Article title:

Computational Analysis of Magnetic Resonance Images Predicts the Osteosarcoma Chemoresponsiveness

Short running title: Prediction of Osteosarcoma Chemoresponsiveness

Djuričić Goran J, Rajković Nemanja, Milošević Nebojša, Sopta JP, Borić I, Dučić S, Apostolović Milan, Radulović Marko

Abstract

Aim: This study aimed to improve osteosarcoma chemoresponsiveness prediction by optimisation of computational analysis of magnetic resonance (MR) images. **Patients & Methods:** Our retrospective predictive model involved osteosarcoma patients with MRI scans performed before OsteoSa MAP neoadjuvant cytotoxic chemotherapy. **Results**: We found that several monofractal and multifractal algorithms were able to classify tumours according to their chemoresponsiveness. The predictive clues were defined as morphological complexity, homogeneity and fractality. The monofractal feature *CV for Λ′(G)* provided the best predictive association (AUC = 0.88; $P<0.001$), followed by *Y-INTforFDB*, r^2*forFDM* and *tumour circularity*. **Conclusions**: This is the first full-scale study to indicate that computational analysis of pretreatment MR images could provide imaging biomarkers for the classification of osteosarcoma according to their chemoresponsiveness.

Tweetable abstract: Fractal analysis of MRI scans was shown to predict the chemosensitivity of osteosarcoma. These findings may eventually lead to improved patient survival by enabling personalised cytotoxic chemotherapy prescription.

Keywords: Computational image analysis; medical image analysis; cancer; cytotoxic chemotherapy; tumor circularity; osteosarcoma; prediction; prognosis, fractal analysis, MRI.

Introduction

Osteosarcoma is a malignant tumour that mostly affects the long bones before 20 years of age. Patients with high-grade osteosarcoma undergo several cytotoxic chemotherapy cycles before surgical intervention. This treatment consists of induction MAP chemotherapy (methotrexate [M], doxorubicin [A], cisplatin [P]) followed by surgical tumour removal or amputation, depending on the local tumour invasiveness.

The fact that the past several decades brought only a minor improvement in osteosarcoma survival rates highlights the need for novel research directions. Cytotoxic chemotherapy is the main obstacle to prolongation of cancer survival because of its limited efficacy [1]. Furthermore, response to chemotherapy is inconsistent due to intratumoral and intertumoral heterogeneity. The shortfalls of existing cytotoxic therapies are difficult to overcome because novel and more efficient anti-cancer therapies are introduced very slowly. Another major drawback is our inability to assess chemosensitivity prior to treatment which leads to indiscriminate cytotoxic chemotherapy prescription and poor response of most tumours [2]. It is generally assumed that early prediction of chemotherapy response in osteosarcoma could lead to avoidance of ineffective treatments and the possible metastasis development in chemoresistant patients. Such patients should receive alternative options, such as experimental trial protocols or amputation [3]. Prediction of the chemotherapy response may thus improve survival by enabling precision treatments.

Functional magnetic resonance imaging (MRI) methods provide significant predictive performance in osteosarcoma but with a delay after the start of chemotherapy [4], while molecular markers [2] such as tumour hypoxia-inducible factor 1 [5] and P-glycoprotein [6] were reported to deliver early chemosensitivity prediction in osteosarcoma. Because such molecular markers' clinical usability

remains uncertain, the research increasingly focuses on the heterogeneity of tumour morphology as the tool for chemosensitivity classification [7]. Tumour macroscopic morphology acquired by MRI presents a rich source of heterogeneity information because it is shaped by the growth patterns of malignant cells and, ultimately, by the sum of molecular interactions within a tumour. Due to the complexity and irregularity of tumour morphology, its quantification is approached by computational analysis [8]. Fractal analysis has been proven useful in medicine for morphological quantification of irregular natural shapes. It is sensitive to size, texture, shape [9], morphological heterogeneity and morphological complexity of tumours [10].

Based on the pressing need to improve the performance of osteosarcoma chemosensitivity prediction, this study optimises the analytical conditions and compares the predictive value of pretreatment MRI by use of fractal analysis, first-order statistics and osteosarcoma ROI geometry.

Methods

Subjects

This noninterventional, retrospective and predictive study was approved by the Ethics committee of the School of Medicine at the University of Belgrade (#29/VI-4) and written informed consent was obtained from the subjects for this study. It is reported according to STROBE guidelines for cross-sectional studies. It conforms with The Code of Ethics of the World Medical Association (Declaration of Helsinki), printed in the British Medical Journal (18 July 1964) and its seventh revision in 2013. We received patient data by the pathology unit in a de-identified and recoded form without direct or indirect identifiers that could enable re-identification. In total, 292 consecutive patients with skeletal pain and tumour diagnosis were evaluated for eligibility. Of these, 71 were osteosarcoma, 49 Ewing sarcoma and 172 were other tumour types. Of the 71 osteosarcoma patients, four were excluded based on low-quality MRI or artefacts, while 13 were excluded because of the axial skeleton localisation in order to assemble a homogeneous long tubular bone group. Inclusion criteria comprised:

histopathologic diagnosis of primary osteosarcoma on tubular long bones and pre-chemotherapy MRI. Presentation of MRI artefacts, pathological fractures and missing data were the exclusion criteria. A chemotherapeutical response was evaluated based on Huvos grading by an expert pathologist with 22 years of experience. All patients received preoperative OsteoSa MAP therapeutic 3x3 weeks protocol. Thirty-five patients were male, aged 5-46 years (median=16) and 19 were female, 7-25 years (median=14). All patients were diagnosed and treated at the Institute for Oncology and Radiology and University Children's Hospital. The sample size calculation was based on a pilot study including 22 patients and required 54 patients with six positive cases for the alpha=0.20, beta=0.20 and AUC effect size of 0.22/0.78 (MedCalc Software, Ostend, Belgium). The actual average significant AUC was 0.22/0.78, for 54 patients, of which eight cases exerted the ≥95% chemotherapy-induced necrosis.

