Predicting HER2 expression status in patients with gastric cancer using 18F-FDG PET/CT radiomics

Xiaojing Jiang (✉ jxj17369476823@163.com)
The Fourth Hospital of Hebei Medical University

Tianyue Li
The Fourth Hospital of Hebei Medical University

Zhaoqi Zhang
The Fourth Hospital of Hebei Medical University

Jianfang Wang
The Fourth Hospital of Hebei Medical University

Meng Dai
The Fourth Hospital of Hebei Medical University

Jingya Han
The Fourth Hospital of Hebei Medical University

Xiaolin Chen
The Fourth Hospital of Hebei Medical University

Jingmian Zhang
The Fourth Hospital of Hebei Medical University

Xinming Zhao
The Fourth Hospital of Hebei Medical University

Research Article

Keywords: 18F-FDG, PET/CT, Gastric cancer, HER2, Radiomics

Posted Date: June 15th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2962066/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background Immunohistochemistry (IHC) is the main method used to detect human epidermal growth factor receptor 2 (HER2) expression levels. However, IHC is invasive and cannot dynamically reflect HER2 expression status. The aim of this study was to construct and verify three types of radiomics models based on $^{18}$F-FDG PET/CT imaging and to evaluate the predictive ability of radiomics models for the expression status of HER2 in patients with gastric cancer (GC).

Methods A total of 118 patients with GC who underwent $^{18}$F-FDG PET/CT imaging before treatment were enrolled in this study. The LIFEx software package was applied to extract PET and CT radiomics features. The minimum absolute contraction and selection operator (LASSO) algorithm was employed to select the best radiomics features. Three machine learning methods, logistic regression (LR), support vector machine (SVM), and random forest (RF) models, were constructed and verified. The Synthetic Minority Oversampling Technique (SMOTE) was applied to address data imbalance.

Results In the training and test sets, the area under the curve (AUC) values of the LR, SVM, and RF models were 0.809 and 0.761, 0.861 and 0.628, and 0.993 and 0.717, respectively, and the Brier scores were 0.118, 0.214, and 0.143, respectively. Among the three models, the LR model exhibited the best prediction performance. The AUC values of the three models significantly improved after SMOTE balanced the data.

Conclusion $^{18}$F-FDG PET/CT-based radiomics models demonstrated good performance in predicting HER2 expression status in patients with GC and can be used to preselect patients who may benefit from HER2-targeted therapy.

Introduction

Gastric cancer (GC) is the most common malignant tumor of the digestive system, with a high incidence and poor prognosis[1, 2]. Studies have confirmed that Human Epidermal Growth Factor Receptor 2 (HER2) is not only correlated with poor prognosis in patients with advanced gastric cancer but also with the occurrence and development of gastric cancer [3, 4, 5]. In recent years, with the development of molecular biology, molecular subtypes of gastric have provided a good opportunity for personalized treatment, which will change treatment strategies for GC and improve the survival rate of patients. Biomarkers, such as HER2, are increasingly driving improvements in treatment regiments increasingly [6]. Studies have shown that approximately 30–50% of gastric cancer patients have HER2 positive expression [7].

At present, HER2 status of test specimens obtained through surgery or gastroscopy is mainly assessed by immunohistochemistry (IHC) or fluorescent in situ hybridization (FISH). However, gastroscopic biopsy is invasive and carries the risk of perforation and bleeding [8, 9], and the application of IHC is greatly limited for tumors that are not easily accessible for biopsies. In addition, GC is highly intra-tumor heterogeneous [10], HER2 expression status in the primary and metastatic lesions is occasionally different, and the expression level of HER2 in the same lesion is sometimes different before and after
treatment. $^{18}$F/$^{68}$Ga-HER2 affibody PET/CT can detect HER2 expression levels in malignant tumors, but it has not been widely applied in clinical practice until now. The National Comprehensive Cancer Network (NCCN), European Society of Medical Oncology (ESMO), and Chinese Society of Clinical Oncology (CSCO) recommend targeted therapy with trastuzumab combined with a standard chemotherapy regimen for HER2-positive GC patients [11, 12]. Therefore, Accurate identification of HER2 status is essential for the treatment of patients with GC. Therefore, it is of great significance to develop a new noninvasive and accurate method to assess HER2 expression levels in primary and metastatic lesions in GC patients dynamically and in real time.

