Prediction of neoadjuvant chemotherapy response in high-grade osteosarcoma: added value of non-tumorous bone radiomics using CT images

Lei Xu[1#], Pengfei Yang[2#], Kun Hu[3], Yan Wu[4], Meng Xu-Welliver[5], Yidong Wan[1], Chen Luo[1], Jing Wang[1], Jinhua Wang[6], Jiale Qin[7], Yi Rong[8], Tianye Niu[3]

[1]Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Institute of Translational Medicine, Zhejiang University, Hangzhou, China; [2]College of Biomedical Engineering &Instrument Science, Zhejiang University, Hangzhou, China; [3]Nuclear & Radiological Engineering and Medical Physics Programs, Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA; [4]Department of Orthopedics, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China; [5]Department of Radiation Oncology, The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA; [6]Department of Radiology, Jiangxi Maternal and Child Health Hospital, Nanchang, China; [7]Women's Hospital, Zhejiang University School of Medicine, Hangzhou, China; [8]Department of Radiation Oncology, Mayo Clinic Arizona, Phoenix, AZ, USA

[#]These authors contributed equally to this work.

Correspondence to: Tianye Niu, PhD. Nuclear & Radiological Engineering and Medical Physics Programs, Woodruff School of Mechanical Engineering, Georgia Institute of Technology, 770 State Street, Boggs 385, Atlanta, GA 30332-0745, USA. Email: tyniu@gatech.edu; Yi Rong, PhD. Department of Radiation Oncology, Mayo Clinic Arizona, 5777 E. Mayo Blvd, Phoenix, AZ 85054, USA. Email: rong.yi@mayo.edu.

Background: This study aimed to determine the impact of including radiomics analysis of non-tumorous bone region of interest in improving the performance of pathological response prediction to chemotherapy in high-grade osteosarcomas (HOS), compared to radiomics analysis of tumor region alone.

Methods: This retrospective study included 157 patients diagnosed with HOS between November 2013 and November 2017 (age range, 5–44 years; mean age, 16.99 ±7.42 years), in which 69 and 88 patients were diagnosed as pathological good response (pGR) and non-pGR, respectively. Radiomics features were extracted from tumor and non-tumorous bone regions based on diagnostic CT images. Pathological response classifiers were developed and validated via leave-one-out cross validation (LOOCV) and independent validation methods by using the area under the receiver operating characteristic curve (AUC) value as the figure of merit.

Results: Using the LOOCV, the classifiers combining features from tumor and non-tumorous regions showed better prediction performance than those from tumor region alone (AUC, 0.8207±0.0043 *vs.* 0.7799±0.0044). The combined classifier also showed better performance than the tumor feature-based classifier in both training and validation datasets [training dataset: 0.791, 95% confidence interval (CI), 0.706–0.860 *vs.* 0.766, 95% CI, 0.679–0.840; validation dataset: 0.816, 95% CI, 0.662–0.920 *vs.* 0.766, 95% CI, 0.606–0.885].

Conclusions: Radiomics analysis of combined tumor and non-tumorous bone features showed improved performance of pathological response prediction to chemotherapy in HOS compared to that of tumor features alone. Moreover, the proposed classifier had the potential to predict pathological response to chemotherapy for HOS patients.

Keywords: High-grade osteosarcoma (HOS); neoadjuvant chemotherapy response; non-tumorous bone features; added value

Introduction

Osteosarcoma is the most common primary malignant bone tumor originated from the primitive bone-forming mesenchymal cells (1). When diagnosed, around 90% of osteosarcoma patients are classified as high-grade osteosarcoma (HOS) (2). Preoperative neoadjuvant chemotherapy combined with definitive surgical resection, as the standard treatment option, has significantly improved the five-year survival rate from less than 20% to more than 60% (3). Unfortunately, there still remains a considerable amount of osteosarcoma patients that might not benefit from chemotherapy, especially those with poor histologic responses from chemotherapy (4,5). Therefore, accurate pre-treatment pathological response prediction to chemotherapy in patients with HOS is essential in determining the best treatment option. If poor response could be predicted, it might be helpful for modifying treatment regimen for chemotherapy and/or surgery (6,7).

Radiomics analysis based on medical images has been used to predict outcomes of osteosarcoma treated with chemotherapy in previous studies (8-10). Song *et al.* found that tumor texture features of baseline 18F-fluorodeoxyglucose uptake measured by positron emission tomography (PET) images could predict pathological response to chemotherapy in patients with localized osteosarcoma of the extremities (10). Davis *et al.* reported that the percentage change of maximum standardized uptake during treatment could be used as metabolic predictors of pathological response in osteosarcoma (9). Wu *et al.* illustrated that the percentage change of maximum standardized uptake during treatment could be used as metabolic predictors of pathological response in osteosarcoma (11). Our previous study demonstrated that delta CT imaging features could be used for individualized pathologic response evaluation after chemotherapy in osteosarcoma (12). These findings showed that imaging features extracted from tumors have the potential to predict chemotherapy outcomes.

Several previous studies suggested that texture features from normal parenchymal structures of solid tumors could reflect the biologic factors associated with tumor development, which could be interpreted by the important role of normal parenchymal stromal cells in tumor formation and development (13-15). Sala *et al.* demonstrated that parenchymal texture features could be used to evaluate or predict the stage and subtype in breast cancer (16). Li *et al.* found that combining quantitative radiomics features from tumors with contralateral parenchyma characterizations can improve the diagnostic accuracy for breast cancer (17). The idea of combined features for breast cancer intrigued our present study of looking into the correlation between non-tumorous bone characteristics and tumor development in osteosarcoma.

This study was developed with a hypothesis that the combination of imaging features extracted from non-tumorous bone region and tumor region on diagnostic CT images could improve the accuracy of pathological response prediction to chemotherapy. Quantitative CT imaging features from both tumor and non-tumorous bone regions were extracted and analyzed. The prediction performance of pathological response classifiers developed based on the combination features was assessed and compared with the classifiers developed based on the tumor features alone.

