

Original Article

Prediction of lymph node metastasis in cervical cancer patients using AdaBoost machine learning model: analysis of risk factors

Nan Li, Erxuan Peng, Fenghua Liu

Department of Gynecology, The Affiliated Cancer Hospital of Zhengzhou University and Henan Cancer Hospital, Zhengzhou 450008, Henan, China

Received November 9, 2024; Accepted February 18, 2025; Epub March 15, 2025; Published March 30, 2025

Abstract: This study focuses on the development and evaluation of machine learning models, particularly the Adaptive Boosting (AdaBoost) algorithm, for predicting lymph node metastasis (LNM) in cervical cancer (CC) patients. The findings show that AdaBoost outperformed traditional statistical methods and other machine learning models, including Random Forest, Support Vector Machine (SVM), and Least Absolute Shrinkage and Selection Operator (LASSO) regression, in predicting LNM. The areas under the curve (AUCs) for the training and validation sets were 0.882 and 0.857, respectively, indicating high prediction efficiency. Multivariate logistic regression identified key independent risk factors for LNM, including FIGO staging, squamous cell carcinoma antigen (SCC-Ag), white blood cell count (WBC), neutrophil count (NEUT), hemoglobin (HGB) level, and prealbumin (PAB) level. These factors are significant in predicting LNM and emphasize their importance in clinical decision-making. AdaBoost's ability to predict LNM preoperatively, without invasive procedures such as lymph node dissection, can reduce treatment risks and improve patient outcomes. While other models, such as XGBoost, showed a marginally higher AUC in training, AdaBoost's performance in validation was comparable ($P=0.18$). Inflammatory and nutritional markers, such as WBC, NEUT, HGB, and PAB, were significant predictors and provide valuable insights into tumor progression. Despite the study's retrospective nature, the integration of larger, multi-center datasets, and multi-modal imaging could further enhance the model's accuracy and generalizability. This high-performance AdaBoost model offers clinical potential for refining personalized treatment strategies for CC patients.

Keywords: Cervical cancer, lymph node metastasis, machine learning, AdaBoost, non-invasive prediction

Introduction

Cervical cancer (CC) ranks among the most prevalent malignant neoplasms among females on a global scale, particularly in developing nations, where its incidence and mortality rates persist at elevated levels [1]. Statistically, in 2022, there were approximately 348,189 new fatal cases globally [2]. In China, CC constitutes a common malignant tumor in the female reproductive system, ranking sixth in terms of incidence and seventh in mortality [3]. The pathogenesis of CC is intricately associated with multiple risk factors, among which the predominant etiological factor is the persistent infection of high-risk human papillomavirus (HPV) [4]. Additionally, socio-economic factors such as early sexual activity, multiple sexual part-

ners, immunosuppression, and smoking are also correlated with the incidence of CC [5].

Lymph node metastasis (LNM) represents one of the crucial indices for prognostic assessment in CC patients and exerts a substantial influence on treatment selection and survival rates [6]. In early-stage CC patients without LNM, the 5-year survival rate can reach as high as 90%, whereas in those with LNM, the 5-year survival rate plummets rapidly to approximately 65% [7]. Consequently, LNM is not only a significant indicator of tumor dissemination but also a pivotal factor influencing clinical staging, therapeutic decision-making, and prognosis [8]. Although postoperative pathological diagnosis serves as the "gold standard" for the confirmation of LNM, this approach relies on postopera-

Predicting cervical cancer lymph node metastasis with AdaBoost

tive evaluation, which means that it can not promptly ascertain the lymph node status of inoperable patients.

Currently, the primary clinical modalities for predicting LNM in CC patients include imaging examinations (magnetic resonance imaging (MRI), computed tomography (CT), positron-emission tomography-computed tomography (PET-CT), etc.) and postoperative pathological assessment [9]. Nevertheless, each of these methods has its inherent limitations. For instance, imaging examinations demonstrate relatively low sensitivity in detecting small-scale or minute lymph node metastases. Besides, they are economically burdensome, thereby precluding their extensive utilization [10]. Although surgical lymph node dissection can yield precise pathological information, it is an invasive procedure associated with certain risks of complications, such as lymphocyst formation, infections, and hemorrhagic events [11]. Moreover, squamous cell carcinoma antigen (SCC-Ag) is frequently employed as an ancillary diagnostic indicator in clinical settings. However, this biomarker exhibits relatively low specificity among non-squamous cell carcinoma patients, thereby circumscribing its application scope [12]. Consequently, the quest for an economical, non-invasive methodology capable of precisely predicting LNM prior to surgical intervention has emerged as a significant research topic in the current scientific landscape.

A growing body of research has indicated that chronic inflammation is closely related to the oncogenesis and progression of tumors. Inflammatory responses play an indispensable and multifaceted role during diverse phases of tumorigenesis, namely initiation, progression, and metastasis [13]. Nutritional status also represents a cardinal prognostic determinant among tumor-afflicted patients. Hypoalbuminemia is usually and robustly linked to malnutrition and cachexia among cancer patients, portending an adverse prognostic outcome [14]. Plasma fibrinogen level, in a similar vein, exhibits a significant correlation with tumor progression and patient survival. Fibrin, beyond its traditional role as a coagulation factor, serves as a potent facilitator in the growth and metastatic dissemination of tumor cells [15]. Consequently, inflammatory indices and nutritional status

markers in the blood might present a novel perspective for LNM prediction.

In recent years, with the advancement of big data and artificial intelligence technologies, machine learning (ML) has progressively manifested substantial application potential in the medical field [16]. In contrast to traditional statistical approaches, ML has the capacity to manage complex data sets and apprehend nonlinear relationships among variables, thereby augmenting the precision of prediction [17]. In the realm of predicting LNM in CC, ML models including eXtreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), and Random Forest have gradually emerged as novel, non-invasive predictive tools, due to their ability to integrate diverse clinical and laboratory variables [18]. These models can efficaciously integrate numerous clinical characteristics and laboratory examination data of patients, thereby enabling a more accurate assessment of the risk of LNM and furnishing clinicians with more reliable decision-making support.

The primary aim of this study is to explore the predictive relationship between key laboratory indices and LNM in CC patients through a comprehensive retrospective analysis of clinicopathological data. While previous studies have investigated individual factors influencing LNM, this study stands out by incorporating multiple ML models, such as AdaBoost and XGBoost, to build a robust, data-driven prediction model. This innovative approach leverages the potential of advanced ML techniques to integrate and analyze a wide range of clinical and laboratory variables, including inflammatory markers, tumor biomarkers, and hematological indices. By assessing the prognostic capacity of these blood-based indices in predicting LNM, our model aims to provide clinicians with a non-invasive, preoperative tool for early and accurate prediction of LNM, which can significantly improve treatment planning and patient outcomes. This research not only contributes to the growing body of knowledge in cancer metastasis prediction but also offers a potential clinical tool that can guide decision-making in the management of CC patients, reducing unnecessary surgical interventions and enhancing personalized care.

Methods and materials

Research subjects

A retrospective analysis was performed on 612 CC patients who were diagnosed and underwent surgical treatment in The Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital during the period from January 2020 to July 2023. This study obtained the approval from the Ethics Committee of The Affiliated Cancer Hospital of Zhengzhou University & Henan Cancer Hospital.

Inclusion and exclusion criteria

Inclusion criteria: Patients diagnosed with CC and subjected to surgical intervention [19]; Patients with the International Federation of Gynecology and Obstetrics (FIGO) staging [20] between IB1 and IIA1; Patients of 18 years of age or older; Patients with comprehensive clinicopathological data and follow-up information.

Exclusion criteria: Patients with concomitant other malignant neoplasms; Patients with clinical data suggesting infectious diseases, hematological disorders, or immune system maladies at the time of initial diagnosis of CC; Patients with severe comorbid conditions who were intolerant to surgical procedures; Patients with other events that could impact the results of routine blood tests, routine coagulation examinations, and liver function assays (use of non-steroidal medications, heparin, or hepatotoxic drugs, etc.).

