X-rays radiomics-based machine learning classification of atypical cartilaginous tumour and high-grade chondrosarcoma of long bones

Salvatore Gitto,[a,b] Alessio Annovazzi,[c] Kitija Nulle,[d] Matteo Interlenghi,[e] Christian Salvatore,[e,f] Vincenzo Anelli,[g] Jacopo Baldi,[h] Carmelo Messina,[a,b] Domenico Albano,[a,i] Filippo Di Luca,[j] Elisabetta Armiraglio,[k] Antonina Parafioriti,[k] Alessandro Luzzati,[a] Roberto Biagini,[h] Isabella Castiglioni,[l] and Luca Maria Sconfienza[a,b,*]

[a]IRCCS Istituto Ortopedico Galeazzi, Milan, Italy
[b]Dipartimento di Scienze Biomediche per la Salute, Università degli Studi di Milano, Milan, Italy
[c]Nuclear Medicine Unit, IRCCS Regina Elena National Cancer Institute, Rome, Italy
[d]Radiology Department, Riga East Clinical University Hospital, Riga, Latvia
[e]DeepTrace Technologies s.r.l., Milan, Italy
[f]Department of Science, Technology and Society, University School for Advanced Studies IUSS Pavia, Pavia, Italy
[g]Radiology and Diagnostic Imaging Unit, IRCCS Regina Elena National Cancer Institute, Rome, Italy
[h]Oncological Orthopaedics Unit, IRCCS Regina Elena National Cancer Institute, Rome, Italy
[i]Dipartimento di Scienze Biomediche, Chirurgiche ed Odontoiatriche, Università degli Studi di Milano, Milan, Italy
[j]Scuola di Specializzazione in Radiodiagnostica, Università degli Studi di Milano, Milan, Italy
[k]UOC Anatomia Patologica, ASST Gaetano Pini - CTO, Milan, Italy
[l]Department of Physics "G. Occhialini", Università degli Studi di Milano-Bicocca, Milan, Italy

eBioMedicine
2024;101: 105018

Published Online xxx
https://doi.org/10.1016/j.ebiom.2024.105018

Summary

Background Atypical cartilaginous tumour (ACT) and high-grade chondrosarcoma (CS) of long bones are respectively managed with active surveillance or curettage and wide resection. Our aim was to determine diagnostic performance of X-rays radiomics-based machine learning for classification of ACT and high-grade CS of long bones.

Methods This retrospective, IRB-approved study included 150 patients with surgically treated and histology-proven lesions at two tertiary bone sarcoma centres. At centre 1, the dataset was split into training (n = 71 ACT, n = 24 high-grade CS) and internal test (n = 19 ACT, n = 6 high-grade CS) cohorts, respectively, based on the date of surgery. At centre 2, the dataset constituted the external test cohort (n = 12 ACT, n = 18 high-grade CS). Manual segmentation was performed on frontal view X-rays, using MRI or CT for preliminary identification of lesion margins. After image pre-processing, radiomic features were extracted. Dimensionality reduction included stability, coefficient of variation, and mutual information analyses. In the training cohort, after class balancing, a machine learning classifier (Support Vector Machine) was automatically tuned using nested 10-fold cross-validation. Then, it was tested on both the test cohorts and compared to two musculoskeletal radiologists' performance using McNemar's test.

Findings Five radiomic features (3 morphology, 2 texture) passed dimensionality reduction. After tuning on the training cohort (AUC = 0.75), the classifier had 80%, 83%, 79% and 80%, 89%, 67% accuracy, sensitivity, and specificity in the internal (temporally independent) and external (geographically independent) test cohorts, respectively, with no difference compared to the radiologists ($p \geq 0.617$).

Interpretation X-rays radiomics-based machine learning accurately differentiates between ACT and high-grade CS of long bones.

Funding AIRC Investigator Grant.

Keywords: Artificial intelligence; Atypical cartilaginous tumour; Bone neoplasm; Chondrosarcoma; Radiomics

*Corresponding author. Dipartimento di Scienze Biomediche per la Salute, Università degli Studi di Milano and IRCCS Istituto Ortopedico Galeazzi, via Cristina Belgioioso 173, 20157 Milan (MI), Italy.
E-mail address: io@lucasconfienza.it (L.M. Sconfienza).

Research in context

Evidence before this study

Previous radiomic studies focused on the classification of atypical cartilaginous tumour and high-grade chondrosarcoma of long bones based on CT and MRI. Accuracy rates of up to 75% and 92% were reported using CT and MRI radiomics, respectively, in geographically independent test cohorts from different institutions.

Added value of this study

Therapeutic strategies for appendicular cartilaginous lesions are entirely different and, particularly, treatment of atypical cartilaginous tumours is progressively shifting from surgery (curettage) to active surveillance, which involves the repeated use of costly advanced imaging such as MRI. Thus, in this study, we attempted to differentiate atypical cartilaginous tumour from high-grade chondrosarcoma of long bones using radiomics based on frontal view X-rays only. In a large population including 150 patients from two tertiary bone tumour centres, our radiomics-based machine learning classifier achieved 80% accuracy in two independent test cohorts (temporally independent and geographically independent), overlapping two musculoskeletal radiologists who read MRI and/or CT in addition to X-rays in all patients. Furthermore, this approach had 83–89% sensitivity, which makes it ideally suited for screening of high-grade chondrosarcoma, especially considering that only X-rays images are needed as inputs for the analysis. Thus, our method may be helpful to identify high-grade chondrosarcoma both at diagnosis and follow-up.

Implications of all the available evidence

Multimodality imaging radiomics-based machine learning is an objective method that may be used in clinical decision making by accurately differentiating atypical cartilaginous tumour from high-grade chondrosarcoma of long bones.

