Am J Transl Res 2023;15(3):1852-1861
www.ajtr.org /ISSN:1943-8141/AJTR0148834

Original Article

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

Yugang Meng[1*], Xiting Yan[2*], Jiangtao Fan[1]

[1]Department of Gynecology Division, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, PR China; [2]Department of Gynecology Division, The Second Affiliated Hospital of Guangxi Medical University, Nanning 530005, Guangxi, PR China. [*]Equal contributors and co-first authors.

Received December 29, 2022; Accepted February 23, 2023; Epub March 15, 2023; Published March 30, 2023

Abstract: Objectives: The prediction model of para-aortic lymph node metastasis (LNM) in patients with early cervical cancer was constructed based on the logistic regression (LR) and random forest (RF) algorithms in the machine learning algorithm. The prediction efficiencies of the two models were compared. Methods: The clinical data of 204 patients with early cervical cancer in the First Affiliated Hospital of Guangxi Medical University were retrospectively collected. The 204 patients were randomly divided into a training set and a verification set according to a ratio of 3:1. The training set was used to build the model. The verification set was used to evaluate model effectiveness. The para-aortic LNM prediction model of early cervical cancer was established by LR and RF. Receiver operating characteristic curve (ROC), sensitivity, and specificity were used to evaluate the prediction performances of the two models. Results: LR analysis showed that tumor diameter > 4 cm, choroidal aneurysm embolism, pelvic lymph node metastasis, and high preoperative squamous cell carcinoma antigen (SCC-Ag) level were risk factors for para-aortic LNM in patients with early cervical cancer (*P* < 0.05). The area under the ROC curve (AUC) was 0.914. The sensitivity, specificity, and accuracy were 92.6%, 66.7%, 87.0%, respectively. The results of the importance analysis of the characteristic variables of the RF showed that the top 5 variables were preoperative SCC-Ag level, tumor diameter > 4 cm, advanced clinical stage, cancer thrombus, and pelvic lymph node metastasis. The AUC of the RF was 0.883. The sensitivity, specificity, and accuracy were 90.7%, 53.3%, 82.6%, respectively. There was no significant difference in AUC between the LR and RF (*P* > 0.05). Conclusions: Both LR and RF models based on machine learning algorithm have great predictive value in predicting early cervical cancer para-aortic lymph node metastasis.

Keywords: Machine learning algorithm, logistic regression, random forest, cervical cancer, lymph node metastasis

Introduction

Cervical cancer is the leading cause of death from malignant tumors in women in developing countries. It is reported that there are 527,600 new cases and 265,700 deaths of cervical cancer in the world every year [1]. Lymph node metastasis (LNM), as one of the major modes of metastasis in cervical cancer, has a significant impact on patient prognosis. The 5-year survival rate of patients with cervical cancer and a positive LNM (35%-69%) is much lower than that of patients with negative metastasis (91%) LNM is recognized as an independent risk factor affecting the prognosis of cervical cancer patients [2, 3]. The recommended treatment for early-stage cervical cancer without LNM is radical surgical resection. Additional radiation therapy can be taken if the postoperative pathology report shows positive LNM or the presence of other high-risk factors [4]. This would make surgical treatment redundant. Combined radiotherapy and chemotherapy after the surgery can cause more serious complications compared with direct radiotherapy and chemotherapy. These complications include gastrointestinal reactions, cystitis, and the radioactive recto vaginal fistula and vesicovaginal fistula [5, 6]. Discovering the indicators for predicting cervical LNM, carrying out accurate clinical staging and evaluation of the disease before the treatment, and selecting a more reasonable and individual treatment plan are the keys to improving the survival rate and reducing recurrence. The lymph node status of early cervical cancer affects the prognosis and the choice of treatment. The gold standard for the diagnosis of LNM is the pathological examination, but pathological examination has a lag, which is not conducive to timely selection of correct treatment. Systematic lymph node dissection increases the difficulty of the operation and the risk of damage to the surrounding organs. It significantly impacts the lives of patients with long-term complications such as lymphocytes and lymphedema. Preoperative attempts were made to investigate and develop predictive models for risk factors for LNM by using various test results to help some patients avoid superimposed damage from surgery and radiotherapy.

Several scholars have studied risk factors or prediction models for LNM in early cervical cancer patients [7]. The inclusion indexes and methods of these studies vary, resulting in either inconsistent prediction results or unsatisfactory prediction accuracy of the established prediction models. This does not provide a clear conclusion worthy of reference. In recent years, with the rapid development of medical informatics and machine learning algorithms, machine learning algorithms have been gradually applied to the construction of disease risk models [8, 9]. In this study, two machine learning algorithms, logistic regression (LR), and Random Forest (RF) were used to construct prediction models of para-aortic lymph node metastasis in patients with early cervical cancer. The prediction model with a higher predictive value is screened out by comparing the two models in predicting the risk of para-aortic lymph node metastasis of early cervical cancer and providing a scientific basis for improving the prognosis of early cervical cancer.

