

ORIGINAL RESEARCH

Open Access



Value of ^{18}F -FDG-PET/CT radiomics combined with clinical variables in the differential diagnosis of malignant and benign vertebral compression fractures

Xun Wang¹, Dandan Zhou², Yu Kong¹, Nan Cheng¹, Ming Gao¹, Guqing Zhang¹, Junli Ma³, Yueqin Chen^{1*} and Shuang Ge^{3*} 

Abstract

Background Vertebral compression fractures (VCFs) are common clinical problems that arise from various reasons. The differential diagnosis of benign and malignant VCFs is challenging. This study was designed to develop and validate a radiomics model to predict benign and malignant VCFs with ^{18}F -fluorodeoxyglucose-positron emission tomography/computed tomography (^{18}F -FDG-PET/CT).

Results Twenty-six features (9 PET features and 17 CT features) and eight clinical variables (age, SUVmax, SUVpeak, SULmax, SULpeak, osteolytic destruction, fracture line, and appendices/posterior vertebrae involvement) were ultimately selected. The area under the curve (AUCs) of the radiomics and clinical–radiomics models were significantly different from that of the clinical model in both the training group (0.986, 0.987 vs. 0.884, $p < 0.05$) and test group (0.962, 0.948 vs. 0.858, $p < 0.05$), while there was no significant difference between the radiomics model and clinical–radiomics model ($p > 0.05$). The accuracies of the radiomics and clinical–radiomics models were 94.0% and 95.0% in the training group and 93.2% and 93.2% in the test group, respectively. The three models all showed good calibration (Hosmer–Lemeshow test, $p > 0.05$). According to the decision curve analysis (DCA), the radiomics model and clinical–radiomics model exhibited higher overall net benefit than the clinical model.

Conclusions The PET/CT-based radiomics and clinical–radiomics models showed good performance in distinguishing between malignant and benign VCFs. The radiomics method may be valuable for treatment decision-making.

Keywords Vertebral compression fractures, Radiomics, Positron emission tomography/computed tomography

Background

Vertebral compression fractures (VCFs) are common clinical problems that can be caused by various reasons [1–4]. According to the nature of the disease, it can be divided into benign and malignant VCFs. Benign VCFs are caused by osteoporosis or trauma, while malignant VCFs are caused by tumors, such as metastatic solid tumors, myeloma, lymphoma, and Langerhans cell histiocytosis (LCH) [1–3, 5]. With increasing age, benign and malignant VCFs can occur simultaneously [1, 6]. Correct

*Correspondence:

Yueqin Chen
chenyueqin010@163.com
Shuang Ge
geshuang616@163.com

¹ Department of Medical Imaging, Affiliated Hospital of Jining Medical University, Guhuai Road, Jining, Shandong, China

² Big Data and Artificial Intelligence, Jining Polytechnic, Jinyu Road, Jining, Shandong, China

³ Department of Radiation Oncology, Affiliated Hospital of Jining Medical University, Guhuai Road, Jining, Shandong, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

identification of benign and malignant VCFs is of great importance in guiding treatment.

Both open biopsy and percutaneous biopsy are invasive with complications of different degrees. Modern radiological imaging techniques, including computed tomography (CT), magnetic resonance imaging (MRI) and bone single photon emission computed tomography/computed tomography (SPECT/CT), may be helpful in the differential diagnosis based on morphology, signal intensity abnormalities, blood flow and metabolic status. MRI has been established as the most relevant imaging technique for the diagnosis of malignant vertebral lesions due to its high sensitivity to bone marrow abnormalities [7, 8]. Some characteristics commonly associated with malignant fractures include the presence of an epidural or paravertebral soft tissue mass, abnormal signal of the pedicle or other posterior elements, diffuse posterior vertebral border convexity and so on [3, 8]. However, when these typical signs are absent, the diagnosis can be challenging. In addition, there are some drawbacks to MRI, including limited availability, absolute and relative contraindications such as pacemakers and claustrophobia.

¹⁸F-fluorodeoxyglucose-positron emission tomography/computed tomography (¹⁸F-FDG-PET/CT) can provide information about both morphological changes and metabolic status of the diseased vertebra. In general, tumor-induced fractures accumulate ¹⁸F-FDG, while benign fractures are not expected to accumulate the same high level of ¹⁸F-FDG [1, 9, 10]. This feature is helpful for the quantitative analysis of the lesions. In the study by Won-Ik Cho et al., the threshold of the maximum standardized uptake value (SUVmax) was 4.25 to differentiate malignant and benign VCFs, with 85% sensitivity and 71% specificity for malignancy [10]. In addition, some studies have found that the patterns of FDG uptake such as increased activity involving the pedicle and posterior element are also of great value in differentiating the nature of VCFs [10, 11]. In addition, many quantitative metabolic parameters of PET/CT, such as the peak standardized uptake value (SUVpeak), lean body mass correction of SUV (SUL), and metabolic tumor volume (MTV) have been widely used in the diagnosis, treatment and prognosis assessment of various tumors [12–16]. However, few studies have involved using them in the differential diagnosis of benign or malignant bone lesions [17, 18].

To date, the diagnosis of the causes of VCFs by CT, MRI, bone SPECT/CT and PET/CT is still largely dependent on the experience of diagnostic physicians. Radiomics can extract a large number of quantitative features from digital medical images in high throughput and provide objective information that is difficult for human eyes to quantify [19–21]. Many studies have

included reports of the high diagnostic performance of radiomics in distinguishing benign and malignant VCFs using CT and MRI [3, 22].

The objective of this study was to predict the benign and malignant nature of VCFs based on PET/CT radiomics and clinical indicators.

Materials and methods

Study participants

Our institutional Ethics Review Board approved this retrospective study and waived the written informed consent requirement. The analysis was conducted on 439 patients diagnosed with vertebral compression fracture or pathological fracture after PET/CT in our hospital from January 2016 to January 2023, with a total of 539 vertebrae. Vertebral compression fracture was defined as (1) a reduction in vertebral height, (2) cortical discontinuity of an endplate or vertebral cortex with impaction into the vertebral body, or (3) buckling of the vertebral cortex [19, 23]. Patients who had lost follow-up, unclear final diagnosis, asymptomatic old compression fractures, poor image quality or corresponding vertebral body treatment (such as vertebroplasty, kyphoplasty, local radiotherapy of vertebra or systemic chemotherapy) before PET/CT were excluded, and 121 patients with 144 vertebrae were enrolled in the study. All enrolled patients had varying degrees of pain in the corresponding vertebral area within 6 weeks before PET/CT examination. The cohort of patients was randomly divided into a training group (n = 100) and a test group (n = 44) at a ratio of 7:3. There were 47 benign VCFs and 53 malignant VCFs in the training group and 20 benign VCFs and 24 malignant VCFs in the test group (Fig. 1).

Image acquisition

¹⁸F-FDG-PET/CT imaging was obtained by a GE Discovery710 instrument, USA. ¹⁸F-FDG was produced by the GE Minitrace cyclotron and FDG synthesis module, and the radiochemical purity was >95%. All patients fasted for more than 6 h, and blood glucose was controlled below 11.1 mmol/L before injection. Patients received an intravenous injection of approximately 3.70–5.55 MBq/Kg body weight of ¹⁸F-FDG, and PET/CT scans were performed from the top of the skull to the upper femur (limbs were scanned if necessary) after a 60 min rest. The CT scanning voltage was 120 keV, tube current was 100 mAs, and the layer thickness was 3.75 mm. PET scanning was performed with three-dimensional acquisition, 2.5 min/bed, and 5~7 beds were collected. An iterative method was used to reconstruct the image.

