

Development of a Predictive Model for Lymph Node Metastasis in Endometrial Cancer Based on Integrated Clinical Characteristics, Serum Tumor Biomarkers, and Related Factors

wei hao zhang

Xinjiang Medical University

ya peng huang

Xinjiang Medical University

jun fei li

Xinjiang Medical University

hui tan

Xinjiang Medical University

hui min jia

XXXXXXXXXXXX

yan wang

x.jwangyan2012@163.com

XXXXXXXXXXXX

Research Article

Keywords: endometrial cancer, lymph node metastasis, CA125, HE4, CA199, model prediction

Posted Date: August 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10554485/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Objective

To investigate the expression levels of serum tumor biomarkers, including carbohydrate antigen 125 (CA125), human epididymis protein 4 (HE4), and carbohydrate antigen 19 – 9 (CA199), as well as other related clinical characteristics in patients with endometrial cancer (EC), and to explore their associations with lymph node metastasis (LNM). Furthermore, a predictive model for LNM in EC was developed based on machine learning algorithms by integrating clinical characteristics and serum tumor biomarkers, aiming to provide support for subsequent clinical decision-making.

Methods

Clinical data from 1,099 patients diagnosed with endometrial cancer at the Affiliated Cancer Hospital of Xinjiang Medical University between 2021 and 2026 were retrospectively collected. Univariate analysis was performed to identify clinical characteristics and serum tumor biomarkers significantly associated with LNM. The selected variables were subsequently incorporated into three machine learning models, including logistic regression (LR), random forest (RF), and extreme gradient boosting (XGBoost), for model development. The dataset was randomly divided into training and validation sets at a ratio of 7:3. The performance of each model was evaluated using the area under the receiver operating characteristic curve (AUC), and the importance of predictive indicators was further analyzed.

Results

Univariate analysis demonstrated that clinical stage and serum levels of CA125, HE4, and CA199 were significantly associated with LNM in patients with EC ($P < 0.05$). Among the three machine learning models, XGBoost showed the best predictive performance, with an AUC of 0.936. Further analysis indicated that FIGO tumor stage had the highest predictive value among individual indicators (AUC = 0.900). Among serum tumor biomarkers, CA125 exhibited superior predictive ability (AUC = 0.703) compared with HE4 (AUC = 0.658) and CA199 (AUC = 0.605). In addition, the combined detection model incorporating the three tumor biomarkers achieved an AUC of 0.730, demonstrating improved predictive performance compared with individual biomarker assessment. The optimal cutoff values for CA125, HE4, and CA199 were also determined, providing quantitative references for clinical risk stratification and individualized surgical decision-making.

Conclusion

The XGBoost model integrating clinical stage and serum tumor biomarkers, including CA125, HE4, and CA199, demonstrated effective predictive performance for LNM in patients with EC. Among the

evaluated biomarkers, CA125 showed superior predictive ability compared with HE4 and CA199. Combined detection of multiple tumor biomarkers further improved the prediction of LNM. This model may provide valuable support for preoperative risk assessment and surgical decision-making in patients with endometrial cancer.

Introduction

Endometrial cancer (EC) is one of the three most common malignant tumors of the female reproductive system and accounts for approximately 20%–30% of malignant tumors of the female genital tract. In recent years, the incidence of EC has shown a continuous upward trend in China. According to the latest cancer statistics, there were approximately 77,700 newly diagnosed cases and 13,500 deaths due to EC in China in 2022 [1]. In some economically developed regions, the incidence of EC has surpassed that of cervical cancer, with an increasing tendency toward younger age at onset [2].

At the national level, increasing attention has been paid to cancer prevention and control strategies. The Healthy China Action—Implementation Plan for Cancer Prevention and Control (2023–2030) proposed expanding screening coverage for high-incidence cancers and promoting precise screening strategies for high-risk populations[3]. In addition, the publication of clinical consensus guidelines, such as the Chinese Expert Consensus on Screening and Prevention of Lynch Syndrome-Associated Endometrial Cancer (2023 Edition), has provided important guidance for standardized screening and clinical management of EC [4].

Lymph node metastasis (LNM) is one of the major metastatic pathways of EC and represents an important prognostic indicator associated with disease progression and clinical outcomes[5]. Accurate preoperative assessment of the risk of LNM is essential for determining surgical strategies and guiding decisions regarding adjuvant therapy. Serum tumor biomarkers have attracted increasing attention in the prediction of LNM due to their advantages of convenience, repeatability, and minimal invasiveness. Previous studies have demonstrated that serum human epididymis protein 4 (HE4) levels are closely associated with LNM in patients with EC. HE4 expression was significantly increased in patients with lymph node metastasis and may serve as an auxiliary indicator for evaluating metastatic risk [6]. Furthermore, the combination of pelvic magnetic resonance imaging with serum CA125 and HE4 detection has shown favorable diagnostic performance for predicting LNM in EC, suggesting that multi-parameter assessment may provide more reliable evidence for clinical lymph node management decisions [7].