Introduction

Chondrosarcomas (CSs) account for 20–30% of primary bone tumours in adulthood.[1] In most cases, they arise *de novo* from the medullary cavity and are referred as primary central conventional CSs.[2] Based upon histopathology, conventional CSs are grouped into low to high grades.[3] In the axial skeleton, all-grade CSs exhibit an aggressive behaviour[3] and will not be discussed in the present study. In long bones, low-grade (grade I) cartilaginous lesions are borderline tumours, which are now termed atypical cartilaginous tumours (ACTs) according to 2020 edition of the World Health Organization (WHO) classification.[3] They are locally aggressive lesions with relatively indolent clinical behaviour and unlikelihood to metastasize.[3] Additionally, the increased prevalence of ACTs secondary to an increase in diagnostic imaging over the past decades, relative to the lack of increase in higher-grade lesions, does not support the previous opinion that ACTs are at risk of malignant transformation.[4,5] Conversely, appendicular high-grade CSs (grades II and III) are malignant lesions and show high recurrence rates after surgery with metastatic potential.[3] Thus, the 2020 WHO classification well connects to therapeutic options which are entirely different between ACT and high-grade CS in long bones. Particularly, surgical excision with wide margins is the current standard of care for high-grade CS, whereas ACT can be managed with curettage for sufficient local control.[6] Alternatively, some reference centres now recommend active surveillance with imaging follow-up (watchful waiting) for asymptomatic ACTs, aiming at preventing overtreatment and morbidity associated with surgery.[6]

As clinical management is now very different, the main challenge is to differentiate ACT from high-grade CS in long bones. Preoperatively, biopsy suffers from sample errors[7] and discrepancies in tumour grading even among specialized bone tumour pathologists.[8] Imaging has added substantially to our ability to differentiate between these tumours. Particularly, X-rays are the first imaging investigation and often performed for surgical planning. Magnetic resonance imaging (MRI) is the method of choice for local staging.[6] Computed tomography (CT) and positron emission tomography-CT provide additional information on bone involvement, such as matrix mineralization and cortical destruction, and standard uptake values, respectively.[6,9] However, interobserver variability in tumour grading has been reported even among expert readers.[10,11] Thus, new imaging-based methods like radiomics may improve our ability to better diagnose and grade cartilaginous bone tumours more objectively.[12]

Radiomics includes the analysis of quantitative parameters extracted from medical images, known as radiomic features.[13–15] Radiomic features can be coupled with machine learning to obtain classification models for the diagnosis of interest.[16–19] Machine learning has already shown high accuracy in differentiating ACT from high-grade CS based on CT[20] and MRI[21,22] radiomics. Our aim was to determine diagnostic performance of X-rays radiomics-based machine learning for classification of ACT and high-grade CS of long bones.

Methods

Ethics

Institutional Review Board approved this retrospective study and waived the need for informed consent (Protocol name: "AI tumori MSK", approved by the Ethics Committees of the participating institutions IRCCS Orthopaedic Institute Galeazzi in Milan and IRCCS Regina Elena National Cancer Institute in Rome, Italy). Patients included in this study granted written permission for anonymized data use for research purposes. After matching imaging, pathological, and surgical data, our database was completely anonymized to delete any connections between data and patients' identity according to the General Data Protection Regulation for Research Hospitals. The Checklist for Artificial Intelligence in Medical Imaging[23] and the Standards for Reporting of Diagnostic Accuracy[24] were followed.

Design and inclusion/exclusion criteria

This retrospective study was conducted 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 surgery and pathology departments. Inclusion criteria were: (i) ACT or primary central high-grade CS (grade II or III) of long bones that was surgically treated with curettage or resection and proven by post-surgical pathology; (ii) pre-operative plain radiographs performed within three months before surgery; (iii) MRI and/or CT available for review and aid in tumour identification. Exclusion criteria were: (i) metacarpal, metatarsal and phalangeal location; (ii) recurrent tumours; (iii) pathological fracture; (iv) multiple enchondromatosis (Ollier disease or Maffucci syndrome). A flowchart of the patient selection process is shown in Fig. 1.

Datasets

One-hundred-fifty patients were consecutively included overall. At centre 1, 120 patients were included and split into training cohort and internal test cohort (temporally independent) based on temporal criteria, namely surgery performed in 2011–2020 and 2021–2022, respectively. Thus, the training and internal test cohorts consisted of 95 (n = 71 ACT; n = 24 high-grade CS) and 25 (n = 19 ACT; n = 6 high-grade CS) patients, respectively. At centre 2, 30 patients (n = 12 ACT; n = 18 high-grade CS) were included and constituted the external test cohort (geographically independent). The training and internal/external test cohorts were employed for model tuning and independent testing on unseen data, respectively. Patient demographics and data regarding tumour location are reported in Table 1. All radiographs were obtained using digital radiography systems (Ysio, Siemens Healthcare, Erlangen, Germany at centre 1 and Clisis Evolution, General Medical Merate S.P.A., Seriate, Italy at centre 2). The median image size (pixel x

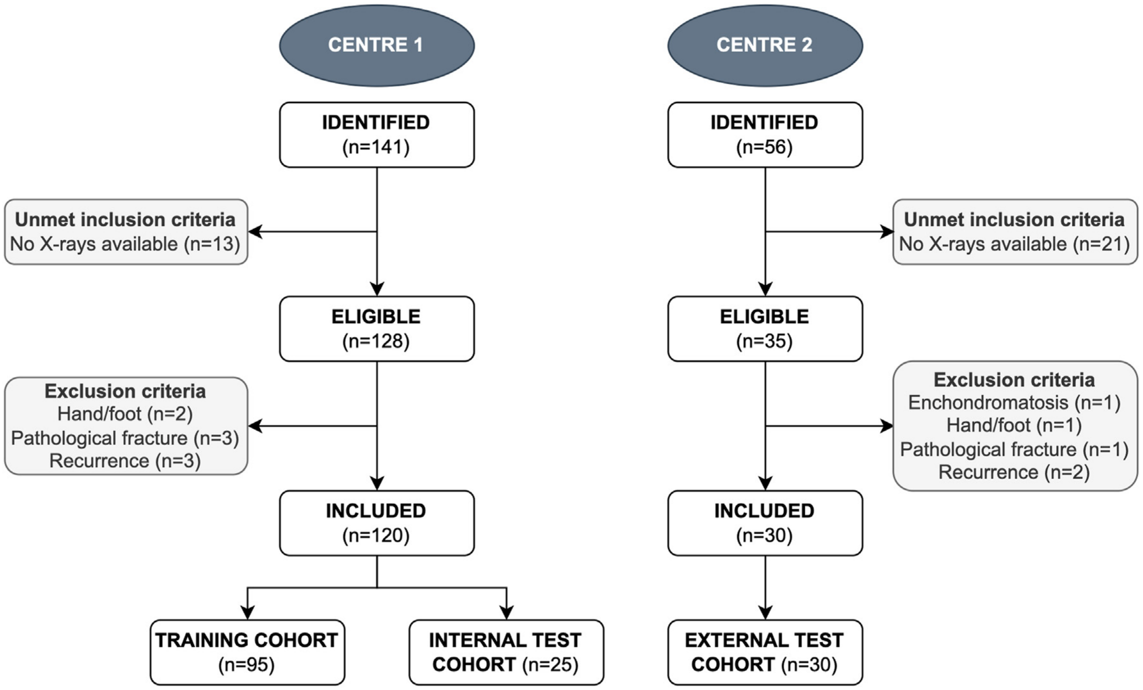

Fig. 1: Flowchart of patient selection.