In recent years, with the rapid development of artificial intelligence (AI) and machine learning (ML), radiomics, as an emerging and noninvasive technology, has played a non-negligible role in the diagnosis and management of cancer [13]. Currently, some researchers have found that radiomics based on $^{18}$F-fluorodeoxyglucose ($^{18}$F-FDG) Positron Emission Tomography/Computed Tomography (PET/CT) imaging has excellent performance in the diagnosis, staging, therapeutic efficacy assessment, and prognosis of various types of cancers [14]. However, no studies have been conducted to predict HER2 expression status of gastric cancer using $^{18}$F-FDG PET/CT-based radiomics.

In this study, we constructed and verified three types of radiomics models based on $^{18}$F-FDG PET/CT imaging and obtained the optimal model to evaluate the prediction ability of radiomics models in HER2 expression status in patients with gastric cancer, which can provide a noninvasive and accurate method to evaluate HER2 expression status and can filter patients who can benefit from HER2-targeted therapy.

**Materials and methods**

**Patient selection**

A total of 118 patients with gastric cancer who underwent $^{18}$F-FDG PET/CT before treatment were retrospectively enrolled at the Fourth Hospital of Hebei Medical University (Shijiazhuang, China) between January 2013 and September 2022, including 87 men and 31 women, aged 22–85 years with an average age of 62.50 ± 11.66 years. The enrolled patients met the following criteria: (1) GC was confirmed by pathology; (2) $^{18}$F-FDG PET/CT scan was performed one week before treatment; (3) HER2 expression levels were detected by IHC and FISH if required; and (4) the clinical data were complete. The exclusion criteria were as follows: (1) the patient had received treatment prior to PET/CT examination, and (2) the lesion had no FDG hypermetabolism on PET/CT images. The clinical and imaging data of each patient were collected. Clinical data included sex, age, pathological type, tumor location, lymph node metastasis, and distant metastasis. These patients were randomly divided into a training set (n = 82) and a test set (n = 36) at a ratio of 7:3.

**Detection of HER2 expression status**
Histological samples for HER2 detection were obtained by surgical resection or gastroscopic biopsy, and the expression level of HER2 was detected by IHC (and FISH, if required). Positive HER2 expression was defined as 3+ on IHC or 2+ on IHC with FISH. Negative HER2 expression was defined as 0, 1+ on IHC or 2+ on IHC and FISH negative [15].

**PET/CT image acquisition**

All patients underwent 18F-FDG PET/CT examination one week before treatment. The PHILIPS GEMINI GXL16 and Vereos PET/CT scanners were used in this study. The radiochemical purity of 18F-FDG was >95%. Patients fasted for at least 6 h before injection with 3.70–5.55 MBq/kg 18F-FDG, and the fasting blood glucose concentration of the patients was controlled below 11.1 mmol/L before the PET/CT examination. PET/CT acquisition was performed from the skull base to the upper femur 50–60 minutes after 18F-FDG injection. The attenuation of PET images was corrected using CT images.

**Radiomics feature extraction**

Radiomics features of PET/CT images in the region of interest (ROI) of gastric lesions were extracted the LIFEx 7.1.0 software package [16]. The image data of all patients were burned by CD. PET/CT images were imported into software using Digital Imaging and Communications in Medicine (DICOM). Two well-experienced nuclear medicine physicians manually delineated the ROI layer by layer on the cross-section of the image using three-dimensional (3D) drawing tools, and 40% of the maximum standardized uptake value (SUVmax) was used as a threshold to optimize the VOI. The spatial resampling interval of PET and CT images was 1 mm. Radiomics features were extracted from the two types of images within the same VOI because of the good matching of the PET and CT images. LIFEx software automatically extracted 134 radiomics features: 67 based on PET and 67 based on CT.