However, previous studies have mainly focused on single-center cohorts, relatively small sample sizes, or individual biomarkers. Reports on predictive models integrating comprehensive clinical characteristics with multiple serum tumor biomarkers based on larger clinical cohorts remain limited. Moreover, existing studies have primarily emphasized the relationship between pathological characteristics and LNM, whereas the application of machine learning approaches integrating multiple clinical variables and biomarkers for preoperative prediction remains insufficiently explored. Therefore,

based on real-world clinical data, this study investigated the associations between serum CA125, HE4, CA199 levels, related clinical characteristics, and LNM in patients with EC. Furthermore, predictive models were developed by integrating clinical and biomarker characteristics using machine learning algorithms, aiming to provide an effective auxiliary tool for preoperative assessment of LNM risk and improve individualized clinical decision-making [8]

Materials and Methods

The clinical data analyzed in this study were obtained from the Affiliated Cancer Hospital of Xinjiang Medical University. Patients who met the eligibility criteria were retrospectively enrolled, and their clinical characteristics, pathological information, and serum tumor biomarker data were collected for analysis. The inclusion criteria were as follows: patients aged ≥ 18 years; patients diagnosed with endometrial cancer by postoperative pathological examination; patients who underwent primary surgical treatment, including extrafascial hysterectomy, modified radical hysterectomy, or radical hysterectomy combined with bilateral salpingo-oophorectomy and pelvic and/or para-aortic lymph node dissection. Patients with incomplete clinical data, unclear tumor staging information, preoperative radiotherapy or chemotherapy, or elevated tumor biomarkers caused by other diseases were excluded. After applying the inclusion and exclusion criteria, a total of 1,099 patients were included in the final analysis. The original data regarding age, height, and weight were further categorized for subsequent statistical analysis.

1.1 Inclusion and Exclusion Criteria

The initial dataset consisted of 2,612 patients diagnosed with endometrial cancer at the Affiliated Cancer Hospital of Xinjiang Medical University between 2021 and 2026. Clinical variables collected included age, disease diagnosis, pathological diagnosis, previous medical history, reproductive history, menstrual history, height, weight, tumor stage, and serum tumor biomarkers, including human epididymis protein 4 (HE4), carbohydrate antigen 199 (CA199), and carbohydrate antigen 125 (CA125) [5, 9, 10]

Patients were excluded if they met any of the following criteria: incomplete clinical information or missing laboratory data; abnormal values; unclear tumor staging; receipt of preoperative chemotherapy or radiotherapy; or elevated serum tumor biomarker levels potentially caused by other diseases, including hepatic cysts, breast cancer, lung cancer, and pancreatic cancer [11].

Finally, 1,099 patients who fulfilled the eligibility criteria were included in this study.

Age was categorized into five groups according to the following criteria:

1. 18–<35 years (assigned value: 1);
2. 35–<45 years (assigned value: 2);
3. 45–<55 years (assigned value: 3);
4. 55–<65 years (assigned value: 4);
5. ≥ 65 years (assigned value: 5).

Body mass index (BMI) was calculated based on height and weight. According to the BMI classification criteria established by the National Health Commission of China, patients were categorized as follows:

1. BMI <18.5 kg/m²: underweight (assigned value: 0);
2. 18.5≤BMI<24 kg/m²: normal weight (assigned value: 1);
3. 24≤BMI<28 kg/m²: overweight (assigned value: 2);
4. BMI≥28 kg/m²: obesity (assigned value: 3).

1.2 Descriptive Statistics

A retrospective analysis was performed among the 1,099 enrolled patients. Regarding reproductive history, 356 patients had no children or unavailable reproductive information, accounting for 32.3% of the study population. A total of 375 patients had one childbirth (34.1%), 219 patients had two childbirths (19.9%), and patients with three to nine childbirths accounted for 13.7%. Among all patients, 842 (76.6%) were premenopausal and 257 (23.3%) were postmenopausal. Based on BMI classification according to the criteria of the National Health Commission of China, 495 patients met the criteria for obesity, accounting for 45.0% of the cohort. Additionally, 115 patients had diabetes mellitus (10.4%), and 224 patients had hypertension (20.3%). According to the FIGO staging criteria for endometrial cancer, 852 patients were classified as stage I, 84 as stage II, 126 as stage III, and 37 as stage IV. Lymph node metastasis was determined based on postoperative pathological examination. Among the enrolled patients, 98 patients were confirmed to have lymph node metastasis, accounting for 9.8% of the total cohort [12]

2. Statistical Methods

Statistical analyses were performed using R software version 4.5.0, and Excel was used for data management. Continuous variables were initially tested for normality. Normally distributed variables were expressed as mean ± standard deviation ($\bar{x} \pm s$) and compared between groups using the independent-samples *t*-test, whereas non-normally distributed variables were presented as median (interquartile range, IQR) and analyzed using the Wilcoxon rank-sum test. Categorical variables were expressed as numbers (percentages) and compared using the Pearson χ^2 test or Fisher's exact test. Postoperative pathological results were used to determine the presence or absence of lymph node metastasis as the outcome variable, and variables with statistical significance in the univariate analysis were included in subsequent model analyses. Logistic regression (LR), random forest (RF), and extreme gradient boosting (XGBoost) models were applied to construct predictive models, with the dataset randomly divided into training and validation sets at a ratio of 7:3. Model performance was evaluated using receiver operating characteristic (ROC) curves, area under the curve (AUC), sensitivity, specificity, Youden index, and 95% confidence intervals. Calibration curves, Brier scores, and expected calibration error (ECE) were used to assess the consistency between predicted probabilities and actual outcomes. All statistical tests were two-sided, and a *P* value <0.05 was considered statistically significant.

Results

3.1 Univariate Analysis

The potential factors associated with lymph node metastasis in patients with endometrial cancer were analyzed using univariate analysis, and the results are presented in Table 1. The results showed that clinical stage, age, and the levels of tumor markers CA125, HE4, and CA199 were significantly different between the lymph node metastasis group and the non-metastasis group ($P < 0.05$). Specifically, patients with lymph node metastasis had a significantly higher proportion of stage III and IV disease compared with those without metastasis, and the levels of CA125, HE4, and CA199 were significantly elevated in the metastasis group. In contrast, menopausal status, BMI category, history of diabetes mellitus, history of hypertension, and number of childbirths showed no statistically significant differences between the two groups ($P > 0.05$). Based on the results of the univariate analysis, clinical stage, age, CA125, HE4, and CA199 were included for subsequent predictive model construction.