	Training (centre 1)	Internal test (centre 1)	External test (centre 2)
Age	51 (45–62) years	57 (48–62) years	58 (45–68) years
Sex	Men: n = 37 Women: n = 58	Men: n = 13 Women: n = 12	Men: n = 15 Women: n = 15
Location	Femur: n = 44 Fibula: n = 9 Humerus: n = 34 Radius: n = 1 Tibia: n = 7	Femur: n = 10 Fibula: n = 2 Humerus: n = 10 Tibia: n = 3	Femur: n = 15 Fibula: n = 7 Humerus: n = 3 Tibia: n = 5
Grading	ACT: n = 71 Grade II CS: n = 14 Grade III CS: n = 10	ACT: n = 19 Grade II CS: n = 4 Grade III CS: n = 2	ACT: n = 12 Grade II CS: n = 12 Grade III CS: n = 6

Age is presented as median and interquartile (1st–3rd) range.

Table 1: **Demographics and clinical data for each cohort of patients, which were collected retrospectively through medical records at the participating institutions.**

pixel) was 2135 × 2508 at centre 1 and 3520 × 3520 at centre 2. The automated exposure control system determined the exposure kVp and mAs.

Radiomics-based machine learning analysis

Radiomics-based machine learning analysis was performed according to the International Biomarker Standardization Initiative guidelines.[25] The Trace4Research© radiomic/AI platform (DeepTrace Technologies, www.deeptracetech.com/files/TechnicalSheet__TRACE4.pdf) was used. In detail, our radiomic workflow included different steps as follows.

1. *Segmentation.* A last-year radiology resident (K.N.) performed manual segmentation by drawing an oval-shaped region of interest (ROI) including the lesion on pre-operative frontal view X-rays (Fig. 2). MRI and/or CT (based on availability) were used for identification of lesion margins in all patients before X-rays image segmentation. The reader knew the study would deal with cartilaginous bone tumours but was blinded to post-surgical pathology and grading.
2. *Pre-processing.* Image pre-processing consisted of resampling to isotropic pixel and intensity discretization. In detail, pixels were resampled to isotropic spacing using a down-sampling scheme by considering an image slice thickness of 1 mm. Intensity discretization was obtained using a fixed number of 64 bins.
3. *Feature extraction.* Radiomic features were extracted from the segmented ROI, including morphology, intensity-based statistics, intensity histogram, and texture (Gray-Level Co-occurrence Matrix, Gray-Level Run Length Matrix, Gray-Level Size Zone Matrix, Neighbourhood Gray Tone Difference Matrix, Neighbouring Gray Level Dependence Matrix) features.
4. *Feature dimensionality reduction.* Dimensionality reduction included stability, coefficient of variation, and mutual information analyses. Stability analysis with respect to different segmentations was performed by statistically comparing features obtained through data augmentation strategies. These strategies included randomly manipulating the segmentations and rotating the original images and segmentations. Intraclass correlation coefficient (ICC) was used for stability assessment, and features were considered stable if ICC was 0.8 or higher. For each stable feature, coefficient of variation was computed, and features with coefficient of variation below 0.1 were removed. Finally, a mutual information analysis was performed on the remaining features. In detail, the mutual information between features and associated class labels was estimated, and features with mutual information below 0.28 were removed. The coefficient of variation (0.1) and mutual information (0.28) thresholds were selected semi-automatically. In detail, thresholds were recursively lowered from default starting values of 0.1 and 0.5, respectively, to retain a minimum number of features (at least five features).
5. *Machine learning analysis.* A supervised machine learning model (3 ensembles of 100 support vector machines combined with principal components analysis and fisher discriminant ratio with majority-vote rule) was automatically tuned and validated using nested 10-fold cross-validation on the training cohort. Feature normalization (range = 0–1) was applied to the training data during cross-validation. In detail, the normalization parameters (minimum and maximum) were extracted from the training folds and used to normalize the validation and testing folds for each fold of cross-validation. Feature normalization was then applied to the independent (internal and external) test cohorts, accordingly. Given the unbalanced nature of the training cohort, the adaptive synthetic sampling method (ADASYN)[26] was used to balance the dataset

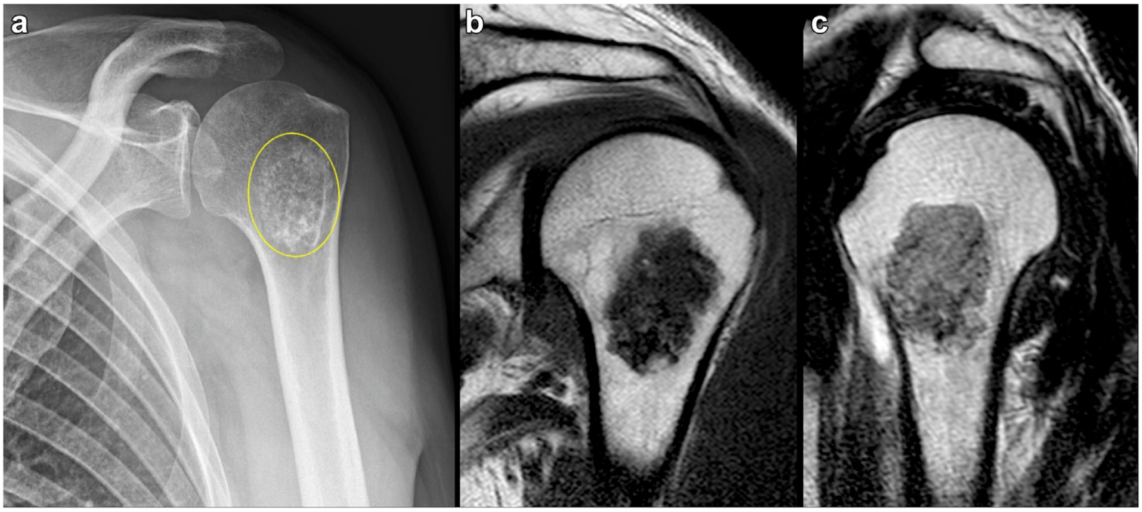

Fig. 2: Manual segmentation. An oval-shaped region of interest including the lesion (yellow oval) was drawn on X-rays (a). Coronal T1-weighted (b) and sagittal T2-weighted (c) MRI sequences were used for preliminary identification of lesion borders before X-rays image segmentation.

by creating new instances from the minority class in the training cohort, thus increasing the number of high-grade CS to n = 71. Hence, the performance of the machine learning model was evaluated on the independent internal and external test cohorts.

