

RESEARCH ARTICLE

Metastasis risk prediction model in osteosarcoma using metabolic imaging phenotypes: A multivariable radiomics model

Heesoon Sheen¹, Wook Kim¹, Byung Hyun Byun², Chang-Bae Kong³, Won Seok Song³, Wan Hyeong Cho³, Ilhan Lim², Sang Moo Lim², Sang-Keun Woo^{1*}

1 Division of applied RI, Korea Institute of Radiological and Medical Sciences (KIRAMS), Seoul, Republic of Korea, **2** Departments of Nuclear Medicine, Korea Institute of Radiological and Medical Sciences (KIRAMS), Seoul, Republic of Korea, **3** Department of Orthopedic Surgery, Korea Institute of Radiology and Medical Sciences, Seoul, Republic of Korea

* skwoo@kirams.re.kr



Abstract

OPEN ACCESS

Citation: Sheen H, Kim W, Byun BH, Kong C-B, Song WS, Cho WH, et al. (2019) Metastasis risk prediction model in osteosarcoma using metabolic imaging phenotypes: A multivariable radiomics model. PLoS ONE 14(11): e0225242. <https://doi.org/10.1371/journal.pone.0225242>

Editor: Giorgio Treglia, Ente Ospedaliero Cantonale, SWITZERLAND

Received: February 14, 2019

Accepted: October 21, 2019

Published: November 25, 2019

Copyright: © 2019 Sheen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by grants from the Korea Institute of Radiological and Medical Sciences (KIRAMS), funded by the Ministry of Science and ICT (MSIT), Republic of Korea (No.50473-2019) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (2019M2D2A1A02057204, 2019R1F1A1060190).

Background

Osteosarcoma (OS) is the most common primary bone tumor affecting humans and it has extreme heterogeneity. Despite modern therapy, it recurs in approximately 30–40% of patients initially diagnosed with no metastatic disease, with the long-term survival rates of patients with recurrent OS being generally 20%. Thus, early prediction of metastases in OS management plans is crucial for better-adapted treatments and survival rates. In this study, a radiomics model for metastasis risk prediction in OS was developed and evaluated using metabolic imaging phenotypes.

Methods and findings

The subjects were eighty-three patients with OS, and all were treated with surgery and chemotherapy for local control. All patients underwent a pretreatment ¹⁸F-FDG-PET scan. Forty-five features were extracted from the tumor region. The incorporation of features into multivariable models was performed using logistic regression. The multivariable modeling strategy involved cross validation in the following four steps leading to final prediction model construction: (1) feature set reduction and selection; (2) model coefficients computation through train and validation processing; and (3) prediction performance estimation. The multivariable logistic regression model was developed using two radiomics features, SUVmax and GLZLM-SZLGE. The trained and validated multivariable logistic model based on probability of endpoint ($P = 1 / (1 + \exp(-Z))$) was $Z = -1.23 + 1.53 * \text{SUVmax} + 1.68 * \text{GLZLM-SZLGE}$ with significant p-values (SUVmax: 0.0462 and GLZLM_SZLGE: 0.0154). The final multivariable logistic model achieved an area under the curve (AUC) receiver operating characteristics (ROC) curve of 0.80, a sensitivity of 0.66, and a specificity of 0.88 in cross validation.

Competing interests: The authors confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Conclusions

The SUVmax and GLZLM-SZLGE from metabolic imaging phenotypes are independent predictors of metastasis risk assessment. They show the association between ^{18}F -FDG-PET and metastatic colonization knowledge. The multivariable model developed using them could improve patient outcomes by allowing aggressive treatment in patients with high metastasis risk.

Introduction

Osteosarcoma (OS), the most common primary bone tumor affecting humans, develops in children and adolescents between the ages of ten and twenty-five years, and in persons with deficient bone remodeling [1,2]. The survival rate of these patients has significantly improved as a result of comprehensive management in the form of intensive chemotherapy and surgery [1,3]. However, despite modern therapy, OS recurs in approximately 30–40% of patients initially diagnosed with no metastatic disease [4,5]. The long-term survival rates of such patients is generally 20% [6,7]. Consequently, individual and early prediction of metastases in the plan of OS management is crucial as it could result in better-adapted treatments and survival rate.

Currently, the most important predictors are tumor grade, size, and presence of skip lesions or distant metastases, which are found from biopsies and microscopic examinations [1,8]. However, they do not always provide sufficient results because the risk of relapse may differ among patients with the same disease stage or chemotherapeutic response owing to the heterogeneity of the tumor population [4,8,9]. Within this context, imaging modalities and, more particularly, ^{18}F -FDG positron emission tomography (PET) are generally used for staging and monitoring various cancers. As its accumulation explains the characteristics of heterogeneity in tumors, it is increasingly being accepted as the prognostic value [10,11]. Moreover, the analysis of radiomics features from ^{18}F -FDG PET imaging has recently been considered as additional information for predicting tumor response to therapy [12] because it could theoretically provide further understanding of the hidden tumor biology as compared with maximum standard uptake value (SUVmax), mean SUV (SUV_{mean}), peak SUV (SUV_{peak}), metabolic tumor volume (MTV), and total lesion glycolysis (TLG).