Data collection

Data source: The clinical data of patients were systematically retrieved from the electronic medical records, including demographic features, laboratory examination results, imaging data, and pathological findings. Clinical indicators: Demographic characteristics involve age and body mass index (BMI). Tumor characteristics include disease types (squamous cell carcinoma, adenocarcinoma, etc.), tumor size (<2 cm, 2-4 cm, or >4 cm), FIGO grade (IB1, IB2, IB3, IIA1, etc.), differentiation degree (well-, moderately-, or poorly-differentiated), and lymphovascular invasion (with/without). Laboratory test results cover squamous cell carcinoma antigen (SCC-Ag), white blood cell count (WBC), neutrophil count (NEUT), lymphocyte count

(LYM), monocyte count (MON), hemoglobin (HGB), platelet count (PLT), alanine aminotransferase (ALT), aspartate aminotransferase (AST), prealbumin (PAB), total protein (TP), triglyceride (TG), high-density lipoprotein (HDL), low-density lipoprotein (LDL), and cystatin C (Cyccs). For the biochemical function assay, a Roche Cobas c702 fully-automatic biochemical analyzer (manufactured by Roche Diagnostics, Germany) was employed. The tumor markers were measured using a Roche Cobas e602 fully-automatic electrochemiluminescence immunoassay analyzer (Roche Diagnostics, Germany). The routine blood examination was carried out using a Siemens ADVIA 2120i fully-automatic blood analyzer (produced by Siemens Healthineers, Germany).

Grouping and comparison

Patients were randomly partitioned into a training set (n=428) and a validation set (n=184) in a 7:3 ratio. The training set served the purpose of model construction, while the validation set was dedicated to model assessment. The openxlsx package in R language was utilized for the grouping procedure, and the equilibrium of baseline characteristics between the two groups was meticulously ensured. During the grouping process, the Shapiro-Wilk normality test, along with the t-test/Wilcoxon rank-sum test, was employed to appraise the discrepancies among numerical variables. For categorical variables, the chi-square test was adopted to guarantee the absence of significant differences between the two groups. The standard for successful grouping was that all *P*-values exceeded 0.05 (**Figure 1**).

Establishment of ML models

Model building: The training and validation sets were utilized to build and evaluate ML models. The mainstream ML models, namely XGBoost, AdaBoost, Random Forest, Support Vector Machine (SVM), and Least Absolute Shrinkage and Selection Operator (LASSO), were selected. For the XGBoost model, main hyper-parameters such as learning rate, maximum depth, and subsample rate are configured. Hyperparameter optimization was achieved through grid-search and cross-validation techniques. In the AdaBoost model, the number of base estimators (*n*_estimators) and the learning rate were used as main hyper-parameters. The base

Predicting cervical cancer lymph node metastasis with AdaBoost

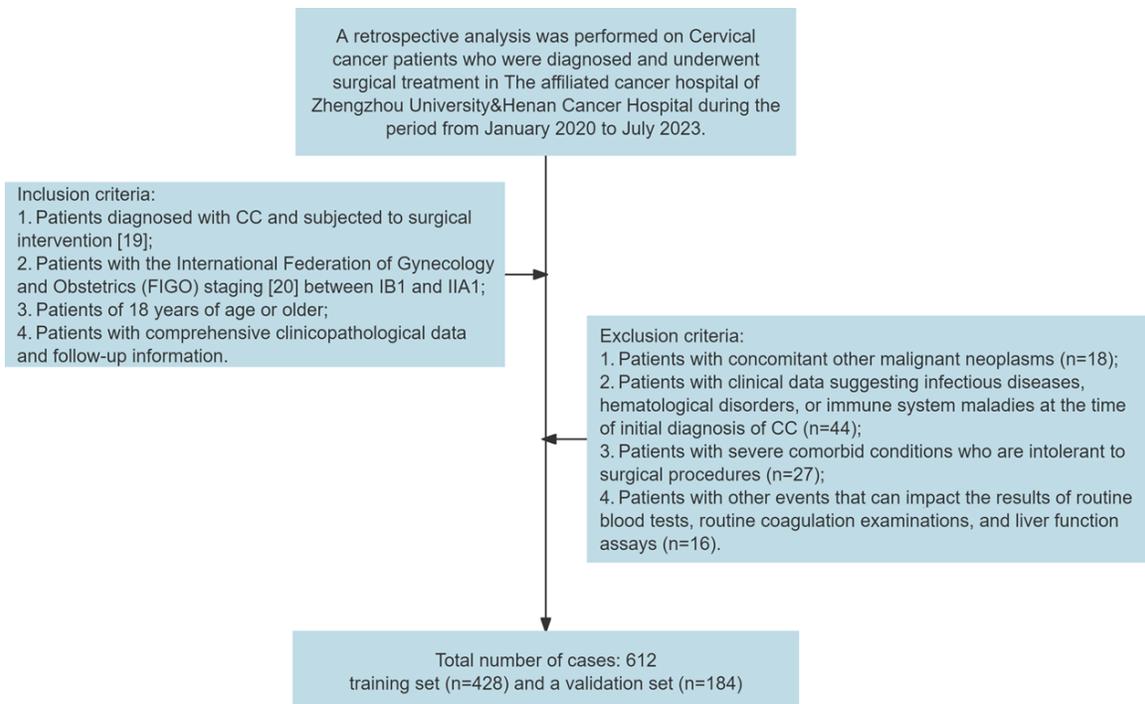


Figure 1. Flow diagram of sample screening.

estimators included 50 to 150 trees with a maximum depth of 1 to 3, and 10-fold cross-validation was used to find the appropriate learning rate and number of iterations. The hyper-parameters of the Random Forest model included the number of trees ($n_{\text{estimators}}$) and the maximum number of features (max_features). The optimal feature set was selected through recursive feature elimination (RFE), and the number of the best trees was selected by minimizing the out-of-bag (OOB) error rate. In the SVM model, the primary parameters subject to optimization were the penalty parameter (C) and the kernel function (kernel). Specifically, the Radial Basis Function (RBF) was employed as the kernel function type. The hyper-parameters were optimized through grid-search methodology, and the model construction was carried out in conjunction with RFE for feature selection. LASSO regression utilized L1 regularization. The regularization parameter (α) was determined through cross-validation, and the lambda path in the cross-validation process was depicted to identify the optimal λ_{min} and $\lambda_{1\text{se}}$ values. Subsequently, the most influential features were sieved via path coefficient analysis and feature importance assessment. All the models were opti-

mized by means of grid-search and cross-validation procedures to achieve the optimal model performance. Moreover, the features exerting the most significant impact on LNM prediction were screened by importance scoring, thereby simplifying the model and enhancing the prediction accuracy.

Outcome measurement

Primary outcomes: Machine-learning-based prediction models for LNM in CC were constructed and validated. Further, multiple metrics including the models' accuracy, sensitivity, specificity, and areas under the curve (AUCs) were comprehensively evaluated. These evaluations provide a quantitative assessment of the model's performance in predicting LNM within the context of CC, enabling a more precise understanding of their potential clinical utility.

Secondary outcomes: The predictive ability of various clinical features and laboratory indicators for LNM was explored, the significant independent risk factors for LNM were screened out, and the most significant variables were determined through feature-importance analysis.