Data and methods

Research design

Medical records of early cervical cancer patients undergoing surgical treatment in the gynecology and oncology department of the First Affiliated Hospital of Guangxi Medical University from June 2019 to June 2021 were collected for a retrospective study. A total of 204 patients with early cervical cancer were included according to inclusion criteria and exclusion criteria. This study was approved by the Ethics Committee of the First Affiliated Hospital of Guangxi Medical University.

Inclusion criteria: patients who met the diagnostic criteria of cervical cancer "Guidelines for the Standardized Treatment of Cervical Cancer and Precancerous Lesions (Trial)" [10], had stage I-II clinical staging, had undergone radical hysterectomy and bilateral lymph node dissection for the first time, had complete clinical data, followed the doctor's advice, and had regular follow-up were included.

Exclusion criteria: pregnant or lactating women; patients with gynecological malignancies or history of malignancies in other sites; patients with coagulopathy; patients with a history of antitumor therapy such as radiotherapy, and chemotherapy.

Methods of examination for para-aortic lymph node metastasis

Biograph Sensation type 16 PET/CT instrument was used for examination. The 18F-FDG was automatically synthesized by Siemens Eclipse type RD cyclotron and positron radiopharmaceuticals synthesis module (Explora FDG1), with radiochemical purity > 95%. The patient had fasted for more than 6 h and maintained a normal blood glucose level before the examination. The patient was then injected with 18F-FDG 3.7 MBg/Kg body mass and underwent routine PET/CT examination after lying in repose for 60 min. CT scan was performed first (tube voltage 140 kV, tube current 120 mA, 0.8 seconds/circumference, layer thickness 3.8 mm). After that, PET emission scanning was performed, using three-dimensional acquisition and implantation, and 3 min was collected in each position. The whole-body scanning range included the upper segment of bilateral thigh to the top of the head. The images were estimated by using the ordered subset maximum expectation method. The size, extent, and number of lymph nodes were evaluated independently by a radiologist and a nuclear medicine physician. The maximum standardized uptake (SUV_{max}) was recorded. The two disagreed and reached an agreement through negotiation. PET/CT diagnostic criteria for para-

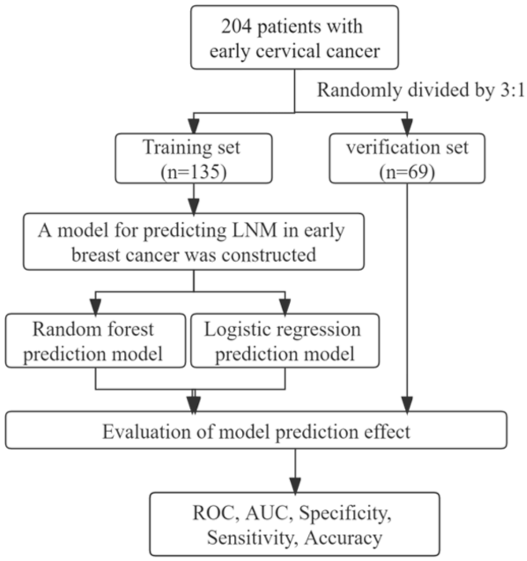

Figure 1. Research flow chart. LNM is lymph node metastasis.

aortic lymph node metastasis were lymph node diameter > 0.5 cm and SUV_{max} ≥ 2.5. A histopathological examination of the para-aortic lymph nodes was performed. If there were tumor cells in the lymph nodes, it was considered to be positive. If there were not any tumor cells present, it was considered to be negative.

Data collection

The primary outcome variable was para-aortic lymph nodes metastasize. Demographic, clinical characteristics, and baseline data that are associated with para-aortic lymph node metastasis were collected. This included age, menopause, abortion, pathological type, clinical stage, tumor diameter, cervical erosion, depth of muscular invasion, cervical canal invasion, intravascular cancer thrombus, uterine invasion, internal iliac lymph node metastasis, pelvic lymph node metastasis, and preoperative Squamous Cell Carcinoma antigen (SCC-Ag). These data are from medical records in the Clinical Information Management System.

Clinical staging was determined according to the International Federation of Gynecology and Obstetrics (FIGO) cervical cancer staging standards [11]. The tumor diameter was based on the size of the cervical lesion seen in the gynecological examination combined with vaginal ultrasound or pelvic nuclear magnetic examination. The maximum diameter of the tumor shall prevail. Intravascular cancer thrombus refers to the embolus formed by malignant tumor cells or clumps that are found in the blood vessels or lymphatic vessels connected by tumor lumps in the tumor tissue specimens removed after surgery. Other variables, such as Lymph node status, cervical erosion status, tumor histological type, and degree of cell differentiation were determined by the chief physician of pathology department based on relevant laboratory or imaging examinations.

