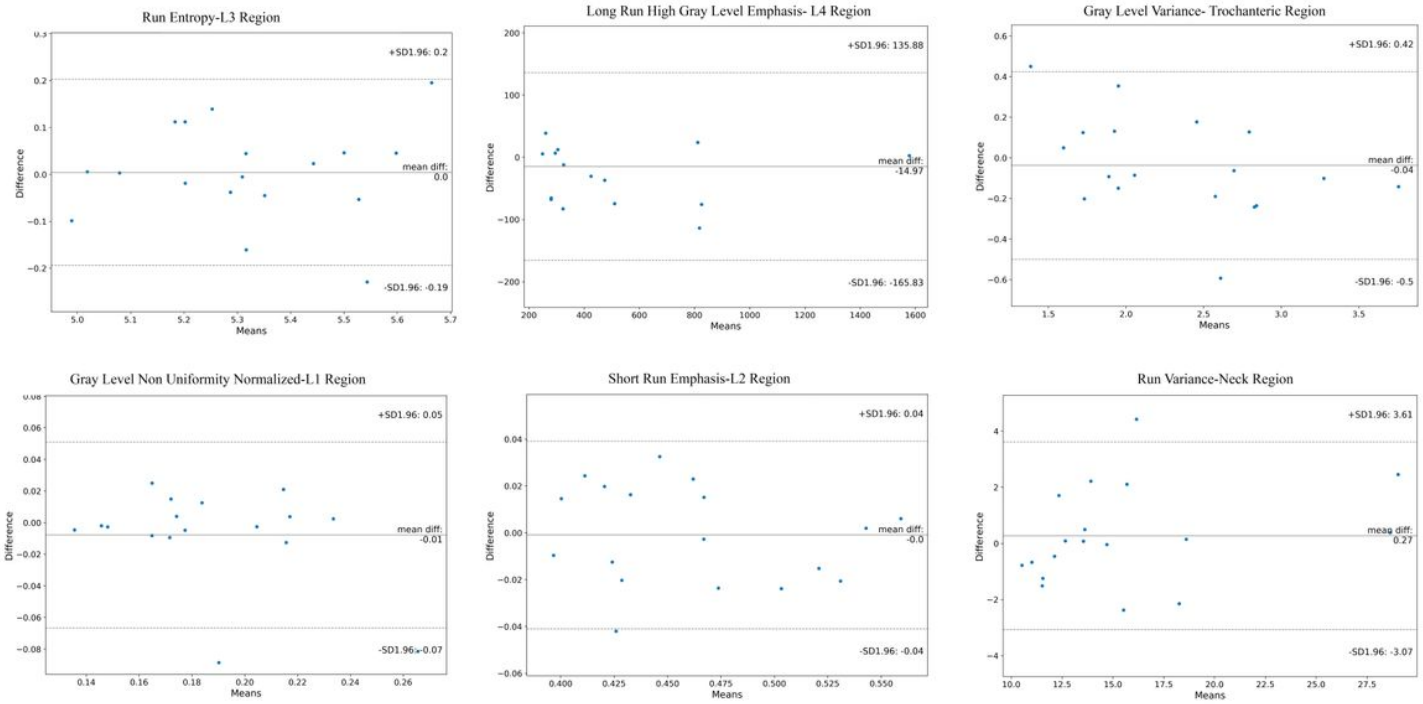

Figure 10

Bland-Altman graph for some first order GLRLM radiomics features in some ROIs. These features have ICC ≧ 85%