Qualitative image assessment

Two musculoskeletal radiologists with 4 (S.G.) and 14 (V.A.) years of work experience in tertiary bone sarcoma centres read the imaging studies from the internal and external test cohorts, respectively, blinded to any information regarding pathology and radiomics-based machine learning analysis. All available imaging studies (X-rays in all cases and CT and/or MRI based on availability) were used for qualitative assessment. The following parameters were assessed to differentiate high-grade CS from ACT and give the final impression: intralesional osteolytic areas without matrix mineralization suggestive of mucoid components, expansion of the medullary canal with cortical thinning and/or remodelling, cortical breakthrough, periosteal reaction, reactive oedema in adjacent bone or soft-tissues (evaluated on MRI), and soft-tissue extension.[27]

Statistical analysis

Statistical analysis was performed using Trace4Research© radiomic/AI platform (DeepTrace Technologies, www.deeptracetech.com/files/TechnicalSheet__TRACE4.pdf). Chi-square and Mann–Whitney *U* tests were performed to evaluate sex and age differences between the two centres, respectively. Mann–Whitney *U* test was also performed to verify feature significance in differentiating ACT from high-grade CS. To account for multiple comparisons, p-values were adjusted using the Bonferroni-Holm method. Sensitivity, specificity, accuracy, receiving operating characteristic (ROC) curve and area under the curve (AUC) were calculated. In the internal and external test cohorts, machine learning performance was compared to qualitative image assessment using McNemar's test. A two-sided p-value <0.05 indicated statistical significance. Radiomics Quality Score was assessed to estimate the methodological rigor of our study, as suggested by Lambin et al.[28]

Role of funding source

This research was supported by the 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)" (L.M.S.). The funding source 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.

Results

There was no significant difference in terms of age (p = 0.263 [Mann–Whitney *U* test]) and sex (p = 0.410 [Chi-square test]) between centre 1 and centre 2. A total of 121 radiomic features were extracted from each lesion. The rate of stable features was 89% (n = 108). Removing low coefficient of variation (n = 17) and low mutual information (n = 86) features yielded a dataset of 5 features (3 morphology and 2 texture features), which passed dimensionality reduction (p < 0.005 [Mann–Whitney *U* test]). The selected features are reported in Table 2. Their distribution is shown in violin and box plots in Fig. 3.

Feature family	Feature nomenclature	Median [95% CI] in high-grade CS class	Median [95% CI] in ACT class
Neighbourhood Grey Tone Difference Matrix	Busyness	3.17 [2.09–4.25]	0.98 [0.75–1.21]
Morphology	Area	3021.56 [1221.87–4821.25]	960.58 [795.56–1125.6]
Morphology	Perimeter To Area Ratio	7.25e-02 [4.90e-02–9.59e-02]	0.13 [0.12–0.14]
Neighbourhood Grey Tone Difference Matrix	Strength	0.25 [0.11–0.39]	0.93 [0.74–1.13]
Morphology	Perimeter	221.13 [162.32–279.94]	120.6 [109.7–131.49]

ACT, atypical cartilaginous tumour; CI, confidence interval; CS, chondrosarcoma. All features were significant (uncorrected and corrected p-value <0.005 [Mann–Whitney *U* test]) in differentiating ACT from high-grade CS. For the definition of each feature, refer to International Biomarker Standardization Initiative official documentation.[25]

Table 2: **Selected radiomic features reported in descending order according to their statistical significance and relevance.**

After tuning on the training cohort (mean AUC = 0.75 [95% confidence interval: 0.70–0.80], shown in ROC curve in Fig. 4), the machine learning classifier had 80% accuracy, 83% sensitivity and 79% specificity in the internal test cohort (temporally independent). It had 80% accuracy, 89% sensitivity and 67% specificity in the external test cohort (geographically independent). The AUC was 0.93 and 0.90 in the internal and external test cohorts, respectively, as shown in ROC curves in Fig. 5. Machine learning performance in the internal and external test cohorts is described in Table 3. Fig. 6 shows different cases of ACT and high-grade CS from the internal test cohort, which were correctly identified or misdiagnosed by the machine learning classifier. The radiologists had 88% (22/25 correctly classified lesions) and 80% (24/30 correctly classified lesions) accuracy in the internal and external test cohorts, respectively, with no statistical difference compared to machine learning (p = 0.683 and p = 0.617 [McNemar's test], respectively). Our radiomics quality score was 44% (Supplementary Table).

Discussion

In this study, our main finding was that machine learning had 80% accuracy in differentiating ACT from high-grade CS of long bones based on radiomic features extracted from frontal view X-rays in both the internal (temporally independent) and external (geographically independent) test cohorts from two different centres, respectively. Machine learning performance did not differ compared to musculoskeletal radiologists working in tertiary bone sarcoma centres ($p \geq 0.617$ [McNemar's test]), who used MRI and/or CT in addition to X-rays for qualitative image assessment.