MR Imaging and patient treatment

T2-weighted fat-suppression MRI was performed by the 1.5-T Siemens Healthcare Magnetom Avanto Syngo MR B15 workstation by the standard clinical protocol [11], followed by an administration of preoperative OsteoSa MAP neoadjuvant chemotherapy (Doxorubicin/Adriamycin 75 mg/m2 + Cisplatin 100 mg/m2 + high-dose Methotrexate 12 g/m^2, in 3x3-week cycles. Surgical removal of a tumour or amputation was subsequently performed, depending on the local tumour invasiveness of critical structures such as blood vessels and joints. Quantification of tumour cell necrosis in response to MAP chemotherapy was done by a pathologist with 18 years of experience in the evaluation of chemotherapy-induced necrosis. Patients with a good histologic response (>90% necrosis) were treated with three additional cycles of MAP. In comparison, patients with <90% necrosis were administered with three additional cycles of MAP supplemented with 14 gr/m^2 high-dose Ifosfamide.

ROI demarcation

Images were exported from Kodak Carestream PACS Client Suite v10.2 in the TIFF 1920×1080 greyscale format and tumour areas manually cropped in Fiji/ImageJ [12] along the borders of individual tumours by the two staff radiologists with ten and sixteen years of experience in

musculoskeletal radiology (G.J.D and I.B., respectively). The observers were not made aware of the other observer's segmentation results.

The manual approach is still considered equal or more reliable in comparison to the automated segmentation [13]. ROIs were introduced in ImageJ software by the creation of ROI masks ("create mask" command), delineation of ROI ("create selection" command) and transfer of ROI selection to the original greyscale image ("restore selection" command) with grayscale tumour area on a white background (Figure 1a-c; 2a). Such ROI delineation was adequate for the calculation of ROI geometry features and pixel intensity statistics by the "measure" command, while fractal analysis by the FracLac plugin required the additional step of non-tumour background flooding with black RGB pixels by the "clear outside" command (Figure 2c). Such ROI demarcation for fractal analysis is based on the fact that FracLac only recognises grey pixels while colour RGB pixels are identified as background.

Feature extraction

ImageJ calculated three types of features:

1.) The first order pixel intensity statistics: *Mean intensity, Skewness, Kurtosis, Solidity, IntDen and RawIntDen.*

2.) Tumour geometry: *Area, Perimeter, Width, Height, Major, Minor, Angle, Circularity, Feret, FeretX, FeretY, FeretAngle, MinFeret, Aspect ratio (AR)* and *Roundness*.

3.) Monofractal: $\overline{FD_B}$ (average box-counting dimension for 12 different starting grid positions), *SD for* FD_B (SD of FD_B for 12 different starting grid positions), FD_{Bmin} (minimal box-counting dimension over 12 different starting grid positions) , FD_{Bmax}, *FD_B with highest* r^2 (FD_B value with the highest r^2 over the FD_B calculated for 12 different starting grid positions, where r^2 refers to the regression line of the box count/size log plot), *SE for* FD_B (SE for box fractal dimension calculated for 12 different grid positions), *Y-INT for* FD_B (Y-intercept of the regression line of the box count/size log plot), $\overline{FD_M}$ (mass fractal dimension averaged over 12 grid positions), *SD for* FD_M (SD for mass fractal dimension), FD_{Mmin}, FD_{Mmax}, FD_M *with highest* r^2, *SE for* FD_M, *Y-INT for* FD_M, $\overline{FD_\chi}$ *with highest* r^2, *SE for* $\overline{FD_\chi}$, *Y-INT for*

FD_{χ}^{-} lacunarity (Λ), Λ' (lacunarity averaged over 12 grid positions), Λ_{min}, Λ_{max}, *CV for* $\Lambda_{(G)}$ (coefficient of variation of $\Lambda_{(G)}$ over 12 grid positions), and *CV for* $\Lambda'_{(G)}$.

4.) Multifractal: α, $f(\alpha)$ and Dq. $F(\alpha)$ is the local fractal dimension corresponding to the local exponent α, while Dq is the generalised dimension. These parameters were distorted by 200 Q values, from -10 to +10 in 0.1 increments to produce multifractal spectra which indicate how the images behave at each distortion. Excel formulas were further used to calculate 16 additional multifractal features from multifractal spectra $f(\alpha)$ *vs* α and Dq vs q: *Dqmax, Dqmin,* $f(\alpha)max$, $f(\alpha)min$, $\Delta f(\alpha)$, $\alpha f(\alpha)max$, $\alpha f(\alpha)min$, $Dq(q=0)$, $Dq(q=1)$ $Dq(q=2)$, $f(\alpha)(q=0)$, $f(\alpha)(q=1)$, $f(\alpha)(q=2)$, αmin, αmax *and* $\Delta\alpha$. Please find the detailed explanation for features that are most relevant for this study in the next subsection. We used the "differential box counting" method for fractal analysis of greyscale images in FracLac plugin version 2016apr for ImageJ, as previously explained in full detail [8]. The grid consisted of boxes sized 5 - 630 pixels in linear 3-pixel increments and it was moved to 12 different positions within an image or ROI. Figure 2d shows an example of a box counting grid within osteosarcoma ROI.

Detailed explanation of fractal features which reached predictive significance

Dqmax was calculated from the Dq vs q multifractal spectrum. Fractal dimension (FD) and lacunarity (Λ) of greyscale images were calculated based on the difference between the maximum and minimum pixel intensities $\delta I_{i,j,\varepsilon}$ within each individual box, at each box size ε. The transformation by adding 1 to the actual calculated pixel intensity difference was done to prevent zero values in later calculations (Equation 1). *Y-INT for* FD_B is the Y-intercept (Equation 3) of the regression line for the box FD (Equation 2) where S stands for the log of box size, C = log of box count, n = number of box sizes and m = slope of the regression line for the box count/size log plot. Monofractal $r^2 for FD_M$ feature is the determination coefficient of the regression line in Equation 4.

$$I\varepsilon = \sum [1 + \delta I_{i,j,\varepsilon}] \quad (1)$$

$$FD_{Bgrey} = \lim_{\varepsilon \to 0} \left(\frac{\ln (I\varepsilon)}{\ln (1/\varepsilon)} \right) = \textit{slope of the regression line} \quad (2)$$

$$\text{Y-INT} = \frac{\sum C - m\sum S}{n} \quad (3)$$

$$FD_{Mgrey} = \lim_{\varepsilon \to 0} \left(\frac{\ln(\bar{I}\varepsilon)}{\ln(1/\varepsilon)} \right) = \textit{slope of the regression line} \quad (4)$$