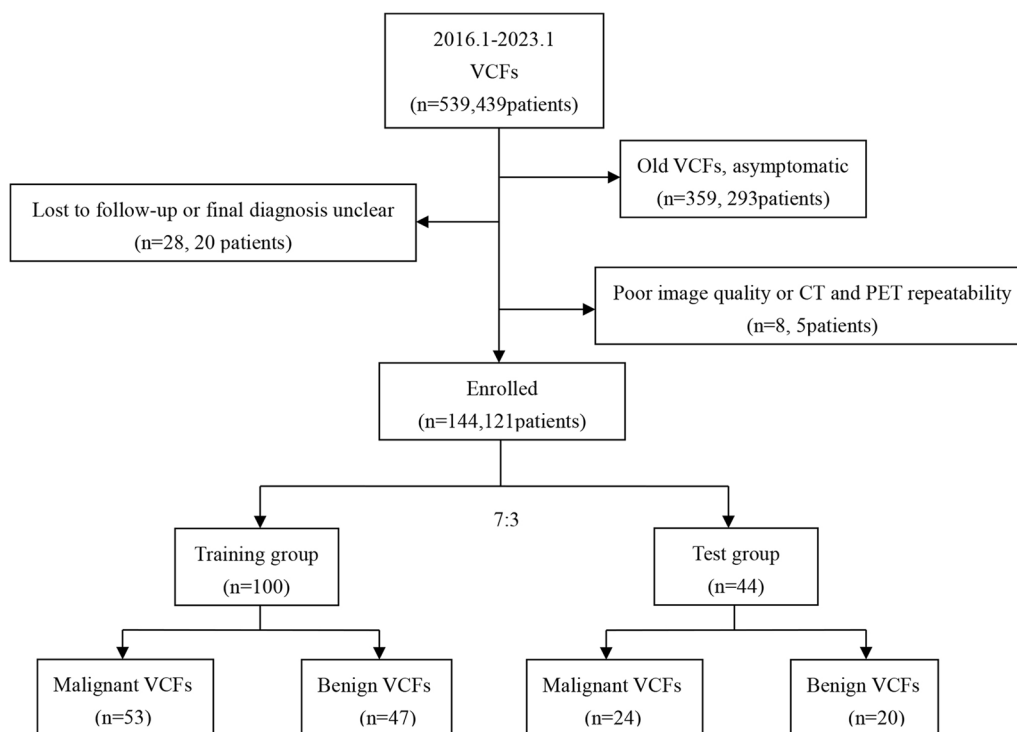


Fig. 1 Flowchart of patient screening in this study. VCFs vertebral compression fractures

Clinical data collection, lesion segmentation and radiomics feature extraction

Two nuclear medicine physicians with more than 5 years of experience, who were blinded to the lesion results, analyzed and recorded the lesion features, including lesion site, number, osteolytic destruction, fracture line, soft tissue mass/swelling, and appendices/posterior vertebrae involvement. If there was a disagreement, the decision was made by a third experienced nuclear medicine physician. Multiple metabolic parameters including SUVmax, SUVpeak, SULmax, and SULpeak were measured by creating a region of interest (ROI) along the lesion edge. Data on the patients’ age, sex and history of malignancy were also collected.

The workflow of the radiomics analysis included lesion segmentation, radiomics feature extraction, feature selection, and model construction (Fig. 2). PET/CT data in digital imaging and communications in medicine (DICOM) format were imported into 3D-slicer software (Version 5.1.0, <https://www.slicer.org>). Two experienced PET/CT diagnostic physicians manually drew the ROI and were blinded to the final diagnosis and clinical history of the patients. For each fractured vertebra, a three-dimensional ROI was drawn along the margin of the cortex of the whole vertebral body and the anterior margin of the bilateral pedicle on sagittal CT images of 3.75-mm thickness. The ROI of PET was sketched

along the edges of the ROI of CT. A total of 107 features including 18 first-order features, 75 texture features and 14 shape features were extracted from the ROIs of PET and CT images using 3D Slicer’s built-in module Slicer-Radiomics. There are five categories of texture features, including the gray-level dependence matrix (GLDM), gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), and neighboring gray tone difference matrix (NGTDM). To ensure the stability and reproducibility of the acquired radiomics features, each set of PET and CT image data ROIs was segmented, and radiomics features were extracted twice. The intraclass correlation coefficient (ICC) for each radiomics feature was calculated. ICC > 0.75 was considered indicative of stability, and features were thereby entered into the statistical analysis that followed.

Feature selection and radiomics model establishment

Student’s t test was conducted if the features were consistent with the normal distribution, and the Mann–Whitney *U* test was conducted if the features were not. A value of *p* < 0.05 was considered statistically significant. For features with high repeatability, the Pearson’s standard correlation coefficient (defined as *corr*) was used to calculate the correlation among features. For those pairs of features showing high correlation coefficient

