

Application of CT Radiomics Combined with Machine Learning Methods in Predicting the Recurrence or Metastasis of Gastrointestinal Stromal Tumors

Qiu-Xia Feng

the First Affiliated Hospital of Nanjing Medical University

Lu-Lu Xu

the First Affiliated Hospital of Nanjing Medical University

Qiong Li

the First Affiliated Hospital of Nanjing Medical University

Xiao-Ting Jiang

the First Affiliated Hospital of Nanjing Medical University

Bo Tang

the First Affiliated Hospital of Nanjing Medical University

Yu-Dong Zhang

the First Affiliated Hospital of Nanjing Medical University

Xi-Sheng Liu (✉ njmu_lxs@163.com)

the First Affiliated Hospital of Nanjing Medical University

Research Article

Keywords: Gastrointestinal stromal tumor, CT, Radiomics, Recurrence, Metastasis, Machine learning

Posted Date: December 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1124517/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

The study aimed to evaluate the diagnostic performance of machine learning-based CT radiomics models for predicting the recurrence and metastasis of gastrointestinal stromal tumors (GISTs) preoperatively.

Methods

A total of 382 patients with histopathological confirmed GISTs were retrospectively included. According to postoperative follow-up, patients were classified into non-recurrence and metastasis group (NRM) and recurrence or metastasis group (RM). Radiomics features were extracted from arterial and portal venous phase CT images. Four feature selection methods and ten machine learning techniques were used to train predicting models on training cohort with internal validation by 10-fold cross-validation. F1 score was used to evaluate the performance of the classification model. The best model of two phase were stacked to build an ensemble model. The area under the curve (AUC), recall, precision, accuracy, and F1 score were used to evaluate the performance of the models and compare with clinical criteria based on diameter.

Results

Eighty machine learning models in two phases were built and the ensemble model was integrated by analysis of variance and Naive Bayes (ANOVA_NB) model in arterial phase which selected only 5 features provided the highest F1 Score of 0.560 and Kruskal Wallis and Adaptive Boosting (KW_ AdaBoost) model in venous phase which selected only 4 features provided the highest F1 Score of 0.500. The AUC of the generated ensemble model and the clinical criteria showed no difference (0.866 vs 0.857; DeLong Test, $P = 0.865$). But the ensemble model had higher accuracy (0.961), recall (0.826), precision (0.905), F1 Score (0.864), and the area under the Precision-Recall curve (0.774; 95%CI, 0.552 - 0.917), compared with clinical criteria, of which, the accuracy was 0.942, recall was 0.367, precision was 0.478, the F1 Score was 0.415 and the area under the Precision-Recall curve was 0.354(95%CI, 0.552 - 0.917).

Conclusions

Our findings highlight the potential of machine learning techniques based on CT radiomics in the prediction of recurrence and metastasis of GISTs preoperatively.

Introduction

Gastrointestinal stromal tumors (GISTs) are the most common mesenchymal tumors of the gastrointestinal tract, with a varying malignant potential[1, 2]. There are many criteria for evaluating postoperative biological behavior of primary resectable GISTs. The most recognized standards include the National Institute of Health (NIH) classification, the Armed Forces Institute of Pathology (AFIP) classification, and the Memorial Sloan Kettering Cancer Center (MSKCC) prognostic nomogram, which all classify GISTs to different recurrence risk groups by mitotic index and tumor size, and AFIP and MSKCC also highlight the importance of tumor site[3]. The biological behavior of GISTs ranges from very low risk to malignant, and different treatment is designed for GISTs with different risk stratification[3, 4]. Fortunately, molecular-targeting therapy using tyrosine kinase inhibitor imatinib mesylate is effective for high-risk GISTs[5, 6]. And neoadjuvant therapy is reported have the potential to increase the complete resection rate and to avoid the surgical rupture by decreasing the tumor size. The non-recurrence and metastasis (NRM) survival rate can be obtained by the neoadjuvant therapy of high risk GISTs[7].

So, accurate preoperative assessment the recurrence and metastatic risk of GISTs has high clinical value, which can provide important clues for personalized treatment of GISTs, like the use of neoadjuvant chemotherapy. No matter NIH, AFIP classification, and MSKCC nomogram, they are all based on the mitotic index and are appropriate for GISTs classification after operation. The higher the mitotic index means the higher the risk of recurrence, and a mitotic count >5/50 HPF is considered as a warning sign of high recurrence rate of GISTs[8]. As we know, the mitotic index can only be measured by anatomic pathology. Although the tumor tissue of gastric stromal tumors could be obtained by endoscopy, the mitotic index is determined for the full volume of the whole tissue. Additionally, small intestinal stromal tumors are not so easy to get the tissue by endoscopy preoperatively. Therefore, clinicians often considered a diameter more than 5 cm of a GIST as a marker for poor prognosis in preoperative decision making[9, 10].

Machine learning (ML) has a variety of applications in the healthcare and medicine field. The multi algorithms have been employed in the cancer diagnosis and prognosis prediction, like breast cancer, oral cancer, cervical cancer, colon cancer, gastric cancer, multiple myeloma, and so on[11, 12]. These ML methods have been regarded as more powerful prognostic biomarkers and diagnostic tools, which can provide additional information to clinical data due to the ability of ML tools to detect key features from complex datasets[13].

However, there are few literatures on the performance of ML methods in the setting of recurrence or metastasis (RM) of GISTs and whether the performance is rival or even surpasses that of clinical criteria based on diameter. Thus, we aim to provide a general overview on various ML tools in predicting of RM of GISTs and to compare their predictive performance to the clinical criteria.

