

ORIGINAL ARTICLE

Open Access



# Acquisition repeatability of MRI radiomics features in the head and neck: a dual-3D-sequence multi-scan study

Cindy Xue<sup>1</sup>, Jing Yuan<sup>1\*</sup>, Yihang Zhou<sup>1</sup>, Oi Lei Wong<sup>1</sup>, Kin Yin Cheung<sup>2</sup> and Siu Ki Yu<sup>2</sup>

## Abstract

Radiomics has increasingly been investigated as a potential biomarker in quantitative imaging to facilitate personalized diagnosis and treatment of head and neck cancer (HNC), a group of malignancies associated with high heterogeneity. However, the feature reliability of radiomics is a major obstacle to its broad validity and generality in application to the highly heterogeneous head and neck (HN) tissues. In particular, feature repeatability of radiomics in magnetic resonance imaging (MRI) acquisition, which is considered a crucial confounding factor of radiomics feature reliability, is still sparsely investigated. This study prospectively investigated the acquisition repeatability of 93 MRI radiomics features in ten HN tissues of 15 healthy volunteers, aiming for potential magnetic resonance-guided radiotherapy (MRgRT) treatment of HNC. Each subject underwent four MRI acquisitions with MRgRT treatment position and immobilization using two pulse sequences of 3D T1-weighted turbo spin-echo and 3D T2-weighted turbo spin-echo on a 1.5T MRI simulator. The repeatability of radiomics feature acquisition was evaluated in terms of the intraclass correlation coefficient (ICC), whereas within-subject acquisition variability was evaluated in terms of the coefficient of variation (CV). The results showed that MRI radiomics features exhibited heterogeneous acquisition variability and uncertainty dependent on feature types, tissues, and pulse sequences. Only a small fraction of features showed excellent acquisition repeatability (ICC > 0.9) and low within-subject variability. Multiple MRI scans improved the accuracy and confidence of the identification of reliable features concerning MRI acquisition compared to simple test-retest repeated scans. This study contributes to the literature on the reliability of radiomics features with respect to MRI acquisition and the selection of reliable radiomics features for use in modeling in future HNC MRgRT applications.

**Keywords:** Radiomics, Magnetic resonance guided radiotherapy, Head and neck, Repeatability, Intraclass correlation coefficient

## Introduction

Head and neck cancer (HNC) represents a group of malignancies associated with high heterogeneity, not only in terms of organs and tissues of origin, but also etiological, molecular, and mutational differences [1]. The global incidence of HNC has been continuously rising in recent decades [2]. Treatment options for HNC treatment

include surgery, radiation therapy (RT), chemotherapy, targeted therapy, immunotherapy, and combinations of these methods. However, the heterogeneity of HNC partially accounts for the frequency of unsatisfactory treatment outcomes, particularly in advanced stages [3].

Traditionally, magnetic resonance imaging (MRI) has played an important role in the diagnosis, prognosis, and treatment planning of HNC by virtue of its superior soft-tissue image contrast and various functional imaging capabilities [4–7]. In recent years, with the introduction of treatment with MRI-integrated linear accelerator (MR-LINAC) systems to clinical use [8–10], the role of MRI in

\*Correspondence: [jyuanbwh@gmail.com](mailto:jyuanbwh@gmail.com)

<sup>1</sup> Research Department, Hong Kong Sanatorium & Hospital, Hong Kong, China

Full list of author information is available at the end of the article

HNC has considerably extended from conventional diagnosis to image-based guidance of radiation delivery in RT, referred to as MR-guided radiotherapy (MRgRT) [9]. Despite in its infancy, MRgRT has shown promise as an innovative technique for HNC treatment [11–13], allowing for better delineation of target organs and organs at risk (OARs), daily treatment plan adaptation, real-time motion monitoring, gating, and tracking for dose delivery, as well as intra- and inter-fractional treatment response evaluation.

There has been an increased demand for developing biomarkers to facilitate personalized diagnosis and treatment of HNC. Radiomics [14–16] has attracted increasing research interest as a multidimensional data mining technique in medical imaging for the diagnosis and prognosis of HNC in recent years [17–21]. Despite the promising results reported in these studies, the reliability of radiomics features remains a major obstacle to the broad validity and generality of radiomics in routine clinical use [22–24].

Image acquisition is a crucial procedure that substantially influences radiomics feature values for all imaging modalities, in particular for MRI [25–30]. Firstly, the image intensity of normal anatomical MRI is semi-quantitative, being comprehensively influenced by many tissue properties such as relaxation times, proton density, fat-water composition, and susceptibility, without representing an exact physical meaning. Second, different hardware and configurations of MRI scanners from various vendors considerably impact image quality and characteristics and thus radiomics features. Third, the variety of MRI pulse sequences, imaging parameters, and reconstruction algorithms also dramatically influences MR image contrast and quality. Moreover, radiomics features values can also vary with organ motion during acquisition and the administration of contrast agents.

To facilitate radiomics in HNC MRgRT, the acquisition repeatability of MRI radiomics features prior to the use of these features must be investigated directly for diagnosis or prognosis modeling. Of note, MRgRT exhibits some unique characteristics compared to diagnostic MRI. In contrast to diagnostic MRI, in which one (for cross-sectional studies) or two (for longitudinal studies) MRI scans per patient are normally involved, multiple MRI scans are required in MRgRT fractions to obtain the required daily anatomical information for treatment adaptation. MRI acquisition in MRgRT relies heavily on 3D pulse sequences to obtain isotropic voxel sizes and better geometric fidelity than 2D sequences. In particular, 3D T2-weighted (T2W) turbo spin-echo (TSE) is heavily utilized in MRgRT without the need for administration of a contrast agent. In addition, patients are typically scanned with flexible RF coils that are compatible with the immobilized treatment position, rather than diagnostic volumetric coils. Finally, radiomics in MRgRT mainly utilizes

within-subject inter-fractional longitudinal radiomics feature variation for response evaluation and treatment adaptation, which is also known as delta-radiomics [31, 32], in contrast to the between-subject radiomics feature difference used in diagnostic radiology to perform lesion differentiation or characterization.

Thus, in this study, several aims were considered in investigating the repeatability of the acquisition of MRI radiomics features for MRgRT applications specifically. They were (1) to identify repeatable MRI features in two pulse sequences of 3D T1-weighted turbo spin-echo (3D-T1-TSE) and 3D T2-weighted turbo spin-echo (3D-T2-TSE), (2) to investigate whether feature acquisition repeatability varies with different HN tissues; (3) to evaluate whether and how a multi-scan study design could impact the determination of repeatable radiomics features compared to the commonly used test-retest (repeated scan) study design, and (4) to establish a benchmark for feature variability in MRI acquisition from normal HN tissues of use in future research on delta-radiomics in MRgRT.

## Methods

This study was approved by the research ethics committee of Hong Kong Sanatorium and Hospital. A total of 15 healthy volunteers (8 men and 7 women with ages ranging from 24 to 40 years) were prospectively recruited for this study. Informed consent was obtained from each subject.

### MRI acquisition

All MRI scans were conducted using a 1.5 T MRI scanner dedicated to radiotherapy simulation (MR-sim) (Magnetom Aera, Siemens Healthineers, Erlangen, Germany). Each volunteer underwent four scans (with an interval of approximately 15 mins between scan) while immobilized with a 5-pin head, neck, and shoulder thermoplastic mask (Orfit Industries, Belgium). For each scan, the volunteers were precisely aligned using a 3-dimensional external laser (DORADOnova MR3T, LAP GmbH Laser Applikationen, Luneburg, Germany) and scanned in the same treatment position on an RT-indexed flat couch top (Diacor, Salt Lake City, Utah, USA). Two flexible 4-channel flexible coils, one 18-channel flexible coil, and a spine coil were used in combination for MRI signal reception. Each scan consisted of a 3D-T1W-TSE sequence followed by a 3D-T2W-TSE sequence. A vendor-provided received B1 field-inhomogeneity correction, i.e., prescan normalization, was conducted to minimize the bias field of the MR images. 3D geometric distortion correction was also enabled on the console for image acquisition. The

**Table 1** Imaging parameters of the two 3D pulse sequences

Scanning Sequence	3D-T1W-TSE	3D-T2W-TSE
TR/TE [ms]	420/7.2	2300/303
Echo train length	40	185
FOV (LR × SI × AP) [mm <sup>3</sup> ]	470 × 470 × 269	470 × 470 × 269
Matrix size (LR × SI × AP)	448 × 448 × 256	448 × 448 × 256
Voxel size [mm <sup>3</sup> ]	1.05 × 1.05 × 1.05	1.05 × 1.05 × 1.05
Acceleration factor (GRAPPA)	3	3
Partial Fourier factor	6/8	6/8
Pixel Bandwidth [Hz/pixel]	657	620
Scan duration (mm:ss)	05:01	05:24

LR Left/right, SI Superior/inferior, AP Anterior/posterior, GRAPPA Generalized autocalibrating partial parallel acquisition

imaging parameters prescribed for each sequence are listed in Table. 1.

