



Machine Learning-Assisted Ensemble Analysis for the Prediction of Response to Neoadjuvant Chemotherapy in Locally Advanced Cervical Cancer

OPEN ACCESS

Yibao Huang[†], Qingqing Zhu[†], Liru Xue, Xiaoran Zhu, Yingying Chen and Mingfu Wu*

Edited by:

Stefano Restaino,
Ospedale Santa Maria della
Misericordia di Udine, Italy

Reviewed by:

Anna Myriam Perrone,
Sant'Orsola-Malpighi Polyclinic, Italy
Carlo Ronsini,
Università degli Studi della Campania
"Luigi Vanvitelli", Italy

*Correspondence:

Mingfu Wu
wu_mingfu@tjh.tjmu.edu.cn

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Gynecological Oncology,
a section of the journal
Frontiers in Oncology

Received: 17 November 2021

Accepted: 28 February 2022

Published: 29 March 2022

Citation:

Huang Y, Zhu Q, Xue L, Zhu X, Chen Y
and Wu M (2022) Machine Learning-
Assisted Ensemble Analysis for the
Prediction of Response to
Neoadjuvant Chemotherapy in Locally
Advanced Cervical Cancer.
Front. Oncol. 12:817250.
doi: 10.3389/fonc.2022.817250

Department of Gynecology, National Clinical Research Center for Obstetrical and Gynecological Diseases, Key Laboratory of Cancer Invasion and Metastasis, Ministry of Education, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

The clinical benefit of neoadjuvant chemotherapy (NACT) before concurrent chemoradiotherapy (CCRT) vs. adjuvant chemotherapy after CCRT is debated. Non-response to platinum-based NACT is a major contributor to poor prognosis, but there is currently no reliable method for predicting the response to NACT (rNACT) in patients with locally advanced cervical cancer (LACC). In this study we developed a machine learning (ML)-assisted model to accurately predict rNACT. We retrospectively analyzed data on 636 patients diagnosed with stage IB2 to IIA2 cervical cancer at our hospital between January 1, 2010 and December 1, 2020. Five ML-assisted models were developed from candidate clinical features using 2-step estimation methods. Receiver operating characteristic curve (ROC), clinical impact curve, and decision curve analyses were performed to evaluate the robustness and clinical applicability of each model. A total of 30 candidate variables were ultimately included in the rNACT prediction model. The areas under the ROC curve of models constructed using the random forest classifier (RFC), support vector machine, eXtreme gradient boosting, artificial neural network, and decision tree ranged from 0.682 to 0.847. The RFC model had the highest predictive accuracy, which was achieved by incorporating inflammatory factors such as platelet-to-lymphocyte ratio, neutrophil-to-lymphocyte ratio, neutrophil-to-albumin ratio, and lymphocyte-to-monocyte ratio. These results demonstrate that the ML-based prediction model developed using the RFC can be used to identify LACC patients who are likely to respond to rNACT, which can guide treatment selection and improve clinical outcomes.

Keywords: locally advanced cervical cancer, neoadjuvant chemotherapy, machine learning analysis, predictive model, pathology response

INTRODUCTION

Cervical cancer is a malignant tumor and major cause of morbidity and mortality, with an estimated 500,000 new cases and 300,000 deaths each year (1, 2). Over the past decade, substantial progress has been made in the early diagnosis and treatment of locally advanced cervical cancer (LACC) (3). The standard treatments for LACC are surgery, radiation therapy, and chemotherapy (CT) (3), but none of these are optimal. Local residual disease following chemoradiotherapy (CRT) can be treated by salvage surgery; however, this is associated with various complications (4, 5). Additionally, the response of LACC patients to radical surgery after radiotherapy and CT is generally poor, while radiotherapy may not be a treatment option in low-income countries (6, 7). As such, there is a need for more effective and accessible treatment options for LACC.

As a potential alternative therapy, platinum-based neoadjuvant (NA)CT has been shown to reduce tumor volume (3, 8). According to the International Federation of Gynecology and Obstetrics (FIGO) classification system of 2009, NACT can be considered for patients with stage IB2 to IIA2 LACC, especially before radical hysterectomy (9–11). Cisplatin-based NACT is associated with improved long-term survival rates in LACC (12–14). However, there are currently no models that can accurately predict the pathologic response to NACT in these patients, although this could facilitate clinical management. Machine learning (ML) is a data analysis method with applications in healthcare (15). Compared to conventional statistical models, ML-based ensemble analysis can ensure robustness of a statistical model and improve its predictive accuracy through iterative algorithms.

There is no consensus on the cutoff for optimal response to NACT (rNACT) in patients with LACC, with overall response rates to NACT ranging from 52 to 95% (16, 17). In this study, we applied ML-based algorithms to establish a model to accurately predict rNACT in LACC patients by using preoperative clinical parameters and inflammatory markers.

METHODS

Patient Selection

Patients who were diagnosed with FIGO stage IB2–IIA2 cervical cancer at the Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology (Wuhan, China) between January 2010 and December 2020 were retrospectively enrolled in this study. The inclusion criteria for patients were as follows: (i) received platinum-based NACT, without adverse effects; (ii) received a standard cycle of NACT before the operation with no other treatment; (iii) underwent systematic physical examination before the operation, including peripheral blood monitoring and imaging examination; and (iv) a complete set of medical data was available. We excluded patients who had severe organ injuries or incomplete clinical parameters, laboratory test results, and imaging findings in their medical records. The study protocol was in compliance with the

provisions of the Helsinki Declaration (2013 revision) and was approved by the Institutional Review Committee of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology (TJ-IRB20210631). All information of the patients was strictly confidential and informed consent was waived due to its traceability. The workflow for LACC patient selection and model construction is summarized in **Figure 1**. The study was approved by the Institutional Review Committee of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology (TJ-IRB20210631).

Data Collection and Quality Assessment

The following data were collected for all patients: age, body mass index, histology, FIGO stage, tumor size before NACT, histologic grade, lymph node metastasis, lymph vascular space invasion, parametrial involvement, surgical margin, neutrophil count ($10^9/l$), lymphocyte count ($10^9/l$), platelet count ($10^9/l$), monocyte count ($10^9/l$), hemoglobin, albumin, globulin, and tumor markers. For variables with missing values, the median was typically used. If $\geq 10\%$ of values were missing for a given variable, it was excluded from variable screening for the final model.

