

OPEN ACCESS

EDITED BY

Ming Li,
Fudan University, China

REVIEWED BY

Marco Curti,
Imaging Institute of Italian Switzerland
(IIMSI), Switzerland
Alessandro Stefano,
Department of Biomedical Sciences,
National Research Council (CNR), Italy

*CORRESPONDENCE

Luca Maria Sconfienza
io@lucascconfienza.it

SPECIALTY SECTION

This article was submitted to
Cancer Imaging and
Image-directed Interventions,
a section of the journal
Frontiers in Oncology

RECEIVED 10 August 2022

ACCEPTED 17 November 2022

PUBLISHED 02 December 2022

CITATION

Gitto S, Corino VDA, Annovazzi A,
Milazzo Machado E, Bologna M,
Marzorati L, Albano D, Messina C,
Serpi F, Anelli V, Ferraresi V, Zoccali C,
Aliprandi A, Parafioriti A, Luzzati A,
Biagini R, Mainardi L and
Sconfienza LM (2022) 3D vs. 2D MRI
radiomics in skeletal Ewing sarcoma:
Feature reproducibility and preliminary
machine learning analysis on
neoadjuvant chemotherapy response
prediction.
Front. Oncol. 12:1016123
doi: 10.3389/fonc.2022.1016123

COPYRIGHT

© 2022 Gitto, Corino, Annovazzi,
Milazzo Machado, Bologna, Marzorati,
Albano, Messina, Serpi, Anelli, Ferraresi,
Zoccali, Aliprandi, Parafioriti, Luzzati,
Biagini, Mainardi and Sconfienza. This is
an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use,
distribution or reproduction is
permitted which does not comply with
these terms.

3D vs. 2D MRI radiomics in skeletal Ewing sarcoma: Feature reproducibility and preliminary machine learning analysis on neoadjuvant chemotherapy response prediction

Salvatore Gitto¹, Valentina D. A. Corino^{2,3}, Alessio Annovazzi⁴,
Estevão Milazzo Machado⁵, Marco Bologna²,
Lorenzo Marzorati², Domenico Albano⁶, Carmelo Messina^{1,6},
Francesca Serpi¹, Vincenzo Anelli⁷, Virginia Ferraresi⁸,
Carmine Zoccali^{9,10}, Alberto Aliprandi¹¹, Antonina Parafioriti¹²,
Alessandro Luzzati⁶, Roberto Biagini¹⁰, Luca Mainardi²
and Luca Maria Sconfienza^{1,6*}

¹Dipartimento di Scienze Biomediche per la Salute, Università degli Studi di Milano, Milan, Italy,

²Department of Electronics, Information and Bioengineering (DEIB), Politecnico Di Milano, Milan, Italy, ³Cardiotech Lab, Centro Cardiologico Monzino IRCCS, Milan, Italy, ⁴Nuclear Medicine Unit, IRCCS Regina Elena National Cancer Institute, Rome, Italy, ⁵International Medical School, Università degli Studi di Milano, Milan, Italy, ⁶IRCCS Istituto Ortopedico Galeazzi, Milan, Italy, ⁷Radiology and Diagnostic Imaging Unit, IRCCS Regina Elena National Cancer Institute, Rome, Italy, ⁸Sarcomas and Rare Tumours Departmental Unit, IRCCS Regina Elena National Cancer Institute, Rome, Italy,

⁹Department of Anatomical, Histological, Forensic and Musculoskeletal System Sciences, Sapienza University of Rome, Rome, Italy, ¹⁰Oncological Orthopaedics Unit, IRCCS Regina Elena National Cancer Institute, Rome, Italy, ¹¹Istituti Clinici Zucchi, Monza, Italy, ¹²Pathology Department, ASST Pini - CTO, Milan, Italy

Objective: The extent of response to neoadjuvant chemotherapy predicts survival in Ewing sarcoma. This study focuses on MRI radiomics of skeletal Ewing sarcoma and aims to investigate feature reproducibility and machine learning prediction of response to neoadjuvant chemotherapy.

Materials and methods: This retrospective study included thirty patients with biopsy-proven skeletal Ewing sarcoma, who were treated with neoadjuvant chemotherapy before surgery at two tertiary sarcoma centres. 7 patients were poor responders and 23 were good responders based on pathological assessment of the surgical specimen. On pre-treatment T1-weighted and T2-weighted MRI, 2D and 3D tumour segmentations were manually performed. Features were extracted from original and wavelet-transformed images. Feature reproducibility was assessed through small geometrical transformations of the regions of interest mimicking multiple manual delineations, and intraclass correlation coefficient >0.75 defined feature

reproducibility. Feature selection also consisted of collinearity and significance analysis. After class balancing in the training cohort, three machine learning classifiers were trained and tested on unseen data using hold-out cross-validation.

