REVIEW

3321 Encouraging specific biomarkers-based therapeutic strategies for hepatocellular carcinoma

ORIGINAL ARTICLE

Clinical and Translational Research

3334 Autophagy-related long non-coding RNA prognostic model predicts prognosis and survival of melanoma patients

3352 Identification of circ_0000375 and circ_0011536 as novel diagnostic biomarkers of colorectal cancer
Yin TF, Du SY, Zhao DY, Sun XZ, Zhou YC, Wang QQ, Zhou GYJ, Yao SK

Retrospective Study

3369 Echocardiography in the diagnosis of Shone’s complex and analysis of the causes for missed diagnosis and misdiagnosis
Li YD, Meng H, Pang KJ, Li MZ, Xu N, Wang H, Li SJ, Yan J

3379 Predictors and prognostic impact of post-operative atrial fibrillation in patients with hip fracture surgery
Bae SJ, Kwon CH, Kim TY, Chang H, Kim BS, Kim SH, Kim HJ

3389 Added value of systemic inflammation markers for monitoring response to neoadjuvant chemotherapy in breast cancer patients
Ke ZR, Chen W, Li MX, Wu S, Jin LT, Wang TJ

3401 Washed microbiota transplantation reduces serum uric acid levels in patients with hyperuricaemia
Cai JR, Chen XW, He YJ, Wu B, Zhang M, Wu LH

Clinical Trials Study

3414 Concurrent chemoradiotherapy using gemcitabine and nedaplatin in recurrent or locally advanced head and neck squamous cell carcinoma

META-ANALYSIS

3426 Effect of enhanced recovery after surgery on inflammatory bowel disease surgery: A meta-analysis
Peng D, Cheng YX, Tao W, Tang H, Ji GY

3436 Accuracy of ultrasound elastography for predicting breast cancer response to neoadjuvant chemotherapy: A systematic review and meta-analysis
Chen W, Fang LX, Chen HL, Zheng JH
## Contents

<table>
<thead>
<tr>
<th>Page</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>3449</td>
<td>Association of chronic obstructive pulmonary disease with mild cognitive impairment and dementia risk: A systematic review and meta-analysis</td>
<td>Zhao LY, Zhou XL</td>
</tr>
<tr>
<td>3472</td>
<td>Difference and similarity between type A interrupted aortic arch and aortic coarctation in adults: Two case reports</td>
<td>Ren SX, Zhang Q, Li PP, Wang XD</td>
</tr>
<tr>
<td>3478</td>
<td>Combination therapy (toripalimab and lenvatinib)-associated toxic epidermal necrolysis in a patient with metastatic liver cancer: A case report</td>
<td>Huang KK, Han SS, He L, Yang LL, Liang BY, Zhen QY, Zhu ZB, Zhang CY, Li HY, Lin Y</td>
</tr>
<tr>
<td>3485</td>
<td>Unusual glomus tumor of the lower leg: A case report</td>
<td>Wang HY, Duan P, Chen H, Pan ZY</td>
</tr>
<tr>
<td>3490</td>
<td>Pulmonary <em>Cladosporium</em> infection coexisting with subcutaneous <em>Corynespora cassiicola</em> infection in a patient: A case report</td>
<td>Wang WY, Luo HB, Hu JQ, Hong HH</td>
</tr>
<tr>
<td>3496</td>
<td>Preoperational diagnosis and management of breast ductal carcinoma <em>in situ</em> arising within fibroadenoma: Two case reports</td>
<td>Wu J, Sun KW, Mo QP, Yang ZR, Chen Y, Zhong MC</td>
</tr>
<tr>
<td>3505</td>
<td>Reconstruction of complex chest wall defects: A case report</td>
<td>Huang SC, Chen CY, Qiu P, Yan ZM, Chen WZ, Liang ZZ, Luo KW, Li JW, Zhang YQ, Huang BY</td>
</tr>
<tr>
<td>3511</td>
<td>Young children with multidrug-resistant epilepsy and vagus nerve stimulation responding to perampanel: A case report</td>
<td>Yang H, Yu D</td>
</tr>
<tr>
<td>3541</td>
<td>Severe gastric insufflation and consequent atelectasis caused by gas leakage using AIR-Q laryngeal mask airway: A case report</td>
<td>Zhao Y, Li P, Li DW, Zhao GF, Li XY</td>
</tr>
<tr>
<td>Page</td>
<td>Title</td>
<td>Authors</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>3547</td>
<td>Hypereosinophilic syndrome presenting as acute ischemic stroke, myocardial infarction, and arterial involvement: A case report</td>
<td>Sun RR, Chen TZ, Meng M</td>
</tr>
<tr>
<td>3553</td>
<td>Cytochrome P450 family 17 subfamily A member 1 mutation causes severe pseudohermaphroditism: A case report</td>
<td>Gong Y, Qin F, Li WJ, Li LY, He P, Zhou XJ</td>
</tr>
<tr>
<td>3573</td>
<td>Qingchang decoction retention enema may induce clinical and mucosal remission in left-sided ulcerative colitis: A case report</td>
<td>Li PH, Tang Y, Wen HZ</td>
</tr>
<tr>
<td>3587</td>
<td>Ultrasound-guided local ethanol injection for fertility-preserving cervical pregnancy accompanied by fetal heartbeat: Two case reports</td>
<td>Kakinuma T, Kakinuma K, Matsuda Y, Ohwada M, Yanagida K, Kaijima H</td>
</tr>
<tr>
<td>3593</td>
<td>Successful apatinib treatment for advanced clear cell renal carcinoma as a first-line palliative treatment: A case report</td>
<td>Wei HP, Mao J, Hu ZL</td>
</tr>
<tr>
<td>3601</td>
<td>Del(5q) and inv(3) in myelodysplastic syndrome: A rare case report</td>
<td>Liang HP, Luo XC, Zhang YL, Liu B</td>
</tr>
<tr>
<td>3609</td>
<td>Papillary thyroid microcarcinoma with contralateral lymphatic skip metastasis and breast cancer: A case report</td>
<td>Ding M, Kong YH, Gu JH, Xie RL, Fei J</td>
</tr>
<tr>
<td>3615</td>
<td>Contrast-enhanced ultrasound manifestations of synchronous combined hepatocellular-cholangiocarcinoma and hepatocellular carcinoma: A case report</td>
<td>Gao L, Huang JY, Lu ZJ, Lu Q</td>
</tr>
<tr>
<td>3624</td>
<td>Thyrotoxicosis after a massive levothyroxine ingestion: A case report</td>
<td>Du F, Liu SW, Yang H, Duan RX, Ren WX</td>
</tr>
<tr>
<td>3630</td>
<td>Pleomorphic adenoma of the left lacrimal gland recurred and transformed into myoepithelial carcinoma after multiple operations: A case report</td>
<td>Huang WP, Li LM, Gao JB</td>
</tr>
</tbody>
</table>
ABOUT COVER

