**Machine Learning for Preoperative Discrimination of Lymph Node Metastasis Status in Endometrial Carcinoma: A Systematic Review and Meta-analysis**

**Abstract**

**Background:** Early identification of lymph node metastasis (LNM) status in endometrial carcinoma (EC) is highly challenging in clinical practice. Mayo criteria have been widely applied, but its early predictive accuracy is debatable. Therefore, some researchers have introduced machine learning (ML) into the early identification of LNM in EC patients, but the predictive performance of ML remains controversial due to diverse models and modeling variables (i.e., clinical features and radiomic features). Therefore, this systematic review and meta-analysis was conducted to systematically assess the performance of ML in the early discrimination of LNM in EC patients.

**Methods:** PubMed, Cochrane, Embase, and Web of Science were systematically searched from database inception to March 12, 2023. The risk of bias in the included studies was assessed by PROBAST. Subgroup analyses were performed by modeling variables (i.e., clinical features, radiomics features, and radiomics plus clinical features) and different ML methods. In addition, the clinical features currently used for the early identification of LNM status in EC patients were comprehensively pooled.

**Results:** Fifty primary studies were included, covering 103,752 EC patients, 12,579 of whom were positive for LNM. The meta-analysis results showed that the ML model based on radiomics plus clinical features performed best, with a pooled C-index of 0.907 (95% CI: 0.886-0.928) in the training set and 0.823 (95% CI: 0.757-0.890) in the validation set, and its sensitivity and specificity were satisfactory. The C-index of the ML model based on clinical features was not inferior to that of the radiomics-based ML model. In addition, it was found that logistic regression was used most frequently and exhibited ideal predictive performance with different categories of modeling variables.

**Conclusions:** Although the predictive performance of the ML model based on radiomics plus clinical features is the highest, no recognized specification for the use of radiomics is available so far. The logistic regression model constructed by clinical features displays ideal sensitivity and specificity. Against this background, large-sample studies covering different races are necessary to develop predictive nomograms based on clinical features, which can be popularized and applied in clinical practice.

**Keywords:** endometrial carcinoma, lymph node metastasis, radiomics, machine learning, systematic review

**Introduction**

Endometrial carcinoma (EC) is the most common gynecologic cancer in high-income countries. Globally, 417,367 women were diagnosed with EC in 2020. It is more prevalent in high-income regions than in low- and middle-income countries, with the highest burdens in North America and Western Europe. The incidence of EC appears to be increasing rapidly (1, 2). The incidence and mortality of EC are 19.2-20.2/100,000 and 2.0-3.7/100,000 in women in Europe as of 2018, respectively (3, 4), seriously threatening their lives.

Surgery is the mainstay of treatment for patients with localized EC. However, the need for lymphadenectomy during surgery is controversial, and it has been debated whether or not to add para-aortic lymphadenectomy to pelvic lymphadenectomy (2, 5). Previously, complete standard lymphadenectomy (i.e., pelvic and para-aortic lymph node dissection and evaluation) was recommended for all patients, but it could result in multiple side effects (6). Therefore, effective preoperative assessment of lymph node metastasis (LNM) is significant in clinical practice. Unfortunately, efficient preoperative assessment methods are lacking. The Mayo criteria have been widely used in clinical practice for predicting the risk of LNM in EC (7), but its true predictive accuracy remains to be further improved.

As the statistical theory gradually improves, machine learning (ML) methods, especially supervised ML methods, have been gradually utilized to diagnose disease states (8, 9), and predict the occurrence (10, 11) and prognosis of diseases (12, 13). In particular, ML appears to be no worse than human clinical practice in screening or diagnosing diseases in some fields (14, 15). In this context, some researchers have used ML methods to identify preoperative LNM status in EC. However, ML encompasses diverse mathematical modeling methods (e.g., logistic regression, random forest, support vector machine, and artificial neural network), and also involves a wide range of modeling variables (e.g., radiomics, clinical features, and pathological imaging). With such a diversity of modeling methods and variables, a comprehensive and systematic understanding of the preoperative diagnostic performance of ML for LNM status in EC is lacking (16). Therefore, the purpose of this systematic review and meta-analysis was to explore the predictive performance of ML for LNM in EC patients. The effective predictive variables were also comprehensively summarized and the predictive performance of clinical and radiomic features for LNM in EC patients was compared.

