

Establishment of Risk Prediction Model and Nomogram for Lymph Node Metastasis of Cervical Cancer: Based on SEER Database

Sufei Wang*, Shiwei Li*, Yong Chen#, Ya Zhang†

Obstetrics and Gynecology, The First Clinical Medical College of Yangtze University, Jingzhou First People's Hospital, Jingzhou, China

Email: †43547235@qq.com

How to cite this paper: Wang, S.F., Li, S.W., Chen, Y. and Zhang, Y. (2023) Establishment of Risk Prediction Model and Nomogram for Lymph Node Metastasis of Cervical Cancer: Based on SEER Database. *Yangtze Medicine*, 7, 105-115.

<https://doi.org/10.4236/ym.2023.72011>

Received: April 23, 2023

Accepted: June 26, 2023

Published: June 29, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Objective: To predict the risk factors of lymph node metastasis in cervical cancer by using large sample clinical data, and to construct and verify the nomogram for predicting lymph node metastasis. **Methods:** A total of 5940 patients with cervical cancer from 2004 to 2015 in the National Cancer Institute Surveillance Epidemiology and End Results database were retrospectively screened and randomly assigned to training group (n = 4172) and validation group (n = 1768). Multivariate Logistic regression analysis was used, and the optimal model was selected according to AIC or BIC and likelihood ratio test, and a nomogram was drawn. The accuracy and robustness of the prediction model were evaluated in three aspects: discrimination, calibration and clinical net benefit. **Results:** The prediction model based on race, tumor tissue differentiation degree, tumor histopathological type, distant metastasis of tumor, tumor diameter and other risk factors was successfully established and a nomogram was constructed. The AUCs of training group and validation group were: 0.736 and 0.714, respectively. And the p-values of the Hosmer-Lemeshow test were 0.28 and 0.11, respectively. The calibration curve was in good agreement with the ideal curve. It had high accuracy and applicability after internal verification. **Conclusion:** A prediction model is constructed based on the risk factors of lymph node metastasis of cervical cancer. The nomogram has a good effective prediction and can provide a theoretical basis for clinicians to assess the disease quickly before surgery.

Keywords

Cervical Cancer Lymph Node Metastasis, SEER Database, Logistic Regression, Nomogram

*First author.

#Second author.

†Corresponding author.

1. Introduction

A common malignant tumor, cervical cancer is also the second malignant tumor of the female reproductive system in Chinese women [1]. Cervical cancer is the fourth malignant tumor with the highest morbidity and mortality worldwide. At present, the treatment of cervical cancer is mainly operation and radiotherapy, while surgery is the main comprehensive treatment for early-stage cervical cancer. And individualized treatment plan can be made according to the relevant risk factors after operation [2]. The main high-risk factors for postoperative recurrence of cervical cancer include positive lymph nodes, para-uterine infiltration, positive resection margins and so on. Lymph node metastasis can occur in the early-stage cervical cancer, the incidence is about 10% - 20%, and once lymphatic metastasis, the 5-year survival rate of patients can be reduced by 20% - 45% [3]. Therefore, lymph node metastasis is an important factor affecting the treatment and prognosis of patients.

In recent years, nomogram is an important tool to evaluate and guide treatment in clinical prediction models, and it is widely used in clinical practice. Nomogram can be constructed by predicting high-risk factors, which can visualize and quantify the occurrence and prognosis of various diseases. At present, there is no risk assessment method for lymph node metastasis of cervical cancer at home and abroad. The purpose of this study is to retrieve a large number of cervical cancer data from SEER database, to explore the related influencing factors of cervical cancer patients with lymph node metastasis, and to establish a prediction model to provide an effective predictive tool for identifying high-risk patients, further scientific guidance for follow-up treatment, and strict grasp of surgical indications, so as to propose individualized and more optimized treatment plans.

2. Objects and Methods

2.1. Research Object

The data of 5941 cases included in this study were extracted from the SEER database (<https://seer.cancer.gov/>) using SEER*STATA software, which stores cancer surveillance data from various parts of the United States, covers approximately 28% of the population of the United States, and provides data for population studies of various cancers [4]. The ICD code C73.9 was used to screen the case data of uterine cancer patients whose pathological code was 8510 from 2004 to 2015. Due to the openness of the SEER database, this study does not require ethical approval or informed consent. Inclusion criteria: The pathological diagnosis was cervical cancer and had complete pathological characteristic data. Exclusion criteria: complicated with other types of malignant tumors. The clinicopathological data are incomplete or unknown.

