Name of Journal: World Journal of Gastrointestinal Oncology
Manuscript NO: 73090
Manuscript Type: ORIGINAL ARTICLE

Retrospective Study
Preoperative prediction of malignant potential of 2-5 cm gastric gastrointestinal stromal tumors by computerized tomography-based radiomics

Sun XF et al. Predicting malignant potential of gastric GISTs

Xue-Feng Sun, Hai-Tao Zhu, Wan-Ying Ji, Xiao-Yan Zhang, Xiao-Ting Li, Lei Tang, Ying-Shi Sun
Abstract
BACKGROUND
The use of endoscopic surgery in treating gastrointestinal stromal tumors (GISTs) between 2 and 5 cm remains controversial considering the potential risk of metastasis and recurrence. Also, surgeons are facing great difficulties and challenges in assessing the malignant potential of 2-5 cm gastric GISTs.

AIM
To develop and evaluate a computerized tomography (CT)-based radiomics for predicting the malignant potential of primary 2-5 cm gastric GISTs.

METHODS
A total of 103 patients with pathologically confirmed gastric GISTs between 2-5 cm were enrolled. The malignant potential was categorized into low grade and high grade according to postoperative pathology results. Preoperative CT images were reviewed by two radiologists. A radiological model was constructed by CT findings and clinical characteristics using logistic regression. Radiomic features were extracted from the preoperative contrast-enhanced CT images in the arterial phase. XGboost method was used to construct a radiomics model for the prediction of malignant potential. Nomogram was established by combing the radiomics score with CT findings. All the models were developed in a training group (n = 69) and evaluated in a test group (n = 34).

RESULTS
The area under curve (AUC) value of the radiological model, radiomics model and nomogram model was 0.753 (95%CI: 0.597-0.909), 0.919 (95%CI: 0.828-1.000) and 0.916 (95%CI: 0.801-1.000) in the training group vs. 0.642 (95%CI: 0.379-0.870), 0.881 (95%CI: 0.772-0.990) and 0.894 (95%CI: 0.773-1.000) in the test group, respectively. The AUC of the nomogram model was significantly larger than that of the radiological model in
both the training group \((Z = 2.795, P = 0.0052)\) and the test group \((Z = 2.785, P = 0.0054)\). The decision curve of analysis showed that the nomogram model produced increased benefit across the entire risk threshold range.

CONCLUSION
Radiomics may be used as an effective tool to predict the malignant potential of 2-5 cm gastric GISTs and assist preoperative clinical decision-making.

**Key Words:** Gastrointestinal stromal tumors; Gastric gastrointestinal stromal tumors; Computed tomography; Malignant potential; Radiomics; Nomogram

Sun XF, Zhu HT, Ji WY, Zhang XY, Li XT, Tang L, Sun YS. Preoperative prediction of malignant potential of 2-5 cm gastric gastrointestinal stromal tumors by computerized tomography-based radiomics. *World J Gastrointest Oncol* 2022; In press

**Core Tip:** The use of endoscopic surgery in gastrointestinal stromal tumors (GISTs) between 2 and 5 cm remains controversial considering the potential risk of metastasis and recurrence. Also, surgeons are facing great difficulties and challenges in assessing the malignant potential of 2-5 cm gastric GISTs. This study aimed to develop and evaluate a computerized tomography-based radiomics for predicting the malignant potential of primary 2-5 cm gastric GISTs.

**INTRODUCTION**
Gastrointestinal stromal tumors (GISTs) are the most common mesenchymal tumors in the gastrointestinal tract and account for the majority of submucosal tumors\(^{[1,2]}\). They most frequently appear in the stomach (50%-60%) and small intestine (30%-35%) and rarely in the colorectum (5%) and esophagus (≤ 1%)\(^{[3,4]}\). GISTs are clinically heterogeneous with varying degrees of malignant potential. Therefore, preoperative
evaluation of the biological behavior of GISTs is important for surgical decision-making[1,3].

Endoscopic resection is an effective and safe method for treating gastric GISTs smaller than 2 cm[6-8]. Nevertheless, whether endoscopic surgery can be used in resecting gastric GISTs between 2 and 5 cm remains controversial considering the potential risk of metastasis and recurrence[6,9]. Also, surgeons are facing great difficulties and challenges in assessing the malignant potential of 2-5 cm gastric GISTs.