Methods

Patient data set

This retrospective study was approved by the Institutional Research Ethics Board. Patient informed consent was waived for the study. All medical images and baseline characteristics were obtained from the Second Affiliated Hospital, Zhejiang University School of Medicine (Zhejiang, China) between November 2013 and November 2017.

The focus of the study was on HOS in extremities, considering that large tumor size can impact the accuracy of non-tumorous bone region delineation. Patients' inclusion criteria are: (I) patients were diagnosed as HOS in extremities; (II) patients underwent diagnostic CT scans before chemotherapy; (III) patients underwent surgical resections after chemotherapy; (IV) patients had confirmed pathological response to chemotherapy by pathological examination; (V) patients had complete baseline characteristics and histologic information. The baseline characteristics were obtained from the Electronic Medical Record System (EMRS), including age, gender, and location of primary tumor, tumor stage, pathologic subtype, any pulmonary metastasis, and type of chemotherapy regimens. The flowchart of patient exclusion criteria was presented in *Figure 1*.

The baseline characteristics of patients with pathological good response (pGR) and non-pGR were described in *Table 1*. A total of 157 patients with HOS were involved in this study, 69 of whom showed pGR to chemotherapy and 88

1186

Xu et al. Added value of non-tumorous radiomics in NC response prediction in HOS
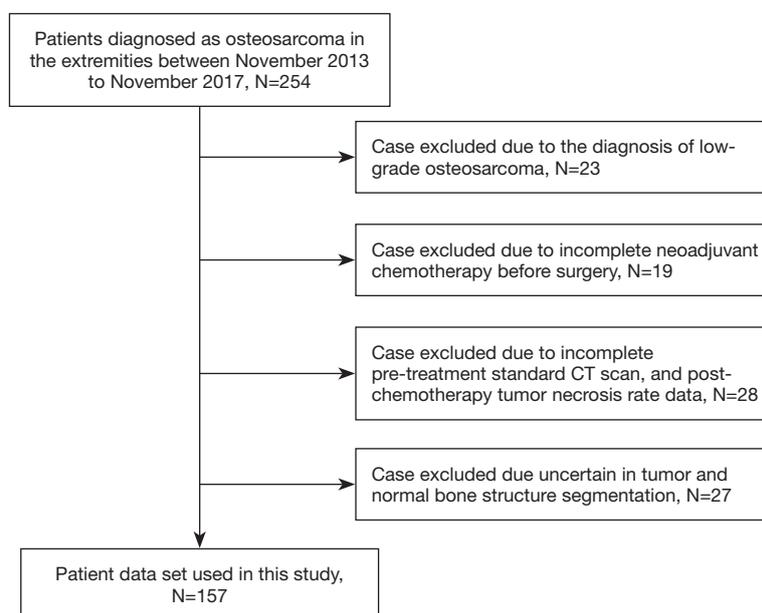
Figure 1 Flowchart showing the number of patients with high-grade osteosarcoma in the extremities used in this study.

showed non-pGR. Specifically, tumor volumes for patients with pGR ranged from 15.53 to 372.70 cm^3 (mean ± standard deviation, 134.82±84.45 cm^3), and 7.02 to 1,142.96 cm^3 (173.57±160.13 cm^3) for patients with non-pGR. Patient dataset was stratified to the training dataset and independent validation dataset randomly at a ratio of 3:1.

CT image acquisition

All images were acquired on five CT scanners, including a 16 multi-slice CT (MSCT) (Siemens Sensation 16, Siemens Medical Systems, Erlangen, Germany), a 64-MSCT (Philips Brilliance 64, Philips Medical Systems, Cleveland, USA), a 128-MSCT (Definition Flash, Siemens Medical Systems), a 160-MSCT (Volume Zoom, Siemens Medical Systems) and a 320-MSCT (Aquilion ONE, Toshiba Medical Systems, Otawara, Japan). All CT scan parameters were as follows: tube voltage of 100–140 kVp, tube current of 80–480 mAs, slice thickness of 3–6 mm, and pixel size of 0.3×0.3–1.0×1.0 mm^2. Standard DICOM format was used for all medical images in this study.

Neoadjuvant chemotherapy, pathological response assessment and surgery

All patients received preoperative chemotherapy followed by surgical resection. The neoadjuvant chemotherapy schema was designed according to the National Comprehensive Cancer Network guidelines. The total duration of neoadjuvant chemotherapy ranged from 8 to 10 weeks.

The conventional three-drug regimen was used consisting of methotrexate, cisplatin, and doxorubicin. This regimen was administered in 9 weeks: week 1 and week 2: methotrexate (8 mg/m^2 in 6 hours for patients <14 years of age, 12 mg/m^2 in 6 hours for patients ≥14 years of age); week 3: cisplatin (100 mg/m^2 in 24 hours) and doxorubicin (60 mg/m^2 in 8 hours); week 5 and week 6: methotrexate; week 7: cisplatin and doxorubicin; week 9: surgery. Patients who suffered severe liver dysfunction or other adverse reactions after the first week of treatment received an alternative treatment schema consisting of methotrexate, ifosfamide, cisplatin and doxorubicin in 11 weeks: week 1: methotrexate; week 2: ifosfamide (2 mg/m^2/day for 5 days); week 4: cisplatin and doxorubicin; week 6: ifosfamide; week 8: ifosfamide; week 9: cisplatin and doxorubicin; week 11: surgery. Patients who suffered tumor progression or newly diagnosed lung metastasis after the first week of treatment received an alternative treatment schema consisting of methotrexate, ifosfamide, cisplatin and doxorubicin in 10 weeks: week 1 and week 2: methotrexate; week 3: cisplatin and doxorubicin; week 5 and week 7: ifosfamide; week 8: cisplatin and doxorubicin; week 10: surgery.