The advent of personalized medicine has increased the need for the improvement of clinically feasible prediction models for treatment decision [13]. After radiomics features from ^{18}F -FDG PET imaging are decided by the relation to a given tumor outcome as prognostic factors, models combining those factors need to be developed in order to make better prediction performance. Multivariable models could further fully characterize intratumoral heterogeneity, even though univariate analysis has been deemed informative. Based on the above considerations, this study was conducted with the objective of developing a multivariable model based on optimal radiomics features from pretreatment ^{18}F -FDG PET imaging in order to assess metastasis risk at initial diagnosis. To achieve this main objective, firstly, the relevant features were selected from a large number of features by the statistical methods for radiomics model. Secondly, multivariable modeling strategies were formulated to develop texture-based models with optimal predictive and generalized properties. Finally, the final optimal model was evaluated using the test dataset. This developed radiomics model and the selected significant features are expected to ultimately encourage physicians to make better decisions for treatment and potentially improve survival rate.

Materials and methods

Patient cohort

A database of eighty-three consecutive patients was retrospectively retrieved between June 2006 and August 2010. The patient qualification requirements were as follows: newly diagnosed, histologically biopsy-proven primary intramedullary osteosarcoma; completion of neoadjuvant chemotherapy and surgery; less than two weeks between ^{18}F -FDG PET/CT scan and initiation of preoperative chemotherapy; and no history of previous treatment except for biopsy. Patients with metastatic osteosarcoma at the present time were excluded from the study. Metastases were either proven by biopsy or diagnosed by an expert physician from the following-up cross-sectional imaging results of ^{18}F -FDG PET/CT imaging, MR imaging, bone scanning, and x-ray for at least six months. Informed consent was waived by the ethics committee because of the retrospective nature of this study and the analysis using anonymous imaging and clinical data. This study was approved by the Institutional Review Board of KIR-AMS (IRB No.: K-1707-001-004 (eIRB NO.: 2017-04-004)). Patient characteristics and histologic features are described in Table 1.

Good historic response was 46.99% of the patients and poor historic response was 53.01% of the patients in neoadjuvant chemotherapy outcomes. Of the patients, 26.51% had metastasis (lung metastasis: 72.22%, bone metastasis: 27.78%) after completion of neoadjuvant chemotherapy and surgery in five years or seven years (one patient). The remaining 73.49% of patients were metastasis-free until a composite endpoint determined by the time from the date of diagnosis to the date of metastasis (Table 2).

All eighty-three qualified patients had performed pre-treatment ^{18}F -FDG PET/CT scans using a Biography6 PET/CT scanner (Siemens Medical Solutions, Erlangen, Germany) at Korean Cancer Center Hospital (KCCH). All patients were instructed to fast for at least six hours before the intravenous administration, with water intake permitted and encouraged.

Table 1. Patient characteristics.

Characteristics	Value
Sex, n (%)	
Female	23 (27.71%)
Male	60 (72.29%)
Age, n (%)	
years \leq 19	67 (80.72%)
years $>$ 19	16 (19.27%)
Location of primary tumor, n (%)	
Humerus	4 (4.82%)
Radius	2 (2.41%)
Femur	44 (53.01%)
Fibula	3 (3.61%)
Tibia	30 (36.14%)
AJCC stage, n (%)	
IIA	35 (42.2%)
IIB	48 (57.8%)
Pathologic subtype, n (%)	
OB (Osteoblastic)	60 (72.29%)
CB (Chondroblastic)	9 (10.84%)
FB (Fibroblastic)	7 (8.43%)
Others	7 (8.43%)

<https://doi.org/10.1371/journal.pone.0225242.t001>

Table 2. Characteristics of clinical outcomes.

Clinical outcomes	Total Number (n)	83 (100%)
Histologic response	Good	39 (46.99%)
Metastasis	Poor	44 (53.01%)
	No	61 (73.49%)
	Yes	22 (26.51%)
Histologic response and Metastasis	Good-MetaFree (Good Response and Metastasis Free)	34 (40.96%)
	Poor-MetaFree (Poor Response and Metastasis Free)	27 (32.53%)
	Good-Meta (Good Response and Metastasis)	5 (6.02%)
	Poor-Meta (Poor Response and Metastasis)	17 (20.48%)

¹⁸F-FDG PET/CT Imaging Data

<https://doi.org/10.1371/journal.pone.0225242.t002>

Prior to administration of ¹⁸F-FDG (injected dose: 7.4 MBq per kg of body weight), a blood glucose level of <7.2 mmol/L was confirmed. The CT scan (130 kVp, 30 mA, 0.6 sec/rotation, pitch 6) without contrast agent and 3D PET scan (16.2 cm axial field of view, 3.5 min/bed position) were obtained for the sites of tumors located in the extremities (the vertex to the upper thigh). PET images were reconstructed using the ordered-subsets expectation maximization (OSEM) algorithm (two iterations and eight subsets) with CT-based attenuation correction after normalization, correction for scatter, random, decay, and dead time, and smoothed using a 5 mm post Gaussian filter to control noise. The ¹⁸F-FDG PET slice thickness was 3.03 mm and its matrix size was 4.063 mm × 4.063 mm.

Tumor volume definition and features extraction

Tumor volumes were segmented and radiomics features in the defined tumors subsequently extracted using the Local Image Features Extraction (LIFEx) version 4.0 software package (<http://www.lifexsoft.org>) [14–16]. The tumor region was drawn using a semi-automated segmentation method with threshold SUV of 2.0 based on our previous report [9] in three-dimensional (3D) images. In the segmented tumors, SUVmax, SUVmean, SUVpeak, Metabolic Tumor Volume (MTV), and Total Lesion Glycolysis (TLG) and features from shape and histogram were calculated as the first-order features.

For texture feature calculation, the numbers of intensity levels were resampled using 64 discrete values between zero and 20 SUVs, corresponding to a sampling bin width of 0.3125 SUVs [14,16,17]. Spatial resampling was 4.1 mm (X-direction), 4.1 mm (Y-direction), and 2.5 mm (Z-direction) in Cartesian coordinates [14].