Material And Methods

Research dataset

Ethical approval for this retrospective study was obtained from our institutional review board and informed consent was waived. By searching of electronic medical records, a total of 382 patients from January 2010 to December 2019 were included in the study (215 men and 167 women). The inclusion criteria were as follows: (1) patients with primary GISTs confirmed by postoperative histopathological examination; (2) availability of standard contrast-enhanced CT images before surgery; (3) patients with complete clinicopathologic data. The exclusion criteria were as follows: (1) patients with other concurrent primary malignant tumors; (2) distant metastasis confirmed by preoperative images; (3) patients received preoperative targeted therapy, such as the use of imatinib; (4) tumor ruptured during or before the operation; (5) unclear lesion in the CT images. Baseline clinical data includes age, clinical symptom, tumor site, size, mitotic rate, Ki67 index, and risk stratification (according to the modified National Institutes of Health criteria). After radical surgery, all patients were routinely followed up with abdomen CT examinations or telephone calls annually. The last follow-up date was June 2020. The endpoint was time to recurrence or metastasis (RM).

The CT images of all patients in the arterial phase and portal venous phase were utilized for tumor analysis and segmentation. CT imaging acquisition procedure was described in the supplementary material (Text S1). Tumor size (maximal diameter) was measured on CT by one radiologist (QXF, who had abdominal radiological experience of 5 years), who was unaware of the clinical and pathological data. Lesion segmentation was semi-automatically performed with a dedicated commercial software package (Frontier, Syngo via, Siemens's healthcare) by one radiologist (BT) and reconfirmed one month later by another radiologist (QXF who had abdominal radiological experience of 3 years). After lesion segmentation, imaging features were analyzed from target volumes using an open-source python package for the extraction of Radiomics features (<https://pyradiomics.readthedocs.io/en/latest/index.html>). Image normalization was performed using a method that remaps the histogram to fit within $\mu \pm 3\sigma$ (μ : gray-level mean between the VOI and σ : gray-level standard deviation). In total, 851 radiomic imaging features were automatically extracted from target volume based on the 7 texture analysis methods available in the software package: first order statistics; shape features; features of the gray level co-occurrence matrix (GLCM); features of gray level run-length matrices (GLRLM); features of the gray level size zone matrix (GLSZM); features of the neighboring gray tone difference matrix (NGTDM) and features of the gray level dependence matrix (GLDM).

Machine learning

All patients were randomly selected to set up a training data set ($n=267$, 16/251 = positive/negative) and the testing data set ($n=115$, 7/108 = positive/negative). To remove the unbalance of the training data set, we used the Synthetic Minority Oversampling Technique (SMOTE) to make positive/negative samples balance. Then normalization was applied to the feature matrix and each feature vector was subtracted by the mean value and was divided by the standard deviation. After normalization process, each vector has mean values of 0 and standard deviations of 1, and the similarity of each feature pair was compared. If the PCC value of the feature pair was larger than 0.99, it would be removed. After this process, the dimension of the feature space was reduced, and each feature was independent to each other.

Before build the model, we evaluated four feature selection methods to select features: analysis of variance (ANOVA), Kruskal Wallis (KW), recursive feature elimination (RFE), and relief.

ANOVA and KW were used to select features according to the corresponding F-value. The goal of RFE was to select features based on a classifier by recursively considering smaller set of the features. And relief selected sub data set and found the relative features according to the label recursively. Then all radiomics features selected were applied to the classifiers to establish the predicting model of different algorithm combinations for the recurrence or metastasis of GISTs. The ten machine learning classifiers were: linear discriminant analysis (LDA), Support Vector Machines (SVM), Random Forest (RF), Adaptive Boosting (AdaBoost), Auto-Encoder (AE) sometimes called multi-layer perceptron (MLP), Gaussian process (GP), Naive Bayes (NB), Logistic Regression (LR), Least Absolute Shrinkage and Selection Operator (LASSO), and Decision Tree (DT). Here we would describe the ten ML tools generally as following. LDA was a linear classifier by fitting class conditional densities to the data and using Bayes' rule. SVM was an effective and robust classifier to build the model. The kernel function has the ability to map the features into a higher dimension to search the hyper-plane for separating the cases with different labels and was easier to explain the coefficients of the features for the final model. RF is an ensemble learning method which combining multiple decision trees at different subset of the training data set and it is an effective method to avoid over-fitting. AdaBoost is a meta-algorithm that conjunct other type of algorithms and combine them to get a final output of boosted classifier. It is sensitive to the noise and the outlier. Over-fitting can also be avoided by AdaBoost. Here we used decision tree as the base classifier for AdaBoost. MLP is based neural network with multi-hidden layers to find the mapping from inputted features to the label. Here we used 1 hidden layer with 100 hidden units. The non-linear activate function was rectified linear unit function and the optimizer was Adam with step 0.001. GP combines the features to build a joint distribution to estimate the probability of the classification. NB is a kind of probabilistic classifiers based on Bayes theorem. NB requires number of parameters linear in the number of features. Logistic regression is a linear classifier that combines all the features. A hyper-plane was searched in the high dimension to separate the samples. Logistic regression with LASSO constrains is also a linear classifier based on logistic regression. L1 norm is added in the final lost function and the weights was constrained, which make the features sparse. DT is a non-parametric supervised learning method and can be used for classification with high interpretation.

To determine the hyper-parameter of model, cross validation with 10-fold was applied on the training data set. The hyper-parameters were set according to the model performance on the validation dataset using one standard error rule. All above processes were implemented with FeAture Explorer Pro (FAEPro, V 0.3.3) on Python (3.7.6)[14].

The study process diagram is shown in Figure 1.

