Test-retest repeatability of myocardial radiomic features from quantitative cardiac magnetic resonance T1 and T2 mapping

Test-retest repeatability of myocardial radiomic features from quantitative cardiac magnetic resonance T1 and T2 mapping

ABSTRACT

Radiomics of cardiac magnetic resonance (MR) imaging has proved to be potentially useful in the study of various myocardial diseases. Therefore, assessing the repeatability degree in radiomic features measurement is of fundamental importance. The aim of this study was to assess test-retest repeatability of myocardial radiomic features extracted from quantitative T1 and T2 maps.

A representative group of 24 subjects (mean age 54 ± 18 years) referred for clinical cardiac MR imaging were enrolled in the study. For each subject, T1 and T2 mapping through MOLLI and T2-prepared TrueFISP acquisition sequences, respectively, were performed at 1.5 T. Then, 98 radiomic features of different classes (shape, first-order, second-order) were extracted from a region of interest encompassing the whole left ventricle myocardium in a short axis slice. The repeatability was assessed performing different and complementary analyses: intraclass correlation coefficient (ICC) and limits of agreement (LOA) (i.e., the interval within which 95% of the percentage differences between two repeated measures are expected to lie).

Radiomic features were characterized by a relatively wide range of repeatability degree in terms of both ICC and LOA. Overall, 44.9% and 38.8% of radiomic features showed ICC values > 0.75 for T1 and T2 maps, respectively, while 25.5% and 23.4% of radiomic features showed LOA between ±10%. A subset of radiomic features for T1 (Mean, Median, 10Percentile, 90Percentile, RootMeanSquared, Imc2, RunLengthNonUniformityNormalized, RunPercentage and

ShortRunEmphasis) and T2 (MaximumDiameter, RunLengthNonUniformityNormalized, RunPercentage, ShortRunEmphasis) maps presented both ICC > 0.75 and LOA between ±5%.

Overall, radiomic features extracted from T1 maps showed better repeatability performance than those extracted from T2 maps, with shape features characterized by better repeatability than first-order and textural features. Moreover, only a limited subset of 9 and 4 radiomic features for T1 and T2 maps, respectively, showed high repeatability degree in terms of both ICC and LOA. These results confirm the importance of assessing test-retest repeatability degree in radiomic feature estimation and might be useful for a more effective/reliable use of myocardial T1 and T2 mapping radiomics in clinical or research studies.

KEYWORDS

Magnetic resonance imaging; cardiac T1 and T2 mapping; radiomics; test-retest repeatability

Test-retest repeatability of myocardial radiomic features from quantitative cardiac magnetic resonance T1 and T2 mapping

Daniela Marfisi[1#^], Marco Giannelli[1^], Chiara Marzi[2], Jacopo Del Meglio[3], Andrea Barucci[4], Luigi Masturzo[1], Claudio Vignali[5], Mario Mascalchi[6,7], Antonio Traino[1], Giancarlo Casolo[3], Stefano Diciotti[8], Carlo Tessa[9]

[1] Unit of Medical Physics, Pisa University Hospital “Azienda Ospedaliero-Universitaria Pisana”, 56126 Pisa, Italy
[2] Department of Statistics, Computer Science, Applications "Giuseppe Parenti", University of Florence, 50134 Florence, Italy
[3] Unit of Cardiology, Azienda USL Toscana Nord Ovest, Versilia Hospital, 55041 Lido di Camaiore, Italy
[4] Institute of Applied Physics “Nello Carrara” (IFAC), Council of National Research (CNR), 50019 Sesto Fiorentino, Italy
[5] Unit of Radiology, Azienda USL Toscana Nord Ovest, Versilia Hospital, 55041 Lido di Camaiore, Italy
[6] Department of Experimental and Clinical Biomedical Sciences “Mario Serio”, University of Florence, 50121 Florence, Italy
[7] Clinical Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network (ISPRO), 50139 Florence, Italy
[8] Department of Electrical, Electronic, and Information Engineering “Guglielmo Marconi”, University of Bologna, 47522 Cesena, Italy
[9] Unit of Radiology, Azienda USL Toscana Nord Ovest, Apuane Hospital, 54100 Massa, Italy
[#] Current address: Department of Medical Physics, Azienda Sanitaria Universitaria Friuli Centrale, 33100 Udine, Italy
[^] These authors contributed equally to this work

Corresponding author: Marco Giannelli, Unit of Medical Physics, Pisa University Hospital “Azienda Ospedaliero-Universitaria Pisana”, via Roma 67, 56126 Pisa, Italy; telephone number 00 39 050993359, e-mail address m.giannelli@ao-pisa.toscana.it

Manuscript type: Original contribution.

Declarations of interest: The Authors declare no competing interests.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Figure 1