Texture features were assessed by four texture matrices: the co-occurrence matrix (CM), the gray-level run length matrix (GRLM), the gray-level zone length matrix (GZLM), and the neighborhood gray-level different matrix (NGLDM). The CM was calculated in 13 directions with one voxel distance relationship between neighboring voxels, and each texture feature calculated from this matrix was the average of the features over the 13 directions in space (X, Y, Z). The GRLM was also calculated for 13 directions via a similar method while the GZLM was computed directly in 3D. NGLDM was computed from the difference of gray-levels between one voxel and its 26 neighbors in 3D and each texture feature was calculated from this matrix.

Forty-five features were extracted from the analysis of the volumes inspected: five conventional features, five histogram features, three shape features, and thirty-one texture features (Table 3).

Table 3. Summary of the first-order, second-order, and high-order features index.

Order of extracted feature	Matrix	Index	Type
First order	Conventional features	<ul style="list-style-type: none"> • SUVmin (minimum SUV) • SUVmax • SUVpeak • SUVmean • TLG 	Global
	Histogram features	<ul style="list-style-type: none"> • Skewness • Kurtosis • Entropy_log10 • Entropy_log2 • Energy 	
	Shape features	<ul style="list-style-type: none"> • Sphericity • Compacity • Volume (MTV) 	
Second order	GLCM (Gray-Level Co-occurrence based on concurrence Matrix)	<ul style="list-style-type: none"> • Homogeneity • Energy • Correlation • Contrast • Entropy • Dissimilarity 	Regional
	GLRLM (Gray-Level Run Length based on voxel-alignment Matrix)	<ul style="list-style-type: none"> • SRE (short-run emphasis) • LRE (long-run emphasis) • LGRE (low grey-level run emphasis) • HGRE (high grey-level run emphasis) • SRLGE (short-run low grey-level emphasis) • SRHGE (short-run high grey-level emphasis) • LRLGE (long-run low grey-level) • LRHGE (long-run high grey-level emphasis) • GLNUr (grey-level non-uniformity for run) • RLNU (run-length non-uniformity) • RP (run percentage) 	Regional
High order	NGLDM (Neighborhood Gray-Level Different based on neighborhood intensity-difference Matrix)	<ul style="list-style-type: none"> • Coarseness • Contrast • Busyness 	Local
	GLZLM (Gray-Level Zone Length based on intensity-size-zone Matrix)	<ul style="list-style-type: none"> • SZE (short-zone emphasis) • LZE (long-zone emphasis) • LGZE (low grey-level zone emphasis) • HGZE (high grey-level zone emphasis) • SZLGE (short-zone low grey-level emphasis) • SZHGE (short-zone high grey-level emphasis) • LZLGE (long-zone low grey-level emphasis) • LZHGE (long-zone high grey-level emphasis) • GLNUz (grey-level nonuniformity for zone) • ZLNU (zone length non-uniformity) • ZP (zone percentage) 	Local

<https://doi.org/10.1371/journal.pone.0225242.t003>

Feature selection and evaluations

All statistical analyses were performed with RStudio software (version 1.1.456; RStudio, Inc., Boston, MA, United States) except for p-value of the overall model fit. Multivariable logistic regression models were developed integrating metastatic event and imaging features. The