Statistical analysis and predictive performance of models

To analyze baseline clinical data, categorical variables were compared by using the χ^2 test or *Fisher* exact test. Continuous variables were compared by using the *Student t* test. As for machine learning model, each model was evaluated by calculating accuracy, the area under the ROC curve (AUC), recall, precision, and F1 Score indicators. These performance measures or indicators were defined and computed as follows:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

TP means true positive, TN means true negative, FP means false positive, and FN means false negative. They are all ensembled in the confusion matrix. Due to our imbalanced data, we use F1 Score as the main performance indicator. Precision-Recall (PR) Curves were draw to compare the performance of machine learning model between the clinical criteria. The AUC of each PR curve was calculated.

A two-sided P value<0.05 was considered to indicate statistical significance. All regular statistical analyses were performed using the Medcalc software (Version 19.1.6). The machine learning algorithms were programmed using were performed on Python (3.7.6)[14].

Result

Baseline clinical characteristics

Among the 382 patients (mean age, 58.6 years old; range: 20–90 years old) included in the present study, 215 patients were male and 167 patients were female. Recurrence or metastasis was found on 23 (6%) patients during following up imaging. The mean RM time was 1.93 years (± 1.21 years) from 0.30 to 4.50 years. The age, gender, and clinical symptom had no significant relationship with RM of GISTs. And the tumor size, mitotic rate, Ki67 index, and risk stratification showed difference between the RM and NRM groups. The baseline characteristics of all patients were summarized in Table 1.

Table 1
Baseline clinicopathological characteristics of GIST patients
between NRM and RM groups

Variable	NRM group (n = 359)	RM group (n = 23)	P
Age(mean±std)	58.7 (12.0)	57.1(12.5)	0.533
Gender			0.087
Male	206 (57.4%)	9 (39.1%)	
Female	153 (42.6%)	14 (60.9%)	
Symptom			0.520
No	84 (23.4%)	6 (26.1%)	
Discomfort	188(52.4%)	13 (56.5%)	
Bleeding	87(24.2%)	4 (17.4%)	
Mitosis			<0.001
≤5/50 HPF	293 (81.6%)	9 (39.1%)	
> 5/50 HPF	66 (18.4%)	14 (60.9%)	
Risk stratification			<0.001
Well differentiated	237 (66.0%)	5 (21.7%)	
Poorly differentiated	122 (34.0%)	18 (78.3%)	
Ki-67 index			<0.001
≤ 5%	264 (73.5%)	8 (34.8%)	
>5%	95 (26.5%)	15 (65.2%)	
Site			0.007
Stomach	288 (80.2%)	13 (56.5%)	
Small intestine	71 (19.8%)	10 (43.5%)	
Diameter(cm±std)	3.7 (2.7)	9.0(4.8)	<0.001

Diagnostic Performance Of Ten ML Techniques

Cross validation with 10-fold was applied on the training data set to determine the hyper-parameter of model. The hyper-parameters were set according to one standard error rule to select the smallest model. And on the testing data set, the performance of the forty smallest ML models in arterial phase were

showed in Table 2. Among the machine learning models, ANOVA_NB which selected only 5 features provided the highest F1 Score of 0.560 than the others. The precision and recall could achieve 0.389 and 1.000, whereas accuracy was 0.904 and AUC of ROC was 0.942(95%CI: 0.8891-0.9851). And on the testing data set, the performance of the forty smallest ML models in venous phase were showed in Table 3. Among the forty models, KW_ AdaBoost which selected only 4 features provided the highest F1 Score of 0.500. The precision and recall could achieve 0.600 and 0.429, whereas accuracy was 0.948 and AUC of ROC was 0.644(95%CI: 0.3509-0.9782). The selected features of each model were provided in the supplementary material (Table S1).

Table 2
The forty ML models' performance in arterial phase

Selection	Classification	Number	AUC	95%CI	Acc	Recall	Precision	F1 Score
ANOVA	AdaBoost	1	0.686	[0.4669-0.8873]	0.687	0.714	0.128	0.217
	AE	4	0.952	[0.8976-0.9927]	0.852	1.000	0.292	0.452
	DT	1	0.628	[0.3989-0.8465]	0.678	0.571	0.105	0.178
	GP	1	0.931	[0.8563-0.9893]	0.913	0.857	0.400	0.545
	LASSO	2	0.948	[0.8949-0.9893]	0.878	1.000	0.333	0.500
	LDA	2	0.946	[0.8919-0.9855]	0.878	1.000	0.333	0.500
	LR	2	0.948	[0.8949-0.9893]	0.878	1.000	0.333	0.500
	NB	5	0.942	[0.8891-0.9851]	0.904	1.000	0.389	0.560
	RF	1	0.687	[0.5000-0.8689]	0.478	1.000	0.105	0.189
	SVM	2	0.947	[0.8927-0.9874]	0.878	1.000	0.333	0.500
KW	AdaBoost	4	0.718	[0.5000-0.9273]	0.913	0.429	0.333	0.375
	AE	8	0.929	[0.8236-1.0000]	0.861	0.857	0.286	0.429
	DT	3	0.608	[0.3973-0.8341]	0.765	0.429	0.115	0.182
	GP	3	0.803	[0.7054-0.8982]	0.730	1.000	0.184	0.311
	LASSO	1	0.956	[0.8976-0.9958]	0.852	1.000	0.292	0.452
	LDA	9	0.540	[0.2320-0.8379]	0.687	0.571	0.108	0.182
	LR	1	0.956	[0.8976-0.9958]	0.852	1.000	0.292	0.452