The pathological features included in this study include: ID, age, race, marital status, primary site of tumor, tumor diameter, tumor tissue differentiation degree, tumor histopathological type, T stage, N stage, M stage, SEER historic stage

A, AJCC Stage Group. The TNM staging of the patients was based on the American Joint Commission on Cancer (AJCC) tumor staging, 6th edition.

2.2. Statistical Methods

EXCEL, SPSS 26.0 and Stata15 software were used to analyze the data. The patients who were included in the study were cleaned by EXCEL, and the data with missing data were removed. In SPSS, the patients were randomly divided into training group and validation group according to the proportion of 7:3. Chi-square test was used to test the difference of clinicopathological features between these training groups and validation groups. The counting data were described by frequency, and the chi-square test was used to compare the two groups. The variables with $P < 0.05$ were included in multivariate Logistic regression analysis by univariate Logistic regression. Four clinical prediction models were obtained by direct input, forward method, backward method and stepwise method. The optimal model was selected according to AIC or BIC and likelihood ratio test, and the nomogram was drawn. The prediction model was evaluated by discrimination, calibration and clinical net benefit. The discrimination was evaluated by the area under (AUC) the receiver operating characteristic curve (ROC curve), calibration curve and Hosmer-Lemeshow goodness-of-fit test to evaluate the calibration, and decision curve analysis (DCA) to evaluate clinical effectiveness. Finally, the nomogram is constructed according to the selection of the optimal model. The difference was statistically significant ($P < 0.05$).

3. Results

3.1. Comparison of the Balance of Basic Clinical Characteristics between Training Set and Validation Set

Training set and validation set patients' lymph node status, age, race, marital status, primary site of tumor, tumor tissue differentiation degree, tumor histopathological type, distant metastasis of tumor, tumor diameter, SEER historic stage, AJCC Stage Group. There was no significant difference in 6th ed. ($P > 0.05$), see [Table 1](#).

3.2. Univariate and Multivariate Analysis of Lymphatic Positive Risk Prediction for Cervical Cancer

Univariate analysis showed that there were 7 suspected risk factors associated with positive lymph nodes of cervical cancer. They were race, marital status, primary site of tumor, tumor tissue differentiation degree, tumor histopathological type, distant metastasis of tumor, tumor diameter ($P < 0.1$, see [Table 2](#)). The influencing factors with statistical significance in univariate analysis were included in multivariate Logistic regression analysis. The results showed that race, tumor tissue differentiation degree, tumor histopathological type, distant metastasis of tumor, tumor diameter There were considered statistically significant ($P < 0.05$, see [Table 2](#)).

Table 1. Clinical information for training sets and validation sets.

Variable	Training set (n = 4172)	Validation set (n = 1768)	P	Variable	Training set (n = 4172)	Validation set (n = 1768)	P
Lymph node status				Tumor tissue differentiation			
Negative	3147	1339	0.80	Level 1	610	232	0.18
Positive	1025	429		Level 2	1890	806	
Age				Level 3	1564	670	
<40	1161	498	0.94	Level 4	108	60	0.98
40 - 60	2069	877		Distant metastasis of tumor			
≥60	942	393		No distant metastasis	3801	1605	0.68
Race				Distant metastasis	371	163	
Whites	3091	1301	0.62	Tumor diameter			
Blacks	510	232		≤6	3420	1456	0.73
Native American	63	27		>6	752	312	
Asian	491	195		SEER stage			
Unidentified race	17	11		Regional	1632	705	0.94
Marital status				Localized	2129	889	
Married	1938	828	0.89	Distant	405	171	
Unmarried and live apart	1118	466		AJCC Stage			
Unmarried	613	268		I	2218	928	0.98
Widowed	323	138		II	568	251	
Unknown	182	68		III	943	402	
Tumor histological type				IV	437	184	
Cervical squamous cell carcinoma	2692	1138	0.88	unstage	6	3	
Squamous adenocarcinoma	161	71		Primary site			
Adenocarcinoma	1206	517		Endocervix	901	405	0.52
Others	113	42		Cervix uteri	3111	1302	
Primary site				Exocervix	82	35	
Endocervix	901	405	0.52	Overlapping lesion of cervix uteri	78	25	
Cervix uteri	3111	1302					
Exocervix	82	35					
Overlapping lesion of cervix uteri	78	25					

Table 2. Univariate analysis and multivariate Logistic analysis of cervical cancer lymph node metastasis.