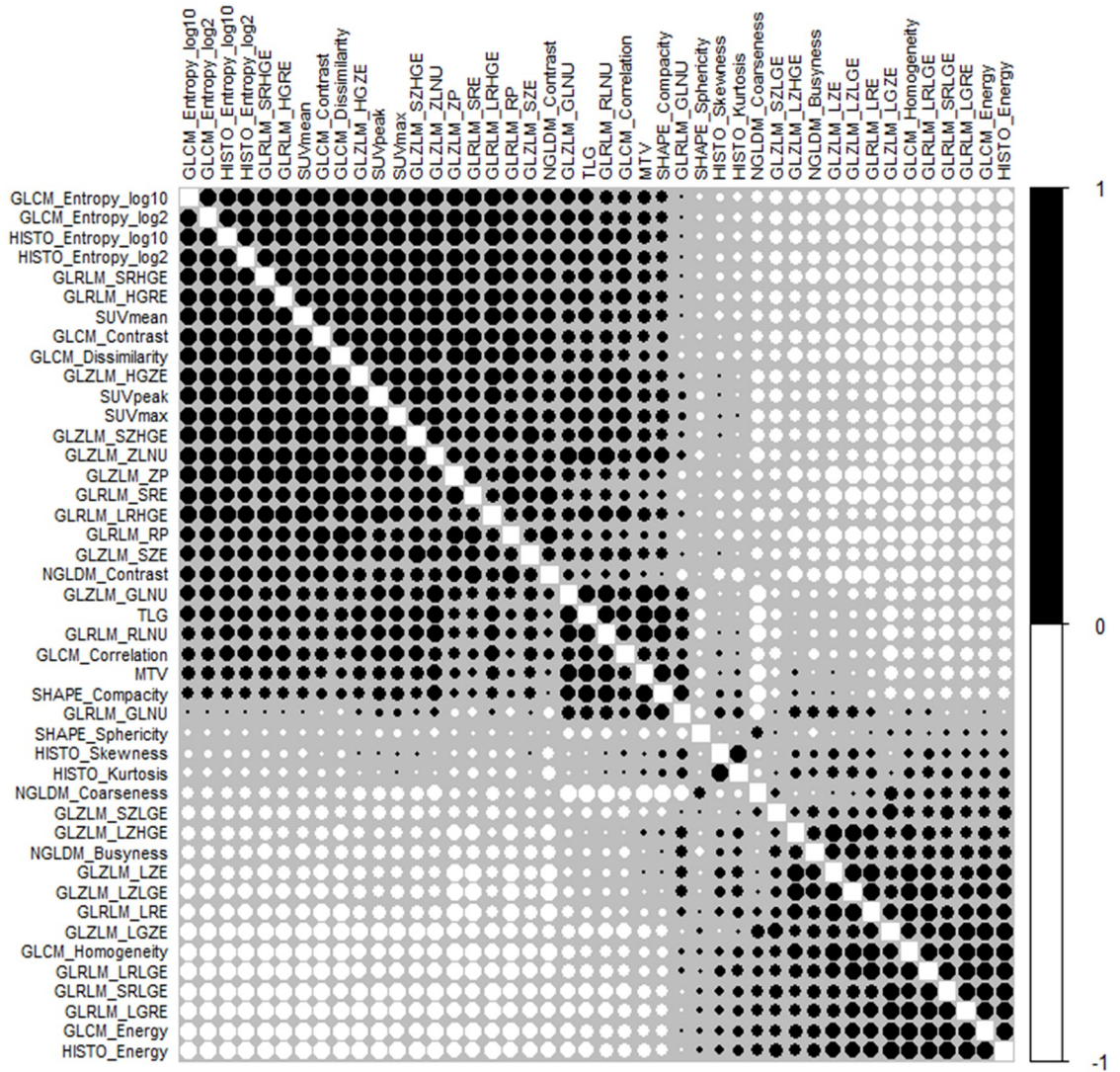


Fig 1. Spearman rank correlation of radiomics features in the training dataset. Forty-five features were extracted from tumor volumes in eighty-three patients. Across all tumors, the correlation of each feature with all other features was investigated via Spearman's rank correlation. The color and size of the circle indicate the degree of correlation. The final radiomics features were decided based on Spearman's correlation coefficient >0.9 and significant p-value >0.01 after Holm-Bonferroni correction.

<https://doi.org/10.1371/journal.pone.0225242.g001>

datasets were split into two random stratified cohorts: a training set (60%) and a test set (40%). Only the training set was used in the process to decide predictive and prognostic features. The 45 features were pre-selected, and then the bilateral correlation between these initial 45 features was evaluated with Spearman's rank correlation coefficient in order to estimate potential redundancy between the features [16,18]. The threshold of testing correlation coefficient was higher than 0.9 [19]. To correct for multiple test comparisons, the Holm-Bonferroni correction method was applied for all p' values: the significance level was lower than a value $p < p'/m$, where p' is 0.05 and m is the number of comparisons [20]. The calculated p-value was 0.01. The Spearman rank correlations for radiomics features are presented in Fig 1. The features from previous procedures were used to evaluate the relationship between them via the multiple backward stepwise elimination method based on Akaike's Information Criterion (AIC) [9] in order to find the subset of features in the dataset resulting in the lowest prediction error. The

determined features resulted from backward stepwise elimination and were assessed for overdispersion. The evaluation value was the ratio of residual deviance to residual degree of freedom (>1) and its p-value (>0.05). Subsequently, Friedman's ANOVA test and odds-ratio were used to estimate their significances. They were also investigated for variable importance (varImp) and multicollinearity using variance inflation factor (VIF)—for which a value over the limit of four would indicate a problematic amount of collinearity. The logistic regression model was constructed again using the last decided features, and subsequently evaluated using the Hosmer-Lemeshow goodness of fit test. The p-value of the overall model fit was computed using MedCalc Statistical Software (version 18.10; MedCalc Software bvba, Ostend, Belgium).

Model development and performance

The ratio of training set to test set was 60% to 40%. The training dataset (metastasis case: 27.45% and no metastasis: 72.55%) was used for training and validation and the test dataset (metastasis case: 25.00% and no metastasis: 75.00%) was utilized only to assess its prediction. In the 10-fold cross-validation method, the training dataset was further split into a training set (90% of the data) and a validate set (10% of the data) randomly. One fold was reserved for validation, and the other nine folds were used to train the model. Subsequently, they were used to predict the target variable in the validating data. This process was repeated ten times, with the performance of each model in predicting the reserved set being tracked using a performance metric such as accuracy. The validated prediction model was applied to the test dataset that had not been used in the previous steps, and then the predicted values were compared to the actual values. Its outputs were evaluated in terms of area under curve (AUC), sensitivity, specificity, accuracy, and precision computed based on receiver-operating-characteristic (ROC) curves and confusion matrix.

Predictive Tendency of Neoadjuvant Chemotherapy Prognosis and Metastasis

To analyze predictive tendency of neoadjuvant chemotherapy prognosis and metastasis, the radiomics features used for the radiomics model were investigated in four patient groups: Good-MetaFree (Good historic response and Metastasis Free), Poor-MetaFree (Poor historic response and Metastasis Free), Good-Meta (Good historic response and Metastasis), and Poor-Meta (Poor historic response and Metastasis) (Table 2).

Results

Feature selection and evaluations

The ratio of the patient's data used for the training set and the test set was 60% to 40%. From 45 radiomics features, eight radiomics features were determined via Spearman's rank correlation coefficient >0.9 and significant p-value >0.01 after the Holm-Bonferroni correction. These radiomics features were also decided after the logistic regression predictive model with backward stepwise elimination method was applied at 57.76 AIC. SUVmax and GLZM-SZLGE were eventually decided (Table 4).