Evaluation Criteria for NACT

For pathologic response assessment, all patients who received NACT were independently examined by 2 pathologists. rNACT was evaluated according to the Response Evaluation Criteria in Solid Tumors (RECIST) v1.1 criteria (18), and was categorized into the following 4 levels according to the presence or absence of

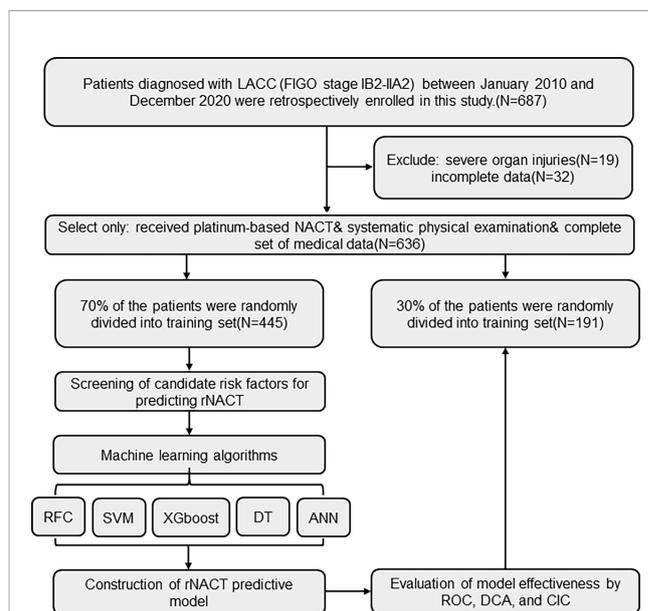


FIGURE 1 | Flow diagram of patient selection and data processing. ANN, artificial neural network; CIC, clinical impact curve; DCA, decision curve analysis; DT, decision tree; FIGO, International Federation of Gynecology and Obstetrics; RFC, random forest classifier; ROC, receiver operating characteristic; SVM, support vector machine; XGboost, eXtreme gradient boosting.

pathologic response: (i) complete response (CR), almost complete disappearance of cancer lesions; (ii) partial response (PR), $\geq 30\%$ decrease in total maximum diameter of cancer lesions; (iii) progressive disease (PD), $\geq 20\%$ increase in total maximum diameter of cancer lesions; and (iv) stable disease (SD), total maximum diameter of cancer lesions defined as either insufficient contraction in line with PR or increased compliance with PD. LACC patients were considered to be rNACT if they were determined as having CR or PR following NACT treatment; meanwhile, patients with SD or PD were regarded as non-rNACT.

Development and Validation of ML-Based Models

Four ML-based algorithms were used to build predictive models. We used the Classification and Regression Training (caret) package to randomly divide the dataset into 2 parts, 70% for model training and 30% for model validation. A total of 5 ML-based algorithms were evaluated for the predictive model. The model variables were screened by 2-step estimation (19) according to the following formula.

$$\hat{m} = \hat{m}_3 \cap \hat{m}_4$$

The characteristic variable was X and the target variable was Y ; these were evenly divided into 2 parts—namely, X_1 , Y_1 and X_2 , Y_2 . Through univariate screening, the variable quantum set m^1 was screened on X_1 and Y_1 , and m^2 was filtered by X_2 and Y_2 . A lasso was then used to re-fit the model, and the filtered variables were marked as m^3 and m^4 . The optimal subset for modeling was obtained based on the intersection of the variable sets. The model was evaluated by inspection, discrimination, and calibration. A receiver operating characteristic (ROC) curve was used to assess the predictive accuracy of the model in the training and validation sets. The discriminatory ability of each model was quantified by the area under the ROC curve (AUC), decision curve analysis (DCA), and clinical impact curve (CIC) analysis.

Statistical Analysis

Median (interquartile range) and frequencies (%) were described for continuous and categorical variables, respectively. The chi-squared test or Mann–Whitney U test was used to compare baseline clinical characteristics between the rNACT and non-rNACT cohorts as appropriate. All analyses were performed using Python v3.9.2 (<https://www.python.org/>) and R v4.0.4 (<http://www.r-project.org/>). All tests were 2-sided, and $p < 0.05$ was considered statistically significant.

RESULTS

Baseline Characteristics of the Study Population and rNACT

The detailed clinical characteristics and pathologic baseline data of 636 patients with LACC are shown in **Table 1**. All patients

received platinum-based NACT, with no serious adverse reactions. For internal validation, patients were randomly divided into a training set ($N = 445$, 70%) and validation set ($N = 191$, 30%) using the caret package. According to the RECIST criteria, 396 (88.9%) and 162 (84.8%) patients showed rNACT in the training and validation sets, respectively, indicating that these patients were sensitive to NACT. Follow-up treatment was determined according to the condition of the patients and included radical surgery, radiation, and concurrent (C)CRT. Of these patients, 614 (96.6%) underwent radical surgery and 91 (14.3%) received radiotherapy or CCRT.

Selection of Candidate Variables Using Different ML-Based Algorithms

Candidate covariates of each algorithm were filtered and 30 were included in the correlation analysis between outcome and independent variables. rNACT was significantly correlated with inflammatory factors and clinical variables, namely, platelet-to-lymphocyte ratio (PLR), neutrophil-to-lymphocyte ratio (NLR), neutrophil-to-albumin ratio (NAR), lymphocyte-to-monocyte ratio (LMR), and tumor size (**Figure 2A**). PLR, LMR, and NAR were important factors in the ML-based model (**Figure 2B**). Consistent with the results of correlation analysis, the 5 top-ranked predictors were PLR, NLR, NAR, LMR, and tumor size.

Construction of ML-Based rNACT Predictive Model

Random forest classifier (RFC) and decision tree (DT) are commonly used ML-based algorithms in supervised learning. The RFC model was constructed using the formula $I(X = x_i) = -\log_2 P(x_i)$, where $I(X)$ is the information for candidate variables and $P(x_i)$ is the probability of x_i (**Figure 3A**). Thirty variables were ordered according to the mean decrease in Gini index (**Supplementary Table 1**); the top 10 ranked variables were used to construct the optimal RFC prediction model, which included PLR, NLR, NAR, LMR, and tumor size. Inflammatory factors, PLR, LMR, and tumor size served as irreplaceable weights at DT branches (**Figure 3B**). Using the iterative algorithm of supervised learning, both RFC and DT models were used for rNACT prediction.