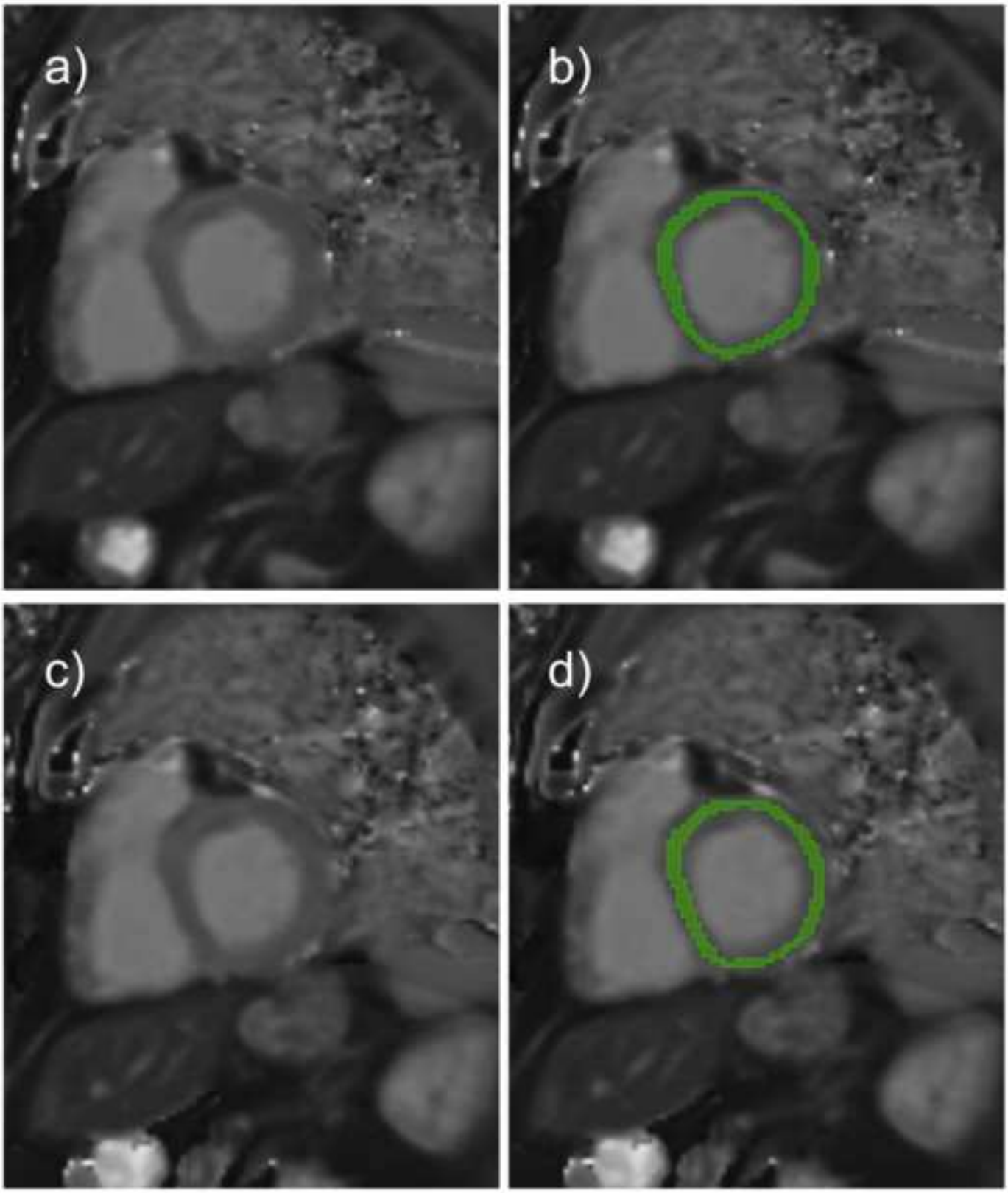


Figure 2

Click here to access/download;Figure;Figure 2.tiff

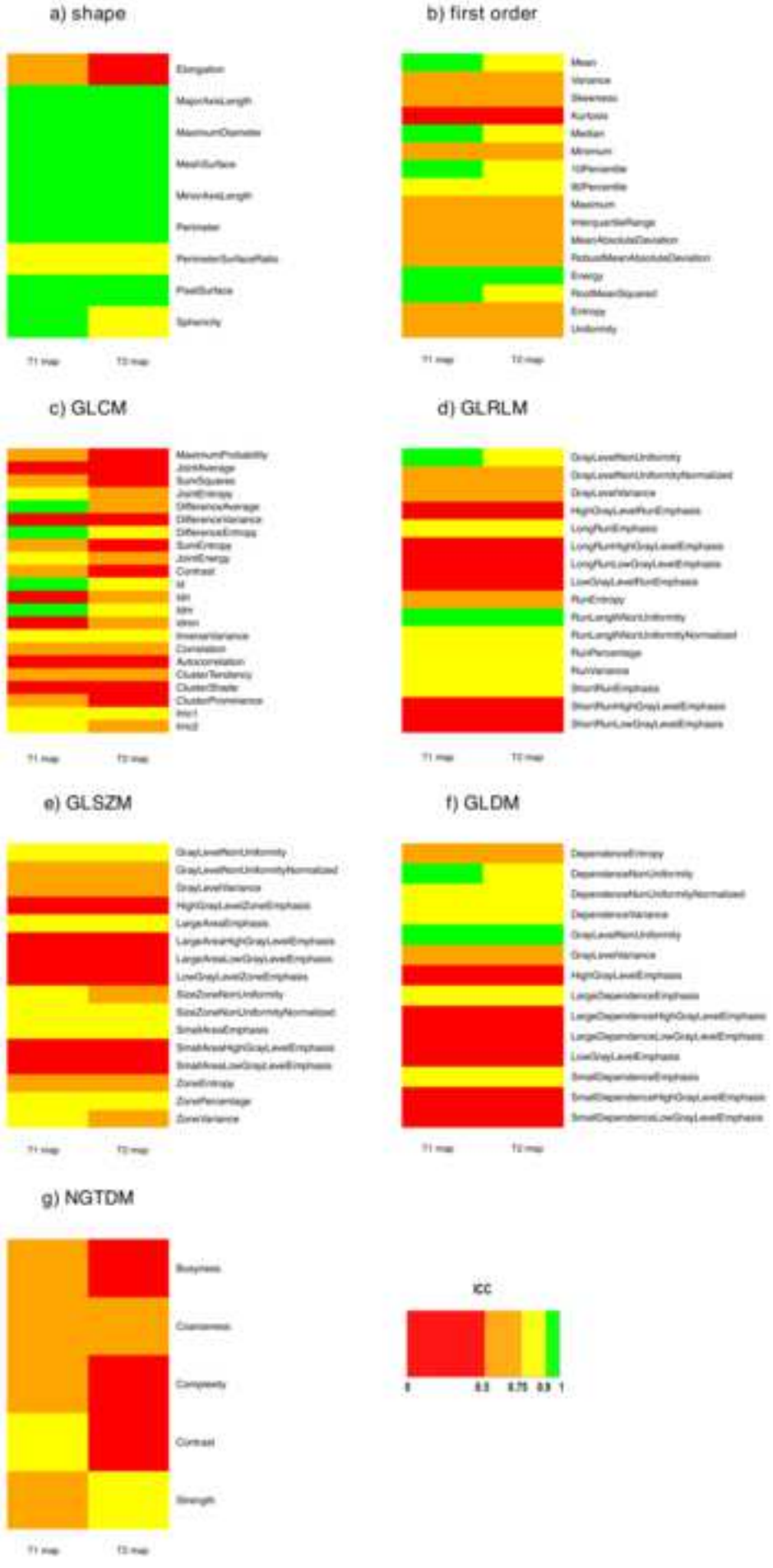


Figure 3

Click here to access/download;Figure;Figure 3.tiff

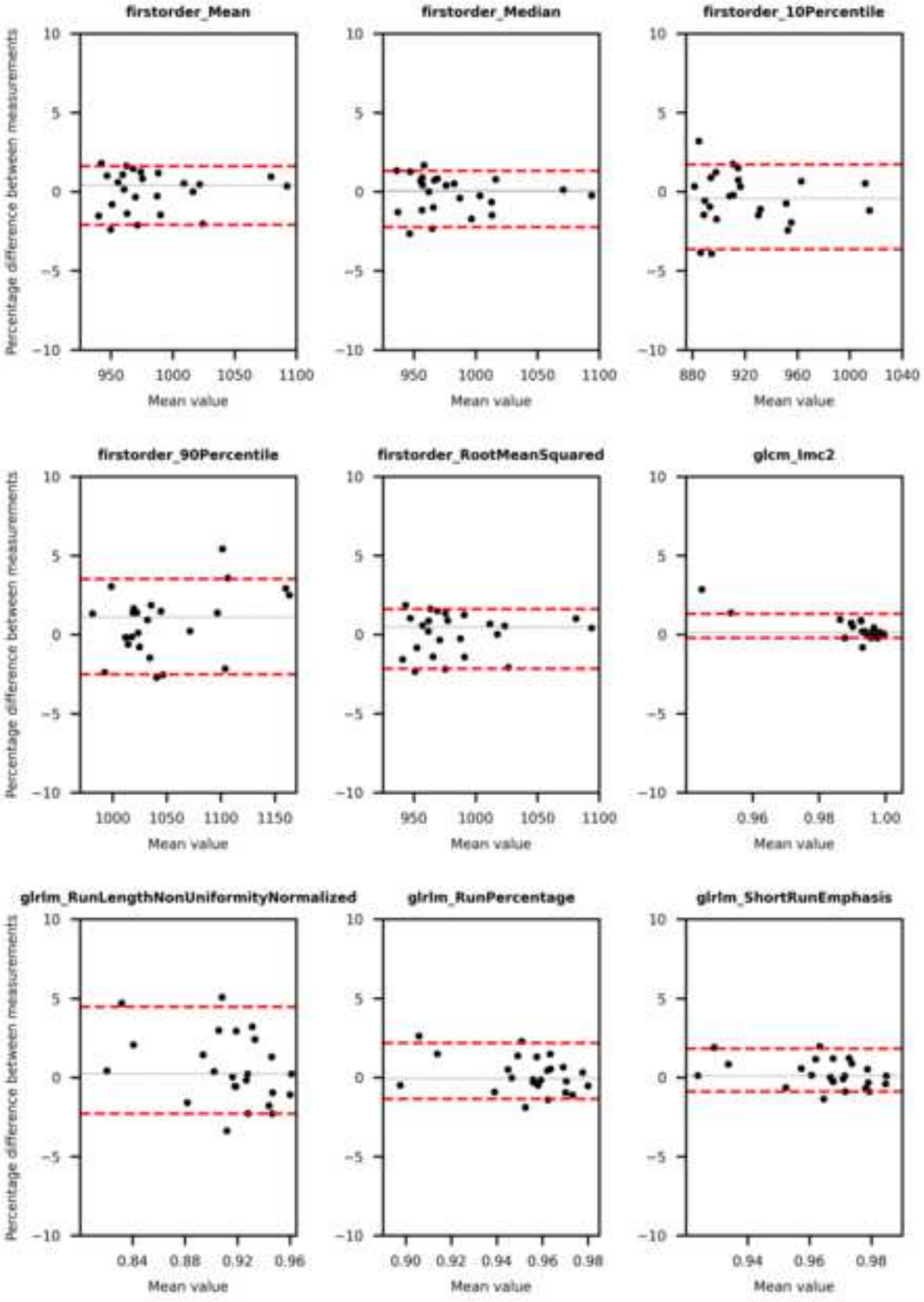
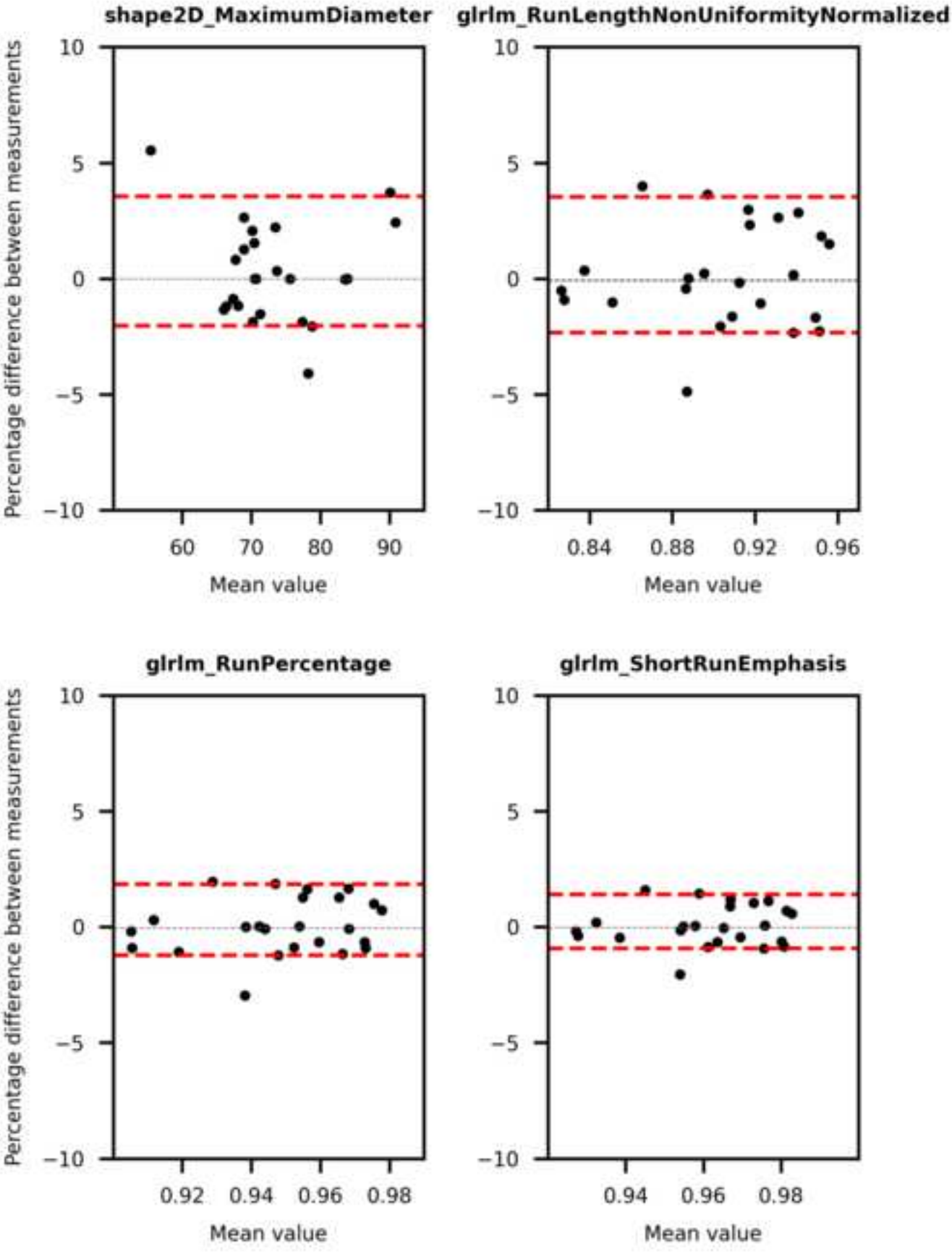


Figure 4

Click here to access/download;Figure;Figure 4.tiff



Test-retest repeatability of myocardial radiomic features from quantitative cardiac magnetic resonance T1 and T2 mapping

1. INTRODUCTION

Radiomics is a valuable technique with the potential to provide additional quantitative information from medical images employing mathematical descriptors (i.e., radiomic features) (1–4). In particular, radiomic features express various properties of a region of interest (ROI) – inherent to shape, volume, distribution of voxel intensity values, and texture (i.e., the spatial arrangement of the voxel intensity values) – and represent various indices which may be related to tissues structural changes or exploited in the field of precision medicine (1).

Nonetheless, the application of radiomics deserves some caution. Several studies have shown that radiomic features are sensitive to differences in acquisition, reconstruction, preprocessing parameters, and ROI positioning (5–18). Moreover, the proper use of radiomic features as imaging biomarkers requires a certain degree of repeatability in their estimate (2). Indeed, repeatability – i.e., the “measurement of precision that occurs with identical or near-identical conditions” (19) – is a fundamental (albeit not sufficient) condition to effectively and reliably employ radiomic features in both cross-sectional and longitudinal studies, as well as in daily clinical practice (2). Indeed, poorly repeatable features with a high degree of variability in their measurement can reduce the statistical power of a study and hinder the interpretation of any change in their values. Previous studies have shown that repeatability degree varies across radiomic features and depends on both imaging modality and anatomical region (10, 20–23). Therefore, the repeatability of radiomic features should be possibly assessed for each specific application before planning

and performing clinical or research studies.