Grid positioning influences the result of a box count. For this reason, grids were placed in 12 different positions. *SD for FDM* was calculated as the standard deviation of FD_M values for each of the 12 different grid positions (Equation 4). Parameters *CV of* $\Lambda_{(G)}$ and *CV of* $\Lambda'_{(G)}$ were calculated as coefficients of variation of $\Lambda_{(G)}$ and $\Lambda'_{(G)}$ obtained for each of the 12 different grid positions. $\Lambda_{(G)}$ is the average on each grid position of lacunarities $\lambda_{(\varepsilon)}$ calculated for all box sizes (Equation 5). $\Lambda'_{(G)}$ is the slope of the λ vs. box size regression line [$\ln\lambda_{(\varepsilon)+1}$ vs $\ln_{(\varepsilon)}$] for each grid position (Equation 6). It thus reflects the coupling of λ with scale. Coefficient of variation of any feature over different grid positions reflects the heterogeneity of an image.

$$\lambda_{(\varepsilon)} = (\text{CV})^2 = (\sigma/\mu)^2 \text{ of differences in pixel intensities for each box size } \delta I = (I_{max} - I_{min} + 1) \quad (5)$$

$$\Lambda'(G) = \lim_{\varepsilon \to 0} \left(\frac{\ln(\lambda(\varepsilon)+1)}{\ln(\varepsilon)} \right) = \textit{slope of the regression line} \quad (6)$$

Inter- and intraobserver agreement

Interobserver agreement of ROI manual segmentation was evaluated by comparison of areas delineated by the two radiologists mentioned above (G.J.D and I.B.). Intraobserver reproducibility was estimated by the two separate segmentations performed by the same observer. The similarity between segmentations was estimated by the Dice similarity coefficient (DSC) which measures a spatial overlap of segmentations by considering the ROI intersection in images A and B, divided by the sum of ROI areas in A and B (Equation 7).

$$\text{DSC} = \frac{2\,|A \cap B|}{|A| + |B|} \quad (7)$$

DSC calculation was performed by analysis of binarized images homogeneously flooded with black pixels in DSCImageCalc v1.2a software [14]. ROIs were subsequently homogeneously flood-filled with black pixels Binary images were produced by binarization of the greyscale (8-bit) images that

were used above for feature extraction. Binarization was performed by the following commands in Fiji/ImageJ software: setThreshold(0, 254); //print(i, mean) and run("Make Binary", "slice"). The ROI area of each image was subsequently flood-filled with black pixels by the Fiji/ImageJ commands: run("Analyze Particles...", "size=100-Infinity show=Masks clear include in situ"); run("Maximum...", "radius=70"); run("Minimum...", "radius=70").

Statistical analysis

Data were analysed in IBM SPSS software package v24 (IBM Corporation) and Stata/MP 13.0 (StataCorp, College Station). Values of the features mentioned in the previous subsection were averaged among the two images available for each of the three anatomical directions for each patient. The prognostic evaluation was performed by receiver operating characteristic (ROC) analysis and binary logistic regression, with the binary histologic response to chemotherapy as the outcome event. Areas under the ROC curves (AUCs) were calculated as a quantitative measure of discrimination efficiency, based on continuous feature values. Discrimination is the capability to stratify chemosensitive and chemoresistant tumours. AUC=0.5 represents chance discrimination, while AUC=0 and AUC=1 designate perfect discrimination. We performed AUC calculation by use of continuous feature values, while the binary logistic regression considered the values categorised by an optimal cutpoint. The bootstrap random resampling technique was applied for bias correction [15]. This procedure tests model stability and reliability by estimating the bias and modification of the original AUC confidence intervals (95% CIs) and *P*-values, as previously explained in detail [15]. The advantage of bootstrap over the split-sample cross-validation is that the entire dataset is used for model development. $P \leq 0.05$ was used as the significance threshold. Univariate and multivariate binary logistic regression were also supplemented by the bootstrap random resampling technique with 10,000 bootstrap resamples (SPSS v24).

Results

The predictive model for chemotherapy response

The model included 54 osteosarcoma patients diagnosed and treated by the National Sarcoma Multidisciplinary Team during the 5-years 2010-2014. Thirty-two osteosarcoma were in the femur, 14 in the tibia, four in the fibula and four in the clavicula. Two expert radiologists selected the two most representative images with the largest tumour area in each of the three acquired MRI slice orientations (Figure 1, A-C), a total of six images per patient. The mean intraobserver DSC was 0.95, with a standard deviation (SD) of ±0.04 (maximum: 0.99; minimum: 0.84). In the comparison between Observer-1 and Observer-2, the mean interobserver DSC was 0.93, with a standard deviation (SD) of ±0.04 (maximum: 0.99; minimum: 0.78).

To establish the predictive consistency of image analysis features, analysis was performed in all three anatomical directions: horizontal (Figure 1A) which divides a body into upper-lower parts, frontal (front-back, (Figure 1B) and sagittal (left-right, Figure 1C).

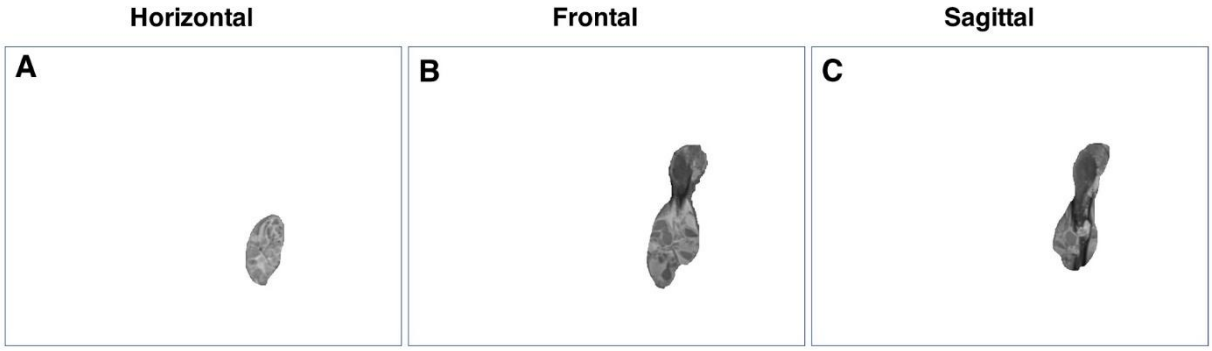

Figure 1. Examples of osteosarcoma MR images recorded in all three anatomical directions. (**A**) Horizontal, (**B**) frontal and (**C**) sagittal. Two characteristic images were analysed per each MRI slice orientation.