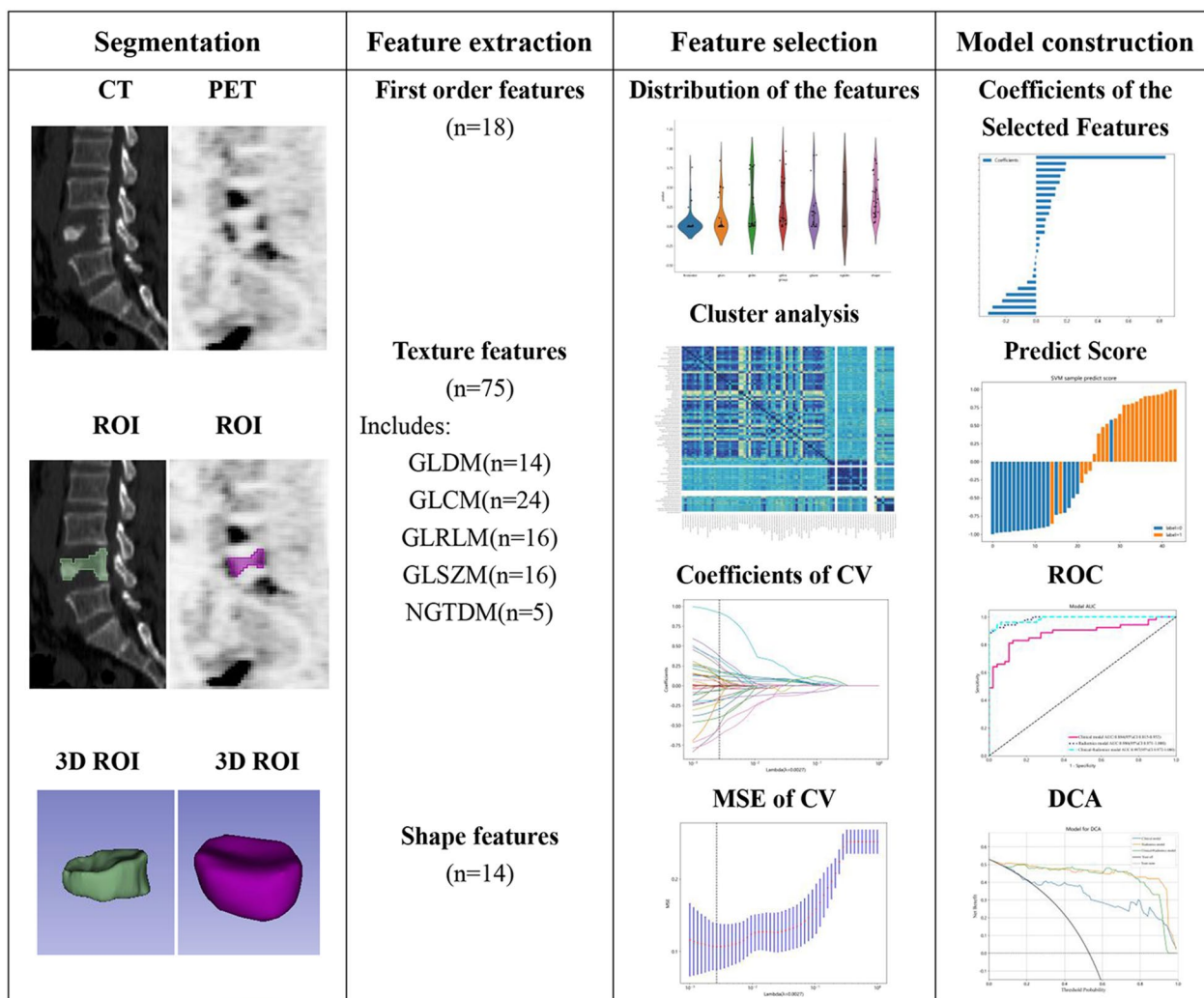


Fig. 2 Workflow of the radiomics analysis. ROI region of interest. 3D ROI three-dimensional ROI. CV cross-validation. MSE mean square error. ROC receiver operating characteristic. DCA decision curve analysis

($corr > 0.9$), the feature with a higher p value was excluded and the maximum redundancy was deleted each time. Least absolute shrinkage and selection operator (LASSO) regression was applied with penalty parameter tuning by tenfold cross-validation, and features with nonzero coefficients were selected. Finally, the features of CT and PET after screening were entered into a support vector machine (SVM) classifier to establish a radiomics signature that distinguished between benign and malignant VCFs. A radiomics model was built based on both CT and PET findings.

Clinical model building and clinical–radiomics model

The building process of the clinical signature was almost the same as the radiomics signature. Student’s t test was used for normally distribution data, and the Mann–Whitney U test was used for nonnormally distributed

data. The Chi-square test was used for categorical variables. Statistically significant characteristics were selected. The same machine learning model in the clinical signature building process was used. Then, the clinical signature and radiomics signature were combined to develop a final clinical–radiomics prediction model. A clinical–radiomics nomogram was developed in this study.

Performance evaluation of the models

The performance of each model was assessed according to the area under the curve (AUC) based on the receiver operating characteristic (ROC) curve analysis, and the sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) were calculated in the training and test groups. The Delong test was used to compare the AUC between the three models.

Decision curve analysis (DCA) was used to determine the clinical benefit of each model.

Statistical analysis

Statistical analyses were performed using SPSS (version 20.0; IBM Corp.) and the “Onekey AI” platform (<https://www.medai.icu>), which is based on PyTorch 1.8.0. Continuous variables are expressed as the mean±standard deviation(SD), and categorical variables are expressed as absolute counts and percentages(%). Statistical significance was defined as a two-sided *p* value < 0.05.

Results

Clinical characteristics

Among the 121 patients, 51 patients were diagnosed with benign VCFs, with 67 vertebral bodies (27 males, 24 females, age 70.6±8.9 years, range 50–92 years), and 70 patients were diagnosed with malignant VCFs, with 77 vertebral bodies (40 males, 30 females, age 60.6±13.4 years, range 13–81 years). The difference between benign and malignant VCFs was statistically significant in age (*p* < 0.05), but not in sex (*p* > 0.05).

All cases were confirmed by histopathology or clinical follow-up examination. Of the 67 benign VCFs, 6 were surgically confirmed and 61 were confirmed by clinical follow-up; among them, 39 cases had a history of malignant tumors. Of the 77 malignant VCFs, 31 were confirmed by puncture or surgery, and 46 were confirmed by comprehensive imaging diagnosis and follow-up. Among them, 64 cases were metastatic solid tumors (28 lung cancers, 7 breast cancers, 6 prostatic cancers, 5 thyroid cancers, 4 colorectal cancers, 4 hepatocellular carcinomas, 2 gastric cancers, 2 renal cancers, 2 esophageal cancers,

cervical cancer, ovarian cancer, pancreatic cancer, and synovial sarcoma), 6 cases were multiple myeloma, 5 cases were lymphoma, and 2 cases were Langerhans cell histiocytosis.

Radiomics feature selection, establishment and performance of the radiomics model

Twenty-six features, consisting of 9 PET features (6 first-order features and 3 texture features) and 17 CT features (3 first-order features, 13 texture features and 1 shape feature) were selected to construct the radiomics model after LASSO regression and tenfold cross-validation (Fig. 3). The details of the selected features are shown in Fig. 4. The formula of the radiomics signature score (rad-score) for each patient is shown in Table 1.

The AUC of the radiomics model for predicting the probability of malignancy of the VCFs was 0.986 (95% confidence interval [CI], 0.9714–1.0000) for the training group and 0.962 (95% CI, 0.9137–1.0000) for the test group (Fig. 5). The accuracy, sensitivity, specificity, PPV, and NPV were 0.940, 0.887, 1.000, 1.000, and 0.887 in the training group and 0.932, 0.917, 0.950, 0.957, and 0.905 in the test group, respectively (Table 2).

Establishment and performance of the clinical model and clinical–radiomics model

The selection of features for establishing the clinical model was based on a *p* value < 0.05 in the training and test groups. SUVmax, SUVpeak, SULmax, SULpeak, age, osteolytic destruction, fracture line and involvement of the appendices/posterior vertebrae met the conditions and were used to build clinical model (Table 3).