The baseline data including demographic and clinical characteristics were collected at the time of admission. The data of abdominal aortic metastasis were collected 3 months after follow-up. All patients were followed up for 3 months. The patients with positive para-aortic lymph node metastases were put into the metastatic group. The remaining patients were classified as the non-metastatic group.

Construction of prediction model

There were 2/3 (135 cases) from the 204 patients who were randomly selected as the training set for constructing the model. The remaining 1/3 (69 cases) were used as the validation set for evaluating the model efficacy. Method of random selection: patients were numbered from 1 to 204 according to the time of diagnosis. There were 135 numbers selected as the training set by using the "sample" function of R software. The remaining numbers were put into the verification set.

The Random Forest model is an integrated machine learning algorithm based on the decision tree. The structure of the model is clear, easy to explain, and stable. It is not easy to overfit. It is commonly used to conduct cluster and regression analysis, and to evaluate the importance of influencing factors. In this study, the "randomForest" package was used for random forest regression analysis. After debugging, the parameters mtry = 5 and ntree = 500 were selected. The remaining parameters were defaulted.

The logistic regression model is a generalized linear regression analysis model. It is often used to predict the probability of disease occurrence. In this study, the "glm" package was used to establish a logistic regression model.

Table 1. Comparison of baseline data of patients with early cervical cancer [n (%)]

Variables	Metastatic group (n = 45)	Non-metastatic group (n = 159)	t/χ^2 Value	*P* Value
Age (years)	45.73±9.38	47.67±7.95	1.387	0.167
Menopause			0.076	0.782
Yes	20 (44.44)	67 (42.14)		
No	25 (55.56)	92 (57.86)		
Abortion			0.491	0.484
≤ 2 times	37 (82.22)	123 (77.36)		
> 2 times	8 (17.78)	36 (22.64)		
Pathological type			0.451	0.798
Squamous carcinoma	32 (71.11)	105 (66.04)		
Adenocarcinoma	9 (20.00)	39 (24.53)		
Squamous adenocarcinoma	4 (8.89)	15 (9.43)		
Clinical stage			18.158	< 0.001
Phase I	11 (24.44)	96 (60.38)		
Phase II	34 (75.56)	63 (39.62)		
Tumor diameter			42.420	< 0.001
≤ 4 cm	8 (17.78)	114 (71.70)		
> 4 cm	37 (82.22)	45 (28.30)		
Cervical erosion			5.5778	0.016
Yes	26 (57.78)	60 (37.74)		
No	19 (42.22)	99 (62.26)		
Depth of muscular invasion			4.672	0.031
≤ 1/2	23 (51.11)	103 (64.78)		
> 1/2	22 (48.89)	56 (35.22)		
Neck invaded			3.784	0.052
Yes	18 (40.00)	42 (26.42)		
No	27 (60.00)	117 (73.58)		
Cancer thrombus			19.989	< 0.001
Yes	33 (73.33)	57 (35.85)		
No	12 (26.67)	102 (64.15)		
Invasion uterine			6.799	0.009
Yes	30 (66.67)	71 (44.65)		
No	15 (33.33)	88 (55.35)		
Internal iliac lymph node metastasis			0.317	0.573
Yes	13 (28.89)	53 (33.33)		
No	32 (71.11)	106 (66.67)		
Pelvic lymph node metastasis			15.110	< 0.001
Yes	38 (84.44)	83 (52.20)		
No	7 (15.56)	76 (47.80)		
Preoperative SCC-Ag (μg/L)	2.43±0.73	1.74±0.32	9.226	< 0.001

Notes: SCC-Ag means Squamous Cell Carcinoma antigen.

Assessment of the prediction model

The prediction performance of the model was verified using the training set data. The receiver operating characteristic (ROC) curve analysis was used to evaluate the discriminatory ability of the model. DeLong's test was used to test the difference between the AUC values of ROC curves of the two models. Hosmer-Lemeshow was used to test the fitting effect of the model to determine whether the model could account for the difference between the predicted and

Table 2. Variable assignment description

Variables	Serial number	Assignment
Clinical stages	X1	0 = Phase I, 1 = Phase II
Tumor diameter	X2	0 = ≤ 4 cm, 1 = > 4 cm
Cervical erosion	X3	0 = No, 1 = Yes
Depth of muscularis invasion	X4	0 = ≤ 1/2, 1 = > 1/2
Cancer thrombus	X5	0 = No, 1 = Yes
Invasion uterine	X6	0 = No, 1 = Yes
Pelvic lymph node metastasis	X7	0 = No, 1 = Yes
Preoperative SCC-Ag	X8	Continuous variable

Notes: SCC-Ag means Squamous Cell Carcinoma antigen.

observed event rates. According to the model prediction results, the confusion matrix of prediction and actual results was obtained. Its accuracy, precision, recall ratio, and F1 score were calculated.