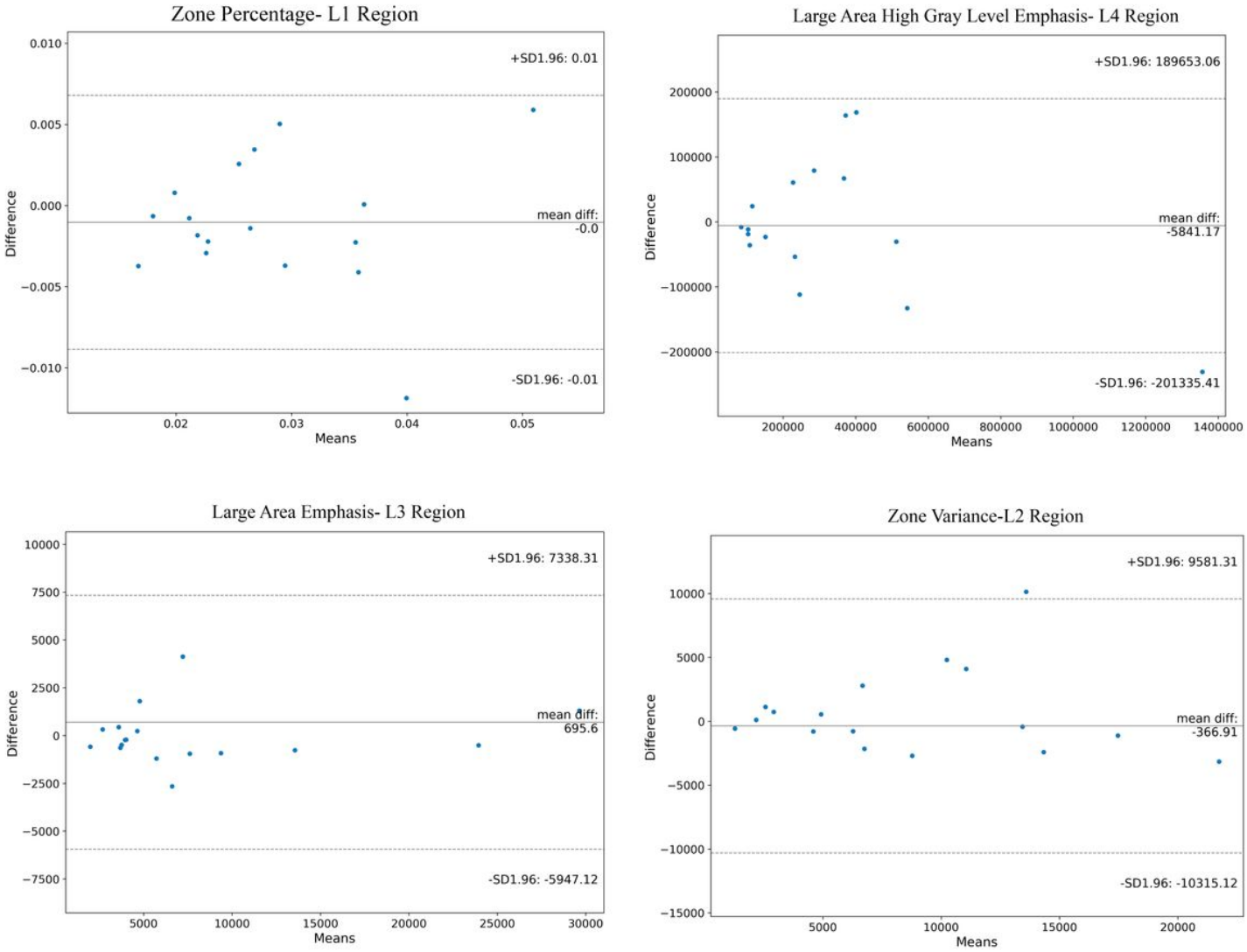

Figure 11

Bland-Altman graph for some first order GLSZM radiomics features in some ROIs. These features have ICC ≥ 85%.

