Preoperative prediction of vessel invasion in locally advanced gastric cancer based on computed tomography radiomics and machine learning

ZHI-WEI HU[1], PAN LIANG[1], ZHI-LI LI[2], LIU-LIANG YONG[1], HAO LU[1], RUI WANG[1] and JIAN-BO GAO[1]

[1]Department of Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan 450052;
[2]Department of Radiology, Henan Provincial People's Hospital Medical Imaging Center, Zhengzhou, Henan 450003, P.R. China

Abstract. Vessel invasion (VI) is an important factor affecting the prognosis of gastric cancer (GC), and the accurate determination of preoperative VI for locally advanced GC is of great clinical significance. Traditional methods for the evaluation of VI require postoperative pathological examination. Noninvasive preoperative evaluation of VI is therefore crucial to determine the best treatment strategy. To determine the value of preoperative prediction of gastric VI based on portal venous phase computed tomography (CT) radiomic features and machine-learning models, a retrospective analysis of 296 patients with locally advanced GC confirmed through pathological examination was performed. They were divided into two groups, VI+ (n=213) and VI- (n=83), based on pathological results. Using pyradiomics to extract two-dimensional radiomic features of the portal venous stage of locally advanced GC, data were divided into training (n=207) and validation sets (n=89), with a ratio of 7:3, and three feature selection methods were cascaded and merged. Finally, least absolute shrinkage and selection operator (LASSO) regression was used for feature screening to obtain the optimal feature subset. Four current representative machine-learning algorithms were used to construct the prediction model, the receiver operating characteristic curve was constructed to evaluate the predictive performance of the model, and the area under the curve (AUC), accuracy, sensitivity, and specificity were calculated. The differentiation degree, and the Lauren's and CA199 classifications were independent risk factors for locally advanced GC VI. Pyradiomics extracted 864 quantitative features of portal vein images of locally advanced GC. After filtering out low variance features using R, 236 features remained.

Next, 18 features were screened using the LASSO algorithm. Extreme gradient boosting (XGBoost), logistic regression, Gaussian naive Bayes, and support vector machine models were constructed based on the 18 best features screened out of the portal venous CT images of advanced GC and three independent risk factors of GC VI in clinical features predicted the training set AUC values of 0.914, 0.897, 0.880, and 0.814, respectively. The predicted validation set AUC values were 0.870, 0.877, 0.859, and 0.773, respectively. The DeLong test results indicated no statistically significant difference in AUC values between the XGBoost and logistic regression models in the training and validation sets. The four machine-learning models showed high predictive performance. The logistic regression model had the highest AUC value in the validation set (0.877), and the accuracy and F1 score were 77 and 87.6%, respectively. CT radiomic features and machine-learning models based on the portal venous phase can be used as a noninvasive imaging method for the preoperative prediction of VI in locally advanced GC. The logistic regression model exhibited the highest diagnostic performance.

Introduction

In China, gastric cancer (GC) is one of the most common malignant tumors of the digestive system. As the fifth most common cancer and third most deadly cancer, >1,000,000 new cases and ~783,000 deaths from GC were reported worldwide in 2018 (1). Gastric cancer onset is insidious, progresses rapidly, and is associated with poor treatment efficacy. Although patients with GC can undergo radical treatment through surgical resection in the early stages, the postoperative recurrence rate is as high as 60%, and the long-term prognosis of patients remains unsatisfactory. The postoperative recurrence of GC has become the primary cause of death in GC patients, which is related to the tumor size, degree of differentiation, pathological type, and TNM stage (2,3). In addition, several studies have confirmed that vessel invasion (VI) is an independent risk factor for the recurrence and metastasis of locally advanced GC (4-8). Therefore, accurate preoperative evaluation of VI in locally advanced GC is of great clinical significance. The evaluation of VI traditionally requires postoperative pathological examination, which delays the

Correspondence to: Dr Jian-Bo Gao, Department of Radiology, The First Affiliated Hospital of Zhengzhou University, 1 Jianshe East Road, Erqi, Zhengzhou, Henan 450052, P.R. China
E-mail: cjr.gaojianbo@vip.163.com

formulation of a preoperative individualized treatment plan (9). Therefore, exploring methods of noninvasive assessment of VI prior to surgery to determine the best treatment strategy for patients with locally advanced GC is particularly important. Radiomics is based on medical imaging images to obtain more objective, quantitative, deeper, and visually complex tumor features. With the continuous development and improvement of radiomics, the diagnostic ability of artificial intelligence (AI)-based methods has become equal to or even better than that of human experts in several different types of cancer. As a result of this success, AI approaches are currently helping solve more complex clinical decision-making challenges, such as tumor identification and prediction, the study of the tumor microenvironment, assessment of the response to different treatment modalities, identification of treatment-related changes, and imaging representations of genotypic features associated with prognosis (10-12).

In the present study, radiomics analysis was performed based on computed tomography (CT) images of the portal venous phase of locally advanced GC, combined with the diagnostic model established by AI. A variety of machine-learning algorithms were used for comparison, aiming to explore the value of machine-learning models based on the radiomics characteristics of portal vein phase CT in predicting VI of locally advanced GC before surgery.