The calculation formula for accuracy rate, accuracy rate, recall rate, and F1 score is as follows:

Accuracy = (TP + TN)/(TP + TN + FN + FP), Precision = TP/(TP + FP), Recall = TP/(TP + FN), F1 value = Accuracy × Recall × 2/(Accuracy + Recall).

TP means True Positive. FP means False Positive. FN means False Negative. TN means True Negative.

Statistical analysis

SPSS Version 23 (IBM Statistical Product and Service Solutions) and R version 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria) were used for data analysis. The statistical data were expressed in percentage (%), and were compared by χ^2 test, or by continuous corrected Chi-square test or Fisher exact probability test when necessary. Quantitative data were described by mean ± standard deviation. A comparison between the groups was performed by t test. P values less than 0.05 indicated significant differences. The flow chart of this study is shown in **Figure 1**.

Results and discussion

Baseline characteristics of para-aortic lymph node metastasis of early cervical cancer

Para-aortic lymph node metastasis was detected in 45 of 204 patients with early cervical cancer, with a positive rate of 22.06%, aged from 26 to 66 years. The proportion of stage II, tumor diameter > 4 cm, cervical erosion, muscle layer infiltration depth > 1/2, cancer thrombus, uterine infiltration, and pelvic lymph node metastasis in the metastasis group was higher than that in the non-metastasis group ($P < 0.05$). The preoperative SCC-Ag in the metastasis group was higher than that in the non-metastatic group ($P < 0.05$). As shown in **Table 1**.

Results of logistic regression model construction

The condition of para-aortic lymph node metastasis was taken as the dependent variable (0 = non-metastasis, 1 = metastasis). The significant features of univariate analysis (clinical stage, tumor diameter, cervical erosion, depth of muscular invasion, cancer thrombus, invasion uterine, pelvic lymph node metastasis, and preoperative SCC-Ag) were used as independent variables for multivariate logistic regression analysis. The variable assignment table is shown in **Table 2**. The logistic regression analysis showed that tumor diameter > 4 cm, vascular tumor plug, pelvic lymph node metastasis, and preoperative SCC-Ag increase were risk factors for lymph node metastasis in patients with early cervical cancer ($P < 0.05$), as shown in **Table 3**. The model formula is P = exp (-10.654 + 2.423X2 + 1.412X5 + 1.500X7 + 2.952X8)/(1 + exp (-10.654 + 2.423X2 + 1.412X5 + 1.500X7 + 2.952X8)).

Results of random forest model construction

The training sample was 135 cases, ntree = 500, mtry = 5, and the data of the test set were classified based on this parameter setting. The out-of-bag error rate of the RF was 8.89%, indicating that the model has good generalization and no over-fitting, as shown in **Figure 2**. The order of importance of variables in the random forest model is shown in **Table 4**. Columns 1 and 0 show the contribution of each variable to the prediction of metastasis and non-metastasis respectively. The average decrease of model accuracy shows the importance of variables to the final prediction.

Table 3. Multivariate logistic regression analysis results

Variable	*B*	*SE*	*P* Value	*OR*	95% CI	
					Lower	Upper
Constant	-10.654	1.582	< 0.001			
Tumor diameter	2.423	0.671	< 0.001	11.285	3.030	42.026
Cancer thrombus	1.412	0.629	0.025	4.103	1.197	14.065
Pelvic lymph node metastasis	1.500	0.629	0.017	4.482	1.306	15.383
Preoperative SCC-Ag	2.952	0.580	< 0.001	19.137	6.134	59.700

Notes: SCC-Ag means Squamous Cell Carcinoma antigen.

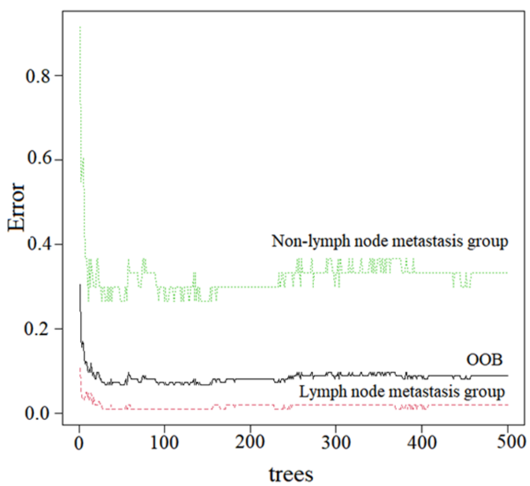

Figure 2. Random Forest model for predicting lymph node metastasis in early cervical cancer: relationship between model error rate and random forest tree orders. OOB is Out-of-Bag error rate.