Features	Univariate Logistic analysis					Multivariate Logistic analysis				
	B	OR	95% CI Lower	95% CI Upper	P	B	OR	95% CI Lower	95% CI Upper	P
Age					0.21					
<40			reference					NA		
40 - 60	0.11	1.12	0.97	1.29	0.11					
≥60	0.14	1.15	0.99	1.36	0.76					
Race					0.04					0.01*
Whites			reference					reference		
Blacks	1.43	4.17	0.99	17.62	0.05	1.16	3.17	0.73	13.77	0.12
American Indians/Alaska Native	1.44	4.23	0.99	17.97	0.05	0.88	2.42	0.55	10.61	0.24
Asian	1.97	7.17	1.6	32.2	0.01	1.65	5.25	1.12	24.52	0.03
Marital status					0.06					
Married			reference							
Unmarried	0.06	1.06	0.77	1.44	0.73					
Divorce and diaspora	0.25	1.28	0.93	1.77	0.13			NA		
Widowed	0.2	1.22	0.88	1.71	0.24					
Unknown	0.19	1.21	0.84	1.75	0.3					
Tumor histological type					<0.05					<0.05*
Cervical squamous cell carcinoma			reference					reference		
Squamous adenocarcinoma	0.23	1.26	0.94	1.68	0.12	-0.12	0.88	0.60	1.29	0.51
Adenocarcinoma	-0.53	0.59	0.51	0.68	0.00	0.09	1.09	0.68	1.77	0.70
Others	0.91	2.49	1.80	3.44	0.00	-0.48	0.61	0.41	0.91	0.02
The primary site of the tumor					0					0.72
Endocervix			reference					reference		
Cervix uteri	-0.46	0.63	0.54	0.74	0.00	-0.16	0.83	0.44	1.59	0.58
Exocervix	-0.68	0.51	0.31	0.84	0.01	0.48	1.13	0.62	2.07	0.69
Overlapping lesion of cervix uteri	0.08	1.08	0.70	1.66	0.73	0.10	1.10	0.87	1.38	0.43
Tumor tissue differentiation					<0.05					<0.05*
Level 1			reference					reference		
Level 2	1.30	3.67	2.78	4.80	0.00	1.61	2.98	2.23	3.98	0.00
Level 3	1.82	6.16	4.68	8.12	0.00	1.41	4.09	3.05	5.49	0.00
Level 4	2.14	8.49	5.67	12.72	0.00	1.80	6.04	3.89	9.39	0.00
Distant metastasis	2.27	9.65	7.92	11.76	0.00	2.02	7.54	6.11	9.29	0.00*
Tumor diameter	1.25	3.48	3.03	4.00	0.00	0.82	2.28	1.95	2.66	0.00*

3.3. Building a Prediction Model

In SPSS and STATA software, the treatment mode of patients with positive

lymph nodes of cervical cancer was taken as dependent variables (assigned: lymph node negative = 0, lymph node positive = 1). Seven variables selected from univariate logistic regression were included in multivariate Logistic regression analysis. Four clinical prediction models were constructed by forward method, backward method and stepwise method. The AIC were: 4035.367/4029.592/4038.9/4034.221. The BIC were: 4105.065/4073.945/4078.574/4070.581. According to the selection of the minimum AIC or BIC value and the comparison of likelihood ratio under the same AIC value, the optimal model was selected, and drew the nomogram according to the prediction variables, namely Nomogram (Figure 1). The corresponding values of each variable could be scored by the nomogram, and then the total score could be obtained by adding the scores of all variables, and a vertical line could be drawn down according to the total score, the estimated probability of lymph node metastasis risk of cervical cancer could be marked.