Materials and methods

Patients. Between July 2011 and December 2020, information was retrospectively collected from 296 patients (214 men and 82 women) with locally advanced GC who were admitted to our hospital and whose diagnoses were confirmed through pathological examination. Their ages varied from 28 to 86 years, with the median age of 60.75 years. The inclusion criteria were as follows: Patients underwent radical surgical resection, locally advanced GC was pathologically confirmed with VI results, and abdominal contrast-enhanced CT (CECT) examination was performed ≤2 weeks before surgery. The exclusion criteria were as follows: Received radiotherapy, chemotherapy, radiochemotherapy, or other nonsurgical treatments before surgery; lesions too small or images where the quality was too poor to be evaluated; incomplete clinicopathological data; and macroscopic VI by the tumor. The pathological evaluation of VI was based on the presence of tumor cell infiltration in the lymphatic vessels and arteriovenous vessels of the peritumor under the microscope (9). Based on postoperative pathological results, 213 patients were confirmed as VI+, and 83 patients were VI-. Patients were divided into training (n=207) and validation sets (n=89) using a stratified sampling method according to the ratio of 7:3. Among them, 149 VI+ and 58 VI- cases were assigned to the training set, while 64 VI+ and 25 VI- cases were assigned to the validation set. The flowchart of case screening is presented in Fig. 1. Clinical data of patients in terms of sex, age, location, TNM stage, degree of differentiation, Lauren's classification, carcinoembryonic antigen levels, and CA125 levels were collected. The study was approved by the Institutional Review Board of The First Hospital of Zhengzhou University (Zhengzhou, China; approval no. 2021-KY-1070-002).

CT image acquisition protocol. A Discovery CT 750HD scanner was used to perform the CECT examination. The primary imaging parameters were as follows: Tube voltage, 100 kVp; tube current, selection automatic mA; pitch, 1.375; detector collimation, 0.625 mm; rotation time, 0.5-0.6 sec; and reconstruction slice thickness, 5 and 1.25 mm. Contrast medium (370 mg I/ml) was injected through the cubital vein with a high-pressure syringe at a flow rate of 3.5 ml/sec and a dose of 1.5 ml/kg body mass. Arterial phase and venous phase scans were performed at 25 and 60 sec after the injection of the contrast agent.

Radiomic analysis. The portal venous phase thin slice images (1.25 mm) of locally advanced GC were introduced into Pyradiomics (version 3.0, https://pyradiomics.readthedocs.io) for imaging feature extraction and analysis.

Tumor segmentation. The tumor region of interest (ROI) was manually delineated based on the portal venous phase CT images. The images were exported in Digital Imaging and Communications in Medicine format. Two abdominal imaging radiologists with 5 years of diagnostic experience (doctor A) and 13 years of diagnostic experience (doctor B), respectively, were asked to select the largest slice of the tumor and determine the contour (Fig. 2). If an agreement could not be reached, the results were judged by another radiologist with 36 years (doctor C) experience in abdominal diagnosis. CT images of 50 cases were randomly selected, and the ROIs were delineated by doctors A and B every 2 weeks, respectively, and radiomics features were extracted. An intraclass correlation coefficient (ICC) >0.75 indicated good consistency.

Radiomics feature extraction and screening. In total, 864 quantitative features of locally advanced GC were extracted by Pyradiomics based on portal venous phase images, including first-order features, shape features, gray co-occurrence matrix features, gray run matrix features, gray region size matrix features, adjacent gray difference matrix features, and gray dependence matrix features. The Mann-Whitney U test and Student's t-test were used for feature selection. After filtering out low variance, the minimum absolute contraction and selection algorithm was used to further screen the optimal feature subset with low co-linearity. Meanwhile, linear fusion was adapted to calculate the score of the radiomics signature for every patient (Rad-score=constant + coefficient x imaging features).

Construction of predictive model. Four machine-learning algorithms, extreme gradient boosting (XGBoost), logical regression, Gaussian naive Bayes (GNB), and support vector machine (SVM) were used to select the optimal feature subset. To evaluate the prediction ability of the model, the receiver operating characteristic (ROC) and calibration curves were used to test the prediction efficiency and fitness of the model. The calibration curve was modified by 1,000 self-weightlifting resampling and then evaluated by the Hosmer-Lemeshow test. The accuracy, sensitivity, and specificity were calculated based on the optimal threshold value in the training set.

Statistical analysis. R version 3.4.3 (http://www.r-project.org) (13) was used for statistical analysis through R Studio (version 1.1.383) (14). Quantitative data were compared using

```
From July 2011 to December 2020, 610 patients with locally advanced gastric cancer confirmed by pathology
```