SUVmax and GLZM-SZLGE were used for the multivariable logistic regression model and then evaluated using ANOVA test, odds ratio, varImp, and VIF (Table 5). SUVmax and GLZM-SZLGE showed less than 0.05 p-value in ANOVA test, higher than 1.0 Odds ratio, and their VIF values were higher than 4.0. The varImp test value of GLZM-SZLGE was higher than that of SUVmax. For overdispersion check, the ratio of residual deviance to residual degree of freedom was 1.078, and p-value was 0.30. The p-value of Hosmer-Lemeshow goodness of fit

Table 4. Radiomics features decided by Spearman’s rank correlation and backward stepwise elimination for use in multivariable regression analysis.

Classification of matrix	Features selected by Spearman’s correlation	Features selected by(Spearman’s correlation+backward stepwise elimination)
Conventional Indices	<ul style="list-style-type: none"> • SUVmax • MTV 	<ul style="list-style-type: none"> • SUVmax
Indices based on intensity histogram	<ul style="list-style-type: none"> • Skewness 	
Indices based on shape	<ul style="list-style-type: none"> • Sphericity 	
GLCM based on gray level co-occurrence matrix	<ul style="list-style-type: none"> • Correlation 	
NGLDM based on gray level neighborhood matrix	<ul style="list-style-type: none"> • Contrast • Busyness 	
GLRLM based on gray level homogeneous run size matrix	<ul style="list-style-type: none"> • GLNU 	
GLZLM based on gray level homogeneous zone size matrix	<ul style="list-style-type: none"> • SZLGE 	<ul style="list-style-type: none"> • SZLGE

<https://doi.org/10.1371/journal.pone.0225242.t004>

Table 5. Evaluation results of two radiomics features for use in multivariable regression analysis. The determined radiomic features were evaluated using ANOVA test, Odds ratio, varImp, and VIF. The results show that they are valid.

Classification of matrix	Index	ANOVA test	Odds Ratio (95% CI)	varImp	VIF
Conventional Indices	SUVmax	0.047	4.64 (1.03 to 20.97)	1.99	3.98
GLZLM	SZLGE	0.009	5.34 (1.38 to 20.73)	2.42	3.98

<https://doi.org/10.1371/journal.pone.0225242.t005>

test was 0.5725. As a result, SUVmax and GLSZM-SZLGE were confirmed for the logistic regression model.

Model development and performance

The trained and validated multivariable logistic model based on probability of endpoint $P = 1 / (1 + \exp(-Z))$ was

$$Z = -1.23 + 1.53 * \text{SUVmax} + 1.68 * \text{GLZLM} - \text{SZLGE}$$

with significant p-values (SUVmax: 0.0462 and GLZLM-SZLGE: 0.0154). The developed predictive model was estimated using a test dataset that had not been used in previous procedures. Its AUC (Fig 2), accuracy, sensitivity, and specificity were 0.8, 0.88, 0.63, and 0.88 (Table 6).

Predictive Tendency of Neoadjuvant Chemotherapy Prognosis and Metastasis

The predictive tendency of historic response prognosis and metastasis was investigated using SUVmax and GLZLM-SZLGE (Fig 3). Both features were negative in the Good-MetaFree group and Poor-MetaFree group, while they were positive in the Poor-Meta group. In the Good-Meta group, SUVmax was negative, but GLZLM-SZLGE was positive.

Discussion

In this study, a multivariable logistic regression imaging model was developed for the prediction of future metastases development at the point of osteosarcoma diagnosis. The model consists of two radiomics features extracted from ¹⁸F-FDGPET. The selection of radiomics features comprises several suggested steps that remove the redundancy between radiomics features and its overfitting. In order to strengthen the clinical impact of the model, the images

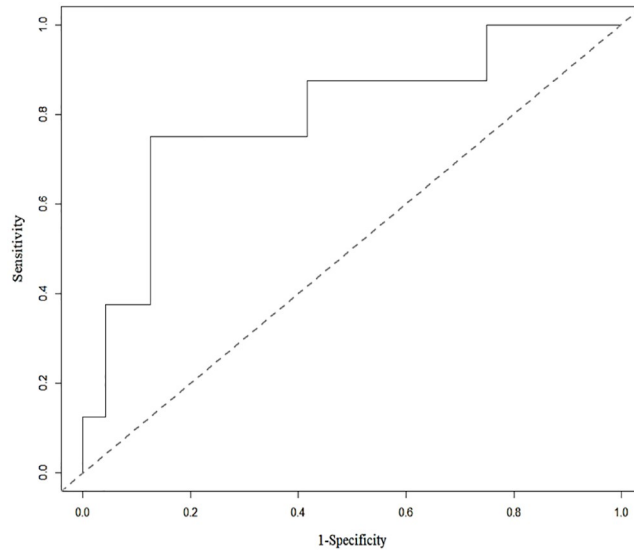


Fig 2. ROC curve for two radiomics features tested using the test dataset: SUVmax + GLZLM_SZLGE. The value shows that the classification based on texture analysis has good predictive value, as the area under the ROC curve is 0.80.

<https://doi.org/10.1371/journal.pone.0225242.g002>

used were acquired and reconstructed under routine clinical conditions. It is expected that the methodology used in this study could be generalized to other types of cancer and tumor outcomes.