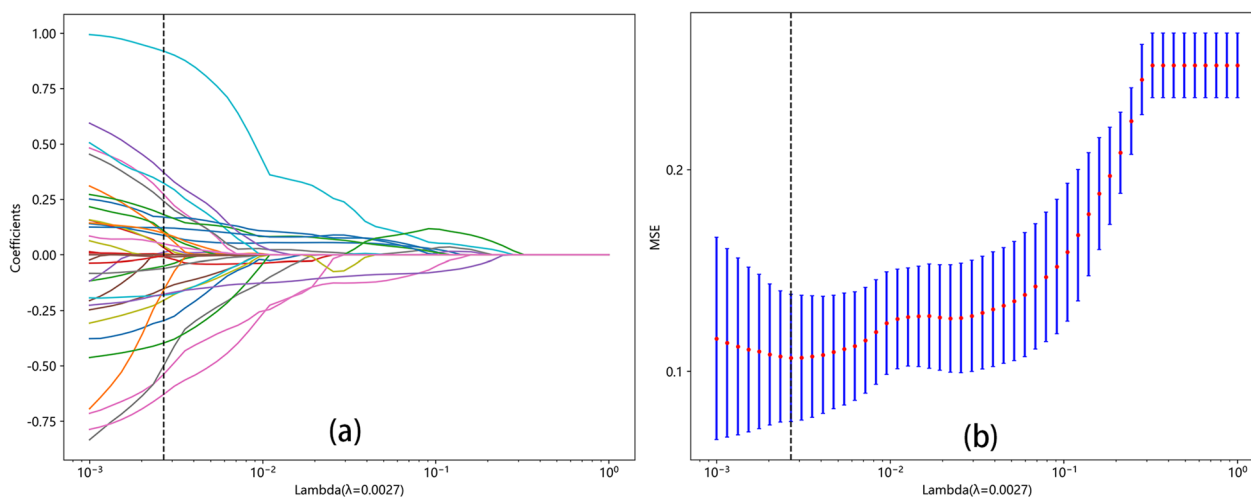


Fig. 3 LASSO regression and tenfold cross-validation were used to select the radiomics features. **a** LASSO coefficient profiles of the radiomic features. **b** Optimal feature selection of CV. LASSO least absolute shrinkage and selection operator. CV cross-validation. MSE mean square error

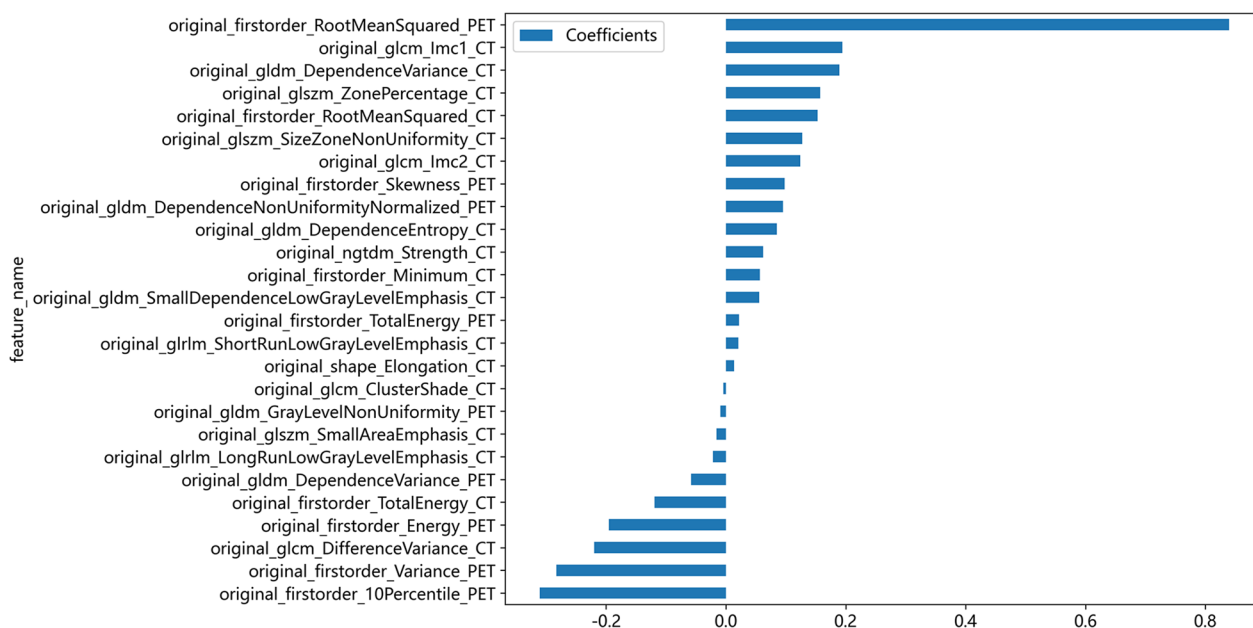


Fig. 4 Histogram of the coefficients of the selected features

Table 1 Formular of radiomics signature score (rad-score)

Rad-score = 0.5351086867374044

+0.056623*firstorder_Minimum_CT	-0.310642*firstorder_10Percentile_PET
+0.153041*firstorder_RootMeanSquared_CT	-0.195190*firstorder_Energy_PET
-0.119239*firstorder_TotalEnergy_CT	+0.839216*firstorder_RootMeanSquared_PET
-0.004347*GLCM_ClusterShade_CT	+0.098178*firstorder_Skewness_PET
-0.219531*GLCM_DifferenceVariance_CT	+0.021632*firstorder_TotalEnergy_PET
+0.194025*GLCM_Imc1_CT	-0.282702*firstorder_Variance_PET
+0.124174*GLCM_Imc2_CT	+0.095305*GLDM_DependenceNonUniformityNormalized_PET
+0.085175*GLDM_DependenceEntropy_CT	-0.057897*GLDM_DependenceVariance_PET
+0.189177*GLDM_DependenceVariance_CT	-0.009464*GLDM_GrayLevelNonUniformity_PET
+0.055343*GLDM_SmallDependenceLowGrayLevelEmphasis_CT	
-0.021720*GLRLM_LongRunLowGrayLevelEmphasis_CT	
+0.020565*GLRLM_ShortRunLowGrayLevelEmphasis_CT	
+0.127269*GLSZM_SizeZoneNonUniformity_CT	
-0.015686*GLSZM_SmallAreaEmphasis_CT	
+0.157129*GLSZM_ZonePercentage_CT	
+0.062160*NGTDM_Strength_CT	
+0.013747*SHAPE_Elongation_CT	

The AUC of the clinical model for predicting the probability of malignancy of the VCFs was 0.884 (95% CI, 0.8153–0.9518) for the training group and 0.858(95% CI, 0.7437–0.9729) for the test group (Fig. 5). The accuracy, sensitivity, specificity, PPV, and NPV were 0.850, 0.830, 0.872, 0.880, and 0.820 in the training group and 0.841, 0.875, 0.800, 0.840, and 0.842 in the test group, respectively (Table 2).

The AUC of the clinical–radiomics model for predicting the probability of malignancy of the VCFs was 0.987 (95% CI, 0.9716–1.0000) for the training group and 0.948 (95% CI, 0.8787–1.0000) for the test group (Fig. 5). The

accuracy, sensitivity, specificity, PPV, and NPV were 0.950, 0.906, 1.000, 1.000, and 0.904 in the training group and 0.932, 0.958, 0.900, 0.920, and 0.947 in the test group, respectively (Table 2).

Performance of the prediction models and nomogram construction

The clinical model, radiomics model and clinical–radiomics model all showed good calibration. The *p* values of the Hosmer–Lemeshow test for the three models were 0.664, 0.787, and 0.422 in the training group and 0.241, 0.237, and 0.051 in the test group, respectively.