Exclusion →
```
Receive other non-surgical treatment before surgery; n=205

Small lesions or poor image quality cannot be assessed; n=18

Incomplete clinicopathological data; n=79

Locally advanced gastric cancer has gross vascular invasion; n=12
```

```
296 Patients were included, 213 VI (+) and 83 VI (−)
```

```
Randomized allocation (ratio 7:3)
```

```
Training set (n=207) 149VI (+); 58VI (−)
```

```
Validation set (n=89) 64VI (+); 25VI (−)
```
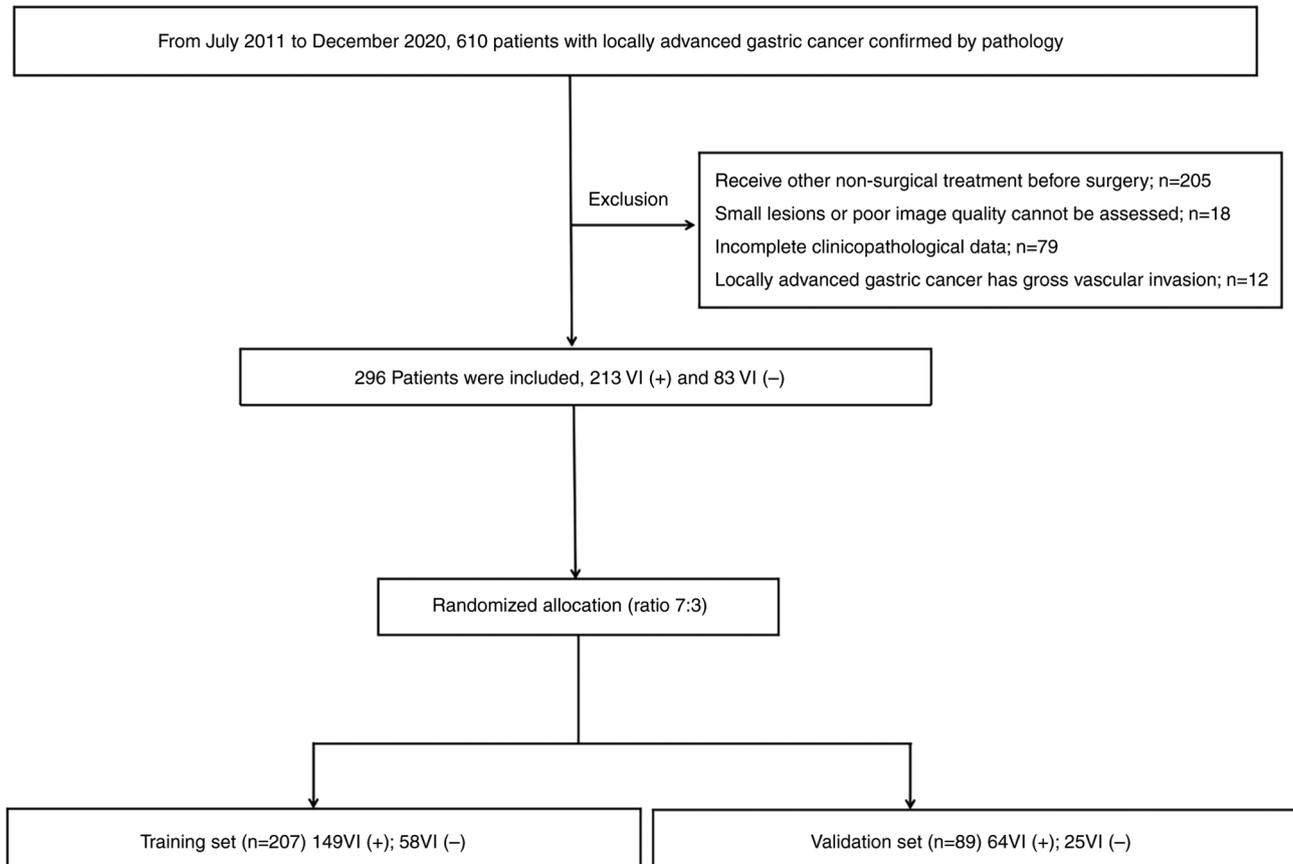
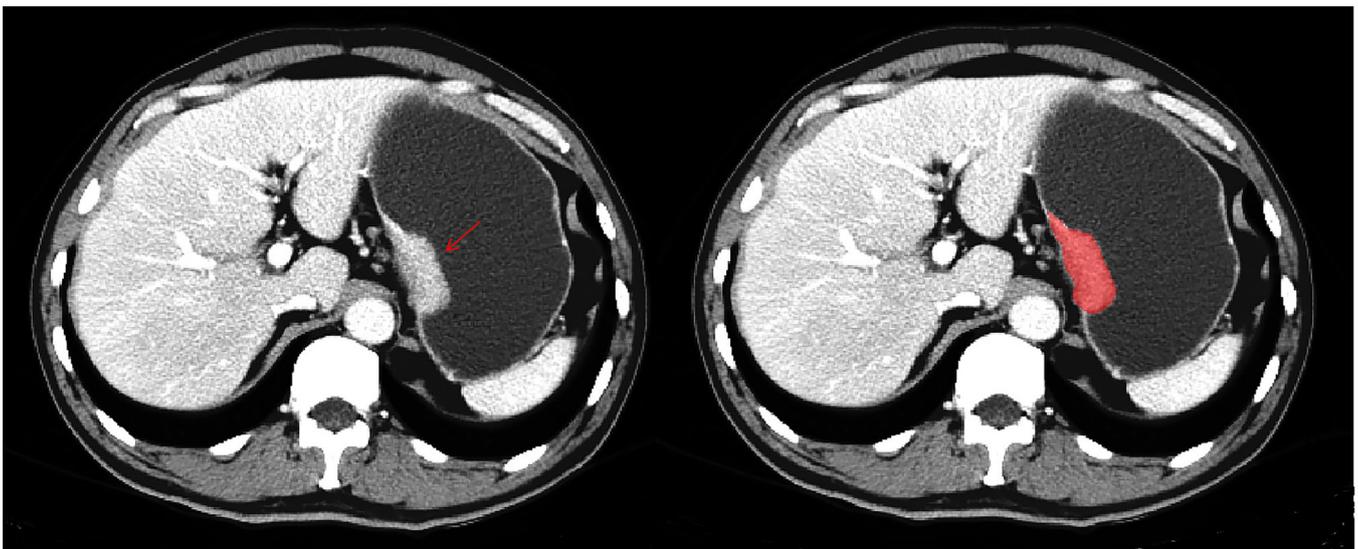Figure 1. Flow chart of patient screening. VI, vascular invasion.

Figure 2. Area of interest was delineated for locally advanced gastric cancer. The venous phase enhanced CT images showed a marked thickening of the stomach wall from the cardia, with uniform enhancement on enhanced scan (red arrow). The red area delineates areas of interest along the tumor. CT, computed tomography.

an independent samples Student's t-test if data were normally distributed or otherwise a Mann-Whitney U test, and qualitative data were analyzed using a χ^2 test. The pROC package was used to draw the ROC curve, and the rms package was applied to draw the calibration curve. The glmnet package was used to realize the least absolute shrinkage and selection operator (LASSO) regression with a five-fold cross-validation method to screen radiomics features. The XGBoost, logical regression, SVM, and GNB algorithms were performed using the packages xgboost, glm, kernlab, and e1071, respectively.

Table I. Comparison of the clinicopathological factors between the VI+ and VI- set of patients with locally advanced gastric cancer.