#### Image registration and volumes-of-interest drawing

The digital imaging and communications in medicine format MR images were imported to the 3D Slicer v 4.10.2 [33] for registration and volume of interest (VOI) drawing. The first-scan 3D-T1W-TSE MRI set was used as the positional reference for image registration. Other images were rigidly registered to the reference 3D-T1W-TSE images to compensate for the residual positional shifts, although these shifts have been reported to be very small (approximately several millimeters) owing to the immobilization of the subject by the thermoplastic mask [34, 35].

Ten spherical or ellipsoidal VOIs of pons ( $2075.34 \pm 675.17 \text{ mm}^3$ ), left (L) and right (R) parotid glands (L:  $5693.71 \pm 2614.65 \text{ mm}^3$ , R:  $5336.53 \pm 2497.18 \text{ mm}^3$ ), mandible ( $1008.58 \pm 526.79 \text{ mm}^3$ ), tongue ( $4077.71 \pm 1030.59 \text{ mm}^3$ ), L/R pterygoid muscle (L:  $2455.70 \pm 992.56 \text{ mm}^3$ , R:  $2290.83 \pm 927.36 \text{ mm}^3$ ), thyroid ( $550.40 \pm 243.95 \text{ mm}^3$ ), and L/R submandibular gland (L:  $2241.13 \pm 966.31 \text{ mm}^3$ , R:  $2255.08 \pm 880.35 \text{ mm}^3$ ) were manually drawn by an MRI physicist on the first-scan (reference) 3D-T1W-TSE images and validated by a second MRI physicist. Then, all VOIs were propagated to other registered image sets and visually checked by both MRI physicists to ensure the tissue coherence of the propagated VOIs. A typical MRI scan setup and VOIs overlaid on the 3D-T1W-TSE and 3D-T2W-TSE images are shown in Fig. 1.

#### Radiomics feature extraction and calculation

3D radiomics features were calculated using PyRadiomics v.2.2.0 [36]. Ninety-three first-order and texture radiomics

features in five categories of gray-level co-occurrence matrix (GLCM), gray-level dependence matrix (GLDM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), and neighboring gray-tone difference matrix (NGTDM), mostly compliant with Image Biomarker Standardization Initiative (IBSI) standards [37, 38] (first-order,  $n = 18$ ; texture\_GLDM,  $n = 24$ ; texture\_GLDM,  $n = 14$ ; texture\_GLRLM,  $n = 16$ ; texture\_GLSZM,  $n = 16$ ; texture\_NGTDM,  $n = 5$ ) were extracted from the original MRI images. Shape radiomics features were not included in the analysis mainly because they were theoretically independent of MRI acquisition and constant VOIs were applied for all MRI datasets. For the mathematical definition of each radiomics feature, the reader is referred to the PyRadiomics documentation (<https://pyradiomics.readthedocs.io/en/latest/features.html>).

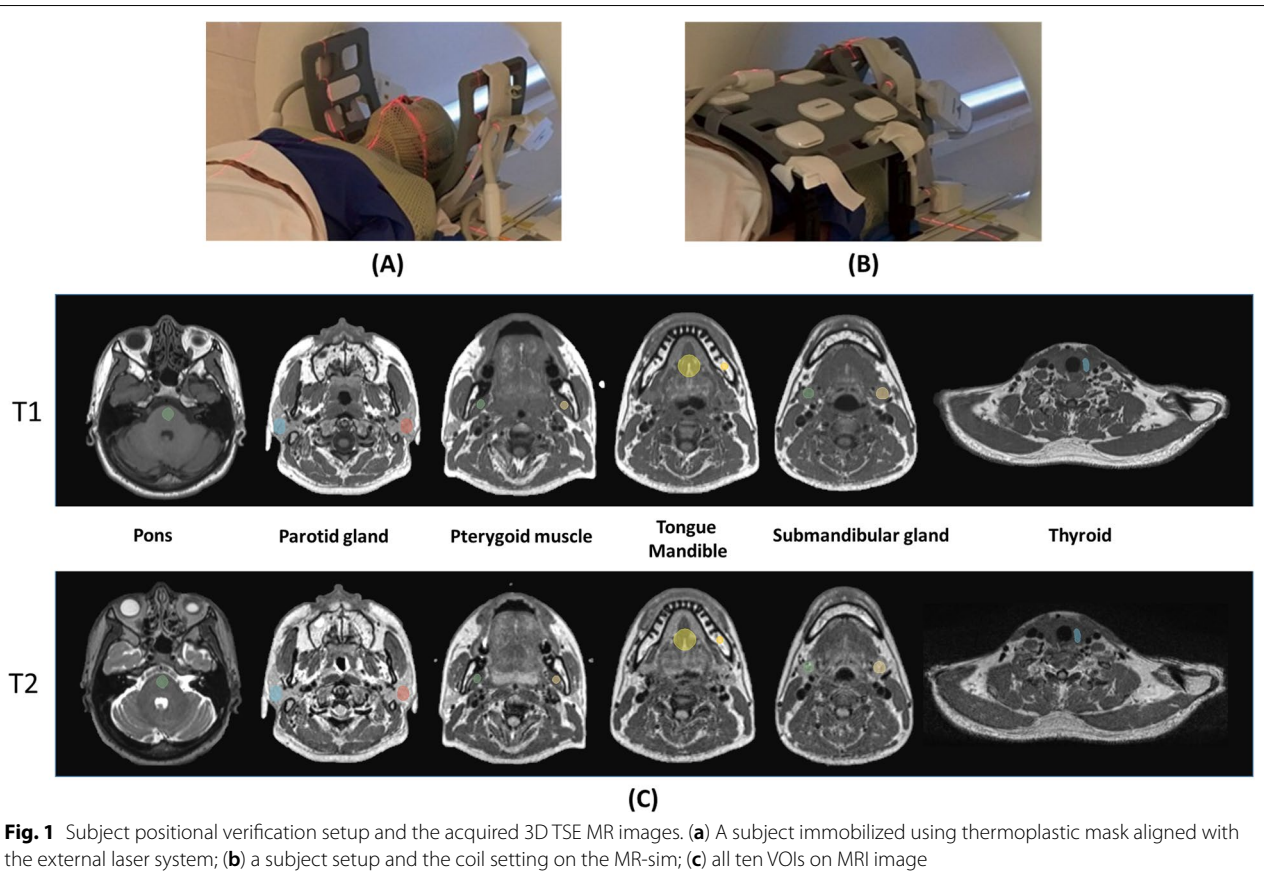
The default bin size of 25 in the software was used to perform image intensity discretization. No scaling of the image voxel size was applied due to the isotropic voxel size of the acquired images. Image intensity normalization was not performed because the images were acquired using a single MRI scanner with fixed imaging parameters. No image denoising, filtering, or other post-processing was conducted prior to radiomics feature calculation to minimize their influence on feature values [39]. Default configuration settings were applied in PyRadiomics for feature calculation unless otherwise specified.

#### Data analysis

##### *Inter-scan acquisition repeatability of radiomics features*

The intraclass correlation coefficient (ICC) (2-way mixed effects, absolute agreement, single rater) calculated based on all four MRI scans was used to assess the acquisition repeatability of radiomics features. The feature repeatability was classified as excellent ( $\text{ICC} > 0.9$ ), good ( $0.9 > \text{ICC} > 0.75$ ), moderate ( $0.75 > \text{ICC} > 0.5$ ), and poor ( $\text{ICC} < 0.5$ ) when the calculated ICC and its 95%CI were both within the thresholds according to Koo and Li [40]. If the 95%CI of the calculated ICC was located across two or more ranges, the corresponding feature was classified as the lowest repeatability class. Based on the ICC classification, the radiomics features showing excellent acquisition repeatability were identified for each VOI and each sequence. Then, features universally showing excellent acquisition repeatability in all VOIs for both pulse sequences were identified.