Analyses were implemented without (Figure 2A, B) and with (Figure 2C, D) tumour ROI delineation. Without ROI, fractal box counting was performed across entire images (Figure 2B), while with ROI, box counting was limited to tumour areas (Figure 2D). The effect of irregular ROI boundaries on the box-counting procedure was minimised by filling the background with colour RGB pixels which are ignored by the performed greyscale fractal analysis (Figure 2C, D). The calculated feature values were assessed for their predictive performance by use of the chemotherapy-induced tumour cell necrosis as the endpoint event (Tables 1, 2 and 3; Figure 3).

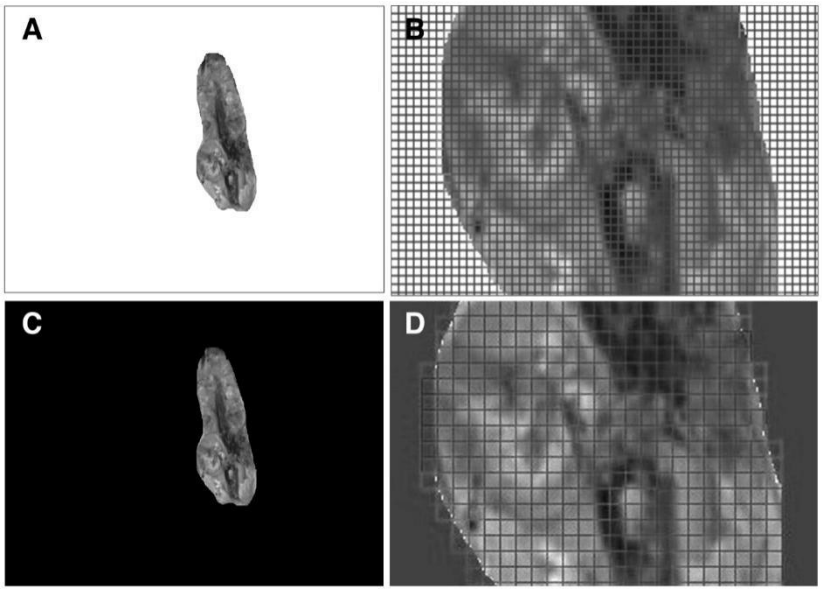

Figure 2. Demarcation of the regions of interest (ROIs) in osteosarcoma MR images. (**A**) Example of an image without strict ROI delineation. The tumour area was in the 8-bit greyscale format while the background was filled with white pixels. (**B**) Magnified image shows that fractal analysis boxes were distributed throughout an image in the absence of ROI delineation. (**C**) Example of the strict tumour ROI delineation achieved by filling the background outside of tumour area with black RGB pixels. The area within tumour ROI was again in the 8-bit greyscale format. (**D**) Magnified image shows how such ROI demarcation with RGB pixels limits the distribution of fractal analysis boxes to the tumour area. This is because FracLac software only analyses grey pixels, while colour RGB pixels are recognized as background.

Predictive performance of the demographic, MRI and clinicopathological parameters

Age, gender, tumour area, tumour volume, metastasis and pain intensity before diagnosis (Table 1) did not associate significantly with the 95% chemotherapy-induced tumour necrosis based on the receiver operating characteristic (ROC) curve analysis (Table 1). This analysis was performed on continuous values except for gender and metastasis occurrence, which are intrinsically categorical parameters (Table 1). By categorizing continuous values with an optimal threshold, the *tumour area* reached significant predictive performance (Table 1; AUC=0.27, $P<0.05$), indicating that increased tumour area predicts higher chemoresistance. A significant predictive association was achieved for 95% necrosis chemoresponsiveness but not for 80% and 90% (not shown).

Table 1. Demographic, MRI and clinicopathological features in the prediction of chemotherapy response.			
Classification		*Continuous*	*Categorized*
Gender	AUC	-	0.38
	95% CI	-	0.15-0.61
	P-value	-	0.30
Age	AUC	0.57	0.69
	95% CI	0.39-0.75	0.52-0.85
	P-value	0.56	0.10
Area	AUC	0.37	0.27
	95% CI	0.19-0.55	0.12-0.42
	P-value	0.27	**0.05**
Volume	AUC	0.30	0.31
	95% CI	0.12-0.48	0.12-0.50
	P-value	0.10	0.11
Metastasis	AUC	-	0.54
	95% CI	-	0.32-0.77
	P-value	-	0.71
Pain intensity	AUC	0.48	0.38
	95% CI	0.25-0.72	0.15-0.61
	P-value	0.90	0.30
Areas under the ROC curves (AUCs) were calculated by use of continuous or categorised feature values, with 95% histological tumour cell necrosis as the chemotherapy-response endpoint. Necrosis thresholds at 80% and 90% were also evaluated but not indicated in the table due to their inferior predictive performance. AUCs 95% confidence intervals (95% CI) were corrected by bootstrap. Significant p-values ≤0.05 are marked in bold style.			

Predictive performance of the tumour geometry, pixel intensity and fractal features

The obtained feature values were averaged for the anatomical directions and axes: horizontal (short bone axis), frontal and sagittal (long bone axis) and among all three directions (Table 2). The tumour geometry and fractal features showed the optimal predictive association with the 95% necrosis rate (Table 2) in comparison to the 80% and 90% necrosis (not shown). The chemotherapy-response achieving 95% necrosis was in between the 90% threshold that is widely considered as treatment success and the pathologic complete response characterised by 100% necrosis [16]. Table 2 presents image analysis features which fulfilled our inclusion criterion for consistent predictive performance. Under this criterion, the feature needs to show significant predictive value either in the horizontal plane (short bone axis) or in both frontal and sagittal planes (long bone axis) by their unmanipulated continuous values. This inclusion criterion was based on the expectation that frontal and sagittal directions provide similar predictive performance because they both belong to bones' long axis.