Comparison Among ML-Based Models

Based on the iterative analysis of baseline characteristics, we used 5 supervised learning models for NACT risk assessment and to optimize predictive performance. As expected, the RFC model was better able to distinguish between LACC patients in the rNACT and non-rNACT cohorts. The AUCs of the RFC model reached a plateau when *** variables were introduced, indicating that the RFC model had the highest predictive accuracy, followed by DT, artificial neural network (ANN), support vector machine (SVM), and eXtreme gradient boosting (XGBoost) models (**Figure 4A**). The predictive performance of ML-based models is summarized in **Table 2**. Consistent with the results of the ROC analysis, the RFC model also showed a robust predictive performance in the DCA (**Figure 4B**).

TABLE 1 | Baseline demographic and clinicopathologic features of patients with LACC with and without a diagnosis of rNACT.

Variables	Overall N = 636	Training set		P-value	Testing set		P-value
		Responders (N = 396)	Non-responders (N = 49)		Responders (N = 162)	Non-responders (N = 29)	
Age, years	47.00 [42.00, 52.00]	47.00 [42.75, 52.00]	49.00 [41.00, 54.00]	0.713	47.00 [43.00, 52.00]	45.00 [37.00, 52.00]	0.229
Weight, kg	55.44 [50.00, 60.00]	55.44 [50.00, 60.00]	55.44 [49.00, 59.00]	0.563	55.44 [50.00, 60.00]	55.00 [53.00, 60.00]	0.642
Height, m	157.40 [154.00, 160.00]	157.40 [154.00, 160.62]	157.40 [155.00, 160.00]	0.86	157.40 [155.00, 160.00]	158.00 [154.00, 160.00]	0.934
Smoking							
Yes	34 (5.3)	22 (5.6)	1 (2.0)	0.48	10 (6.2)	1 (3.4)	0.883
No	602 (94.7)	374 (94.4)	48 (98.0)		152 (93.8)	28 (96.6)	
FIGO stage							
Ib	4 (0.6)	2 (0.5)	1 (2.0)	0.356	1 (0.6)	0 (0.0)	0.845
Ib2	402 (63.2)	256 (64.6)	28 (57.1)		99 (61.1)	19 (65.5)	
IIa	12 (1.9)	7 (1.8)	2 (4.1)		3 (1.9)	0 (0.0)	
IIa2	218 (34.3)	131 (33.1)	18 (36.7)		59 (36.4)	10 (34.5)	
Histology							
SCC	542 (85.2)	340 (85.9)	44 (89.8)	0.235	134 (82.7)	24 (82.8)	0.921
ADC	68 (10.7)	42 (10.6)	3 (6.1)		19 (11.7)	4 (13.8)	
AdCa	9 (1.4)	5 (1.3)	2 (4.1)		2 (1.2)	0 (0.0)	
Other	17 (2.7)	9 (2.3)	0 (0.0)		7 (4.3)	1 (3.4)	
Tumor grade							
G1	37 (5.8)	28 (7.1)	1 (2.0)	0.063	7 (4.3)	1 (3.4)	0.04
G2	278 (43.7)	186 (47.0)	19 (38.8)		64 (39.5)	9 (31.0)	
G3	241 (37.9)	138 (34.8)	26 (53.1)		59 (36.4)	18 (62.1)	
Unknown	80 (12.6)	44 (11.1)	3 (6.1)		32 (19.8)	1 (3.4)	
Tumor size	3.60 [2.50, 4.10]	3.50 [2.30, 4.00]	6.00 [5.00, 7.00]	<0.001	3.50 [2.50, 4.00]	6.00 [5.00, 7.00]	<0.001
Lymphatic invasion							
Yes	85 (13.4)	56 (14.1)	12 (24.5)	0.152	13 (8.0)	4 (13.8)	0.597
No	528 (83.0)	326 (82.3)	36 (73.5)		142 (87.7)	24 (82.8)	
Unknown	23 (3.6)	14 (3.5)	1 (2.0)		7 (4.3)	1 (3.4)	
Parametrial invasion							
Yes	38 (6.0)	21 (5.3)	7 (14.3)	0.033	7 (4.3)	3 (10.3)	0.374
No	598 (94.0)	375 (94.7)	42 (85.7)		155 (95.7)	26 (89.7)	
Pelvic lymph metastasis							
Yes	103 (16.2)	56 (14.1)	22 (44.9)	<0.001	18 (11.1)	7 (24.1)	0.159
No	510 (80.2)	326 (82.3)	26 (53.1)		137 (84.6)	21 (72.4)	
Unknown	23 (3.6)	14 (3.5)	1 (2.0)		7 (4.3)	1 (3.4)	
Paraortic lymph metastasis							
Yes	18 (2.8)	11 (2.8)	4 (8.2)	0.128	3 (1.9)	0 (0.0)	0.74
No	595 (93.6)	371 (93.7)	44 (89.8)		152 (93.8)	28 (96.6)	
Unknown	23 (3.6)	14 (3.5)	1 (2.0)		7 (4.3)	1 (3.4)	
SCC	4.30 [0.80, 4.30]	4.30 [0.80, 4.30]	2.20 [0.60, 4.30]	0.041	4.30 [1.12, 4.30]	1.30 [0.70, 4.30]	0.028
P53							
Positive	32 (5.0)	16 (4.0)	1 (2.0)	0.382	11 (6.8)	4 (13.8)	0.282
Negative	14 (2.2)	11 (2.8)	0 (0.0)		2 (1.2)	1 (3.4)	
Unknown	590 (92.8)	369 (93.2)	48 (98.0)		149 (92.0)	24 (82.8)	
CEA	5.15 [3.44, 5.15]	5.15 [3.10, 5.15]	5.15 [3.61, 5.15]	0.705	5.15 [4.16, 5.15]	5.15 [5.15, 5.15]	0.96
CA125	42.00 [42.00, 42.00]	42.00 [42.00, 42.00]	42.00 [42.00, 42.00]	0.008	42.00 [28.73, 42.00]	42.00 [42.00, 42.00]	0.289
CA199	32.19 [32.19, 32.19]	32.19 [32.19, 32.19]	32.19 [32.19, 32.19]	0.605	32.19 [28.69, 32.19]	32.19 [32.19, 32.19]	0.521
CCRT							
Yes	91 (14.3)	16 (4.0)	44 (89.8)	<0.001	9 (5.6)	22 (75.9)	<0.001
No	530 (83.3)	371 (93.7)	5 (10.2)		147 (90.7)	7 (24.1)	
Unknown	15 (2.4)	9 (2.3)	0 (0.0)		6 (3.7)	0 (0.0)	
Surgical method							
Laparoscopy	485 (76.3)	300 (75.8)	38 (77.6)	0.888	127 (78.4)	20 (69.0)	0.424
Open	129 (20.3)	83 (21.0)	10 (20.4)		28 (17.3)	8 (27.6)	
Unknown	22 (3.5)	13 (3.3)	1 (2.0)		7 (4.3)	1 (3.4)	
Surgical margin							
Positive	11 (1.7)	4 (1.0)	2 (4.1)		4 (2.5)	1 (3.4)	
Negative	603 (94.8)	379 (95.7)	46 (93.9)	0.194	151 (93.2)	27 (93.1)	0.935
Positive	22 (3.5)	13 (3.3)	1 (2.0)		7 (4.3)	1 (3.4)	
LND	31.00 [25.00, 38.00]	31.00 [25.00, 38.00]	32.00 [27.00, 37.00]	0.309	32.00 [25.00, 39.75]	30.00 [26.00, 34.00]	0.302