The frequency of 2 to 5 cm gastric GISTs metastases with mitotic counts larger than 5/50 high power fields (HPFs) and smaller than 5/50 HPFs are 16% and 1.9%, respectively[10]. Based on mitotic counts, several risk stratification systems have been proposed to assess the recurrence risk after complete resection of primary GISTs[10-12]. Gastric GISTs are generally associated with a better prognosis than nongastric GISTs[10]. The modified National Institutes of Health (NIH) criteria classify GISTs into four categories (very low, low, intermediate, and high risk) according to tumor size, mitotic count, tumor size, and tumor rupture. The modified NIH criteria have become a commonly accepted risk stratification tool for GISTs due to their important value in assessing prognosis after operation[13-15]. However, these criteria are only postoperatively applied as the mitosis count of the specimen available after excision is a significant criterion factor.

Preoperative prediction of the malignant potential and prognosis of these GISTs is crucial for clinical decision-making. Preoperative biopsy is a common method for determining the characteristics of suspected lesions. Yet, this method has several disadvantages, such as the lack of adequate tissue for fine-needle biopsy, the possible failure to obtain mitosis counts with improper sampling, or the underestimation of mitotic grades, which increases the difficulty of response evaluation during follow-up. On the other hand, with the recent remarkable development of imaging technology, non-invasive real-time imaging tools, such as computerized tomography (CT), magnetic resonance imaging, and endoscopic ultrasound (EUS), have been increasingly applied for assessing the potential malignancy and prognosis of a variety of tumors, including
GISTs. For example, Chen et al.\textsuperscript{[13]} have indicated that CT features were more useful than EUS features for predicting mitotic counts. Therefore, exploring the association between CT features and GISTs risk stratification may influence the surgical treatment decision for 2-5 cm gastric GISTs. Nevertheless, subjective assessments may overlook abundant information hidden in the images and may be limited by overreliance on observers’ experience.

As a combination of texture analysis and machine learning methods, radiomics has been widely used in the field of assisted tumor diagnosis, staging, and prognosis prediction\textsuperscript{[16,17]}. Many studies have indicated that radiomics features can be used to comprehensively assess the biological behavior of malignant cells, improving the accuracy of diagnosis, prognosis, and prediction\textsuperscript{[18-20]}. Radiomics has also been used to preoperatively predict the malignant potential of GISTs\textsuperscript{[21]}. However, the study on 2-5 cm gastric GISTs has not yet been reported.

The aim of the current study was to propose a radiomics model for predicting the malignant potential of 2-5 cm gastric GISTs by preoperative enhanced CT images. The method may be helpful for preoperative design of individualized treatment strategy for patients with 2-5 cm gastric GISTs.

**MATERIALS AND METHODS**

*Subjects*

This retrospective study was approved by our institutional review board and patient’s informed consent was waived. A total of 695 gastric GISTs patients with histologically confirmed 2-5 cm gastric GISTs who were treated at our hospital were consecutively enrolled between January 2010 and December 2019. The inclusion criteria were the following: (1) Patients who underwent surgery for primary gastric GISTs with curative intent; (2) patients who underwent standard contrast-enhanced CT less than 15 d before surgical resection; (3) patients with complete clinicopathologic data; and (4) patients with a tumor size of 2-5 cm.
The exclusion criteria were: (1) Patients who received imatinib therapy or other tyrosine kinase inhibitor as a neoadjuvant therapy before surgery; and (2) patients who had tumor rupture before or during surgery.

Finally, 592 patients were excluded due to the above reasons, and 103 patients were included in this study (48 males and 55 females; mean age, 58.31 ± 9.20 years). The included patients were randomly divided into a training group (n = 69) and a test group (n = 34) in a portion of 2:1 ratio with equal proportions of positive and negative samples. The inclusion and exclusion criteria are shown in Figure 1.

**CT imaging**

Contrast-enhanced CT examinations were performed using one of the following CT scanners: GE LightSpeed VCT (n = 62) or GE Discovery CT750 HD (n = 41). All the patients were fasted for at least 8 h before the examination. They were given 6 g of gas production powder orally before the examination to ensure the adequate expansion of the gastric cavity. CT images were obtained during breath-holding. Both scanners used 5 mm slice thickness, 5 mm slice increment, 0.9 pitch, 120 kV tube voltage, and 300 mA tube current.