Selection	Classification	Number	AUC	95%CI	Acc	Recall	Precision	F1 Score
	NB	1	0.956	[0.8976-0.9958]	0.852	1.000	0.292	0.452
	RF	1	0.801	[0.5994-0.9583]	0.878	0.714	0.294	0.417
	SVM	1	0.956	[0.8976-0.9958]	0.852	1.000	0.292	0.452
Relief	AdaBoost	20	0.878	[0.7676-0.9709]	0.748	1.000	0.194	0.326
	AE	1	0.612	[0.4340-0.7839]	0.539	0.857	0.103	0.185
	DT	18	0.712	[0.5020-0.9306]	0.835	0.571	0.200	0.296
	GP	8	0.782	[0.5473-0.9564]	0.861	0.714	0.263	0.385
	LASSO	2	0.730	[0.5804-0.8677]	0.687	0.857	0.146	0.250
	LDA	18	0.612	[0.4340-0.7839]	0.539	0.857	0.103	0.185
	LR	18	0.859	[0.7593-0.9473]	0.722	1.000	0.180	0.304
	NB	5	0.451	[0.2679-0.6369]	0.391	0.857	0.080	0.146
	RF	8	0.851	[0.7235-0.9505]	0.652	1.000	0.149	0.259
	SVM	20	0.840	[0.7328-0.9286]	0.713	1.000	0.175	0.298
RFE	AdaBoost	4	0.503	[0.3382-0.6717]	0.365	1.000	0.088	0.161
	AE	20	0.808	[0.6036-0.9749]	0.652	0.857	0.133	0.231
	DT	15	0.521	[0.4299-0.7013]	0.852	0.143	0.083	0.105
	GP	2	0.946	[0.8925-0.9851]	0.878	1.000	0.333	0.500
	LASSO	2	0.948	[0.8949-0.9893]	0.878	1.000	0.333	0.500

Selection	Classification	Number	AUC	95%CI	Acc	Recall	Precision	F1 Score
	LDA	5	0.653	[0.3125-0.9255]	0.687	0.714	0.128	0.217
	LR	3	0.939	[0.8803-0.9832]	0.852	1.000	0.292	0.452
	NB	9	0.583	[0.1881-0.9482]	0.826	0.571	0.191	0.286
	RF	8	0.773	[0.6137-0.9088]	0.522	1.000	0.113	0.203
	SVM	4	0.681	[0.3280-0.9591]	0.922	0.571	0.400	0.471

Table 3
The forty ML models' performance in venous phase

Selection	Classification	Number	AUC	95%CI	Acc	Recall	Precision	F1 Score
ANOVA	AdaBoost	1	0.667	[0.4775-0.8296]	0.583	0.857	0.113	0.200
	AE	2	0.734	[0.4212-0.9455]	0.678	0.857	0.143	0.245
	DT	3	0.670	[0.4595-0.8889]	0.757	0.571	0.138	0.222
	GP	9	0.694	[0.4272-0.9127]	0.800	0.714	0.192	0.303
	LASSO	8	0.720	[0.4255-0.9109]	0.670	0.857	0.140	0.240
	LDA	8	0.722	[0.4189-0.9204]	0.670	0.857	0.140	0.240
	LR	3	0.733	[0.4302-0.9444]	0.678	0.857	0.143	0.245
	NB	2	0.734	[0.4212-0.9436]	0.687	0.857	0.146	0.250
	RF	15	0.759	[0.5443-0.9143]	0.704	0.857	0.154	0.261
	SVM	18	0.694	[0.4264-0.8691]	0.722	0.857	0.162	0.273
KW	AdaBoost	4	0.644	[0.3509-0.9782]	0.948	0.429	0.600	0.500
	AE	16	0.705	[0.4027-0.9105]	0.774	0.714	0.172	0.278
	DT	4	0.624	[0.4730-0.8563]	0.922	0.286	0.333	0.308
	GP	6	0.765	[0.6018-0.9273]	0.557	1.000	0.121	0.215
	LASSO	6	0.684	[0.3780-0.9255]	0.861	0.571	0.235	0.333
	LDA	1	0.598	[0.3571-0.8386]	0.783	0.429	0.125	0.194
	LR	8	0.693	[0.3869-0.9246]	0.870	0.571	0.250	0.348

Selection	Classification	Number	AUC	95%CI	Acc	Recall	Precision	F1 Score
	NB	1	0.598	[0.3571-0.8386]	0.783	0.429	0.125	0.194
	RF	11	0.737	[0.5255-0.9292]	0.652	0.857	0.133	0.231
	SVM	6	0.630	[0.3716-0.8860]	0.861	0.429	0.200	0.273
Relief	AdaBoost	3	0.639	[0.3700-0.8649]	0.583	0.714	0.098	0.172
	AE	4	0.622	[0.3363-0.8378]	0.644	0.714	0.114	0.196
	DT	4	0.544	[0.4545-0.7271]	0.896	0.143	0.143	0.143
	GP	3	0.717	[0.4618-0.9167]	0.826	0.571	0.191	0.286
	LASSO	4	0.623	[0.3363-0.8401]	0.644	0.714	0.114	0.196
	LDA	3	0.615	[0.3272-0.8333]	0.644	0.714	0.114	0.196
	LR	4	0.623	[0.3363-0.8401]	0.644	0.714	0.114	0.196
	NB	13	0.721	[0.4352-0.9286]	0.661	0.857	0.136	0.235
	RF	6	0.711	[0.4702-0.9109]	0.835	0.714	0.227	0.345
	SVM	3	0.619	[0.3363-0.8363]	0.644	0.714	0.114	0.196
RFE	AdaBoost	1	0.688	[0.4286-0.8938]	0.783	0.714	0.179	0.286
	AE	3	0.683	[0.4073-0.8983]	0.861	0.571	0.235	0.333
	DT	2	0.677	[0.4643-0.8823]	0.896	0.429	0.273	0.333
	GP	1	0.712	[0.4801-0.9000]	0.626	0.857	0.125	0.218
	LASSO	1	0.618	[0.3363-0.8353]	0.644	0.714	0.114	0.196