(Continued)

TABLE 1 | Continued

Variables	Overall N = 636	Training set		P-value	Testing set		P-value
		Responders (N = 396)	Non-responders (N = 49)		Responders (N = 162)	Non-responders (N = 29)	
Platelet count	251.50 [209.00, 276.40]	239.00 [204.75, 276.40]	327.00 [301.00, 370.00]	<0.001	245.50 [204.50, 276.00]	347.00 [318.00, 386.00]	<0.001
Leukocyte count	6.30 [4.92, 7.43]	6.36 [5.03, 7.43]	5.15 [4.17, 7.43]	0.01	6.37 [4.89, 7.43]	6.02 [5.14, 6.53]	0.351
Lymphocyte count	1.63 [1.27, 1.77]	1.63 [1.28, 1.82]	1.40 [0.99, 1.63]	0.006	1.63 [1.33, 1.76]	1.48 [1.26, 1.65]	0.54
Monocyte count	0.43 [0.30, 0.47]	0.43 [0.31, 0.46]	0.37 [0.24, 0.45]	0.059	0.45 [0.32, 0.51]	0.41 [0.25, 0.47]	0.576
Hemoglobin	112.00 [102.00, 122.00]	112.00 [102.00, 122.00]	112.00 [102.00, 119.00]	0.455	112.00 [103.00, 123.00]	114.00 [112.00, 123.00]	0.311
Neutrophil count	3.98 [2.84, 4.74]	4.06 [2.90, 4.92]	3.20 [2.23, 4.38]	0.025	4.06 [2.82, 4.62]	3.62 [2.97, 4.49]	0.564
NLR	2.67 [1.74, 3.47]	2.37 [1.63, 2.97]	10.00 [9.00, 11.00]	<0.001	2.35 [1.72, 2.82]	11.00 [9.00, 12.00]	<0.001
PLR	156.52 [123.27, 180.89]	146.19 [118.82, 171.15]	280.00 [230.00, 356.00]	<0.001	149.72 [119.02, 168.10]	253.00 [222.22, 314.00]	<0.001
LMR	6.80 [4.66, 9.18]	7.22 [5.30, 9.41]	2.91 [1.64, 3.75]	<0.001	7.59 [5.60, 9.41]	3.72 [2.63, 4.66]	<0.001
PNR	63.11 [49.17, 85.37]	63.11 [46.68, 79.01]	104.06 [76.35, 140.00]	<0.001	63.11 [50.83, 78.84]	97.71 [70.83, 122.36]	<0.001
Fibrinogen	4.28 [3.14, 5.56]	3.95 [3.01, 5.00]	6.72 [6.13, 7.18]	<0.001	4.11 [2.99, 5.25]	6.39 [6.05, 7.04]	<0.001
Albumin	39.40 [36.50, 41.20]	39.40 [36.30, 41.00]	39.30 [36.20, 40.80]	0.822	39.40 [37.23, 41.20]	40.10 [38.30, 42.20]	0.343
GGT	23.00 [15.00, 29.55]	22.00 [15.00, 29.55]	29.55 [19.00, 37.00]	0.012	24.00 [15.00, 29.55]	20.00 [15.00, 29.55]	0.42
Globulin	31.20 [28.50, 33.02]	31.20 [28.48, 33.00]	31.30 [29.40, 34.90]	0.165	31.20 [28.52, 32.27]	31.20 [28.70, 32.30]	0.887
A/G	1.26 [1.15, 1.39]	1.26 [1.15, 1.39]	1.26 [1.11, 1.32]	0.343	1.26 [1.19, 1.41]	1.29 [1.21, 1.39]	0.823
PNI	47.55 [43.50, 49.66]	47.55 [43.50, 49.75]	47.20 [42.00, 48.00]	0.245	47.55 [43.60, 50.44]	48.10 [46.10, 49.85]	0.52
NAR	0.26 [0.17, 0.34]	0.23 [0.15, 0.31]	1.27 [0.98, 1.51]	<0.001	0.25 [0.15, 0.32]	1.12 [0.58, 1.44]	<0.001

Data are shown as median [interquartile range] or n (%).

ADC, adenocarcinoma; AdCa, adenosquamous carcinoma; A/G, albumin-to-globulin ratio; CA125, cancer antigen 125; CA199, cancer antigen 199; CCRT, concurrent chemoradiotherapy; CEA, carcinoembryonic antigen; FBG, fasting blood glucose; FIGO, International Federation of Gynecology and Obstetrics; GGT, γ -glutamyltransferase; IQR, interquartile range; LMR, lymphocyte-to-monocyte ratio; LND, lymph node dissection; NAR, neutrophil-to-albumin ratio; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; PNI, prognostic nutrition index; PNR, platelet-to-neutrophil ratio; SCC, squamous cell carcinoma.