Editorial Board Member of *World Journal of Clinical Cases*, Chi-Yuan Yeh, MD, PhD, Assistant Professor, Chief Doctor, radiation oncology, Tungs’ Taichung MetroHarbor Hospital, Taichung 43503, Taiwan. peteryeh46@gmail.com

AIMS AND SCOPE

The primary aim of *World Journal of Clinical Cases* (WJCC, *World J Clin Cases*) is to provide scholars and readers from various fields of clinical medicine with a platform to publish high-quality clinical research articles and communicate their research findings online.

WJCC mainly publishes articles reporting research results and findings obtained in the field of clinical medicine and covering a wide range of topics, including case control studies, retrospective cohort studies, retrospective studies, clinical trials studies, observational studies, prospective studies, randomized controlled trials, randomized clinical trials, systematic reviews, meta-analysis, and case reports.

INDEXING/ABSTRACTING

The WJCC is now indexed in Science Citation Index Expanded (also known as SciSearch®), Journal Citation Reports/Science Edition, Scopus, PubMed, and PubMed Central. The 2021 Edition of Journal Citation Reports® cites the 2020 impact factor (IF) for WJCC as 1.337; IF without journal self cites: 1.301; 5-year IF: 1.742; Journal Citation Indicator: 0.33; Ranking: 119 among 169 journals in medicine, general and internal; and Quartile category: Q3. The WJCC’s CiteScore for 2020 is 0.8 and Scopus CiteScore rank 2020: General Medicine is 493/793.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: Hua-Ge Yu; Production Department Director: Xiang Li; Editorial Office Director: Jin-Lei Wang.

NAME OF JOURNAL

*World Journal of Clinical Cases*

ISSN

ISSN 2307-8960 (online)

LAUNCH DATE

April 16, 2013

FREQUENCY

Thrice Monthly

EDITORS-IN-CHIEF

Bao-Gan Peng, Jerzy Tadeusz Chudek, George Kontogeorgos, Maurizio Serati, Ja Hyeon Ku

EDITORIAL BOARD MEMBERS

https://www.wjgnet.com/2307-8960/editorialboard.htm

PUBLICATION DATE

April 16, 2022

COPYRIGHT

© 2022 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS

https://www.wjgnet.com/bpg/gerinfo/204

GUIDELINES FOR ETHICS DOCUMENTS

https://www.wjgnet.com/bpg/GerInfo/287

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH

https://www.wjgnet.com/bpg/gerinfo/240

PUBLICATION ETHICS

https://www.wjgnet.com/bpg/GerInfo/288

PUBLICATION MISCONDUCT

https://www.wjgnet.com/bpg/gerinfo/208

ARTICLE PROCESSING CHARGE

https://www.wjgnet.com/bpg/gerinfo/242

STEPS FOR SUBMITTING MANUSCRIPTS

https://www.wjgnet.com/bpg/gerinfo/239

ONLINE SUBMISSION

https://www.ffipublishing.com

© 2022 Baishideng Publishing Group Inc. All rights reserved. 7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA

E-mail: bpgoffice@wjgnet.com https://www.wjgnet.com
Retrospective Study

Added value of systemic inflammation markers for monitoring response to neoadjuvant chemotherapy in breast cancer patients

Zi-Rui Ke, Wei Chen, Man-Xiu Li, Shun Wu, Li-Ting Jin, Tie-Jun Wang

Abstract

BACKGROUND
Complete response after neoadjuvant chemotherapy (rNACT) elevates the surgical outcomes of patients with breast cancer, however, non-rNACT have a higher risk of death and recurrence.

AIM
To establish novel machine learning (ML)-based predictive models for predicting probability of rNACT in breast cancer patients who intends to receive NACT.

METHODS
A retrospective analysis of 487 breast cancer patients who underwent mastectomy or breast-conserving surgery and axillary lymph node dissection following neoadjuvant chemotherapy at the Hubei Cancer Hospital between January 1, 2013, and October 1, 2021. The study cohort was divided into internal training and testing datasets in a 70:30 ratio for further analysis. A total of twenty-four variables were included to develop predictive models for rNACT by multiple ML-based algorithms. A feature selection approach was used to identify optimal predictive factors. These models were evaluated by the receiver operating characteristic (ROC) curve for predictive performance.

RESULTS
Analysis identified several significant differences between the rNACT and non-rNACT groups, including total cholesterol, low-density lipoprotein, neutrophil-to-lymphocyte ratio, body mass index, platelet count, albumin-to-globulin ratio, platelet-to-lymphocyte ratio, and lymphocyte-to-monocyte ratio. The areas under the curve of the six models ranged from 0.81 to 0.96. Some ML-based models performed better than models using conventional statistical methods in both ROC
curves. The support vector machine (SVM) model with twelve variables introduced was identified as the best predictive model.

**CONCLUSION**

By incorporating pretreatment serum lipids and serum inflammation markers, it is feasible to develop ML-based models for the preoperative prediction of rNACT and therefore facilitate the choice of treatment, particularly the SVM, which can improve the prediction of rNACT in patients with breast cancer.

**Key Words:** Breast cancer; Neoadjuvant chemotherapy; Clinical response; Machine learning; Prediction

©The Author(s) 2022. Published by Baishideng Publishing Group Inc. All rights reserved.

---

**Core Tip:** For predicting response after neoadjuvant chemotherapy (rNACT), some machine learning-based models performed better than models using conventional methods, and the support vector machine model performed best. Preoperative serum lipids and serum inflammation markers have contributed to predicting rNACT in breast cancer patients. These results suggested the need to raise awareness of the importance of minimally-invasive approaches for monitoring breast cancer patients who intended to undergo neoadjuvant chemotherapy. However, the current study needs to be validated with caution and require external validation in the future.

**INTRODUCTION**

Worldwide, breast cancer is a major cause of human suffering and high mortality among women[1]. Neoadjuvant chemotherapy (NACT) as a treatment for early breast cancer, can make breast conserving surgery more feasible, and may achieve more than the same chemotherapy after surgery to eradicate micrometastasis[2]. More than 65% of the patients treated with NACT have a response, and more than 15% have achieved a complete clinical response. Although some trials use the old chemotherapy regimen, more than 15% of the patients have undergone partial chemotherapy[2-4]. In other words, most patients who cannot achieve a complete pathological response after NACT may face a higher risk of death and recurrence. Therefore, it is necessary to develop a practical, convenient and efficient tool to predict the pathological response of patients with NACT breast cancer.

Machine learning (ML)-based integrated analysis is a new computer-based method, which has been widely used in medical data management in the past decade[5]. It appears at the intersection of statistics and computer science. The former attempts to learn relationships from data, while the latter emphasizes efficient computational algorithms[6,7]. Compared with traditional statistical prediction models such as logistic regression (LR), ML depends on a predetermined model. It can potentially find the interaction between variables and iteratively learn the update algorithm from the data[8,9]. Previously, several conventional predictive models have been made for predicting after NACT in breast cancer patients, including LR, GLM[10-14]. However, few reports have incorporated multiple ML-based ensemble analyses for predicting response after NACT (rNACT).

In this study, we aimed to develop a rNACT risk prediction model for breast cancer patients that utilizes pretreatment serum lipids and serum inflammation markers to stratify patients by rNACT risk on admission. We then analyzed the predictive performance of these ML-based models in a deviation cohort and then verified performance in an internal and external validation cohort.