Evaluation of the prediction performances and comparison of the two models

According to the verification results of the model, the confusion matrix of prediction and actual results were obtained, as shown in **Tables 5** and **6**. The sensitivity, specificity, and accuracy of LR calculated based on the confusion matrix were 92.6%, 66.7%, and 87.0%, respectively. The sensitivity, specificity, and accuracy of RF were 90.7%, 53.3%, and 82.6%, respectively, as shown in **Table 7**. The ROC curve results of the training set showed that the AUC of LR was 0.974, and the AUC of RF was 0.998, as shown in **Figure 3**. The ROC curve results of the validation set showed that the AUC of LR was 0.914 and the AUC of RF was 0.883, as shown in **Figure 4**. The DeLong test showed that there was no significant difference in LR and RF AUC between the training set and the validation set (*P* > 0.05).

The DeLong test showed that there was no significant difference between the AUC of LR and RF in the training set and the validation set (*P* > 0.05).

Discussion

LNM is an important way for cervical cancer tumor cells to spread. The five-year overall survival rate of early-stage cervical cancer has exceeded 90% [12]. The survival of patients is drastically shortened once LNM occurs [13]. Hosaka et al. [14] reported that in patients with stage IB-IIB cervical cancer, the 5-year overall survival (OS) rate was 94.8% in patients without lymph node metastasis. The 5-year OS in patients with lymph node metastasis decreased to 62%. The prognosis was correlated with the number of metastatic lymph nodes. The number of lymph nodes increased, and the prognosis of patients became worse. Radical hysterectomy combined with pelvic and abdominal aortic lymph node dissection is the routine treatment. There are 80% of patients with early-stage cervical cancer who are lymph node-negative. The efficacy of lymphadenectomy for these patients remains controversial. The incidence rate of LNM in patients with early-stage cervical cancer was reported as 7%-20% [15, 16]. The results of this study showed that the rate of para-aortic LNM in patients with early-stage cervical cancer was 22.06%. This was like previous studies. Based on these data, it is estimated that nearly 80% of patients with early cervical cancer undergo unnecessary lymph node dissection. Patients with lymph node metastasis suffer the risk of surgery and postoperative complications and suffer the adverse effects of postoperative adjuvant chemoradiotherapy. It is of great clinical value to identify lymph node status as early as possible and provide active treatment for this population.

Previous studies on the risk of early cervical cancer lymph node metastasis were limited to

Table 4. Index importance of random forest model

Variable	0	1	Mean Decrease Accuracy	Mean Decrease Gini
Preoperative SCC-Ag	26.025	29.080	32.799	19.776
Tumor diameter	4.327	11.977	11.251	4.906
Clinical stages	2.948	12.614	11.023	3.223
Cancer thrombus	5.926	7.768	9.048	3.333
Pelvic lymph node metastasis	0.865	9171	7.024	2.167
Cervical erosion	3.824	5.631	6.713	2.329
Invasion uterine	4.175	5.474	6.498	2.036
Depth of muscularis invasion	3.970	-1.870	2.424	1.068

Note: 0 represents non-lymph node metastasis and 1 represents lymph node metastasis.

Table 5. Confusion matrix of the logistic regression model for training set and validation set predictions

Logistic regression model predicts		Observed	
		Positive	Negative
Actual (training set)	Positive	25	5
	Negative	4	112
Actual (validation set)	Positive	10	5
	Negative	4	50

Table 6. Confusion matrix of the random forest model for the training set and validation set predictions

Random forest models predict		Observed	
		Positive	Negative
Actual (training set)	Positive	27	3
	Negative	1	115
Actual (validation set)	Positive	8	7
	Negative	5	49

Table 7. Comparison of prediction performances between the two models

	Prediction performances	Logistic regression	Random forest
Training set	Accuracy	0.938	0.973
	Sensitivity	0.966	0.991
	Specificity	0.833	0.900
	Recall	0.966	0.991
	Precision	0.958	0.975
	F1 Value	0.961	0.983
	AUC	0.974	0.998
Validation set	Accuracy	0.870	0.826
	Sensitivity	0.926	0.907
	Specificity	0.667	0.533
	Recall	0.926	0.907
	Precision	0.909	0.875
	F1 Value	0.917	0.890
	AUC	0.914	0.883

the exploration of risk factors, examination methods, and diagnosis [2, 3, 7, 17-19]. There are few studies on prediction models based on risk factors [20]. In this study, the LR and RF model were developed by combining patients' general information, etiology, comorbidities, and laboratory findings. The predictive efficacy of the two models for early cervical cancer para-aortic lymph node metastasis was compared. The results showed that the AUC of the LR and RF were 0.914 and 0.883, respectively. The differences between them were not significant. The two models performed well in terms of sensitivity, specificity, and accuracy. The differences were not significant. This suggested that the two models could help clinicians predict the risk of para-aortic lymph node metastasis in early cervical cancer. The RF model can be used as a supplement to the LR.