Table 2. Performance of the calculated features in the prediction of chemotherapy-response.						
Classification		**Horizontal orientation**	**Frontal orientation**	**Sagittal orientation**	**Frontal/Sagittal average**	**Front./Sag./Horiz. average**
	Without ROI					
Y-INT for FDB	AUC	0.31	0.24	0.27	0.26	0.24
	95% CI	0.09-0.52	0.09-0.40	0.09-0.44	0.08-0.43	0.08-0.40
	P-value	0.10	**0.03**	**0.05**	**0.04**	**0.03**
r² for FDM	AUC	0.34	0.23	0.27	0.21	0.24
	95% CI	0.14-0.54	0.05-0.41	0.10-0.43	0.05-0.38	0.28-0.41
	P-value	0.18	**0.02**	**0.05**	**0.02**	**0.03**
	With ROI					
Min. Feret	AUC	0.37	0.23	0.20	0.21	0.25
	95% CI	0.15-0.55	0.06-0.41	0.06-0.35	0.06-0.38	0.08-0.41
	P-value	0.21	**0.03**	**0.01**	**0.02**	**0.03**
Circularity	AUC	0.78	0.42	0.40	0.60	0.51
	95% CI	0.63-0.94	0.21-0.63	0.16-0.59	0.36-0.84	0.25-0.71
	P-value	**0.01**	0.50	0.39	0.42	0.94
SD for FDM	AUC	0.68	0.76	0.73	0.85	0.83
	95% CI	0.47-0.84	0.54-0.95	0.55-0.90	0.71-0.95	0.70-0.95
	P-value	0.08	**0.04**	**0.04**	**0.003**	**0.001**
CV for Λ(G)	AUC	0.68	0.78	0.75	0.76	0.77
	95% CI	0.48-0.80	0.56-0.99	0.59-0.89	0.57-0.91	0.66-0.93
	P-value	0.12	**0.03**	**0.03**	**0.04**	**0.01**
CV for Λ'(G)	AUC	0.74	0.77	0.88	0.88	0.86
	95% CI	0.55-0.96	0.63-0.92	0.79-0.98	0.77-1.0	0.71-0.97
	P-value	**0.04**	**0.03**	**0.001**	**0.001**	**0.001**
Dq_{max}	AUC	0.21	0.54	0.44	0.36	0.39
	95% CI	0.08-0.36	0.31-0.77	0.17-0.66	0.12-0.61	0.16-0.62
	P-value	**0.01**	0.73	0.59	0.22	0.36

Areas under the ROC curves (AUCs) were calculated by use of the 95% histological tumour cell necrosis as the chemotherapy-response endpoint and feature values calculated within tumour ROIs. Confidence intervals (95% CI) of AUCs were corrected by bootstrap. Significant *p*-values ≤0.05 are marked in bold style.
HR: hazard ratio; CI: confidence interval; *Y-INT for FDB*: Y-axis intersection of the regression line for box *FD*; r² *for FD*: determination coefficient for the regression line for mass *FD*; *Min. Feret*: minimum diameter; CV: coefficient of variation; *Λ(G)*: an average of *λ* values calculated for different box sizes; *Λ'(G)*: the slope of the regression line for *λ* vs box size; SD: standard deviation; FD_M: mass fractal dimension; Dq_{max}: maximum of the generalized fractal dimension.

We emphasise the use of unmanipulated continuous values because their categorisation often introduces a bias. By analysis of images without ROI delineation, two monofractal features: *Y-INT for DB* and *r² for DM* satisfied the inclusion criterion (not shown). However, analysis within strictly delineated tumour ROI delivered six features satisfying the inclusion criterion, including monofractal: *SD for FDM, CV for Λ(G)* and *CV for Λ'(G)*, multifractal: Dq_{max} and tumour geometry: *MinFeret* and *circularity* (Table 2). The areas under the ROC curve (AUC) below 0.5 for Y-INT for DB, *r² for DM*,

MinFeret and Dq_{max} indicated that high values of these features associated with a poor histologic chemotherapy response (Table 2). Yet, *SD for FDM, CV for Λ(G), CV for Λ'(G)* and *circularity* associated with a good histologic response by their AUC >0.5 (Table 2). The obtained predictive performance of the two independent features (Table 3) is illustrated in Figure 3A by showing how their raw numerical values separate good and poor responders and by AUC plots (Figure 3B).

Table 3. Identification of the predictively non-redundant features.

Classifier	P-value	HR	95%CI
Age	**0.001**	6.1×10^{15}	6.3×10^{7} - 2.0×10^{17}
CV for Λ'(G)	**0.001**	3.2×10^{14}	1.4×10^{7} - 2.1×10^{16}
Circularity	**0.001**	9.8×10^{9}	2.8×10^{7} - 6.4×10^{11}

Multivariate binary regression analysis was performed by the inclusion of categorized values of demographic, MRI, clinical and image analysis features obtained within tumour ROI. Categorization of the continuous feature values was performed by optimal cutpoints. The entry criterion was p≤0.10 achieved in ROC analysis while the stay criterion was p<0.05 in the stepwise multivariate binary logistic regression. P-values and 95% confidence intervals were corrected by bootstrap.
HR: hazard ratio; CI: confidence interval; CV: coefficient of variation; Λ'(G): the slope of the regression line for λ vs box size.

Identification of the predictively most relevant computational analysis features was accomplished by their adjustment for demographic, MRI and clinical parameters using the stepwise multivariate binary logistic regression analysis. *CV for Λ'(G)* and *circularity* were thereby identified as the predictively independent image analysis features (Table 3). Their predictive performance is directly presented in Figure 3A, without any statistical arbitration, by simply showing how increasing numerical values of each feature separate good- and poor-chemotherapy responders. The predictive performance is also illustrated by ROC curves (Figure 3B), whereby the curve of *CV for Λ'(G)* looks particularly convincing because it is far removed from the central no-discrimination line. A ROC curve is a graphical plot that illustrates the predictive ability of a binary classifier system, as its discrimination threshold changes. It is created by plotting the true positive rate or sensitivity, against the false positive rate (1-specificity) at various threshold settings.