The association of selected radiomics features during processing with metastases development in osteosarcoma is shown in Tables 3 and 4. The partial volume effect was not an issue in this study because all tumor volumes were higher than 8.39 cm³ (matrix size: 4.063 mm × 4.063 mm and slice thickness: 3.03 mm). Eight independent radiomics features resulted from Spearman rank correlation based on all matrix classifications. The results indicate that SUVmax was correlated to other SUVs (SUVmean and SUVpeak) and twenty-seven other texture radiomics features were associated with homogeneity, uniformity, inertia, randomness, dissimilarity, run-length matrix, and zone-size matrix of gray level. Further, it was related to global radiomics features and higher radiomics features, but it was not correlated to any feature from NGLCM. On the other hand, MTV was interrelated with six features related to TLG, shape, edge, and similarity of intensity values, and it was related to coarseness from NGLDM. Skewness (asymmetry of the histogram) based on histogram was correlated to its kurtosis (flatness of the histogram). The remaining six radiomics features (Shape-Sphericity, GLCM-Correlation, GLRLM-GLNU, NGLCM-Contrast, NGLCM-Busyness, and GLZLM-SZLGE) were not dependent on any other features. These features were investigated via varImp and the backward stepwise elimination method based on AIC.

Table 6. Estimation results of the predictive multivariable logistic model using the test dataset. AUV, accuracy, and specificity are good values while sensitivity is a fair value.

Predictive multivariable logistic model	probability of metastasis, $P = \frac{1}{1+e^{-Z}}$			
	$Z = -1.23 + 1.53 \cdot \text{SUVmax} + 1.68 \cdot \text{GLZLM_SZLGE}$			
Estimation item	AUC	Accuracy	Sensitivity	Specificity
Results	0.80	0.81	0.63	0.88

<https://doi.org/10.1371/journal.pone.0225242.t006>

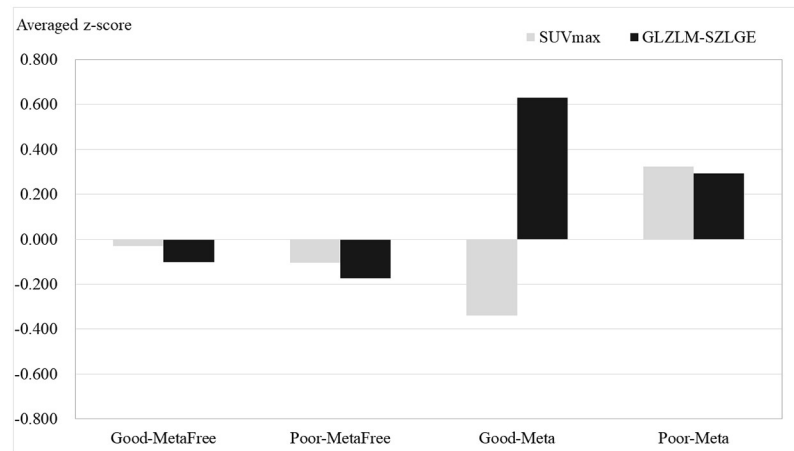


Fig 3. Histogram of the averaged z-scores of SUVmax and GLZM-SZLGE for the historic response prognosis and metastasis. Regardless of historic response outcome, GLZM-SZLGE presents positive value for the metastasis patient group, while it shows negative value for the metastasis free patient group. SUVmax has no related tendency.

<https://doi.org/10.1371/journal.pone.0225242.g003>

SUVmax and GLZM-SZLGE were finally decided as predictors for the multivariable logistic regression model, and were verified based on the results of Friedman's ANOVA test (p -value < 0.05), odds-ratio (> 1.0), multicollinearity ($VIF < 4.0$), and varImp. The developed multivariable logistic model with these predictors was suggested as a good fit model without overdispersion after a Hosmer-Lemeshow goodness of fit test and overdispersion evaluation. After the finalized multivariable logistic model was trained and validated using the 10-folds cross-validation method, it was applied to the test dataset. For this predictive model, good evaluation results were obtained: AUC, accuracy, and specificity were 0.8, 0.81, and 0.88, respectively. The sensitivity (0.63) was lower than other values, although it was fair. This occurred because of the insufficient number of metastasis cases—only eight. As a result, the multivariable logistic model built via linear combination of two verified radiomics features (SUVmax and GLZM-SZLGE) was confirmed with high predictive potential for metastases in osteosarcoma.

In this study, SUVmax and GLZM-SZLGE were finalized as the predictors for the radiomics model. SUVmax reflects the tumor aggressiveness and is an independent prognostic factor [4]. As a result, the positive correlation of SUVmax with metastases confirms that high-uptake in ^{18}F FDG-PET can play an important role in the characterization of aggressive tumors. These results are congruent with published papers that state that it could be a clinical indicator of sarcoma [20–24]. It was also used as an indicator of histopathologic response after neoadjuvant chemotherapy in extremity osteosarcoma in our previous report [25]. The other predictor, GLZM-SZLGE, increases when the short runs with low gray value are dominant. The positive correlation of GLZM-SZLGE indicates that many short runs of low gray value dominate in the fine-grained texture of the tumor. Consequently, higher SUVmax and GLZM-SZLGE result in higher possibility of tumor metastasis. They could imply tumor biological phenomenon based on the Warburg effect [26–28]. Tumor cells at sites close to blood vessels are relatively sufficiently oxygenated, whereas those located further away are hypoxic. This phenomenon is important for secondary tumor events as hypoxia can independently induce the epithelial–mesenchymal transition (EMT), a crucial step in cancer progression and metastasis via a number of mechanisms such as HIF-1 α expression in osteosarcoma and osteosarcoma cells [29–31]. Considering our results, SUVmax can be recognized as the most proliferative region by tumor cells residing closer to blood vessels. In contrast, GLAM-SZLAG can

be considered a hypoxic region by those residing further from it. It can be concluded that SUVmax and GLAM-SZLAG related with metastasis could provide information about tumor latent biology. Additionally, these two features could differentiate between two AJCC stages (IIA and IIB) with less than 0.05 p-value when the correlation between those stages and the AJCC stage commonly used for predicting outcome was considered (S1 Table).