To assess whether multi-scan substantially affected the calculated ICC values and feature repeatability determination, the ICCs were also calculated based on the first two and the first three MRI scans, and compared to the corresponding ICCs based on all four MRI scans.



**Fig. 1** Subject positional verification setup and the acquired 3D TSE MR images. (a) A subject immobilized using thermoplastic mask aligned with the external laser system; (b) a subject setup and the coil setting on the MR-sim; (c) all ten VOIs on MRI image

The intra-subject radiomics feature variability due to multi-scan acquisition was quantified in terms of the coefficient of variation ( $CV_{\text{intra-subject}}$ ), defined as the ratio of the standard deviation (SD) to the mean of the radiomics feature values calculated from four MRI scans. Similarly, inter-subject radiomics feature variability was quantified by  $CV_{\text{inter-subject}}$ , defined as the ratio of the SD to the mean of radiomics feature values across all subjects.

#### Statistical analysis

Descriptive statistics were represented in the form of mean  $\pm$  SD. The Mann-Whitney U-test was conducted to compare the (4-scan derived) ICC values between T1 and T2 pulse sequences. The Kruskal-Wallis test was used to compare ICCs derived from two, three, and four MRI scans for each sequence. The Mann-Whitney U-test and Wilcoxon signed-rank test were also conducted to compare the CVs for different VOIs and feature categories in each pulse sequence. A  $p$ -value smaller than 0.05 indicated statistical significance. All statistical tests were conducted using RStudio 2021.09.0 Build 351 (RStudio PBC, Boston, MA, USA).

## Results

#### Inter-scan acquisition repeatability of radiomics features

The repeatability of radiomics feature acquisition assessed by ICC varied with feature categories, VOIs, and pulse sequences. Figure 2 shows boxplots of ICC in different VOIs for both pulse sequences. As shown in Fig. 2, the ICC values were significantly different (ANOVA,  $p < 0.001$ ) between different VOIs for both pulse sequences. In general, ICCs associated with 3D-T1W-TSE ( $0.418 \pm 0.270$ ) were significantly lower ( $p < 0.001$ ) than those associated with 3D-T2W-TSE ( $0.473 \pm 0.249$ ), implying that better feature acquisition repeatability could be obtained with the 3D-T2W-TSE sequence. For 3D-T1W-TSE, the VOI of the R parotid gland showed the highest ICCs ( $0.539 \pm 0.250$ ), whereas the VOI of the thyroid showed the lowest ( $0.283 \pm 0.355$ ). In comparison, for 3D-T2W-TSE, the VOI of the R parotid gland also showed the highest ICCs ( $0.548 \pm 0.285$ ), whereas the VOI of the R pterygoid muscle showed the lowest ICCs ( $0.316 \pm 0.258$ ). The ICCs of the paired VOIs of L/R parotid gland, L/R pterygoid muscle, and L/R submandibular gland did not differ significantly (all  $p > 0.05$ ) for both pulse sequences.