Selection	Classification	Number	AUC	95%CI	Acc	Recall	Precision	F1 Score
	LDA	1	0.618	[0.3363-0.8353]	0.644	0.714	0.114	0.196
	LR	1	0.618	[0.3363-0.8353]	0.644	0.714	0.114	0.196
	NB	1	0.646	[0.3452-0.8874]	0.652	0.714	0.116	0.200
	RF	3	0.683	[0.4534-0.8958]	0.739	0.714	0.152	0.250
	SVM	1	0.618	[0.3363-0.8353]	0.644	0.714	0.114	0.196

Comparison Of The Ensemble Model And The Clinical Criteria

We combined the ANOVA_NB model and the KW_ AdaBoost model to an ensemble model for RM prediction of GISTs. The performance of the ensemble model and the clinician criteria based on diameter were shown in Figure 2 and Table 4. For the ensemble model, accuracy was 0.961, recall was 0.826, precision was 0.905, the F1 Score was 0.864 and the area under the PR curve was 0.774(95%CI, 0.552 - 0.917). For the clinical criteria, accuracy was 0.942, recall was 0.367, precision was 0.478, the F1 Score was 0.415 and the area under the PR curve was 0.354(95%CI, 0.175 - 0.533).

Table 4
Comparison of the ensemble learning model and the clinical criteria

Classification	AUC (95%CI)	Acc	Recall	Precision	F1 score	F1auc(95%CI)
Ensemble model	0.866(0.827 - 0.898)	0.961	0.826	0.905	0.864	0.774(0.552 - 0.917)
Clinical criteria	0.857(0.818 - 0.891)	0.942	0.367	0.478	0.415	0.354(0.175 - 0.533)
Note. AUC means the area under the ROC curve						
F1auc means the area under the PR curve						

We found that the model ensembled by the combination of ANOVA_NB and the KW_ AdaBoost classifiers was valuable for the RM prediction of GISTs, although the AUC of the ensemble model and the clinical criteria showed no difference (0.866 vs 0.857; DeLong Test, P = 0.865), seen in Figure 3.

Discussion

Our study combined four feature selection methods and ten machine learning techniques to predict RM of GISTs identified based on radiomics of contrast CT and then stacked the arterial and venous phase best model to an ensemble model by logistic regression. The ANOVA_NB and KW_ AdaBoost models were found outperformed other feature selections and ML techniques. Ultimately, we compared the ensemble model's diagnostic performance to that of clinical criteria based on diameter. It provided superior F1 Score, accuracy, recall, and precision measures compared with clinical criteria.

Patients with GISTs who had recurrence or metastasis had a higher mortality rate when compared with the patients who did not have[15]. Similar to previous study, the pathological features including Ki67 and mitotic index were tended to be related to the prognosis of GISTs, and the tumors in the recurrence or metastasis group tended to be larger in diameter (mean size: 9 cm) in our research. And clinically, a GIST more than 10 cm with any mitotic rate is considered to have higher risk of recurrence, subsequently requiring target drug therapy[16]. There are many nomograms and models for predicting the biological behavior of GISTs. Moreover, the clinical pathological classification of GISTs is to predict the risk or probability of metastasis of GISTs like modified NIH and AFIP criteria[3]. And although approximately 50% of the patients with primary GIST who have complete tumor resection will survive more than 5 years, moreover, it may take more than ten years for a patient with GISTs to have metastasis[17], surgery alone may not be enough for the treatment[18]. However, few studies obtained long-term survival follow-up of stromal tumors and directly predict the recurrence or metastasis of GISTs.

In our study, 23 cases were found RM and one of them was found metastasis nearby 5 years. The proportion of recurrent cases in our study is low, which is inconsistent with previous reports by Ronald P[18] and Chairat[15], but similar to the recent report by Ao[19]. We ruminated that it might result from our country's clinicians deepening understanding of GISTs in recent years, combined with the detection of kit and PDGFR gene mutation and the application of targeted medicine in high-risk stromal tumors. All of these lead to the decline of recurrent and metastatic incidence rate in our study. However, based on this imbalanced data, different from Ao's evaluation of the performance of the model by using ROC curve, SMOTE technique was used to balance the data and we also used F1 Score and PR curve to evaluate the performance of our designed model. The results of our study showed that even though both the ensemble model and the clinical criteria had high AUC, there were still huge differences between the two models on PR curve, and the ensemble model designed by ML techniques was better for predicting RM of GISTs.

As seen in the literature, the application of radiomics based on CT or MRI have been found in the differential diagnosis of GISTs, risk stratification and prediction of prognosis after surgical resection, and evaluation of mutational status in GISTs[20]. Various feature selection methods of KW and Relief algorithms and ML techniques, including SVM, LASSO, LR, DT, and RF had been used in GISTs diagnosis and prognosis prediction[19, 21–23]. And all these methods and techniques were reported to have achieved promising application in GISTs differential diagnosis and biological behavior prognosis prediction with the range of AUC from 0.754 to 0.962[20].

In addition to the two methods mentioned above, we also added RF and ANOVA in selecting features after data cleansing since the number of features is high and identifying the influencing features is of paramount importance in clinical disease diagnosis and prognosis predicting. Moreover, except for SVM, LASSO, LR, DT, and RF classifiers, AdaBoost had been used in breast tumors and brain tumors classification[24, 25], MLP was applied in bladder cancer diagnosis[26], breast cancer detection and diagnosis[27], classification of skin cancer[28], and specific Borrmann classification in advanced gastric cancer[12], GP regression was proved to be of some value for predicting the survival time of a cancer patient based on his/her genome-wide gene expression, LDA could be considered as appropriate tool for classifying bladder cancer cases and searching for important biomarkers[29], and NB was capable of achieving 98% accuracy of predicting breast cancer and 90% of predicting lung cancer[30]. However, a single model may not make the best predictions and may be subject to errors such as variance and bias and combining several models to a single model could reduce these errors and improve the predictions. So, all of ten ML techniques were proposed in our study to evaluate the ability of predicting the recurrence and metastasis of GISTs and they were compared with each other and the best model in two phase was choose to stack an ensemble model. Definitely, the ensemble model showed good performance in RM prediction of GISTs with high accuracy, recall, precision, and F1 Score in our study.