Internal Validation of the Optimal Predictive Model

To further validate the performance of the RFC model, we also used CIC to evaluate predictive accuracy. The CIC analysis

revealed rNACT stratification in the training set (Figure 5A). This was supported by the risk factors for rNACT identified in the validation set (Figure 5B), indicating that the selected features were highly relevant to rNACT.

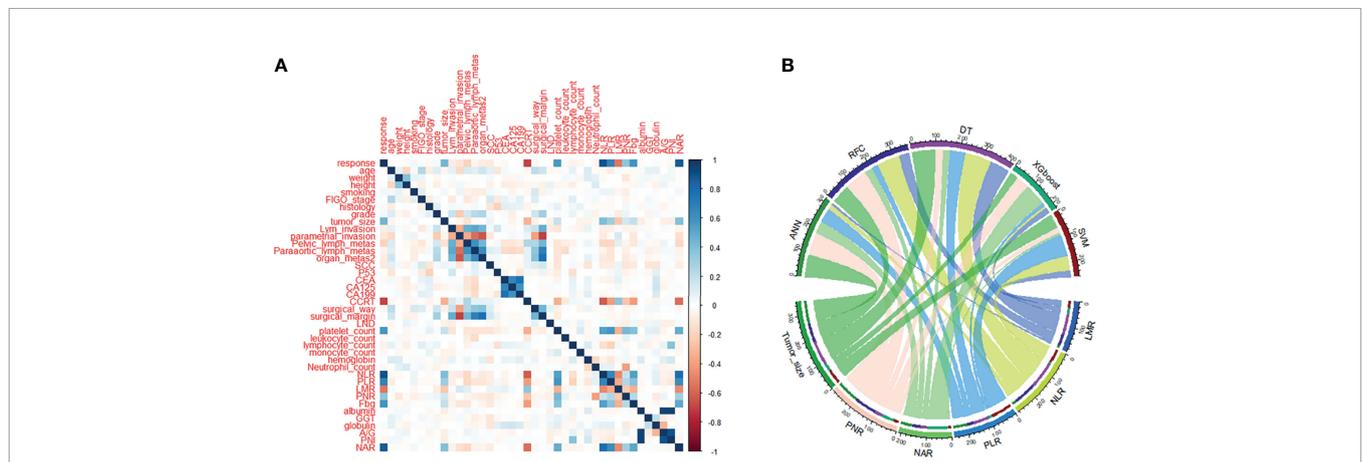


FIGURE 2 | Variable screening and weight allocation. (A) Correlation matrix analysis of candidate features. (B) Weight distribution of the candidate variables of each ML-based model. ADC, denocarcinoma; A/G, albumin-to-globulin ratio; ANN, artificial neural network; CA125, cancer antigen 125; CA199, cancer antigen 199; CCRT, concurrent chemoradiotherapy; CEA, carcinoembryonic antigen; CIC, clinical impact curve; DCA, decision curve analysis; DT, decision tree; FIGO, International Federation of Gynecology and Obstetrics; GGT, serum γ -glutamyltransferase; LMR, lymphocyte-to-monocyte ratio; LND, lymph node dissection; NAR, neutrophil-to-albumin ratio; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; PNI, prognostic nutrition index; PNR, platelet-to-neutrophil ratio; RFC, random forest classifier; ROC, receiver operating characteristic curve; SCC, squamous cell carcinoma; SVM, support vector machine; XGboost, eXtreme gradient boosting.

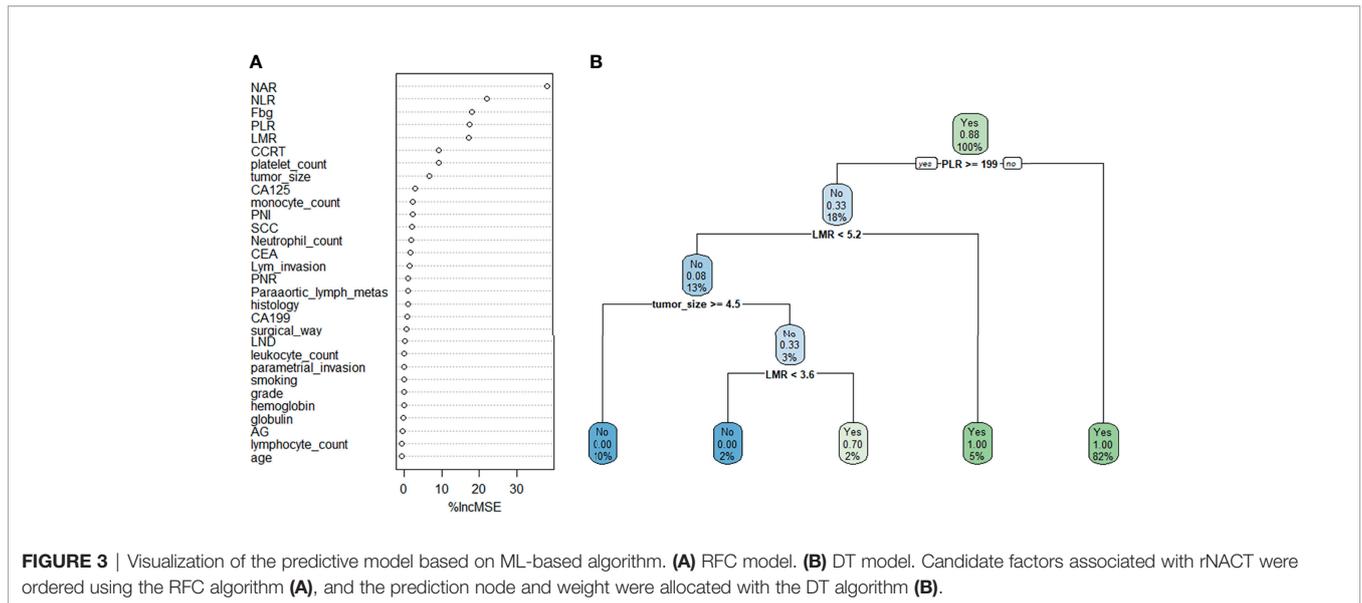


FIGURE 3 | Visualization of the predictive model based on ML-based algorithm. **(A)** RFC model. **(B)** DT model. Candidate factors associated with rNACT were ordered using the RFC algorithm **(A)**, and the prediction node and weight were allocated with the DT algorithm **(B)**.