Factor	VI+ (n=213)	VI- (n=83)	χ^2/t	P-value
Sex, n			-1.341	0.084
Male	148	66		
Female	65	17		
Age, years	63 (54, 69)	61 (54, 67)	-0.646	0.518
Tumor location, n			-3.386	<0.001
Cardia	98	56		
Gastric body	30	13		
Gastric antrum	85	14		
TNM, n			-5.833	<0.001
I	9	20		
II	37	32		
III/IV	167	31		
Differentiation, n			-7.65	<0.001
Middle	50	67		
Low	163	16		
Lauren, n			-1.896	0.042
Intestinal	75	41		
Diffuse	46	14		
Mixed	92	28		
CEA, ng/ml[a]	2.24 (1.43, 4.06)	2.77 (1.68, 4.70)	1.633	0.103
CA125, kU/l[a]	8.40 (6.14, 11.97)	10.60 (7.31, 15.77)	2.562	0.010
CA199, kU/l[a]	9.95 (4.97, 23.24)	14.60 (7.19, 27.20)	2.092	0.036

[a]Data are presented as the median (first quartile, third quartile). VI, vascular invasion.

Results

Clinicopathological features. Statistically significant differences were found in tumor location, TNM stage, differentiation degree, Lauren's classification, CA125, and CA199 levels between VI+ and VI- patients with locally advanced GC (all P<0.05, Table I), but no statistically significant differences in sex, age, and CEA were observed (P>0.05, Table I). Multivariate logistic regression model analysis showed that the degree of differentiation, Lauren's classification, and CA199 levels were independent risk factors for predicting VI in locally advanced GC (Table II).

Radiomics feature extraction repeatability assessment. The extracted radiomics features were based on the ROIs delineated by the doctors two times, which showed good intragroup consistency, and with ICC values >0.75. Radiomics features based on ROIs outlined by the two doctors also exhibited good intergroup consistency.

Radiomic feature screening. A total of 864 radiomics features were extracted based on the ROI from the CT images in the portal venous phase. After filtering out low variance features, 236 radiomics features were obtained. Next, 18 optimal radiomics features were finally selected by the LASSO algorithm, including two original features (original-gldm-Dependence Non-Uniformity Normalized,

Table II. Multivariate logistic regression model analysis.

Factor	P-value	Odds ratio	95% Confidence interval
Differentiation	0.003	13.651	7.265-25.650
Lauren	0.042	1.349	1.011-1.799
CA199	0.044	1.796	1.406-2.186

original-glrlm-ShortRunEmphasis) and 16 wavelet features (wavelet-LLLglszm Zone Entropy, wavelet-HLHgldm Dependence Non Uniformity Normalized, wavelet-LLLngtdm Contrast, wavelet-LLLglrlm Short Run Low Gray Level Emphasis, wavelet-LLHglszm Small Area Low Gray Level Emphasis, wavelet-LHHglszm Zone Variance, wavelet-LHL first order Median, wavelet-LLLglcm Imc1, wavelet-LHHglrlm Gray Level Variance, wavelet-LHHglrlm Gray Level Non Uniformity Normalized, wavelet-LLH first order Total Energy, wavelet-HHLgldm Small Dependence Low Gray Level Emphasis, wavelet-HLHglszm Low Gray Level Zone Emphasis, wavelet-HLHglszm High Gray Level Zone Emphasis, wavelet-HHHngtdm Contrast, and wavelet-LLLglszm High Gray Level Zone Emphasis). In the training set, the Rad-score value [presented as median

Table III. Summary of the performance of the different models in the training and test sets.

	Training set			Validation set		
Model	AUC (95% CI)	Accuracy (95% CI)	F1 (95% CI)	AUC (95% CI)	Accuracy (95% CI)	F1 (95% CI)
XGBoost	0.914 (0.875-0.953)	0.816 (0.774-0.857)	0.885 (0.860-0.910)	0.870 (0.769-0.971)	0.787 (0.751-0.823)	0.876 (0.852-0.900)
Logistic	0.897 (0.853-0.940)	0.823 (0.801-0.846)	0.871 (0.852-0.891)	0.877 (0.788-0.964)	0.777 (0.691-0.863)	0.840 (0.751-0.930)
GNB	0.880 (0.832-0.928)	0.799 (0.775-0.823)	0.850 (0.828-0.873)	0.859 (0.755-0.961)	0.770 (0.717-0.823)	0.876 (0.818-0.935)
SVM	0.814 (0.755-0.873)	0.707 (0.659-0.755)	0.763 (0.704-0.821)	0.773 (0.647-0.898)	0.662 (0.550-0.774)	0.795 (0.750-0.841)

AUC, area under the curve; CI, confidence interval; XGBoost, extreme gradient boosting; SVM, support vector machine; GNB, Gaussian naive Bayes.
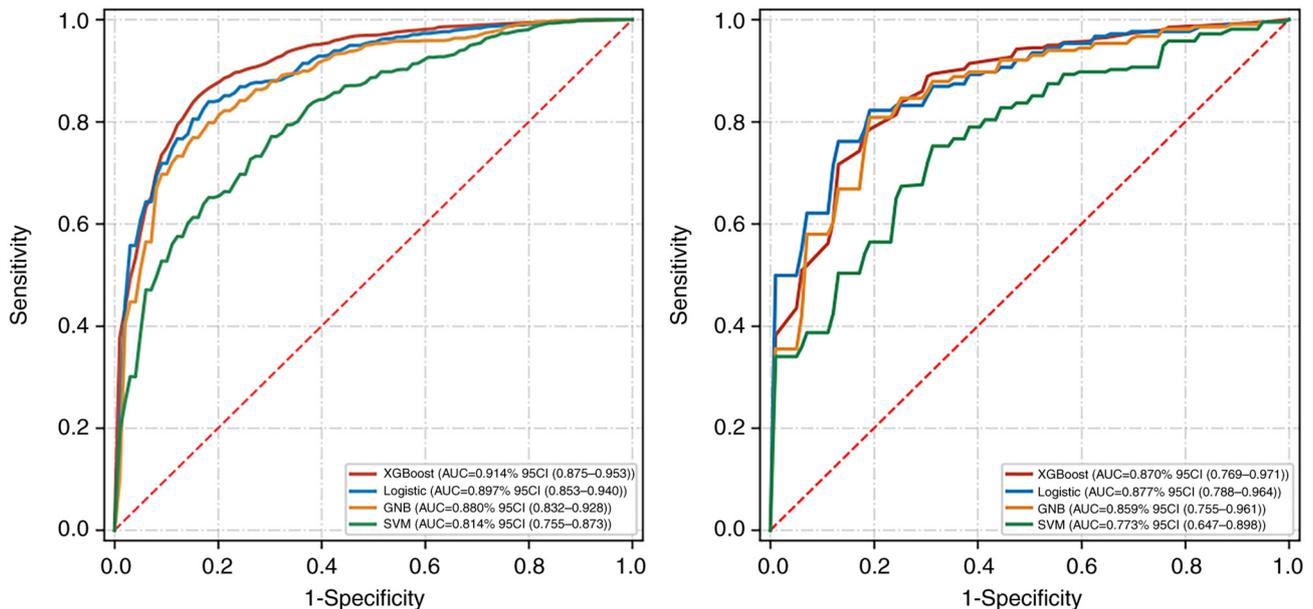