Histology analysis of pathological response to

Table 1 Baseline characteristics of patients with high-grade osteosarcoma

Baseline characteristics	pGR (N=69)	Non-pGR (N=88)
Age, years		
Mean [range]	17.65±7.57 [5–43]	16.48±7.30 [7–44]
≤15	33	47
>15	36	41
Sex		
Male	41	50
Female	28	38
Location of primary tumor		
Humerus	8	5
Femur	38	58
Tibia and fibula	23	24
Radius and ulna	0	1
Tumor stage		
Localized	57	76
Metastatic	12	12
Pathologic subtype		
Osteoblastic	58	58
Chondroblastic	3	16
Fibroblastic	7	7
Telangiectatic	0	5
Others	1	2
New pulmonary metastasis		
Yes	3	2
No	66	86
Chemotherapy regimen		
MTX, DDP and ADM	47	66
MTX, IFO, DDP, and ADM	19	17
MTX, IFO, DDP, and ADM	3	5

Age is presented as means ± standard deviations, with ranges in parentheses. MTX, methotrexate; DDP, cisplatin; ADM, doxorubicin; IFO, ifosfamide.

chemotherapy was performed by two experienced pathologists blinded to the CT images, using the criteria of Bacci *et al.* (18). The pGR was defined as the tumor necrosis percentages ≥90%; while non-pGR was defined as the tumor necrosis percentages <90% (19).

About three weeks after the last cycle of chemotherapy, patients underwent surgery. The type of surgery (amputation and limb salvage) was chosen depending on the comprehensive consideration involving location and extension of the tumor, and neurovascular bundle involvement. Simultaneous surgery for primary and metastatic lesions was performed for patients with metastatic lesions. Margins of the resections were defined according to the Enneking's criteria (20).

Radiomics analysis of tumor regions

Radiomics analysis of osteosarcoma tumor regions was conducted according to the following three steps. The tumor was delineated by one experienced orthopedic surgeon using the ITK-SNAP software, reviewed and modified if necessary, by the second experienced orthopedic surgeon and an experienced radiologist (21). The two orthopedic surgeons and the radiologist were blinded to the pathological response to chemotherapy. Radiomics features were extracted from the confirmed tumor regions on CT images following image pre-processing. Subsequent classification was implemented to differentiate patients with pGR and non-pGR by using a pathological response classifier.

A total of 491 image features were extracted from each tumor region. Prior to feature extraction, image pre-processing was performed including voxel resampling and intensity normalization. All CT images were resampled to the same voxel size of $0.5×0.5×5$ mm³. Voxel intensity within the tumor region was discretized to a range of 64 intensities. The feature pool was categorized as intensity statistics features, geometry features, texture features, and wavelet features. The intensity statistics features measured the voxel intensities statistical distribution within the tumor region. The geometry features described the 3D shape features of the tumor region. The texture features
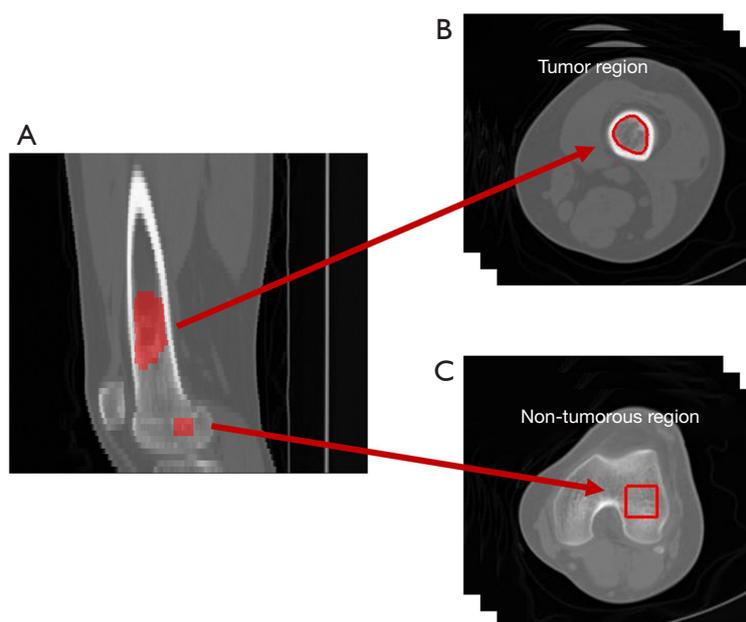
Figure 2 Illustration of the regions of interest segmentation on CT images. Radiomics analysis was performed in regions of interest from both tumor region (top, B) and non-tumorous bone region (bottom, C). Images of (B) and (C) were in the transverse plane. Image (A) was in the frontal plane.

describing the spatial intensity correlation and distributions of the voxels within the tumor region were calculated based the gray-level co-occurrence matrix (GLCM), the gray-level size-zone matrix (GLSZM), the gray-level run-length matrix (GLRLM), and the neighborhood gray-tone-difference matrix (NGTDM). The feature extraction process was conducted based on open-source Radiomics packages in MATLAB 2017b (MathWorks, Natick, MA, USA) (22-24). Details in feature extraction were described comprehensively in the Supplementary Material (Appendix 1).

Radiomics analysis of non-tumorous regions

For non-tumorous bone feature extraction, regions of interest of 30×30×3 voxels (15×15×15 mm³, cubic area) were delineated in the region of cancellous bone structures where the tumor was located. The compact bone region was not included in the regions of interest given that the compact bone was too thin to obtain feature information. Since imaging features were extracted based on the non-tumorous bone region with a cubic size, the geometry features were not included. The feature pool for the non-tumorous bone region included three categories, involving intensity statistics

features, texture features, and wavelet features. Radiomics analysis of non-tumorous regions was performed on the extracted 484 non-tumorous bone features. An example of regions of interest for the tumor region and non-tumorous bone region was shown in *Figure 2*.