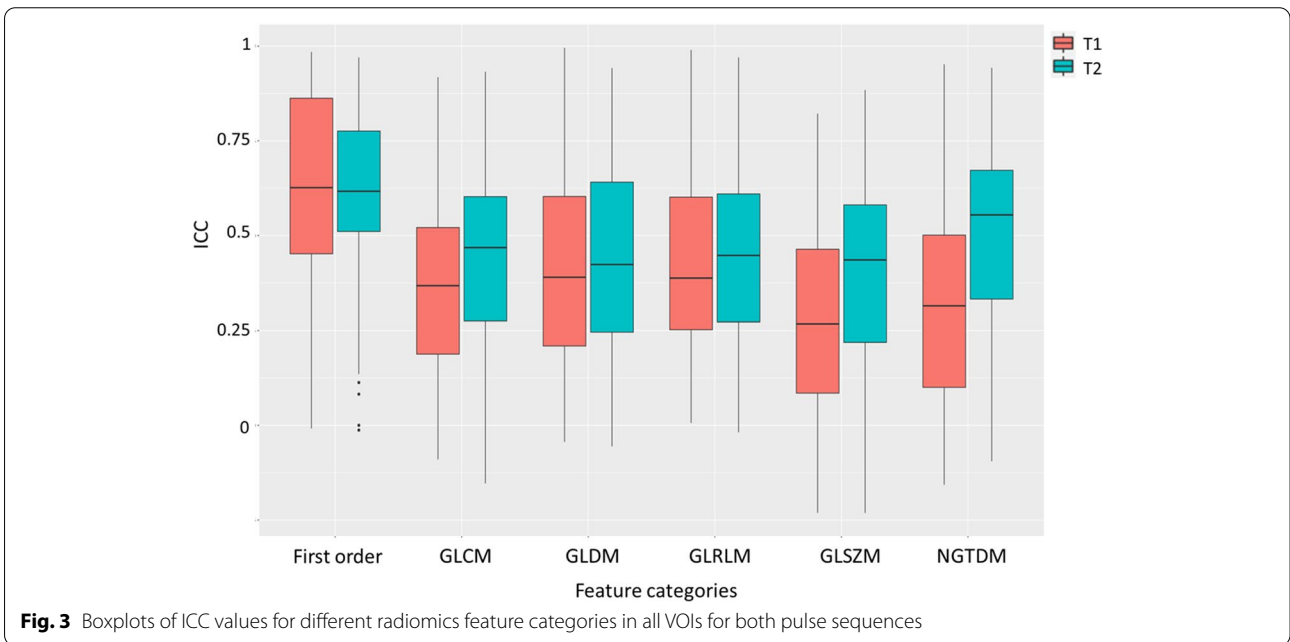
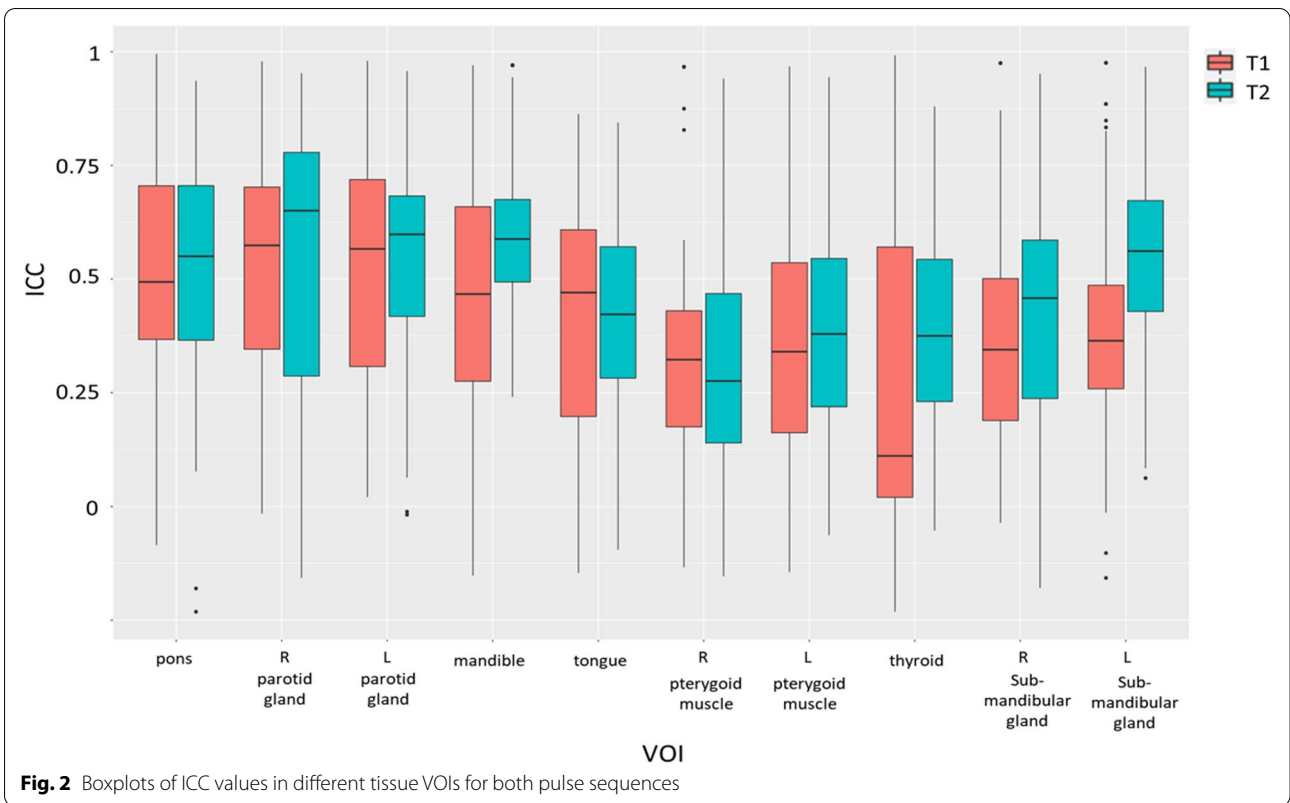
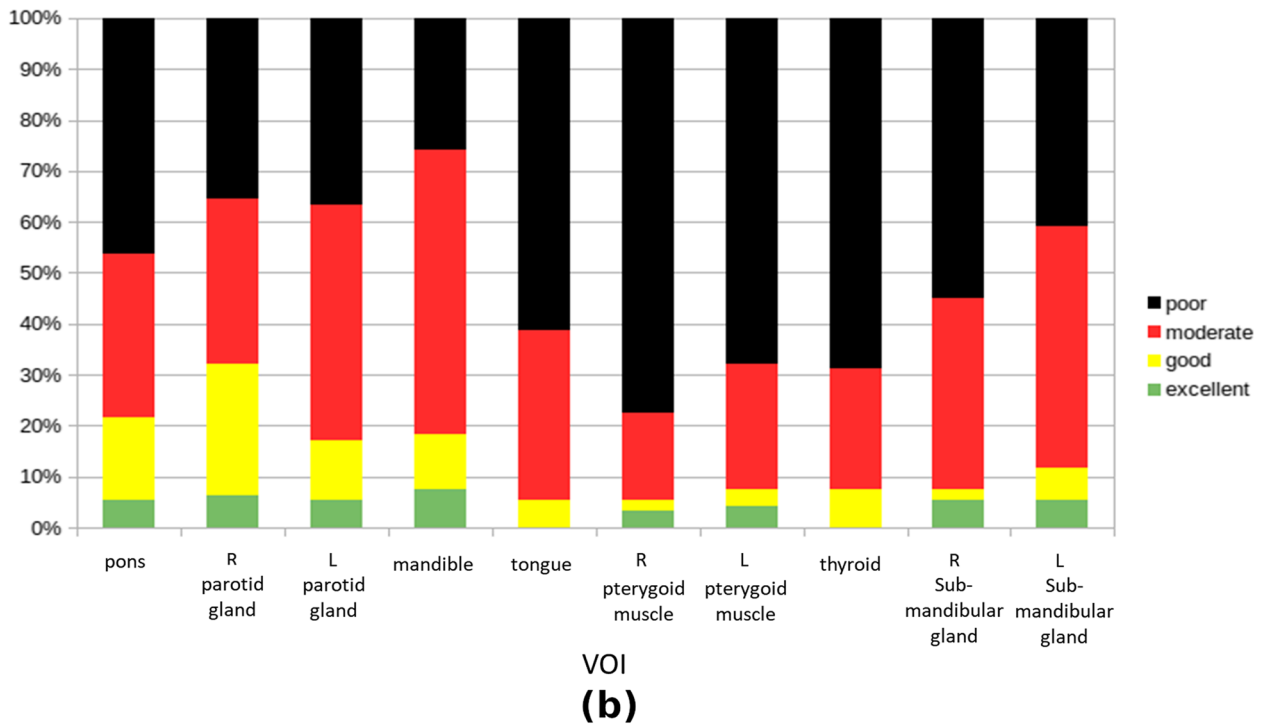
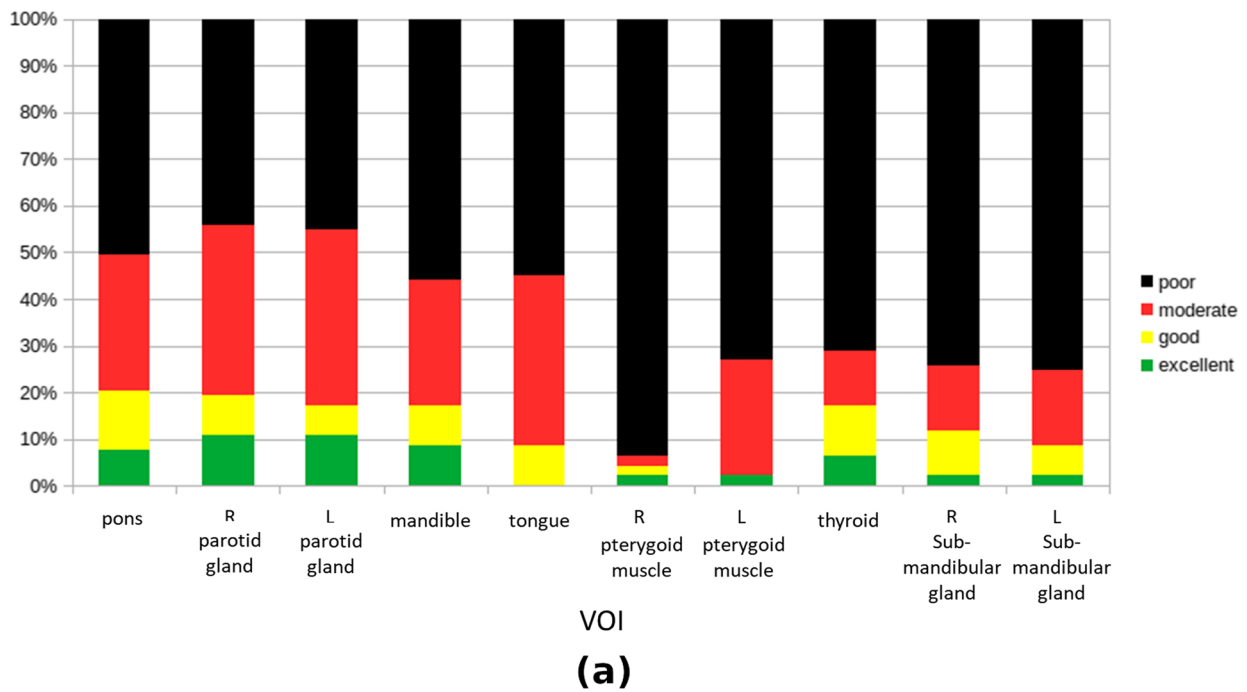


Figure 3 shows boxplots of ICCs of different radiomics feature categories for all VOIs for both pulse sequences. The ICCs of first-order radiomics features were significantly higher ( $p < 0.001$ ) than those

of texture radiomics features for both sequences. The GLSZM features exhibited the lowest ICCs (T1:  $0.279 \pm 0.242$ ; T2:  $0.387 \pm 0.262$ ) in all feature categories for both sequences.



**Fig. 4** Percentages of excellent, good, moderate, and poor ICCs of radiomics features in different tissue VOIs for (a) 3D-T1W-TSE and (b) 3D-T2W-TSE pulse sequences

Figure 4 illustrates the percentages of radiomics features showing excellent, good, moderate, and poor ICCs in different VOIs for the 3D-T1W-TSE and 3D-T2W-TSE sequences. There were only ( $5.27\% \pm 4.00\%$ ) and ( $4.41\% \pm 2.66\%$ ) radiomics features that showed excellent repeatability for 3D-T1W-TSE and 3D-T2W-TSE, respectively, in different VOIs. For both sequences, the VOI of the R parotid gland exhibited the largest number of excellent repeatability features (T1W: 10.75%, 10/93; T2W: 7.53%, 7/93), while tongue had no excellent repeatability feature at all.

Figure 5 shows the radiomics features with their acquisition repeatability classifications for each VOI and each sequence. Only four features showed mostly good or excellent ICCs in all VOIs (except for the tongue), without significant differences in the ICC between the two sequences. They were `firstorder_Energy`, `firstorder_TotalEnergy`, `GLDM_GrayLevelNonUniformity`, and `GLRLM_GrayLevelNonUniformity`. These four features were highly repeatable and robust to image acquisition using the two 3D pulse sequences.