Figure 3. Schematic diagram of ROC of different models in the training set and validation set. ROC, receiver operating characteristic curve; AUC, area under the curve; CI, confidence interval; XGBoost, extreme gradient boosting; SVM, support vector machine; GNB, Gaussian naive Bayes.

(lower quartile, upper quartile)] of VI+ patients [0.763 (0.681, 0.831)] was significantly higher than that of VI- patients [0.672 (0.616, 0.753)]. In the validation set, the Rad-score value of VI+ patients [0.746 (0.668, 0.800)] was higher than that of VI- patients [0.721 (0.625, 0.794)], and the difference between these two groups was statistically significant.

Value of CT radiomics model in predicting VI in advanced GC. The machine-learning model was constructed by combining the 18 best image radiomics features screened using venous CT images of locally advanced GC and the three independent risk factors of GC VI in clinical features. The diagnostic efficacy and ROC curve of different machine-learning models for predicting VI of locally advanced GC are shown in Table III and Fig. 3. The AUC values of prediction models based on

XGBoost, logical regression, GNB, and SVM in training sets were 0.914, 0.897, 0.880, and 0.814, respectively. The AUC values of prediction models in verification sets were 0.870, 0.877, 0.859, and 0.773, respectively. Calibration curves of the validation set showed that the predicted results of the four machine-learning models corresponded with the pathological results (Fig. 4). The results of the DeLong test indicated no significant difference in AUC value between the XGBoost and logical regression models either in the training set or in the validation set. Among the four machine-learning models, the logical regression model showed the highest AUC value of 0.877 (95% confidence interval: 0.788-0.964) in the validation set. The accuracy and F1 scores of the logical regression model were 77.7 and 84.0%, respectively (Table III). Finally, the nomogram is shown in Fig. 5.
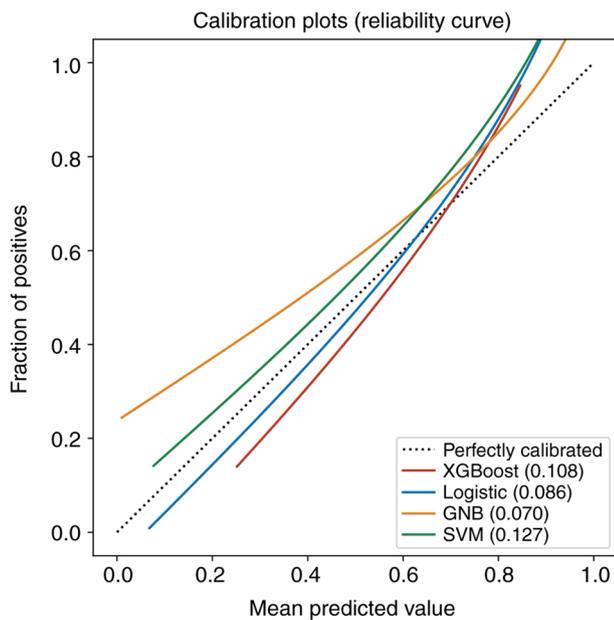
Figure 4. Calibration curves of four machine learning models in the validation sets. XGBoost, extreme gradient boosting; SVM, support vector machine; GNB, Gaussian naive Bayes.

Discussion

In this study, 18 radiomics features screened using venous CT images of locally advanced GC were used to construct image omics labels, and three clinical features screened by multivariate logistic regression were incorporated into the model for training and testing, indicating the rationality of this study in selecting the optimal feature subset. At the same time, XGBoost, logistic, GNB, and SVM models were constructed in this study. Their AUC values and accuracy, sensitivity, and specificity in the test set were good, indicating that the four machine-learning models had high predictive efficiency, suggesting the feasibility of feature selection and model training in this study. In the test set, the best efficacy and highest accuracy in terms of the AUC value of the differentiation index was shown by the logistic model, followed by the XGBoost, GNB, and SVM models. However, no statistical significance in the AUC value difference between the four models was found. The AUC of the XGBoost model in the training set was 0.914, but decreased to 0.870 in the test set, showing an obvious overfitting phenomenon. The AUCs of the logistic model in the training and test sets were 0.897 and 0.877, respectively, which were relatively stable with no overfitting phenomenon and continued to have good fitting and prediction abilities. In this study, the calibration curves of the four models were good, indicating that the results predicted by the four models showed high consistency with the pathological results. Based on these results, the logistic model is the best model to predict the VI status of locally advanced GC.