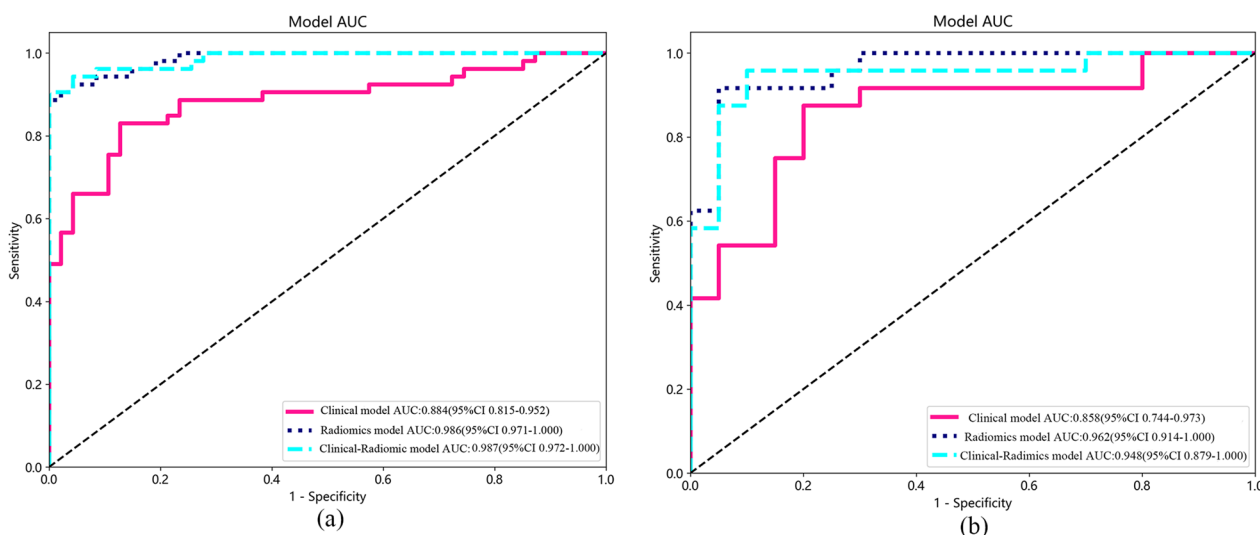


Fig. 5 AUCs of the prediction models. **a** The training group. **b** The test group. *AUC* area under the curve

Table 2 Performance of the prediction models

Prediction model	AUC	95%CI	Accuracy	Sensitivity	Specificity	PPV	NPV	Group
Clinical model	0.884	0.8153–0.9518	0.850	0.830	0.872	0.880	0.820	Train
Radiomics model	0.986	0.9714–1.0000	0.940	0.887	1.000	1.000	0.887	Train
Clinical–radiomics model	0.987	0.9716–1.0000	0.950	0.906	1.000	1.000	0.904	Train
Clinical model	0.858	0.7437–0.9729	0.841	0.875	0.800	0.840	0.842	Test
Radiomics model	0.962	0.9137–1.0000	0.932	0.917	0.950	0.957	0.905	Test
Clinical–radiomics model	0.948	0.8787–1.0000	0.932	0.958	0.900	0.920	0.947	Test

AUC area under curve; *CI* confidence interval; *PPV* positive predictive value; *NPV* negative predictive value

The Delong test was used to compare the AUCs of the three models. In both the training and test groups, the radiomics model and clinical–radiomics model were significantly different from the clinical model ($p < 0.05$), but there was no significant difference between the radiomics model and clinical–radiomics model ($p > 0.05$). The DCA demonstrated that the radiomics model and clinical–radiomics model could provide higher overall net benefit than the clinical model (Fig. 6). A nomogram based on the rad-score and clinical risk factors was developed (Fig. 7).

Discussion

Imaging plays a crucial role in the diagnosis of VCFs. Early and accurate differential diagnosis of benign and malignant VCFs can allow clinicians to effectively choose appropriate treatment plans and potentially provide improved outcomes. Previous studies have reported several CT features that were more frequently found in benign VCFs with statistical significance, including fracture lines, sclerotic bands beneath the end plate, and diffuse thin paraspinous soft tissue thickening

[1, 24]. In addition, the intravertebral vacuum phenomenon has never been visualized in malignant VCFs, although its occurrence is uncommon and not statistically significant [1]. CT findings that are predictive of malignant VCFs include osteolytic destruction and epidural or focal paravertebral soft tissue masses [24, 25]. One study showed an accuracy of 89.7% in the differentiation of malignant from osteoporotic vertebral fractures based on the CT scoring system [24]. In our study, osteolytic destruction, fracture line and appendices/posterior vertebrae involvement had statistical significance in differentiating benign and malignant VCFs, but there was no significant difference in soft tissue mass/swelling. A possible reason is that benign VCFs can present as paravertebral or epidural hemorrhage with soft tissue edema. Malignant VCFs usually present as soft tissue masses. However, if there is inflammation around the tumor and/or no tumor infiltrating the cortex, malignant VCFs may also present as similar smooth soft tissue swelling [1]. In our study, we did not distinguish between soft tissue masses and swelling, which might have influenced the results.

Table 3 Clinical variables of benign and malignant VCFs in the training and test groups

Clinical variable	Training (n = 100)		p Value*	Test (n = 44)		P Value
	Benign (n = 47)	Malignant (n = 53)		Benign (n = 20)	Malignant (n = 24)	
SULpeak	2.79 ± 0.90 [†]	5.72 ± 2.73	< 0.001	2.73 ± 1.02	5.46 ± 3.08	< 0.001
SULmax	3.45 ± 1.12	7.35 ± 3.59	< 0.001	3.36 ± 1.15	7.16 ± 4.05	< 0.001
SUVpeak	3.58 ± 1.08	7.37 ± 3.50	< 0.001	3.43 ± 1.27	6.92 ± 3.90	< 0.001
SUVmax	4.43 ± 1.35	9.46 ± 4.51	< 0.001	4.23 ± 1.45	9.08 ± 5.08	< 0.001
Age, year	71.68 ± 9.29	59.89 ± 14.40	< 0.001	69.00 ± 8.74	62.25 ± 11.76	0.04
Osteolytic destruction			< 0.001			0.003
No	39 (82.98) [‡]	14 (26.42)		18 (90.00)	10 (41.67)	
Yes	8 (17.02)	39 (73.58)		2 (10.00)	14 (58.33)	
Fracture line			< 0.001			0.025
No	29 (61.70)	51 (96.23)		13 (65.00)	23 (95.83)	
Yes	18 (38.30)	2 (3.77)		7 (35.00)	1 (4.17)	
Soft tissue mass/swelling			0.097			0.974
No	35 (74.47)	30 (56.60)		16 (80.00)	18 (75.00)	
Yes	12 (25.53)	23 (43.40)		4 (20.00)	6 (25.00)	
Appendices/posterior vertebrae involvement			< 0.001			0.034
No	38 (80.85)	15 (28.30)		17 (85.00)	12 (50.00)	
Yes	9 (9.15)	38 (71.70)		3 (15.00)	12 (50.00)	

SUVmax maximum standardized uptake value; SUVpeak peak of standardized uptake value; SULmax maximum lean body mass correction of SUV; SULpeak peak of lean body mass correction of SUV