DISCUSSION

Inaccurate risk stratification of cancer patients can affect clinical decision-making and outcomes. Given the excellent performance of ML-based algorithms in the classification of rNACT, the RFC, DT, ANN, XGboost, and SVM algorithms were used in our study to establish a predictive model for rNACT in LACC patients. There were 2 major findings to our study. First, we achieved accurate risk stratification of LACC patients who received NACT based on markers of systemic inflammation. Second, we developed and validated a novel ML-based predictive model that is superior to existing prediction algorithms to identify LACC patients who would benefit from NACT.

Systemic inflammation plays a critical role in promoting the progression and metastasis of many cancers (20–23), and was shown to be associated with the development, progression, and metastasis of cervical cancer (24). We therefore examined pre-NACT treatment peripheral blood-related inflammatory marker

levels in patients with LACC, and the results were consistent with previous findings (13). As expected, systemic inflammatory markers such as neutrophils, lymphocytes, and platelets and also albumin, C-reactive protein, and other biochemical markers were useful in predicting rNACT in LACC patients. NACT effectively reduces serum levels of tumor markers and NLR and prolongs survival time (25). We compared patients who responded to NACT with those who did not respond and found that in the former, the levels of inflammatory biomarkers were significantly altered compared to before NACT treatment whereas in non-responders, there were no differences between pre- and post-treatment levels. Additionally, we found that PLR, prognostic nutrition index, and LMR were significantly associated with rNACT. Thus, changes in the level of cooperative inflammatory factors can predict the response of LACC patients to NACT.

rNACT was related to NAR and the concentration of fibrinogen, a coagulation factor. The latter has prognostic value

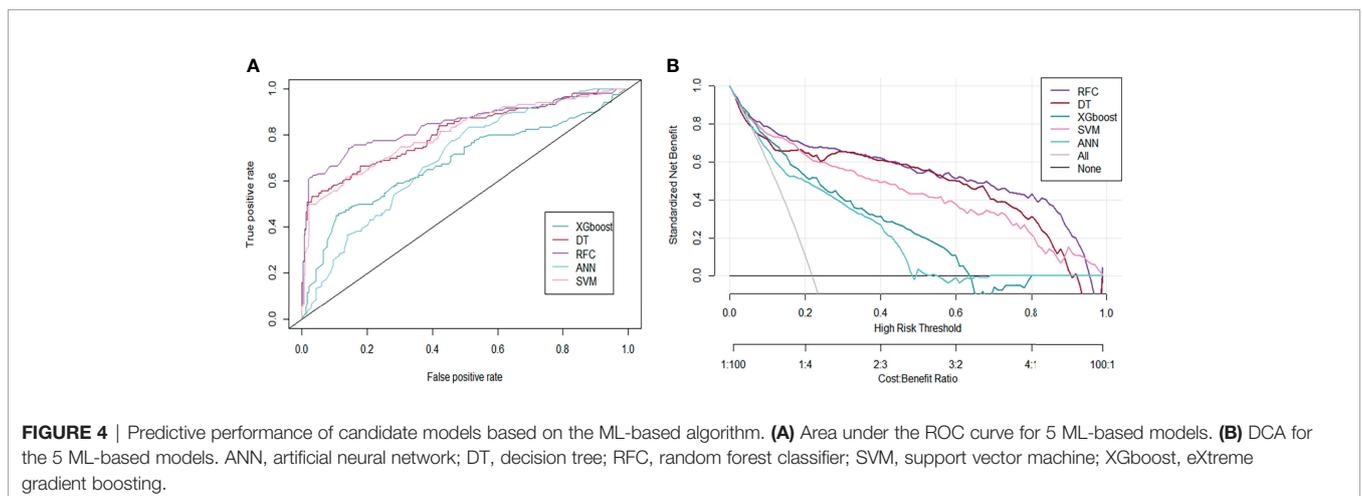
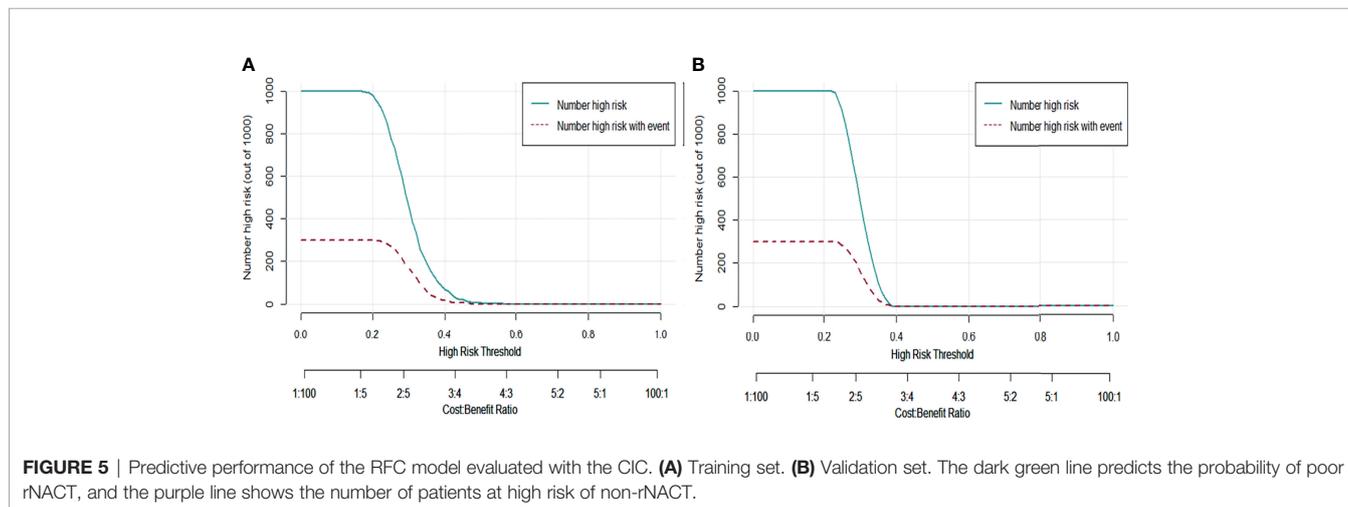


FIGURE 4 | Predictive performance of candidate models based on the ML-based algorithm. **(A)** Area under the ROC curve for 5 ML-based models. **(B)** DCA for the 5 ML-based models. ANN, artificial neural network; DT, decision tree; RFC, random forest classifier; SVM, support vector machine; XGboost, eXtreme gradient boosting.