Fig 3 shows the predictive tendency of historic response prognosis and metastasis. Whether the historic response outcome was good or poor, the averaged z-score of GLZLM-SZLGE was positive in the metastasis patient group. In contrast, it was negative in the metastasis free patient group. SUVmax did not show the relative tendency. Our results highlighted the Good-Meta group, which indicated good historic response of neoadjuvant chemotherapy but the primary tumor was metastasized in the endpoint. The conventional value, SUVmax, could not drive a high possibility of metastasis because it was too low. However, the radiomics feature, GLZLM-SZLGE, could predict the possibility of metastasis. We demonstrated that the texture-based radiomics model could supplement clinical prognostic factors for optimal prediction. However, multinomial logistic regression for the histological response and metastasis could not be achieved owing to insufficient patient cohorts.

In this study, the potential association between radiomics features and known metastatic colonization was presented, but biological information from the same patient was not used. However, this is required in order to assess the association and predictive power of the radiomics features for tumor cell information acquired from a patient. In addition, a larger patient cohort is essential in order to develop a more powerful model. Further, in order to facilitate clinical implementation of a texture-based decision-support system, harmonization and standardization on data and methods are also obviously required.

Conclusion

Texture biomarkers derived from ^{18}F -FDG images have been attracting attention as promising tools for characterizing spatial heterogeneity in order to predict tumor consequences at an early stage in humans. Verified selection of texture extraction and multivariable logistic modeling approaches have been proposed for the development of tumor outcome prediction models from a number of radiomics features in this study. The findings show that ^{18}F -FDG image features could perform as strong prognostic and predictive factors of metastasis in osteosarcoma and could provide observations about their inherent biology. The findings also indicate the priority of optimizing radiomics features to build up their predictive value and to determine the relation between features and biology. This model achieved good prediction performance evaluations in cross-validation, and its predictive properties were confirmed using an independent test dataset. However, further validation is required for stronger and more robust predictive properties. The established methodology can be applied to other tumors and could ultimately induce progress in treatment personalization and patient survival.

Supporting information

S1 Table. Average values of SUVmax and GLZLM_SZLGE and their p-value results from two-sample t-test in IIA and IIB.
(DOCX)

Author Contributions

Conceptualization: Heesoon Sheen, Sang-Keun Woo.

Data curation: Byung Hyun Byun, Chang-Bae Kong, Won Seok Song, Wan Hyeong Cho, Ilhan Lim, Sang Moo Lim.

Formal analysis: Heesoon Sheen.

Investigation: Heesoon Sheen.

Methodology: Heesoon Sheen, Wook Kim.

Software: Heesoon Sheen.

Supervision: Sang-Keun Woo.

Validation: Heesoon Sheen.

Visualization: Heesoon Sheen.

Writing – original draft: Heesoon Sheen.

Writing – review & editing: Byung Hyun Byun, Sang-Keun Woo.

References

1. Lindsey BA, Markel JE, Kleinerman ES. Osteosarcoma Overview. *Rheumatol Ther. Springer Healthcare* 2017; 4:25–43.
2. Durfee RA, Mohammed M, Luu HH. Review of osteosarcoma and current management. *Rheumatol Ther. Springer Healthcare* 2016; 3:221–43.
3. Bailly C, Leforestier R, Campion L, Thebaud E, Moreau A, Kraeber-Bodere F, et al. Prognostic value of FDG-PET indices for the assessment of histological response to neoadjuvant chemotherapy and outcome in pediatric patients with Ewing sarcoma and osteosarcoma. Tsokos M, editor. *PLoS One* 2017 12:e0183841 <https://doi.org/10.1371/journal.pone.0183841> PMID: 28841702
4. Analytics G, Nailon WH, Rai P, Srinivasan G, Shobha G, Thibault G, et al. Role of SUVmax obtained by 18F-FDG PET/CT in patients with a solitary pancreatic lesion: Predicting malignant potential and proliferation. *Eur J Nucl Med Mol Imaging* 2015; 53:1149–54.
5. Whelan JS, Jinks RC, McTiernan A, Sydes MR, Hook JM, Trani L, et al. Survival from high-grade localised extremity osteosarcoma: Combined results and prognostic factors from three European osteosarcoma intergroup randomised controlled trials. *Ann Oncol.* 2012; 23:1607–16. <https://doi.org/10.1093/annonc/mdr491> PMID: 22015453
6. Daw NC, Chou AJ, Jaffe N, Rao BN, Billups CA, Meyers PA, et al. Survival after recurrence of osteosarcoma: A 20-year experience at a single institution. *Br J Cancer.* Nature Publishing Group; 2006; 112:278–82.
7. Daw NC, Chou AJ, Jaffe N, Rao BN, Billups CA, Rodriguez-Galindo C, et al. Recurrent osteosarcoma with a single pulmonary metastasis: A multi-institutional review. *Br J Cancer.* Nature Publishing Group; 2015; 112:278–82.
8. Heare T, Hensley MA, Dell’Orfano S. Bone tumors: Osteosarcoma and Ewing’s sarcoma. *Curr Opin Pediatr.* 2009; 21:365–72. <https://doi.org/10.1097/MOP.0b013e32832b1111> PMID: 19421061
9. Byun BH, Kong C-B, Park J, Seo Y, Lim I, Choi CW, et al. Initial metabolic tumor volume measured by 18F-FDG PET/CT can predict the outcome of osteosarcoma of the extremities. *J Nucl Med.* 2013; 54:1725–32. <https://doi.org/10.2967/jnumed.112.117697> PMID: 23949909
10. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJR. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging.* 2013; 40:133–40. <https://doi.org/10.1007/s00259-012-2247-0> PMID: 23064544
11. Asselin MC, O’Connor JPB, Boellaard R, Thacker NA, Jackson A. Quantifying heterogeneity in human tumours using MRI and PET. *Eur J Cancer.* 2012; 48:447–55. <https://doi.org/10.1016/j.ejca.2011.12.025> PMID: 22265426
12. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges J-P, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med.* 2011; 52:369–78. <https://doi.org/10.2967/jnumed.110.082404> PMID: 21321270
13. Lambin P, Van Stiphout RGPM, Starmans MHW, Rios-Velazquez E, Nalbantov G, Aerts HJWL, et al. Predicting outcomes in radiation oncology-multifactorial decision support systems. *Nat Rev Clin Oncol.* 2013; 10:27–40. <https://doi.org/10.1038/nrclinonc.2012.196> PMID: 23165123