Results: 1303 (77%) 3D and 620 (65%) 2D radiomic features were reproducible. 4 3D and 4 2D features passed feature selection. Logistic regression built upon 3D features achieved the best performance with 85% accuracy (AUC=0.9) in predicting response to neoadjuvant chemotherapy.

Conclusion: Compared to 2D approach, 3D MRI radiomics of Ewing sarcoma had superior reproducibility and higher accuracy in predicting response to neoadjuvant chemotherapy, particularly when using logistic regression classifier.

KEYWORDS

artificial intelligence, Ewing sarcoma, machine learning, magnetic resonance imaging, radiomics, texture analysis

Introduction

Ewing sarcoma is the second most common bone sarcoma in children and adolescents after osteosarcoma (1). The standard of treatment in non-metastatic patients is represented by neoadjuvant chemotherapy, surgical resection and adjuvant chemotherapy (1). Radiotherapy can be administered to achieve local control post-operatively if surgical margins are inadequate or alone when complete resection is not feasible, particularly in certain anatomic locations (1). Five-year survival is 60-75% in localized disease, but it decreases substantially to < 20% in case of local or distant relapse (2). The extent of response to neoadjuvant chemotherapy predicts survival (3). Unfortunately, this can be assessed only after surgery based on histopathology (1). However, it would be desirable to know how the patient responded before surgery to tailor subsequent treatment accordingly. Furthermore, pathological response data are unavailable in non-operated patients treated with radiotherapy.

Radiomics includes extraction and analysis of quantitative parameters from medical images, known as radiomic features (4). Several features are routinely extracted in radiomic studies, but many of them are redundant or not informative for the clinical question of interest (5). Thus, a selection of radiomic features is necessary and feature reproducibility or stability assessment is an essential step of this process (5). However, a

recent systematic review dealing with radiomics of musculoskeletal sarcomas highlighted that more than half of the papers lacked a preliminary analysis of feature reproducibility (6). In studies where this analysis was performed, intraclass correlation coefficient (ICC) was the most employed method to assess reproducibility (6). After selection, the optimal subset of radiomic features can be combined with machine learning algorithms for prediction purposes, including therapy response or outcome prediction (7, 8).

The aim of this study is twofold: (i) to assess the reproducibility of 3D and 2D radiomic features of skeletal Ewing sarcoma extracted from pre-treatment magnetic resonance imaging (MRI); and (ii) to preliminarily investigate 3D vs. 2D radiomics-based machine learning performance for prediction of response to neoadjuvant chemotherapy.

Materials and methods

Ethics

Institutional Review Board approved this retrospective study and waived the need for informed consent. All included patients granted written permission for anonymized data use for research purposes at the time of the MRI. After matching imaging, pathological and surgical data, our database was anonymized to delete any connections between data and patients' identity according to the General Data Protection Regulation for Research Hospitals.

Abbreviations: AUC, area under the curve; ICC, intraclass correlation coefficient; k-NN, k-nearest neighbours; LR, logistic regression; MRI, magnetic resonance imaging; RF, random forest; ROI, region of interest.

Dataset description

This study was designed to meet the numerical requirements of reliability analyses in terms of patients involved, namely a minimum of 30, according to the ICC guidelines by Koo et al. (9). Thirty patients were retrospectively included at two tertiary bone sarcoma centres (centre 1, IRCCS Orthopaedic Institute Galeazzi, Milan, Italy; centre 2, IRCCS Regina Elena National Cancer Institute, Rome, Italy). Information was retrieved through medical records from the orthopaedic and pathology departments. Inclusion criteria were: (i) biopsy-proven Ewing sarcoma of the bone treated with neoadjuvant chemotherapy before surgery; (ii) response to neoadjuvant chemotherapy evaluated after surgery based on histopathology; (iii) 1.5- or 3-T MRI performed within two months before chemotherapy including turbo spin echo T1-weighted and T2-weighted sequences. Externally obtained MRI scans of patients referred to centres 1 and 2 were also included in this study as long as the minimal MRI protocol was available (MRI sequence parameters are detailed in [Supplementary Material](#)). Exclusion criteria were: (i) recurrent tumours; (ii) metastatic tumours; and (iii) image noise (motion or metal artifacts) affecting tumour segmentation. Neoadjuvant chemotherapy regimen consisted of 3-6 cycles including five to six drug combinations, namely doxorubicin, cyclophosphamide, ifosfamide, vincristine, dactinomycin and etoposide, as per current guidelines (1). Histopathologic response to neoadjuvant chemotherapy was assessed using the grading system by Picci et al. (also known as the “Bologna system”), where grade I (macroscopic residual tumour) defined poor response and grade II or III (microscopic or no residual disease, respectively) defined good response (10, 11). In our population of study, 7 patients were poor responders (all included at centre 1) and 23 were good responders (n=10 included at centre 1 and n=13 included at centre 2). Clinical data and demographics of the study population are detailed in [Table 1](#).