**Radiomics feature screening and model building**

In the training set, the optimum features were screened from 134 radiomics features to build Logistic regression (LR) model, Random forest (RF) model and Support vector machine(SVM) models. LR is a linear model without error terms, which is similar to multiple linear regression analysis. LR effectively avoids the problem of inaccurate assumptions of data distribution, but is prone to problems such as insufficient assembly and low classification accuracy. In addition, this method is not suitable when the data feature is missing or the feature space is large. The SVM is a supervised learning method that is widely used in classification and regression analyses. RF is an integrated supervised learning method that consists of multiple decision trees corresponding to different sub-datasets [17]. The predicted performances of the three models were validated using the test set. In addition, the prediction performance of the three models was evaluated by comparing the predicted probabilities with the actual probabilities, and a calibration curve was plotted.

**Statistical analysis**

Quantitative data that conformed to a normal distribution are expressed as $x \pm s$, and quantitative data that did not conform to the normal distribution are expressed as $M (P25, P75)$. Independent sample t-
tests and Mann-Whitney U tests were used to compare continuous variables (age), and chi-square test was used to compare categorical variables (sex, lesion location, pathological type, lymph node metastasis, and distant metastasis). SPSS 26.0 Statistics software was used for the above analysis.

Mann-Whitney U tests were used to screen out the different features between the HER2-positive and HER2-negative groups in the training set and the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm and 10-fold cross-validation screened the best features, which were used to establish three prediction models. The predictive performance of the models was validated using the test set. Receiver operating characteristic (ROC) curves, area under the curve (AUC), sensitivity, specificity, and accuracy were used to evaluate the predictive performances of the three models. The calibration curve of the model was plotted to evaluate the accuracy of the model. A Synthetic Minority Oversampling Technique (SMOTE) was used to balance the data. The above analysis was performed using Python software (version 3.8.8).

All $P$ values $< 0.05$ were considered statistically significant.

**Results**

**Clinical characteristics of patients**

Among the 118 patients with GC, there were no significant differences in age ($t = -1.540, P = 0.126$), sex ($\chi^2 = 0.020, P = 0.887$), tumor location ($\chi^2 = 0.358, P = 1.000$), pathological type ($\chi^2 = 3.270, P = 0.590$), lymph node metastasis ($\chi^2 = 0.020, P = 0.887$), or distant metastasis between the training and test sets. In terms of age, sex, tumor location, pathological type, lymph node metastasis, and distant metastasis, no significant differences were noted between the HER2-positive and HER2-negative groups in the training and test sets. Table 1 summarized the basic clinical characteristics of the patients with GC.
Table 1
Distribution of clinical characteristics of GC patients in the training and test sets