*The differences were assessed by Mann–Whitney U test or student T-test

[†] Mean ± SD: mean ± standard deviation

[‡] Percentage

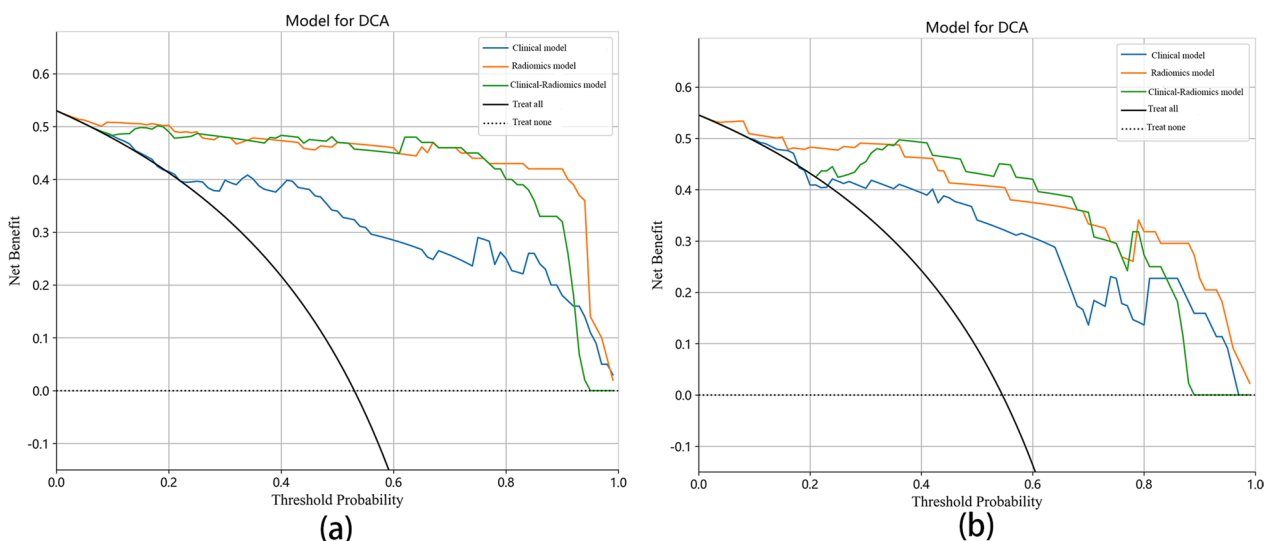


Fig. 6 DCA of the prediction model. **a** The training group. **b** The test group. DCA decision curve analysis

Conventional metabolic parameters of PET/CT such as max and peak of SUV and SUL have been used to quantify intratumoral heterogeneity [26]. Some previous studies involved investigating the diagnostic accuracies of PET/CT for differentiating between malignant and

benign VCFs and revealed sensitivity ranging from 86 to 100% and specificity ranging from 29.4 to 92.8% [10, 27, 28]. In a meta-analysis, the results reported that PET/CT had excellent diagnostic accuracy in the detection of malignant VCFs, with a sensitivity of 0.96, specificity of

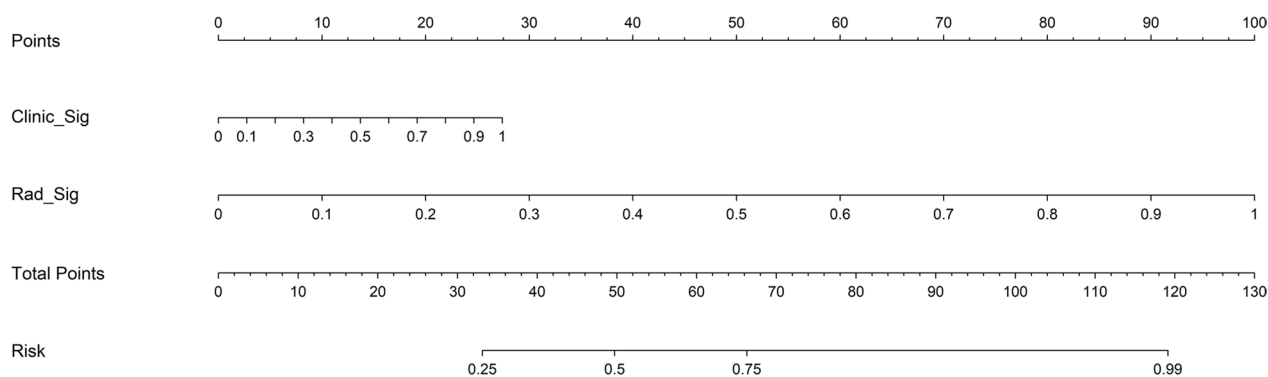


Fig. 7 Nomogram to predict the malignancy of VCFs

0.77 and AUC of 0.94 [29]. However, a single metabolic parameter (SUVmax) was chosen for analysis in all of these studies. In our study, we chose multiple parameters for analysis. The SUVmax value of benign VCFs was lower than that of malignant VCFs, which was useful in the differential diagnosis of VCFs and consistent with previous studies [1, 29, 30]. The SUVpeak, SULmax, and SULpeak values of malignant VCFs were higher than those of benign VCFs ($p < 0.05$), indicating that malignant VCFs were associated with higher metabolic activity on PET/CT.

Age was significantly different in distinguishing between benign and malignant VCFs. The mean age in the malignant VCF group was less than that in the benign VCF group, which was consistent with previous studies [2, 19, 24]. In our study, the clinical model including age, SUVmax, SUVpeak, SULmax, SULpeak, osteolytic destruction, fracture line and appendices/posterior vertebrae involvement showed a sensitivity of 0.830, specificity of 0.872, and AUC of 0.884 in the training group and sensitivity of 0.875, specificity of 0.800, and AUC of 0.858 in the test group, all of which were values lower than those from the meta-analysis by Kim SJ et al. [29]. This may be because asymptomatic old compression fractures were excluded in our study, and the metabolic parameters of acute VCFs on PET/CT may be higher than those of old VCFs.

Radiomics has been used in many clinical studies, including tumor molecular characteristics, patient prognosis and response to therapy [31]. A number of studies have shown that radiomics and deep learning based on CT and MRI have good diagnostic performance in distinguishing benign and malignant VCFs. The AUCs of the radiomics score on CT for predicting the malignancy probability of VCFs were 0.852–0.97 [19, 25]. The AUC and accuracy of machine learning based on MRI to identify benign versus malignant indistinguishable VCFs were 0.86 and 87.61%,

respectively [32]. Liu B et al. believed that automatic deep learning networks showed better diagnostic performance than radiologists in identifying benign or malignant VCFs, and were potentially useful tools for future clinical applications [33]. In our study, we developed and validated a radiomics model and a clinical–radiomics model to predict the malignancy of VCFs on PET/CT. The discrimination performance of the radiomics model and clinical–radiomics model was higher than that of the clinical model in both the training group (AUC: 0.986, 0.987 vs. 0.884, $p < 0.05$) and the test group (AUC: 0.962, 0.948 vs. 0.858, $p < 0.05$).

We believe that radiomics-based models can improve the diagnostic accuracy and efficiency of diagnostic physicians because radiomics models use mathematical algorithms to describe lesions more objectively and can complement radiologists by providing quantitative information that is not available through visual analysis. Radiomics features are believed to reflect intraregional heterogeneity [21, 34]. In our study, the radiomics model included 9 PET features (6 first-order features and 3 texture features) and 17 CT features (3 first-order features, 13 texture features and 1 shape feature). Among the 26 features, the root mean squared, skewness, dependence variance, size zone nonuniformity and dependence entropy were the features with high weight coefficients. Root mean squared as entropy-derived data has emerged as one of the most relevant radiomics features for tumor aggressiveness [35]. The skewness represents the asymmetry of the gray distribution. Higher skewness has been reported as a predictive feature of reduced survival and genetic mutations in lung and colorectal cancer [36–38]. In our study, we found that malignant VCFs had a higher discretized intensity skewness than benign VCFs, which was consistent with the results of Choong Guen Chee’s study [19]. Furthermore, we created a nomogram including the rad-score and clinical risk factors that can depict

the prediction results and provide an easy-to-use method for individualized prediction of benign and malignant VCF.