**MATERIALS AND METHODS**

**Patients**

Between January 1, 2013 and October 1, 2021, we retrospectively collated data from consecutive patients who had been diagnosed with breast cancer at the Hubei Cancer Hospital. All patients had received NACT before surgery. This study was approved by the Institutional Ethics Committee of the Hubei
Cancer Hospital (Reference: LLHBCH2021YN-021), in compliance with the Declaration of Helsinki. Written informed consent was obtained from all participants before any treatment. We confirmed that the data from all the patients were anonymized in this study. The inclusion and exclusion criteria were summarized in Figure 1. The study cohort was divided into internal training and testing datasets in a 70:30 ratio for further analysis.

**Blood data collection**
The blood samples of all patients were taken from the fasting state before chemotherapy, and the blood tests were operated by professional personnel to ensure that the blood test results were not biased. The results of the blood test are as follows: Blood routine, liver and kidney function, electrolytes, and blood lipids.

*Figure 1 The flow chart of patient selection and data process. NACT: Neoadjuvant chemotherapy; ROC: Receiver operating characteristic.*
Figure 2 Statistical analysis of features included in machine learning based models. A: Heatmap representing the correlation between candidate variables included in predictive models using Spearman’s correlation coefficient; B: Scaled importance rank of all features included in predictive models for identifying risk of response after neoadjuvant chemotherapy (nNACT) in breast cancer patients; C-J: Box and jitter plots showing distribution of continuous features included in predictive models between nNACT and non-nNACT groups. BMI: Body mass index; PR: Partial response; NLR: Neutrophil-to-lymphocyte ratio; LMR: Lymphocyte-to-monocyte ratio; PLR: Platelet-to-lymphocyte ratio; A/G: Albumin-to-globulin ratio; LDL: Low-density lipoprotein; HDL: High-density lipoprotein; TC: Total cholesterol; TG: Triglyceride.

Evaluating the safety and efficacy of NACT

According to RECIST (version 1.1) criteria, the efficacy of NACT is defined as follows: (1) Cardiol Res. The tumor is disappeared completely; (2) Partial response (PR). The diameter of the tumor is reduced (≥30%); (3) Progressive disease (PD). The diameter of the tumor was reduced (≥20%); and (4) Stable disease (SD). The diameter of the tumor was altered between PR and PD. Collectively, patients were considered to be responsive to NACT provided that they were evaluated as CR or PR after NACT treatment. On the contrary, patients with SD or PD were regarded as non-responsive to NACT.
Validation and comparison of the predictive model. A: Area under the receiver operating characteristic curve (AUC) to assess the performance of response after neoadjuvant chemotherapy (rNACT) risk prediction of machine learning based models; B: AUC to assess the performance of rNACT risk prediction of generalized linear model (GLM); C: Discriminative evaluation of support vector machine in predicting rNACT; D: Discriminative evaluation of GLM in predicting rNACT. SVM: Support vector machine; GLM: Generalized linear model.

Development and validation of ML-based models

Four ML-based algorithms were performed to build predictive models, we used the caret package to randomly divide the data set into two parts, 70% for model training and 30% for model testing. A total of 6 ML-based algorithms were executed to establish the predictive model. According to the principle of "two-step estimation"[16], We obtained the prediction model through variable screening and algorithm, as follows: M is the intersection of M3 and M4. The characteristic variable is marked as X and the target variable is marked as Y. The X and Y were evenly divided into two parts, namely X1, Y1, and X2, Y2. Through univariate screening, the variable quantum set M1was screened on X1 and Y1, and M2was filtered by X2 and Y2. Then, a lasso was used to fit the model again, and the filtered variables were marked asM3 and M4. Briefly, by sorting the intersection of variable sets, the optimal subset modeling is obtained. The model was evaluated by inspection, discrimination, and calibration. The receiver operating characteristic (ROC) curve was used to evaluate the recognition ability of the prediction model in the training data set and the test data set. The discrimination ability of each model was quantified by the area under the ROC curve (AUC), decision curve analysis, and clinical impact curve (CIC).

Statistical analysis

Continuous variables are expressed as mean ± SD and compared using the two-tailed t-test or the Mann-Whitney test. Categorical variables were compared using the chi-square test or Fisher's exact test. Univariate and multivariate logistic analyses were used to explore the risk factors for rNACT. Several ML-based algorithms were applied to predict rNACT, including support vector machine (SVM), random forest (RF), Naive Bayes (NB), neural network (NN), decision tree (DT), and generalized linear...
Monitoring response to neoadjuvant chemotherapy

RESULTS

Clinicopathological characteristics
During the period of enrollment, 287 consecutive patients with breast cancer underwent mastectomy or breast-conserving surgery and axillary lymph node dissection following NACT. Besides, 201 patients were validated as external data sets for the prediction model. Demographics and baseline data were summarized in Table 1. According to the RECIST (version 1.1) criteria, rNACT was identified in 255 (88.9%) and 32 (11.1%) patients with non-rNACT in the internal whole cohort. In the external cohort, 176 (88.0%) patients were confirmed to have rNACT, and 24 (12.0%) patients represented non-rNACT. Overall, most patients with breast cancer presenting with rNACT were positively associated with pretreatment serum lipids and serum inflammation markers. No statistically significant difference was detected between two cohorts with regard to age, menopause, grade, smoking, estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) ($P \geq 0.05$).