TABLE 1 Clinical data and demographics of the study population.

Clinical data and demographics

Age [median (1 st -3 rd quartiles)]	18 (12-27) years
Sex	14 men; 16 women
Tumour location	Lower extremity: n=13 Pelvis: n=7 Spine: n=6 Upper extremity: n=4
Histopathologic grading (10)	Grade I (poor response): n=7 Grade II (good response): n=15 Grade III (good response): n=8

Years of age are approximated to the nearest whole number.

Segmentation

All DICOM images were extracted and imported to the freely available, open-source software ITK-SNAP (v3.8) (12) for segmentation. On T1-weighted and T2-weighted MRI, manual segmentations were obtained by drawing both 2D regions of interest (ROIs) along tumour borders on the slice showing the largest tumour area and 3D ROIs including the whole tumour volume. The “polygon mode” ITK-SNAP tool was used for all segmentations ([Figure 1](#)). A musculoskeletal radiologist performed all segmentations. The radiologist knew the study would deal with Ewing sarcoma but was blinded to any information regarding disease course.

Pre-processing and feature extraction

Image pre-processing and radiomic feature extraction were performed using PyRadiomics 3.0 (13). Five pre-processing steps were employed, including (i) grey-level normalization, (ii) pixel resampling and rescaling using cubic B-spline interpolation (resultant pixel size, 2x2x2 mm³), (iii) grey-level discretization (bin count, 32), (iv) denoising through the implementation of a 3D Gaussian filter, and (v) inhomogeneity correction using the open-source software 3D slicer (14). The default parameters of PyRadiomics were used, except for discretization. Particularly, intensity discretization with a fixed bin number was used instead of the fixed bin size discretization (default in PyRadiomics). The features extracted from each original image were grouped as follows (<https://pyradiomics.readthedocs.io/en/latest/features.html>): first-order features; shape-based features; grey-level cooccurrence matrix features; grey-level run length matrix features; grey-level size zone matrix features; grey-level dependence matrix features; neighbouring grey tone difference matrix features. First-order and textural features were extracted from both the original volumes and the 8 volumes obtained by first level wavelet decomposition (or 4 wavelet decompositions for 2D images) (15). For the full list of radiomic features, refer to PyRadiomics official documentation.

Reproducibility analysis and feature selection

Reproducibility analysis was performed to evaluate the robustness of radiomic features. Feature reproducibility was assessed through small geometrical transformations of the ROIs mimicking multiple manual delineations and the potential sources of intra- and inter-observer variability when multiple ROIs are drawn by the same or different readers, respectively (16, 17). Different translations of the ROI in the positive and negative direction of the x and y axes were applied.

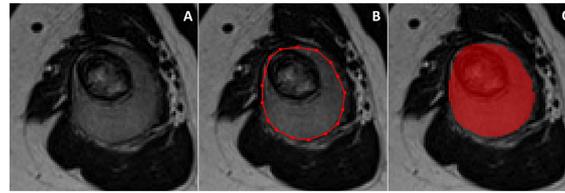


FIGURE 1

On the slice showing the largest tumor area (A), 2D manual segmentation was obtained by drawing a polygonal region of interest (B) to include the whole tumor area (C). The same procedure was repeated slice by slice to obtain 3D segmentation including the whole tumor volume.