For both 3D-T1W-TSE and 3D-T2W-TSE, the ICCs calculated from two, three, and four repeated MRI scans were not significantly different from each other (all  $p > 0.05$ ). However, the 95% CIs associated with the ICCs calculated from fewer repeated MRI scans widened significantly. Boxplots of ICC values based on two or three MRI scans in different tissue VOIs for both pulse sequences are shown in Supplementary Fig. 1. The ICCs calculated from two or three MRI scans moderately affected the repeatability of radiomics feature acquisition. Fewer radiomics features showed excellent or good repeatability, but more features showed poor repeatability due to the much wider 95% CI of the ICC (indicating a larger uncertainty of the calculated ICC value) calculated from fewer scans. The percentages of radiomics features showing excellent, good, moderate, and poor ICCs based on two and three MRI scans in different VOIs for two sequences were demonstrated by the bar plots given in Supplementary Fig. 2.

The heatmaps shown in Fig. 6 depict the  $CV_{\text{intra-subject}}$  of all radiomics features in different VOIs for both pulse sequences. The  $CV_{\text{intra-subject}}$  of features were  $24.57\% \pm 25.36\%$  and  $18.06\% \pm 23.34\%$  for 3D-T1W-TSE and 3D-T2W-TSE, respectively, exhibiting pronounced variability in the values of radiomics features in multi-scan acquisitions, with a significant difference ( $p < 0.001$ ) between the two sequences. In comparison, the  $CV_{\text{intra-subject}}$  was significantly lower than the  $CV_{\text{inter-subject}}$  (3D-T1W-TSE:  $49.70\% \pm 72.97\%$ ; 3D-T2W-TSE:  $51.22\% \pm 142.09\%$ ) ( $p < 0.05$ ). The results showed a significant difference in  $CV_{\text{intra-subject}}$  between the first-order and texture features for both 3D-T1W-TSE

and 3D-T2W-TSE ( $p < 0.05$ ). The boxplots in Fig. 7 illustrate the within-subject feature  $CV_{\text{intra-subject}}$  with regard to the tissue VOIs. There was a significant difference in  $CV_{\text{intra-subject}}$  ( $p < 0.05$ ) between tissues. The VOI of the mandible showed the smallest  $CV_{\text{intra-subject}}$  (T1:  $13.44\% \pm 9.97\%$ ; T2:  $12.80\% \pm 9.26\%$ ), whereas the VOI of the thyroid showed the largest CV (T1:  $47.17\% \pm 43.37\%$ ; T2:  $25.36\% \pm 30.79\%$ ) in both sequences.

## Discussion

This study prospectively investigated the multi-scan acquisition repeatability and variability of IBSI-compliant MRI radiomics features for two pulse sequences of 3D-T1W-TSE and 3D-T2W-TSE in a cohort of healthy volunteers, to support potential applications of MRI radiomics in HNC MRgRT. To the best of our knowledge, the present work is the first to investigate the repeatability of radiomics features for the acquisition of MRI data on the HN specifically.

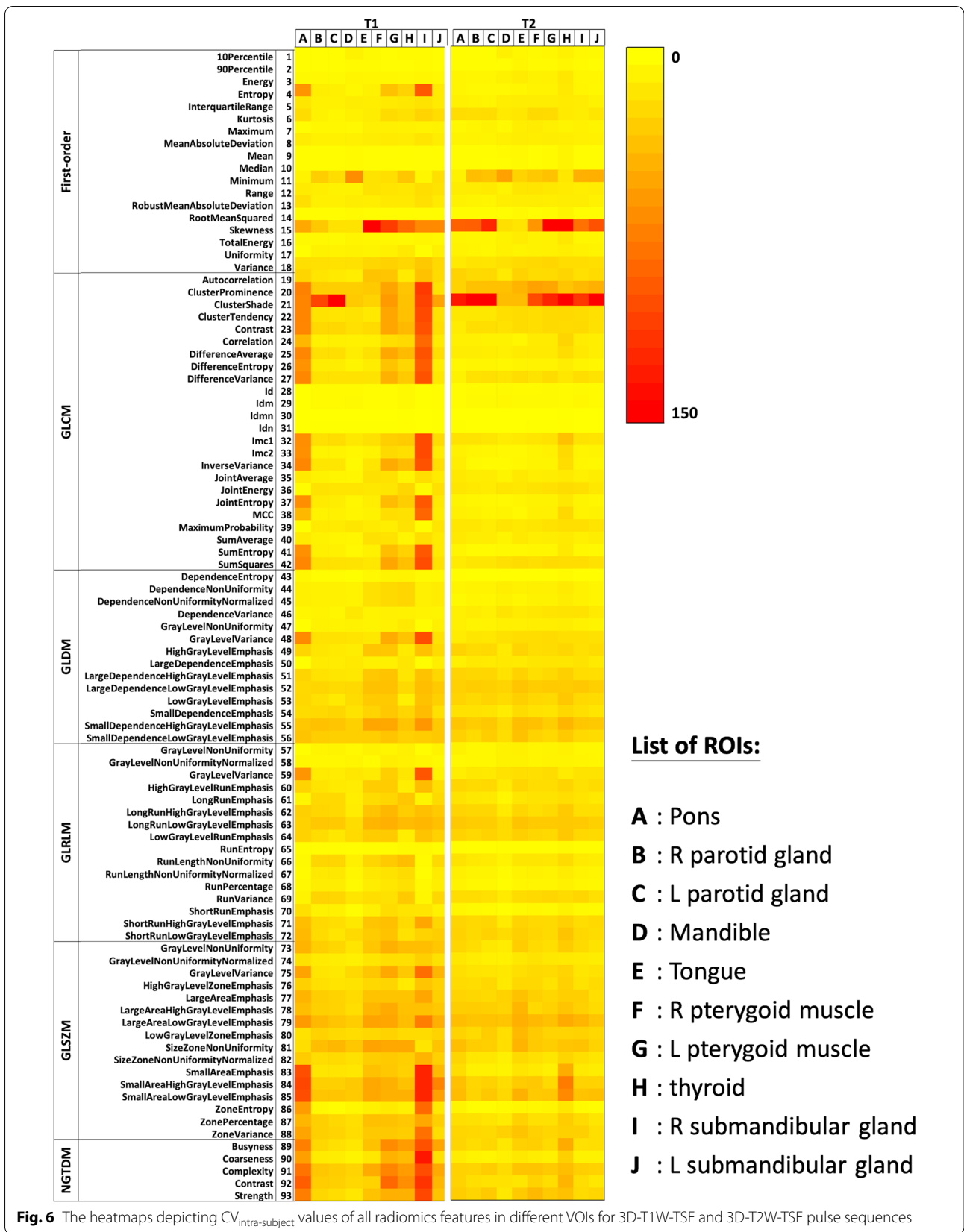
In recent years, an increasing number of studies have reported promising preliminary results on the use of radiomics for the diagnosis and prognosis of HNC [17–21]. However, the broad reproducibility, validity, and generality of radiomics remain open to question or challenge owing to the wide variety of confounding factors that could substantially impact every procedure of the complicated radiomics workflow and lead to uncertainty, instability, or unreliability of the results of radiomics analysis [22–24]. To date, the accumulated evidence remains insufficient to justify the deployment of radiomics in routine clinical practice.

This study specifically addressed an important source of MRI radiomics feature unreliability due to image acquisition, and identified repeatable features from two 3D sequences, which are expected to be helpful in the selection of reliable radiomics features and thus support modeling in future HNC MRgRT applications. The CVs presented in the results should also be useful to establish a reference benchmark of MRI radiomics feature variability for acquisition from normal HN tissues for future research on MRgRT delta-radiomics.