Leave-one-out cross validation (LOOCV) and feature selection

To reduce selection bias within the training dataset, feature selection and temporary classifier construction were performed concurrently using the LOOCV method (17,25,26). The temporary classifier was constructed using features selected within each cross-validation iteration process. Note that in each iteration process, we further divided the training dataset into two separate sub-datasets: *N-1* patients as the temporary training set and one patient as the temporary validation set. *N* was the overall number of patients in the training dataset in our study. The temporary classifier was developed using the temporary training set and validated using the temporary validation set. These training and validation procedures were repeated *N* times. Model performance was measured using the area under the receiver operating characteristic curve (AUC) value

averaged during the LOOCV process (mean ± standard deviation).

During each LOOCV iteration process, the maximum relevance minimum redundancy (mRMR) feature selection approach was used to select a compact feature set (27,28). The mRMR method aimed to generate a feature set which had the maximum relevance to the actual label and the minimum redundancy to other features. Based on the rule of thumb, the number of predictors should remain less than 1/10 of the case number in the group (29,30). Thus, to prevent over fitting, the potential feature number was limited to 12 when deriving a temporary classifier using the multivariable logistic regression algorithm. The optimal temporary classifier was determined from 11 classifiers with feature number ranging from 2 to 12. The optimal temporary classifier was identified with the AUC value.

Furthermore, to investigate individual importance of the tumor and non-tumorous bone features, features used in the temporary classifier for more than 50% times were recorded during the LOOCV process. To investigate the additional value of non-tumorous bone features in assessing treatment response to chemotherapy, the temporary classifiers were constructed using tumor features alone, non-tumorous bone features alone, and combined features, respectively.

Classifier construction and independent validation

We developed a classifier to predict individual pathological response status using the tumor features that were selected more than 50% of the LOOCV process with multivariable logistic regression algorithm. For comparison, we further developed a pathological response classifier using the combined features selected more than 50% of the LOOCV process. Then the proposed classifiers were validated using an independent validation dataset.

Reproducibility validation of pathological response classifier

To assess the reproducibility of the region selection of the bone region and the proposed classifier, we re-delineated the non-tumorous bone ROIs of patients from the independent validation dataset. The re-delineation process was conducted by the same orthopedic surgeon following the same procedure. Furthermore, we extracted the non-tumorous bone features based on the re-delineated ROIs. Then features were normalized using values calculated based on the training dataset. We validated the proposed pathological response classifiers using re-delineated feature

dataset. The performance difference between the re-delineated validation dataset and the independent validation dataset was measured using the Delong test.

Classifier construction integrating the clinical features with imaging features

We developed a clinical-imaging classifier using both imaging features and clinical features with multivariable logistic regression algorithm. The clinical features included age, sex, location of primary tumor, tumor stage, pathologic subtype, and new pulmonary metastasis. For the tumor features based clinical-imaging classifier, clinical and radiomics features were combined with those selected more than 50% of the LOOCV process. For the combined imaging features based clinical-imaging classifier, we included the clinical features and combined imaging features selected more than 50% of the LOOCV process. The proposed classifiers were validated using the independent validation dataset.

Statistical analysis

Receiver operating characteristic (ROC) analysis was used to evaluate the performance of the proposed classifiers, with AUC value to differentiate patients with pGR and non-pGR (31). The maximum AUC value was 1.0, indicating a perfect prediction. The minimum value of 0.5 indicated no predictive power. The Mann-Whitney U-test was used to assess performance difference in AUC values. The Delong test was used to access the deference in ROC curves. A threshold P value less than 0.05 was defined as the significant difference in the two-tailed analysis. The Holm-Bonferroni method was conducted to correct for multiple comparisons. All statistical analysis was performed with R software (Version 3.4.1, www.R-project.org) and MedCalc (Version 15.2.2, www.medcalc.org). The mRMR algorithm was conducted using the "mRMRe" package in R software. The ROC analysis was conducted using the "pROC" package.

Results

Radiomics analysis of tumor features

As shown in *Table 2*, an AUC of 0.7799±0.0044 (range, 0.7649–0.7913) was obtained for the pathological response classifier developed using the tumor features alone

Table 2 Classification performance with tumor volume features, bone structure features, and combined features in the assessment of neoadjuvant chemotherapy response

Radiomics features	AUC	AUC difference	P value
Tumor volume features	0.7799±0.0044	–	–
Bone structure features	0.7219±0.0121	–	–
Combined features	0.8207±0.0043	–	–
Tumor volume features *vs.* bone structure features	–	0.0580±0.0130	<0.0001
Tumor volume features *vs.* combined features	–	0.0408±0.0046	<0.0001
Bone structure features *vs.* combined features	–	0.0987±0.0121	<0.0001

Data were presented with means ± standard deviations. P value was calculated using the Mann-Whitney U-test. The Holm-Bonferroni method was conducted to correct for multiple comparisons. AUC, area under the receiver operating characteristic curve.

during the LOOCV process. ROC analysis indicated that individual tumor feature showed inferior classification performance in pathological response prediction, with an AUC of 0.5494±0.0339 (range, 0.5008–0.6304). In the LOOCV process, we found eleven tumor features were selected more than 50% times, including *LZE-tumor*, *SRE-LLL-tumor*, *corrp-LLH-tumor*, *SZE-LLH-tumor*, *Complexity-LHL-tumor*, *corrm-LHH-tumor*, *SRHGE-LHH-tumor*, *svarh-HLH-tumor*, *corrp-HHL-tumor*, *cshad-HHH-tumor*, *SZHGE-HHH-tumor*. The feature *SRE-LLL-tumor* indicated short run emphasis value extracted based on GLRLM from the wavelet decomposed image using low-pass, low-pass and low-pass filters. The full names for all other abbreviations were provided in Table S1, Supplementary Material (Appendix 1).

Radiomics analysis non-tumorous features

When we performed radiomics analysis of non-tumorous bone regions, an AUC of 0.7219±0.0121 (range, 0.6916–0.7492) was obtained using the pathological response classifier developed based on the non-tumorous bone features only during the LOOCV process. Individual non-tumorous feature showed poor classification performance in pathological response prediction with an AUC of 0.5341±0.0288 (range, 0.5002–0.6270) using ROC analysis. Two non-tumorous bone features were selected more than 50% of the time during the LOOCV process, including *Entropy-bone*, *corrm-LLH-bone*.