The entity of the translation was 10% of the length of the bounding box including the tumour (Figure 2). For each patient, radiomic features were extracted from five different ROIs, namely the original and 4 translated ROIs. ICC was used to quantify reproducibility, and radiomic features were considered stable if $ICC > 0.75$ (9). Among stable features, further dimensionality reduction was performed through collinearity and significance analysis. Collinearity was evaluated using Pearson correlation coefficient (r). The threshold for collinearity was set as $r = 0.8$. If a pair of features had high collinearity, the one with higher collinearity with others was excluded. Significance analysis addressed the ability of radiomic features to discriminate between good and poor responders. Wilcoxon rank-sum test was used. Radiomic features with statistically different distribution between these two groups (p -value < 0.05) were kept and ranked by their p -value.

classifiers, namely k -nearest neighbours (k -NN, $k = 1$ linear nearest neighbour search with Euclidean distance function and no distance weighting), logistic regression (LR) and random forest (RF). To evaluate the unbiased performance of the classifiers, a cross-validation approach was used with a hold-out partition of 80-20 (80% for training and 20% for test sets, repeated 100 times). The same hold-out partition was used for 3D and 2D radiomics-based analysis. An imbalance between groups existed in our data (23 good vs. 7 poor responders), which could adversely affect classification performance. Therefore, prior to the training of the classifiers, class balancing was performed in the training cohort using the synthetic minority oversampling technique (18). This technique is used to artificially oversample the minority class, namely the poor responders in our study. After tuning on the training cohort, the classifiers were tested on unseen data, namely the test set.

Machine learning-based classification

Based on the optimal subsets of radiomic features, classification models were built using three machine learning

Statistical analysis

Continuous data are presented as median and interquartile (1^{st} - 3^{rd}) range. Categorical data are presented as value counts

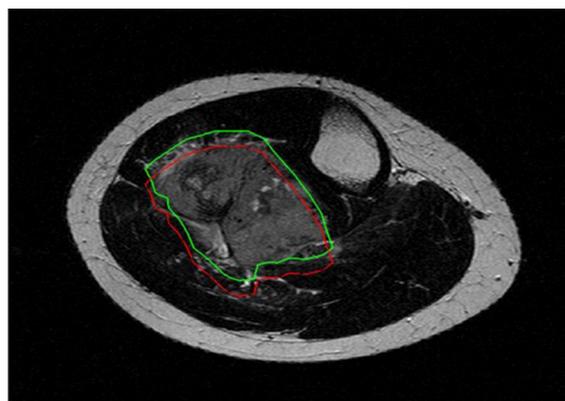


FIGURE 2

Original (green) and translated (red) version of 2D region of interest drawn around tumor borders on T2-weighted axial image.

and proportions. Differences in age and sex between good and poor responders were evaluated using Mann-Whitney U and chi-square tests, respectively. The reproducibility of 2D and 3D features was compared using t-test. The performances of the classifiers were evaluated and compared using area under the curve (AUC), accuracy, sensitivity, specificity, precision and balanced accuracy (i.e., the average between sensitivity and specificity) averaged on the test sets. Since LR and RF return continuous values, a binarization was necessary (the threshold for binarization was selected on the training set and then used on the test set). A two-tailed p-value < 0.05 was considered statistically significant. The statistical analysis was performed using MATLAB 2018b (Mathworks, Natick, MA, USA).

Results

No statistical difference in age ($p = 0.841$; 19 [12–29] vs. 18 [13–23] years) and sex ($p = 0.526$; 10 men and 13 women vs. 4 men and 3 women) was found between good and poor responders, respectively. A total of 1702 3D and 958 2D features were extracted per each patient. In detail, they were 851 3D and 479 2D features obtained from each MRI sequence. Among them, 1303 (77%) 3D and 620 (65%) 2D radiomic features were stable to geometrical transformations of the ROI (for 3D: 92% of the first order, 39% textural, 76% wavelet; for 2D: 89% of the first order, 33% textural, 61% wavelet), as shown in Figure 3.

After excluding features with high collinearity, the numbers of 3D and 2D features were further reduced to 101 and 58, respectively. Significance analysis yielded to four 3D and four 2D features, which were finally selected. In particular, the selected 3D features were all computed on the wavelet-transformed images and the selected 2D features were all obtained from the original images, respectively, as detailed in Table 2.

The evaluation metrics for all classifiers are reported in Table 3. LR and RF classifiers achieved better accuracies when using 3D features compared to 2D features. Conversely, k-NN classifier showed better accuracy when built upon 2D features.

As shown in column charts in Figure 4, only 7 patients were correctly classified with all 2D radiomics-based classifiers when using a percentage threshold of 80%. An improvement was obtained with 3D radiomics, as 14 patients were correctly classified at least 80% of times with all classifiers. LR built upon 3D features achieved the best performance with 85% sensitivity, 87% specificity and 85% accuracy (AUC=0.9) in predicting response to neoadjuvant chemotherapy, with 24 patients correctly classified at least 80% of the times.