However, several shortcomings in our research should be noted. First, the positive group with recurrence or metastasis type was relatively small, and it may affect the statistical stability. We will increase the sample size next. Second, in the ensemble model of machine learning, we only tried to integrate the two best models, not all models. And other ensemble learning methods will be researched and compared with each other. Third, because this was a retrospective single-center study, the patient population and imaging methods were heterogeneous and selection bias may exist, this may result in lacking external data validation, which is essential in the future research.

Conclusion

Our study indicates that CT radiomics combined with machine learning methods is of powerful preoperative value for recurrence or metastasis prediction of GISTs, and may help better stratify patients for clinicians to select optimal treatment strategies and an individualized management to improve clinical outcomes.

List Of Abbreviations

AdaBoost, Adaptive Boosting; AE, Auto-Encoder; AFIP, Armed Forces Institute of Pathology; AUC, area under the curve; ANOVA, analysis of variance; DT, Decision Tree; GP, Gaussian process; GIST, gastrointestinal stromal tumor; KW, Kruskal Wallis; LASSO, Least Absolute Shrinkage and Selection Operator; LDA, linear discriminant analysis; LR, Logistic Regression; ML, Machine learning; MLP, multi-layer perceptron; MSKCC, Memorial Sloan Kettering Cancer Center; NB, Naive Bayes; NIH, National Institute of Health; NRM, non-recurrence and metastasis; RM, recurrence or metastasis; RF, Random

Forest; RFE, recursive feature elimination; SMOTE, Synthetic Minority Oversampling Technique; SVM, Support Vector Machines.

Declarations

Ethics approval and consent to participate Ethical approval for this retrospective study was obtained from the institutional review board of the First Affiliated Hospital of Nanjing Medical University and informed consent was waived. All procedures were performed in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments.

Consent for publication Not applicable

Availability of data and materials All data generated or analysed during this study are included in this published article [and its supplementary information files].

Conflicts of interest The authors declare that they have no conflict of interest.

Funding This study is supported by a Key Social Development Program for the Ministry of Science and Technology of Jiangsu Province (BE2017772, XSL)

Authors' contributions Conceptualization: Xi-Sheng Liu; Material preparation: Lu-Lu Xu, Bo Tang, and Xiao-Ting Jiang; Methodology: Yu-dong Zhang; Data collection and analysis: Bo Tang, Qiu-Xia Feng, and Qiong Li; Writing - original draft preparation: Qiu-Xia Feng; Writing - review and editing: Xi-Sheng Liu; Study supervision and guarantors: Xi-Sheng Liu.

Acknowledgements The authors of this manuscript declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled “**Application of CT radiomics combined with machine learning methods in predicting the recurrence or metastasis of gastrointestinal stromal tumors**”.

References

1. Nishida T, Blay JY, Hirota S, Kitagawa Y, Kang YK: **The standard diagnosis, treatment, and follow-up of gastrointestinal stromal tumors based on guidelines.** *Gastric Cancer* 2016, **19**(1):3–14.
2. Trinh VQ, Dashti NK, Cates JMM: **A proposed risk assessment score for gastrointestinal stromal tumors based on evaluation of 19,030 cases from the National Cancer Database.** *J Gastroenterol* 2021, **56**(11):964–975.
3. Greco A, Rossi S, Ruffolo C, Pauletti B, Massani M: **Evidence for improvements to risk stratification in high-risk gastrointestinal stromal tumor patients.** *Gastrointestinal Cancer Targets and Therapy* 2018,