## Methods

**1. Study registration**

 This study was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA 2020) and was prospectively registered with PROSPERO (ID:)

**2. Eligibility criteria**

**Inclusion criteria**

1. The patients were diagnosed with EC;
2. Case-control studies, cohort studies, nested case-control studies, and case-cohort studies;
3. A complete ML model for predicting LNM status was established;
4. External validation or independent validation was lacking even if the corresponding ML model was established;
5. Different types of ML were constructed from the same dataset;
6. English-language studies.

**Exclusion criteria**

1. Meta-analysis, review, guidelines, and expert opinion;
2. Only risk factors were analyzed, and no complete risk model was established;
3. The following outcome indicators (ROC, C-statistic, C-index, sensitivity, specificity, accuracy, recovery, precision, confusion matrix, diagnostic fourfold table, F1 score, and calibration curve) were lacking for the predictive accuracy of the risk model;
4. Studies on the validation of a mature scale;
5. Studies on the accuracy of a single factor in predicting LNM.

**3. Data sources and search strategy**

PubMed, Embase, Web of Science, and Cochrane were systematically searched from database inception to March 12, 2023 using subject terms plus free terms, without restrictions on publication regions. The search strategy is detailed in Fig. 1.

**4. Study selection and data extraction**

The duplicate publications were removed automatically and then manually by importing the retrieved studies into Endnote X9. Then the titles or abstracts of the remaining studies were read to obtain the initially eligible studies. Finally, the full texts were read to include eligible studies.

Before data extraction, a standardized data extraction form was developed to record the title, first author, year of publication, study type (case-control, retrospective/prospective cohort study, nested cohort study, and case-cohort study), patient source (single-center, multicenter, and registry database), FIGO stage, number of LNM cases (training set, and validation set ), total number of cases, generation method of validation set (internal validation: random sampling, k-fold cross-validation, and leave-one-out; external validation: prospective, and multicenter; overfitting method: k-fold cross-validation, and bootstrap), missing value handling method, variable screening/feature selection method, types of mathematical models and modeling variables.

The study screening, data extraction, and cross-checking were carried out independently by two investigators (RZL and YJY). Disagreement was resolved by consultation with a third investigator (LYQ).

**5. Risk of bias (RoB) assessment**

The RoB in the included studies was assessed using PROBAST. This tool covered several questions in four domains (participants, predictive variables, results, and statistical analyses), reflecting the overall RoB and applicability (17). The four domains contained two, three, six, and nine specific questions, respectively, answered by yes/probably yes, no/probably no, or no information. A domain was considered high RoB if at least one specific question in the domain was answered by no/probably no. A domain was considered low RoB if all of the specific questions in the domain were answered by yes/probably yes. The RoB for a domain was considered unclear if all specific questions in the domain were not answered by no/probably no and at least one question was answered by no information. PROBAST was used to evaluate the ML models in the studies.

The RoB assessment and cross-checking were carried out independently by two investigators (RZL and YJY). Disagreement was resolved by consultation with a third investigator (LYQ).

**6. Outcomes**

The C-index was used as an outcome indicator to reflect the overall accuracy of the model. However, the C-index cannot reflect the accuracy of the model in predicting LNM, especially in case of a significant imbalance between the number of LNM and non-LNM samples. A high C-index may be attributed to the high accuracy of the model in predicting negative events (non-LNM). Therefore, the sensitivity and specificity of the ML model in predicting LNM were also included. In addition, the modeling variables were also pooled. As logistic regression is the main method for establishing the ML model in clinical practice, the odds ratio (OR) of each modeling variable for constructing logistic regression was also incorporated as the outcome indicator, in order to try to construct a logistic regression risk equation for LNM.