Variable importance and candidate features selection
By feature selection, the twenty-four variables for each algorithm were screened by their predictive importance. As depicted in Figure 2A, only twelve of the candidate features were eventually chosen for modeling, among which eight features had a positive association with rNACT, including PLT, monocyte count (MONO), neutrophil-to-lymphocyte ratio (NLR), lymphocyte-to-monocyte ratio (LMR), platelet-to-lymphocyte ratio (PLR), low-density lipoprotein (LDL), A/G, and total cholesterol (TC). Four features were negatively correlated with rNACT, including high-density lipoprotein (HDL), triglyceride (TG), BMI, and age. The weight of the top eight variables was shown in Figure 2B. The pretreatment serum lipids and serum inflammation markers also showed significant differences between rNACT and non-rNACT groups (Figure 2C-J). Multivariable logistic analysis using raw data of the candidate features proved that the features selected by stepwise analysis exhibited similar risk implications (Supplementary Table 1). Based on our results, NLR (OR: 1.02, 95% CI: 0.78-1.26), LMR (OR: 1.44, 95% CI: 1.32-1.56), PLR (OR: 2.54, 95% CI: 1.81-6.94), PLT (OR: 1.13), BMI (OR: 1.23, 95% CI: 1.09-1.38), A/G (OR: 1.69, 95% CI: 1.24-2.14), TC (OR: 0.71, 95% CI: 0.26-1.16), and TG (OR: 0.42, 95% CI: 0.17-0.68) were positively correlated with rNACT.

Comparison Between ML-Based Models
A total of twelve preoperative variables were used to develop predictive models for rNACT based on six algorithms. The predictive performance of all models was shown in Figure 3A and B and Table 2. The best performance was observed in the SVM model (AUC = 0.96, 95% CI: 0.91-1.01), which performed similarly to RF model (AUC = 0.94, 95% CI: 0.87-1.01), superior than NB model (AUC = 0.86, 95% CI: 0.79-0.93), NN model (AUC = 0.88, 95% CI: 0.82-0.94), DT model (AUC = 0.83, 95% CI: 0.77-0.89), and GLM (AUC = 0.81, 95% CI: 0.71-0.91). All ML-based models were better than conventional model. Furthermore, the optimal model SVM showed superior to the traditional linear model in discrimination (Figure 3C and D).

Internal and external validation of the optimal predictive model
To further validate the performance of the SVM model, we also adopted CIC to evaluate the prediction efficiency, as illustrated in Figure 4A, the CIC demonstrated that the stratification of rNACT could be distinguished in the training cohorts. These results were also parallel to risk factors of rNACT delineated in the validation cohorts (Figure 4B), indicating that the selected features were highly relevant to rNACT.

DISCUSSION
Reliable markers of chemosensitivity help select patients who most benefit from NACT[19]. Previous studies on the candidate predictors of NACT efficacy in breast cancer patients are discordant, suggesting that the potential predictors to predict efficacy is insufficient[20-23]. In addition, whilst many studies report the predictive outcomes of breast cancer patients who have received NACT, however,
## Table 1 Demographics and baseline characteristics of the breast cancer patients undergoing neoadjuvant chemotherapy