While in principle radiomics can be applied to any sequences performed in routine clinical cardiac magnetic resonance (MR) imaging, T1 and T2 mapping are particularly attractive given their inherent quantitative nature. Recently, some promising studies have applied radiomic analysis in cardiac MR imaging to assess whether this tool could reveal myocardial phenotypic alterations in different cardiac diseases (24–40), with only a subset of these studies actually employing T1 and T2 mapping techniques (31–38). Nonetheless, to our knowledge, only two preliminary studies have assessed the repeatability of cardiac MR imaging radiomic features measurement (23, 41). In particular, the study by Raisi-Estabragh et al. (23) has evaluated test-retest repeatability of cardiac MR radiomic features from cine images using a multi-center and multi-vendor dataset. Overall, they have shown a wide variation in the repeatability degree within each considered class of radiomic features, with only slight differences between cardiac phases. Jang et al. (41) have investigated test-retest repeatability in phantoms, healthy participants, and patients referred for clinical cardiac MR imaging. In particular, for the *in vivo* investigation, they have observed varying patterns of radiomic features repeatability among acquisition sequences, with radiomic features from T1 and T2 mapping showing a similar repeatability degree to those derived from conventional acquisition sequences. However, they have reported only summary results across radiomic features, providing no information regarding the specific repeatability degree of each single radiomic feature.

Therefore, the present study aimed to assess in detail and comprehensively test-retest repeatability of myocardial radiomic features derived from quantitative cardiac MR T1 and T2 mapping.

2. MATERIAL AND METHODS

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the local Ethics Committee. Informed consent was obtained from all subjects involved in the study.

2.1. Subjects

We prospectively enrolled 6 healthy participants (2 females, 4 males) with no cardiovascular disease and 18 patients (4 females, 14 males) referred for clinical cardiac MR imaging, as reported in Table 1. This yielded a representative group of 24 subjects (54 ± 18 years), for whom a comprehensive MR imaging examination, including both T1 and T2 mapping sequences, was performed.

2.2. Cardiac MR imaging

All cardiac MR imaging examinations were performed using a 1.5 T MR scanner system (MAGNETOM Avanto, Siemens Healthcare, Erlangen, Germany) equipped with gradients of 45 mT/m strength and a 12-channel surface phased array coils.

Cine images were performed in the 2- and 4-chamber view planes (3 slices each) and in short-axis view (8-14 slices encompassing the entire left ventricular myocardium), using a TrueFISP sequence (TR = 2.5 ms, TE = 1.2 ms, slice thickness = 8 mm).