Table 1

Univariate analysis of factors associated with lymph node metastasis in 1,099 patients with endometrial cancer

Factor	Normality Test	Metastasis Group (%)n = 98	Non-metastasis Group(%)n = 1099	Statistical Test	Statistic	P-value
Menopause	-			Pearson Chi-square Test	0.23	0.632
Yes		21(21.43)	236(23.58)			
No		77(78.57)	765(76.42)			
BMI Group	-			Pearson Chi-square Test	4.13	0.248
BMI < 18.5		0(0)	13(1.3)			
18.5 ≤ BMI < 24		29(29.59)	222(22.18)			
24 ≤ BMI < 28		26(26.53)	314(31.37)			
BMI ≥ 28		43(43.88)	452(45.15)			
Diabetes	-			Pearson Chi-square Test	2.16	0.141
Yes		6(6.12)	109(10.89)			
no		92(93.88)	892(89.11)			
Hypertension	-			Pearson Chi-square Test	2.46	0.117
yes		14(14.29)	210(20.98)			
no		84(85.71)	791(79.02)			
Clinical Stage	-			Pearson Chi-square Test	589.73	< 0.0001
I		12(12.24)	840(83.92)			
II		6(6.12)	78(7.79)			
III		70(71.43)	56(5.59)			
IV		10(10.2)	27(2.7)			

Factor	Normality Test	Metastasis Group (%)n = 98	Non-metastasis Group(%)n = 1099	Statistical Test	Statistic	P-value
age	Non-normal	-	-	Wilcoxon Rank Sum Test	41096	0.008
Number of Children	Non-normal	-	-	Wilcoxon Rank Sum Test	50837	0.5338
CA125(U/ml)	Non-normal	-	-	Wilcoxon Rank Sum Test	29168.5	< 0.001
HE4(pmol/L)	Non-normal	-	-	Wilcoxon Rank Sum Test	33507.5	< 0.001
CA199(U/ml)	Non-normal	-	-	Wilcoxon Rank Sum Test	38707.5	< 0.001

3.2 Model Construction

3.2.1 Construction of the Optimal Prediction Model

After variable selection, the 1,099 patients were randomly divided into a training set and a validation set at a ratio of 7:3.

The training set consisted of 769 patients, while the validation set included 330 patients. Three commonly used machine learning algorithms were applied for model construction:

Logistic regression (LR);

Random forest (RF);

Extreme gradient boosting (XGBoost).

Receiver operating characteristic (ROC) curves were generated, and the area under the curve (AUC) was calculated to evaluate the predictive performance of each model. The model with the highest AUC value was considered the optimal predictive model. According to the results obtained using R software, the AUC values were:

LR: 0.898;

RF: 0.915;

XGBoost: 0.936.

Among the three models, XGBoost achieved the highest AUC value and was therefore selected as the optimal model. The ROC curves of the three machine learning models are shown in Fig. 1.

3.2.2 Construction of Prediction Models Based on Important Indicators

Based on the optimal model identified above, the four statistically significant variables obtained from univariate analysis were further incorporated into the XGBoost model to evaluate their individual predictive performance. The prediction models for each factor were established using R software. The AUC value, sensitivity, specificity, Youden index, and 95% confidence interval (CI) were calculated. The detailed results are presented in Table 2 and Fig. 2.

Table 2
Predictive performance of individual indicators for lymph node metastasis.

Predictive Factor	AUC Value	Sensitivity	Specificity	Youden's Index	95%CI
Clinical Stage	0.901	0.867	0.884	0.751	0.863 ~ 0.936
CA125(U/ml)	0.703	0.684	0.633	0.317	0.645 ~ 0.760
HE4(pmol/L)	0.658	0.653	0.599	0.252	0.603 ~ 0.714
CA199(U/ml)	0.605	0.480	0.765	0.245	0.536 ~ 0.674

According to the model fitting results, clinical stage was closely associated with the occurrence of lymph node metastasis in patients with endometrial cancer and represented the most important predictive factor among all evaluated variables. Among the included serum tumor biomarkers, CA125 achieved an AUC value of 0.703, indicating that CA125 was also an important predictive indicator and exhibited the highest predictive performance among the evaluated tumor biomarkers. The AUC values of the other two tumor biomarkers, HE4 and CA199, were both greater than 0.600, suggesting that they may also provide valuable information for clinical assessment and decision-making. Previous studies by Guo Sheng et al. and other researchers have demonstrated that combined detection of CA125 and HE4, as well as combined detection of CA125, HE4, and CA724, showed higher predictive AUC values and improved accuracy in predicting lymph node metastasis compared with individual tumor biomarkers alone [13, 14]. Therefore, in the present study, the combined detection of the three tumor biomarkers (CA125, HE4, and CA199) was further incorporated into the optimal predictive model. After model validation, the combined detection of the three tumor biomarkers included in this study achieved an AUC value of 0.73 in the optimal XGBoost model[15], which was higher than the AUC values obtained from the individual detection of each tumor biomarker. The detailed results are presented in Fig. 3 and Table 3.

Table 3
Predictive performance of combined detection of three serum tumor biomarkers for lymph node metastasis.

Predictive Factor	AUC Value	Sensitivity	Specificity	Youden's Index	95%CI
Combined Detection	0.714	0.833	0.559	0.392	0.643 ~ 0.827

It should be noted that although the sample size included in this study was relatively large, the number of patients with lymph node metastasis was relatively small, with only 98 cases, accounting for 9.8% of the total study population. Therefore, calibration analysis was further performed based on the fitted model to evaluate the consistency between the overall predicted probabilities and the actual observed outcomes (Fig. 4). According to the calibration curve, the Brier score of the model was 0.047, indicating good agreement between the predicted probabilities generated by the model and the actual outcomes. However, the expected calibration error (ECE) was 0.19, suggesting that mild overestimation or underestimation occurred in certain probability intervals. This phenomenon may be attributed to the relatively low incidence of lymph node metastasis in the study population, resulting in a limited number of actual positive cases within each predicted probability interval. Consequently, greater fluctuations were observed in the intermediate probability range of the calibration curve.