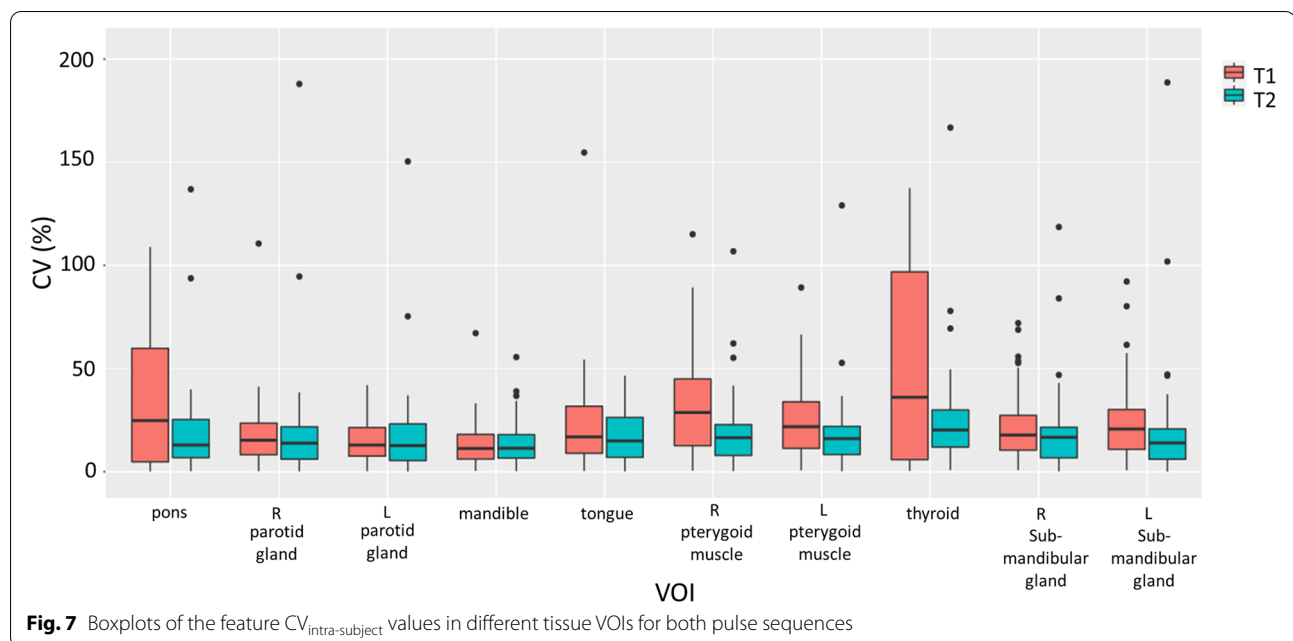
Some notable findings were observed that deserve discussion. First, only a very small proportion of the investigated features could achieve excellent multi-scan repeatability measured by ICC, which is generally consistent with the results of many previous MRI radiomics studies, although in different anatomies [25, 26, 30, 41–43]. However, the proportion of features showing excellent ICC obtained in this study was even smaller than that reported in prior works. In addition to high heterogeneity of the multiple HN tissues, this finding could also be attributed to the ICC calculation based on the four-scan MRI data and more stringent ICC classification by its 95% CI. First-order







**Fig. 6** The heatmaps depicting CV<sub>intra-subject</sub> values of all radiomics features in different VOIs for 3D-T1W-TSE and 3D-T2W-TSE pulse sequences



features were more repeatable than texture features, which also accords with previous studies [26, 44–47]. Second, regarding sequence dependence, the mean feature ICC was better with 3D-T2W-TSE than with 3D-T1W-TSE for most tissue VOIs (Fig. 2) and for most feature categories (Fig. 3). Fewer features showed poor ICC ( $<0.5$ ) in 3D-T2W-TSE than in 3D-T1W-TSE (Fig. 4). These results might be partially explained by the fact that T2W MR Images typically exhibit a wider voxel intensity range than T1W MR Images. That is, T1W MR Images have more uniform intensities in many tissues, which causes many texture features to show small inter-subject differences in their values, thus leading to low ICC values. However, the number of features showing excellent ICC was not necessarily larger in 3D-T2W-TSE for the different VOIs (Fig. 4). Third, feature repeatability was also found to be tissue-dependent. Different tissues could exhibit substantially different feature ICCs owing to their intrinsic properties, heterogeneities, and thus different image representations. This finding indicates that different features might be chosen for the construction of tissue-specific radiomics models for clinical use. Next, multiple MRI scans provided additional information on feature repeatability compared to the simple test-retest repeated scans. Fewer scans were found to have a wider 95%CI and thus a larger uncertainty of the calculated ICC values, which might lead to overestimation of feature repeatability if based on the ICC value alone. However, by referring to the 95%CI of ICC for feature repeatability classification, the highly repeatable features for MRI acquisition could be more

accurately identified via multi-scan than by dual-scan. In addition, multi-scan enabled the calculation of feature  $CV_{\text{intra-subject}}$  to assess within-subject feature value variability, whereas dual-scan could only evaluate the feature value difference between two scans. This within-subject feature variability is of particular importance for individualized radiomics analysis in MRgRT studies, in which longitudinal MRI datasets obtained from multiple treatment fractions of each subject are used. It is crucial for MRgRT to accurately differentiate the true variations in the value of radiomics features due to response to irradiation from variation or uncertainty due to image acquisition. Otherwise, radiomics analysis could result in false positives discovery. The  $CV_{\text{intra-subject}}$  results obtained in this study revealed that many features were subjected to pronounced value variations in multiple MRI scans. The  $CV_{\text{intra-subject}}$ s were much larger than the previously reported values in a longitudinal phantom study using the same model MRI scanner [48]. This is not surprising because in vivo tissues exhibit much higher heterogeneity and are subjected to much more intra- and inter-scan tissue property change, positional variation, deformation, and motion. Thus, it is vital to carefully select repeatable acquisition features from each patient for reliable delta-radiomics analysis. However, the  $CV_{\text{intra-subject}}$  was still smaller than the  $CV_{\text{inter-subject}}$ , which indicated that although acquisition-induced feature variability can impact inter-subject radiomics analysis for diagnosis purpose, its impact might not be as great as in within-subject longitudinal radiomics analysis.

This study involves several strengths and limitations. In addition to its prospective nature, this is the first study on MRI radiomics feature acquisition repeatability study in the HN dedicated to MRgRT applications, as noted above. Dedicated 3D pulse sequences for MRgRT were used, and multiple MRI scans were conducted under MRgRT treatment positioning (e.g., 3D laser alignment, flat couch, mask immobilization, etc.) and setting (RF coil selection and coil setting with dedicated holders and bridges). The use of mask immobilization maximized tissue coherence in the VOIs, alleviating the influence of image registration on feature quantification [26]. The multiple-scan study design not only reflected the pattern of MRI use in clinical MRgRT practice but also extended the capability to calculate within-subject feature CV and increased the ICC calculation confidence for more reliable feature selection. The adoption of IBSI-compliant features increased the transparency of the feature calculation. Confounding factors other than image acquisition were excluded as much as possible in the study design and data analysis to minimize their influence on feature values. As such, the results may be considered to faithfully reveal the feature repeatability and variability purely with respect to the acquisition. Paired tissue analysis for calibration was helpful in ensuring the validity and rigor of the quantification results.

However, this study does involve some limitations. This pilot study was mainly limited to the inclusion of only healthy volunteers and a small sample size. Malignant HN tumors might exhibit substantially different between-subject radiomics feature heterogeneities as well as their within-subject measurement uncertainties compared to normal HN tissues, but they could not be revealed because the study was conducted on healthy volunteers. The logistical difficulty and ethical concerns involved in conducting such a multi-scan study on real HNC patients should be recognized. However, despite the absence of malignant tumors in healthy volunteers, it is expected that this pilot study may still be useful for future clinical applications because radiomics can also be used in a variety of OARs, which are thought to be normal tissues but are inevitably irradiated, to assess their toxicities or the quality-of-life of HNC patients treated with MRgRT. The small sample size limited the statistical power of the tests. Although this study was designed with a focus on MRgRT applications, all MRI scans were acquired using an MRI simulator instead of an actual MR-LINAC. Considering the difference in the configurations of MRI hardware and implementations of 3D pulse sequence on a 1.5 T MR-LINAC from those of a 1.5 T MRI simulator from a different vendor, radiomics features and their repeatability/variability

characteristics obtained on a 1.5 T MR-LINAC might be considerably different from those obtained in this study. Meanwhile, the approximately 1 mm isotropic spatial resolution obtained in this study was higher than that normally used for daily MRI acquisition in MRgRT. Although it is helpful for better tissue delineation and registration, this high resolution might also influence feature repeatability and variability. This study specifically addressed an important source of MRI radiomics feature unreliability due to multiple acquisitions, but did not assess the influence of many other confounding factors on feature values, such as image reconstruction, segmentation, and other image post-processing methods [49]. Although image acquisition has been found to be more impactful on feature reliability than other sources in MRI radiomics [24], it is equally important to investigate the impact of other factors on the reliability of radiomics features. Meanwhile, only a small subset of first-order and texture original radiomics features was included for analysis in this study among the thousands of radiomics features in the original and transformed images proposed for radiomics modeling in the medical literature. In particular, shape features indicating the geometric characteristics of various tissues (such as size or volume) are conventionally used as imaging markers in cancer staging [50] and treatment response evaluation [51], but were not included in the present work. Although features in the transformed domains might provide more candidates for radiomics modeling and different information on tissue properties, it has been found that image transformation does not necessarily improve radiomics features reliability [24]. Therefore, the absence of transformed features in this study may not seriously compromise the validity and interpretation of the results.