Our results are clinically relevant as therapeutic strategies for ACT and high-grade CS are now entirely different in long bones. Particularly, wide resection with negative margins remains the therapy of choice for high-grade CS.[6] On the other hand, the treatment of ACT has been changing remarkably over time. Intralesional curettage has been standard of care for several years in tertiary sarcoma centres.[6] Nowadays, however, treatment is shifting from surgery to active surveillance

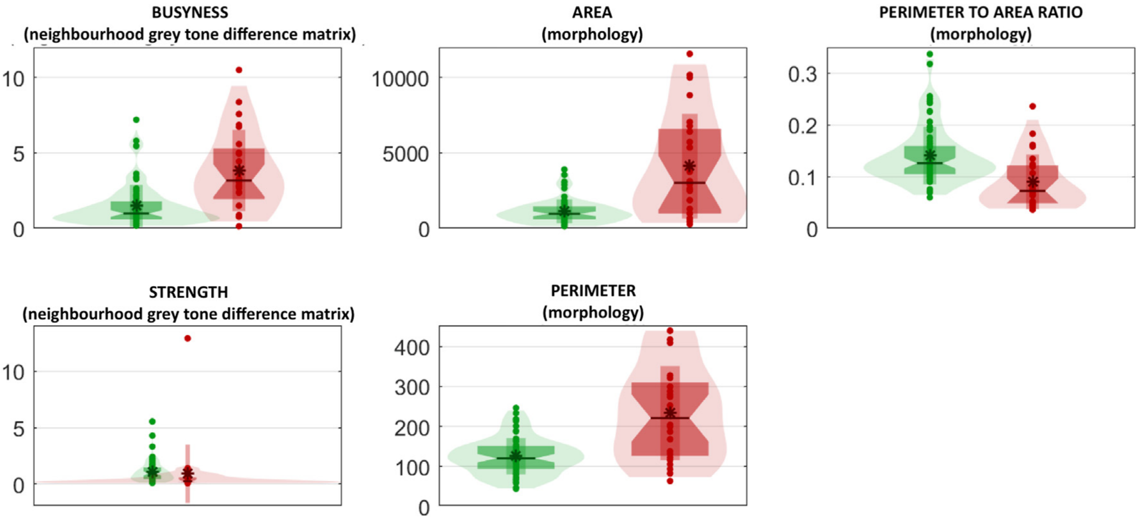

Fig. 3: Violin and box plots of the selected radiomic features, which passed dimensionality reduction. Violin and box plots of high-grade CS and ACT classes are reported in red and green, respectively.

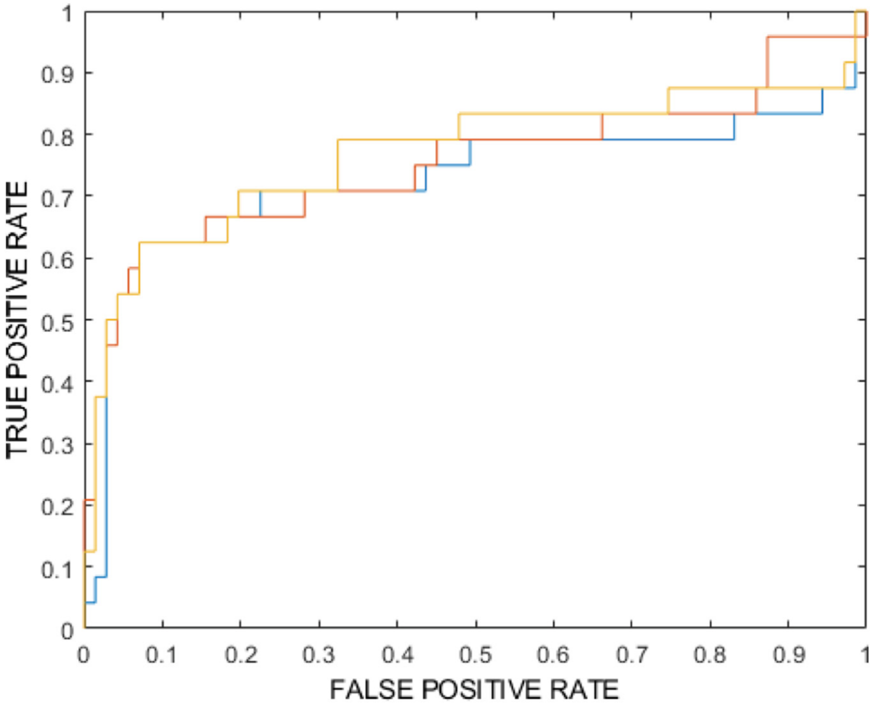

Fig. 4: ROC curve for the machine learning model consisting of 3 ensembles of support vector machine classifiers from the testing folds of cross validation.

based on new insights on the natural course of ACTs, with low rates of tumour growth and no reported transformation into high-grade CS at follow-up.[29–33] Thus, the difference in therapeutic strategies between enchondroma and ACT is progressively disappearing. MRI is recommended for ACT follow-up in centres where a conservative strategy is favoured over surgery,[29–32] with some institutions also performing dynamic contrast-enhanced MRI sequences to improve accuracy.[12] Nonetheless, the strategies for active

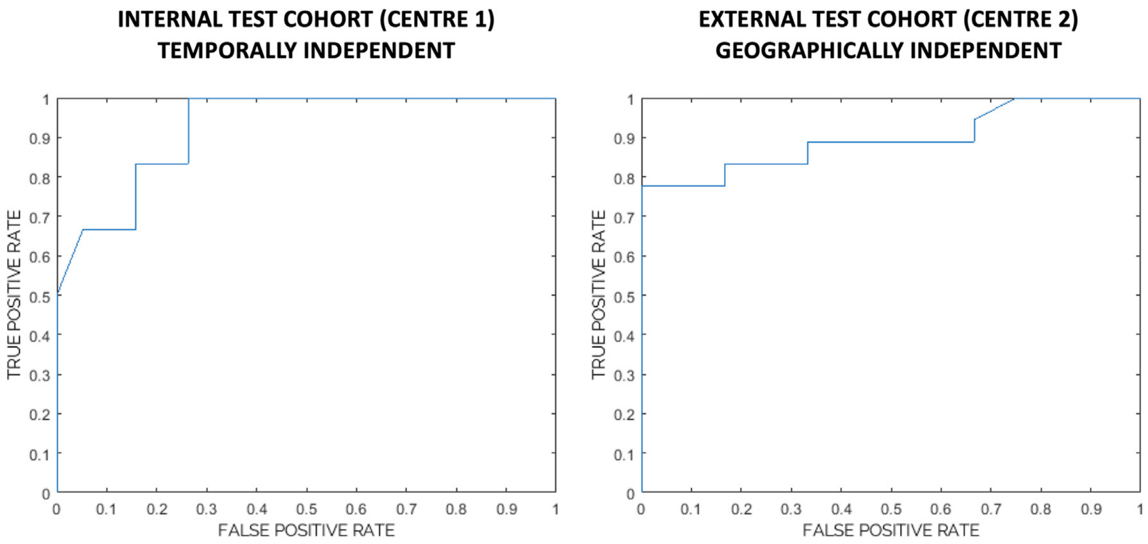

Fig. 5: ROC curves showing the machine learning classifier performances in the internal (temporally independent) and external (geographically independent) test cohorts from centre 1 and centre 2, respectively.