Discussion

Endometrial cancer (EC) is currently one of the most common gynecological malignant tumors, and its incidence has shown a continuously increasing trend with improvements in living standards. Because lymph node metastasis lacks a universally accepted consensus guideline for assessment and represents an important indicator associated with poor prognosis, there is an urgent clinical need to accurately predict the possibility of lymph node metastasis before surgery. Such prediction may assist clinicians in selecting the most appropriate surgical treatment strategy and achieving optimal therapeutic outcomes.

Therefore, based on currently available clinical diagnostic and treatment data, establishing a predictive model for lymph node metastasis in patients with EC is of considerable clinical importance [16].

In this study, we selected 10 potential related factors commonly investigated in previous domestic and international studies, including patient age category, menstrual status, BMI category, history of diabetes mellitus, history of hypertension, number of previous childbirths, clinical stage, and three serum tumor biomarkers, namely CA199, CA125, and HE4. These variables were systematically evaluated.

Three commonly used machine learning models were applied to predict lymph node metastasis using the same dataset. All three models demonstrated relatively high predictive performance based on ROC curve analysis, with AUC values of 0.898 for logistic regression (LR), 0.915 for random forest (RF), and 0.936 for extreme gradient boosting (XGBoost), suggesting that the included clinical dataset had good predictive potential. Among the three models, XGBoost achieved the highest AUC value, indicating

superior predictive performance compared with LR and RF. Previous machine learning studies in other malignant tumors have also demonstrated that XGBoost models exhibited better predictive performance than other machine learning algorithms. For example, Nan Zihan, Zhang, and Ren et al. [17–19] reported that XGBoost showed advantages in predicting clinical outcomes of malignant tumors. The findings of the present study are consistent with previous research, further supporting the reliability and predictive capability of XGBoost-based models. Based on validation using the optimal model, this study demonstrated that CA125 achieved an AUC value of 0.703, with a sensitivity of 0.684 and a specificity of 0.663, representing the highest AUC value among the three evaluated tumor biomarkers. CA125 is a glycoprotein secreted by coelomic epithelial cells[20] and is widely expressed in tissues including the peritoneum, cervical epithelium, and endometrium. Elevated CA125 levels have been closely associated with malignant transformation and tumor progression, making it one of the earliest identified tumor-associated antigens. The present study further confirmed that CA125 was closely associated with lymph node metastasis in patients with EC. Therefore, CA125 measurement may contribute to early assessment of tumor progression and facilitate optimization of clinical treatment strategies [16] HE4 was initially isolated from human epididymal tissue. It is expressed at low levels in normal tissues and benign tumors but is abnormally elevated in both tumor tissues and serum samples of patients with endometrial cancer. Therefore, HE4 has been considered an important biomarker for predicting metastasis in patients with EC. The findings of this study are consistent with previous studies conducted by Wang Yingxiang, Romera, Maurya, and other researchers [21–23], further supporting the potential clinical value of HE4 in assessing lymph node metastasis risk. CA199 is a carbohydrate antigen expressed on the surface of cell membranes [24]. Although it is commonly used for the diagnosis and differential diagnosis of gastrointestinal tumors, it has also been increasingly applied in the evaluation of gynecological malignancies in recent years.

In the present study, CA199 showed relatively high specificity (76.5%) but relatively low sensitivity (48.0%), indicating that its predictive performance as an individual biomarker for lymph node metastasis was limited. Therefore, combined detection of multiple tumor biomarkers was considered in this study[25]. Using the optimal XGBoost model, the combined detection of CA125, HE4, and CA199 was incorporated into the prediction model. The ROC curve analysis demonstrated that the AUC value of the combined biomarker model reached 0.73, which was higher than the predictive performance of any individual tumor biomarker. The sensitivity of the combined model reached 83.3%, while the specificity was 55.9%. These findings suggest that integrating multiple serum tumor biomarkers may improve the ability to identify patients at risk of lymph node metastasis compared with single-marker assessment. Furthermore, this study analyzed the optimal cutoff values associated with lymph node metastasis prediction. According to the XGBoost model results, the optimal cutoff value was 24.8 U/ml for CA125, 56.7 pmol/L for HE4, and 31.42 U/ml for CA199. These cutoff values were similar to those reported in previous studies by Wang Yingxiang, Zhao Jian, and other researchers[21, 25], suggesting that these thresholds may provide additional reference values for clinical risk assessment.

Through predictive modeling analysis, this study provides potential evidence for clinical differential assessment, which may assist clinicians in making more accurate judgments, improving the allocation

of medical resources, and reducing unnecessary medical costs, thereby benefiting both healthcare providers and patients[26]. However, several limitations should be acknowledged. Although this study included a relatively large sample size, the number of positive lymph node metastasis cases remained limited. In future studies, larger cohorts with more positive cases should be included, and external validation using independent datasets should be performed to further improve the robustness and generalizability of the predictive model [27].

Declarations

Ethics approval and consent to participate. This retrospective study was approved by the Ethics Committee of the Affiliated Cancer Hospital of Xinjiang Medical University (Approval No. S-2025253, dated 26 September 2025). All procedures performed in this study were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Written informed consent was obtained from all individual participants included in the study.

Consent to participate. Informed consent was obtained from all individual participants included in this study.

Funding. This work was supported by the Xinjiang Strategic Talent Cultivation Plan – Leading Science and Technology Talent Program (“Development of a Data-Driven Intelligent Platform for Early Warning Models of Endometrial Cancer”, Grant No. XJRC-2025-KJ-PY-KJLJ-123, 2026/01–2028/12, CNY 1,260,000, PI: Yan Wang).

Conflict of interest. The authors declare that they have no conflicts of interest.