4. Patel SR, Reichardt P: **An updated review of the treatment landscape for advanced gastrointestinal stromal tumors.** *Cancer* 2021, **127**(13):2187–2195.
5. Ishikawa T, Kanda T, Kameyama H, Wakai T: **Neoadjuvant therapy for gastrointestinal stromal tumor.** *Transl Gastroenterol Hepatol* 2018, **3**:3.
6. Iwatsuki M, Harada K, Iwagami S, Eto K, Ishimoto T, Baba Y, Yoshida N, Ajani JA, Baba H: **Neoadjuvant and adjuvant therapy for gastrointestinal stromal tumors.** *Ann Gastroenterol Surg* 2019, **3**(1):43–49.
7. Bednarski BK, Araujo DM, Yi M, Torres KE, Lazar A, Trent JC, Cormier JN, Pisters PW, Lev DC, Pollock RE *et al*: **Analysis of prognostic factors impacting oncologic outcomes after neoadjuvant tyrosine kinase inhibitor therapy for gastrointestinal stromal tumors.** *Ann Surg Oncol* 2014, **21**(8):2499–2505.
8. Parab TM, DeRogatis MJ, Boaz AM, Grasso SA, Issack PS, Duarte DA, Urayeneza O, Vahdat S, Qiao JH, Hinika GS: **Gastrointestinal stromal tumors: a comprehensive review.** *J Gastrointest Oncol* 2019, **10**(1):144–154.
9. Stanek M, Pisarska M, Rzepa A, Radkowiak D, Major P, Budzynski A: **Laparoscopic treatment of large gastrointestinal stromal tumors (> 5 cm).** *Wideochir Inne Tech Maloinwazyjne* 2019, **14**(2):170-175.
10. Zhang H, Liu Q: **Prognostic Indicators for Gastrointestinal Stromal Tumors: A Review.** *Transl Oncol* 2020, **13**(10):100812.
11. Uhlig J, Uhlig A, Kunze M, Beissbarth T, Fischer U, Lotz J, Wienbeck S: **Novel Breast Imaging and Machine Learning: Predicting Breast Lesion Malignancy at Cone-Beam CT Using Machine Learning Techniques.** *AJR Am J Roentgenol* 2018, **211**(2):W123-W131.
12. Wang S, Dong D, Zhang W, Hu H, Li H, Zhu Y, Zhou J, Shan X, Tian J: **Specific Borrmann classification in advanced gastric cancer by an ensemble multilayer perceptron network: a multicenter research.** *Med Phys* 2021, **48**(9):5017–5028.
13. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI: **Machine learning applications in cancer prognosis and prediction.** *Comput Struct Biotechnol J* 2015, **13**:8–17.
14. Song Y, Zhang J, Zhang YD, Hou Y, Yan X, Wang Y, Zhou M, Yao YF, Yang G: **FeAture Explorer (FAE): A tool for developing and comparing radiomics models.** *PLoS One* 2020, **15**(8):e0237587.
15. Supsamutchai C, Wilasrusmee C, Hiranyatheeb P, Jirasiritham J, Rakchob T, Choikrua P: **A cohort study of prognostic factors associated with recurrence or metastasis of gastrointestinal stromal tumor (GIST) of stomach.** *Ann Med Surg (Lond)* 2018, **35**:1–5.
16. Fletcher CD, Berman JJ, Corless C, Gorstein F, Lasota J, Longley BJ, Miettinen M, O'Leary TJ, Remotti H, Rubin BP *et al*: **Diagnosis of gastrointestinal stromal tumors: A consensus approach.** *Hum Pathol* 2002, **33**(5):459–465.
17. Fujinaga A, Ohta M, Masuda T, Itai Y, Nakanuma H, Kawasaki T, Kawano Y, Hirashita T, Endo Y, Inomata M: **Recurrence of gastric gastrointestinal stromal tumor 12 years after repeat hepatectomies for liver metastases: report of a case.** *Clin J Gastroenterol* 2021, **14**(6):1637–1641.

18. DeMatteo RP, Lewis JJ, Leung D, Mudan SS, Woodruff JM, Brennan MF: **Two hundred gastrointestinal stromal tumors: recurrence patterns and prognostic factors for survival.** *Ann Surg* 2000, **231**(1):51–58.
19. Ao W, Cheng G, Lin B, Yang R, Liu X, Zhou S, Wang W, Fang Z, Tian F, Yang G *et al*: **A novel CT-based radiomic nomogram for predicting the recurrence and metastasis of gastric stromal tumors.** *Am J Cancer Res* 2021, **11**(6):3123–3134.
20. Cannella R, La Grutta L, Midiri M, Bartolotta TV: **New advances in radiomics of gastrointestinal stromal tumors.** *World J Gastroenterol* 2020, **26**(32):4729–4738.
21. Mao H, Zhang B, Zou M, Huang Y, Yang L, Wang C, Pang P, Zhao Z: **MRI-Based Radiomics Models for Predicting Risk Classification of Gastrointestinal Stromal Tumors.** *Front Oncol* 2021, **11**:631927.
22. Chen T, Ning Z, Xu L, Feng X, Han S, Roth HR, Xiong W, Zhao X, Hu Y, Liu H *et al*: **Radiomics nomogram for predicting the malignant potential of gastrointestinal stromal tumours preoperatively.** *Eur Radiol* 2019, **29**(3):1074–1082.
23. Wang J, Xie Z, Zhu X, Niu Z, Ji H, He L, Hu Q, Zhang C: **Differentiation of gastric schwannomas from gastrointestinal stromal tumors by CT using machine learning.** *Abdom Radiol (NY)* 2021, **46**(5):1773–1782.
24. Minz A, Mahobiya C: **MR Image Classification Using Adaboost for Brain Tumor Type.** In: *IEEE International Advance Computing Conference: 2017; 2017*: 701-705.
25. Huang Q, Chen Y, Liu L, Tao D, Li X: **On Combining Biclustering Mining and AdaBoost for Breast Tumor Classification.** *IEEE Transactions on Knowledge and Data Engineering* 2019:1–1.
26. Lorencin I, Andelic N, Spanjol J, Car Z: **Using multi-layer perceptron with Laplacian edge detector for bladder cancer diagnosis.** *Artif Intell Med* 2020, **102**:101746.
27. A M, Ms B: **An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN).** *Clinical eHealth* 2021, **4**:1-11.
28. Jusman Y, Firdiantika IM, Dharmawan DA, Purwanto K: **Performance of Multi Layer Perceptron and Deep Neural Networks in Skin Cancer Classification.** In: *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech): 2021; 2021*.
29. Raeisi Shahraki H, Bemani P, Jalali M: **Classification of Bladder Cancer Patients via Penalized Linear Discriminant Analysis.** *Asian Pac J Cancer Prev* 2017, **18**(5):1453–1457.
30. Kamel H, Abdulah D, Al-Tuwaijari JM: **Cancer Classification Using Gaussian Naive Bayes Algorithm.** In: *2019 International Engineering Conference (IEC): 2019; 2019*.

Figures



Figure 1

Study-process diagram.