Radiomics analysis of combined features

The radiomics analysis of combined features showed higher AUC values (0.8207±0.0042, range, 0.8099 to

0.8347) than the analysis of tumor features alone during the LOOCV process. The AUC difference between the classifiers based on combined features and tumor features was 0.0408±0.0046, with a range from 0.0259 to 0.0538. Statistical significance was observed for the AUC difference with P value <0.0001 using Mann-Whitney U-test.

Ten image features were selected more than 50% during the LOOCV process, including seven tumor features (*SRE-LLL-tumor*, *SZE-LLH-tumor*, *SRHGE-LHH-tumor*, *svarh-HLH-tumor*, *corrp-HHL-tumor*, *cshad-HHH-tumor*, *SZHGE-HHH-tumor*) and three non-tumorous features (*corrm-LLH-bone*, *SRLGE-HLH-bone*, *RLV-HHH-bone*). Of these features, seven tumor features were all selected more than 50% times for the tumor features-based classifier; a bone feature of *corrm-LLH-bone* was selected more than 50% times for the non-tumorous bone features-based classifier.

Classifier construction and independent validation

For the pathological response classifier developed using the eleven tumor features, AUC was 0.766 [95% confidence interval (CI), 0.679–0.840; significance level, P value <0.0001] for the training dataset, and 0.766 (95% CI, 0.606–0.885; significance level, P value =0.0004) for the independent validation dataset. By contrast, the combined features-based classifier showed better performance than the tumor features-based classifier in both training and validation datasets (training dataset: 0.791, 95% CI, 0.706–0.860; significance level, P value <0.0001; validation dataset: 0.816, 95% CI, 0.662–0.920; significance level, P value <0.0001). The ROC curves of both tumor features and combined features-based classifiers in the training dataset and validation dataset were showed in *Figure 3*.
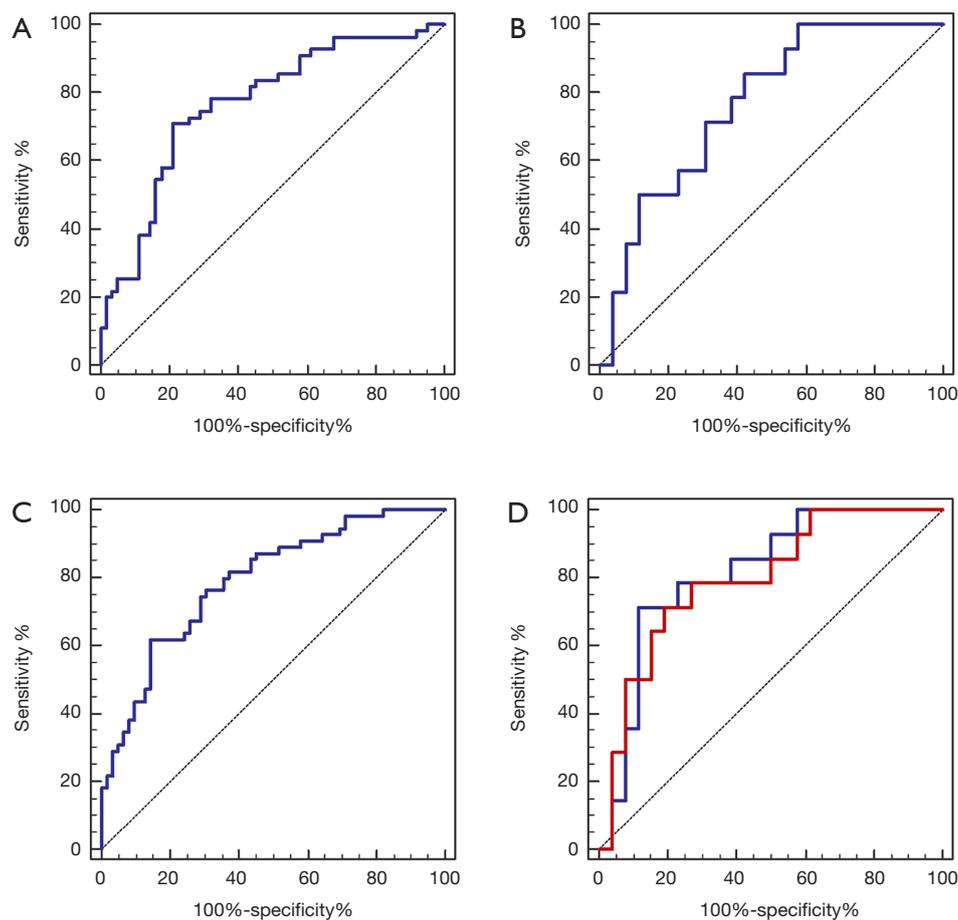
Figure 3 The ROC curves of the tumor features-based classifier in the training dataset (A) and validation dataset (B). The ROC curves of the combined features-based classifier in the training dataset (C), validation dataset (D, blue line), and re-delineation validation dataset (D, red line). ROC, receiver operating characteristic.

Reproducibility validation of pathological response classifier

In the reproducibility validation test of the proposed classifiers using re-delineated ROIs, a promising performance was achieved with an AUC of 0.797 (95% CI, 0.640–0.907; significance level, P value =0.0001). We also observed a non-significant performance difference between the re-delineation dataset and the independent validation dataset (P value =0.4553). The ROC curves of the combined features-based classifier in the re-delineation validation dataset were showed in *Figure 3*.