TABLE 2 | Receiver operating characteristic curve analysis for predicting response to neoadjuvant chemotherapy in each machine learning-based model.

Model	AUC		Number of candidate variables
	Mean	95% CI	
RFC	0.847	0.286–1.408	8
SVM	0.811	0.245–1.377	9
DT	0.807	0.225–1.389	8
ANN	0.682	0.118–1.246	7
XGboost	0.697	0.113–1.281	9

ANN, artificial neural network; AUC, area under the receiver operating characteristic curve; CI, confidence interval; DT, decision tree; RFC, random forest classifier; SVM, support vector machine; XGboost, eXtreme gradient boosting.



in many cancer types, namely, hepatocellular carcinoma (26), gastrointestinal stromal tumors (27), and colorectal cancer (28), and was found to predict the levels of inflammatory factors in our study. In the weighting of the prediction model, it was an index that optimized the accuracy and robustness of rNACT prediction by the RFC model.

Our ML-based model incorporated clinical parameters and laboratory test results according to previous reports (12–14); clinical indicators including larger tumor size and earlier stage were shown to be independent predictors of rNACT (12). We therefore evaluated the predictive accuracy of the models and found that systemic inflammatory markers had a large weight in each model. The RFC model allowed calculation of risk level based on all variables collected from medical records, and yielded the highest predictive accuracy. The RFC uses the bootstrapping resampling technique to reduce variance (29). DCA and CIC analysis were used to assess the predictive performance of the RFC model, and the results showed that the model was able to discriminate between rNACT and non-rNACT cohorts. Thus, the model can be used to identify LACC patients who may benefit from NACT before surgery.

There were some limitations to this study. First, there was selection bias as only patients from a tertiary referral hospital were included. Second, although the predictive model was validated in our study, it may not be applicable to other patient populations; therefore, it must be tested using external data. Third, the ML-based model was only for patients with stage

Ib2 to IIa2 LACC; further research is needed to determine whether it can be applied to patients at different stages.

CONCLUSION

We developed a ML-based algorithm to identify factors that can predict rNACT in patients with LACC. The model constructed using the RFC had the highest predictive accuracy, with PLR, NLR, NAR, LMR, and tumor size being the most important predictors. Thus, a combination of clinical data and systemic inflammatory markers may aid clinicians in individual risk assessment of rNACT.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Institutional Review Committee of Tongji

Hospital, Tongji Medical College, Huazhong University of Science and Technology. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

YBH and QQZ: study conception and planning, statistical analysis, data interpretation, manuscript drafting, and final approval of manuscript. LRX, XRZ, and YYC: data interpretation and final approval of the manuscript. MFW: study conception and planning, statistical analysis, data interpretation, manuscript drafting, and final approval of the manuscript. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

- Wang L, Zhao Y, Wang Y, Wu X. The Role of Galectins in Cervical Cancer Biology and Progression. *BioMed Res Int* (2018) 2018:2175927. doi: 10.1155/2018/2175927
- Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, et al. Estimates of Incidence and Mortality of Cervical Cancer in 2018: A Worldwide Analysis. *Lancet Glob Health* (2020) 8:e191–203. doi: 10.1016/s2214-109x(19)30482-6
- Klyuchko KO, Gargin VV. Influence of Neoadjuvant Chemoradiotherapy for Locally Advanced Cervical Cancer. *Pol Merkur Lekarski* (2020) 48:406–9.
- van Kol KGG, Ebisch RMF, Piek MJ, Zusterzeel PLM, Vergeldt TFM, Bekkers RLM. Salvage Surgery for Patients With Residual Disease After Chemoradiation Therapy for Locally Advanced Cervical Cancer: A Systematic Review on Indication, Complications, and Survival. *Acta Obstet Gynecol Scand* (2021) 100:1176–85. doi: 10.1111/aogs.14093
- Horeweg N, Mittal P, Gradowska PL, Boere I, Chopra S, Nout RA. Adjuvant Systemic Therapy After Chemoradiation and Brachytherapy for Locally Advanced Cervical Cancer: A Systematic Review and Meta-Analysis. *Cancers* (2021) 13:1880. doi: 10.3390/cancers13081880
- Ferrandina G, Gallotta V, Federico A, Fanfani F, Ercoli A, Chiantera V, et al. Minimally Invasive Approaches in Locally Advanced Cervical Cancer Patients Undergoing Radical Surgery After Chemoradiotherapy: A Propensity Score Analysis. *Ann Surg Oncol* (2021) 28:3616–26. doi: 10.1245/s10434-020-09302-y
- Yin M, Zhao F, Lou G, Zhang H, Sun M, Li C, et al. The Long-Term Efficacy of Neoadjuvant Chemotherapy Followed by Radical Hysterectomy Compared With Radical Surgery Alone or Concurrent Chemoradiotherapy on Locally Advanced-Stage Cervical Cancer. *Int J Gynecol Cancer* (2011) 21:92–9. doi: 10.1111/IGC.0b013e3181fe8b6e
- Voinea S, Herghelegiu CG, Sandru A, Ioan RG, Bohiltea RE, Bacalbaşa N, et al. Impact of Histological Subtype on the Response to Chemoradiation in Locally Advanced Cervical Cancer and the Possible Role of Surgery. *Exp Ther Med* (2021) 21:93. doi: 10.3892/etm.2020.9525
- Lapresa M, Parma G, Portuesi R, Colombo N. Neoadjuvant Chemotherapy in Cervical Cancer: An Update. *Expert Rev Anticancer Ther* (2015) 15:1171–81. doi: 10.1586/14737140.2015.1079777
- Li R, Lu ST, Si JG, Liu B, Wang H, Mei YY, et al. Prognostic Value of Responsiveness of Neoadjuvant Chemotherapy Before Surgery for Patients With Stage IB(2)/IIA(2) Cervical Cancer. *Gynecol Oncol* (2013) 128:524–9. doi: 10.1016/j.ygyno.2012.11.006
- Feng X, Chen H, Li L, Gao L, Wang L, Bai X. Postoperative Adjuvant Chemotherapy Improved the Prognosis in Locally Advanced Cervical

FUNDING

This research was supported by the Applied Basic Research Program of WMBST (no. 2019020701011436).