Contrast enhanced scanning was performed for all the subjects. They were intravenously administered with 70-100 mL of a nonionic contrast agent (iohexol, 300 mg I/L; General Electric) at a rate of 2.5-3.5 mL/s. For the arterial phase, a delay time of 30 s was used. Venous phase and delayed phase scanning were performed 60 and 120 s after the contrast agent injection.

Axial, sagittal, and coronal multiplanar reconstructions images were obtained with a reconstruction thickness of 2.5 mm. CT images were sent to the picture archiving and communication system (PACS) for interpretation at the workstations.

**CT findings and radiological model**

Two radiologists with 14 and 5 years of experience in abdominal imaging independently reviewed all images. In case of disagreement, the two readers jointly
reviewed the findings to reach a consensus for further analysis. The radiologists were blinded to the pathological data.

The following CT findings were recorded: tumor size (cm), location (cardiac region, fundus, body, or antrum), necrosis (present or absent), ulceration (present or absent), calcification (present or absent), growth pattern, tumor contour (irregular or regular) and tumor margin (poorly or well-defined). Tumor size was defined as the maximal diameter on the transverse, coronal, or sagittal plane. Ulceration was defined as a focal mucosal defect/indentation filled with air or fluid or when contrast material was found on the endoluminal surface of the lesion. Growth patterns were classified as endoluminal, exophytic, or mixed. The tumor contour was considered as either regular/round/ovoid or irregular/lobulated. The mean CT value (Hounsfield unit) was measured in the plain phase, arterial phase, venous phase, and delayed phase. Univariate analysis was used to select useful CT findings. A radiological model was constructed by the selected CT findings using backward logistic regression.

**Tumor delineation**

The regions of GISTs were manually delineated by a junior radiologist (with 5 years of experience in abdominal imaging diagnosis) with 3D Slicer (version 4.8.1) in the axial direction. A senior radiologist (with 14 years of experience) evaluated the delineations and made modifications if needed. Delineation was performed on each slice of CT images from the artery phase so as to cover the whole tumor. Both radiologists were blinded to the risk classification of patients. One example is shown in Figure 2.

**Feature extraction**

Pyradiomic (version 3.0.1) was used to extract 851 features from the region of interest (ROI), including 14 shape features, 18 first-order features, 75 s-order (texture) features (24 gray level co-occurrence matrix features, 14 gray level dependence matrix features, 16 gray level run length matrix features, 16 gray level size zone matrix features, 5
neighboring gray-tone difference matrix features), and their 8 kinds of wavelet transforms \([(18 + 75) \times 8 + 18 + 75 + 14 = 851]\).

**Low-grade and high-grade malignant potential**

According to the NIH-modified criteria,\(^{11}\) mitotic counts > 5/50 HPFs were categorized into high grade, and mitotic counts < 5/50 HPFs were categorized into low grade. Patients were then divided into the very low/Low-risk group (low-grade malignant potential group, \(n = 82\)) and the moderate/high-risk group (high-grade malignant potential group, \(n = 21\)). Low-grade was labeled 0, and high-grade was labeled 1 as the ground truth for training and test.

**Radiomics model**

First, a t-test examination was performed to compare all the features between the high-grade and low-grade groups. The features with a \(P\) value > 0.05 were removed. Second, the correlation was calculated between each pair of the features. If the absolute value of correlation was > 0.5, the feature with a smaller T value in the t-test was removed. Third, the XGboost algorithm was used to construct a model with remaining features and ground truth.

Due to the small sample size, the maximum estimator number and the maximum depth were set to 3 to avoid overfitting. A 3-fold cross-validation was used to determine the optimal tree number and depth. After cross-validation, the whole training group was trained again by the fixed hyperparameter to obtain the predictive model. A radiomics score was generated by the model for each patient. Finally, the model was assessed in the test group.

**Nomogram model**

Logistic regression was performed in the training group to classify the high-grade and low-grade by combining radiomics scores with CT findings. Nomogram was used to
visualize the combination of radiomics score and the selected CT findings. A risk score was generated by the nomogram and evaluated in the test group.