The results of the importance analysis of the characteristic variables of the RF showed that the top 5 variables were preoperative SCC-Ag level, tumor diameter > 4 cm, advanced clinical stage, cancer thrombus, and pelvic lymph node metastasis. LR analysis showed that tumor diameter > 4 cm, choroidal aneurysm embolism, pelvic lymph node metastasis, and increased preoperative SCC-Ag level were risk factors for para-aortic LNM in patients with early cervical cancer. The results of the two models were similar. This indicated that the prediction results were stable and reliable. Park et al. [21] showed that with the increase in tumor diameter, the probability of LNM gradually increased. The study showed that tumor

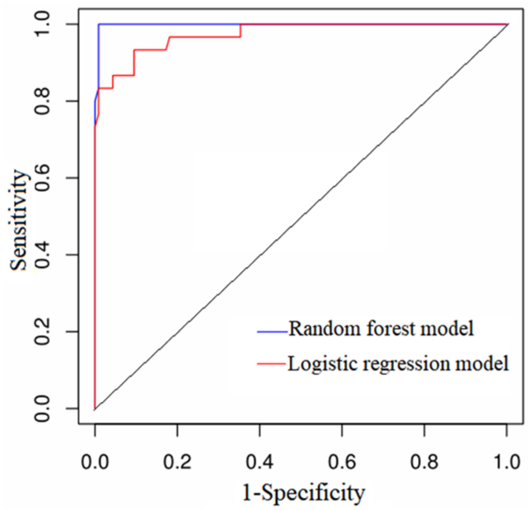

Figure 3. ROC curves of two model training set.

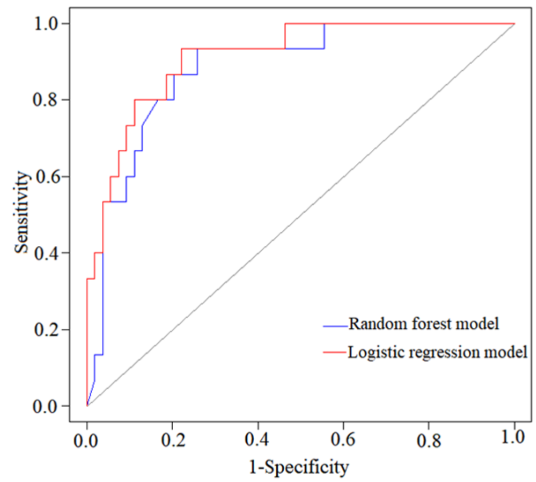

Figure 4. ROC curves of two model validation sets.

size was positively correlated with lymph node metastasis, with the lymph node metastasis rate of 6.0% in patients with tumor diameter ≤ 2 cm, 18.4% in patients with tumor diameter 2-4 cm, and 36.6% in patients with tumor diameter 4-6 cm. In this study, when the tumor size was limited to 4 cm, the lymph node metastasis rate of patients with tumor diameter larger than 4 cm was 82.22%. The multivariate analysis showed that tumor diameter greater than 4 cm was an independent risk factor for para-aortic lymph node metastasis of early cervical cancer (*OR* = 11.285, 95% CI: 3.030-42.026). The results were like those of Song J et al. [22]. Vascular cancer thrombus is the presence of malignant tumor cells between two layers of vascular endothelium [23]. After the tumor cells invade the vascular system of the interstitial cervical space, the tumor clots drain lymphatically to the regional lymph nodes. They proceed to the next level of lymph node groups or travel directly to the presacral lymph nodes or para-aortic lymph nodes in the form of "jumping metastases" and spread to other tissues and organs with the vascular circulation. Yanaranop M et al. [24] reported tumor diameter > 2 cm and vascular cancer thrombus as risk factors associated with LNM in stage IB1-IIA cervical adenocarcinoma. They suggested that surgery can be narrowed when the tumor was < 2 cm and there was no vascular cancer thrombus. The results of this study are like the above conclusions. Some researchers believed that high SCC-Ag was associated with para-aortic LNM of cervical cancer [25, 26]. A meta-analysis based on 17 studies and 3985 patients [27] showed that SCC-Ag levels had a sensitivity of 70.0%, specificity of 63.0%, and AUC of 0.73 for predicting early cervical cancer LNM. Van de Lande et al. [26] found that the accuracy of SCC-Ag > 1.65 ng/ml for the diagnosis of LNM in patients with stage IB1 and stage IB2 + IIA cervical cancer was 76% and 53%, respectively. This study was expressed as mean values. The mean preoperative SCC-Ag value was 2.43 μg/L in patients with positive lymph nodes in early cervical cancer compared to 1.74 μg/L in patients with negative lymph nodes. This was higher than the level reported by Van de Lande et al. [26]. Other studies have found that patients with positive pelvic lymph node status with SCC-Ag > 3.0 ng/mL had a higher risk (OR = 14.0). The above results suggested that preoperative SCC-Ag can be used as a good parameter to predict the risk of para-aortic lymph node metastasis. More studies are needed regarding the SCC-Ag threshold.