Predicting cervical cancer lymph node metastasis with AdaBoost

Statistical methods

SPSS 26.0 was employed for data processing and statistical analysis. For data visualization, the ggplot2 package of R was used, and Sankey diagrams were constructed using the ggalluvial package. For count data, chi-square tests were applied to compare the discrepancies among groups, with results presented in terms of frequencies and percentages. For measurement data, the normality of distribution was first assessed using the Shapiro-Wilk test. When the data followed a normal distribution, they were expressed as mean \pm standard deviation, and independent-sample t-tests were performed. The non-normally distributed data were expressed as median and interquartile range (IQR), and Z-tests were applied for inter-group comparisons. When comparing multiple groups, either one-way ANOVA or Kruskal-Wallis tests were used as appropriate. Multivariate Logistic regression analysis was performed to identify independent risk factors significantly associated with LNM. Variables were selected based on clinical relevance and statistical significance ($P < 0.05$). The results were presented as odds ratios (ORs) with 95% confidence intervals (CIs). Stepwise regression methods (forward selection) were employed to optimize the model. To assess model performance, we calculated the AUC, specificity, sensitivity, and accuracy. A higher AUC value reflects a better discriminatory power. The Delong test was used to compare AUCs between different models (e.g., AdaBoost, XGBoost), and the model with the highest AUC was selected as optimal. Confusion matrices were used to evaluate classification performance, including accuracy, recall, and F1-score. K-fold cross-validation ($K=10$) was conducted on the training set to evaluate model stability and generalization ability, ensuring that the model did not overfit the data. The cross-validation process helped assess the reliability and applicability of the model in diverse scenarios.

Results

Comparison of clinical characteristics between patients in training and validation sets

In the present study, an in-depth comparison of the clinical characteristics was conducted between patients in the training ($n=428$) and

validation ($n=184$) sets. The findings revealed that age ($P=0.279$), BMI ($P=0.303$), disease type ($P=0.862$), tumor size ($P=0.333$), FIGO grading ($P=0.323$), differentiation degree ($P=0.476$), lymphovascular invasion ($P=0.680$), history of hypertension ($P=0.875$), history of diabetes ($P=0.711$), and LNM ($P=0.481$) were not significantly different between the groups (for detailed information, refer to **Table 1**).

Comparison of patient measurement data between training and validation sets

In the comparison of measurement data between the training and validation sets, we found no marked differences in the distribution of various indices ($P > 0.05$), including SCC-Ag ($P=0.811$), WBC ($P=0.511$), NEUT ($P=0.904$), LYM ($P=0.493$), MON ($P=0.834$), HGB ($P=0.958$), PLT ($P=0.603$), ALT ($P=0.722$), AST ($P=0.755$), PAB ($P=0.877$), TP ($P=0.871$), TG ($P=0.870$), HDL ($P=0.493$), LDL ($P=0.344$), and Cysc ($P=0.681$), indicating that the two groups of patients were consistent in these indicators (**Table 2**).

Comparison of clinical characteristics between metastatic and non-metastatic patients in the training set

Through the comparison of count data and measurement data between the metastatic and non-metastatic patients in the training set, it was found that the majority of clinical factors did not exhibit statistically significant discrepancies between the two groups. Concerning the count data, no significant differences were manifested in age ($P=0.867$), BMI ($P=0.273$), disease type ($P=0.426$), history of hypertension ($P=0.252$), and history of diabetes ($P=0.176$). Nevertheless, statistically significant differences were presented in tumor size ($P=0.008$), FIGO grading ($P=0.001$), differentiation degree ($P=0.015$), and lymphovascular invasion ($P=0.004$) (**Table 3**). In terms of measurement data, no significant differences were detected in LYM ($P=0.418$), MON ($P=0.474$), PLT ($P=0.261$), ALT ($P=0.773$), TP ($P=0.788$), TG ($P=0.306$), HDL ($P=0.152$), and Cysc ($P=0.420$) between the LNM and non-LNM groups. However, significant differences were evident in SCC-Ag ($P < 0.001$), WBC ($P=0.028$), NEUT ($P=0.045$), HGB ($P=0.005$), AST ($P=0.025$), PAB ($P=0.001$), and LDL ($P=0.034$) between groups (**Table 4**).

Predicting cervical cancer lymph node metastasis with AdaBoost

Table 1. Comparison of count data between patients in the validation and training sets

Variable	Total	Training set (n=428)	Validation set (n=184)	Statistic	P
Age					
<45	279	189	90	1.173	0.279
≥45	333	239	94		
BMI					
<24 kg/m ²	208	151	57	1.062	0.303
≥24 kg/m ²	404	277	127		
Disease type					
Squamous cell carcinoma	544	380	164	0.298	0.862
Adenocarcinoma	60	43	17		
Others	8	5	3		
Tumor size					
<2 cm	302	203	99	2.199	0.333
2-4 cm	273	199	74		
>4 cm	37	26	11		
FIGO grading					
IB1	273	181	92	3.480	0.323
IB2	217	157	60		
IB3	37	26	11		
IIA1	85	64	21		
Differentiation degree					
Well differentiated	238	160	78	1.484	0.476
Moderately differentiated	231	164	67		
Poorly differentiated	143	104	39		
Lymphovascular invasion					
Without	130	89	41	0.170	0.680
With	482	339	143		
History of hypertension					
Without	102	72	30	0.025	0.875
With	510	356	154		
History of diabetes					
Without	71	51	20	0.137	0.711
With	541	377	164		
LNM					
Without	110	80	30	0.497	0.481
With	502	348	154		

Note: BMI, body mass index; FIGO, International Federation of Gynecology and Obstetrics; LNM, lymph node metastasis.

Multivariate Logistic regression analysis for screening risk factors of LNM and the association between risk factors and LNM

In this study, the measurement data were subjected to dichotomization based on the cut-off value, then Logistic regression analysis was carried out after assignment of the clinical variables (**Table 5**). Logistic regression analysis found that FIGO staging (P=0.014, OR=1.517),

SCC-Ag (P<0.001, OR=24.057), WBC (P=0.015, OR=0.416), NEUT (P<0.001, OR=0.281), HGB (P=0.005, OR=0.379), and PAB (P=0.018, OR=0.460) were significant associated with LNM. In contrast, variables such as tumor size (P=0.857), differentiation degree (P=0.131), lymphovascular invasion (P=0.326), AST (P=0.196), and LDL (P=0.077) did not demonstrate statistical significance (**Table 6**). Moreover, the Sankey diagram (**Figure 2**) presented the rela-

Predicting cervical cancer lymph node metastasis with AdaBoost

Table 2. Patient measurement data in training and validation sets

Variable	Method	Training set (n=428)	Verification set (n=184)	Statistic	P
SCC-Ag (ng/mL)	Mann-Whitney U	2.85 [1.91, 3.84]	2.83 [1.73, 4.05]	0.239	0.811
WBC ($\times 10^9/L$)	t-test	6.01 \pm 1.66	5.91 \pm 1.78	-0.658	0.511
NEUT ($\times 10^9/L$)	Mann-Whitney U	3.12 [2.29, 4.09]	3.18 [2.11, 4.35]	0.121	0.904
LYM ($\times 10^9/L$)	t-test	1.87 \pm 0.64	1.83 \pm 0.58	-0.686	0.493
MON ($\times 10^9/L$)	t-test	0.41 \pm 0.15	0.41 \pm 0.15	-0.210	0.834
HGB (g/L)	t-test	121.15 \pm 12.74	121.08 \pm 14.11	-0.053	0.958
PLT ($\times 10^9/L$)	t-test	239.80 \pm 57.59	242.73 \pm 66.31	0.520	0.603
ALT (U/L)	t-test	13.89 \pm 5.29	14.06 \pm 5.16	0.356	0.722
AST (U/L)	Mann-Whitney U	19.88 [11.93, 29.10]	19.37 [12.33, 28.79]	0.312	0.755
PAB (mg/L)	t-test	262.52 \pm 61.47	261.70 \pm 59.13	-0.155	0.877
TP (g/L)	t-test	67.60 \pm 5.01	67.66 \pm 4.61	0.163	0.871
TG (mmol/L)	Mann-Whitney U	1.14 [0.57, 1.74]	1.12 [0.56, 1.74]	0.164	0.870
HDL (mmol/L)	t-test	1.14 \pm 0.24	1.15 \pm 0.26	0.686	0.493
LDL (mmol/L)	Mann-Whitney U	2.67 [2.17, 3.13]	2.59 [2.20, 2.97]	0.947	0.344
Cycc (mg/L)	t-test	0.79 \pm 0.20	0.80 \pm 0.21	0.411	0.681

Note: SCC-Ag, squamous cell carcinoma antigen; WBC, white blood cell count; NEUT, neutrophil count; LYM, lymphocyte count; MON, monocyte count; HGB, hemoglobin; PLT, platelet count; ALT, alanine aminotransferase; AST, aspartate transaminase; PAB, prealbumin; TP, total protein; TG, triglyceride; HDL, high-density lipoprotein; LDL, low-density lipoprotein; Cycc, cystatin C.

relationship between LNM and diverse clinical variables. Notably, the LNM status demonstrated a significant correlation with FIGO staging, SCC-Ag, WBC, NEUT, HGB, and PAB levels, thereby providing an intuitive visualization of the distribution of these variables under varying conditions.