Decision curve of analysis
Decision curve of analysis (DCA) was performed to study the benefit of radiomics model. Net benefit was calculated by subtracting the proportion of all patients who were false positive from the proportion of those who were true positive, weighted by the relative harm of forgoing treatment compared with the negative consequences of unnecessary treatments. Standardized net benefit scaled the net benefit into the range between 0 and 1. The relative harm was the ratio of the harm of false positive harm to false negative harm.

Statistical analysis
Independent sample t-test was used to compare the continuous variables in the low and high malignant potential groups. Chi-squared tests or Fisher’s exact tests were applied for categorical variables. Receiver operating characteristic (ROC) curves were used to evaluate the predictive model. The cutoff value between low-grade and high-grade was selected by maximizing the Youden index (sensitivity + specificity-1). The area under curve (AUC) was compared by the DeLong method.

RESULTS
Patients characteristics
The clinical characteristics and CT findings between the low-grade and high-grade malignant potential groups are analyzed in Table 1. In the univariate analysis, tumor diameter, necrosis, ulceration, tumor margin, and tumor contour significantly differed between the different risk stratification groups (all $P < 0.05$). No significant differences were found in other subjective features between the two groups, including tumor location, growth pattern, calcification, density, and the degree of enhancement in each phase of CT between the different risk stratification groups (all $P \geq 0.05$).
compares the basic characteristics between the training and the test group. Moreover, there was no significant difference in age, gender, and ground truth between the two groups.

**Prediction by radiological model**

A radiological model was constructed by backward logistic regression using 5 selected CT findings including tumor diameter, necrosis, ulceration, tumor margin, and tumor contour. Two features were retained in the final model, including the largest diameter \((P = 0.032; \text{OR} = 1.082, 95\% \text{CI}: 1.007-1.163)\) and ulceration \((P = 0.061; \text{OR} = 3.618, 95\% \text{CI}: 0.943-13.876)\). The performance of this radiological model is summarized in Table 3. The AUC value was 0.753 (95\% CI: 0.597-0.909) for the training group and 0.642 (95\% CI: 0.379-0.870) for the test group.

**Prediction by radiomics model**

After the removal of features via \(t\)-test and correlation, 13 features remained. XGboost method selected 4 features by 3-fold cross-validation with an optimal learning rate of 0.03. The 4 selected features and their importance were: gray-level-non-uniformity (wavelet-HHH glszm feature type) with an importance of 0.703, mean-absolute-deviation (wavelet-HHH first-order feature type) with an importance of 0.154, small-dependence-low-gray-level-emphasis (wavelet-LHH gldm feature type) with an importance of 0.098, and maximum (wavelet-LHL_firstorder) with an importance of 0.045. Figure 3 shows the 2 trees (estimators) for classification. The radiomics score is the summation of the scores from the 2 trees. The prediction results by radiomics score are summarized in Table 3. The AUC of the prediction by radiomics model was 0.919 (95\% CI: 0.828-1.000) for the training group and 0.881 (95\% CI: 0.772-0.990) for the test group.

**Prediction by nomogram model**
Three CT findings were selected by linear regression to combine with the radiomics score above, including necrosis, calcification, and ulcer. Nomogram was plotted as shown in Figure 4. The prediction result by the risk calculated from the nomogram is also summarized in Table 3. The AUC predicted by the nomogram model was 0.916 (95%CI: 0.801-1.000) for the training group and 0.894 (95%CI: 0.773-1.000) for the test group. The ROC curves of the radiological model, radiomics model, and nomogram model were plotted as shown in Figure 5. The AUC of the nomogram model was significantly larger than that of the radiological model in both the training group (Z = 2.795, \( P = 0.0052 \)) and the test group (Z = 2.785, \( P = 0.0054 \)).

**DCA**

Figure 6 shows the result of DCA. The y-axis measured the net benefit. The red line represents the prediction by the nomogram model. The blue line represents the assumption that all patients have high-grade malignant potential GISTs. The horizontal green line represents the assumption that all patients have low-grade malignant potential GTSTs. A 95% confidence interval (dashed line) was determined by 1000 bootstraps. The results showed that the nomogram model produced increased benefit across the whole risk threshold range.