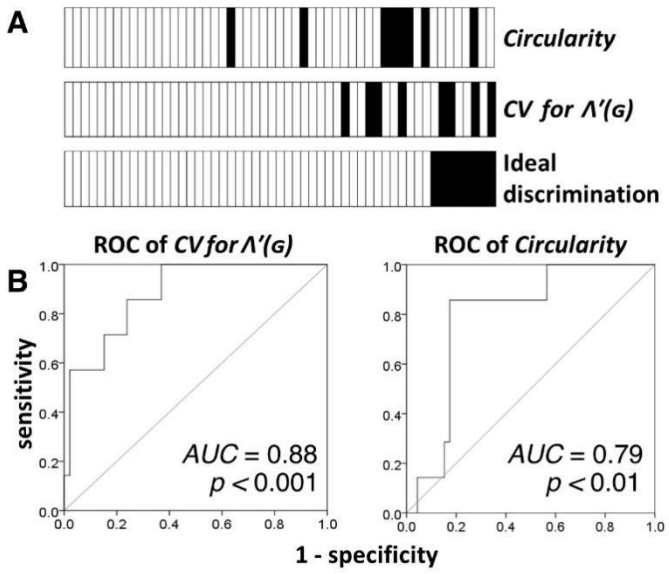

Figure 3. Illustration of the obtained predictive performance. (**A**) Continuous values of the predictively independent features are ordered sequentially and aligned with the actual chemotherapy response. Black fields indicate patients with at least 95% necrosis response to cytotoxic Osteosa MAP chemotherapy. Ideal discrimination where AUC=1.0 is shown for orientation. Values for *CV for Λ'(G)* and circularity were ordered from lowest (left) to highest (right). Thereby, increasing values of *CV for Λ'(G)* and circularity indicate a better response to chemotherapy. **(B)** Receiver operating characteristic curves further reveal efficiency in discrimination between chemosensitive and chemoresistant tumours by continuous feature values of *CV for Λ'(G)* and *circularity*. Values are for the sagittal plane for of *CV for Λ'(G)* while *circularity* values are for the horizontal plane. ROC analysis was performed by use of the 95% histological tumour cell necrosis as the chemotherapy-response endpoint and feature values calculated within tumour ROIs.
ROC: Receiver operating characteristic.

Discussion

Although tumour morphology is the obvious source of clues for chemotherapy response prediction [8], its investigation is still underexploited in radiological practice. We report that computational analysis of pretreatment T2-weighted MR images collected sufficient predictive clues to stratify osteosarcoma by their sensitivity to OsteoSa MAP chemotherapy. Such early chemosensitivity prediction is clinically highly relevant because it could improve survival by enabling individual adjustment of therapy protocols.

The main monofractal features, fractal dimension (*FD)* and lacunarity (*Λ*), did not provide any predictive value in this study. Instead, the predictive significance was reached by their derivatives: *SD for* FD_M, *CV for* $\Lambda_{(G)}$, *CV for* $\Lambda'_{(G)}$, *Y-INT for* D_B and r^2 *for* D_M. The common trait for the features: SD *for*

$FD_{M,}$ $CVfor\Lambda_{(G)}$ and $CVfor\Lambda'_{(G)}$ was that they offered predictive value only if calculated within ROI boundaries. These features reflect the degree of variation for *FD* and Λ among the 12 different box-counting grid positions. We assume that high variation of *FD* and Λ reflects an increased morphological heterogeneity of a tumour. Consequently, based on the obtained results, decreased morphological heterogeneity (or increased homogeneity) of osteosarcoma predicted higher chemoresistance. This finding was in line with the previous report showing that low Λ was predictive of higher osteosarcoma chemoresistance [17]. $CVfor\Lambda'_{(G)}$ achieved the best predictive performance by its AUC of 0.88, while the similar $CVfor\Lambda_{(G)}$ only achieved an AUC of 0.78. The difference between $\Lambda'_{(G)}$ and $\Lambda_{(G)}$ is that Λ represents lacunarity (λ) averaged among all box sizes, while $\Lambda'_{(G)}$ represents the dependency of lacunarity λ and the scale of measurement (box size). $\Lambda'_{(G)}$ is thus a more robust estimate of heterogeneity because it includes λ as the intrinsic measure of heterogeneity and its dependency on box size as another heterogeneity sensor ($\Lambda'_{(G)}$). Furthermore, the coefficient of variation of a parameter calculated for different grid positions also reflects the heterogeneity of an image. Taken together, the excellent predictive performance of $CVfor\Lambda'_{(G)}$ could be explained by its incorporation of the three measures of morphological heterogeneity.

Other predictive clues identified in this study include complexity, as measured by $Y\text{-}INTforD_B$ and fractality, based on r^2forD_M, whereby increased structural complexity and fractality predicted higher chemoresistance. It is important to note that fractality estimation by r^2forD_M refers to restricted self-similarity defined as unchanging complexity among all scales and not to the structural self-similarity as in mathematical fractals [10]. To our knowledge, the above fractal features have never been employed in the study of medical images for predictive or prognostic purposes. Therefore, although most studies focus on *FD* and Λ [7, 17, 18], our current results provide an example of how the depth of analysis benefits the clinical relevance of medical images.

Fractal features are very abstract due to their complex calculation. We therefore also calculated the simple tumour ROI geometry to find that increasing values of size descriptors *tumour area* and *min.*

feret (minimum diameter) significantly indicated higher chemoresistance. This was consistent with previous reports that increased tumour size prognosticated poor outcome for osteosarcoma [19]. Furthermore, higher tumour *circularity* indicated higher chemosensitivity in the current study. Shape biomarkers have not been previously investigated in osteosarcoma, possibly because growth and shape are by far more restricted for bone tumours than for tumours growing in soft tissues. However, shape descriptors have been rarely investigated even in soft tissue tumours, therefore, the association between tumour circularity and chemosensitivity has not been directly investigated in any tumour type. Higher tumour circularity was reported to associate with increased Ki67 staining in breast cancer [20], implying that the higher proliferation rate of circular tumours might explain their increased sensitivity to cytotoxic chemotherapy. Our result showing the predictive value of osteosarcoma circularity thus points to the need for closer investigation of the tumour shape descriptors in the prognosis of disease outcome and prediction of chemoresponsiveness.