14. Orlhac F, Nioche C, Buvat I. Texture—User Guide Local Image Features Extraction. 2018.
15. Nioche C, Orlhac F, Boughdad S, Reuze S, Goya-Outi J, Robert C, et al. Lifex: A freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* 2018; 78:4786–9. <https://doi.org/10.1158/0008-5472.CAN-18-0125> PMID: 29959149
16. Orlhac F, Maisonobe J, Garcia CA, Vanderlinden B. Tumor texture analysis in 18 F-FDG PET: Relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med.* 2014; 55:414–23. <https://doi.org/10.2967/jnumed.113.129858> PMID: 24549286
17. Orlhac F, Soussan M, Chouahnia K, Martinod E, Buvat I. 18F-FDG PET-derived textural indices reflect tissue-specific uptake pattern in non-small cell lung cancer. Adusumilli PS, editor. *PLoS One* 2015; 10:1–16.
18. Kirienco M, Cozzi L, Antunovic L, Lozza L, Fogliata A, Voulaz E, et al. Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery. *Eur J Nucl Med Mol Imaging* 2018; 45:207–17. <https://doi.org/10.1007/s00259-017-3837-7> PMID: 28944403
19. Hatt M, Tixier F, Pierce L, Kinahan PE, Cheze Le Rest C, Visvikis D, et al. Characterization of PET/CT images using texture analysis: The past, the present. . . any future? *Eur J Nucl Med Mol Imaging* 2017; 44:151–65. <https://doi.org/10.1007/s00259-016-3427-0> PMID: 27271051
20. Vallières 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. <https://doi.org/10.1088/0031-9155/60/14/5471> PMID: 26119045
21. Eary JF, Sullivan FO, Sullivan JO, Conrad EU. Spatial heterogeneity in sarcoma 18 F-FDG uptake as a predictor of patient outcome. *J Nucl Med.* 2009; 49:1973–9.
22. Eary JF, O'Sullivan F, Powitan Y, Chandhury K, Vernon C, Bruckner JD, et al. Sarcoma tumor FDG uptake measured by PET and patient outcome: A retrospective analysis. *Eur J Nucl Med.* 2002; 29:1149–54.
23. Hwang JP, Lim I, Kong CB, Jeon DG, Byun BH, Kim B II, et al. Prognostic value of SUVmax measured by pretreatment fluorine-18 fluorodeoxyglucose positron emission tomography/computed tomography in patients with Ewing sarcoma. *PLoS One.* 2016; 11:1–9.
24. Skamene SR, Rakheja R, Dalhstrom KR, Roberge D, Nahal A, Charest M, et al. Metabolic activity measured on PET/CT correlates with clinical outcomes in patients with limb and girdle sarcomas. *J Surg Oncol.* 2014; 109:410–4. <https://doi.org/10.1002/jso.23523> PMID: 24310279
25. Kong CB, Byun BH, Lim I, Choi CW, Lim SM, Song WS, et al. 18F-FDG PET SUVmax as an indicator of histopathologic response after neoadjuvant chemotherapy in extremity osteosarcoma. *Eur J Nucl Med Mol Imaging.* 2013; 40:728–36. <https://doi.org/10.1007/s00259-013-2344-8> PMID: 23361860
26. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: The metabolic requirements of cell proliferation. *Science* 2009; 324:1029–33. <https://doi.org/10.1126/science.1160809> PMID: 19460998
27. Keith B, Simon MC. Hypoxia-inducible factors, stem cells, and cancer. *Cell.* 2007; 129:465–72. <https://doi.org/10.1016/j.cell.2007.04.019> PMID: 17482542
28. Gatenby RA, Gillies RJ. Why do cancers have high aerobic glycolysis? *Nat Rev Cancer.* Nature Publishing Group; 2004; 4:891–9.
29. Cao J, Wang Y, Dong R, Lin G, Zhang N, Wang J, et al. Hypoxia-induced WSB1 promotes the metastatic potential of osteosarcoma cells. *Cancer Res.* 2015; 75:4839–51. <https://doi.org/10.1158/0008-5472.CAN-15-0711> PMID: 26424695
30. Guan G, Zhang Y, Lu Y, Liu L, Shi D, Wen Y, et al. The HIF-1 α /CXCR4 pathway supports hypoxia-induced metastasis of human osteosarcoma cells. *Cancer Lett.* 2015; 357:254–64. <https://doi.org/10.1016/j.canlet.2014.11.034> PMID: 25444927
31. Thiery JP, Acloque H, Huang RYJ, Nieto MA. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell.* 2009; 39:871–90.