**DISCUSSION**

GISTs initiate from very early forms of Cajal cells in the gastrointestinal tract wall[22]. GISTs have complex and unpredictable biological behavior, with KIT or PDGFRA being the main pathogenetic pathways[23]. Up-to-date clinical practice guidelines suggest that the standard treatment for localized GISTs is complete surgical excision. R0 excision (microscopically negative margins) is the goal, especially for patients with a high risk of recurrence. According to recent studies, when surgery is technically challenging (rectum, duodenum, and GEJ surgeries) and preoperative cytoreduction may facilitate tumor R0 excision, preoperative imatinib should be considered. Imatinib is currently the first-line molecular targeted drug for the treatment of GISTs, and can be used in
combination with KIT and PDGFRA\textsuperscript{[24]}. The current guidelines recommend more than three years of adjuvant treatment for high-risk GISTs patients\textsuperscript{[25]}. Patients with low malignant potential (low and very low risk) generally have a good prognosis and do not require further adjuvant imatinib therapy\textsuperscript{[26-28]}. The majority of GISTs < 2 cm usually have risk of metastasis and their mitotic counts are < 5 per 50 HPFs in general. Conversely, for GISTs between 2.5 cm, there is a ten-fold difference in metastasis frequency between low-mitosis and high-mitosis groups\textsuperscript{[10]}. According to the current diagnosis and treatment paradigm, individualized preoperative prediction of recurrence is particularly important for 2.5 cm GISTs. While the modified NIH consensus criteria are frequently used to estimate the risk of recurrence, the key criteria are only postoperatively accessible. A biopsy may provide preoperative estimation. However, a core needle biopsy may not provide an accurate mitotic count and a full-scale malignant potential assessment of the tumor. Therefore, a new robust risk assessment standard is needed.

Contrast-enhanced CT is the standard imaging method for the pretreatment and follow-up evaluation of GISTs. Several studies have investigated the predictive value of multiple CT findings for the malignant potential of GISTs\textsuperscript{[13,29-31]}. The results varied, possibly due to the different inclusion criteria and subjective assessment standards. A previous study noted that CT findings were predictors of risk stratification for GISTs\textsuperscript{[29]}. In the present study, univariate analysis revealed that high-grade malignant potential tumors tended to have an irregular shape, indistinct tumor margins, necrosis, and ulceration, consistent with previous studies\textsuperscript{[30,32]}. Our results also showed that high-grade malignant potential tumors frequently displayed tumors with a larger size. Tateishi \textit{et al}\textsuperscript{[33]} reported that an extrinsic epicenter and an unclear border were the most significant predictors for high-grade tumors, according to multiple stepwise logistic regression analysis. In our series, tumor size, shape, margins, the presence of necrosis, and ulceration were statistically significant factors for risk stratification of 2.5 cm gastric GISTs in the univariate analysis. Nevertheless, our radiological model showed that only the largest diameter and the presence of ulceration were independent predictors in
backward logistic regression for high malignant potential. Limited by the inadequate predictive power of subjective CT findings, the AUC of the radiological model (0.642 for the test group) was unsatisfactory for clinical application.

Compared with subjective CT findings, both our radiomics and nomogram models had greater predictive power, as indicated by higher AUC values. Significant AUC difference was found between the radiological model and nomogram model despite a small test sample. This demonstrated that our radiomics approach with quantitative analysis had an advantage over the subjective CT findings. Unlike the radiomics models proposed by Chen et al\[21\], the present study focused on the GISTs with the largest diameter of 2-5 cm. According to the modified NIH criteria, the risk stratification of gastric GISTs is mainly based on the size of the tumor and mitotic count. GISTs larger than 5 cm tend to be classified into the high-risk group. It is more challenging to predict the potential risk of smaller GISTs. Therefore, it is clinically important to construct a prediction model, especially for the 2-5 cm GISTs. In this study, the ground truth of risk was determined only by mitotic counts. Mitotic counts > 5/50 HPFs were categorized into high-grade malignant potential, and mitotic counts < 5/50 HPFs were categorized into low-grade malignant potential. Therefore, the impact of tumor size was excluded, which was reasonable because 2-5 cm GISTs tended to have a uniform tumor size. In this study, although the largest diameter showed a significant difference in t-test examination and was included in the radiological mode, the CT findings were not selected in the final nomogram model. This indicated that tumor size was not crucial for predicting the potential risk for 2-5 cm GISTs. It is possible that manual measurement of 2-5 cm GISTs on CT images was relatively unstable compared with the quantitative features from radiomics models.