The primary reason for the poor prognosis of locally advanced GC is postoperative recurrence and metastasis, which has become the primary cause of death. VI is closely related to the pathological characteristics of tumor growth and metastasis. Previous studies have confirmed that the detection of VI is of great clinical value in the diagnosis and treatment of GC. The VI status of GC is generally confirmed based on the postoperative pathological examination, and the obtained results cannot fully reflect the heterogeneity of locally advanced GC. Therefore, some limitations in the guiding of preoperative management of GC patients exist.

A previous study on primary liver cancer found that the nomogram constructed using the random forest method is a potential biomarker for predicting microvascular invasion and relapse-free survival in solitary hepatocellular carcinoma (15). Similarly, in a study of intrahepatic cholangiocarcinoma with mass formation (16), both CECT radiomics analysis and radiomics factors could predict microvascular invasion, and the nomogram could further improve the prediction efficiency. Previous studies have reported that CT radiomics demonstrates good performance in the pathological grading of gastric neuroendocrine tumors, prognostic prediction of GC patients, and neoadjuvant chemotherapy response in advanced GC patients (17-19). Other studies have shown that the normalized iodine concentration based on energy spectrum CT can reflect the angiogenesis of GC (20,21), which is closely related to the recurrence and prognosis of GC. Based on these studies, the prediction of VI of advanced GC based on radiomics should be feasible. To the best of our knowledge, reports on the preoperative prediction of VI in GC are rare (5,7). Therefore, a preliminary study on the VI of GC based on radiomics and machine learning was performed, and these results may form the basis of further research.

Radiomics has excellent clinical application value in the grading and staging of GC, predicting prognosis, and evaluating response (5). Portal venous phase CT images of locally advanced GC effectively reflect the biological behavior of the tumor, and VI can promote the growth and metastasis of GC cells. Therefore, radiomics analysis based on portal venous phase CT images is helpful for the prediction of VI status in locally advanced GC. In this study, original features and wavelet features based on high-pass and low-pass were screened from portal venous phase CT images, reflecting tumor morphological differences and image inhomogeneity. In a study on the prediction of the VI in borderline pancreatic cancer patients after neoadjuvant therapy, Ahmed *et al* (22) found that a circumferential interface ≥180 degrees, contour deformity ≥grade 3, and/or a contact length of tumor >2 cm may be important predictors, and its performance of prediction yielded AUCs of 0.85-0.88 and 0.92- -0.87 for arterial and venous invasion, respectively. However, the sample size was small, and the clinical indicators were not included. Yang *et al* (23) developed a nomogram model integrating the clinical and radiomics features to predict the microvascular invasion of hepatocellular carcinoma, which showed that the AUC values of the model in the training and validation sets were 0.943 and 0.850, respectively, and the predictive performance of the model was lower than that of the prediction models constructed by XGBoost, logistic regression, and GNB in this study. Li *et al* (24) explored the value of the nomogram based on prothrombin induced by vitamin K absence-II, α-fetoprotein, and tumor size for predicting microvascular invasion in early hepatocellular carcinoma. The results showed that the model can predict microvascular invasion in early hepatocellular carcinoma. However, the study included
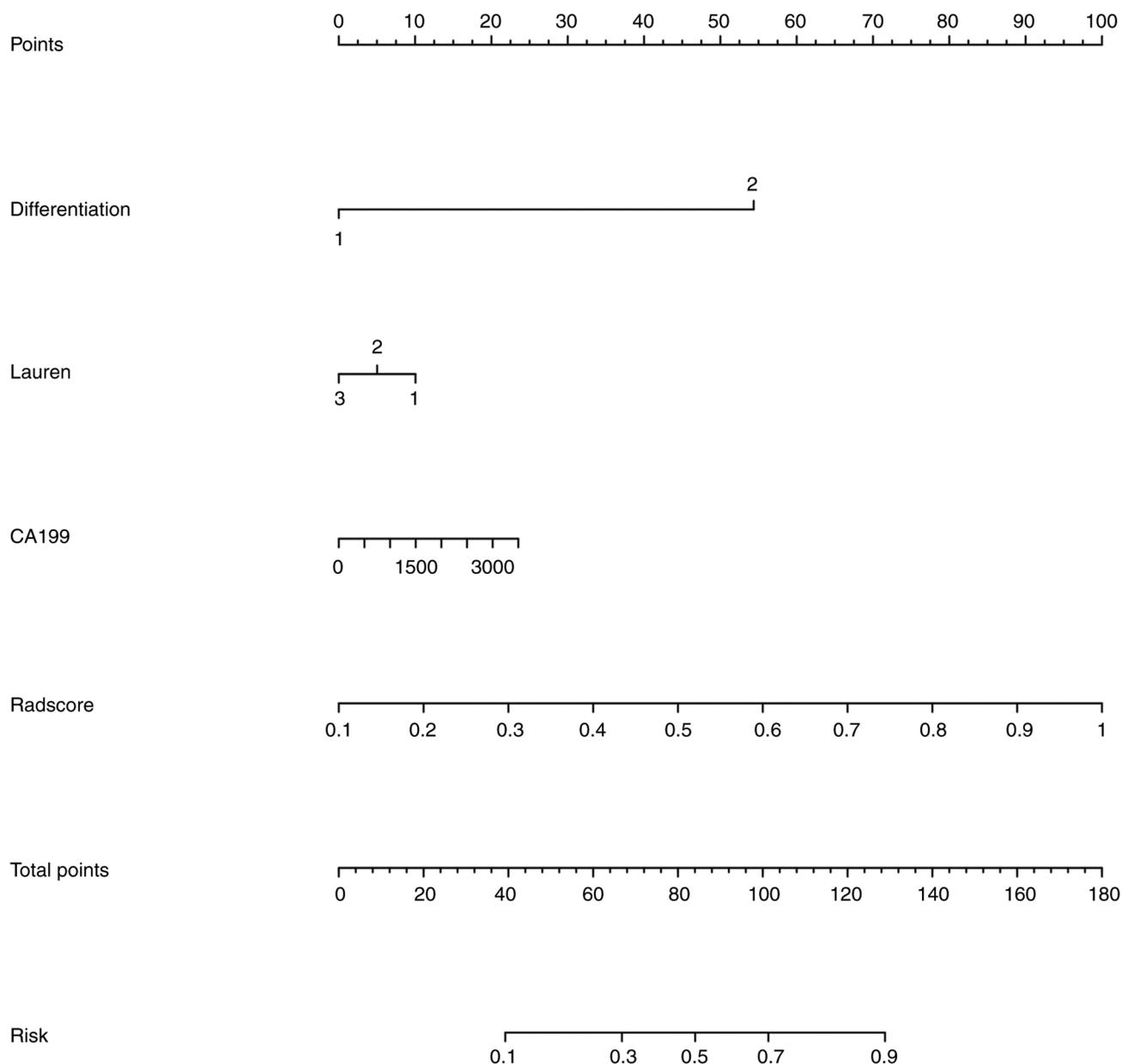
Figure 5. Nomograph prediction model for fusion of clinical features and imaging Radscore.