	Internal test cohort (centre 1)	External test cohort (centre 2)
Correctly classified high-grade CS (true positive)	n = 5	n = 16
Correctly classified ACT (true negative)	n = 15	n = 8
ACT misdiagnosed as high-grade CS (false positive)	n = 4	n = 4
High-grade CS misdiagnosed as ACT (false negative)	n = 1	n = 2
Accuracy	80%	80%
Sensitivity	83%	89%
Specificity	79%	67%
AUC	0.93	0.90

Table 3: **Performance of the machine learning classifier evaluated on the internal (temporally independent) and external (geographically independent) test cohorts.**

surveillance are still debated, including frequency, duration and employed imaging methods. Based on our preliminary findings, X-rays combined with radiomics-based machine learning may potentially be helpful to prevent overutilization of costly advanced imaging such as MRI for active surveillance of ACTs.

Radiomic studies to date have focused on the classification of ACT and high-grade CS based on CT[20] and MRI,[21,22,34] using radiomics-based nomograms[34] or machine learning approaches.[20–22] Particularly, in a recent study, Gitto et al.[22] focused on the differentiation of ACT and grade II CS of long bones using T1-weighted MRI radiomics-based machine learning. This approach showed 92% accuracy in the external test cohort, with no difference compared to an experienced bone tumour radiologist.[22] Previously, the same authors' group performed a similar analysis using CT radiomics-based machine learning.[20] In the latter study, machine learning achieved 75% accuracy in differentiating ACT from high-grade CS of long bones in the external test cohort, overlapping an expert radiologist.[20] In these previous studies, a two-centre population ranging from 120 to 158 patients was split into training and external test cohorts for machine learning model tuning and independent testing, respectively, based on geographical criteria.[20,22] In the current study including 150 patients

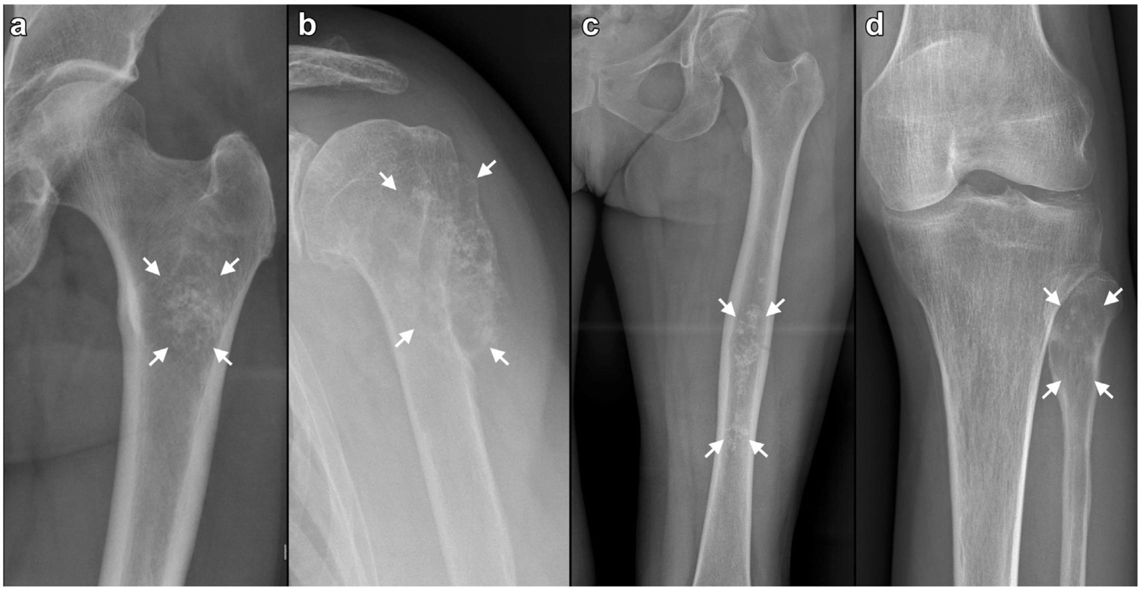

Fig. 6: X-rays images show two different cases of ACT of the proximal femur (a) and high-grade CS (grade II) of the proximal humerus (b), which were correctly identified by the machine learning classifier. No aggressive features are noted in the proximal femur (a), whereas cortical breakthrough and extra-osseous extension are indicative of high-grade lesion in the humerus (b). X-rays images also show two different cases of ACT of the mid-femoral diaphysis (c) and high-grade CS (grade II) of the proximal fibula (d), which were respectively misdiagnosed as high-grade CS and ACT by the classifier. In the ACT of the mid-femoral diaphysis (c), mild expansion of the medullary canal with no cortical destruction is noted. In the high-grade CS of the fibula (d), expansion of the medullary canal with cortical thinning is observed. Cortical destruction is not well seen on X-rays but was demonstrated on MRI (not shown) in the latter case. Arrows point at the lesion in all images.

with ACT and appendicular high-grade CS from two tertiary bone sarcoma centres, we performed our radiomic analysis using frontal view X-rays and tested our machine learning classifier on two independent test cohorts, namely the temporally-independent internal and geographically-independent external test cohorts. Our machine learning classifier achieved 80% accuracy in both test cohorts, overlapping two musculoskeletal radiologists who read MRI and/or CT in addition to X-rays in all patients. Furthermore, this approach had 83–89% sensitivity, which makes it ideally suited for screening of high-grade CS, especially considering that only X-rays images are needed as inputs for the analysis.