Data availability. The data that support the findings of this study are available from the Affiliated Cancer Hospital of Xinjiang Medical University, but restrictions apply to the availability of these data, which were used under licence for the current study and so are not publicly available. The data are, however, available from the authors upon reasonable request and with the permission of the Ethics Committee of the Affiliated Cancer Hospital of Xinjiang Medical University.

Authors’ contributions. W.H.Z. was responsible for manuscript writing, data analysis, and figure preparation. Y.P.H. performed initial data cleaning and table compilation. J.F.L. and H.T. contributed to manuscript proofreading and revision of details. Corresponding authors H.M.J. and Y.W. provided the study concept, funding support, and technical guidance. All authors read and approved the final manuscript.

References

1. Zheng RS, Chen R, Han BF, Wang SM, Li L, Sun KX et al. Cancer incidence and mortality in China in 2022: analysis of national cancer statistics. 2024:1–11.

2. Luo XZ, Chen XJ, Feng WW, Tang XD, Wang HY, Wang ZQ, et al. Guidelines for minimally invasive and tumor-free management of diagnosis and treatment of endometrial cancer (2025 edition). *Chin J Practical Gynecol Obstet*. 2025;41:905–13.
3. Healthy China Action—Implementation. Plan for Cancer Prevention and Control (2023–2030). *China Cancer*. 2023;32:887–90.
4. Zhang GN et al. Chinese expert consensus on screening and prevention of Lynch syndrome-associated endometrial cancer (2023 edition). *Chin J Practical Gynecol Obstet*. 2023.
5. Hao M, Zheng XD, Wei XJ, Ma N, Jiang MY, Zhu XQ, et al. Factors associated with lymphovascular space invasion in low-grade endometrioid endometrial carcinoma. *Chin J Minim Invasive Surg*. 2024;24:529–33.
6. Zhang Y, Zhang J, Wang L, Wang F. Clinical significance of serum human epididymis protein 4 and carbohydrate antigen 125 in predicting lymph node metastasis of endometrial cancer. *J Practical Med*. 2020;36:1340–13431348.
7. Niu WW, Yao YF. Diagnostic value of pelvic magnetic resonance imaging combined with HE4 and CA125 detection for predicting lymph node metastasis in patients with endometrial cancer. *Chin J Health Eng*. 2023;22:675–7.
8. Wu SW, Xie YQ, Feng YF, Zhang XH, Zhang MM, Yan SJ, et al. Analysis of related factors associated with lymph node metastasis and prognosis in stage I–III endometrial cancer. *J Practical Obstet Gynecol*. 2021;37:689–93.
9. Saarelainen KS, Peltonen, Nina Lehtimäki et al. Terho,. Predictive value of serum human epididymis protein 4 and cancer antigen 125 concentrations in endometrial carcinoma. *American Journal of Obstetrics and Gynecology*. 2013;209:142.e1-142.e6.
10. Zanotti L, Bignotti E, Calza S, Bandiera E, Ruggeri G, Galli C et al. Human epididymis protein 4 as a serum marker for diagnosis of endometrial carcinoma and prediction of clinical outcome.:1–10.
11. Wang AN, Jia HQ, Lu GM, Long ZQ. Value of serum human epididymis protein 4, carbohydrate antigen 125, and risk of ovarian malignancy algorithm index in predicting lymph node metastasis of endometrial cancer. *Chin J Practical Gynecol Obstet*. 2020;36:996–1000.
12. Wang DD, Wu YL, Han L, Zhang FL. Analysis of factors associated with lymph node metastasis and establishment of predictive model in endometrial cancer. *China Med Pharm*. 2025;15:129–132136.
13. Guo S, Shang Y, Niu LN, Wei B. Clinical value of serum CA125 and HE4 levels in evaluating lymph node metastasis in patients with endometrial cancer. *Chin J Family Plann*. 2025;33:1885–9.
14. Jiang LH, Min B, Chang Y, Song JJ. Application value of serum CA724, CA125, and HE4 in diagnosis of endometrial cancer and lymph node metastasis. *Practical J Cancer*. 2026;41:371–373381.
15. Reijnen C, IntHout, Joanna, Massuger GLFA, et al. Diagnostic Accuracy of Clinical Biomarkers for Preoperative Prediction of Lymph Node Metastasis in Endometrial Carcinoma: A Systematic Review and Meta-Analysis. *Oncologist*. 2019;24:e880–90.
16. Wang J, Yu H, He JX, Wang H, Wang XF, Yu JL. Significance of serum and endometrial tissue CA125 and HE4 levels in preoperative evaluation and pathological characteristics of endometrial cancer.

Maternal Child Health Care China. 2021;36:2506–8.