**7. Synthesis methods**

The C-index and its standard error (SE) and 95% confidence interval (95% CI) should be provided for the meta-analysis of the C-index. However, many included studies did not provide the SE and 95% CI of C-index, and thus the SE was estimated with reference to the study by Debray TP et al. (18). Meta-analyses were also performed on sensitivity and specificity, for which a diagnostic fourfold table was required. However, only the outcome indicators of sensitivity, specificity, precision, and accuracy were provided in the included studies, and thus we developed the diagnostic fourfold table by combining the number of cases with LNM and the total number of cases. In addition, some included studies only provided the receiver operating characteristic (ROC) curve of the ML model, for which Origen2021 was used to extract the sensitivity and specificity on the ROC curve, and the optimal Youden index was used to select the sensitivity and specificity. A diagnostic fourfold table was developed based on the number of cases with LNM and the total number of cases. Moreover, if the included studies converted continuous variables into categorical variables or kept them in the original continuous state, we conducted meta-analyses for continuous and categorical variables when pooling the OR of modeling variables in logistic regression, respectively.

In meta-analyses of C-index and OR of modeling variables, a random-effects model was adopted when the heterogeneity index *I2* was ≥50%, and a fixed-effects model when *I2* was <50%. A bivariate mixed-effects model was used for meta-analyses of sensitivity and specificity.

In addition, the modeling variables included clinical features, radiomic features, and radiomics plus clinical features. Diversified ML methods were included. Therefore, subgroup analyses were performed by different modeling variables and ML methods. R4.2.0 (R Development Core Team, Vienna, http://www.R-project.org) was used for this meta-analysis, and P<0.05 was considered statistically significant.

## Results

**1. Search results**

A total of 3,033 studies were retrieved, of which 782 duplicate publications were automatically marked by Endnote. Endnote can only mark studies with a completely consistent title and author name, but a large number of duplicate publications had minor differences in the two aspects, making automatic marking difficult. Therefore, 356 duplicates were manually removed. After reading the titles or abstracts of the remaining studies, 62 studies were initially eligible. After reading the full texts downloaded, 50 studies (19-68) were finally included (Fig. 1).



Fig. 1 Study selection process

**2. Study characteristics**

The 50 included studies covered 103,752 patients with EC, of whom 12,579 were positive for LNM. These studies were published from 2013 to 2023, mainly from 2016 to 2023. These eligible studies consisted of case-control studies and cohort studies. Among them, there were only 10 cohort studies (28, 35, 39, 49, 51-53, 62-64), 11 multicenter studies (26, 30, 33, 45, 46, 60-64, 66), and one study (40) based on the SEER database, and other studies were single-center studies. The median number of cases was 342 (IQR: 200-661), and it was 300 (IQR: 154-533) in the training set. Only 29 studies (19-21, 23, 25, 26, 29-33, 36-38, 40, 43, 45, 47-50, 52, 54, 55, 58, 60-63) had independent validation sets, of which 11 (21, 30, 31, 33, 36, 45, 55, 61-63, 65) adopted external validation, and the rest studies adopted internal validation with random sampling. Nine studies (20-22, 31, 32, 35, 45, 52) did not explicitly describe the variable screening method, and univariate and multivariate logistic regression was mainly used.

Fourteen studies (22, 23, 25, 26, 32, 34, 36, 37, 47-50, 52, 55) used radiomics as modeling variables and included 27 models, containing a diversity of mathematical algorithms: logistic regression (n=16), ridge regression (n=2), J48 (Decision tree (n=1)), random forest (n=1), support vector machine (n=2), XGBoost (n=1), artificial neural network (n=2), Hoeffding Tree (n=1), and ResNet-18 (n=1). The modeling variables in the remaining studies were from clinical features and 41 models were included, which contained the following mathematical algorithms: logistic regression (n=38), ridge regression (n=1), random forest (n=1), and support vector machine (n=1). The basic information is shown in Table 1.