Some limitations of this study need to be addressed. First, this study was retrospective. Although prospective design provides the highest level of evidence supporting the clinical validity and usefulness of radiomics,[28] a relatively large number of patients with an uncommon disease and imaging data already available could be included retrospectively in our study. Second, only patients with cartilaginous bone tumours were included, and our method could not be applied to X-rays images showing different bone lesions or none. However, our study was intentionally focused on grading of lesions, which were already classified as cartilaginous tumours using routinely available tools, such as MRI. Additionally, this method could provide complementary information to different machine learning approaches for identification and classification of bone lesions on X-rays, which were mainly based on deep learning algorithms and achieved encouraging results in previous studies.[35–41] Third, compared to high-grade CS, ACT were over-represented in centre 1 and under-represented in centre 2, resulting in unbalanced classes. However, in centre 1, this reflected the true prevalence of these lesions in clinical practice, and class balancing was performed to artificially oversample the minority class in the training cohort.[26] In centre 2, where a selection bias related to the unavailability of X-rays images in one third of the identified patients occurred, the unbalance was relatively small (40% ACT and 60% high-grade CS) and acceptable for a machine learning analysis.[42] Fourth, only frontal view X-rays images were included in our radiomics-based machine learning analysis, and lateral views were not evaluated. Although our intention was to keep our model as simple as possible by focusing on a single image, lateral view X-rays deserve further investigation. Fifth, segmentation was performed manually following the preliminary identification of lesion margins on MRI and/or CT. Although radiomic feature stability was assessed as a dimensionality reduction method in our analysis, interobserver variability could ideally be reduced using deep learning-based automated segmentation,[43,44] which is a promising technique and will be employed in future studies. Finally, our radiomics quality score was 44%. This was in line with average values described in a recent systematic review focusing on radiomics quality score applications[45] and higher compared to median radiomics quality scores reported in another systematic review addressing the quality of radiomic studies on bone chondrosarcoma.[46] Nonetheless, further improvements in methodological quality can still be obtained.

In conclusion, radiomics-based machine learning is an objective method, using frontal view X-rays images only, that may be used in the management of cartilaginous bone lesions by accurately differentiating between ACT and high-grade CS of long bones. As active surveillance of ACTs is an increasingly favoured option over surgery, our method may be helpful to identify high-grade CS both at diagnosis and follow-up. Our large population of study and the good performance obtained using independent data from two different centres ensure the generalizability of our results. Future prospective investigations will verify the transferability of our findings into clinical practice.

Contributors

Salvatore Gitto: conceptualization; data curation; formal analysis; investigation; methodology; project administration; validation; visualization; writing – original draft.

Alessio Annovazzi: data curation; investigation; writing – review and editing.

Kitija Nulle: data curation; investigation; writing – review and editing.

Matteo Interlenghi: software; writing – review and editing.

Christian Salvatore: software; writing – review and editing.

Vincenzo Anelli: resources; writing – review and editing.

Jacopo Baldi: resources; writing – review and editing.

Carmelo Messina: investigation; writing – review and editing.

Domenico Albano: investigation; writing – review and editing.

Filippo Di Luca: investigation; writing – review and editing.

Elisabetta Armiraglio: resources; writing – review and editing.

Antonina Parafioriti: resources; writing – review and editing.

Alessandro Luzzati: resources; writing – review and editing.

Roberto Biagini: resources; writing – review and editing.

Isabella Castiglioni: software; supervision; validation; writing – review and editing.

Luca Maria Sconfienza: conceptualization; funding acquisition; project administration; resources; supervision; validation; writing – review and editing.

All authors read and approved the final version of the manuscript.

Data sharing statement

Deidentified datasets are available in the online repository Zenodo (Zenodo.org). The model, its implementation instructions and all required files for data extraction and processing are available from the corresponding author by request.

Declaration of interests

Matteo Interlenghi: CTO and employee of DeepTrace Technologies. DeepTrace Technologies is a spin-off of Scuola Universitaria Superiore IUSS, Pavia, Italy; shareholder in DeepTrace Technologies. Christian Salvatore: CEO of DeepTrace Technologies. DeepTrace Technologies is a spin-off of Scuola Universitaria Superiore IUSS, Pavia, Italy; shareholder in DeepTrace Technologies. Isabella Castiglioni: Shareholder in DeepTrace Technologies. All other authors declare that they have no conflicts of interest to disclose.

Acknowledgements

This research was supported by the 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)" (L.M. Sconfienza). The funding source 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.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2024.105018.

References

1 Murphey MD, Walker EA, Wilson AJ, Kransdorf MJ, Temple HT, Gannon FH. From the archives of the AFIP: imaging of primary chondrosarcoma: radiologic-pathologic correlation. *Radiographics*. 2003;23:1245–1278.

2 Gelderblom H, Hogendoorn PCW, Dijkstra SD, et al. The clinical approach towards chondrosarcoma. *Oncologist*. 2008;13:320–329.

3 WHO Classification of Tumours Editorial Board. *WHO classification of tumours: soft tissue and bone tumours*. Lyon, France: International Agency for Research on Cancer Press; 2020.

4 Davies AM, Patel A, Botchu R, Azzopardi C, James S, Jeys L. The changing face of central chondrosarcoma of bone. One UK-based orthopaedic oncology unit's experience of 33 years referrals. *J Clin Orthop Trauma*. 2021;17:106–111.

5 van Praag Veroniek VM, Rueten-Budde AJ, Ho V, et al. Incidence, outcomes and prognostic factors during 25 years of treatment of chondrosarcomas. *Surg Oncol*. 2018;27:402–408.

6 Strauss SJ, Frezza AM, Abecassis N, et al. Bone sarcomas: ESMO–EURACAN–GENTURIS–ERN PaedCan clinical practice guideline for diagnosis, treatment and follow-up. *Ann Oncol*. 2021;32:1520–1536.

7 Hodel S, Laux C, Farei-Campagna J, Götschi T, Bode-Lesniewska B, Müller DA. The impact of biopsy sampling errors and the quality of surgical margins on local recurrence and survival in chondrosarcoma. *Cancer Manag Res*. 2018;10:3765–3771.

8 Eefting D, Schrage YM, Geirnaerdt MJA, et al. Assessment of interobserver variability and histologic parameters to improve reliability in classification and grading of central cartilaginous tumors. *Am J Surg Pathol*. 2009;33:50–57.

9 Annovazzi A, Anelli V, Zoccali C, et al. 18F-FDG PET/CT in the evaluation of cartilaginous bone neoplasms: the added value of tumor grading. *Ann Nucl Med*. 2019;33:813–821.

10 Skeletal Lesions Interobserver Correlation among Expert Diagnosticians (SLICED) Study Group. Reliability of histopathologic and radiologic grading of cartilaginous neoplasms in long bones. *J Bone Joint Surg Am*. 2007;89:2113–2123.

11 Zamora T, Urrutia J, Schweitzer D, Amenabar PP, Botello E. Do orthopaedic oncologists agree on the diagnosis and treatment of cartilage tumors of the appendicular skeleton? *Clin Orthop Relat Res*. 2017;475:2176–2186.