17. Nan ZH, Han LN, Li SW, Zhu ZY, Zhu QQ, Duan Y, et al. Prediction of prognosis in patients with malignant tumors complicated with acute respiratory failure using an interpretable machine learning model: based on the eICU Collaborative Research Database. *Med J Peking Union Med Coll Hosp*. 2026;17:98–108.
18. Zhang J, Wang, Dengfeng, Peng L, et al. Preoperative evaluation and a nomogram prediction model for pelvic lymph node metastasis in endometrial cancer. *Eur J Surg Oncol*. 2024;50:108230.
19. Ren Z, Chen, Banghong H. The value of machine learning in preoperative identification of lymph node metastasis status in endometrial cancer: a systematic review and meta-analysis. *Front Oncol*. 2023;13:1289050.
20. Kang S, Nam J-H. Preoperative assessment of lymph node metastasis in endometrial cancer: A Korean Gynecologic Oncology Group study. *Cancer*. 2017;123:263–72.
21. Wang YX, Gu J, Wang J. Predictive value of inflammatory markers in differentiating endometrial lesions. *J Sun Yat-sen Univ (Medical Sciences)*. 2024;45:923–9.
22. Espiau Romera A, Cuesta Guardiola T, Benito Vielba M, De Bonroostro Torralba C, Coronado Martín PJ, Baquedano Mainar L. HE4 tumor marker as a predictive factor for lymphatic metastasis in endometrial cancer. *Int J Gynecol Obstet*. 2020;149:265–8.
23. Maurya V, Singh, Shruti, Singh S, et al. Tumor biomarkers HE4 and CA125 alongside enhanced computed tomography imaging in assessing lymph node metastasis in endometrial cancer. *Bioinformation*. 2024;20:1453–60.
24. Li Y, Cong P, Wang P, et al. Risk factors for pelvic lymph node metastasis in endometrial cancer. *Arch Gynecol Obstet*. 2019;300:1007–13.
25. Zhao J, Gao JR. Predictive value of CA199, VEGF, and HE4 detection for lymph node metastasis in endometrial cancer. *Chin J Family Plann*. 2020;28:1901–1904,1908,1937.
26. Liu X, Weiha X, Qiao J, et al. Prediction of Lymph Node Metastasis in Endometrial Cancer Based on Color Doppler Ultrasound Radiomics. *Acad Radiol*. 2024;31:4499–508.
27. Wu H, Feng, Haiqin, Miao X, et al. Construction and validation of a prognostic model based on 11 lymph node metastasis-related genes for overall survival in endometrial cancer. *Cancer Med*. 2022;11:4641–55.

Figures

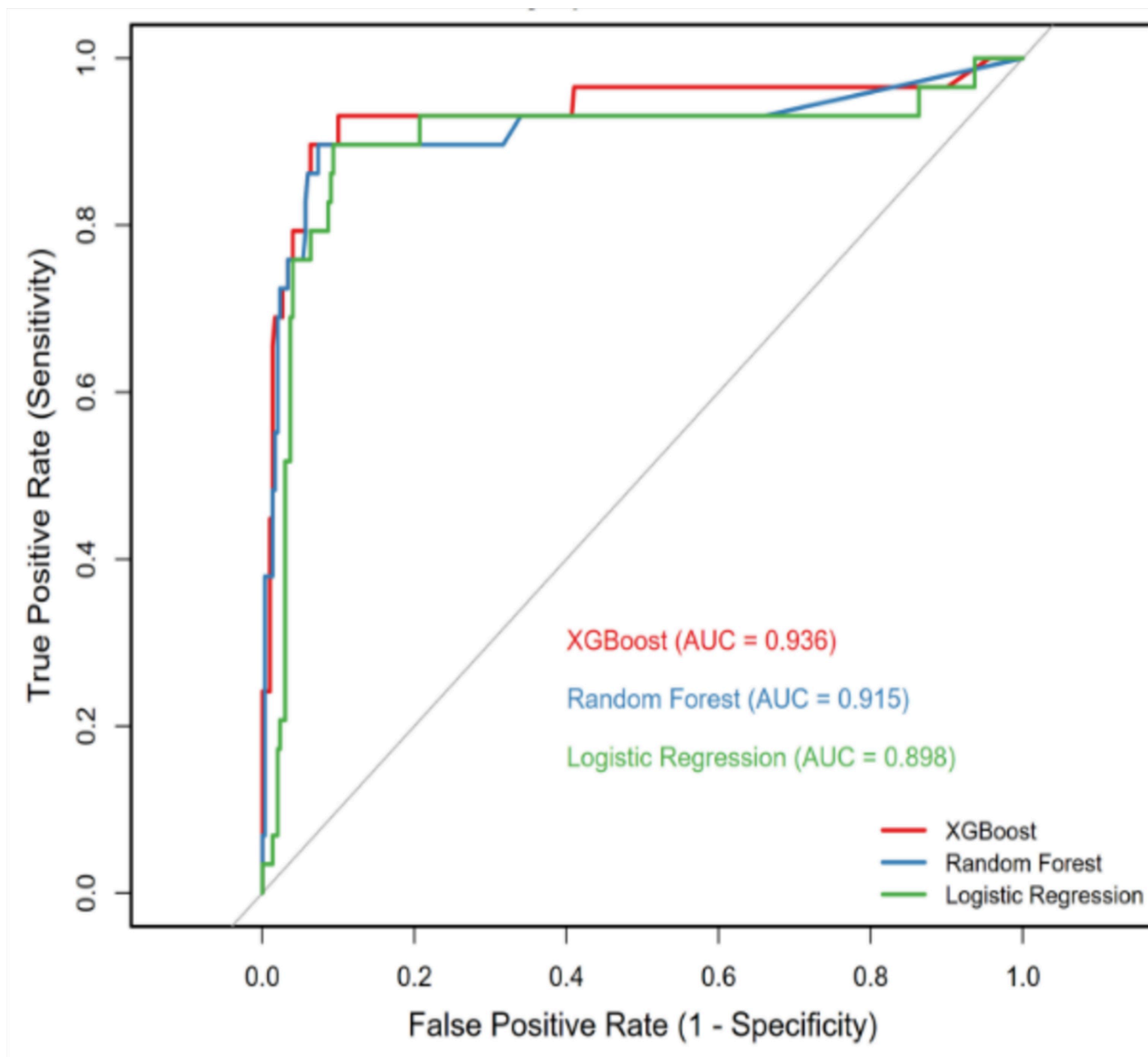


Figure 1

Receiver operating characteristic curves of three machine learning models for predicting lymph node metastasis in patients with endometrial cancer.