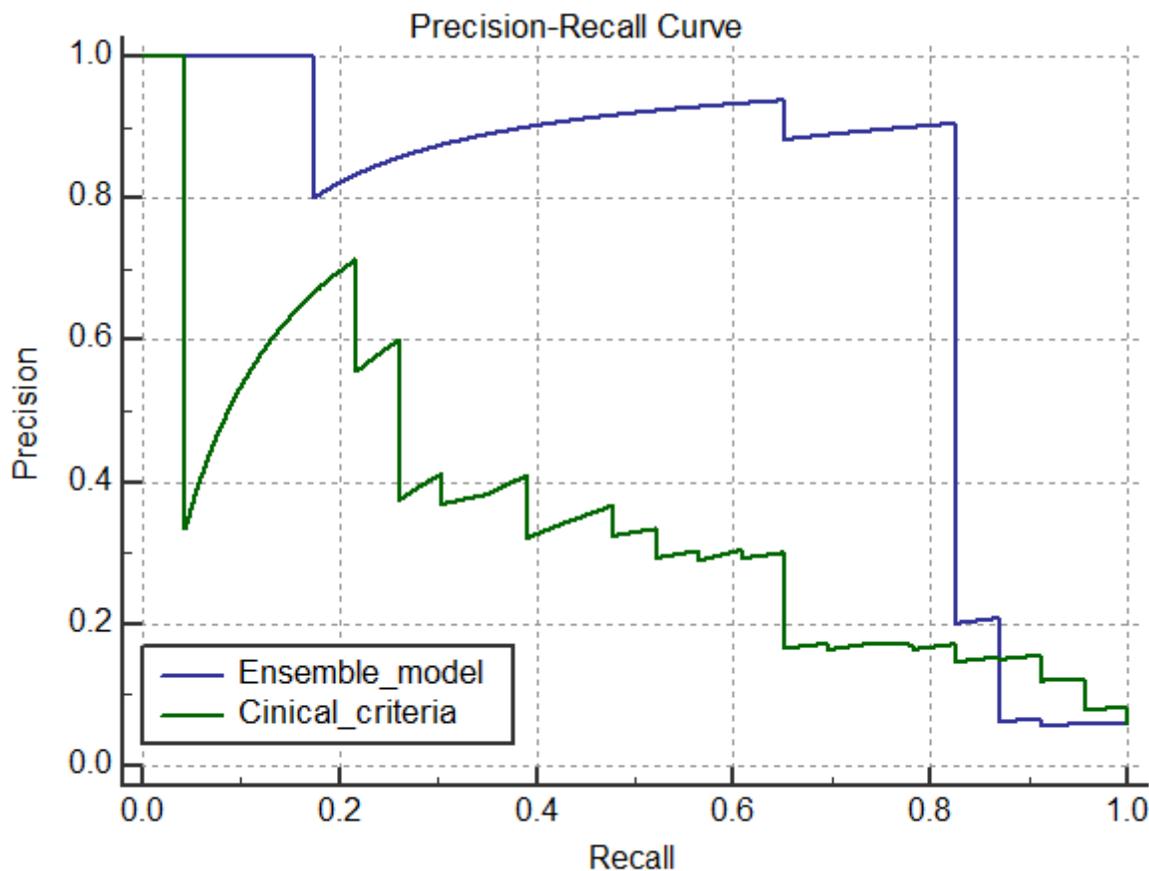


Figure 2

Precision-Recall curves for the ensemble model and clinical criteria

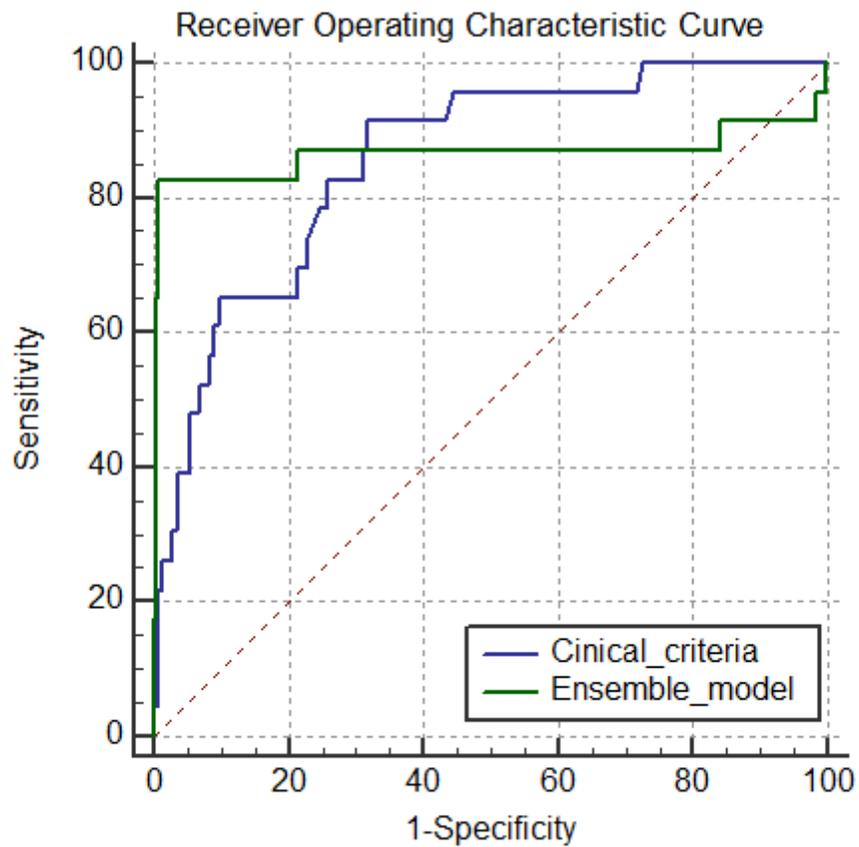


Figure 3

Receiver Operating Characteristic curves for the ensemble model and clinical criteria

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TextS1.docx](#)
- [TableS1.docx](#)
- [rawdataArterial.xlsx](#)
- [rawdataVenous.xlsx](#)