**3. RoB**

Among the original studies, 39 studies were case-control studies, in which the ML models were considered to have a high RoB in populations according to PROBAST. Whether the assessment of predictors was conducted under the condition of known LNM status in a large number of single-center case-control studies was unclear, in which the ML models were also considered to have a high RoB in prediction factors. Since the LNM status was confirmed by biopsy, the RoB was rarely rated as high in results. In addition, according to the statistical analysis, the number of cases in the training set needed to satisfy EVP≥20 and an independent validation set was required with >100 cases. Thus, the RoB was mostly high in the statistical analysis. The results of RoB assessment are shown in Fig. 2.

Fig. 2 Results of RoB assessment by PROBAST

**4. Meta-analysis**

**4.1 Mayo criteria**

The accuracy of the Mayo criteria for predicting LNM in EC patients was validated by eight datasets from the included studies. The meta-analysis results showed a C-index of 0.690 (95% CI: 0.640-0.740), a sensitivity of 0.81 (95% CI: 0.66-0.90), and a specificity of 0.59 (95% CI: 0.38-0.77) in the training set. The detailed results are shown in Tables 1 and 2.

**4.2 ML models for LNM status based on clinical features**

In the included studies, 41 ML models were established based on clinical features. There were 36 ML models in the training set with a pooled C-index of 0.824 (95% CI: 0.806-0.843), and 19 ML models in the validation set with a pooled C-index of 0.793 (95% CI: 0.756-0.829). In the training set, the pooled sensitivity was 0.81 (95% CI: 0.77-0.84) and the specificity was 0.75 (95% CI: 0.71-0.79). The pooled sensitivity and specificity were 0.75 (95% CI: 0.67-0.82) and 0.78 (95% CI: 0.74-0.82), respectively, in the validation set. The detailed results are displayed in Tables 1 and 2.

**4.3 Radiomics-based ML models for LNM status**

Sixteen ML models were established based on radiomic features in the included studies. The pooled C-index was 0.798 (95% CI: 0.758-0.837) in the training set and 0.810 (95% CI: 0.770-0.850) in the validation set. In the training set, the pooled sensitivity was 0.82 (95% CI: 0.79-0.85) and the specificity was 0.83 (95% CI: 0.79-0.87). The pooled sensitivity and specificity were 0.77 (95% CI: 0.64-0.87) and 0.84 (95% CI: 0.74-0.91), respectively, in the validation set. The detailed results are displayed in Tables 1 and 2.

**4.4 ML model for LNM status based on radiomics plus clinical features**

The included studies contained 11 ML models established based on radiomics plus clinical features, with a pooled C-index of 0.907 (95% CI: 0.886-0.928) in the training set and 0.823 (95% CI: 0.757-0.890) in the validation set. In the training set, the pooled sensitivity was 0.88 (95% CI: 0.84-0.92) and the specificity was 0.83 (95% CI: 0.79-0.87). The pooled sensitivity and specificity were 0.77 (95% CI: 0.64-0.87) and 0.84 (95% CI: 0.74-0.91), respectively, in the validation set. The detailed results are displayed in Tables 1 and 2.

**4.5 Subgroup analyses**

Subgroup analyses were performed by the clinical features, radiomics, and radiomics plus clinical features. Logistic regression was used most frequently to construct models based on different modeling variables, and a visual nomogram was also constructed in most of the studies. The meta-analysis results showed that the predictive performance of logistic regression was high and was not inferior to that of other ML models based on the same type of variables.

For the logistic regression models constructed based on clinical features, the pooled C-index, sensitivity, and specificity were 0.828 (95% CI: 0.809-0.846), 0.81 (95% CI: 0.78-0.84), and 0.75 (95% CI: 0.71-0.79), respectively, in the training set, and 0.806 (95% CI: 0.758-0.853), 0.81 (95% CI: 0.78-0.84), and 0.75 (95% CI: 0.71-0.79), respectively, in the validation set.