In the radiomics model, four features were selected to construct the decision trees by XGboost. The feature with the largest importance showed the gray level nonuniformity from the gray-level size zone matrix. It was used as the root node for both two decision trees. A gray-level zone was defined as the number of connected voxels that share the same gray level intensity. Gray level nonuniformity measures the variability of gray-
level intensity values in the image, with a lower value indicating more homogeneity in intensity values. Therefore, signal inhomogeneity inside the tumor region in the arterial phase of CT images is important for predicting the potential risk of 2-5 cm GISTs. Due to the small training samples, only 4 features and 2 trees with a depth of 2 were included in the radiomics model to avoid overfitting. The similar accuracy between the training and test group indicated a good fitting for both radiomics and nomogram models. In the nomogram, 3 CT findings were combined with the radiomics score to calculate the risk. This provides a simple way to incorporate the subjective findings with the result of machine learning. Although the presence of calcification was not selected in the t-test or logistic regression, it appeared useful in the nomogram. Probably, the mutual effect of calcification and radiomics score contributed to the improvement of the prediction accuracy.

The present study had some limitations. First, our data were collected retrospectively, so further prospective research was needed. Second, this study was a single-center study. Although two scanners were used, the scanning parameters were the same. Third, a relatively small sample size limited the complexity of machine learning models. In addition, we did not have information on whether the patients experienced recurrence or death due to the lack of long-term follow-up. Nevertheless, to the best of our knowledge, this was the first study that predicted the malignant potential of 2-5 cm gastric GISTs patients by radiomics. More cohort validation and more integrable factors such as KIT and PDGFRA mutations should be considered in future research\[3,33]\.

CONCLUSION

In this study, we developed a radiomics model and a nomogram to predict the malignant potential of 2-5 cm gastric GISTs. The models revealed more accurate predictive power compared to subjective CT findings.
1. Chang Li, Wenhao Fu, Li Huang, Yingqian Chen, Pei Xiang, Jian Guan, Canhui Sun. "A CT-based nomogram for predicting the malignant potential of primary gastric gastrointestinal stromal tumors preoperatively", Abdominal Radiology, 2021


3. Tao Chen, Lili Xu, Xiaoyu Dong, Yue Li, Jiang Yu, Wei Xiong, Guoxin Li. "The roles of CT and EUS in the preoperative evaluation of gastric gastrointestinal stromal tumors larger than 2 cm", European Radiology, 2019


5. link.springer.com

6. Huali Li, Gang Ren, Rong Cai, Jian Chen, Xiangru Wu, Jianxi Zhao. "A correlation research of Ki67 index, CT
<table>
<thead>
<tr>
<th>7</th>
<th>Yuanjing Gao, Yanwen Luo, Chenyang Zhao, Mengsu Xiao, Li Ma, Wenbo Li, Jing Qin, Qingli Zhu, Yuxin Jiang. &quot;Nomogram based on radiomics analysis of primary breast cancer ultrasound images: prediction of axillary lymph node tumor burden in patients&quot;, European Radiology, 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Gang Peng, Bingcang Huang, Xiaodan Yang, Maohua Pang, Na Li. &quot;Preoperative CT feature of incomplete overlying enhancing mucosa as a high-risk predictor in gastrointestinal stromal tumors of the stomach&quot;, European Radiology, 2020</td>
</tr>
<tr>
<td>11</td>
<td>readthedocs.org</td>
</tr>
</tbody>
</table>
Miao-Ping Zhu, Qiao-Ling Ding, Jian-Xia Xu, Chun-Yan Jiang, Jing Wang, Chao Wang, Ri-Sheng Yu.
"Building contrast-enhanced CT-based models for preoperatively predicting malignant potential and Ki67 expression of small intestine gastrointestinal stromal tumors (GISTs)", Abdominal Radiology, 2021
Crossref