ROC curve analysis and performance comparison of multiple models based on the training and validation sets

In the model analysis of the training set, the ROC curve performances of different models exhibit variations (**Table 7**). XGBoost had an AUC of 0.896 (95% CI: 0.858-0.933), demonstrating the highest predictive performance, followed by LASSO regression (AUC=0.877), Adaboost (AUC=0.882), SVM (AUC=0.740), and Random Forest (AUC=0.738). The specificity and sensitivity of XGBoost were 84.48% and 80.00%, respectively, with the Youden index reaching 64.48% and its accuracy rate being 83.64%. Through the Delong test (**Table 8**), we found no statistically significant difference in AUC between XGBoost and Adaboost ($P=0.18$). However, significant differences in AUC exist between XGBoost and Random Forest, SVM, and LASSO regression ($P<0.001$), indicating the superior performance of XGBoost over the latter models. See **Figure 3**.

In the validation set, the AUCs of XGBoost and AdaBoost were 0.855 (95% CI: 0.788-0.922) and 0.857 (95% CI: 0.791-0.924), respectively, showing no significant difference in performance between them ($P=0.862$) (**Tables 9 and 10**). The AUC of LASSO regression was 0.859, and that of Random Forest and SVM was both 0.751. The Delong test indicated no statistically significant difference in AUC between XGBoost and AdaBoost or LASSO regression ($P>0.05$), but there was significant difference between XGBoost and Random Forest and SVM ($P<0.001$), further indicating that the performance of XGBoost and Adaboost in the validation set is superior to that of Random Forest and SVM.

Discussion

This study aims to construct and evaluate the performance of multiple ML models, especially the AdaBoost model, in predicting LNM in CC patients. The research findings demonstrate that the AdaBoost model exhibits excellent performance in predicting LNM and outperforms traditional statistical models as well as other commonly employed ML models, such as Random Forest, SVM, and LASSO regression. Specifically, its AUCs in the training set and the validation set were 0.882 and 0.857 respectively, indicating highly efficient predictive per-

Predicting cervical cancer lymph node metastasis with AdaBoost

Table 3. Comparison of counting data between metastatic and non-metastatic patients in the training set

Variable	Total	LNM group (n=80)	Non-LNM group (n=348)	Statistic	P
Age					
<45	189	36	153	0.028	0.867
≥45	239	44	195		
BMI					
<24 kg/m ²	151	24	127	1.201	0.273
≥24 kg/m ²	277	56	221		
Disease type					
Squamous cell carcinoma	380	69	311	1.707	0.426
Adenocarcinoma	43	9	34		
Others	5	2	3		
Tumor size					
<2 cm	203	29	174	9.758	0.008
2-4 cm	199	41	158		
>4 cm	26	10	16		
FIGO grading					
IB1	181	24	157	15.523	0.001
IB2	157	27	130		
IB3	26	10	16		
IIA1	64	19	45		
Differentiation degree					
Well differentiated	160	20	140	8.343	0.015
Moderately differentiated	164	41	123		
Poorly differentiated	104	19	85		
Lymphovascular invasion					
Without	89	26	63	8.185	0.004
With	339	54	285		
History of hypertension					
Without	72	10	62	1.314	0.252
With	356	70	286		
History of diabetes					
Without	51	6	45	1.828	0.176
With	377	74	303		

Note: LNM, Lymph node metastasis; BMI, body mass index; FIGO, International Federation of Obstetrics and Gynecology.

formance. Moreover, multivariate Logistic regression analysis has disclosed the significant predictive impacts of several key independent risk factors on LNM, including FIGO stage, SCC-Ag, WBC, NEUT, HGB, and PAB. These factors are of significant statistical significance in the model, suggesting their crucial roles in predicting LNM in CC.

The utilization of the AdaBoost model for predicting LNM in CC patients holds profound clinical significance. To begin with, preoperative

non-invasive evaluation of the LNM status can substantially enhance the treatment planning and prognostic assessment of patients. Traditional lymph node dissection surgery, despite being the “gold standard” for the diagnosis of LNM, entails certain risks of complications due to its invasive nature, such as lymphocele formation, infections, and hemorrhages. In contrast, the AdaBoost model, as a non-invasive instrument, is capable of accurately predicting the LNM status prior to surgery, thereby reducing unnecessary surgical procedures and miti-

Predicting cervical cancer lymph node metastasis with AdaBoost

Table 4. Comparison of measurement data between metastatic and non-metastatic patients in the training set

Variable	Method	LNM group (n=80)	Non-LNM group (n=348)	Statistic	P
SCC-Ag (ng/mL)	Mann-Whitney U	5.33 [2.64, 9.08]	2.71 [1.64, 3.65]	6.874	<0.001
WBC ($\times 10^9/L$)	t-test	5.51 \pm 1.42	6.00 \pm 1.84	2.207	0.028
NEUT ($\times 10^9/L$)	Mann-Whitney U	2.71 [1.29, 4.71]	3.27 [2.33, 4.31]	2.006	0.045
LYM ($\times 10^9/L$)	Mann-Whitney U	1.81 [1.40, 2.16]	1.81 [1.46, 2.23]	0.810	0.418
MON ($\times 10^9/L$)	t-test	0.42 \pm 0.12	0.41 \pm 0.15	-0.716	0.474
HGB (g/L)	t-test	116.35 \pm 17.15	122.17 \pm 13.11	-2.851	0.005
PLT ($\times 10^9/L$)	t-test	235.06 \pm 67.93	244.51 \pm 65.89	-1.128	0.261
ALT (U/L)	t-test	14.21 \pm 6.35	14.02 \pm 4.85	-0.289	0.773
AST (U/L)	Mann-Whitney U	23.59 [13.20, 31.69]	18.73 [12.21, 27.34]	2.249	0.025
PAB (mg/L)	t-test	242.70 \pm 49.18	266.06 \pm 60.41	3.222	0.001
TP (g/L)	t-test	67.54 \pm 4.59	67.69 \pm 4.63	0.269	0.788
TG (mmol/L)	Mann-Whitney U	1.04 [0.73, 1.57]	1.16 [0.56, 1.78]	1.024	0.306
HDL (mmol/L)	t-test	1.12 \pm 0.23	1.16 \pm 0.26	1.435	0.152
LDL (mmol/L)	t-test	2.48 \pm 0.43	2.64 \pm 0.65	2.124	0.034
CyCs (mg/L)	t-test	0.81 \pm 0.23	0.79 \pm 0.20	-0.807	0.420

Notes: LNM, Lymph node metastasis; SCC-Ag, squamous cell carcinoma antigen; WBC, white blood cell count; NEUT, neutrophil count; LYM, lymphocyte count; MON, monocyte count; HGB, hemoglobin; PLT, platelet count; ALT, alanine aminotransferase; AST, aspartate transaminase; PAB, prealbumin; TP, total protein; TG, triglyceride; HDL, high-density lipoprotein; LDL, low-density lipoprotein; CyCs, cystatin C.