For the logistic regression models constructed based on radiomics, the pooled C-index, sensitivity, and specificity were 0.850 (95% CI: 0.815-0.884), 0.83 (95% CI: 0.75-0.88), and 0.79 (95% CI: 0.76-0.82), respectively, in the training set, and 0.818 (95% CI: 0.775-0.861), 0.77 (95% CI: 0.60-0.88), and 0.86 (95% CI: 0.75-0.93), respectively, in the validation set.

For the logistic regression models constructed based on radiomics + clinical features, the pooled C-index, sensitivity, and specificity were 0.905 (95% CI: 0.885-0.925), 0.90 (95% CI: 0.84-0.94), and 0.80 (95% CI: 0.72-0.86), respectively, in the training set, and 0.842 (95% CI: 0.793-0.891), 0.90 (95% CI: 0.84-0.94), and 0.80 (95% CI: 0.72-0.86), respectively, in the validation set.

Table 1 Meta-analysis results of C-index of ML for predicting LNM in EC patients

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Model type | Training set | Validation set |
| n | C-index (95% CI) | n | C-index (95% CI) |
| Radiomics | 　 | 　 | 　 | 　 | 　 |
| 　 | Logistic regression | 9 | 0.850(0.815-0.884) | 6 | 0.818(0.775-0.861) |
| 　 | Artificial neural network | 2 | 0.762(0.687-0.837) | 1 | 0.710(0.515-0.905) |
| 　 | Support vector machine | 2 | 0.727(0.661-0.793) | 1 | 0.754(0.607-0.901) |
| 　 | Ridge regression | 1 | 0.730(0.605-0.855) | 　 | 　 |
| 　 | HoeffdingTree | 1 | 0.688(0.585-0.791) | 　 | 　 |
| 　 | Convolutional neural network | 1 | 0.701(0.599-0.803) | 　 | 　 |
| 　 | Overall | 16 | 0.798(0.758-0.837) | 8 | 0.810(0.770-0.850) |
| Radiomics plus clinical features | 　 | 　 | 　 | 　 | 　 |
| 　 | Logistic regression | 7 | 0.905(0.885-0.925) | 4 | 0.842(0.793-0.891) |
| 　 | Random forest | 1 | 0.935(0.906-0.964) | 2 | 0.903(0.866-0.939) |
| 　 | XGBoost | 1 | 0.800(0.680-0.920) | 1 | 0.720(0.700-0.740) |
| 　 | Ridge regression | 1 | 0.800(0.700-0.900) | 1 | 0.750(0.550-0.950) |
| 　 | Convolutional neural network | 1 | 0.938(0.913-0.963) | 1 | 0.770(0.718-0.822) |
|   | Overall | 11 | 0.907(0.886-0.928) | 9 | 0.823(0.757-0.890) |
| Clinical features | 　 | 　 | 　 | 　 | 　 |
| 　 | Logistic regression | 33 | 0.828(0.809-0.846) | 14 | 0.806(0.758-0.853) |
| 　 | Random forest | 1 | 0.810(0.739-0.881) | 1 | 0.820(0.735-0.905) |
| 　 | Support vector machine | 1 | 0.810(0.739-0.881) | 1 | 0.760(0.663-0.857) |
| 　 | Bayesian network | 　 | 　 | 3 | 0.833(0.794-0.873) |
| 　 | Ridge regression | 1 | 0.610(0.458-0.762) | 　 | 　 |
| 　 | Overall | 36 | 0.824(0.806-0.843) | 19 | 0.793(0.756-0.829) |
| Mayo |  |  |  | 8 | 0.690(0.640-0.740) |

Note: (1) n: number of models; (2) Radiomics, radiomics plus clinical features, and clinical features represent modeling variables, where radiomics plus clinical features represent the combination of radiomics and clinical features as a modeling variable.