Dq_{max} was the only multifractal feature providing predictively significant performance. Remarkably, its predictive significance was achieved in the short bone axis, while the six monofractal features consistently achieved their predictive performance only in the long bone axis. By its AUC of 0.21, Dq_{max} was inferior to the best performing monofractal $CVfor\Lambda'_{(G)}$. The fact that monofractal analysis provided better predictive performance than multifractal analysis was surprising because multifractals were particularly designed for investigation of unevenly distributed complexity in irregular natural forms by calculating both global *Dq* and local *f(α)* dimensions. On the other hand, the monofractal analysis only calculates the global fractal dimension (FD). The multifractal analysis also collects richer structural information by calculation of *Dq* and *f(α)* for each of the 200 values of the distortion factor *q* [9, 21]. It is therefore remarkable that of the 16 multifractal features calculated in this study, the global Dq_{max}, but not the local *f(α)* dimensions, achieved predictive significance.

We report the importance of tumour ROI delineation for the predictive performance of osteosarcoma MR image analysis. The advantage of analysis within ROI boundaries is in the strict focus on the relevant tumour area with the possible downside that irregular ROI boundaries influence the box

counting fractal analysis. We made an effort to minimise this edge effect by filling the background outside of ROI with RGB pixels which are not recognised by the used software and thus cannot influence the box counting procedure. Furthermore, we also avoided the edge effect by analysis across entire images and by calculation of the pixel statistics and tumour morphological features that are not affected by irregular tumour boundaries. Therefore, the finding that analysis within strict limits of tumour ROIs still offered the best predictive performance was interesting and suggested that the box counting error may be limited by the relatively small number of boxes at the ROI boundary. Nevertheless, the AUCs of 0.21 and 0.26 achieved in images without delineated ROI by $r^2 for D_M$ and *Y-INTfor*D_B features were inferior but remained in the good range (0.7-0.8 or 0.2-0.3). We explain this good predictive performance by the fact that images without ROI still emphasised the tumour area because the tissue structure surrounding osteosarcoma was removed by replacement with white pixels whose number inevitably depended on *tumour size.* Therefore, the values of $r^2 for D_M$ and *Y-INTfor*D_B might have been influenced not only by the grey pixels in the tumour and white pixels in the background but also by *tumour area,* the feature which showed predictive significance in this patient group. Besides such pronounced tolerance to the mode of ROI delineation, the robustness of the performed MR image analysis was further potentiated by the finding that chemosensitivity prediction could be obtained by different types of image analysis algorithms ranging from the sophisticated fractal analysis to the simple tumour geometry and pixel intensity statistics.

With its best predictive AUC of 0.88, the current study surpassed the early predictive discrimination (AUC=0.57, p=0.32) reported for the 18F-FDG PET and MRI functional imaging [22] and the later similar study reaching an AUC value of 0.82 [23], while fractal analysis of T2-weighted MRI reached the predictive AUC of 0.20 (an equivalent of 0.80) [17]. The achieved predictive performance is important because the reliability of discrimination between the responders and non-responders is essential for the introduction of the personalisation of cytotoxic chemotherapy in routine clinical practice. For orientation, AUC values span from 0.5 (chance discrimination) to 1.0 or 0.0 (perfect discrimination), while 0.6 or 0.4 are considered as fair, 0.7 or 0.3 as good, 0.8 or 0.2 as excellent and 0.9 or 0.1 as almost perfect discrimination.

Advantages of this study include an internal validation performed by bootstrap, which suggested that the model is generalisable. To improve the reliability of predictive estimates, we also performed separate analysis in the three anatomical directions, in addition to standard statistical analysis. This has enabled us to establish the predictive consistency for each calculated feature. Another advantage was that predictive evaluation based on ROC analysis considers continuous data values, without any need for data categorisation, which often introduces a bias. Benefits further include optimisation of ROI delineation, calculation and comparison of diverse features and analysis of greyscale instead of binary images. Moreover, the computational analysis of routinely collected MRI provides remarkable cost-effectiveness.

Limitations

Although the group size of 54 patients satisfied the sample size requirement and the patient group was highly homogenised, the patient number was nevertheless a limitation. This was due to the low osteosarcoma annual incidence rate of ~3 cases per million inhabitants. Another limitation of the predictive model used in this study was its retrospective design. Additional studies in external and extended patient groups are needed to further characterise the clinical validity of the reported approach in the prediction of osteosarcoma chemoresponsiveness.

Although the computational analysis enables an objective description of irregular tumour morphology, feature values are calculated in areas that are segmented subjectively, to the best knowledge of a radiologist, or by use of software that is also not certain to deliver an ideal ROI demarcation. Therefore, ROI segmentation is one of the most critical phases of the computational analysis of MR images. Our assessment of the intra- and interobserver manual segmentation reproducibility showed high reliability, in line with the previous single report of osteosarcoma segmentation [13]. However, ROI demarcation remains the limitation of this type of MRI analysis.

Conclusions

In this study, we provide improvement of osteosarcoma chemoresponsiveness prediction by optimisation of the computational MRI analysis. Reliable prediction of chemoresponsiveness is essential for gaining survival benefits by precision treatment of chemoresistant tumours with alternative options, such as experimental trial protocols or amputation. The proposed methodology outperforms the previously obtained prediction performance. We also quantified the benefit of tumour ROI demarcation and identified the clues predictive of osteosarcoma chemoresponsiveness as tumour circularity, diameter, complexity, homogeneity and fractality. The early prediction of the chemotherapy response has an application potential in routine clinical practice to optimise therapeutic protocols for each osteosarcoma patient.

- **Future perspective**

  Studies in larger and external patient cohorts are needed to confirm the findings pointing to the value of fractal MRI analysis in early prediction of osteosarcoma responsiveness to cytotoxic chemotherapy. We expect that this approach could be subsequently included in routine clinical practice to improve personalised treatments and thus avoid unnecessary treatment of chemoresistant patients.