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Training set (n = 82)</th>
<th>P</th>
<th>Test set (n = 36)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HER2 negative (n = 68)</td>
<td></td>
<td>HER2 positive (n = 14)</td>
<td></td>
</tr>
<tr>
<td>Age (mean ± SD) (years)</td>
<td>62.76 ± 11.31</td>
<td>0.182</td>
<td>60.87 ± 11.91</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td>HER2 negative (n = 30)</td>
<td></td>
<td>HER2 positive (n = 6)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>0.954</td>
<td></td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>50 (73.50%)</td>
<td></td>
<td>22 (73.30%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>18 (26.50%)</td>
<td></td>
<td>8 (26.70%)</td>
<td></td>
</tr>
<tr>
<td>Pathological type</td>
<td>0.664</td>
<td></td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>52 (76.50%)</td>
<td></td>
<td>22 (73.30%)</td>
<td></td>
</tr>
<tr>
<td>Mucinous adenocarcinoma</td>
<td>4 (5.90%)</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Signet ring cell carcinoma</td>
<td>1 (1.50%)</td>
<td></td>
<td>2 (6.70%)</td>
<td></td>
</tr>
<tr>
<td>Neuroendocrine carcinoma</td>
<td>1 (1.50%)</td>
<td></td>
<td>2 (6.70%)</td>
<td></td>
</tr>
<tr>
<td>Tubular adenocarcinoma</td>
<td>2 (2.90%)</td>
<td></td>
<td>3 (10.0%)</td>
<td></td>
</tr>
<tr>
<td>Mixed carcinoma</td>
<td>8 (11.80%)</td>
<td></td>
<td>1 (3.30%)</td>
<td></td>
</tr>
<tr>
<td>Tumor location</td>
<td>0.842</td>
<td></td>
<td>0.429</td>
<td></td>
</tr>
<tr>
<td>Proximal</td>
<td>40 (58.80%)</td>
<td></td>
<td>18 (60.00%)</td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>11 (16.20%)</td>
<td></td>
<td>1 (3.30%)</td>
<td></td>
</tr>
<tr>
<td>Distal</td>
<td>15 (22.10%)</td>
<td></td>
<td>11 (36.70%)</td>
<td></td>
</tr>
<tr>
<td>Whole</td>
<td>2 (2.90%)</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Lymph node metastasis</td>
<td>0.382</td>
<td></td>
<td>0.079</td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>14 (79.40%)</td>
<td></td>
<td>12 (40.00%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>54 (79.40%)</td>
<td></td>
<td>18 (60.00%)</td>
<td></td>
</tr>
<tr>
<td>Distant metastasis</td>
<td>0.412</td>
<td></td>
<td>0.343</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>14 (79.40%)</td>
<td></td>
<td>22 (73.30%)</td>
<td></td>
</tr>
<tr>
<td>Characteristic</td>
<td>Training set(n = 82)</td>
<td></td>
<td>Test set(n = 36)</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------</td>
<td>---</td>
<td>-------------------</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>HER2 negative</td>
<td>HER2 positive</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>((n = 68))</td>
<td>((n = 14))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>54(79.40%)</td>
<td>4(28.60%)</td>
<td>8(26.70%)</td>
<td>3(50.00%)</td>
</tr>
</tbody>
</table>

Feature extraction and selection

The features described met the definition described by the Imaging Biomarker Standardization Initiative (IBSI) in our study [18]. There were 60 first-order, 10 shape, and 64 second-order statistical features. We first selected the features with significant differences between the HER2-positive and HER2-negative groups using the Mann-Whitney U test and obtained a total of 25 features with a \(P\) value < 0.05. Then, the LASSO algorithm and ten-fold cross-validation were used to further screen the optimal radiomics features from the 25 radiomics features. Six optimal features were selected: Conventional_SUVbwkurtosis, Conventional_SUVbwExcessKurtosis, Conventional_HUkurtosis, Conventional_HUExcessKurtosis, GLRLM_LRHGE, and GLRLM_RLNU. Among the six features, the first two were derived from PET, while the 4 remaining features were related to CT. The process of LASSO algorithm-filtering features is shown in Fig. 1.

Model construction and predictive performance of the three models

To evaluate the performance of radiomics features in predicting HER2 expression status in patients with GC, we constructed and validated three models based on the three machine learning methods (Fig. 2). In the training and test sets, the AUC values of the LR model used to predict the different expression status of HER2 in GC patients were 0.809 and 0.761, respectively, and the AUC values of the SVM model were 0.861 and 0.628, respectively. The AUC values of the RF model were 0.993 and 0.717, respectively. In terms of AUC, there were significant differences between the RF and SVM models in the training set \((P = 0.023)\). In the test set, the difference in the AUC between the LR and SVM models was also statistically significant \((P = 0.049)\). There was no significant difference between the LR and RF models in the training \((P = 0.068)\) and test \((P = 0.194)\) sets. The LR and RF models perform better than the SVM model. The predictive abilities of the three models are presented in Table 2.
### Table 2
Comparison of the three models before and after SMOTE

<table>
<thead>
<tr>
<th>Model</th>
<th>SMOTE</th>
<th>Training set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AUC</td>
<td>AUC</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>no</td>
<td>0.809</td>
<td>0.761</td>
</tr>
<tr>
<td></td>
<td>SMOTE</td>
<td>0.900</td>
<td>0.916</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>no</td>
<td>0.861</td>
<td>0.628</td>
</tr>
<tr>
<td></td>
<td>SMOTE</td>
<td>0.952</td>
<td>0.876</td>
</tr>
<tr>
<td>Random forest</td>
<td>no</td>
<td>0.993</td>
<td>0.717</td>
</tr>
<tr>
<td></td>
<td>SMOTE</td>
<td>1.000</td>
<td>0.941</td>
</tr>
</tbody>
</table>