Both T1 and T2 maps were obtained in the short-axis view, acquiring a single basal slice located where myocardial thickness was maximum (with reduced partial volume effect) and where myocardial changes are assumed to be more severe in various pathologies. T1 mapping was performed utilizing a modified look-locker inversion recovery (MOLLI) pulse sequence with a 3-3-5 acquisition scheme (42).

Pulse sequence parameters were as follows: TE/TR = 1.14/2.5 ms, flip angle = 35°, matrix size = 124 × 192, in-plane resolution ranged from 1.77 mm × 1.77 mm to 2.29 mm × 2.29 mm, typical field of view = 380 mm × 273 mm, slice thickness = 8 mm. T2 maps were obtained using a T2-prepared TrueFISP sequence (43) with the following parameters: T2 preparation time = 0/24/55 ms, TR = 4 × *R*–*R*, flip angle = 70°, matrix size = 126 × 192, in-plane resolution ranged from 1.77 mm × 1.77 mm to 2.29 mm × 2.29 mm, typical field of view = 380 mm × 276 mm, slice thickness = 8 mm.

For the purpose of the present investigation, each subject underwent two repeated cardiac MR imaging examinations on the same session, with a short break of a few minutes and repositioning.

2.3. T1 and T2 maps preprocessing

For each subject and each repeated acquisition, an ROI covering the entire myocardium was manually delineated by a single radiologist expert in cardiac MR imaging (15 years of experience) using 3D Slicer (Version 4.11.2) (44, 45). ROIs were delineated separately on T1 and T2 maps, avoiding voxels with potential partial volume effect (see Figure 1).

Given that the original in-plane spatial resolution ranged from 1.77 mm x 1.77 mm to 2.29 mm x 2.29 mm across patients for both T1 and T2 maps due to the need to adapt the field of view to the patient's body size, voxel sizes were resampled to an in-plane isotropic spatial resolution of 2.0 mm using the B-spline interpolation algorithm (with the origins of interpolation and original image grids aligned together (46)).

T1 and T2 maps discretization was carried out following a fixed bin width approach as recommended by the Image Biomarker Standardization Initiative (IBSI)

guidelines when dealing with quantitative data (46). Bin width values were chosen so that the number of quantization levels of T1 and T2 maps was within the range of 30-130 for each subject (9, 47, 48). In particular, given that the median (across subjects) ranges of T1/T2 values (i.e., the difference between the maximum and the minimum T1/T2 values of voxels within the ROI) were 295/29 ms and 261/28 ms for test and retest acquisition, respectively, bin widths of 5 ms and 0.55 ms were employed for T1 and T2 maps, respectively.

All preprocessing steps and subsequent radiomic features estimation were carried out by using the open-source PyRadiomics library (49) (Version 3.0.1) with Python (Version 3.7.3).

2.4. Radiomic features estimation

Since MOLLI and T2-prepared sequences allowed obtaining T1 and T2 maps, respectively, on only one slice for single acquisition (42, 43), the 2D versions of radiomic features were considered. Specifically, a total of 98 features were extracted from each segmented ROI: 9 2D-shape features, 16 first-order features (14 intensity-based statistical features and 2 intensity histogram features, namely Entropy and Uniformity), and 73 second-order features (i.e., textural features) from gray level co-occurrence matrix (GLCM, 22 features), gray level run length matrix (GLRLM, 16 features), gray level size zone matrix (GLSZM, 16 features), gray level dependence matrix (GLDM, 14 features, with coarseness parameter $\alpha = 0$), and neighborhood gray tone difference matrix (NGTDM, 5 features) classes. Second-order features estimation was performed according to the Chebyshev norm with a distance of 1 pixel. GLCM and GLRLM features were computed from each 2D directional matrix (i.e., at 0°, 45°, 90°, and 135°) and averaged over 2D directions.

All radiomic features were computed following definitions provided by the IBSI, with shape features computed in 2D instead of the proposed 3D version. It is worth noting that the first-order feature of Kurtosis calculated by PyRadiomics followed the IBSI except for an offset value (i.e., 3).

2.5. Data analysis

The intraclass correlation coefficient (ICC) was used to assess the repeatability degree (19, 50, 51). In particular, the one-way random effects model for absolute agreement was selected. Accordingly, for each radiomic feature, the ICC for two repeated acquisitions was calculated as:

$$ICC = \frac{MSBS - MSWS}{MSBS + MSWS} \quad (1)$$

where MSBS = mean square between subjects and MSWS = mean square within subjects (52). The ICC expresses the agreement between repeated measurements with respect to the variance between subjects (i.e., a relative estimation of repeatability). Radiomic features were stratified based on the degree of agreement between repeated acquisitions (53): poor (ICC $\leq$ 0.5), moderate (0.5 < ICC $\leq$ 0.75), good (0.75 < ICC $\leq$ 0.9), and excellent (0.9 < ICC $\leq$ 1) agreement.

In addition, for each radiomic feature, the limits of agreement (LOA), as originally proposed by Bland and Altman (19, 54), were obtained. The LOA identify the interval within which 95% of the differences between two repeated measures are expected to lie (54). Thus, following a nonparametric approach, they were estimated using the 2.5^{th} and 97.5^{th} percentiles of the percentage differences between two repeated measurements (54).

Then, given that ICC and LOA are different and complementary indices for characterizing repeatability, radiomic features showing a good degree of repeatability

both in terms of ICC (i.e., ICC > 0.75) and LOA (i.e., LOA between ±5%) were identified.

All statistical analyses were performed according to the Quantitative Imaging Biomarker Alliance (QIBA) recommendations (19, 55) and were carried out by using R (Version 3.6.2) software package in the RStudio (Version 1.2.5033) environment (56).

3. RESULTS

ICC results are reported in Figure 2 for both T1 and T2 maps. Overall, 44.9% (44/98) and 38.8% (38/98) of radiomic features showed ICC values > 0.75 for T1 and T2 maps, respectively. As shown in Figure 2, all radiomic features with a good or excellent agreement for T2 mapping presented the same (or even better) behavior for T1 mapping. On the other hand, all features with ICC $\leq$ 0.75 for T1 mapping, showed the same (or even worse) behavior for T2 mapping.

Shape features exhibited the best repeatability performance, with most features (7/9 and 6/9 for T1 and T2 mapping, respectively) presenting ICC > 0.9, and only one feature (i.e., Elongation) showing moderate/poor agreement between repeated measures in T1/T2 mapping (Figure 2). On the other hand, first-order features showed large variations of ICC values. Most first-order features (10/16) had the same degree of repeatability, with ICC $\leq$ 0.75, for both T1 and T2 maps. Some of the remaining features (i.e., Mean, Median, 10Percentile, and RootMeanSquared) had higher ICC values (i.e., ICC > 0.9) for T1 than for T2 maps.

Overall, some appreciable differences between the repeatability performance of T1 and T2 mapping were found for textural features. In particular, 30/73 and 24/73 textural features showed good or excellent ICC values for T1 and T2 maps,

respectively. Moreover, looking at each texture class separately, it can be observed (Figure 2) that, for the GLCM class, no textural features derived from T2 maps had ICC > 0.9, while 4 features derived from T1 maps (i.e., DifferenceAverage, DifferenceEntropy, Id, and Idm) showed an excellent test-retest repeatability. As for GLRLM, GLSZM, and GLDM classes, the same behavior for features estimated from T1 maps and those estimated from T2 maps was found, with either an excellent/good or moderate/poor agreement between feature values from repeated acquisitions (SizeZoneNonUniformity and ZoneVariance from GLSZM class were the only exceptions, with ICC > 0.75 for T1 maps and ICC $\leq$ 0.75 for T2 maps). The NGTDM class was found to have the worst repeatability performance, with only one feature per map (i.e., Contrast and Strength for T1 and T2 maps, respectively) showing good test-retest agreement.