only 43 patients, only clinical features were included in the nomogram model, and the validation set was lacking. Given this, CT-based radiomics has distinct value in the prediction of preoperative VI. Nevertheless, the determination of optimal features for evaluating preoperative VI is still controversial, and accounting for specificity and sensitivity using a single feature is difficult. Related research on the preoperative prediction of VI in GC using machine-learning models has not yet been reported.

Machine learning has important clinical value in the preoperative prediction of VI in locally advanced GC. In this study, the AUC values of the XGBoost, logistic regression, and GNB models constructed based on CT images in a portal venous phase were higher than 0.80 for the prediction of the VI of locally advanced GC in the validation set, indicating good diagnostic performance, while the AUC of the SVM model in the validation set was relatively low.

Currently, numerous reports have been published on the use of machine-learning algorithms to predict tumor VI (25-27). Using radiomics and clinical and CT image features, Jiang et al (26) developed the XGBoost model and three-dimensional convolutional neural network (3D-CNN) model to predict microvascular invasion in hepatocellular carcinoma. The AUC value of the 3D-CNN model was higher than that of the XGBoost model in both the training and validation sets, which was 0.952 vs. 0.980 and 0.887 vs. 0.906, respectively. Moreover, its predictive performance was higher than that of the prediction model in this study. Although radiomics is an advanced technique for image analysis, the extracted features are usually based on the manual delineation of ROI and rely on personal experience; thus, it cannot represent the most accurate decision (27). The most important advantage of the 3D-CNN model is the high efficiency in identifying microvascular invasion, which can be done automatically with minimal work, time, and materials. For

the construction of the 3D-CNN model, only the original image is required, and clinical data and radiological or radiomics features are not required. In addition, Liu *et al* (28) showed that the deep-learning and SVM models based on arterial phase CT images and clinical features can predict microvascular invasion in hepatocellular carcinoma, and the deep-learning model exhibited better performance with an AUC value of 0.845. The reason why the predictive performance of the previous study was lower than that of the present study may be due to the difference in sample sizes between the two studies. With a smaller sample size, 'falling into the local optimal solution' readily occurs, and achieving the global optimal parameter value is thus difficult.

This study had several limitations. This study preliminarily discussed the value of predicting VI in advanced gastric cancer based on preoperative CT images, However, due to the lack of data related to postoperative recurrence and prognosis, further research is being conducted taking into account these factors. The present study was a single-center study with a small sample, and the stability and practicability of the models require further confirmation. The machine-learning models in this study were based only on internal data, and further external validation is required. Finally, the models were only based on CT images in the portal venous phase, and CT images in other phases should also be compared.

In conclusion, machine-learning models based on portal venous phase CT images can effectively predict the VI status of locally advanced GC before surgery. The predictive performance of the logistic regression model was relatively good and is expected to be an advanced means for the preoperative noninvasive prediction of VI status in locally advanced GC.

Acknowledgements

Not applicable.

Availability of data and materials

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

Authors' contributions

ZWH and JBG confirm the authenticity of all the raw data. PL and RW are responsible for the delineation and classification of the data. ZWH, ZLL, PL, and HL are responsible for data collection. LLY and JBG are responsible for the construction and statistical analysis of the machine learning model. ZWH, ZLL, and PL are responsible drafting and editing the manuscript. JBG gave final approval of the version to be published. All authors listed have read and approved the final manuscript.