Discussion

This retrospective study compared the reproducibility of MRI radiomic features obtained from 2D and 3D segmentations and investigated radiomics-based machine learning prediction of response to neoadjuvant chemotherapy in Ewing sarcoma of the bone. 3D segmentations yielded higher rates of reproducible features compared to 2D approach. Additionally, LR built upon 3D features was the most accurate classifier in differentiating good from poor responders and achieved 85% accuracy (AUC = 0.9).

Despite its great potential to non-invasively quantify several tumour characteristics, radiomics still faces challenges preventing clinical transferability (8). A great variability in radiomic features has emerged as a major issue across studies, and segmentation is the most critical step (4). Therefore, in radiomic studies, preliminary methodological analyses are advisable to assess the robustness of different segmentation approaches and avoid biases due to non-reproducible, noisy features. This is in line with recent literature emphasizing the importance of reproducibility in artificial intelligence and radiology (19). Several strategies can be used to assess feature reproducibility, such as changes in image acquisition parameters (20) and multiple ROI delineations performed by different readers (21, 22), which are however time-consuming. In our study, feature reproducibility was evaluated through a time-saving method based on geometrical transformations of the ROIs mimicking multiple manual delineations (16, 17). In 3D

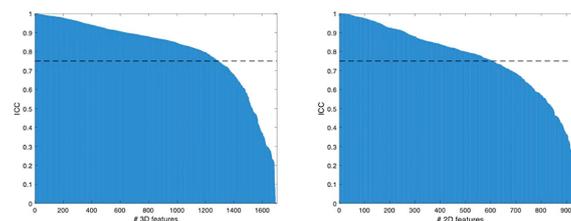


FIGURE 3
Radiomic features ranked by their ICC values.

TABLE 2 List of selected features by feature class, ROI, MRI sequence and source image.

Selected feature	Feature class	ROI	Sequence	Source
Correlation	GLCM	3D	T1w	Wavelet
Minimum	First order	3D	T2w	Wavelet
Small Dependence High Gray Level Emphasis	GLDM	3D	T2w	Wavelet
Cluster Shade	GLCM	3D	T2w	Wavelet
Minimum	First order	2D	T1w	Original
Strength	NGTDM	2D	T1w	Original
Sphericity	Shape	2D	T2w	Original
Strength	NGTDM	2D	T2w	Original

GLCM, Gray Level Co-occurrence Matrix; GLDM, Gray Level Dependence Matrix; NGTDM, neighbouring grey tone difference matrix.

TABLE 3 Evaluation metrics for LR, RF and k-NN classifiers using 2D and 3D radiomic features.

	3D - LR	3D - RF	3D - k-NN	2D - LR	2D - RF	2D - k-NN
Specificity	0.87 ± 0.34*	0.19 ± 0.39*	0.53 ± 0.50	1 ± 0.10	0.6 ± 0.49	0.54 ± 0.50
Sensitivity	0.85 ± 0.15*	0.92 ± 0.16*	0.66 ± 0.25*	0.47 ± 0.25	0.62 ± 0.24	0.74 ± 0.25
Balanced accuracy	0.86 ± 0.18*	0.55 ± 0.21	0.59 ± 0.25	0.73 ± 0.13	0.61 ± 0.26	0.64 ± 0.26
AUC	0.9 ± 0.13	0.72 ± 0.32	0.59 ± 0.25	0.86 ± 0.20	0.65 ± 0.32	0.64 ± 0.26
Accuracy	0.85 ± 0.13*	0.77 ± 0.14*	0.63 ± 0.20*	0.57 ± 0.20	0.62 ± 0.20	0.7 ± 0.21
Precision	0.97 ± 0.08*	0.82 ± 0.10*	0.87 ± 0.17	1 ± 0.03	0.88 ± 0.16	0.88 ± 0.14

*p < 0.05 comparison 2D vs. 3D.