This study has several limitations. First, this was a single-center retrospective study with a relatively small sample size, especially in the test group. Second, most of the malignant VCFs were metastatic tumors, while the proportion of myeloma and lymphoma was small, which might lead to potential confounding factors. Third, the slice thickness of 3.75 mm might be a limitation to this study. Thinner axial slices could increase reliability. Furthermore, we used radiomics for the analysis, and expanding the sample size for deep learning might make the results more reliable and meaningful. Finally, we only performed internal validation. Additional external validation is required to confirm the robustness and generalization of our model.

Conclusions

In summary, the radiomics model and clinical–radiomics model combining clinical parameters with radiomics scores based on ^{18}F -FDG-PET/CT can be used to predict the malignancy of vertebral compression fractures with high diagnostic accuracy. The predictive models can serve as potential decision support tools for clinicians and nuclear medicine physicians and help facilitate the appropriate management of patients with VCFs.

Abbreviations

^{18}F -FDG-PET/CT	^{18}F -fluorodeoxyglucose-positron emission tomography/computed tomography
VCFs	Vertebral compression fractures
LCH	Langerhans cell histiocytosis
SUV	Standardized uptake value
SUL	Lean body mass correction of SUV
MTV	Metabolic tumor volume
TLG	Total lesion glycolysis
ROI	Region of interest
DICOM	Digital imaging and communications in medicine
GLDM	Gray-level dependence matrix
GLCM	Gray-level co-occurrence matrix
GLRLM	Gray-level run length matrix
GLSZM	Gray-level size zone matrix
NGTDM	Neighboring gray tone difference matrix
ICC	Intraclass correlation coefficient
LASSO	Least absolute shrinkage and selection operator
SVM	Support vector machine
AUC	Area under the curve
PPV	Positive predictive value
NPV	Negative predictive value
DCA	Decision curve analysis
CI	Confidence interval
OVCF	Osteoporotic vertebral compression fracture

Acknowledgements

Some of our experiments were carried out on the Onekey AI platform. We thank Onekey AI and its developers for their help in this scientific research work.

Author contributions

XW, DDZ, YQC and SG contributed to designing and performing experiments, analyzing data, interpreting results, and writing the manuscript. MG and YK contributed to performing the experiments and editing the manuscript. YQC, YK and GQZ contributed to interpreting the results and editing the manuscript. NC and JLM contributed to designing the experiments, analyzing the data, interpreting the results, and editing the manuscript. All the authors have read and approved the final manuscript.

Funding

This study was funded by the Jining City of Science and Technology Bureau (Key research and development project: 2022YXNS029).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. The Ethics Committee of the Affiliated Hospital of Jining Medical University approved the study with authorization number 2023-03-C018. Specific informed consent was waived because of the observational and retrospective study design.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 15 June 2023 Accepted: 20 September 2023

Published online: 11 October 2023

References

1. Mauch JT, Carr CM, Cloft H, Diehn FE. Review of the imaging features of benign osteoporotic and malignant vertebral compression fractures. *AJNR Am J Neuroradiol*. 2018;39:1584–92. <https://doi.org/10.3174/ajnr.A5528>.
2. Alsoof D, Anderson G, McDonald CL, Basques B, Kuris E, Daniels AHJTA-MoM. Diagnosis and management of vertebral compression fracture. *Am J Med*. 2022;135(7):815–21. <https://doi.org/10.1016/j.amjmed.2022.02.035>.
3. Frighetto-Pereira L, Rangayyan RM, Metzner GA, de Azevedo-Marques PM, Nogueira-Barbosa MH. Shape, texture and statistical features for classification of benign and malignant vertebral compression fractures in magnetic resonance images. *Comput Biol Med*. 2016;73:147–56. <https://doi.org/10.1016/j.compbio.2016.04.006>.
4. Yu T, Zheng S, Cheng X, Zhao J. Osteoporotic vertebral fracture misdiagnosed as metastatic vertebral fracture. *J College Phys Surg Pakistan*. 2022;32:ss171–3. <https://doi.org/10.29271/jcpsp.2022.Supp0.S5171>.
5. Del Lama RS, Candido RM, Chiari-Correia NS, Nogueira-Barbosa MH, de Azevedo-Marques PM, Tinós R. Computer-aided diagnosis of vertebral compression fractures using convolutional neural networks and radiomics. *J Digit Imag*. 2022;35:446–58. <https://doi.org/10.1007/s10278-022-00586-y>.
6. Li Y, Zhang Y, Zhang E, Chen Y, Wang Q, Liu K, et al. Differential diagnosis of benign and malignant vertebral fracture on CT using deep learning. *Eur Radiol*. 2021;31:9612–9. <https://doi.org/10.1007/s00330-021-08014-5>.
7. Lecouvet FE. Whole-Body MR. Imaging: musculoskeletal applications. *Radiology*. 2016;279:345–65. <https://doi.org/10.1148/radiol.2016142084>.
8. Chiari-Correia NS, Nogueira-Barbosa MH, Chiari-Correia RD, Azevedo-Marques PM. A 3D radiomics-based artificial neural network model for benign versus malignant vertebral compression fracture classification in MRI. *J Digit Imag*. 2023. <https://doi.org/10.1007/s10278-023-00847-4>.