### Calibration curve

A good model should not only be accurate, but also well calibrated. A calibration curve was used to assess the consistency between the model prediction probability and actual event rates. Ideally, the calibration curve should be as close to the diagonal as possible, meaning that the predicted event rate is more realistic. A well-calibrated classifier is a probabilistic classifier whose output of the predicted probability method can be interpreted directly as the confidence level. Calibration can be graphically evaluated by the predictions on the x-axis and the results on the y-axis. The Brier score is a quadratic scoring function that calculates the square variance between the actual binary result and prediction. The Brier scores ranged from 0 (perfect model) to 1. In the test set, the Brier scores were 0.118, 0.137, and 0.214 in the LR, RF, and SVM models, respectively. We can see that the error between the predicted and real values is smaller in the LR and RF models than in the SVM model. (Fig. 3).

### Oversampled balance data

SMOTE is an improvement over a random oversampling algorithm that artificially synthesizes new samples based on a few oversampling samples and generates new datasets [19]. When the ratio of negative to positive samples is close to 1:1, the bias of large samples can be effectively avoided [20]. Before using SMOTE to balance the data in this study, 118 samples were included, and the proportion of negative (98 cases) to positive (20 cases) was approximately 5:1. To achieve better prediction, the SMOTE method was used in this study to deal with the imbalance of data samples; 196 samples were included in total (negative: 98 cases, positive: 98 cases). Based on the training set, 12 optimal image features (PET: 5; CT: 7) were screened using the LASSO algorithm. The remaining methods were the same as those described above to construct the three prediction models (Fig. 4). After using the oversampling method, the AUC of each model significantly improved. The AUC values in the training set (LR, 0.900; SVM, 0.952; RF, 1.000) and those in the test set (LR, 0.916; SVM, 0.876; RF, 0.941) showed that the RF model
had the best performance. The AUC, accuracy, sensitivity, and specificity of the three models before and after using the oversampling technique are presented in Table 2. At the same time, calibration curve analysis showed that the LR and RF models had a higher consistency between the predicted probability and actual event incidence (Brier score: 0.114 and 0.106, respectively) when compared with the SVM model (Brier score: 0.132). The calibration curve is shown in Fig. 5.

Discussion

In recent years, targeted therapy has been increasingly applied to cancer treatment. Many studies have shown that HER2 overexpression is associated with poor prognosis and low survival rate in patients with gastric cancer both before and after surgery, and such patients may benefit from HER2-targeted therapy[21, 22]. In addition, some scholars have confirmed that HER2 overexpression is not only related to the prognosis of patients with gastric cancer but is also closely related to the occurrence and development of gastric cancer [23]. Therefore, it can be seen that the expression level of HER2 is crucial in guiding the choice of individual treatment strategies in GC patients. However, not all patients are able to detect HER2 expression status due to the cost and equipment of IHC examination in daily clinical practice. Some lesions could not be sampled because of their location. In addition, the tumor heterogeneity of gastric cancer is relatively high, with inconsistencies in tissue, time, and space, which can be as high as 5–30% [24]. For example, different biopsy sites in the same lesion can lead to different HER2 expression states. The expression status of HER2 in primary and metastatic lesions is sometimes not identical. The HER2 status can be partially altered before and after treatment for gastric cancer. Moreover, the detection of HER2 by conventional IHC method is affected by different pathologists. Therefore, there is an urgent need to explore a new method that can accurately, standardizedly, noninvasively, and dynamically evaluate HER2 expression in patients with gastric cancer. $^{18}$F/$^{68}$Ga-HER2 antibody PET/CT can detect HER2 expression in malignant tumors, however, it has not been widely applied in clinical practice to date.