Classifier construction integrating the clinical features and imaging features

Integration of clinical features with tumor feature achieved

an AUC of 0.814 (95% CI, 0.731–0.880; significance level, P value <0.0001) for the training dataset and 0.728 (95% CI, 0.564–0.856; significance level, P value =0.0080) for the validation dataset. Integrating the clinical features with the combined features, we observed an AUC of 0.820 (95% CI, 0.738–0.885; significance level, P value <0.0001) for the training dataset and 0.811 (95% CI, 0.656 to 0.917; significance level, P value =0.0001) for the validation dataset. The ROC curves of these two clinical-imaging classifiers in both training dataset and validation dataset were provided in *Figure 4*. The difference between AUC values was 0.0059 (95% CI, –0.0311–0.0428) for the training dataset and 0.0827 (95% CI, –0.0248–0.190) for the validation dataset between the clinical/tumor classifier *vs*. clinical/combined classifier.
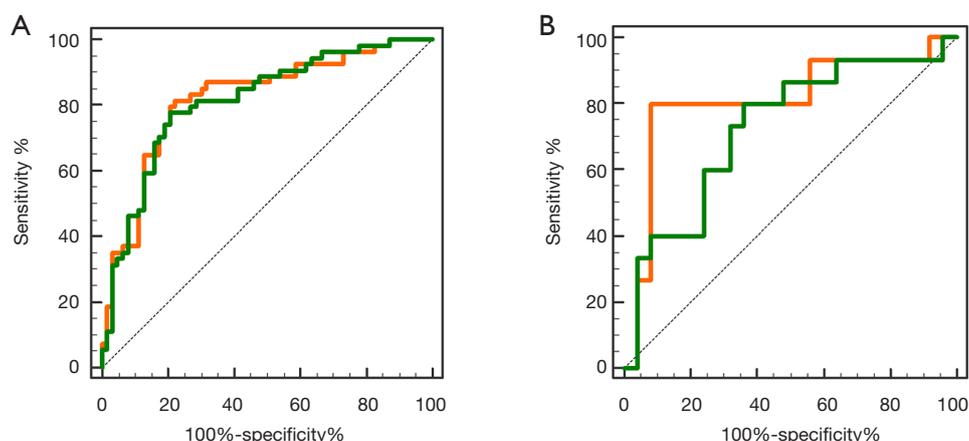
Figure 4 The ROC curves of two clinical-imaging classifiers in training dataset (A) and validation dataset (B). The orange line indicated the ROC curves for the clinical-combined features based classifier. The green line indicated the ROC curves for the clinical-tumor features based classifier. ROC, receiver operating characteristic.

Discussion

In this study, we compared the performance of pathological response prediction to chemotherapy with the use of tumor features alone and the combination of tumor and non-tumorous bone features. Results showed statistically significant improvement in pathological response prediction performance with the combined features.

Literature showed a wide spectrum of accuracy in radiomics studies for pathological response prediction to chemotherapy in patients with osteosarcoma, probably due to the difference in data sets and model construction algorithms. Song *et al*. achieved a high AUC of 0.918 with a relatively small data set of thirty-five patients (10), whereas Davis *et al*. reported an AUC of 0.807 using a data set of thirty-four patients (9). However, both these studies only analyzed image features based on intratumoral heterogeneity. To the best of our knowledge, the present study is the first study to predict pathological response by combining tumor features and non-tumorous bone features.

For the features selected more than 50% in the LOOCV process, ten out of eleven tumor-based features, one out of two non-tumorous bone feature, and ten out of ten combined features were all wavelet features. This indicated the vital role of wavelet image features in the pathological response prediction. This observation was consistent with previous studies which also used wavelet-based features in the proposed radiomics models (32-34). The wavelet transformation splitting images into different subimages with different frequency components may further explore

the spatial heterogeneity at multiple scales within the regions of interest (33).

The results of this study suggested that non-tumorous bone features could strengthen pathological response prediction to chemotherapy in osteosarcoma. Previous studies have demonstrated that the development of osteosarcoma was related to bone microenvironment and complex bone cell-tumor interactions (15,35,36). The improvement of pathological response prediction accuracy using combined features supports that tumor cell and nearby normal bone matrix interaction is important for treatment response.

For the non-tumorous bone features, only two features were selected from the LOOCV process, including *Entropy-bone*, *corrm-LLH-bone*. For the radiomics analysis of combined features, three bone features (*corrm-LLH-bone*, *SRLGE-HLH-bone*, *RLV-HHH-bone*) were retained. The feature of *corrm-LLH-bone* was selected from both processes, indicating stable predictive power in this non-tumorous bone feature. This feature measured pixel correlation to its neighbor over the whole region of interest. This indicated that the neoadjuvant chemotherapy to the osteosarcoma might cause local influence to non-tumorous regions, resulting in changes in local texture.

A few limitations to be noted: (I) the study was a retrospective study with data from single institution and small sample size, which limited the use of LOOCV method and small independent validation set. (II) Only CT images were studied, yet MR imaging might be more commonly

used for patients with HOS nowadays. (III) Patients included in this study underwent three-drug regimen, whereas two regiments of three-drug regimen and two-drug regimen (cisplatin and doxorubicin) may be a preferred treatment by NCCN. The validation using patients treated by different regimens was valuable. (IV) Clinical and imaging features were studied, yet a variety of other relevant signatures, such as biological features, molecular features, immune features, etc. might affect results of the current study. Nevertheless, this is the first pilot study of using quantitative features of tumor and non-tumorous bone regions based on medical images for predicting pathological response to chemotherapy in osteosarcoma. Prospective and multicentric studies are underway involving more imaging modalities, regimens and features to further validate the findings of this study.

Conclusions

Radiomics analysis of combined tumor and non-tumorous bone features showed improved performance of pathological response prediction to chemotherapy in HOS compared to that of tumor features alone. Moreover, the proposed classifier had the potential to predict the pathological response to chemotherapy for HOS patients.

Acknowledgments

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/qims-20-681). The authors have no conflicts of interest to declare.