## Conclusion

This prospective study has rigorously investigated the multi-scan acquisition repeatability and variability of MRI radiomics features for 3D-T1W-TSE and 3D-T2W-TSE sequences in normal HN tissues of a cohort of fifteen healthy volunteers using a 1.5 T MRI simulator with MRgRT treatment positioning and coil setting, focusing on providing a benchmark for future MRgRT radiomics studies in HNC. Radiomics features repeatable to each sequence were identified, as measured by ICC. Within-subject variability in multiple MRI acquisitions was quantified. These results are expected to advance the understanding of the reliability of radiomics features with respect to MRI acquisition and the reliable radiomics features selection for modeling in HNC MRgRT applications.

## Abbreviations

HNC: Head and neck cancer; HN: Head and neck; RT: Radiation therapy; MRI: Magnetic resonance imaging; MR-LINAC: MRI-integrated linear accelerator; MRgRT: Magnetic resonance guided radiotherapy; OARs: Organs at risk; T2W: T2-weighted; TSE: Turbo spin-echo; VOI: Volume of interest; LR: Left/right; SI: Superior/inferior; AP: Anterior/posterior; GRAPPA: Generalized autocalibrating partial parallel acquisition; IBSI: Image Biomarker Standardization Initiative; GLCM: Gray-level co-occurrence matrix; GLDM: Gray-level dependence matrix; GLRLM: Gray-level run length matrix; GLSZM: Gray-level size zone matrix; NGTDM: Neighboring gray-tone difference matrix; ICC: Intraclass correlation coefficient; CV: Coefficient of variation; SD: Standard deviation; ANOVA: Analysis of variance.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42492-022-00106-3>.

### Additional file 1.

## Acknowledgements

Not applicable.

## Authors' contributions

JY provided the conception and design; KYC and SKY provided administrative support; JY, OLV and YZ provided the study materials or patients; CX, OLV and YZ made the collection and assembly of data; CX and JY contributed to data analysis and interpretation; all authors contributed to manuscript writing and editing; the authors read and approved the final manuscript.

## Funding

This study was supported by hospital research project, No. REC-2019-09.

## Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available because the subjects did not provide written consent for their data to be publicly shared.

## Declarations

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Research Department, Hong Kong Sanatorium & Hospital, Hong Kong, China.

<sup>2</sup>Medical Physics Department, Hong Kong Sanatorium & Hospital, Hong Kong, China.

Received: 18 November 2021 Accepted: 23 February 2022

Published online: 01 April 2022

## References

- Chow LQM (2020) Head and neck cancer. *N Engl J Med* 382(1):60–72. <https://doi.org/10.1056/NEJMra1715715>
- Simard EP, Torre LA, Jemal A (2014) International trends in head and neck cancer incidence rates: differences by country, sex and anatomic site. *Oral Oncol* 50(5):387–403. <https://doi.org/10.1016/j.oraloncology.2014.01.016>
- Gatta G, Botta L, Sánchez MJ, Anderson LA, Pierannunzio D, Licitra L (2015) Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: the EURO CARE-5 population-based study. *Eur J Cancer* 51(15):2130–2143. <https://doi.org/10.1016/j.ejca.2015.07.043>
- Dai YL, King AD (2018) State of the art MRI in head and neck cancer. *Clin Radiol* 73(1):45–59. <https://doi.org/10.1016/j.crad.2017.05.020>
- Law BKH, King AD, Ai QY, Poon DMC, Chen WT, Bhatia KS et al (2018) Head and neck tumors: amide proton transfer MRI. *Radiology* 288(3):782–790. <https://doi.org/10.1148/radiol.2018171528>
- Yuan J, Lo G, King AD (2016) Functional magnetic resonance imaging techniques and their development for radiation therapy planning and monitoring in the head and neck cancers. *Quant Imaging Med Surg* 6(4):430–448. <https://doi.org/10.21037/qims.2016.06.11>
- Law BKH, King AD, Bhatia KS, Ahuja AT, Kam MKM, Ma BB et al (2016) Diffusion-weighted imaging of nasopharyngeal carcinoma: can pretreatment DWI predict local failure based on long-term outcome? *AJNR Am J Neuroradiol* 37(9):1706–1712. <https://doi.org/10.3174/ajnr.A4792>
- Hunt A, Hansen VN, Oelfke U, Nill S, Hafeez S (2018) Adaptive radiotherapy enabled by MRI guidance. *Clin Oncol* 30(11):711–719. <https://doi.org/10.1016/j.clon.2018.08.001>
- Legendijk JJW, Raaymakers BW, Van den Berg CAT, Moerland MA, Philipens ME, van Vulpen M (2014) MR guidance in radiotherapy. *Phys Med Biol* 59(21):R349–R369. <https://doi.org/10.1088/0031-9155/59/21/R349>
- Raaymakers BW, Legendijk JJW, Overweg J, Kok JGM, Raaijmakers AJE, Kerkhof EM et al (2009) Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept. *Phys Med Biol* 54(12):N229–N237. <https://doi.org/10.1088/0031-9155/54/12/N01>
- Chen AM, Hsu S, Lamb J, Yang Y, Agazaryan N, Steinberg ML et al (2018) MRI-guided radiotherapy for head and neck cancer: initial clinical experience. *Clin Transl Oncol* 20(2):160–168. <https://doi.org/10.1007/s12094-017-1704-4>
- Boeke S, Mönnich D, van Timmeren JE, Balermipas P (2021) MR-guided radiotherapy for head and neck cancer: current developments, perspectives, and challenges. *Front Oncol* 11:616156. <https://doi.org/10.3389/fonc.2021.616156>
- Henke LE, Contreras JA, Green OL, Cai B, Kim H, Roach MC et al (2018) Magnetic resonance image-guided radiotherapy (MRIGRT): a 4.5-year clinical experience. *Clin Oncol* 30(11):720–727. <https://doi.org/10.1016/j.clon.2018.08.010>
- Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. *Radiology* 278(2):563–577. <https://doi.org/10.1148/radiol.2015151169>
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RGPM, Granton P et al (2012) Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48(4):441–446. <https://doi.org/10.1016/j.ejca.2011.11.036>
- Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14(12):749–762. <https://doi.org/10.1038/nrclinonc.2017.141>
- Tanadini-Lang S, Balermipas P, Guckenberger M, Pavic M, Riesterer O, Vuong D et al (2020) Radiomic biomarkers for head and neck squamous cell carcinoma. *Strahlenther Onkol* 196(10):868–878. <https://doi.org/10.1007/s00066-020-01638-4>
- Ou D, Blanchard P, Rosellini S, Levy A, Nguyen F, Leijenaar RTH et al (2017) Predictive and prognostic value of CT based radiomics signature in locally advanced head and neck cancers patients treated with concurrent chemoradiotherapy or bioradiotherapy and its added value to human papillomavirus status. *Oral Oncol* 71:150–155. <https://doi.org/10.1016/j.oraloncology.2017.06.015>
- Yuan Y, Ren JL, Shi YQ, Tao XF (2019) MRI-based radiomic signature as predictive marker for patients with head and neck squamous cell carcinoma. *Eur J Radiol* 117:193–198. <https://doi.org/10.1016/j.ejrad.2019.06.019>
- Mes SW, van Velden FHP, Peltenburg B, Peeters CFW, Te Beest DE, Van De Wiel MA et al (2020) Outcome prediction of head and neck squamous cell carcinoma by MRI radiomic signatures. *Eur Radiol* 30(11):6311–6321. <https://doi.org/10.1007/s00330-020-06962-y>
- Jethanandani A, Lin TA, Volpe S, Elhalawani H, Mohamed ASR, Yang P et al (2018) Exploring applications of radiomics in magnetic resonance imaging of head and neck cancer: a systematic review. *Front Oncol* 8:131. <https://doi.org/10.3389/fonc.2018.00131>
- Traverso A, Wee L, Dekker A, Gillies R (2018) Repeatability and reproducibility of radiomics features: a systematic review. *Int J Radiat Oncol Biol Phys* 102(4):1143–1158. <https://doi.org/10.1016/j.ijrobp.2018.05.053>
- Yip SSF, Aerts HJWL (2016) Applications and limitations of radiomics. *Phys Med Biol* 61(13):R150–R166. <https://doi.org/10.1088/0031-9155/61/13/R150>
- Xue C, Yuan J, Lo GG, Chang ATY, Poon DMC, Wong OL et al (2021) Radiomics feature reliability assessed by intraclass correlation coefficient:

- a systematic review. *Quant Imaging Med Surg* 11(10):4431–4460. <https://doi.org/10.21037/qims-21-86>
25. Fiset S, Welch ML, Weiss J, Pintilie M, Conway JL, Milosevic M et al (2019) Repeatability and reproducibility of MRI-based radiomic features in cervical cancer. *Radiother Oncol* 135:107–114. <https://doi.org/10.1016/j.radonc.2019.03.001>
  26. Shiri I, Hajianfar G, Sohrabi A, Abdollahi H, Shayesteh SP, Geramifar P et al (2020) Repeatability of radiomic features in magnetic resonance imaging of glioblastoma: test-retest and image registration analyses. *Med Phys* 47(9):4265–4280. <https://doi.org/10.1002/mp.14368>
  27. Jang J, Ngo LH, Mancio J, Kucukseymen S, Rodriguez J, Pierce P et al (2020) Reproducibility of segmentation-based myocardial radiomics features with cardiac MRI. *Radiol Cardiothorac Imaging* 2(3):e190216. <https://doi.org/10.1148/rct.2020190216>
  28. Dreher C, Kuder TA, König F, Mlynarska-Bujny A, Tenconi C, Paech D et al (2020) Radiomics in diffusion data: a test-retest, inter- and intra-reader DWI phantom study. *Clin Radiol* 75(10):798.e713–798.e722. <https://doi.org/10.1016/j.crad.2020.06.024>
  29. Yuan J, Xue C, Lo G, Wong OL, Zhou YH, Yu SK et al (2021) Quantitative assessment of acquisition imaging parameters on MRI radiomics features: a prospective anthropomorphic phantom study using a 3D-T2W-TSE sequence for MR-guided-radiotherapy. *Quant Imaging Med Surg* 11(5):1870–1887. <https://doi.org/10.21037/qims-20-865>
  30. Xue C, Yuan J, Poon DM, Zhou YH, Yang B, Yu SK et al (2021) Reliability of MRI radiomics features in MR-guided radiotherapy for prostate cancer: repeatability, reproducibility, and within-subject agreement. *Med Phys* 48(11):6976–6986. <https://doi.org/10.1002/mp.15232>
  31. Boldrini L, Cusumano D, Chiloiro G, Casà C, Masciocchi C, Lenkowicz J et al (2019) Delta radiomics for rectal cancer response prediction with hybrid 0.35 T magnetic resonance-guided radiotherapy (MRgRT): a hypothesis-generating study for an innovative personalized medicine approach. *Radiol Med* 124(2):145–153. <https://doi.org/10.1007/s11547-018-0951-y>
  32. Carvalho S, Leijenaar RTH, Troost EGC, van Elmpt W, Muratet JP, Denis F et al (2016) Early variation of FDG-PET radiomics features in NSCLC is related to overall survival-the "delta radiomics" concept. *Radiat Oncol* 118:S20–S21. [https://doi.org/10.1016/S0167-8140\(16\)30042-1](https://doi.org/10.1016/S0167-8140(16)30042-1)
  33. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S et al (2012) 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging* 30(9):1323–1341. <https://doi.org/10.1016/j.mri.2012.05.001>
  34. Zhou YH, Wong OL, Cheung KY, Yu SK, Yuan J (2019) A pilot study of highly accelerated 3D MRI in the head and neck position verification for MR-guided radiotherapy. *Quant Imaging Med Surg* 9(7):1255–1269. <https://doi.org/10.21037/qims.2019.06.18>
  35. Zhou YH, Yuan J, Wong OL, Fung WWK, Cheng KF, Cheung KY et al (2018) Assessment of positional reproducibility in the head and neck on a 1.5-T MR simulator for an offline MR-guided radiotherapy solution. *Quant Imaging Med Surg* 8(9):925–935. <https://doi.org/10.21037/qims.2018.10.03>
  36. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V et al (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77(21):e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
  37. Zwanenburg A, Leger S, Vallières M, Löck S (2019) Image biomarker standardisation initiative. *arXiv preprint arXiv:1612.07003*. <https://doi.org/10.48550/arXiv.1612.07003>
  38. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A et al (2020) The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 295(2):328–338. <https://doi.org/10.1148/radiol.2020191145>
  39. Scalco E, Belfatto A, Mastropietro A, Rancati T, Avuzzi B, Messina A et al (2020) T2w-MRI signal normalization affects radiomics features reproducibility. *Med Phys* 47:1680–1691. <https://doi.org/10.1002/mp.14038>
  40. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15(2):155–163. <https://doi.org/10.1016/j.jcmm.2016.02.012>
  41. Merisaari H, Taimen P, Shiradkar R, Ettala O, Pesola M, Saunavaara J et al (2020) Repeatability of radiomics and machine learning for DWI: short-term repeatability study of 112 patients with prostate cancer. *Magn Reson Med* 83(6):2293–2309. <https://doi.org/10.1002/mrm.28058>
  42. Pandey U, Saini J, Kumar M, Gupta R, Ingahlalkar M (2020) Normative baseline for radiomics in brain MRI: evaluating the robustness, regional variations, and reproducibility on FLAIR images. *J Magn Reson Imaging* 53(2):394–407. <https://doi.org/10.1002/jmri.27349>
  43. Ta D, Khan M, Ishaque A, Seres P, Eurich D, Yang YH et al (2020) Reliability of 3D texture analysis: a multicenter MRI study of the brain. *J Magn Reson Imaging* 51(4):1200–1209. <https://doi.org/10.1002/jmri.26904>
  44. Hu PP, Wang JZ, Zhong HY, Zhou Z, Shen LJ, Hu WG et al (2016) Reproducibility with repeat CT in radiomics study for rectal cancer. *Oncotarget* 7(44):71440–71446. <https://doi.org/10.18632/oncotarget.12199>
  45. Li ZR, Duan HC, Zhao K, Ding YH (2019) Stability of MRI radiomics features of hippocampus: an integrated analysis of test-retest and inter-observer variability. *IEEE Access* 7:97106–97116. <https://doi.org/10.1109/ACCESS.2019.2923755>
  46. Rai R, Holloway LC, Brink C, Field M, Christiansen RL, Sun Y et al (2020) Multicenter evaluation of MRI-based radiomic features: a phantom study. *Med Phys* 47(7):3054–3063. <https://doi.org/10.1002/mp.14173>
  47. Vuong D, Tanadini-Lang S, Huellner MW, Veit-Haibach P, Unkelbach J, Andratschke N et al (2019) Interchangeability of radiomic features between [18F]-FDG PET/CT and [18F]-FDG PET/MR. *Med Phys* 46(4):1677–1685. <https://doi.org/10.1002/mp.13422>
  48. Wong OL, Yuan J, Zhou YH, Yu SK, Cheung KY (2021) Longitudinal acquisition repeatability of MRI radiomics features: an ACR MRI phantom study on two MRI scanners using a 3D T1W TSE sequence. *Med Phys* 48(3):1239–1249. <https://doi.org/10.1002/mp.14686>
  49. Zhao BS (2021) Understanding sources of variation to improve the reproducibility of radiomics. *Front Oncol* 11:633176. <https://doi.org/10.3389/fonc.2021.633176>
  50. Huang SH, O'Sullivan B (2017) Overview of the 8th edition TNM classification for head and neck cancer. *Curr Treat Options in Oncol* 18(7):40. <https://doi.org/10.1007/s11864-017-0484-y>
  51. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 45(2):228–247. <https://doi.org/10.1016/j.ejca.2008.10.026>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)