Tables 2 and 3 summarize the results of LOA analysis for radiomic features derived from T1 and T2 maps, respectively. Notably, LOA were found to vary appreciably within each feature class. In general, 25 T1 mapping and 23 T2 mapping features out of 98 radiomic features showed a repeatability degree with LOA between ±10%. Bland-Altman plots for all the radiomic features with both LOA between ±5% and ICC > 0.75 are shown in Figures 3 and 4 for T1 and T2 maps, respectively.

Radiomic features of the shape class showed relatively low variability in LOA results. None of the shape features had LOA between ±5%, except for MaximumDiameter computed on T2 maps (which also had ICC > 0.75, as can be seen in Figure 4). Four shape features (i.e., MajorAxisLength, MinorAxisLength, MaximumDiameter, and Perimeter) had LOA between ±10%, while LOA for the other shape features did not exceed ±30% (Tables 2 and 3).

First-order features showed large variations in LOA results (from approximately ±2% to ±80% or more), with only 6/16 and 5/16 features having LOA between ±10%, for T1 and T2 maps, respectively. Moreover, except for Entropy, first-order features with LOA of approximately ±10% (or even ±5% for T1 mapping, as can be seen in Figure 3) had ICC > 0.75.

Overall, 16/73 and 14/73 T1 and T2 mapping derived textural features, respectively, were characterized by LOA between ±10%. None of these features (neither for T1 nor for T2 maps) belonged to the NGTDM class, which had poor repeatability (i.e., LOA approximately ranged between ±89% and ±100%, for T1 and T2 maps, respectively). Among GLCM features with ICC > 0.75, only JointEntropy, DifferenceEntropy, and Imc2 extracted from T1 maps also presented LOA between ±10%, while no GLCM features from T2 maps satisfied these requirements (Tables 2 and 3). Regarding GLRLM, GLSZM, and GLDM classes, the LOA analysis showed similar behavior for textural features derived from T1 and T2 maps, with the same radiomic features showing LOA between ±10% (Tables 2 and 3). Among these features, ShortRunEmphasis, RunLengthNonUniformityNormalized, and RunPercentage from GLRLM class were the only textural features that showed both LOA between ±5% and ICC > 0.75 (in addition to Imc2 from GLCM estimated from T1 maps).

4. DISCUSSION

While several previous studies have investigated radiomic features repeatability in MR imaging, particularly in gastrointestinal and genitourinary imaging, as well as in neuroradiology (10, 22, 57–61), their results are not directly applicable in clinical cardiac MR studies. So far, only two previous studies have assessed the repeatability

of radiomic features measurement in cardiac MR imaging (23, 41). Specifically, Raisi-Estabragh et al. (23) have evaluated the repeatability of radiomic features on the test-retest scanning, concerning cine balanced steady-state free precession sequences in short-axis view, using a multi-center and multi-vendor dataset. While this study has shown that the repeatability performances vary greatly within each feature class, the authors have not employed quantitative T1 and T2 mapping, as performed in the present study. On the other hand, Jang et al. (41) have carried out a repeatability study in which different acquisition sequences have been used (i.e., cine balanced state free precession, T1-weighted, T2-weighted, T1 mapping, and T2 mapping). Overall, this study, evaluating test-retest repeatability through ICC analysis in a group of patients, as well as in phantom and a group of healthy subjects, has identified 26.4% and 34.8% of repeatable features, with ICC $\geq$ 0.8, extracted from T1 and T2 maps, respectively. However, Jang et al. (41) did not report ICC values for each single radiomic feature, which is a relevant information when applying radiomics in clinical or research studies. Therefore, we feel a further and more comprehensive evaluation of the repeatability of radiomic features measurement in cardiac MR imaging applications can be useful.

Accordingly, we addressed repeatability of radiomic features derived from myocardial T1 and T2 mapping. Notably, we used two different statistical methods based on ICC and LOA analysis, in line with QIBA recommendations (19), obtaining complementary information on the repeatability degree in radiomic features measurement. Specifically, ICC, which expresses the agreement between repeated measures with respect to the variance between subjects, is suitable for assessing the potential effect of repeatability in cross-sectional studies. Indeed, low repeatability in terms of ICC can yield a reduced statistical power of the study. On the other hand, for

longitudinal studies, high ICC values should be interpreted with some caution (55), given that large variability of radiomic features between subjects can yield increased ICC values. In this regard, given that LOA do not depend on the variability between subjects, the LOA analysis can be particularly suitable to assess radiomic features capability of revealing changes in longitudinal studies.

Overall, both ICC and LOA analyses indicated that radiomic features derived from T1 maps were characterized by better repeatability than those derived from T2 maps, regardless of the feature class. Nonetheless, for both T1 and T2 maps, the repeatability degree within each class varied greatly across features, with very different LOA values and large differences in ICC values. Only shape and NGTDM features showed relatively low variations in LOA and ICC values, with the best and the worst repeatability performance, respectively.

The combination of ICC and LOA results allowed the identification of a subset of radiomic features for T1 (i.e., Mean, Median, 10Percentile, 90Percentile, RootMeanSquared from first-order class, Imc2 from GLCM class, and RunLengthNonUniformityNormalized, RunPercentage, ShortRunEmphasis from GLRLM class) and T2 maps (i.e., MaximumDiameter from shape class and RunLengthNonUniformityNormalized, RunPercentage, ShortRunEmphasis from GLRLM class) that showed good repeatability performance both in terms of ICC (> 0.75) and LOA (±5%). Notably, to avoid possible bias due to overfitting and obtain reliable results, a high degree of repeatability might represent one of the possible criteria for selecting features in studies employing radiomics and artificial intelligence methods. Nonetheless, this does not necessarily imply a high discriminative or predictive power in clinical or research studies, which might also depend on the specific effect size associated with selected features.

While the previous study by Raisi-Estabragh et al. (23) has assessed the repeatability performances of various radiomic features from different classes, no direct comparison with our results can be made, since they did not acquire quantitative T1 and T2 maps. The findings of our study might appear not completely in agreement with the preliminary study by Jang et al. (41), which seems to suggest better repeatability performance for T2 mapping- than for T1 mapping-derived features. Several factors could explain this aspect. Indeed, their analysis was based only on ICC. Also, they have utilized an ICC cut-off value of 0.8 to define whether a radiomic feature measurement was repeatable, without however indicating the specific features satisfying this requirement. Moreover, they have not included shape features in their evaluation (i.e., only first-order and textural features were considered). In addition, differences in magnetic field strength (i.e., 3 T in their study versus 1.5 T in our study) and acquisition sequence parameters, as well as in the choice of preprocessing parameters (such as resampling voxel sizes and bin width), might have further contributed.

We recognize some potential limitations of our study. First, we considered a relatively small number of subjects. Second, our results on T1 and T2 mapping are not directly applicable to other cardiac MR imaging acquisition sequences. Third, given inherent difficulties due to organ motion, quantitative cardiac MR T1 and T2 mapping is usually performed by using specific sequences such as MOLLI (42) and T2-prepared TrueFISP (43), respectively, which allow obtaining only one slice for single acquisition. While we carried out hence 2D radiomic analysis (as typically performed in clinical studies), whole-heart coverage and 3D radiomic analysis might provide a more comprehensive evaluation of disease burden and potentially increase the diagnostic performance of cardiac MR T1 and T2 mapping in collaborative

patients. Fourth, in this exquisitely technical study of test-retest repeatability on a representative group of subjects, we did not assess feature discriminant ability for specific clinical applications, which requires planning and performing further tailored clinical studies. Finally, the selected cut-off values (i.e., ICC > 0.75 and LOA between ±5%) to identify radiomic features with repeatable characteristics were arbitrarily chosen, albeit reasonably conservative.

5. CONCLUSIONS

Since we observed that myocardial T1- and T2-derived radiomic features of different classes were characterized by a relatively wide range of repeatability degrees assessed in terms of both ICC and LOA, our results confirm the importance of performing such an analysis toward a more reliable and effective use of radiomics in clinical applications of cardiac MR T1 and T2 mapping. Overall, radiomic features extracted from T1 maps showed better repeatability performance than those extracted from T2 maps, with shape features characterized by better repeatability than first-order and textural features. Moreover, we identified a subset of 9 and 4 radiomic features, for T1 and T2 mapping, respectively, which showed good/excellent repeatability in terms of both ICC (> 0.75) and LOA (±5%). Such highly repeatable features represent hence potential candidates for feature selection in clinical or research studies, although a high degree of repeatability does not necessarily imply an effective discriminative or predictive power, which can vary also with the specific clinical application.