12 van de Sande MAJ, van der Wal RJP, Navas Cañete A, et al. Radiologic differentiation of enchondromas, atypical cartilaginous tumors, and high-grade chondrosarcomas—improving tumor-specific treatment: a paradigm in transit? *Cancer*. 2019;125:3288–3291.

13 Gitto S, Cuocolo R, Emili I, 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;34:820–832.

14 Gitto S, Cuocolo R, Albano D, 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.

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

16 Gitto S, Bologna M, Corino VDA, et al. Diffusion-weighted MRI radiomics of spine bone tumors: feature stability and machine learning-based classification performance. *Radiol Med*. 2022;127:518–525.

17 Gitto S, Corino VDA, Annovazzi A, et al. 3D vs. 2D MRI radiomics in skeletal Ewing sarcoma: feature reproducibility and preliminary machine learning analysis on neoadjuvant chemotherapy response prediction. *Front Oncol*. 2022;12:1016123.

18 Gitto S, Interlenghi M, Cuocolo R, et al. MRI radiomics-based machine learning for classification of deep-seated lipoma and atypical lipomatous tumor of the extremities. *Radiol Med*. 2023;128:989–998.

19 Gitto S, Serpi F, Albano D, et al. AI applications in musculoskeletal imaging: a narrative review. *Eur Radiol Exp*. 2024. https://doi.org/10.1186/s41747-024-00422-8.

20 Gitto S, Cuocolo R, Annovazzi A, et al. CT radiomics-based machine learning classification of atypical cartilaginous tumours and appendicular chondrosarcomas. *EBioMedicine*. 2021;68:103407.

21 Gitto S, Cuocolo R, Albano D, et al. MRI radiomics-based machine-learning classification of bone chondrosarcoma. *Eur J Radiol*. 2020;128:109043.

22 Gitto S, Cuocolo R, van Langevelde K, et al. MRI radiomics-based machine learning classification of atypical cartilaginous tumour and grade II chondrosarcoma of long bones. *EBioMedicine*. 2022;75:103757.

23 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.

24 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology*. 2015;277:826–832.

25 Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295:328–338.

26 He H, Bai Y, Garcia EA, Li S. ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE; 2008:1322–1328.

27 Douis H, Singh L, Saifuddin A. MRI differentiation of low-grade from high-grade appendicular chondrosarcoma. *Eur Radiol*. 2014;24:232–240.

28 Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749–762.

29 Omlor GW, Lohnherr V, Lange J, et al. Outcome of conservative and surgical treatment of enchondromas and atypical cartilaginous tumors of the long bones: retrospective analysis of 228 patients. *BMC Musculoskelet Disord*. 2019;20:134.

30 Scholte CHJ, Dorleijn DMJ, Krijvenaar DT, van de Sande MAJ, van Langevelde K. Wait-and-scan: an alternative for curettage in atypical cartilaginous tumours of the long bones. *Bone Joint J*. 2024;106-B:86–92.

31 Deckers C, Schreuder BHW, Hannink G, de Rooy JWJ, van der Geest ICM. Radiologic follow-up of untreated enchondroma and atypical cartilaginous tumors in the long bones. *J Surg Oncol*. 2016;114:987–991.

32 Deckers C, de Rooy JWJ, Flucke U, Schreuder HWB, Dierselhuis EF, van der Geest ICM. Midterm MRI follow-up of untreated enchondroma and atypical cartilaginous tumors in the long bones. *Cancers (Basel)*. 2021;13:4093.

33 Deckers C, van Zeijl NT, van Hooff ML, et al. Active surveillance of atypical cartilaginous tumours of bone: short term quality of life measurements. *J Orthop Surg Res*. 2023;18:208.

34 Li X, Lan M, Wang X, et al. Development and validation of a MRI-based combined radiomics nomogram for differentiation in chondrosarcoma. *Front Oncol*. 2023;13:1090229.

35 Consalvo S, Hinterwimmer F, Neumann J, et al. Two-phase deep learning algorithm for detection and differentiation of ewing sarcoma and acute osteomyelitis in paediatric radiographs. *Anticancer Res*. 2022;42:4371–4380.

36 He Y, Pan I, Bao B, et al. Deep learning-based classification of primary bone tumors on radiographs: a preliminary study. *EBioMedicine*. 2020;62:103121.

37 Pan D, Liu R, Zheng B, et al. Using machine learning to unravel the value of radiographic features for the classification of bone tumors. *Biomed Res Int*. 2021;2021:8811056.

38 Pan C, Lian L, Chen J, Huang R. FemurTumorNet: bone tumor classification in the proximal femur using DenseNet model based on radiographs. *J Bone Oncol*. 2023;42:100504.

39 Park C-W, Oh S-J, Kim K-S, et al. Artificial intelligence-based classification of bone tumors in the proximal femur on plain radiographs: system development and validation. *PLoS One*. 2022;17:e0264140.

40 von Schacky CE, Wilhelm NJ, Schäfer VS, et al. Development and evaluation of machine learning models based on X-ray radiomics for the classification and differentiation of malignant and benign bone tumors. *Eur Radiol*. 2022;32:6247–6257.

41 von Schacky CE, Wilhelm NJ, Schäfer VS, et al. Multitask deep learning for segmentation and classification of primary bone tumors on radiographs. *Radiology*. 2021;301:398–406.

42 Sollini M, Antunovic L, Chiti A, Kirienko M. Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. *Eur J Nucl Med Mol Imaging*. 2019;46: 2656–2672.

43 Do N-T, Jung S-T, Yang H-J, Kim S-H. Multi-level Seg-Unet model with global and patch-based X-ray images for knee bone tumor detection. *Diagnostics (Basel)*. 2021;11:691.

44 Breden S, Hinterwimmer F, Consalvo S, et al. Deep learning-based detection of bone tumors around the knee in X-rays of children. *J Clin Med*. 2023;12:5960.

45 Spadarella G, Stanzione A, Akinci D'Antonoli T, et al. Systematic review of the radiomics quality score applications: an EuSoMII Radiomics Auditing Group Initiative. *Eur Radiol*. 2022;33: 1884–1894.

46 Zhong J, Hu Y, Ge X, et al. A systematic review of radiomics in chondrosarcoma: assessment of study quality and clinical value needs handy tools. *Eur Radiol*. 2022;33:1433–1444.