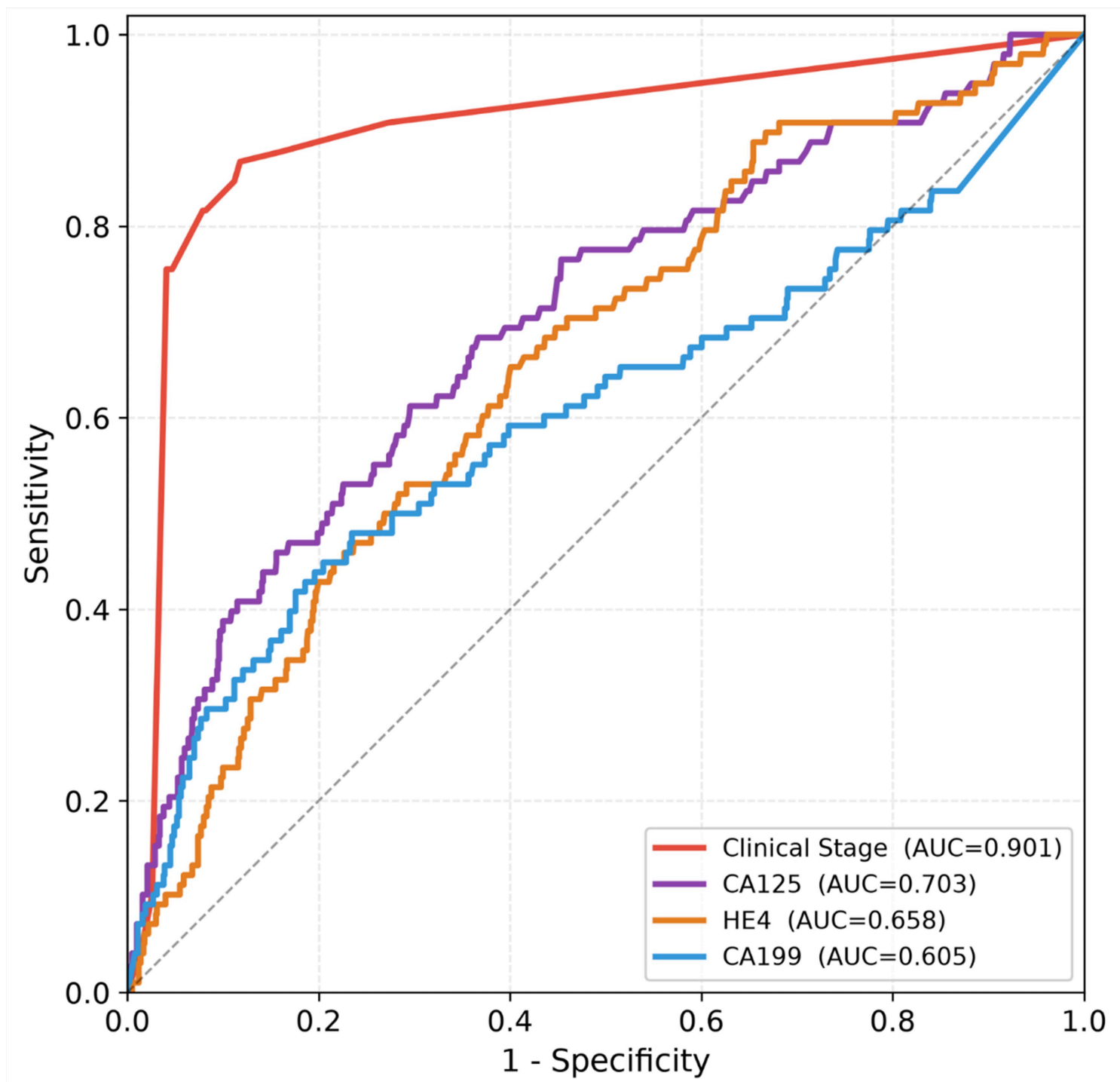


Figure 2

Receiver operating characteristic curves of significant predictors in the XGBoost model.