Conclusion

In this study, machine learning algorithm was used to construct the lymph node prediction model for early cervical cancer. The results and model obtained have some clinical references and values. Based on tumor diameter, choroidal aneurysm embolism, pelvic lymph node metastasis, and SCC-Ag factors, the AUC, sensitivity, specificity, and accuracy of LR model prediction were 0.914, 92.6%, 66.7%, 87.0%, respectively. The AUC, sensitivity, specificity, and accuracy of RF model prediction were

0.883, 90.7%, 53.3%, and 82.6%, respectively. Both LR and RF models based on machine learning algorithm have great predictive value in predicting para-aortic lymph node metastasis of early cervical cancer. There are some shortcomings: (1) Due to time , all the study subjects in this study were from the same hospital. This did not fully represent all early cervical cancer patients in China. (2) The selection of possible influencing factors of para-aortic LNM in this study was based on literature summary, empirical judgment, and expert consultation. Some important factors may have been missed. This needs to be improved in the future. (3) There was no external validation in this study. The accuracy of the prediction model screening can be verified in the future by screening out high-risk groups through the prediction model.

Disclosure of conflict of interest

None.

Address correspondence to: Yugang Meng and Jiangtao Fan, Department of Gynecology Division, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, Guangxi, PR China. Tel: +86-13407732997; E-mail: yorkmengyg@163.com (YGM); Tel: +86-13737085001; E-mail: jiangtao_fan1969@163.com (JTF)

References

[1] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015; 65: 87-108.
[2] Chen Y, Fang C, Zhang K, Deng Q and Zhang P. Distribution patterns of lymph node metastasis in early stage invasive cervical cancer. Medicine (Baltimore) 2020; 99: e22285.
[3] Jeong SY, Park H, Kim MS, Kang JH, Paik ES, Lee YY, Kim TJ, Lee JW, Kim BG, Bae DS and Choi CH. Pretreatment lymph node metastasis as a prognostic significance in cervical cancer: comparison between disease status. Cancer Res Treat 2020; 52: 516-523.
[4] Ikeda M, Shida M, Shigeta S, Nagase S, Takahashi F, Yamagami W, Katabuchi H, Yaegashi N, Aoki D and Mikami M. The trend and outcome of postsurgical therapy for high-risk early-stage cervical cancer with lymph node metastasis in Japan: a report from the Japan Society of Gynecologic Oncology (JSGO) guidelines evaluation committee. J Gynecol Oncol 2021; 32: e44.
[5] Chen SW, Liang JA, Yang SN and Lin FJ. Early stage cervical cancer with negative pelvic lymph nodes: pattern of failure and complication following radical hysterectomy and adjuvant radiotherapy. Eur J Gynaecol Oncol 2004; 25: 81-86.
[6] Lapuz C, Kondalsamy-Chennakesavan S, Bernshaw D, Khaw P and Narayan K. Stage IB cervix cancer with nodal involvement treated with primary surgery or primary radiotherapy: patterns of failure and outcomes in a contemporary population. J Med Imaging Radiat Oncol 2016; 60: 274-282.
[7] Wu C, Li L, Xiao X, Sun A, Lin W and Li A. Risk factors of regional lymph node metastasis in patients with cervical cancer. Open Life Sci 2019; 14: 208-213.
[8] Islam MM, Wu CC, Poly TN, Yang HC and Li YJ. Applications of machine learning in fatty live disease prediction. Stud Health Technol Inform 2018; 247: 166-170.
[9] Liu H, Li J, Leng J, Wang H, Liu J, Li W, Liu H, Wang S, Ma J, Chan JC, Yu Z, Hu G, Li C and Yang X. Machine learning risk score for prediction of gestational diabetes in early pregnancy in Tianjin, China. Diabetes Metab Res Rev 2021; 37: e3397.
[10] National Health Commission of The People's Republic of China. Chinese guidelines for diagnosis and treatment of cervical cancer 2018 (English version). Chin J Cancer Res 2019; 31: 295-305.
[11] Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. Int J Gynaecol Obstet 2009; 105: 103-104.
[12] Quinn MA, Benedet JL, Odicino F, Maisonneuve P, Beller U, Creasman WT, Heintz AP, Ngan HY and Pecorelli S. Carcinoma of the cervix uteri. FIGO 26th annual report on the results of treatment in gynecological cancer. Int J Gynaecol Obstet 2006; 95 Suppl 1: S43-103.
[13] Xu M, Xie X, Cai L, Xie Y, Gao Q and Sun P. Risk factor assessment of lymph node metastasis in patients with FIGO stage IB1 cervical cancer. Front Oncol 2022; 12: 809159.
[14] Hosaka M, Watari H, Mitamura T, Konno Y, Odagiri T, Kato T, Takeda M and Sakuragi N. Survival and prognosticators of node-positive cervical cancer patients treated with radical hysterectomy and systematic lymphadenectomy. Int J Clin Oncol 2011; 16: 33-38.
[15] Wang Y, Yao T, Yu J, Li J, Chen Q and Lin Z. Can pelvic lymphadenectomy be omitted in patients with stage IA2, IB1, and IIA1 squamous cell cervical cancer? Springerplus 2016; 5: 1262.
[16] Tanaka T, Sasaki S, Tsuchihashi H, Terai Y, Yamamoto K, Yamada T and Ohmichi M. Which is better for predicting pelvic lymph node metas-