Table 5. Assignment table

Variable	Assignment
FIGO staging	IB1=0, IB2=1, IB3=2, IIA1=3
Tumor size	<2 cm =0, 2-4 cm =1, >4 cm =2
Differentiation degree	Well differentiated =0, moderately differentiated =1 and poorly differentiated =2
Lymphovascular invasion	With =1, without =0
SCC-Ag (ng/mL)	<5.15=0, \geq 5.15=1
WBC ($\times 10^9/L$)	<6.185=0, \geq 6.185=1
NEUT ($\times 10^9/L$)	<1.79=0, \geq 1.79=1
HGB (g/L)	<114.5=0, \geq 114.5=1
AST (U/L)	<25.345=0, \geq 25.345=1
PAB (mg/L)	<249.255=0, \geq 249.255=1
LDL (mmol/L)	<2.965=0, \geq 2.965=1

Note: FIGO, International Federation of Obstetrics and Gynecology; SCC-Ag, squamous cell carcinoma antigen; WBC, white blood cell count; NEUT, neutrophil count; HGB, hemoglobin; AST, aspartate aminotransferase; PAB, prealbumin; LDL, low-density lipoprotein.

gating the treatment risks for patients. The literature has indicated that LNM prediction based on ML models can effectively facilitate clinical decision-making [21]. Moreover, Deng et al. [8] demonstrated that a preoperative multivariate Logistic regression model integrating clinical and pathological data could effectively predict the LNM risk in early-stage CC, thereby preventing excessive surgical interventions. Similar research endeavors have also devel-

oped prediction models by integrating 3D-PDU parameters and clinical features, further corroborating the clinical practicability of non-invasive tools [22].

The early identification of lymph node status is of paramount significance for the formulation of personalized treatment strategies. In CC patients, those diagnosed with no LNM in the early stage can achieve a 5-year survival rate

Predicting cervical cancer lymph node metastasis with AdaBoost

Table 6. Risk factors

Variable	β	SE	P	OR value	Lower	Upper
FIGO staging	0.417	0.169	0.014	1.517	1.081	2.108
Tumor size	0.054	0.296	0.857	1.055	0.589	1.894
Differentiation degree	0.241	0.159	0.131	1.272	0.931	1.742
Lymphovascular invasion	-0.369	0.376	0.326	0.691	0.334	1.469
SCC-Ag (ng/mL)	3.180	0.402	<0.001	24.057	11.246	54.739
WBC ($\times 10^9/L$)	-0.877	0.360	0.015	0.416	0.200	0.827
NEUT ($\times 10^9/L$)	-1.268	0.352	<0.001	0.281	0.140	0.560
HGB (g/L)	-0.969	0.343	0.005	0.379	0.192	0.741
AST (U/L)	0.437	0.338	0.196	1.548	0.792	3.003
PAB (mg/L)	-0.777	0.329	0.018	0.460	0.239	0.874
LDL (mmol/L)	-0.772	0.437	0.077	0.462	0.186	1.044

Note: FIGO, International Federation of Obstetrics and Gynecology; SCC-Ag, squamous cell carcinoma antigen; WBC, white blood cell count; NEUT, neutrophil count; HGB, hemoglobin; AST, aspartate aminotransferase; PAB, prealbumin; LDL, low-density lipoprotein.

as high as 90%. However, for patients with LNM, the 5-year survival rate drops markedly to around 65%. Studies have demonstrated that non-invasive models can accurately predict LNM and optimize the selection of the surgical extent, reduce unnecessary lymph node dissections, and guide the application of radiotherapy and chemotherapy, thereby improving patient prognosis [23]. The findings of this study are in concordance with those presented in the extant literature and also proffer novel perspectives. Numerous studies have revealed that traditional statistical models have certain limitations in predicting tumor metastasis, primarily due to their ineptitude in handling high-dimensional and convoluted non-linear relationships [24]. In contrast, ML models, especially ensemble learning approaches such as AdaBoost and XGBoost, can more effectively capture the intricate patterns within the data and augment prediction accuracy. For instance, Yang et al. [25] developed a Random Forest-based model for predicting LNM in cervical squamous cell carcinoma, and the results indicated that the model exhibited excellent performance when handling multi-dimensional clinical data. Additionally, investigations based on MRI data also accentuated the advantages of the Random Forest model in predicting LNM [26].

In this study the AdaBoost model has demonstrated high AUC values, favorable specificity, and high-level accuracy, manifesting its capability to capture nonlinear relationships within a

complex data milieu. Specifically, its AUC in the training set was 0.882 and in the validation set was 0.857, both of which were significantly superior to those of Random Forest (AUC=0.738 and 0.751) and SVM (AUC=0.740 and 0.751). The literature indicates that the AdaBoost model performs exceptionally well in handling data-imbalance issues, particularly in scenarios where the incidence of LNM is low, while also maintaining high levels of sensitivity and specificity [27]. Zuo et al. [28] reported that, among a multiplicity of ML algorithms, AdaBoost exhibited optimal performance in predicting the recurrence risk of breast cancer; moreover, the utilization of the SHAP method enhanced the model's interpretability, thereby rendering it more conducive to clinical decision support. Additionally, the AdaBoost-based prototype tree model developed by Liang et al. [29] not only demonstrated outstanding predictive performance in prognosticating pathological images of colorectal cancer but also resolved the "accuracy-interpretability trade-off" dilemma of the model by visualizing the decision-making process, which is in line with the application of the AdaBoost model in predicting LNM of CC in the current study. Although XGBoost had a slightly higher AUC (0.896) than AdaBoost in the training set, AdaBoost's performance was equivalent to that of XGBoost in the validation set (AUC=0.857 vs. 0.855), with no statistical significance (P=0.18). Furthermore, the AdaBoost model is characterized by lower complexity and a rational consumption of computa-