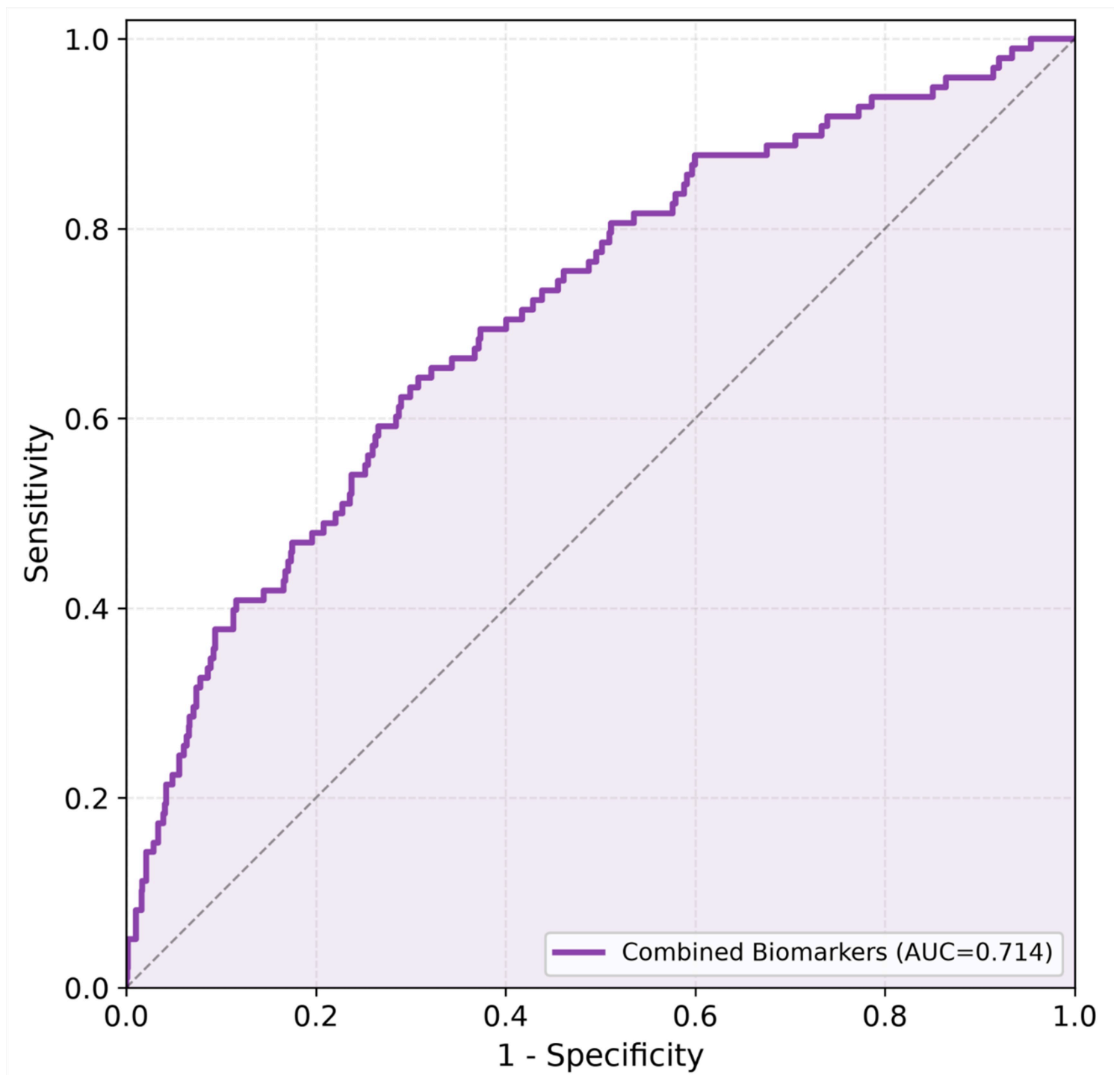


Figure 3

ROC Curve of combined serum tumor biomarkers in XGboost model

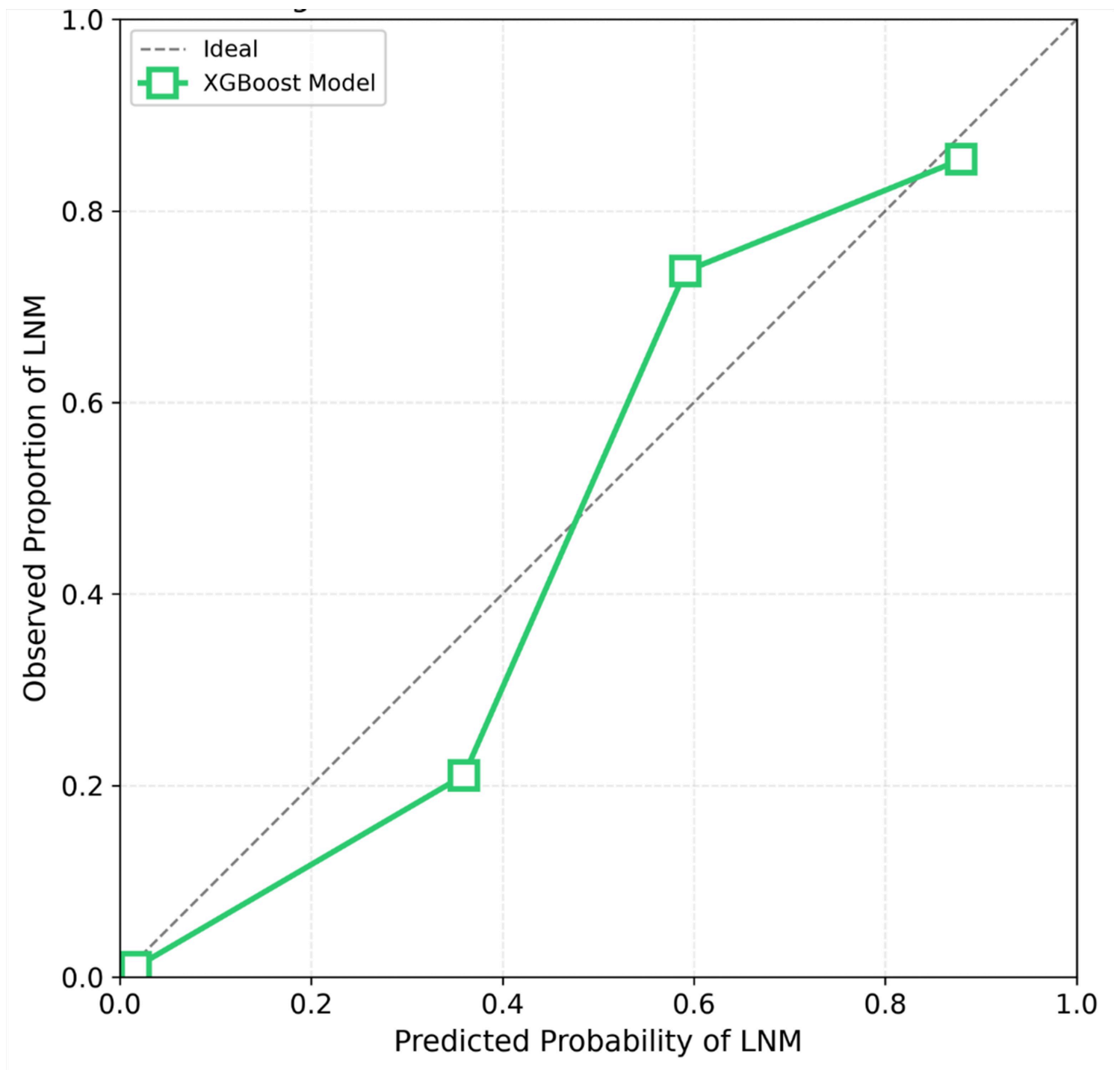


Figure 4

Calibration curve of the combined tumor biomarker model in the XGBoost model