REFERENCES

1. Gillies RJ, Kinahan PE, Hricak H: Radiomics: images are more than pictures, they are data. *Radiology* 2016; 278:563–577.

2. O'Connor JPB, Aboagye EO, Adams JE, et al.: Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol* 2017; 14:169–186.

3. Li R, Xing L, Napel S, Rubin D (Eds): *Radiomics and Radiogenomics: Technical Basis and Clinical Applications*. Boca Raton, FL: CRC Press, Taylor & Francis Group; 2019.

4. Lambin P, Rios-Velazquez E, Leijenaar R, et al.: Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; 48:441–446.

5. Traverso A, Wee L, Dekker A, Gillies R: Repeatability and reproducibility of radiomic features: a systematic review. *Int J of Radiat Oncol Biol Phys* 2018; 102:1143–1158.

6. Pfaehler E, Zhovannik I, Wei L, et al.: A systematic review and quality of reporting checklist for repeatability and reproducibility of radiomic features. *Phys Imaging Radiat Oncol* 2021; 20:69–75.

7. Linsalata S, Borgheresi R, Marfisi D, et al.: Radiomics of patients with locally advanced rectal cancer: Effect of preprocessing on features estimation from computed tomography imaging. *BioMed Res Int* 2022; 2022:2003286.

8. Shafiq-ul-Hassan M, Latifi K, Zhang G, Ullah G, Gillies R, Moros E: Voxel size and gray level normalization of CT radiomic features in lung cancer. *Sci Rep* 2018; 8:10545.

9. Leijenaar RTH, Nalbantov G, Carvalho S, et al.: The effect of SUV discretization in quantitative FDG-PET radiomics: the need for standardized methodology in tumor texture analysis. *Sci Rep* 2015; 5:11075.

10. Schwier M, van Griethuysen J, Vangel MG, et al.: Repeatability of multiparametric prostate MRI radiomics features. *Sci Rep* 2019; 9:9441.

11. Traverso A, Kazmierski M, Welch ML, et al.: Sensitivity of radiomic features to inter-observer variability and image pre-processing in Apparent Diffusion Coefficient (ADC) maps of cervix cancer patients. *Radiother Oncol* 2020; 143:88–94.

12. Marfisi D, Tessa C, Marzi C, et al.: Image resampling and discretization effect on the estimate of myocardial radiomic features from T1 and T2 mapping in hypertrophic cardiomyopathy. *Sci Rep* 2022; 12:10186.

13. Jang J, El-Rewaidy H, Ngo LH, et al.: Sensitivity of myocardial radiomic features to imaging parameters in cardiac MR imaging. *J Magn Reson Imaging* 2021; 54:787–794.

14. Marzi C, Marfisi D, Barucci A, et al.: Collinearity and dimensionality reduction in radiomics: Effect of preprocessing parameters in hypertrophic cardiomyopathy magnetic resonance T1 and T2 mapping. *Bioengineering* 2023; 10:80.

15. Le EPV, Rundo L, Tarkin JM, et al.: Assessing robustness of carotid artery CT angiography radiomics in the identification of culprit lesions in cerebrovascular events. *Sci Rep* 2021; 11:3499.

16. Gevaert O, Mitchell LA, Achrol AS, et al.: Glioblastoma multiforme: Exploratory radiogenomic analysis by using quantitative image features. *Radiology* 2014; 273:168–174.

17. Bologna M, Corino VDA, Montin E, et al.: Assessment of stability and discrimination capacity of radiomic features on apparent diffusion coefficient images. *J Digit Imaging* 2018; 31:879–894.

18. Hu W, Wu X, Dong D, et al.: Novel radiomics features from CCTA images for the functional evaluation of significant ischaemic lesions based on the coronary fractional flow reserve score. *Int J Cardiovasc Imaging* 2020; 36:2039–2050.

19. Sullivan DC, Obuchowski NA, Kessler LG, et al.: Metrology Standards for Quantitative Imaging Biomarkers. *Radiology* 2015; 277:813–825.

20. Desseroit M-C, Tixier F, Weber WA, et al.: Reliability of PET/CT shape and heterogeneity features in functional and morphologic components of non–small cell lung cancer tumors: A repeatability analysis in a prospective multicenter cohort. *J Nucl Med* 2017; 58:406–411.

21. Van Timmeren JE, Leijenaar RTH, Van Elmpt W, et al.: Test–retest data for radiomics feature stability analysis: Generalizable or study-specific? *Tomography* 2016; 2:361–365.

22. Crombé A, Buy X, Han F, Toupin S, Kind M: Assessment of repeatability, reproducibility, and performances of T2 mapping-based radiomics features: A comparative study. *J Magn Reson Imaging* 2021; 54:537–548.

23. Raisi-Estabragh Z, Gkontra P, Jaggi A, et al.: Repeatability of cardiac magnetic resonance radiomics: A multi-centre multi-vendor test-retest study. *Front Cardiovasc Med* 2020; 7:586236.

24. Baessler B, Mannil M, Oebel S, Maintz D, Alkadhi H, Manka R: Subacute and chronic left ventricular myocardial scar: accuracy of texture analysis on nonenhanced cine MR images. *Radiology* 2018; 286:103–112.

25. Baessler B, Mannil M, Maintz D, Alkadhi H, Manka R: Texture analysis and machine learning of non-contrast T1-weighted MR images in patients with hypertrophic cardiomyopathy - preliminary results. *Eur J Radiol* 2018; 102:61–67.

26. Amano Y, Suzuki Y, Yanagisawa F, Omori Y, Matsumoto N: Relationship

between extension or texture features of late gadolinium enhancement and ventricular tachyarrhythmias in hypertrophic cardiomyopathy. *BioMed Res Int* 2018; 2018:4092469.

27. Cheng S, Fang M, Cui C, et al.: LGE-CMR-derived texture features reflect poor prognosis in hypertrophic cardiomyopathy patients with systolic dysfunction: preliminary results. *Eur Radiol* 2018; 28:4615–4624.

28. Schofield RS, Ganeshan B, Fontana M, et al.: Texture analysis of cardiovascular magnetic resonance cine images differentiates aetiologies of left ventricular hypertrophy. *Clin Radiol* 2019; 74:140–149.

29. Amano Y, Yanagisawa F, Omori Y, et al.: Detection of myocardial tissue alterations in hypertrophic cardiomyopathy using texture analysis of T2-weighted short inversion time inversion recovery magnetic resonance imaging. *J Comput Assist Tomogr* 2020; 44:341–345.

30. Alis D, Guler A, Yergin M, Asmakutlu O: Assessment of ventricular tachyarrhythmia in patients with hypertrophic cardiomyopathy with machine learning-based texture analysis of late gadolinium enhancement cardiac MRI. *Diagn Interv Imaging* 2020; 101:137–146.

31. Baessler B, Luecke C, Lurz J, et al.: Cardiac MRI texture analysis of T1 and T2 maps in patients with infarctlike acute myocarditis. *Radiology* 2018; 289:357–365.

32. Baessler B, Luecke C, Lurz J, et al.: Cardiac MRI and texture analysis of myocardial T1 and T2 maps in myocarditis with acute versus chronic symptoms of heart failure. *Radiology* 2019; 292:608–617.

33. Neisius U, El-Rewaidy H, Nakamori S, Rodriguez J, Manning WJ, Nezafat R: Radiomic analysis of myocardial native T1 imaging discriminates between hypertensive heart disease and hypertrophic cardiomyopathy. *JACC Cardiovas Imaging* 2019; 12:1946–1954.

34. Wang J, Yang F, Liu W, et al.: Radiomic analysis of native T1 mapping images discriminates between MYH7 and MYBPC3-related hypertrophic cardiomyopathy. *J Magn Reson Imaging* 2020; 52:1714–1721.