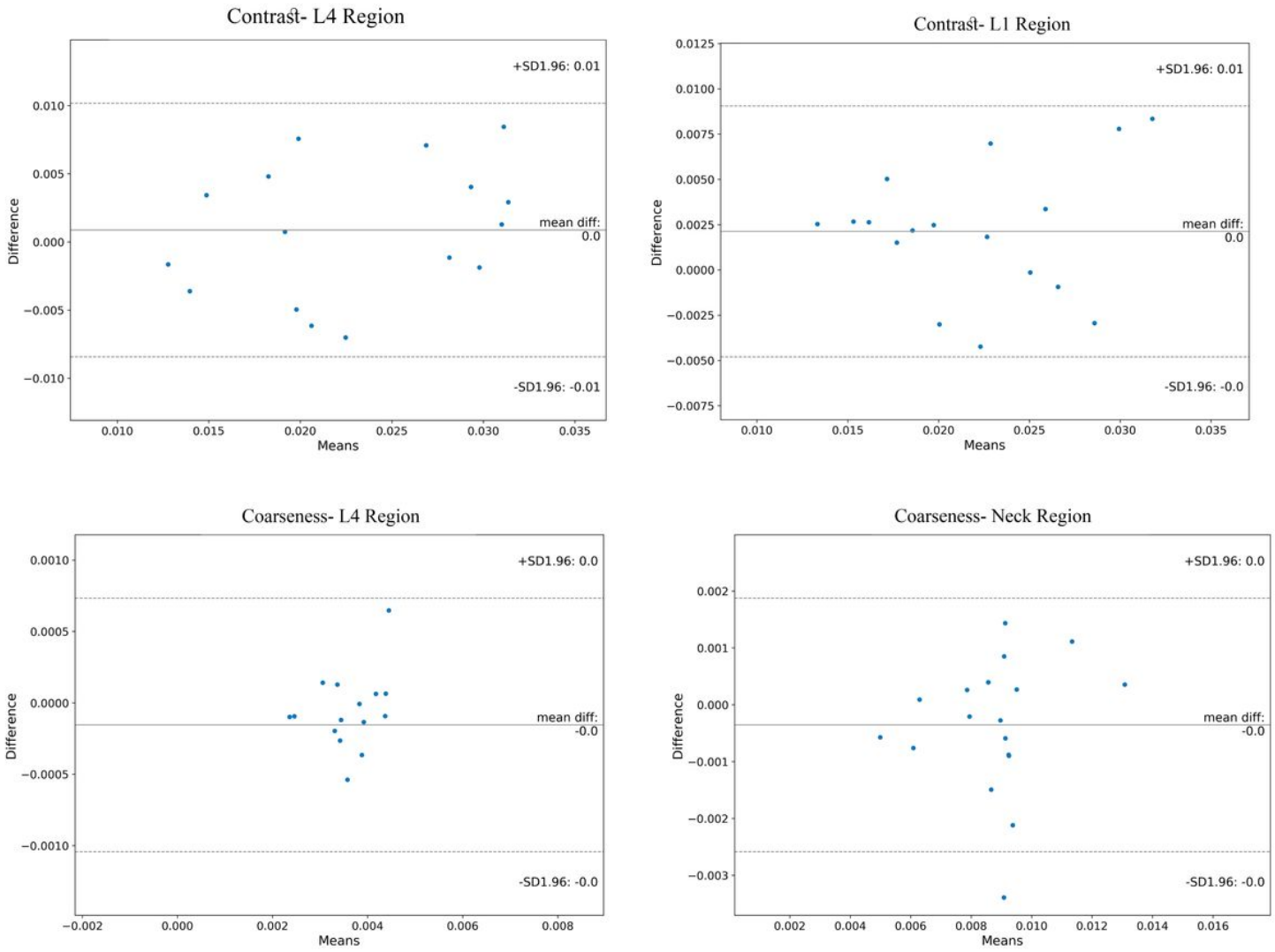

Figure 12

Bland-Altman graph for some first order NGTDM radiomics features in some ROIs. These features have the highest ICC.

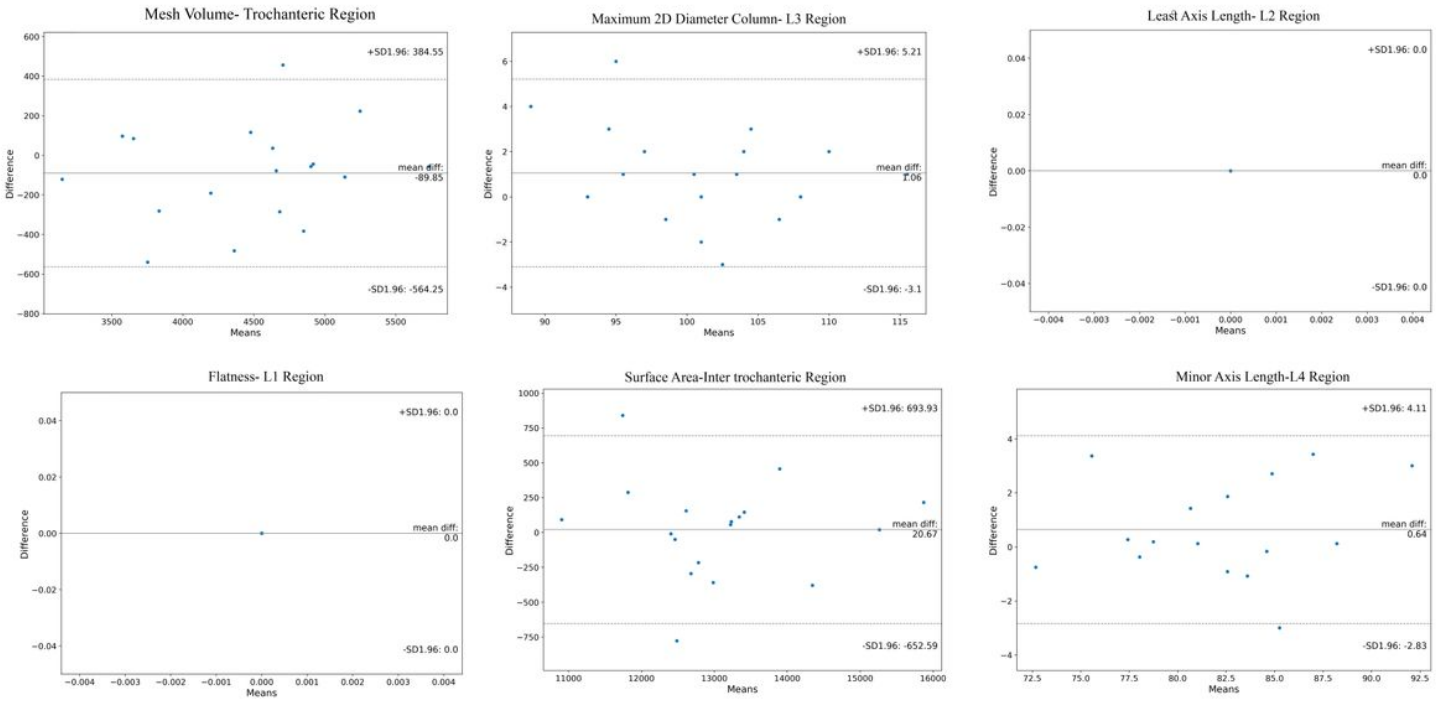

Figure 13

Bland-Altman graph for some first order Shape radiomics features in some ROIs. These features have ICC ≥ 85%.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytables.doc