9. Capdevila-Reniu A, Navarro-López M, López-Soto A. Osteoporotic vertebral fractures: a diagnostic challenge in the 21(st) century. *Revista clinica espanola*. 2021;221:118–24. <https://doi.org/10.1016/j.rceng.2019.09.013>.
10. Cho WJ, Chang UK, Jo NS. Comparison of MR imaging and FDG-PET/CT in the differential diagnosis of benign and malignant vertebral compression fractures. *J Neurosurg Spine*. 2011;14:177–83. <https://doi.org/10.3171/2010.10.SPINE10175>.
11. Bredella MA, Essary B, Torriani M, Ouellette HA, Palmer WE. Use of FDG-PET in differentiating benign from malignant compression fractures. *Skeletal Radiol*. 2008;37:405–13. <https://doi.org/10.1007/s00256-008-0452-5>.
12. Sarikaya I, Sarikaya A. Assessing PET parameters in oncologic (18)F-FDG studies. *J Nucl Med Technol*. 2020;48:278–82. <https://doi.org/10.2967/jnmt.119.236109>.
13. Grut H, Line PD, Labori KJ, Schulz A, Dueland S. Survival after liver resection and liver transplantation for colorectal liver metastases: a comparative analysis stratified by metabolic tumor volume assessed by (18)F-FDG PET/CT. *HPB: Off J Int Hepato Pancreato Biliary Assoc*. 2022;24:1055–62. <https://doi.org/10.1016/j.hpb.2021.11.015>.
14. Imajo M, Norikane T, Yamamoto Y, Maeda Y, Saitoh K, Kato K, et al. Relationship between [(18)F]FDG PET/CT and metabolomics in patients with colorectal cancer. *Metabol Off J Metabol Soc*. 2022;18:91–7. <https://doi.org/10.1007/s11306-022-01952-1>.
15. Filippi L, Bianconi F, Schillaci O, Spanu A, Palumbo B. The role and potential of (18)F-FDG PET/CT in malignant melanoma: prognostication, monitoring response to targeted and immunotherapy, and radiomics. *Diagnostics*. 2022;12:929–50. <https://doi.org/10.3390/diagnostics12040929>.
16. Hicks RJ. The value of the standardized uptake value (SUV) and metabolic tumor volume (MTV) in lung cancer. *Sem Nucl Med*. 2022;52:734–44. <https://doi.org/10.1053/j.semnucmed.2022.04.007>.
17. Choi S, Kim YI, Lee GD, Choi S, Kim HR, Kim YH, et al. Diagnostic value of 18F-FDG PET/CT in discriminating between benign and malignant lesions of the ribs. *Medicine*. 2022;101:e29867. <https://doi.org/10.1097/md.00000000000029867>.
18. Shen CT, Qiu ZL, Sun ZK, Wei WJ, Song HJ, Zhang XY, et al. Dual time-point (18)F-FDG PET/CT imaging with multiple metabolic parameters in the differential diagnosis of malignancy-suspected bone/joint lesions. *Oncotarget*. 2017;8:71188–96. <https://doi.org/10.18632/oncotarget.17140>.
19. Chee CG, Yoon MA, Kim KW, Ko Y, Chung HW, JER. Combined radiomics-clinical model to predict malignancy of vertebral compression fractures on CT. *Eur Radiol*. 2021;31:6825–34. <https://doi.org/10.1007/s00330-021-07832-x>.
20. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749–62. <https://doi.org/10.1038/nrclinonc.2017.141>.
21. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278:563–77. <https://doi.org/10.1148/radiol.2015151169>.
22. Yoda T, Maki S, Furuya T, Yokota H, Matsumoto K, Takaoka H, et al. Automated differentiation between osteoporotic vertebral fracture and malignant vertebral fracture on MRI using a deep convolutional neural network. *Spine*. 2022;47:E347–e52. <https://doi.org/10.1097/brs.0000000000004307>.
23. Lentle B, Trollip J, Lian K. The radiology of osteoporotic vertebral fractures redux. *J Clin Densitom*. 2016;19:40–7. <https://doi.org/10.1016/j.jocd.2015.08.009>.
24. Li Z, Guan M, Sun D, Xu Y, Li F, Xiong W, JBMD. A novel MRI- and CT-based scoring system to differentiate malignant from osteoporotic vertebral fractures in Chinese patients. *BMC Musculoskelet Disord*. 2018;19:406–12. <https://doi.org/10.1186/s12891-018-2331-0>.
25. Duan S, Hua Y, Cao G, Hu J, Cui W, Zhang D, et al. Differential diagnosis of benign and malignant vertebral compression fractures: Comparison and correlation of radiomics and deep learning frameworks based on spinal CT and clinical characteristics. *Eur J Radiol*. 2023;165:110899. <https://doi.org/10.1016/j.ejrad.2023.110899>.
26. Liu X, Xiang K, Geng GY, Wang SC, Ni M, Zhang YF, et al. Prognostic value of intratumor metabolic heterogeneity parameters on (18)F-FDG PET/CT for patients with colorectal cancer. *Contrast Media Mol Imag*. 2022;2022:2586245. <https://doi.org/10.1155/2022/2586245>.
27. Bredella MA, Essary B, Torriani M, Palmer OJSR. Use of FDG-PET in differentiating benign from malignant compression fractures. *Skeletal Radiol*. 2008;37:405–13. <https://doi.org/10.1007/s00256-008-0452-5>.
28. Aggarwal A, Salunke P, Shekhar BR, Chhabra R, Singh P, Bhattacharya A, et al. The role of magnetic resonance imaging and positron emission tomography-computed tomography combined in differentiating benign from malignant lesions contributing to vertebral compression fractures. *Surg Neurol Int*. 2013;4:S323–6. <https://doi.org/10.4103/2152-7806.112619>.
29. Kim SJ, Lee JS, JWN. Diagnostic performance of F-18 FDG PET or PET/CT for differentiation of benign from malignant vertebral compression fractures: A meta-analysis. *World Neurosurg*. 2020;137:626–33. <https://doi.org/10.4103/2152-7806.112619>.
30. He X, Zhao L, Guo X, Zhao L, Wu J, Huang J, et al. Differential diagnostic value of (18)F-FDG PET/CT for benign and malignant vertebral compression fractures: comparison with magnetic resonance imaging. *Cancer Manage Res*. 2018;10:2105–15. <https://doi.org/10.2147/cmar.S168374>.
31. Forghani R, Chatterjee A, Reinhold C, Pérez-Lara A, Romero-Sanchez G, Ueno Y, et al. Head and neck squamous cell carcinoma: prediction of cervical lymph node metastasis by dual-energy CT texture analysis with machine learning. *Eur Radiol*. 2019;29:6172–81. <https://doi.org/10.1007/s00330-019-06159-y>.
32. Zhang H, Yuan G, Wang C, Zhao H, Zhu K, Guo J, et al. Differentiation of benign versus malignant indistinguishable vertebral compression fractures by different machine learning with MRI-based radiomic features. *Eur Radiol*. 2023;33:5069–76. <https://doi.org/10.1007/s00330-023-09678-x>.
33. Liu B, Jin Y, Feng S, Yu H, Zhang Y, Li Y. Benign vs malignant vertebral compression fractures with MRI: a comparison between automatic deep learning network and radiologist's assessment. *Eur Radiol*. 2023;33:5060–8. <https://doi.org/10.1007/s00330-023-09713-x>.
34. Gao X, Tham IWK, Yan J. Quantitative accuracy of radiomic features of low-dose (18)F-FDG PET imaging. *Transl Cancer Res*. 2020;9:4646–55. <https://doi.org/10.21037/tcr-20-1715>.
35. Costa G, Cavinato L, Fiz F, Sollini M, Chiti A, Torzilli G, et al. Mapping tumor heterogeneity via local entropy assessment: making biomarkers visible. *J Digit Imag*. 2023. <https://doi.org/10.1007/s10278-023-00799-9>.
36. Coppola F, Mottola M, Lo Monaco S, Cattabriga A, Cocozza MA, Yuan JC, et al. The heterogeneity of skewness in T2W-based radiomics predicts the response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Diagnostics*. 2021;11:795–807. <https://doi.org/10.3390/diagnostic11050795>.
37. Ahn SY, Park CM, Park SJ, Kim HJ, Song C, Lee SM, et al. Prognostic value of computed tomography texture features in non-small cell lung cancers treated with definitive concomitant chemoradiotherapy. *Investig Radiol*. 2015;50:719–25. <https://doi.org/10.1097/rli.0000000000000174>.
38. Lubner MG, Stabo N, Lubner SJ, del Rio AM, Song C, Halberg RB, et al. CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. *Abdom Imag*. 2015;40:2331–7. <https://doi.org/10.1007/s00261-015-0438-4>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.