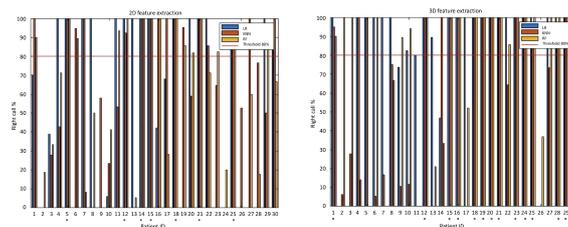


FIGURE 4

Column charts illustrating 2D and 3D radiomics-based machine learning classification performance. On the left, when using a percentage threshold of 80%, it is possible to note that only 7 patients were correctly classified with all classifiers in 2D analysis. On the right, an improvement was obtained in 3D analysis, where 14 patients were correctly classified at least 80% of times with all classifiers (correctly classified patients are marked with an asterisk). With 3D LR classifier (the best performing one) 24 patients were correctly classified at least 80% of times.

and 2D segmentations, 77% and 65% MRI radiomic features were respectively stable to these transformations and then showed good overall reproducibility, although a certain degree of variability existed and highlighted the need for a preliminary reliability assessment.

With the introduction of neoadjuvant chemotherapy, the survival rates of Ewing sarcoma have substantially improved, although they are still low in patients with poor response (2). Accurate response evaluation using non-invasive approaches would be of paramount importance before surgery, as

chemotherapeutic regimen can be adjusted and surgical options range from limb-sparing to radical surgery (1). Several imaging methods have been proposed to identify good and poor responders. Promising results have been obtained using diffusion MRI (23), dynamic contrast-enhanced MRI (24) or positron emission tomography-computed tomography (25). However, these methods often use a mean value to depict whole tumours, potentially overlooking tumour heterogeneity. Radiomics and machine learning have the potential to overcome these limitations (26–29), thus improving prognosis prediction.

Our results showed that MRI radiomics-based machine learning could predict neoadjuvant chemotherapy response with up to 85% accuracy. Similar good performance has been described in previous studies dealing with neoadjuvant chemotherapy response prediction in osteosarcoma (30–32), where the superior predictive capacity of LR was also reported (30), as we observed with Ewing sarcoma. Based on our preliminary findings, we expect that radiomics-based machine learning may facilitate individualized assessment of neoadjuvant chemotherapy response and provide an effective tool for clinical decision-making.

Some limitations of this study need to be acknowledged. First, the design is retrospective, as a prospective analysis is not strictly necessary for radiomic studies (8). Second, the sample size was small. However, Ewing sarcoma of the bone is very rare, and our aim was first to evaluate the reproducibility of 3D and 2D MRI radiomic features. Consequently, our population met the minimum numerical requirements of a reliability analysis in terms of patients involved, according to ICC guidelines (9). Secondly, machine learning prediction of neoadjuvant chemotherapy response was preliminarily explored. Third, good responders were over-represented compared to poor responders in our population of study. However, this reflected the incidence of good and poor response in our clinical practice, and class balancing was performed to artificially oversample the minority class in the training cohort (18). Fourth, the retrospective design accounts for the exclusion of contrast-enhanced MRI, which was not available in all our cases but helps in diagnosis of Ewing sarcoma (33) and deserves investigation in future radiomic studies. Fifth, a validation of the machine learning classifiers was not performed in an independent or external test cohort. The number of patients included at both our institutions was too small to be split into training and external test cohorts. Given the rarity of Ewing sarcoma, further multicentre investigations are warranted to involve more than two institutions, possibly located in different countries. This would allow for machine-learning model training, validation and independent testing on separate datasets, thus moving forward to clinical application.

In conclusion, 3D MRI radiomics of Ewing sarcoma had superior reproducibility and higher accuracy in predicting response to neoadjuvant chemotherapy, particularly when using LR classifier, compared to 2D approach. Although 3D segmentations are relatively time-consuming, they should be preferred in future studies needed to validate our machine learning model for therapy response prediction. If our preliminary findings are confirmed in a larger population, this method will allow for an individualized assessment of therapy response and subsequent treatment planning before surgery is performed, thus offering clinical decision support and having enormous potential for precision medicine.

Data availability statement

The raw data supporting the conclusions of this article will be made available on request by the authors, without undue reservation.

Ethics statement

Institutional Review Board approved this retrospective study and waived the need for informed consent.