Ethical Statement: This study was approved by the Institutional Research Ethics Board of our institution. The requirement of informed consent was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International

References

1. Durnali A, Alkis N, Cangur S, Yukruk FA, Inal A, Tokluoglu S, Seker MM, Bal O, Akman T, Inanc M, Isikdogan A, Demirci A, Helvaci K, Oksuzoglu B. Prognostic factors for teenage and adult patients with high-grade osteosarcoma: an analysis of 240 patients. Med Oncol 2013;30:624.

2. Zaikova O, Sundby Hall K, Styring E, Eriksson M, Trovik CS, Bergh P, Bjerkehagen B, Skorpil M, Weedon-Fekjaer H, Bauer HC. Referral patterns, treatment and outcome of high-grade malignant bone sarcoma in Scandinavia-- SSG Central Register 25 years' experience. J Surg Oncol 2015;112:853-60.

3. Bacci G, Longhi A, Fagioli F, Briccoli A, Versari M, Picci P. Adjuvant and neoadjuvant chemotherapy for osteosarcoma of the extremities: 27 year experience at Rizzoli Institute, Italy. Eur J Cancer 2005;41:2836-45.

4. Bielack SS, Kempf-Bielack B, Delling G, Exner GU, Flege S, Helmke K, Kotz R, Salzer-Kuntschik M, Werner M, Winkelmann W, Zoubek A, Jurgens H, Winkler K. Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. J Clin Oncol 2002;20:776-90.

5. Davis AM, Bell RS, Goodwin PJ. Prognostic factors in osteosarcoma: a critical review. J Clin Oncol 1994;12:423-31.

6. Cheon GJ, Kim MS, Lee JA, Lee SY, Cho WH, Song WS, Koh JS, Yoo JY, Oh DH, Shin DS, Jeon DG. Prediction Model of Chemotherapy Response in Osteosarcoma by F-18-FDG PET and MRI. J Nucl Med 2009;50:1435-40.

7. Coffin CM, Lowichik A, Zhou H. Treatment effects in pediatric soft tissue and bone tumors - Practical considerations for the pathologist. Am J Clin Pathol 2005;123:75-90.

8. Kim S, Shin J, Kim DY, Choi GH, Kim MJ, Choi JY. Radiomics on Gadoxetic Acid-Enhanced Magnetic Resonance Imaging for Prediction of Postoperative Early and Late Recurrence of Single Hepatocellular Carcinoma. Clin Cancer Res 2019;25:3847-55.

9. Davis JC, Daw NC, Navid F, Billups CA, Wu J, Bahrami A, Jenkins JJ, Snyder SE, Reddick WE, Santana VM, McCarville MB, Guo J, Shulkin BL. (18)F-FDG Uptake During Early Adjuvant Chemotherapy Predicts Histologic Response in Pediatric and Young Adult Patients with Osteosarcoma. J Nucl Med 2018;59:25-30.

10. Song H, Jiao Y, Wei W, Ren X, Shen C, Qiu Z, Yang Q, Wang Q, Luo QY. Can pretreatment (18)F-FDG PET tumor texture features predict the outcomes of osteosarcoma treated by neoadjuvant chemotherapy? Eur Radiol 2019;29:3945-54.

11. Wu Y, Xu L, Yang P, Lin N, Huang X, Pan W, Li H, Lin P, Li B, Bunpetch V, Luo C, Jiang Y, Yang D, Huang M, Niu T, Ye Z. Survival Prediction in High-grade Osteosarcoma Using Radiomics of Diagnostic Computed Tomography. EBioMedicine 2018;34:27-34.

12. Lin P, Yang PF, Chen S, Shao YY, Xu L, Wu Y, Teng W, Zhou XZ, Li BH, Luo C, Xu LM, Huang M, Niu TY, Ye ZM. A Delta-radiomics model for preoperative evaluation of Neoadjuvant chemotherapy response in high-grade osteosarcoma. Cancer Imaging 2020;20:7.

13. Korkaya H, Liu S, Wicha MS. Breast cancer stem cells, cytokine networks, and the tumor microenvironment. J Clin Invest 2011;121:3804-9.

14. Weigelt B, Bissell MJ. Unraveling the microenvironmental influences on the normal mammary gland and breast cancer. Semin Cancer Biol 2008;18:311-21.

15. Alfranca A, Martinez-Cruzado L, Tornin J, Abarrategi A, Amaral T, de Alava E, Menendez P, Garcia-Castro J, Rodriguez R. Bone microenvironment signals in osteosarcoma development. Cell Mol Life Sci 2015;72:3097-113.

16. Sala E, Solomon L, Warren R, McCann J, Duffy S, Luben R, Day N. Size, node status and grade of breast tumours: association with mammographic parenchymal patterns. Eur Radiol 2000;10:157-61.

17. Li H, Mendel KR, Lan L, Sheth D, Giger ML. Digital Mammography in Breast Cancer: Additive Value of Radiomics of Breast Parenchyma. Radiology 2019;291:15-20.

18. Bacci G, Bertoni F, Longhi A, Ferrari S, Forni C, Biagini R, Bacchini P, Donati D, Manfrini M, Bernini G, Lari S. Neoadjuvant chemotherapy for high-grade central osteosarcoma of the extremity. Histologic response to preoperative chemotherapy correlates with histologic subtype of the tumor. Cancer 2003;97:3068-75.

19. Huvos AG. Bone tumors: diagnosis, treatment and prognosis. 1987.

20. Enneking WF, Spanier SS, Goodman MA. A system for the surgical staging of musculoskeletal sarcoma. Clin Orthop Relat Res 1980;(153):106-20.

21. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 2006;31:1116-28.

22. Vallieres M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Phys Med Biol 2015;60:5471-96.

23. Zhang R, Xu L, Wen X, Zhang JH, Yang PF, Zhang LX, Xue X, Wang XL, Huang Q, Guo CG, Shi YJ, Niu TY, Chen F. A nomogram based on bi-regional radiomics features from multimodal magnetic resonance imaging for preoperative prediction of microvascular invasion in hepatocellular carcinoma. Quant Imaging Med Surg 2019;9:1503-15.