Predicting cervical cancer lymph node metastasis with AdaBoost

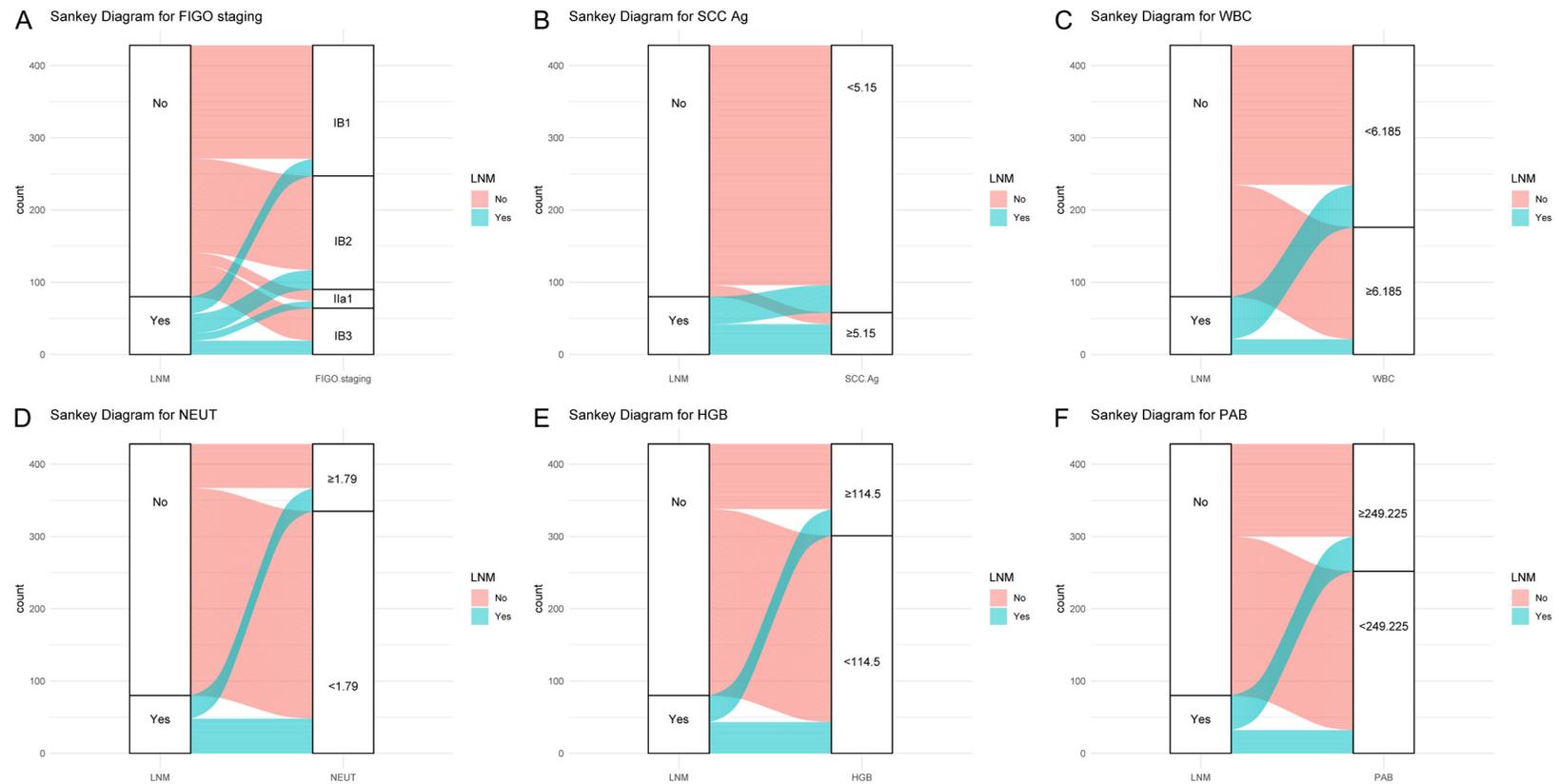


Figure 2. The relationship between LNM and various clinical variables illustrated by the Sankey diagram. A. FIGO staging: the distribution of patients in different FIGO stages (IB1, IB2, IIA1, IIA2, and IIB) according to the LNM status. B. SCC-Ag: the relationship between LNM and SCC-Ag levels (≤ 1.5 vs. > 1.5). C. WBC: the flow relationship between LNM and (WBC (≤ 9.86 vs. > 9.86)). D. NEUT: the relationship between LNM and NEUT (≤ 7.79 vs. > 7.79). E. HGB: the connection between LNM and HGB level (≤ 117.4 vs. > 117.4). F. PAB: the distribution of patients with different PAB levels (≤ 204.22 and > 204.22) according to the LNM status. Note: LNM, lymph node metastasis; FIGO, International Federation of Obstetrics and Gynecology; SCC-Ag, squamous cell carcinoma antigen; WBC, white blood cell count; NEUT, neutrophil count; HGB, hemoglobin; PAB, prealbumin.

Predicting cervical cancer lymph node metastasis with AdaBoost

Table 7. ROC curve parameters of models in the training set

Marker	AUC	95% CI	Specificity	Sensitivity	Youden index	Accuracy	Precision	F1 Score
XGBoost	0.896	0.858-0.933	84.48%	80.00%	64.48%	83.64%	80.00%	64.65%
AdaBoost	0.882	0.843-0.922	88.51%	71.25%	59.76%	85.28%	71.25%	64.41%
Random Forest	0.738	0.682-0.794	95.11%	52.50%	47.61%	87.15%	52.50%	60.43%
SVM	0.74	0.683-0.796	95.40%	52.50%	47.90%	87.38%	52.50%	60.87%
LASSO	0.877	0.837-0.917	72.41%	86.25%	58.66%	75.00%	86.25%	56.33%

Note: ROC, receiver operating characteristic; XGBoost, eXtreme Gradient Boosting; AdaBoost, Adaptive Boosting; SVM, Support Vector Machine; LASSO, Least Absolute Shrinkage and Selection Operator.

Table 8. AUC comparison of models in the training set

Marker 1	Marker 2	Z value	P value	AUC difference	95% CI
XGBoost	AdaBoost	1.342	0.18	0.013	-0.006-0.032
XGBoost	Random Forest	6.118	<0.001	0.157	0.107-0.208
XGBoost	SVM	6.904	<0.001	0.156	0.112-0.200
XGBoost	LASSO	2.058	0.04	0.018	0.001-0.036
AdaBoost	Random Forest	5.861	<0.001	0.144	0.096-0.193
AdaBoost	SVM	6.375	<0.001	0.143	0.099-0.187
AdaBoost	LASSO	0.594	0.553	0.005	-0.012-0.022
Random Forest	SVM	-0.059	0.953	-0.001	-0.049-0.046
Random Forest	LASSO	-5.936	<0.001	-0.139	-0.185 - -0.093
SVM	LASSO	-6.497	<0.001	-0.138	-0.179 - -0.096

Note: AUC, area under the curve; XGBoost, eXtreme Gradient Boosting; AdaBoost, Adaptive Boosting; SVM, Support Vector Machine; LASSO, Least Absolute Shrinkage and Selection Operator.

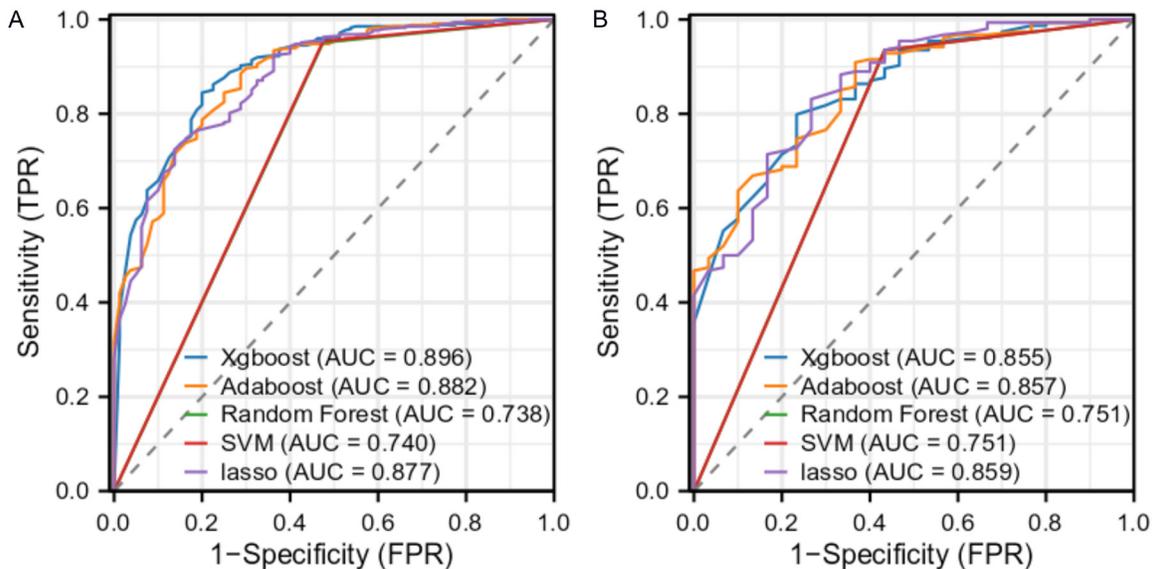


Figure 3. ROC curve representation of the performance of different machine-learning models in the training and validation sets. A. ROC curve representation of the performance of XgBoost, AdaBoost, Random Forest, SVM, and LASSO regression models in the training set. B. Performance of the same models in the validation set. Note: ROC, receiver operating characteristic; XGBoost, eXtreme Gradient Boosting; AdaBoost, Adaptive Boosting; SVM, Support Vector Machine; LASSO, Least Absolute Shrinkage and Selection Operator.

tional resources, making it appropriate for clinical applications.