Author contributions

SG: Conceptualization, data curation, funding acquisition, investigation, methodology, project administration, visualization, writing – original draft. VC: Data curation, formal analysis, methodology, software, validation, writing – review and editing. AAN: Data curation, writing – review and editing. EM: Investigation, writing – review and editing. MB: Methodology, writing – review and editing. LoM: Methodology, writing – review and editing. DA: Investigation, writing – review and editing. CM: Investigation, writing – review and editing. FS: Investigation, writing – review and editing. VA: Resources, writing – review and editing. VF: Resources, writing – review and editing. CZ: Resources, writing – review and editing. AAL: Resources, writing – review and editing. AP: Resources, writing – review and editing. AL: Resources, writing – review and editing. RB: Resources, writing – review and editing. LuM: Methodology, supervision, writing – review and editing. LS: Conceptualization, funding acquisition, project administration, resources, supervision, writing – review and editing. All authors read and approved the final version of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the Post-Doctoral Fellowship awarded by Fondazione Umberto Veronesi (SG) and Investigator Grant awarded by Fondazione AIRC per la Ricerca sul Cancro for the project "RADIOmics-based machine-learning classification of BOne and Soft Tissue Tumors (RADIO-BOSTT)" (LS). The funding sources provided financial support without any influence on the study design; on the collection, analysis, and interpretation of data; and on the writing of the report. The first and last authors had the final responsibility for the decision to submit the paper for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.1016123/full#supplementary-material>

References

1. Strauss SJ, Frezza AM, Abecassis N, Bajpai J, Bauer S, Biagini R, et al. Bone sarcomas: ESMO-EURACAN-GENTURIS-ERN PaedCan clinical practice guideline for diagnosis, treatment and follow-up. *Ann Oncol* (2021) 32:1520–36. doi: 10.1016/j.annonc.2021.08.1995
2. Stahl M, Ranft A, Paulussen M, Bölling T, Vieth V, Bielack S, et al. Risk of recurrence and survival after relapse in patients with Ewing sarcoma. *Pediatr Blood Cancer* (2011) 57:549–53. doi: 10.1002/pbc.23040
3. Wunder JS, Paulian G, Huvos AG, Heller G, Meyers PA, Healey JH. The histological response to chemotherapy as a predictor of the oncological outcome of operative treatment of Ewing sarcoma. *J Bone Joint Surg Am* (1998) 80:1020–33. doi: 10.2106/00004623-199807000-00011
4. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. *Radiology* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
5. Kocak B, Durmaz ES, Ates E, Kilickesmez O. Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol* (2019) 25:485–95. doi: 10.5152/dir.2019.19321
6. Gitto S, Cuocolo R, Albano D, Morelli F, Pescatori LC, Messina C, et al. CT and MRI radiomics of bone and soft-tissue sarcomas: a systematic review of reproducibility and validation strategies. *Insights Imaging* (2021) 12:68. doi: 10.1186/s13244-021-01008-3
7. Fanciullo C, Gitto S, Carlicchi E, Albano D, Messina C, Sconfienza LM. Radiomics of musculoskeletal sarcomas: A narrative review. *J Imaging* (2022) 8:45. doi: 10.3390/jimaging8020045
8. Lubner MG, Smith AD, Sandrasegaran K, Sahani DV, Pickhardt PJ. CT texture analysis: Definitions, applications, biologic correlates, and challenges. *Radiographics* (2017) 37:1483–503. doi: 10.1148/rg.2017170056
9. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012
10. Picci P, Böhlring T, Bacci G, Ferrari S, Sangiorgi L, Mercuri M, et al. Chemotherapy-induced tumor necrosis as a prognostic factor in localized ewing's sarcoma of the extremities. *J Clin Oncol* (1997) 15:1553–9. doi: 10.1200/JCO.1997.15.4.1553
11. Righi A, Pacheco M, Palmerini E, Carretta E, Gambarotti M, Longhi A, et al. Histological response to neoadjuvant chemotherapy in localized Ewing sarcoma of the bone: A retrospective analysis of available scoring tools. *Eur J Surg Oncol* (2021) 47:1778–83. doi: 10.1016/j.ejso.2021.02.009
12. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015
13. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
14. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging* (2012) 30:1323–41. doi: 10.1016/j.mri.2012.05.001
15. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006
16. Bologna M, Corino VDA, Montini E, Messina A, Calareso G, Greco FG, et al. Assessment of stability and discrimination capacity of radiomic features on apparent diffusion coefficient images. *J Digit Imaging* (2018) 31:879–94. doi: 10.1007/s10278-018-0092-9
17. Gitto S, Bologna M, Corino VDA, Emili I, Albano D, Messina C, et al. Diffusion-weighted MRI radiomics of spine bone tumors: feature stability and machine learning-based classification performance. *Radiol Med* (2022) 127:518–25. doi: 10.1007/s11547-022-01468-7
18. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res* (2002) 16:321–57. doi: 10.1613/jair.953
19. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): A guide for authors and reviewers. *Radiol Artif Intell* (2020) 2:e200029. doi: 10.1148/ryai.2020200029
20. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol (Madr)* (2010) 49:1012–6. doi: 10.3109/0284186X.2010.498437
21. Gitto S, Cuocolo R, Emili I, Tofanelli L, Chianca V, Albano D, et al. Effects of interobserver variability on 2D and 3D CT- and MRI-based texture feature reproducibility of cartilaginous bone tumors. *J Digit Imaging* (2021) 34:820–32. doi: 10.1007/s10278-021-00498-3
22. Kocak B, Durmaz ES, Kaya OK, Ates E, Kilickesmez O. Reliability of Single-Slice-based 2D CT texture analysis of renal masses: Influence of intra- and interobserver manual segmentation variability on radiomic feature reproducibility. *AJR Am J Roentgenol* (2019) 213:377–83. doi: 10.2214/AJR.19.21212
23. Degnan AJ, Chung CY, Shah AJ. Quantitative diffusion-weighted magnetic resonance imaging assessment of chemotherapy treatment response of pediatric osteosarcoma and Ewing sarcoma malignant bone tumors. *Clin Imaging* (2018) 47:9–13. doi: 10.1016/j.clinimag.2017.08.003
24. Kubo T, Furuta T, Johan MP, Adachi N, Ochi M. Percent slope analysis of dynamic magnetic resonance imaging for assessment of chemotherapy response of osteosarcoma or Ewing sarcoma: systematic review and meta-analysis. *Skeletal Radiol* (2016) 45:1235–42. doi: 10.1007/s00256-016-2410-y
25. Annovazzi A, Ferraresi V, Anelli V, Covelto R, Vari S, Zoccali C, et al. [18F] FDG PET/CT quantitative parameters for the prediction of histological response to induction chemotherapy and clinical outcome in patients with localised bone and soft-tissue Ewing sarcoma. *Eur Radiol* (2021) 31:7012–21. doi: 10.1007/s00330-021-07841-w
26. Gitto S, Cuocolo R, Albano D, Chianca V, Messina C, Gambino A, et al. MRI Radiomics-based machine-learning classification of bone chondrosarcoma. *Eur J Radiol* (2020) 128:109043. doi: 10.1016/j.ejrad.2020.109043
27. Gitto S, Cuocolo R, Annovazzi A, Anelli V, Acquasanta M, Cincotta A, et al. CT radiomics-based machine learning classification of atypical cartilaginous tumours and appendicular chondrosarcomas. *EBioMedicine* (2021) 68:103407. doi: 10.1016/j.ebiom.2021.103407