24. Li Y, Tan G, Vangel M, Hall J, Cai W. Influence of feature calculating parameters on the reproducibility of CT radiomic features: a thoracic phantom study. Quant Imaging Med Surg 2020;10:1775-85.

25. Kohavi R. editor. A study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI; 1995: Montreal, Canada.

26. Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognition 2015;48:2839-46.

27. Yu L, Liu H. Efficient feature selection via analysis of relevance and redundancy. J Mach Learn Res 2004;5:1205-24.

28. Unler A, Murat A, Chinnam RB. mr2PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification. Inf Sci 2011;181:4625-41.

29. Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer; 2015.

30. Wei J, Yang G, Hao X, Gu D, Tan Y, Wang X, Dong D, Zhang S, Wang L, Zhang H, Tian J. A multi-sequence and habitat-based MRI radiomics signature for preoperative prediction of MGMT promoter methylation in astrocytomas with prognostic implication. Eur Radiol 2019;29:877-88.

31. Pesce LL, Metz CE. Reliable and computationally efficient maximum-likelihood estimation of "proper" binormal ROC curves. Acad Radiol 2007;14:814-29.

---

32. Xu L, Yang P, Liang W, Liu W, Wang W, Luo C, Wang J, Peng Z, Xing L, Huang M, Zheng S, Niu T. A radiomics approach based on support vector machine using MR images for preoperative lymph node status evaluation in intrahepatic cholangiocarcinoma. Theranostics 2019;9:5374-85.

33. Wu S, Zheng J, Li Y, Yu H, Shi S, Xie W, Liu H, Su Y, Huang J, Lin T. A Radiomics Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer. Clin Cancer Res 2017;23:6904-11.

34. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, Liang C, Tian J, Liang C. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. Radiology 2016;281:947-57.

35. Broadhead ML, Clark J, Myers DE, Dass CR, Choong PF. The molecular pathogenesis of osteosarcoma: a review. Sarcoma 2011;2011:959248.

36. Rubio R, Abarrategi A, Garcia-Castro J, Martinez-Cruzado L, Suarez C, Tornin J, Santos L, Astudillo A, Colmenero I, Mulero F, Rosu-Myles M, Menendez P, Rodriguez R. Bone environment is essential for osteosarcoma development from transformed mesenchymal stem cells. Stem Cells 2014;32:1136-48.

The feature pool in this study

In this study, we extracted 491 image features based on tumor volume, including four categories: intensity statistics features, geometry features, texture features, and wavelet features. We extracted 484 image features based on non-tumorous bone region, including three categories: intensity statistics features, texture features, and wavelet features. The feature extraction process was conducted based on open-source Radiomics packages in MATLAB 2017b (MathWorks, Natick, MA, USA).

Table S1 Full names and abbreviations of imaging features in this study

Full names	Abbreviations
Intensity statistics features	
Variance	–
Skewness	–
Kurtosis	–
Mean	–
Energy	–
Entropy	–
Uniformity	–
Geometry features	
Max diameter	MaD
Surface volume ratio	SVR
Compactness1	Cpt1
Compactness2	Cpt2
Surface area	SA
Spherical disproportion	SphDisp
Sphericity	–
Texture features (grey-level co-occurrence matrix)	
Autocorrelation	autoc
Contrast	contr
Correlation	corrm
Correlation2	corrp
Cluster prominence	cprom
Cluster shade	cshad
Dissimilarity	dissi
Energy	energ
Entropy	entro
Homogeneity	homom
Homogeneity2	homop
Maximum probability	maxpr

Table S1 (*continued*)

Table S1 (*continued*)

Full names	Abbreviations
Sum of squares variance	sosvh
Sum average	savgh
Sum variance	svarh
Sum entropy	senth
Difference variance	dvarh
Difference entropy	denth
Information measure of correlation1	inf1h
Information measure of correlation2	inf2h
Inverse difference normalized	indnc
Inverse difference moment normalized	idmnc
Texture features (grey-level run-length matrix)	
Short run emphasis	SRE
Long run emphasis	LRE
Grey-level non-uniformity	GLN
Run-length non-uniformity	RLN
Run percentage	RP
Low grey-level run emphasis	LGRE
High grey-level run emphasis	HGRE
Short run low grey-level emphasis	SRLGE
Short run high grey-level emphasis	SRHGE
Long run low grey-level emphasis	LRLGE
Long run high grey-level emphasis	LRHGE
Grey-level variance	GLV
Run-length variance	RLV
Texture features (grey-level size zone matrix)	
Small zone emphasis	SZE
Large zone emphasis	LZE
Grey-level non-uniformity	GLN
Zone-size non-uniformity	ZSN
Zone percentage	ZP
Low grey-level zone emphasis	LGZE
High grey-level zone emphasis	HGZE
Small zone low grey-level emphasis	SZLGE
Small zone high grey-level emphasis	SZHGE

Table S1 (*continued*)

Table S1 (*continued*)

Full names	Abbreviations
Large zone low grey-level emphasis	LZLGE
Large zone high grey-level emphasis	LZHGE
Grey-level variance	GLV
Zone-size variance	ZSV
Texture features (neighbourhood grey-tone difference matrix)	
Coarseness	–
Contrast	–
Busyness	–
Complexity	–
Strength	–

Wavelet features

The wavelet features were extracted based the imaged decomposed by undecimated wavelet transform. Two kinds of filter of high-pass and low-pass was used in undecimated wavelet transform. By using the filters along three directions (x-, y- and z-direction), we could obtain 8 decomposed images. Then, the textural features were calculated based on the decomposed images. For example, we used the abbreviations 'LLL' to represent the low pass, low pass, and low pass filters in three axes for the original image during the wavelet translation. The same descriptions were used for all other imaged after wavelet translation. The wavelet feature name was defined as the combination of feature name and wavelet filter name. A number of 424 features were obtained through undecimated wavelet transform.