Summary points
○ Computational analysis of tumour morphology extracted predictive information. ○ Monofractal features achieved the best prediction of osteosarcoma chemoresponsiveness. ○ Strict ROI demarcation was necessary for optimal predictive performance. ○ *CV for Λ′(G)* feature provided the best predictive performance by AUC=0.88. ○ The predictive clues included tumour's morphological complexity, homogeneity and fractality.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Harrison DJ, Geller DS, Gill JD, Lewis VO, Gorlick R. Current and future therapeutic approaches for osteosarcoma. *Expert review of anticancer therapy* 18(1), 39-50 (2018).
•• Provides a detailed overview of therapeutic approaches used in osteosarcoma. This information is important for understanding of how could early prediction of chemoresponsiveness provide the survival benefits.
2. Raimondi L, De Luca A, Costa V *et al*. Circulating biomarkers in osteosarcoma: new translational tools for diagnosis and treatment. *Oncotarget* 8(59), 100831-100851 (2017).
3. Papakonstantinou E, Stamatopoulos A, D IA *et al*. Limb-salvage surgery offers better five-year survival rate than amputation in patients with limb osteosarcoma treated with neoadjuvant chemotherapy. A systematic review and meta-analysis. *J Bone Oncol* 25 100319 (2020).
4. Song H, Jiao Y, Wei W *et al*. Can pretreatment (18)F-FDG PET tumor texture features predict the outcomes of osteosarcoma treated by neoadjuvant chemotherapy? *Eur. Radiol.* 29(7), 3945-3954 (2019).
5. Chen Y, Yang Y, Yuan Z, Wang C, Shi Y. Predicting chemosensitivity in osteosarcoma prior to chemotherapy: An investigational study of biomarkers with immunohistochemistry. *Oncology letters* 3(5), 1011-1016 (2012).
6. Zhao ZG, Ding F, Liu M, Ma DZ, Zheng CK, Kan WS. Association between P-glycoprotein expression and response to chemotherapy in patients with osteosarcoma: a systematic and meta-analysis. *Journal of cancer research and therapeutics* 10 Suppl C206-209 (2014).
7. Cusumano D, Dinapoli N, Boldrini L *et al*. Fractal-based radiomic approach to predict complete pathological response after chemo-radiotherapy in rectal cancer. *La Radiologia medica* 123(4), 286-295 (2018).
8. Rajkovic N, Kolarevic D, Kanjer K, Milosevic NT, Nikolic-Vukosavljevic D, Radulovic M. Comparison of Monofractal, Multifractal and gray level Co-occurrence matrix algorithms in analysis of Breast tumor microscopic images for prognosis of distant metastasis risk. *Biomedical microdevices* 18(5), 83 (2016).
• An exemplary study of the application of fractal geometry in the analysis of tumour histological specimens. The main difference in comparison to MRI is that tumour histology specimens fill an entire image and thus do not need ROI demarcation.
9. Duarte-Neto P, Stosic B, Stosic T, Lessa R, Milosevic MV, Stanley HE. Multifractal properties of a closed contour: a peek beyond the shape analysis. *PloS one* 9(12), e115262 (2014).
10. Grizzi F, Castello A, Qehajaj D, Russo C, Lopci E. The Complexity and Fractal Geometry of Nuclear Medicine Images. *Mol Imaging Biol* 21(3), 401-409 (2019).
•• Explains a theoretical background for the usefulness of fractal geometry in medical image analysis.
11. Bajpai J, Gamnagatti S, Kumar R *et al*. Role of MRI in osteosarcoma for evaluation and prediction of chemotherapy response: correlation with histological necrosis. *Pediatr. Radiol.* 41(4), 441-450 (2011).
12. Schindelin J, Arganda-Carreras I, Frise E *et al*. Fiji: an open-source platform for biological-image analysis. *Nature methods* 9(7), 676-682 (2012).
13. Dionisio FCF, Oliveira LS, Hernandes MA *et al*. Manual and semiautomatic segmentation of bone sarcomas on MRI have high similarity. *Braz. J. Med. Biol. Res.* 53(2), e8962 (2020).
•• This study reports a similar performance of the manual segmentation performed in our current study and computational segmentation.
14. Lawton T. Software for Determining Similarity Coefficients for the Analysis of Image Segmentations. *Journal of Open Research Software* 5(1), 28 (2017).
15. Efron B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* 7(1), 1-26 (1979).

16. Amit P, Malhotra A, Kumar R, Kumar L, Patro DK, Elangovan S. Evaluation of static and dynamic MRI for assessing response of bone sarcomas to preoperative chemotherapy: Correlation with histological necrosis. *Indian J Radiol Imaging* 25(3), 269-275 (2015).
•• This pilot study provides an initial investigation of the usefulness of MRI fractal analysis in the prediction of osteosarcoma chemoresponsiveness. It proposes an alternative to ROI demarcation of MRI, however, the current study provided a better predictive performance by use of standard ROI.
17. Djuricic GJ, Radulovic M, Sopta JP, Nikitovic M, Milosevic NT. Fractal and Gray Level Cooccurrence Matrix Computational Analysis of Primary Osteosarcoma Magnetic Resonance Images Predicts the Chemotherapy Response. *Frontiers in oncology* 7 246 (2017).
18. Ma J, Wang R, Yu Y, Xu X, Duan H, Yu N. Is fractal dimension a reliable imaging biomarker for the quantitative classification of an intervertebral disk? *Eur. Spine J.* doi:10.1007/s00586-020-06370-2 (2020).
19. Yang QK, Lai QY, Wang Y, Wang Y, Yao ZX, Zhang XJ. Establishment and validation of prognostic nomograms to predict overall survival and cancer-specific survival for patients with osteosarcoma. *Neoplasma* doi:10.4149/neo_2020_200617N639 (2020).
20. Jiang S, Hong YJ, Zhang F, Li YK. Computer-aided evaluation of the correlation between MRI morphology and immunohistochemical biomarkers or molecular subtypes in breast cancer. *Scientific reports* 7(1), 13818 (2017).
•• Provides insight into the prediction potential of textural analysis of (18)F-FDG scan images.
21. Karperien AL, Jelinek HF. Fractal, multifractal, and lacunarity analysis of microglia in tissue engineering. *Frontiers in bioengineering and biotechnology* 3 51 (2015).
22. Cheon GJ, Kim MS, Lee JA *et al*. Prediction model of chemotherapy response in osteosarcoma by 18F-FDG PET and MRI. *J. Nucl. Med.* 50(9), 1435-1440 (2009).
23. Jeong SY, Kim W, Byun BH *et al*. Prediction of Chemotherapy Response of Osteosarcoma Using Baseline (18)F-FDG Textural Features Machine Learning Approaches with PCA. *Contrast Media Mol Imaging* 2019 3515080 (2019).