Multivariate Logistic regression analysis has identified FIGO staging, SCC-Ag, WBC, NEUT,

Predicting cervical cancer lymph node metastasis with AdaBoost

Table 9. ROC curve parameters of models in the validation set

Marker	AUC	95% CI	Specificity	Sensitivity	Youden index	Accuracy	Precision	F1 Score
XGBoost	0.855	0.788-0.922	79.87%	76.67%	56.54%	79.35%	76.67%	54.76%
AdaBoost	0.857	0.791-0.924	90.91%	63.33%	54.24%	86.41%	63.33%	60.32%
Random Forest	0.751	0.659-0.843	93.51%	56.67%	50.17%	87.50%	56.67%	59.65%
SVM	0.751	0.659-0.843	93.51%	56.67%	50.17%	87.50%	56.67%	59.65%
LASSO	0.859	0.789-0.930	83.12%	73.33%	56.45%	81.52%	73.33%	56.41%

Note: ROC, receiver operating characteristic; XGBoost, eXtreme Gradient Boosting; AdaBoost, Adaptive Boosting; SVM, Support Vector Machine; LASSO, Least Absolute Shrinkage and Selection Operator.

Table 10. AUC comparison of models in the validation set

Marker 1	Marker 2	Z value	P value	AUC difference	95% CI
XGBoost	AdaBoost	-0.174	0.862	-0.003	-0.033-0.028
XGBoost	Random Forest	3.443	<0.001	0.104	0.045-0.163
XGBoost	SVM	3.443	<0.001	0.104	0.045-0.163
XGBoost	LASSO	-0.277	0.782	-0.005	-0.038-0.028
AdaBoost	Random Forest	3.291	<0.001	0.107	0.043-0.170
AdaBoost	SVM	3.291	<0.001	0.107	0.043-0.170
AdaBoost	LASSO	-0.152	0.879	-0.002	-0.027-0.023
Random Forest	SVM	0	1	0	0.000-0.000
Random Forest	LASSO	-3.562	<0.001	-0.109	-0.168 - -0.049
SVM	LASSO	-3.562	<0.001	-0.109	-0.168 - -0.049

Note: AUC, area under the curve; XGBoost, eXtreme Gradient Boosting; AdaBoost, Adaptive Boosting; SVM, Support Vector Machine; LASSO, Least Absolute Shrinkage and Selection Operator.

HGB, and PAB as independent risk factors for LNM in CC. These factors possess well-defined biological significance within the clinical context, thereby facilitating the elucidation of their importance in LNM prediction. First of all, FIGO staging, as an important indicator of tumor spread, reflects both the local and distant spread of cancer. A plethora of studies have substantiated that elevated FIGO stages (IB2, IB3, IIA1, etc.) are usually concomitant with an augmented risk of LNM, which may be ascribed to the amplification of tumor volume, the exacerbation of the infiltration depth, and the augmentation of tumor cell invasion [25]. Therefore, FIGO staging constitutes a pivotal factor in the evaluation of the LNM risk.

Furthermore, SCC-Ag is a tumor marker with high specificity, and an elevation in its level is usually closely related to the tumor burden of CC. Research has indicated that high levels of SCC-Ag may suggest a higher proliferation rate of cancer cells and metastatic potential, thereby significantly augmenting the risk of LNM. The high sensitivity and high specificity of SCC-Ag in detection confer upon its great significance in

clinical applications, especially in the independent and reliable prediction of LNM [24]. The research by Jiang et al. [30] further demonstrated that the combined utilization of SCC-Ag and semi-quantitative parameters of ^{18}F -FDG PET/CT can significantly enhance the predictive ability of LNM. Specifically, by augmenting the predictive value of parameters such as TLG, the combined application of SCC-Ag and SUV_{peak} exhibits high diagnostic efficacy. Additionally, Xu et al. [31] demonstrated the superiority of the comprehensive parameters of K_{trans}, f value, and SCC-Ag in predicting pelvic LNM in early-stage CC by integrating multi-parameter MRI data and SCC-Ag, attaining a sensitivity of 79.1% and a specificity of 94.0%. Chen et al. [32] also noted that optimizing the threshold of SCC-Ag plays a crucial role in the sensitivity and specificity of LNM prediction for patients with early-stage cervical squamous cell carcinoma, providing guidance for the application of SCC-Ag in the management of early-stage CC patients.

WBC and NEUT are conventional hematological indices that mirror the inflammatory status of

the body. In recent years, a multitude of studies have unraveled the intimate nexus between inflammatory responses and tumor progression. In the tumor microenvironment, inflammatory cells can potentiate the proliferation and migration of tumor cells through the secretion of cytokines, chemokines, and other relevant biomolecules [24]. Elevated levels of WBC and NEUT may signify the presence of a potent inflammatory response in the body, which may promote the invasion of tumor cells into lymph nodes. Chronic inflammation not only exerts an impact on immune function but also amplifies the incidence of metastasis by remodeling the microenvironment structure. These mechanisms are of particular importance in CC patients. Nutritional status markers such as HGB and PAB have also been identified as critical factors influencing tumor metastasis. Low HGB levels are usually associated with malnutrition and chronic illness in patients, which may precipitate a decline in immune function and render patients more vulnerable to tumor invasion [33]. As a more sensitive nutritional indicator, a decrease in PAB level indicates potential nutritional insufficiency and cachexia, a condition that further debilitates the body's ability to resist tumor invasion. Consequently, patients with suboptimal nutritional status are more predisposed to LNM. Collectively, the importance of these blood indices in LNM prediction is tightly related to their physiological and pathological basis. These blood-based indices, which are readily accessible in clinical practice, can not only provide reliable LNM-predictive information preoperatively but also furnish a scientific rationale for the formulation of personalized treatment protocols.

Despite the advancements in this research, several limitations remain. First, the utilization of single-center retrospective data in this study, coupled with a relatively diminutive sample size, may potentially circumscribe the generalization capacity of the model. Second, biases may exist in the feature selection process. Despite the application of multivariate Logistic regression and RFE methods for screening significant features, there remains a likelihood of overlooking some crucial variables. Specifically, imaging data, which can provide detailed anatomical and functional information about the tumor and surrounding tissues, and genomic data, which holds the key to understanding the genetic basis of the disease and its potential

progression, might not have been comprehensively incorporated. Future research endeavors should incorporate a greater volume of multi-modal data, including MRI images and heart rate variability (HRV), to further augment the prognostic capability of the model.

Conclusion

This study successfully developed an effective AdaBoost-based prediction model for LNM in CC, utilizing retrospective analysis and various ML approaches. The AdaBoost model demonstrated strong predictive performance and can serve as a valuable non-invasive tool in clinical settings. Future research, incorporating multi-center data and further optimization with advanced algorithmic techniques, is expected to enhance the model's accuracy and clinical applicability, thereby providing more reliable guidance for personalized treatment strategies in CC.

Disclosure of conflict of interest

None.

Address correspondence to: Fenghua Liu, Department of Gynecology, The Affiliated Cancer Hospital of Zhengzhou University and Henan Cancer Hospital, No. 127 Dongming Road, Zhengzhou 450008, Henan, China. Tel: +86-0371-65587224; E-mail: liufenghuadoctor@126.com

References

- [1] Voelker RA. Cervical cancer screening. *JAMA* 2023; 330: 2030.
- [2] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024; 74: 229-263.
- [3] Xia C, Dong X, Li H, Cao M, Sun D, He S, Yang F, Yan X, Zhang S, Li N and Chen W. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin Med J (Engl)* 2022; 135: 584-590.
- [4] Zhen S, Qiang R, Lu J, Tuo X, Yang X and Li X. CRISPR/Cas9-HPV-liposome enhances antitumor immunity and treatment of HPV infection-associated cervical cancer. *J Med Virol* 2023; 95: e28144.
- [5] Ramachandran D and Dörk T. Genomic risk factors for cervical cancer. *Cancers (Basel)* 2021; 13: 5137.