tases in patients with cervical cancer: fluorodeoxyglucose-positron emission tomography/computed tomography or a sentinel node biopsy? A retrospective observational study. Medicine (Baltimore) 2018; 97: e0410.
[17] Huang BX and Fang F. Progress in the study of lymph node metastasis in early-stage cervical cancer. Curr Med Sci 2018; 38: 567-574.
[18] Lu Y, Wei JY, Yao DS, Pan ZM and Yao Y. Application of carbon nanoparticles in laparoscopic sentinel lymph node detection in patients with early-stage cervical cancer. PLoS One 2017; 12: e0183834.
[19] Tax C, Rovers MM, de Graaf C, Zusterzeel PL and Bekkers RL. The sentinel node procedure in early stage cervical cancer, taking the next step; a diagnostic review. Gynecol Oncol 2015; 139: 559-567.
[20] Kim DY, Shim SH, Kim SO, Lee SW, Park JY, Suh DS, Kim JH, Kim YM, Kim YT and Nam JH. Preoperative nomogram for the identification of lymph node metastasis in early cervical cancer. Br J Cancer 2014; 110: 34-41.
[21] Park JY, Kim DY, Kim JH, Kim YM, Kim YT and Nam JH. Outcomes after radical hysterectomy according to tumor size divided by 2-cm interval in patients with early cervical cancer. Ann Oncol 2011; 22: 59-67.
[22] Song J, Hu Q, Huang J, Ma Z and Chen T. Combining tumor size and diffusion-weighted imaging to diagnose normal-sized metastatic pelvic lymph nodes in cervical cancers. Acta Radiol 2019; 60: 388-395.
[23] Kodama J, Kusumoto T, Nakamura K, Seki N, Hongo A and Hiramatsu Y. Factors associated with parametrial involvement in stage IB1 cervical cancer and identification of patients suitable for less radical surgery. Gynecol Oncol 2011; 122: 491-494.
[24] Yanaranop M, Sathapornteera N and Nakrangsee S. Risk factors of pelvic lymph node metastasis in cervical adenocarcinoma following radical hysterectomy and pelvic lymphadenectomy. J Med Assoc Thai 2014; 97 Suppl 11: S87-95.
[25] Sun JR, Zhang YN, Sun XM, Feng SY and Yan M. Prediction model of pelvic lymph node metastasis in early stage cervical cancer and its clinical value. Minerva Chir 2011; 66: 537-545.
[26] van de Lande J, Davelaar EM, von Mensdorff-Pouilly S, Water TJ, Berkhof J, van Baal WM, Kenemans P and Verheijen RH. SCC-Ag, lymph node metastases and sentinel node procedure in early stage squamous cell cervical cancer. Gynecol Oncol 2009; 112: 119-125.
[27] Zhou Z, Li W, Zhang F and Hu K. The value of squamous cell carcinoma antigen (SCCa) to determine the lymph nodal metastasis in cervical cancer: a meta-analysis and literature review. PLoS One 2017; 12: e0186165.