35. Neisius U, El-Rewaidy H, Kucukseymen S, et al.: Texture signatures of native myocardial T_1 as novel imaging markers for identification of hypertrophic cardiomyopathy patients without scar. *J Magn Reson Imaging* 2020; 52:906–919.

36. Shi R-Y, Wu R, An D-AL, et al.: Texture analysis applied in T1 maps and extracellular volume obtained using cardiac MRI in the diagnosis of hypertrophic cardiomyopathy and hypertensive heart disease compared with normal controls. *Clin Radiol* 2021; 76:236.e9-236.e19.

37. Antonopoulos AS, Boutsikou M, Simantiris S, et al.: Machine learning of native T1 mapping radiomics for classification of hypertrophic cardiomyopathy phenotypes. *Sci Rep* 2021; 11:23596.

38. Zhang J, Xu Y, Li W, et al.: The predictive value of myocardial native T1 mapping radiomics in dilated cardiomyopathy: A study in a chinese population. *Magn Reson Imaging* 2023; 58:772–779.

39. Avard E, Shiri I, Hajianfar G, et al.: Non-contrast cine cardiac magnetic resonance image radiomics features and machine learning algorithms for myocardial infarction detection. *Comput Biol Med* 2022; 141:105145.

40. Durmaz ES, Karabacak M, Ozkara BB, et al.: Radiomics-based machine learning models in STEMI: a promising tool for the prediction of major adverse cardiac events. *Eur Radiol* 2023; 33:4611–4620.

41. Jang J, Ngo LH, Mancio J, et al.: Reproducibility of segmentation-based myocardial radiomic features with cardiac MRI. *Radiol Cardiothorac Imaging* 2020; 2:e190216.

42. Messroghli DR, Radjenovic A, Kozerke S, Higgins DM, Sivananthan MU, Ridgway JP: Modified Look-Locker inversion recovery (MOLLI) for high-resolutionT1 mapping of the heart. *Magn Reson Med* 2004; 52:141–146.

43. Giri S, Chung Y-C, Merchant A, et al.: T2 quantification for improved detection of myocardial edema. *J Cardiovasc Magn Reson* 2009; 11:56.

44. Kikinis R, Pieper SD, Vosburgh KG: 3D Slicer: a platform for subject-specific image analysis, visualization, and clinical support. In *Intraoperative Imaging and Image-Guided Therapy*. Edited by Jolesz FA. New York, NY: Springer New York; 2014:277–289.

45. Fedorov A, Beichel R, Kalpathy-Cramer J, et al.: 3D Slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging* 2012; 30:1323–1341.

46. Zwanenburg A, Leger S, Vallières M, Löck S: Image biomarker standardisation initiative. 2016.

47. Tixier F, Le Rest CC, Hatt M, et al.: Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med* 2011; 52:369–378.

48. Yip SSF, Aerts HJWL: Applications and limitations of radiomics. *Phys Med Biol* 2016; 61:R150–R166.

49. van Griethuysen JJM, Fedorov A, Parmar C, et al.: Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017; 77:e104–e107.

50. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86:420–428.

51. McGraw KO, Wong SP: Forming inferences about some intraclass correlation

coefficients. *Psychological Methods* 1996; 1:30–46.

52. Liljequist D, Elfving B, Skavberg Roaldsen K: Intraclass correlation – A discussion and demonstration of basic features. *PLoS ONE* 2019; 14:e0219854.

53. Xue C, Yuan J, Lo GG, et al.: Radiomics feature reliability assessed by intraclass correlation coefficient: a systematic review. *Quant Imaging Med Surg* 2021; 11:4431–4460.

54. Bland JM, Altman DG: Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8:135–160.

55. Raunig DL, McShane LM, Pennello G, et al.: Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. *Stat Methods Med Res* 2015; 24:27–67.

56. R: a language and environment for statistical computing. [https://www.R-project.org/]

57. Gourtsoyianni S, Doumou G, Prezzi D, et al.: Primary rectal cancer: Repeatability of global and local-regional MR imaging texture features. *Radiology* 2017; 284:552–561.

58. Baessler B, Weiss K, Pinto dos Santos D: Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. *Invest Radiol* 2019; 54:221–228.

59. Fiset S, Welch ML, Weiss J, et al.: Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. *Radiother Oncol* 2019; 135:107–114.

60. Peerlings J, Woodruff HC, Winfield JM, et al.: Stability of radiomics features in apparent diffusion coefficient maps from a multi-centre test-retest trial. *Sci Rep* 2019; 9:4800.

61. Shiri I, Hajianfar G, Sohrabi A, et al.: Repeatability of radiomic features in magnetic resonance imaging of glioblastoma: test–retest and image registration analyses. *Med Phys* 2020; 47:4265–4280.

FIGURE CAPTIONS:

Figure 1. ROI delineation in test-retest acquisitions for a representative subject. T1 maps from test (a) and retest (c) acquisitions, with their corresponding segmentations ((b) and (d), respectively) of the entire myocardium.

Figure 2. Test-retest repeatability results, in terms of ICC, of each radiomic feature from different classes (i.e., shape, first-order, GLCM, GLRLM, GLSZM, GLDM, and NGTDM), for both T1 and T2 maps. The heatmaps show stratified ICC values.

Figure 3. Bland-Altman plots for radiomic features derived from T1 maps, with both ICC > 0.75 and LOA between ± 5%. The percentage difference (%) between two repeated measurements on the same subject is plotted against their mean value. Dashed red lines indicate the 2.5^{th} and 97.5^{th} percentile values, while dashed black line represents the median value of the percentage differences between repeated measurements.

Figure 4. Bland-Altman plots for radiomic features derived from T2 maps, with both ICC > 0.75 and LOA between ± 5%. The percentage difference (%) between two repeated measurements on the same subject is plotted against their mean value. Dashed red lines indicate the 2.5th and 97.5th percentile values, while dashed black line represents the median value of the percentage differences between repeated measurements.

Table 1. Enrolled subjects: clinical indication for cardiac MR imaging

Clinical indication	Number of subjects
Amyloidosis	1
Acute myocarditis	2
Suspicion of right ventricular cardiomyopathy	2
Suspicion of left ventricular cardiomyopathy	4
Prior myocardial infarction	4
Hypertrophic cardiomyopathy	5
Healthy subjects	6