In this study, we established and validated three predictive models of HER2 expression status in patients with GC based on $^{18}$F-FDG PET/CT radiomics features before treatment. It could be seen that among the three radiomics models, the LR and RF models had a better performance in predicting the expression status of HER2 in GC patients and also performed better in classification accuracy. After the data were balanced by SMOTE, the AUC of the three models increased in the training and test sets. All had good predictive performance, and the RF model showed the best performance. The results showed that the predictive models of HER2 expression status in GC patients based on $^{18}$F-FDG PET/CT radiomics features had relatively high sensitivity, specificity, and accuracy, and had potential value in clinical use.

Among many tumor markers, HER2 status is one of the most important factors in the clinical diagnosis and treatment of gastric cancer. At present, many radiomics studies on HER2 have made progress in breast cancer, esophageal cancer, and lung cancer [25]. Currently, the main applications of machine learning technology in tumor imaging include two categories: radiomics and convolutional neural networks [26]. convolutional neural network is prone to overfitting when making disease classification
predictions when applied to a small sample size, resulting in poor generalization ability of the model [27]. However, in clinical practice, the sample size of complete medical image data is relatively small, and the use of convolutional neural network is limited to some extent [28]. Therefore, in this study, radiomics was used to construct the model. Reviewing relevant studies, we found that the sample size limit of radiomics is relatively low, and the performance of the model is better. Li et al. [29] combined a radiomics model based on CT image features with the level of Carcinoembryonic Antigen (CEA) and obtained the best model to predict HER2 expression in GC patients; the AUC values of the best model in the training and test sets were 0.799 and 0.771, respectively. Similarly, Wang et al. [30] conducted a retrospective analysis of 101 patients with adenocarcinoma of the esophagogastric junction and obtained 7 optimal features based on portal-phase CT images. The Nomogram, which consisted of the best features with T-staging from CT, also had very good predictive performance for HER2 status (AUC, 0.946 and 0.903 in the training and test sets, respectively). This is also similar to the predictive performance of the composite model constructed by Ma et al. [31] based on three stage CT images of 745 GC patients (AUC, 0.85 and 0.84 in the training and test sets, respectively). In addition, Wang et al. [32] retrospectively analyzed CT images of 132 GC patients and established a RF model to predict HER2 expression in GC patients. The model was built based on the radiomics features of arterial-phase CT, and the AUC values were 0.756 and 0.830 in the training and test sets, respectively. However, the above studies were mainly based on CT images, and PET images were not included in the study; therefore, the prediction model based on CT images has limited clinical application. In this study, three radiomics models based on $^{18}$F-FDG PET/CT imaging were constructed, and the results demonstrated that the three radiomics models had good predictive performance for HER2 status, especially the LR and RF models. The clinical variables are not included in the construction of the prediction model because the research results show that clinical variables such as gender and age do not have the ability to predict the expression of HER2.

Several studies have shown that $^{18}$F-FDG PET/CT plays an important role in predicting HER2 expression. Chen et al. [33] reviewed the $^{18}$F-FDG PET/CT images of 64 patients with gastric cancer before surgery. The results showed that the SUVmax of lesions in patients with HER2-positive gastric adenocarcinoma was significantly higher than that in HER2-negative lesions (SUVmax: 8.619 ± 5.878 vs 2.600 ± 2.036). This is similar to the results of Bai et al. [34], where the SUVmax was significantly different between HER2-positive and HER2-negative gastric cancer lesions. With the development of artificial intelligence, the applications of radiomics, machine learning, and deep learning in the medical field have become increasingly extensive and mature. At present, research on gastric cancer based on $^{18}$F-FDG PET/CT imaging has penetrated many aspects, such as the diagnosis and differential diagnosis of gastric cancer, staging, lymph node metastasis, and prognosis. However, no studies have reported on the prediction of HER2 expression in gastric cancer. In this study, PET imaging features were included to predict HER2 expression in the patients with gastric cancer. In addition, it can be seen from the study that for different sample proportions, the three machine learning models have their own advantages.