<table>
<thead>
<tr>
<th>Variables</th>
<th>Dummy variables</th>
<th>Training cohort</th>
<th>Testing cohort</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall (n = 287)</td>
<td>Yes (n = 255)</td>
<td>No (n = 32)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall (n = 200)</td>
<td>Yes (n = 176)</td>
<td>No (n = 24)</td>
<td></td>
</tr>
<tr>
<td>Age [median (IQR)], yr</td>
<td>61.00 (55.00, 70.00)</td>
<td>61.00 (54.00, 70.00)</td>
<td>64.00 (58.75, 69.25)</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>61.00 (55.75, 70.25)</td>
<td>61.00 (55.00, 70.25)</td>
<td>64.00 (57.75, 70.25)</td>
<td>0.28</td>
</tr>
<tr>
<td>Menopausal (%)</td>
<td>Yes</td>
<td>35 (12.2)</td>
<td>30 (11.8)</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>252 (87.8)</td>
<td>225 (88.2)</td>
<td>27 (84.4)</td>
</tr>
<tr>
<td>T stage (%)</td>
<td>T1-2</td>
<td>216 (75.3)</td>
<td>196 (76.9)</td>
<td>20 (62.5)</td>
</tr>
<tr>
<td></td>
<td>T3-4</td>
<td>71 (24.7)</td>
<td>59 (23.1)</td>
<td>12 (37.5)</td>
</tr>
<tr>
<td>N stage (%)</td>
<td>N0-1</td>
<td>71 (24.7)</td>
<td>58 (22.7)</td>
<td>13 (40.6)</td>
</tr>
<tr>
<td></td>
<td>N2-3</td>
<td>216 (75.3)</td>
<td>197 (77.3)</td>
<td>19 (59.4)</td>
</tr>
<tr>
<td>Grade (%)</td>
<td>I-II</td>
<td>195 (67.9)</td>
<td>175 (68.6)</td>
<td>20 (62.5)</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>92 (32.1)</td>
<td>80 (31.4)</td>
<td>12 (37.5)</td>
</tr>
<tr>
<td>Histology (%)</td>
<td>IDC</td>
<td>111 (38.7)</td>
<td>106 (41.6)</td>
<td>5 (15.6)</td>
</tr>
<tr>
<td></td>
<td>ILC</td>
<td>91 (31.7)</td>
<td>79 (31.0)</td>
<td>12 (37.5)</td>
</tr>
<tr>
<td></td>
<td>IMC</td>
<td>48 (16.7)</td>
<td>41 (16.1)</td>
<td>7 (21.9)</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>37 (12.9)</td>
<td>29 (11.4)</td>
<td>8 (25.0)</td>
</tr>
<tr>
<td>Molecular subtyping (%)</td>
<td>HER2-LuB</td>
<td>108 (37.6)</td>
<td>100 (39.2)</td>
<td>8 (25.0)</td>
</tr>
<tr>
<td></td>
<td>HER2+</td>
<td>71 (24.7)</td>
<td>66 (25.9)</td>
<td>5 (15.6)</td>
</tr>
<tr>
<td></td>
<td>HER2+LuB</td>
<td>29 (10.1)</td>
<td>24 (9.4)</td>
<td>5 (15.6)</td>
</tr>
<tr>
<td></td>
<td>LuA</td>
<td>54 (18.8)</td>
<td>41 (16.1)</td>
<td>13 (40.6)</td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>25 (8.7)</td>
<td>24 (9.4)</td>
<td>1 (3.1)</td>
</tr>
<tr>
<td>BMI (%)</td>
<td>≤ 18</td>
<td>21 (7.3)</td>
<td>21 (8.2)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>≥ 27</td>
<td>36 (12.5)</td>
<td>31 (12.2)</td>
<td>5 (15.6)</td>
</tr>
<tr>
<td>Smoking (%)</td>
<td>No</td>
<td>258 (89.9)</td>
<td>231 (90.6)</td>
<td>27 (84.4)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>29 (10.1)</td>
<td>24 (9.4)</td>
<td>5 (15.6)</td>
</tr>
<tr>
<td>ER (%)</td>
<td>Negative</td>
<td>103 (35.9)</td>
<td>91 (35.7)</td>
<td>12 (37.5)</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>184 (64.1)</td>
<td>164 (64.3)</td>
<td>20 (62.5)</td>
</tr>
<tr>
<td>PR (%)</td>
<td>Negative</td>
<td>165 (57.5)</td>
<td>141 (53.3)</td>
<td>24 (75.0)</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>122 (42.5)</td>
<td>114 (44.7)</td>
<td>8 (25.0)</td>
</tr>
<tr>
<td>HER2 (%)</td>
<td>Negative</td>
<td>157 (54.7)</td>
<td>140 (54.9)</td>
<td>17 (53.1)</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>130 (45.3)</td>
<td>115 (45.1)</td>
<td>15 (46.9)</td>
</tr>
<tr>
<td>PLT [median (IQR)] × 10^3/L</td>
<td>200.00 (154.50, 268.50)</td>
<td>187.00 (152.50, 245.50)</td>
<td>388.00 (334.50, 454.00)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Neutrophil [median (IQR)] × 10^3/L</td>
<td>4.35 (3.52, 5.13)</td>
<td>4.54 (3.76, 5.25)</td>
<td>3.09 (2.02, 3.39)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>MONO [median (IQR)] × 10^3/L</td>
<td>0.40 (0.25, 0.57)</td>
<td>0.38 (0.24, 0.50)</td>
<td>0.88 (0.80, 0.94)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>Lymphocyte [median (IQR)] × 10^3/L</td>
<td>3.09 (2.92, 3.30)</td>
<td>3.13 (2.96, 3.32)</td>
<td>1.64 (1.14, 2.16)</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>NLR [median (IQR)]</td>
<td>1.43 (1.20, 1.67)</td>
<td>1.42 (1.20, 1.64)</td>
<td>1.56 (1.21, 2.68)</td>
<td>0.03</td>
</tr>
<tr>
<td>LMR [median (IQR)]</td>
<td>7.62 (5.59, 8.51)</td>
<td>8.51 (6.18, 8.66)</td>
<td>1.86 (1.15, 1.51)</td>
<td>&lt; 0.01</td>
</tr>
</tbody>
</table>
Table 2 Performance for response to neoadjuvant chemotherapy risk prediction of models in breast cancer patients