3.4. Verification of Prediction Model

The verification of the prediction model was mainly based on the discrimination and calibration of the model. The model discrimination was evaluated by drawing a prediction model to predict the ROC curve of cervical cancer lymph node metastasis risk. The AUC of the training set was 0.736 (95% CI (0.72)) (Figure 2(a)). The AUC of the verification set was 0.714 (95% CI (0.685, 0.742)) (Figure 2(b)). It showed that the prediction model had good discriminant ability. At the

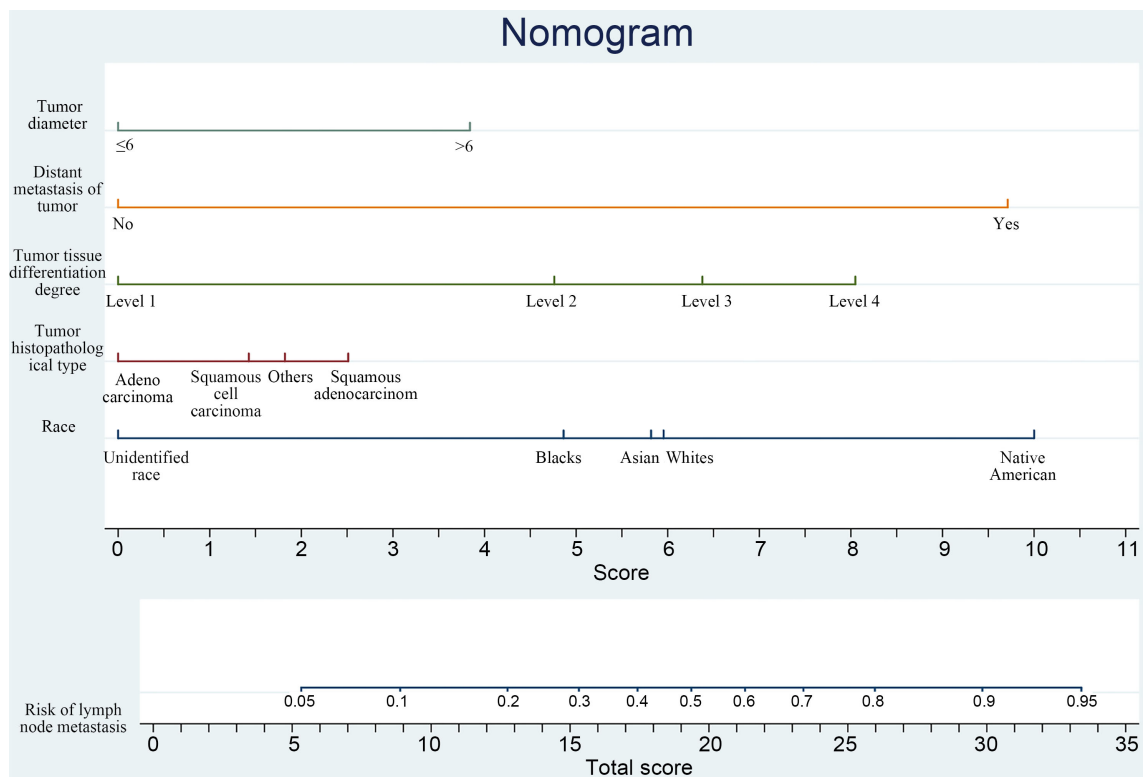


Figure 1. Nomogram.

same time, the Hosmer Lemeshow goodness-of-fit test showed a good degree of fit (training set 0.28; verification queue 0.11), which showed that the prediction probability of the model was basically consistent with the actual probability, and the model had a good calibration. In addition, the calibration curves of the training set and the verification set showed moderate consistency, and the correction ability of the prediction model was good (Figure 3 calibration curve). To sum up, the Nomogram of the prediction model had medium prediction ability.

3.5. Clinical Application

The clinical validity of the prediction model was evaluated by DCA. The DCA of the occurrence probability nomogram of cervical cancer lymph node metastasis was shown in Figure 4. The results showed that if the threshold probability of the patient and the doctor are more than 20% respectively, in the current study, the risk of positive lymph nodes in patients with cervical cancer using this nomogram will be more beneficial than that of all patients implementing intervention programs. Within this range, the net benefit of the prediction model was significantly higher than that of the two extremes. All patients have received clinical intervention.