Radiomics features are usually divided into three categories: shape, first-order, and second-order features. Among the six best features selected in this study, SUVbwKurtosis and SUVbwExcessKurtosis are
radiomics features based on PET images, reflecting cell metabolic activities. Kurtosis, also known as the peak coefficient, represents the peak of the probability density distribution curve at the mean value. Intuitively, the kurtosis reflects the cusp of the peak. Multiple studies have shown that kurtosis, as a texture feature, has high application value in cancer differentiation, tumor heterogeneity differentiation, and tumor staging [35, 36]. In this study, HER2-positive patients had higher kurtosis values than HER2-negative patients, which may be related to the heterogeneity of GC. The remaining four features were all based on the CT images. GLRLM is a basic feature of CT, which is obtained by recording the occurrence of multiple consecutive same pixel values in a one-dimensional direction in the image. A larger length distribution indicates a shorter run length and finer texture. GLRLM_LRE is a method for measuring the run-length distribution, with larger values indicating longer run lengths and coarser structural textures. GLRLM_GLN is always used to measure the intensity value of the image, and a lower GLN value is related to a higher intensity value of the image. GLRLM_RLN measures the similarity of the run lengths in the entire image, and a lower value indicates more uniform run lengths in the image. GLRLM_RLNU describes the length of the grayscale level inhomogeneity run, with lower values representing more uniform run lengths in the image. Lee et al. [37] divided thymus epithelial tumors (TETs) into three subgroups (low-risk thymoma, high-risk thymoma, and thymoma) based on pathological results, which showed differences in texture heterogeneity on PET/CT images. This study showed that most features derived from GLRLM showed good differentiation ability. Similarly, Kunimatsu et al. [38] compared glioblastoma and primary central nervous system lymphoma using texture analysis based on MRI images, and the results also proved that GLRLM features had good distinguishing performance.

In this study, PET/CT image data were obtained using two instruments with different parameters. A few studies have found that the radiomics features change with different parameters such as reconstruction layer thickness and algorithm, but a number of studies in recent years have shown that there is no significant statistical difference between the radiomics features extracted from two different devices [39, 40].

In conclusion, this study shows that $^{18}$F-FDG PET/CT radiomics features before treatment are of great value in predicting HER2 expression status and provide a noninvasive and dynamic detection method to determine HER2 expression levels. This has important significance in guiding the clinical formulation of individualized treatment schemes and in evaluating and predicting the efficacy of HER2-targeted therapy.

Declarations

Ethical approval and consent to participate

All procedures performed in the study and involving human participants were carried out in accordance with the ethical standards of the institutional and/or national research committee and with the principles of the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.
This retrospective analysis was approved by the Institutional Review Board of the Fourth Hospital of Hebei Medical University ( Approval No. 2020KY116).

The written informed consent was obtained from all participants.

Consent for publication

Not applicable.

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Hebei Provincial Department of Science and Technology in China (N0.20377728D)

Authors’ contributions

XJ and TL were involved in drafting and final editing the report. They contributed equally to this work and share first authorship; ZZ and JW segmented gastric cancer on PET/CT scan; MD, JH and XC conducted statistical analysis; JZ and XZ, the corresponding authors, designed the study and supervised the entire project and work. All authors critically reviewed and approved the final version of the manuscript.

Acknowledgments

The authors thank all their coworkers involved in the study for their support and assistance.

References


**Figures**
Figure 1

The LASSO algorithm and tenfold cross-validation were used to extract the optimal subgroup of radiomic features. a The vertical line indicates the value of the characteristic parameters corresponding to the minimum of $\lambda$ in cross-validation. b The penalty diagram of the characteristic coefficient of radiomics.

Figure 2
ROC curves for LR model, SVM model, RF model in predicting different expression status of HER2. a The ROC curve of the training set. b The ROC curve of the test set.

Figure 3

Calibration curves of three prediction models.
**Figure 4**

ROC curves for three models in predicting different expression status of HER2 after SMOTE. 

- **a** The ROC curve of the training set.
- **b** The ROC curve of the test set.
Figure 5

Calibration curves of three prediction models after SMOTE.