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC (95%CI)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPV</th>
<th>NPV</th>
<th>Kappa</th>
<th>Brier</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.96 (0.91-1.01)</td>
<td>96.58</td>
<td>45.28</td>
<td>88.63</td>
<td>75.00</td>
<td>0.68</td>
<td>0.06</td>
</tr>
<tr>
<td>RF</td>
<td>0.94 (0.87-1.01)</td>
<td>94.44</td>
<td>68.75</td>
<td>86.67</td>
<td>68.75</td>
<td>0.65</td>
<td>0.07</td>
</tr>
<tr>
<td>NB</td>
<td>0.86 (0.79-0.93)</td>
<td>96.36</td>
<td>35.80</td>
<td>83.14</td>
<td>75.00</td>
<td>0.62</td>
<td>0.07</td>
</tr>
<tr>
<td>NN</td>
<td>0.88 (0.82-0.94)</td>
<td>93.15</td>
<td>25.00</td>
<td>93.15</td>
<td>53.15</td>
<td>0.62</td>
<td>0.07</td>
</tr>
<tr>
<td>DT</td>
<td>0.83 (0.77-0.89)</td>
<td>94.50</td>
<td>24.14</td>
<td>74.12</td>
<td>65.63</td>
<td>0.59</td>
<td>0.07</td>
</tr>
<tr>
<td>GLM</td>
<td>0.81 (0.71-0.91)</td>
<td>95.70</td>
<td>23.76</td>
<td>69.80</td>
<td>75.00</td>
<td>0.57</td>
<td>0.08</td>
</tr>
</tbody>
</table>

AUC: Area under the curve, area under the receiver operating characteristics curve; PPV: Positive predictive value; NPV: Negative predictive value, 95%CI: 95% confidence interval; SVM: Support vector machine; RF: Random forest; NB: Naive bayes ; NN: Neural network; DT: Decision tree; GLM: Generalized linear model.

relatively few have investigated the individual contribution of multiple models to accuracy, especially prediction efficiency[24-27]. Whilst this study indicates that ML-based predictive algorithms should be included in NACT risk assessments in breast cancer patients, it also highlights the importance of conducting newly predictive models for clinical management.

Supervised ML algorithms have been a dominant method in the data mining field[17]. In recent years, ML-based algorithms were widely used for the evaluation of disease prognosis[28-30]. In this study, extensive variables were made to identify those predictive that applied more than one supervised ML algorithm on rNACT prediction. Based on the ML algorithm, we employed a variety of statistical, probabilistic, and optimization methods to learn from experience and detect useful patterns from large, unstructured, and complex datasets. To sum up, we extracted the data from the patient’s medical records as much as possible. With the help of different algorithms, such as automated text categorisation [31], network intrusion detection[32], optimizing manufacturing process[33], etc., we finally obtained meaningful candidate variables. Given the growing applicability and effectiveness of supervised ML algorithms on predictive disease modeling. Interestingly, we found that the SVM algorithm is applied most robust in predicting rNACT, which denotes superior performance than the conventional linear prediction model. Besides, the remaining machine prediction models are better than GLM. Therefore, our research demonstrated that, compared with the traditional model, machine learning modeling prediction rNACT could obtain better prediction performance.
Inflammation is associated with the development and malignant progression of most cancers[34]. Inflammatory blood markers have emerged as potential prognostic factors in various cancers, such as NLR, LMR, and PLR. Activated inflammatory cells are sources of reactive oxygen species and reactive nitrogen intermediates that can promote cancer initiation[35]. In breast cancer, pretreatment NLR values are associated with patient prognosis[36]. Similarly, our study indicated that NLR, LMR, and PLR values can be reliably used to predict breast patient responses to NACT treatment, which can effectively stratify patients based upon their likelihood of achieving rNACT. Besides, we also found that pretreatment abnormal A/G ratio, which might be attributable to rNACT. Indeed, a low pretreatment A/G ratio is associated with poor prognosis in human cancers[37]. The importance of lipids in tumor progression, invasion, and metastasis has been described in the previous studies[38]. High triglycerides and low levels of HDL are observed to promote tumor growth[39]. In the present study, we observed that LDL, TC, and BMI were highly associated with rNACT, consistent with previous studies[37,38]. Collectively, clinicians can more effectively weigh the relative costs and benefits of pretreatment serum lipids and serum inflammation markers to ensure that they act in the optimal choice of breast cancer patients.