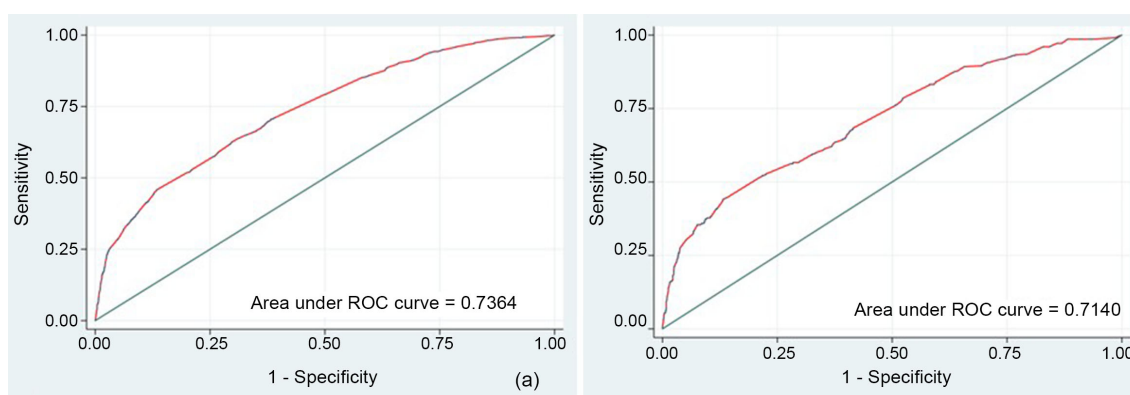


Figure 2. ROC curve classification of two sets of nomograms.

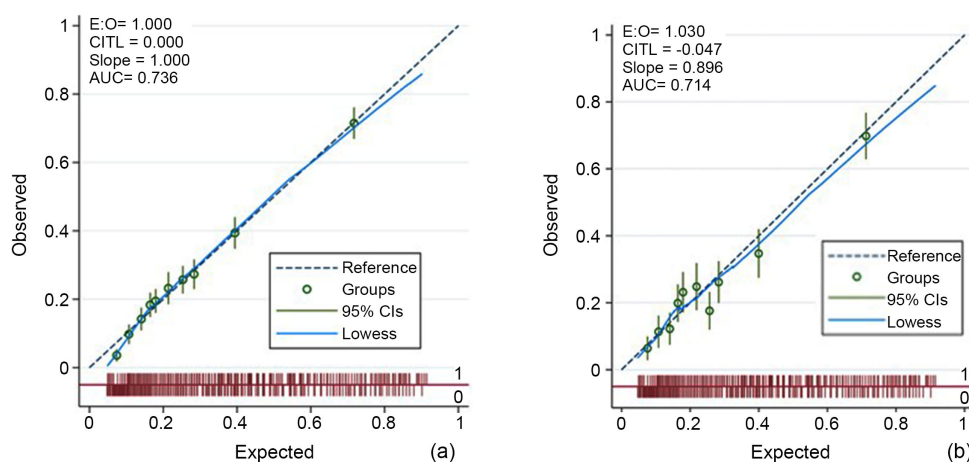


Figure 3. Calibration curves of two sets of nomograms.

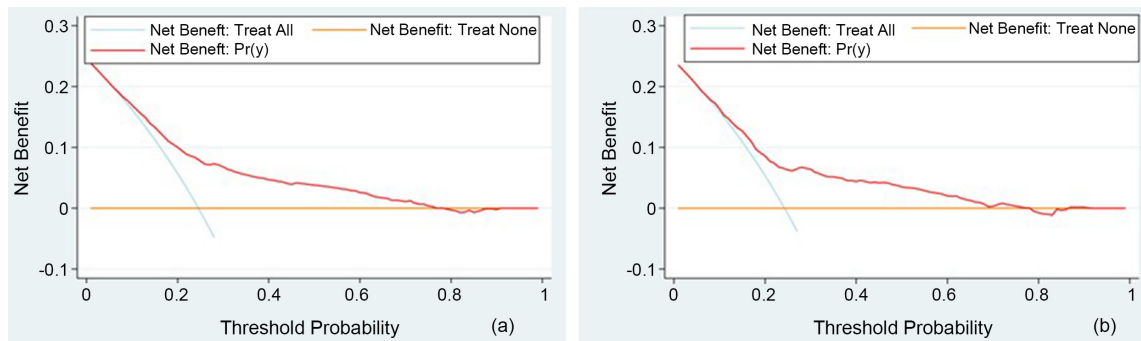


Figure 4. Decision curve analysis of two groups of nomograms.