ACKNOWLEDGMENTS

The authors thank all study participants for consenting to the use of their medical records, and the staff members of Tongji Hospital for their assistance with data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.817250/full#supplementary-material>

- Cancer Patients With Optimal Response to Neoadjuvant Chemotherapy. *Front Oncol* (2020) 10:608333:608333. doi: 10.3389/fonc.2020.608333
- Huang Y, Liu L, Cai J, Yang L, Sun S, Zhao J, et al. The Efficacy and Response Predictors of Platinum-Based Neoadjuvant Chemotherapy in Locally Advanced Cervical Cancer. *Cancer Manag Res* (2020) 12:10469–77. doi: 10.2147/CMAR.S270258
- Chen W, Wang B, Zeng R, Wang T. Development and Validation of a Nomogram for the Estimation of Response to Platinum-Based Neoadjuvant Chemotherapy in Patients With Locally Advanced Cervical Cancer. *Cancer Manag Res* (2021) 13:1279–89. doi: 10.2147/CMAR.S293268
- Ou Z, Zhao D, Li B, Wang Y, Liu S, Zhang Y. A Preoperative Nomogram for Predicting Chemoresistance to Neoadjuvant Chemotherapy in Patients With Locally Advanced Cervical Squamous Carcinoma Treated With Radical Hysterectomy. *Cancer Res Treat* (2021) 53:233–42. doi: 10.4143/crt.2020.159
- Beniczky S, Karoly P, Nurse E, Rylvlin P, Cook M. Machine Learning and Wearable Devices of the Future. *Epilepsia* (2021) 62:S116–24. doi: 10.1111/epi.16555
- Dueñas-Gonzalez A, López-Graniel C, González-Enciso A, Cetina L, Rivera L, Mariscal I, et al. A Phase II Study of Multimodality Treatment for Locally Advanced Cervical Cancer: Neoadjuvant Carboplatin and Paclitaxel Followed by Radical Hysterectomy and Adjuvant Cisplatin Chemoradiation. *Ann Oncol* (2003) 14:1278–84. doi: 10.1093/annonc/mdg333
- Gupta S, Maheshwari A, Parab P, Mahantshetty U, Hawaldar R, Sastri Chopra S, et al. Neoadjuvant Chemotherapy Followed by Radical Surgery Versus Concomitant Chemotherapy and Radiotherapy in Patients With Stage IB2, IIA, or IIB Squamous Cervical Cancer: A Randomized Controlled Trial. *J Clin Oncol* (2018) 36:1548–55. doi: 10.1200/jco.2017.75.9985
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1). *Eur J Cancer* (2009) 45:228–47. doi: 10.1016/j.ejca.2008.10.026
- Fan J, Lv J. A Selective Overview of Variable Selection in High Dimensional Feature Space. *Stat Sin* (2010) 20:101–48.
- Rossi S, Basso M, Strippoli A, Schinzari G, D'Argento E, Larocca M, et al. Are Markers of Systemic Inflammation Good Prognostic Indicators in Colorectal Cancer? *Clin Colorectal Cancer* (2017) 16:264–74. doi: 10.1016/j.clcc.2017.03.015
- Jensen TO, Riber-Hansen R, Schmidt H, Hamilton-Dutoit SJ, Steiniche T. Tumor and Inflammation Markers in Melanoma Using Tissue Microarrays: A Validation Study. *Melanoma Res* (2011) 21:509–15. doi: 10.1097/CMR.0b013e32834a3899
- Busch EL, Whitsel EA, Kroenke CH, Yang YC. Social Relationships, Inflammation Markers, and Breast Cancer Incidence in the Women's Health Initiative. *Breast* (2018) 39:63–9. doi: 10.1016/j.breast.2018.03.013

23. Terlikowska KM, Dobrzycka B, Terlikowski R, Sienkiewicz A, Kinalski M, Terlikowski SJ. Clinical Value of Selected Markers of Angiogenesis, Inflammation, Insulin Resistance and Obesity in Type 1 Endometrial Cancer. *BMC Cancer* (2020) 20:921. doi: 10.1186/s12885-020-07415-x
24. Vitkauskaitė A, Urbonienė D, Celiesiute J, Jariene K, Skrodeniene E, Nadisauskiene RJ, et al. Circulating Inflammatory Markers in Cervical Cancer Patients and Healthy Controls. *J Immunotoxicol* (2020) 17:105–9. doi: 10.1080/1547691x.2020.1755397
25. Wang X, Chen J, Sun W, Zhu M, Li D, Chen G. Influences of Neoadjuvant Chemotherapy on Clinical Indicators, Prognosis and Neutrophil/Lymphocyte Ratio of Stage IB2–IIB Cervical Cancer. *J BUON* (2020) 25:757–63.
26. Gan W, Yi Y, Fu Y, Huang J, Lu Z, Jing C, et al. Fibrinogen and C-Reactive Protein Score is a Prognostic Index for Patients With Hepatocellular Carcinoma Undergoing Curative Resection: A Prognostic Nomogram Study. *J Cancer* (2018) 9:148–56. doi: 10.7150/jca.22246
27. Lu J, Chen S, Li X, Qiu G, He S, Wang H, et al. Gastrointestinal Stromal Tumors: Fibrinogen Levels are Associated With Prognosis of Patients as Blood-Based Biomarker. *Medicine* (2018) 97:e0568. doi: 10.1097/md.00000000000010568
28. Tang L, Liu K, Wang J, Wang C, Zhao P, Liu J. High Preoperative Plasma Fibrinogen Levels are Associated With Distant Metastases and Impaired Prognosis After Curative Resection in Patients With Colorectal Cancer. *J Surg Oncol* (2010) 102:428–32. doi: 10.1002/jso.21668
29. Rigatti SJ. Random Forest. *J Insur Med* (2017) 47:31–9. doi: 10.17849/insm-47-01-31-39.1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Huang, Zhu, Xue, Zhu, Chen and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.