There are multiple strengths to this study. First, our observations were limited to retrospective studies from a single-center, these findings need further multi-institutional validation with larger sample size. Second, our nomograms were merely validated via an internal training set, external verification using independent patient set is necessary. Third, this is a retrospective study that could not completely avoid missing data and measurement biases, more candidate useful biomarkers may be needed to develop predictive models in the future.

**CONCLUSION**

In summary, for predicting rNACT, some ML-based models performed better than models using conventional methods, and the SVM model performed best. Preoperative serum lipids and serum inflammation markers have contributed to predicting rNACT in breast cancer patients. These results suggested the need to raise awareness of the importance of minimally-invasive approaches for monitoring breast cancer patients who intended to undergo NACT. However, the current study needs to be validated with caution and require external validation in the future.

**ARTICLE HIGHLIGHTS**

**Research background**

Complete response after neoadjuvant chemotherapy (rNACT) elevates the surgical outcomes of patients with breast cancer, however, non-rNACT have a higher risk of death and recurrence.

**Research motivation**

In this study, we aimed to develop an rNACT risk prediction model for breast cancer patients that
utilizes pretreatment serum lipids and serum inflammation markers to stratify patients by rNACT risk on admission. We then analyzed the predictive performance of these ML-based models in a deviation cohort and then verified performance in an internal and external validation cohort.

**Research objectives**
In this study, we aimed to develop an rNACT risk prediction model for breast cancer patients that utilizes pretreatment serum lipids and serum inflammation markers to stratify patients by rNACT risk on admission. We then analyzed the predictive performance of these ML-based models in a deviation cohort and then verified performance in an internal and external validation cohort.

**Research methods**
A retrospective analysis of 487 breast cancer patients who underwent mastectomy or breast-conserving surgery and axillary lymph node dissection following NACT at the Hubei Cancer Hospital between January 1, 2013, and October 1, 2021. The study cohort was divided into internal training and testing datasets in a 70:30 ratio for further analysis. A total of twenty-four variables were included to develop predictive models for rNACT by multiple ML-based algorithms. A feature selection approach was used to identify optimal predictive factors. These models were evaluated by the receiver operating characteristic (ROC) curve for predictive performance.

**Research results**
Analysis identified several significant differences between the rNACT and non-rNACT groups, including total cholesterol, low-density lipoprotein, neutrophil-to-lymphocyte ratio, body mass index, platelet count, albumin-to-globulin ratio (A/G), platelet-to-lymphocyte ratio, and lymphocyte-to-monocyte ratio. The areas under the curve of the six models ranged from 0.81 to 0.96. Some ML-based models performed better than models using conventional statistical methods in both ROC curves. The support vector machine (SVM) model with twelve variables introduced was identified as the best predictive model.

**Research conclusions**
By incorporating pretreatment serum lipids and serum inflammation markers, it is feasible to develop ML-based models for the preoperative prediction of rNACT and therefore facilitate the choice of treatment, particularly the SVM, which can improve the prediction of rNACT in patients with breast cancer.

**Research perspectives**
For predicting rNACT, some ML-based models performed better than models using conventional methods, and the SVM model performed best. Preoperative serum lipids and serum inflammation markers have contributed to predicting rNACT in breast cancer patients. These results suggested the need to raise awareness of the importance of minimally-invasive approaches for monitoring breast cancer patients who intended to undergo NACT. However, the current study needs to be validated with caution and require external validation in the future.

**ACKNOWLEDGEMENTS**
The authors gratefully acknowledge all of our participants for sharing their medical records. The authors also wish to thank the staff members at Hubei Cancer Hospital for their assistance with data collection.

**FOOTNOTES**

**Author contributions:** Ke ZR and Chen W originated the idea, data analysis and writing; Li MX, Wu S, Jin LT and Wang TJ contributed to the data analysis and writing; all authors have read and approved the manuscript.

**Institutional review board statement:** This study was approved by the Institutional Ethics Committee of the Hubei Cancer Hospital (Reference: LLHBCH2021YN-021), in compliance with the Declaration of Helsinki.

**Informed consent statement:** Patients were not required to give informed consent to the study because the analysis used anonymous clinical data that were obtained after each patient agreed to treatment by written consent.

**Conflict-of-interest statement:** None of the authors have any conflicts of interest to declare.

**Data sharing statement:** No additional data are available.
REFERENCES


17. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease


38 Hashemi SA, Bathaie SZ, Mohagheghi MA. Crocetin and crocin decreased cholesterol and triglyceride content of both breast cancer tumors and cell lines. Avicenna J Pharmocol 2018; 10: 384-397 [PMID: 32850295]