4. Discussion

Cervical cancer is one of the four most common female malignant tumors in the world. It ranked second in cancer-related deaths among young women between the ages of 20 and 39 in 2020. The global incidence of cervical cancer is about 500,000 cases every year. More than one-fourth of all new cases and fatalities worldwide occur in China. With the exception of cervical and uterine body cancers, survival rates for all of the most prevalent cancers have increased since the middle of the 1970s. As we all know, the main way of metastasis of cervical cancer is through direct spread and lymph node metastasis, and many studies have shown that lymph node metastasis is an independent risk factor to evaluate the prognosis of cervical cancer. There were studies have shown that patients with pathological risk factors (lymph node metastasis, para-uterine infiltration, positive resection margins of vaginal stump) have a higher recurrence rate [5]. Moreover, lymph node status has been included in the staging criteria in the 2018 FIGO guidelines. There is currently controversy over the scope of cervical cancer surgery. The status of lymph nodes has guiding significance for the scope of surgery and postoperative adjuvant therapy. Preoperative evaluation of lymph node status in clinical work is very important for patients' treatment. Accurate preoperative assessment of lymph node status can reduce unnecessary lymph node dissection and reduce operative complications caused by lymph node dissection, such as pelvic vascular injury, Lymphocyst, Chylous fistula and so on. The examination methods for evaluating lymph node metastasis of cervical cancer in clinical practice mainly include imaging examination: MRI, PET-CT, PET-MRI, etc. [6] but its sensitivity is low. Studies by Sato [7] and others showed that CA125 was valuable in evaluating preoperative lymph node metastasis. In addition, SLNB (sentinel lymph node biopsy) is considered to be the most positive rate for lymph node metastasis. However, it is only applicable to cervical tumors with a diameter of less than 2cm, which cannot be widely carried out in clinical practice due to its invasive examination [6]. Currently, there is no quantitative index for comprehensive judgment of cervical cancer lymph node metastasis. Valuable high-risk factors of cervical cancer lymph node metastasis are calculated by statistics, and an effective prediction model is established. It can help clinicians identify high-risk patients and guide individualized treatment.

The SEER database (Surveillance by the National Cancer Institute/Epidemiology and final Database) collects a large number of patient information data, including demographics/disease diagnosis/tumor staging/treatment information and prognosis, which can provide a large amount of systematic data for clinicians. This study used a large population provided by SEER database and constructed and validated a nomogram for predicting lymph node metastasis of cervical cancer.

In this study, cervical cancer cases from 2004 to 2015 were screened by SEER database. Five risk factors for cervical cancer lymph node metastasis were calculated by statistical method: race, tumor tissue differentiation degree, tumor histopathological type, distant metastasis of tumor, tumor diameter. And clinical prediction model was constructed ($P < 0.1$, see **Table 2**). Similar studies have shown that the 5-year survival rate of cervical cancer patients younger than 35 years is lower than average. Moreover, the cancerous lesion is large, the recurrence rate is high, and the prognosis is poor. [8] Some similar studies have suggested that cervical cancer lymph node metastasis is associated with age. It is different from the results of this study, which may be due to the different number of cases and selection bias. Studies by Zhuang Jinman [9] and others showed that no cervical erosion, tumor maximum diameter > 3 cm, Para-uterine infiltration and lymphatic vascular interstitial infiltration were the high-risk factors affecting lymph node metastasis in early-stage cervical cancer. There are some similar predictive factors to this study. A retrospective study of 302 cases of cervical cancer by Yi-Fang Dai [10] *et al.* showed that there was no significant correlation between lymph node metastasis rate and tumor size (>4 cm), but it was related to tumor differentiation, depth of uterine myometrium infiltration, lymphatic vascular infiltration and other factors. However, the study of Zhuang *et al.* showed that the risk of lymph node metastasis of tumor maximum diameter > 3 cm was 1.98 times higher than that of tumor maximum diameter ≤ 3 cm. That is, the larger the maximum diameter of the tumor, the greater the possibility of lymph node metastasis. Zhuang *et al.*'s study included three factors: maximum diameter of tumor, para-uterine infiltration and infiltration of lymphatic vascular space to construct a nomogram to predict pelvic lymph node metastasis. In this study, a more detailed prediction model was constructed based on a large sample of clinical data, and the results showed that the model has better prediction ability. Finally, the results of decision curve analysis showed that our model is also of guiding significance for clinicians to make clinical decisions.