28. Gitto S, Cuocolo R, van Langevelde K, van de Sande MAJ, Parafioriti A, Luzzati A, et al. MRI Radiomics-based machine learning classification of atypical cartilaginous tumour and grade II chondrosarcoma of long bones. *EBioMedicine* (2022) 75:103757. doi: 10.1016/j.ebiom.2021.103757
29. Chianca V, Cuocolo R, Gitto S, Albano D, Merli I, Badalyan J, et al. Radiomic machine learning classifiers in spine bone tumors: A multi-software, multi-scanner study. *Eur J Radiol* (2021) 137:109586. doi: 10.1016/j.ejrad.2021.109586
30. Chen H, Zhang X, Wang X, Quan X, Deng Y, Lu M, et al. MRI-Based radiomics signature for pretreatment prediction of pathological response to neoadjuvant chemotherapy in osteosarcoma: a multicenter study. *Eur Radiol* (2021) 31:7913–24. doi: 10.1007/s00330-021-07748-6
31. Zhang L, Ge Y, Gao Q, Zhao F, Cheng T, Li H, et al. Machine learning-based radiomics nomogram with dynamic contrast-enhanced MRI of the osteosarcoma for evaluation of efficacy of neoadjuvant chemotherapy. *Front Oncol* (2021) 11:758921. doi: 10.3389/fonc.2021.758921
32. Zhong J, Zhang C, Hu Y, Zhang J, Liu Y, Si L, et al. Automated prediction of the neoadjuvant chemotherapy response in osteosarcoma with deep learning and an MRI-based radiomics nomogram. *Eur Radiol* (2022) 32:6196–206. doi: 10.1007/s00330-022-08735-1
33. Murphey MD, Senchak LT, Mambalam PK, Logie CI, Klassen-Fischer MK, Kransdorf MJ. From the radiologic pathology archives: Ewing sarcoma family of tumors: Radiologic-pathologic correlation. *Radiographics* (2013) 33:803–31. doi: 10.1148/rg.333135005