Ethics approval and consent to participate

The present study was approved by the Research and Ethics Committee of The First Affiliated Hospital of Zhengzhou University (approval no. 2021-KY-1070-002). Informed consent was obtained from all patients.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A: Global cancer statistics 2018:GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68: 394-424, 2018.
2. Qiu WW, Chen QY, Zheng WZ, He QC, Huang ZN, Xie JW, Wang JB, Lin JX, Lu J, Cao LL, *et al*: Postoperative follow-up for gastric cancer needs to be individualized according to age, tumour recurrence pattern, and recurrence time. Eur J Surg Oncol 48: 1790-1798, 2022.
3. Jin Y, Xu Y, Li Y, Chen R and Cai W: Integrative radiogenomics approach for risk assessment of postoperative and adjuvant chemotherapy benefits for gastric cancer patients. Front Oncol 11: 755271, 2021.
4. Wang J, Zhong L, Zhou X, Chen D and Li R: Value of multiphase contrast-enhanced CT with three-dimensional reconstruction in detecting depth of infiltration, lymph node metastasis, and extramural vascular invasion of gastric cancer. J Gastrointest Oncol 12: 1351-1362, 2021.
5. Meng Y, Huang X, Liu J, Chen J, Bu Z, Wu G, Xie W, Jeen F, Huang L, Tian C, *et al*: A novel nomogram for individually predicting of vascular invasion in gastric cancer. Technol Cancer Res Treat 20: 15330338211004924, 2021.
6. Gresta LT, Rodrigues-Júnior IA, de Castro LP, Cassali GD and Cabral MM: Assessment of vascular invasion in gastric cancer: A comparative study. World J Gastroenterol 19: 3761-3769, 2013.
7. Yang L, Chu W, Li M, Xu P, Wang M, Peng M, Wang K and Zhang L: Radiomics in gastric cancer: First clinical investigation to predict lymph vascular invasion and survival outcome using ^{18}F-FDG PET/CT images. Front Oncol 12: 836098, 2022.
8. Yang S, Zou X, Li J, Zhang A, Zhu L, Hu X and Li C: The application value of ceMDCT in the diagnosis of gastric cancer extramural vascular invasion and its influencing factors. J Healthc Eng 2022: 4239600, 2022.
9. Rodríguez-Perálvarez M, Luong TV, Andreana L, Meyer T, Dhillon AP and Burroughs AK: A systematic review of microvascular invasion in hepatocellular carcinoma: Diagnostic and prognostic variability. Ann Surg Oncol 20: 325-339, 2013.
10. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, Back T, Chesus M, Corrado GS, Darzi A, *et al*: International evaluation of an AI system for breast cancer screening. Nature 577: 89-94, 2020.
11. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, van Ginneken B, Karssemeijer N, Litjens G, van der Laak JAWM; the CAMELYON16 Consortium; Hermsen M, Manson QF, *et al*: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 318: 2199-2210, 2017.
12. Bera K, Braman N, Gupta A, Velcheti V and Madabhushi A: Predicting cancer outcomes with radiomics and artificial intelligence in radiology. Nat Rev Clin Oncol 19: 132-146, 2022.
13. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, 2012. URL http://www.R-project.org/.
14. RStudio Team: RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, 2015. URL http://www.rstudio.com/.
15. Chong HH, Yang L, Sheng RF, Yu YL, Wu DJ, Rao SX, Yang C and Zeng MS: Multi-scale and multi-parametric radiomics of gadoxetate disodium-enhanced MRI predicts microvascular invasion and outcome in patients with solitary hepatocellular carcinoma ≤5 cm. Eur Radiol 31: 4824-4838, 2021.

16. Xiang F, Wei S, Liu X, Liang X, Yang L and Yan S: Radiomics analysis of contrast-enhanced CT for the preoperative prediction of microvascular invasion in mass-forming intrahepatic cholangiocarcinoma. Front Oncol 11: 774117, 2021.

17. Tan X, Yang X, Hu S, Ge Y, Wu Q, Wang J and Sun Z: Prediction of response to neoadjuvant chemotherapy in advanced gastric cancer: A radiomics nomogram analysis based on CT images and clinicopathological features. J Xray Sci Technol 31: 49-61, 2023.

18. Wang R, Liang P, Yu J, Han YJ and Gao JB: Diagnostic efficacy of a combined diagnostic model based on extreme gradient boosting algorithm in differentiating the pathological grading of gastric neuroendocrine neoplasms. Zhonghua Yi Xue Za Zhi 101: 2717-2722, 2021 (In Chinese).

19. Gao X, Ma T, Bai S, Liu Y, Zhang Y, Wu Y, Li H and Ye Z: A CT-based radiomics signature for evaluating tumor infiltrating Treg cells and outcome prediction of gastric cancer. Ann Transl Med 8: 469, 2020.

20. Chen XH, Ren K, Liang P, Chai YR, Chen KS and Gao JB: Spectral computed tomography in advanced gastric cancer: Can iodine concentration non-invasively assess angiogenesis? World J Gastroenterol 23: 1666-1675, 2017.

21. Liang P, Ren XC, Gao JB, Chen KS and Xu X: Iodine concentration in spectral CT: Assessment of prognostic determinants in patients with gastric adenocarcinoma. AJR Am J Roentgenol 209: 1033-1038, 2017.

22. Ahmed SA, Mourad AF, Hassan RA, Ibrahim MAE, Soliman A, Aboeleuon E, Elbadee OMA, Hetta HF and Jabir MA: Preoperative CT staging of borderline pancreatic cancer patients after neoadjuvant treatment: Accuracy in the prediction of vascular invasion and resectability. Abdom Radiol (NY) 46: 280-289, 2021.

23. Yang L, Gu D, Wei J, Yang C, Rao S, Wang W, Chen C, Ding Y, Tian J and Zeng M: A radiomics nomogram for preoperative prediction of microvascular invasion in hepatocellular carcinoma. Liver Cancer 8: 373-386, 2019.

24. Li H, Li T, Hu J and Liu J: A nomogram to predict microvascular invasion in early hepatocellular carcinoma. J Cancer Res Ther 17: 652-657, 2021.

25. Xu X, Zhang HL, Liu QP, Sun SW, Zhang J, Zhu FP, Yang G, Yan X, Zhang YD and Liu XS: Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. J Hepatol 70: 1133-1144, 2019.

26. Jiang YQ, Cao SE, Cao S, Chen JN, Wang GY, Shi WQ, Deng YN, Cheng N, Ma K, Zeng KN, et al: Preoperative identification of microvascular invasion in hepatocellular carcinoma by XGBoost and deep learning. J Cancer Res Clin Oncol 147: 821-833, 2021.

27. Hosny A, Parmar C, Quackenbush J, Schwartz LH and Aerts HJWL: Artificial intelligence in radiology. Nat Rev Cancer 18: 500-510, 2018.

28. Liu SC, Lai J, Huang JY, Cho CF, Lee PH, Lu MH, Yeh CC, Yu J and Lin WC: Predicting microvascular invasion in hepatocellular carcinoma: A deep learning model validated across hospitals. Cancer Imaging 21: 56, 2021.