Table 2. Limits of agreement analysis for radiomic features derived from T1 maps

	2.5th percentile	97.5th percentile		2.5th percentile	97.5th percentile
Shape			**GLRLM**		
Elongation	-9.05	12.38	GrayLevelNonUniformity	-41.31	36.63
MajorAxisLength	-6.38	6.72	GrayLevelNonUniformityNormalized	-33.75	32.78
MaximumDiameter	-7.14	5.04	GrayLevelVariance	-62.90	61.43
MeshSurface	-23.74	26.27	HighGrayLevelRunEmphasis	-106.35	71.73
MinorAxisLength	-7.31	10.71	LongRunEmphasis	-7.00	5.19
Perimeter	-3.87	8.44	LongRunHighGrayLevelEmphasis	-104.70	69.92
PerimeterSurfaceRatio	-25.31	29.32	LongRunLowGrayLevelEmphasis	-81.51	87.38
PixelSurface	-21.31	25.11	LowGrayLevelRunEmphasis	-73.85	88.76
Sphericity	-17.28	11.72	RunEntropy	-8.52	6.98
			RunLengthNonUniformity	-19.91	22.76
First-order			RunLengthNonUniformityNormalized	-2.75	4.85
Mean	-2.22	1.71	RunPercentage	-1.61	2.43
Variance	-61.06	60.25	RunVariance	-36.02	47.16
Skewness	-1117.83	950.28	ShortRunEmphasis	-1.10	1.93
Kurtosis	-89.12	73.61	ShortRunHighGrayLevelEmphasis	-106.34	72.61
Median	-2.46	1.48	ShortRunLowGrayLevelEmphasis	-74.11	89.58
Minimum	-9.03	13.60			
10Percentile	-3.87	2.38	**GLSZM**		
90Percentile	-2.62	4.37	GrayLevelNonUniformity	-42.57	41.86
Maximum	-23.05	15.35	GrayLevelNonUniformityNormalized	-30.99	36.63
InterquartileRange	-28.24	41.74	GrayLevelVariance	-69.06	65.33
MeanAbsoluteDeviation	-25.09	30.81	HighGrayLevelZoneEmphasis	-106.14	72.11
RobustMeanAbsoluteDeviation	-26.32	37.57	LargeAreaEmphasis	-24.04	15.72
Energy	-24.74	25.41	LargeAreaHighGrayLevelEmphasis	-97.20	70.24
RootMeanSquared	-2.24	1.73	LargeAreaLowGrayLevelEmphasis	-91.58	90.39
Entropy	-7.76	8.00	LowGrayLevelZoneEmphasis	-73.64	92.58
Uniformity	-34.19	31.12	SizeZoneNonUniformity	-20.60	20.21
			SizeZoneNonUniformityNormalized	-9.72	11.88
GLCM			SmallAreaEmphasis	-4.36	5.23
MaximumProbability	-37.26	56.30	SmallAreaHighGrayLevelEmphasis	-105.89	76.25
JointAverage	-62.30	38.52	SmallAreaLowGrayLevelEmphasis	-74.69	97.27
SumSquares	-56.41	61.65	ZoneEntropy	-8.80	5.64
JointEntropy	-7.00	7.39	ZonePercentage	-5.33	7.54
DifferenceAverage	-19.20	16.88	ZoneVariance	-55.40	56.70
DifferenceVariance	-86.63	49.80			
DifferenceEntropy	-4.73	5.79	**GLDM**		
SumEntropy	-7.22	7.22	DependenceEntropy	-9.37	5.50
JointEnergy	-35.61	39.05	DependenceNonUniformity	-20.49	21.89
Contrast	-63.95	38.63	DependenceNonUniformityNormalized	-15.27	17.66
Id	-12.38	12.34	DependenceVariance	-51.27	52.68
Idn	-4.21	3.83	GrayLevelNonUniformity	-39.56	34.18
Idm	-23.01	21.80	GrayLevelVariance	-60.68	60.51
Idmn	-1.42	1.78	HighGrayLevelEmphasis	-106.41	71.60
InverseVariance	-31.59	28.39	LargeDependenceEmphasis	-27.78	22.53
Correlation	-31.49	78.65	LargeDependenceHighGrayLevelEmphasis	-98.94	70.75
Autocorrelation	-107.11	75.49	LargeDependenceLowGrayLevelEmphasis	-110.66	107.71
ClusterTendency	-50.77	73.00	LowGrayLevelEmphasis	-74.05	87.43
ClusterShade	-4580.71	360.37	SmallDependenceEmphasis	-7.68	9.88
ClusterProminence	-120.90	150.56	SmallDependenceHighGrayLevelEmphasis	-106.91	78.84
Imc1	-20.21	22.98	SmallDependenceLowGrayLevelEmphasis	-76.11	93.63
Imc2	-0.46	2.00			
NGTDM					
Busyness	-89.07	81.13			
Coarseness	-22.78	41.35			
Complexity	-67.66	60.03			
Contrast	-66.84	45.59			
Strength	-60.73	84.15			

Table 3. Limits of agreement analysis for radiomic features derived from T2 maps

	2.5th percentile	97.5th percentile		2.5th percentile	97.5th percentile
Shape			**GLRLM**		
Elongation	-10.69	8.28	GrayLevelNonUniformity	-41.99	39.84
MajorAxisLength	-8.26	7.84	GrayLevelNonUniformityNormalized	-33.67	31.29
MaximumDiameter	-2.92	4.50	GrayLevelVariance	-71.71	86.21
MeshSurface	-31.30	22.82	HighGrayLevelRunEmphasis	-84.49	107.32
MinorAxisLength	-7.07	6.81	LongRunEmphasis	-5.39	6.36
Perimeter	-4.87	5.71	LongRunHighGrayLevelEmphasis	-84.66	106.85
PerimeterSurfaceRatio	-25.31	31.71	LongRunLowGrayLevelEmphasis	-121.99	112.97
PixelSurface	-30.98	21.75	LowGrayLevelRunEmphasis	-112.76	107.58
Sphericity	-17.55	14.03	RunEntropy	-8.14	8.05
			RunLengthNonUniformity	-30.88	21.29
First-order			RunLengthNonUniformityNormalized	-3.42	3.79
Mean	-5.00	6.61	RunPercentage	-1.96	1.91
Variance	-70.80	87.22	RunVariance	-51.96	45.50
Skewness	-187.92	135.76	ShortRunEmphasis	-1.41	1.50
Kurtosis	-75.84	98.78	ShortRunHighGrayLevelEmphasis	-85.72	107.37
Median	-5.84	5.44	ShortRunLowGrayLevelEmphasis	-109.92	106.13
Minimum	-16.73	17.80			
10Percentile	-6.93	4.64	**GLSZM**		
90Percentile	-10.69	11.10	GrayLevelNonUniformity	-42.77	36.24
Maximum	-18.17	42.25	GrayLevelNonUniformityNormalized	-33.89	30.07
InterquartileRange	-48.87	52.17	GrayLevelVariance	-69.95	84.50
MeanAbsoluteDeviation	-41.04	44.58	HighGrayLevelZoneEmphasis	-84.18	105.06
RobustMeanAbsoluteDeviation	-48.64	45.67	LargeAreaEmphasis	-21.67	25.49
Energy	-33.73	22.51	LargeAreaHighGrayLevelEmphasis	-82.96	103.13
RootMeanSquared	-5.18	8.42	LargeAreaLowGrayLevelEmphasis	-137.31	122.39
Entropy	-9.66	9.67	LowGrayLevelZoneEmphasis	-102.68	102.53
Uniformity	-33.34	31.80	SizeZoneNonUniformity	-42.29	25.74
			SizeZoneNonUniformityNormalized	-12.39	10.66
GLCM			SmallAreaEmphasis	-5.89	4.75
MaximumProbability	-41.45	42.47	SmallAreaHighGrayLevelEmphasis	-89.01	105.18
JointAverage	-53.16	64.22	SmallAreaLowGrayLevelEmphasis	-84.01	122.11
SumSquares	-75.00	92.73	ZoneEntropy	-6.41	6.06
JointEntropy	-8.86	5.50	ZonePercentage	-7.76	7.03
DifferenceAverage	-37.85	54.17	ZoneVariance	-65.76	59.55
DifferenceVariance	-90.45	113.92			
DifferenceEntropy	-14.14	11.99	**GLDM**		
SumEntropy	-9.44	7.48	DependenceEntropy	-6.21	5.51
JointEnergy	-30.22	42.75	DependenceNonUniformity	-39.46	26.78
Contrast	-80.97	104.16	DependenceNonUniformityNormalized	-17.16	16.68
Id	-24.68	20.71	DependenceVariance	-54.24	56.14
Idn	-2.32	2.56	GrayLevelNonUniformity	-42.19	40.98
Idm	-30.84	31.01	GrayLevelVariance	-71.86	87.03
Idmn	-1.10	1.20	HighGrayLevelEmphasis	-84.50	108.27
InverseVariance	-29.47	30.15	LargeDependenceEmphasis	-24.88	25.51
Correlation	-54.29	68.31	LargeDependenceHighGrayLevelEmphasis	-91.42	105.70
Autocorrelation	-90.86	109.49	LargeDependenceLowGrayLevelEmphasis	-150.11	122.52
ClusterTendency	-87.51	93.99	LowGrayLevelEmphasis	-115.32	109.38
ClusterShade	-156.86	240.71	SmallDependenceEmphasis	-10.44	9.20
ClusterProminence	-130.28	170.78	SmallDependenceHighGrayLevelEmphasis	-93.54	108.74
Imc1	-32.71	29.70	SmallDependenceLowGrayLevelEmphasis	-87.31	116.29
Imc2	-4.37	2.24			
NGTDM					
Busyness	-100.10	86.96			
Coarseness	-34.07	46.10			
Complexity	-96.71	104.95			
Contrast	-54.35	76.24			
Strength	-81.19	88.89			