Predicting cervical cancer lymph node metastasis with AdaBoost

- [6] He H, He M, Zhou Q, Tang Y, Wang J, Li X and Zou D. Genetic analysis of cervical cancer with lymph node metastasis. *J Gynecol Oncol* 2024; 35: e102.
- [7] Yang S, Liu C, Li C and Hua K. Nomogram predicting lymph node metastasis in the early-stage cervical cancer. *Front Med (Lausanne)* 2022; 9: 866283.
- [8] Deng YR, Chen XJ, Xu CQ, Wu QZ, Zhang W, Guo SQ and Li LX. A preoperative nomogram predicting risk of lymph node metastasis for early-stage cervical cancer. *BMC Womens Health* 2023; 23: 568.
- [9] Zhao D, Li B, Zheng S, Ou Z, Zhang Y, Wang Y, Liu S, Zhang G and Yuan G. Separate lateral parametrial lymph node dissection improves detection rate of parametrial lymph node metastasis in early-stage cervical cancer: 10-year clinical evaluation in a single center in China. *Chin J Cancer Res* 2020; 32: 804-814.
- [10] Wang G, Li X, Li L, Liu D, Sun R, Zhang Q, Geng C, Gong H and Gao X. Clinical value of ultrasonic imaging in diagnosis of hypopharyngeal cancer with cervical lymph node metastasis. *Oncol Lett* 2019; 18: 5917-5922.
- [11] Bizzarri N, Obermair A, Hsu HC, Chacon E, Collins A, Tsibulak I, Mutombo A, Abu-Rustum NR, Balaya V, Buda A, Cibula D, Covens A, Fanfani F, Ferron G, Frumovitz M, Guani B, Kocian R, Kohler C, Leblanc E, Lecuru F, Leitao MM Jr, Mathevet P, Mueller MD, Papadia A, Pareja R, Plante M, Querleu D, Scambia G, Tanner E, Zapardiel I, Garcia JR and Ramirez PT. Consensus on surgical technique for sentinel lymph node dissection in cervical cancer. *Int J Gynecol Cancer* 2024; 34: 504-509.
- [12] Chen W, Xiu S, Xie X, Guo H, Xu Y, Bai P and Xia X. Prognostic value of tumor measurement parameters and SCC-Ag changes in patients with locally-advanced cervical cancer. *Radiat Oncol* 2022; 17: 6.
- [13] Duan W, Wang W and He C. A novel potential inflammation-nutrition biomarker for predicting lymph node metastasis in clinically node-negative colon cancer. *Front Oncol* 2023; 13: 995637.
- [14] Zhang J, Gong J, Liu H, Zhou W, Cai M and Zhang C. Prognostic nutritional index predicts lateral lymph node metastasis and recurrence free survival in papillary thyroid carcinoma. *BMC Cancer* 2024; 24: 1039.
- [15] Wakatsuki K, Matsumoto S, Migita K, Ito M, Kunishige T, Nakade H, Nakatani M, Kitano M and Sho M. Preoperative plasma fibrinogen is associated with lymph node metastasis and predicts prognosis in resectable esophageal cancer. *World J Surg* 2017; 41: 2068-2077.
- [16] Haug CJ and Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. *N Engl J Med* 2023; 388: 1201-1208.
- [17] Xiouras C, Cameli F, Quilló GL, Kavousanakis ME, Vlachos DG and Stefanidis GD. Applications of artificial intelligence and machine learning algorithms to crystallization. *Chem Rev* 2022; 122: 13006-13042.
- [18] Al Mudawi N and Alazeb A. A model for predicting cervical cancer using machine learning algorithms. *Sensors (Basel)* 2022; 22: 4132.
- [19] Marth C, Landoni F, Mahner S, McCormack M, Gonzalez-Martin A and Colombo N; ESMO Guidelines Committee. Cervical cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2017; 28: iv72-iv83.
- [20] Berek JS, Matias-Guiu X, Creutzberg C, Fotopoulou C, Gaffney D, Kehoe S, Lindemann K, Mutch D and Concin N; Endometrial Cancer Staging Subcommittee, FIGO Women's Cancer Committee. FIGO staging of endometrial cancer: 2023. *Int J Gynaecol Obstet* 2023; 162: 383-394.
- [21] Zhao H, Wang Y, Sun Y, Wang Y, Shi B, Liu J and Zhang S. Hematological indicator-based machine learning models for preoperative prediction of lymph node metastasis in cervical cancer. *Front Oncol* 2024; 14: 1400109.
- [22] Dong S, Peng YQ, Feng YN, Li XY, Gong LP, Zhang S, Du XS and Sun LT. Based on 3D-PDU and clinical characteristics nomogram for prediction of lymph node metastasis and lymphovascular space invasion of early cervical cancer preoperatively. *BMC Womens Health* 2024; 24: 438.
- [23] Xu M, Xie X, Cai L, Liu D and Sun P. Preoperative scoring system for the prediction of risk of lymph node metastasis in cervical cancer. *Sci Rep* 2024; 14: 23860.
- [24] Meng Y, Yan X and Fan J. Construction of prediction model of lymph node metastasis of early cervical cancer based on machine learning algorithm and its application: experience of 204 cases in a single center. *Am J Transl Res* 2023; 15: 1852-1861.
- [25] Yang S, Zhao J, Zhao H, Hu Y and Zhu H. Development of a nomogram for predicting pelvic lymph node metastasis in cervical squamous cell carcinoma. *Int J Gynaecol Obstet* 2023; 160: 1020-1027.
- [26] Liu S, Zhou Y, Wang C, Shen J and Zheng Y. Prediction of lymph node status in patients with early-stage cervical cancer based on radiomic features of magnetic resonance imaging (MRI) images. *BMC Med Imaging* 2023; 23: 101.
- [27] Monthatip K, Boonnag C, Muangmool T and Charoenkwan K. A machine learning-based prediction model of pelvic lymph node metastasis in women with early-stage cervical cancer. *J Gynecol Oncol* 2024; 35: e17.

Predicting cervical cancer lymph node metastasis with AdaBoost

- [28] Zuo D, Yang L, Jin Y, Qi H, Liu Y and Ren L. Machine learning-based models for the prediction of breast cancer recurrence risk. *BMC Med Inform Decis Mak* 2023; 23: 276.
- [29] Liang M, Wang R, Liang J, Wang L, Li B, Jia X, Zhang Y, Chen Q, Zhang T and Zhang C. Interpretable inference and classification of tissue types in histological colorectal cancer slides based on ensembles adaptive boosting prototype tree. *IEEE J Biomed Health Inform* 2023; 27: 6006-6017.
- [30] Jiang CZ, Zheng K, Zhang YY, Yang J, Ye H and Peng X. (18)F-FDG PET/CT semi-quantitative parameters combined with SCC-Ag in predicting lymph node metastasis in stage I-II cervical cancer. *Front Oncol* 2024; 14: 1278464.
- [31] Xu X, Liu F, Zhao X, Wang C, Li D, Kang L, Liu S and Zhang X. The value of multiparameter MRI of early cervical cancer combined with SCC-Ag in predicting its pelvic lymph node metastasis. *Front Oncol* 2024; 14: 1417933.
- [32] Chen XB, Li WL, Tan CY and Xu N. The optimal threshold for squamous-cell carcinoma antigen to predict lymph node metastasis in early-stage cervical squamous-cell carcinoma. *Asian J Surg* 2024; 47: 3909-3911.
- [33] Guan W, Wang Y, Zhao H, Lu H, Zhang S, Liu J and Shi B. Prediction models for lymph node metastasis in cervical cancer based on preoperative heart rate variability. *Front Neurosci* 2024; 18: 1275487.