Although the prediction ability of this study is good, there are still the following shortcomings. 1) This study is a retrospective study, excluding a large number of incomplete cases, which may lead to selective bias. Therefore, more large samples of prospective studies are needed for further verification. 2) This study is based on the SEER database, there is no vascular invasion/nerve tissue invasion and para-uterine invasion and other data, so there are fewer risk factors. Therefore, more risk factors should be included in the following research to fur-

ther improve the prediction ability of the model. 3) The samples included in this study cover early and late-stage cases, so the score of distant tumor metastasis in the model construction is relatively high.

5. Conclusion

In summary, this study identified the risk factors associated with lymphatic metastasis in patients with cervical cancer by analyzing a large number of data obtained from the SEER database. Finally, we constructed a model with high predictive performance based on the five best risk factors, and constructed a nomogram to help clinicians assess the risk of lymphatic metastasis. Through individual risk assessment, clinicians and patients can choose personalized treatment plans and take necessary intervention measures in advance to extend the survival time of patients.

Conflicts of Interest

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

References

- [1] Sung, H., Ferlay, J., Siegel, R.L., *et al.* (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, **71**, 209-249. <https://doi.org/10.3322/caac.21660>
- [2] Bhatla, N., Aoki, D., Sharma, D.N., *et al.* (2018) Cancer of the Cervix Uteri. *International Journal of Gynecology & Obstetrics*, **143**, 22-36. <https://doi.org/10.1002/ijgo.12611>
- [3] Liu, X., Wang, W., Hu, K., *et al.* (2020) A Risk Stratification for Patients with Cervical Cancer in Stage III C1 of the 2018 FIGO Staging System. *Scientific Reports*, **10**, 362. <https://doi.org/10.1038/s41598-019-57202-3>
- [4] Wang, X., Yu, G.Y., Chen, M., *et al.* (2018) Pattern of Distant Metastases in Primary Extrahepatic Bile Duct Cancer: A SEER Based Study. *Cancer Medicine*, **7**, 5006-5014. <https://doi.org/10.1002/cam4.1772>
- [5] Bedford, S. (2009) Cervical Cancer: Physiology, Risk Factors, Vaccination and Treatment. *British Journal of Nursing*, **18**, 80. <https://doi.org/10.12968/bjon.2009.18.2.37874>
- [6] Zhang, Y., Ma, X. and Chen, J.J. (2022) Advances in Preoperative Evaluation of Lymph Node Metastasis in Patient with Cervical Cancer. *Chinese Computer Medical Imaging*, **28**, 443-445.
- [7] Sato, K., Mizuuchi, H., Mori, Y., *et al.* (1994) Usefulness of CA125 Determination in the Diagnosis of Lymph Node Metastasis in Post Menopausal Uterine Endometrial Carcinoma. *Nihon Sanka Fujinka Gakkai Zasshi*, **46**, 331-336.
- [8] Jarruwale, P., Huang, K.G., Banavides, D.R., *et al.* (2012) Factors Related to Sentinel Node Identification in Cervical Cancer. *Gynecology and Minimally*, **1**, 19-22. <https://doi.org/10.1016/j.gmit.2012.08.001>
- [9] Zhuang, J.M., Lu, W.T., Huang, Y.X., *et al.* (2019) Risk Factors of Lymph Node

Metastasis in Early-Stage Cervical Cancer Patients and Build of a Nomogram Prediction Model. *Cancer Research on Prevention and Treatment*, **46**, 5.

- [10] Dai, Y.F., Mu, X., Zhong, L.Y., *et al.* (2018) Prognostic Significance of Solitary Lymph Node Metastasis in Patients with Stages IA2 to IIA Cervical Carcinoma. *Journal of International Medical Research*, **46**.
<https://doi.org/10.1177/0300060518785827>