Table 2 Meta-analysis results of sensitivity and specificity of ML for predicting LNM in EC patients

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Model type | Training set | Validation set |
| n | Sen (95% CI) | Spe (95% CI) | n | Sen (95% CI) | Spe (95% CI) |
| Radiomics | 　 | 　 | 　 | 　 | 　 | 　 | 　 |
| 　 | Logistic regression | 10 | 0.83(0.75-0.88) | 0.79(0.76-0.82) | 7 | 0.77(0.60-0.88) | 0.86(0.75-0.93) |
| 　 | Artificial neural network | 2 | 0.77-0.86 | 0.66-0.94 | 1 | 0.89 | 0.75 |
| 　 | Support vector machine | 2 | 0.75-0.81 | 0.75-0.87 | 1 | 0.71 | 0.72 |
| 　 | Ridge regression | 1 | 0.7 | 0.86 | 　 | 　 | 　 |
| 　 | HoeffdingTree | 1 | 0.81 | 0.87 | 　 | 　 | 　 |
| 　 | Convolutional neural network | 2 | 0.80-0.83 | 0.90-0.91 | 　 | 　 | 　 |
| 　 | Overall | 18 | 0.82(0.79-0.85) | 0.83(0.79-0.87) | 9 | 0.77(0.64-0.87) | 0.84(0.74-0.91) |
| Radiomics plus clinical features | 　 | 　 | 　 | 　 | 　 | 　 | 　 |
| 　 | Logistic regression | 7 | 0.90(0.84-0.94) | 0.80(0.72-0.86) | 6 | 0.78(0.62-0.88) | 0.87(0.78-0.93) |
| 　 | Artificial neural network | 1 | 0.92 | 0.84 | 2 | 0.85-0.89 | 0.75-0.83 |
| 　 | Ridge regression | 1 | 0.71 | 0.73 |  |  |  |
| 　 | Convolutional neural network | 1 | 0.83 | 0.91 |  |  |  |
| 　 | Overall | 10 | 0.88(0.84-0.92) | 0.81(0.75-0.86) | 6 | 0.81(0.70-0.89) | 0.84(0.76-0.89) |
| Clinical features | 　 | 　 | 　 | 　 | 　 | 　 | 　 |
| 　 | Logistic regression | 31 | 0.81(0.78-0.84) | 0.75(0.71-0.79) | 13 | 0.74(0.66-0.80) | 0.79(0.75-0.82) |
| 　 | Random forest | 1 | 0.67 | 0.78 | 1 | 0.48 | 0.87 |
| 　 | Bayesian network | 　 | 　 | 　 | 2 | 0.87-0.94 | 0.68-0.70 |
| 　 | Overall | 32 | 0.81(0.77-0.84) | 0.75(0.71-0.79) | 16 | 0.75(0.67-0.82) | 0.78(0.74-0.82) |
| Mayo |  |  |  |  | 7 | 0.81(0.66-0.90) | 0.59(0.38-0.77) |

**4.6 Modeling variables in logistic regression**

The predictive performance of logistic regression models was not inferior to other ML models established by the same type of modeling variables. Therefore, the modeling variables included in logistic regression were summarized, and the meta-analysis results showed that grade, histological type, myometrial invasion, cervical stromal invasion, LVSI, CA125, CA153, CA199, Ki67, P53, tumor size, ER, enlarged lymph nodes, mitosis, and SII were effective predictors for LNM status in EC (P<0.05) (Table 3).

Table 3 Meta-analysis results of OR of modeling variables used to construct logistic regression models for predicting LNM in EC

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Factors | Value | n | OR (95% CI) | I2 (%) |
| Age | 　 | 　 | 　 | 　 |
| 　 | Per 1 year | 2 | 1.026(0.941-1.120) | 23.2  |
| 　 | >60 | 1 | 1.214(0.660-2.232) | NA |
| Grade | 　 | 　 | 　 | 　 |
| 　 | Grade2 | 7 | 2.258(2.039-2.500) | 0.0  |
| 　 | Grade3 | 7 | 2.979(1.625-5.459) | 81.0  |
| 　 | Grade2/3 | 5 | 2.015(1.130-3.592) | 67.9  |
| Histological type | Non-endometrioid | 7 | 2.535(1.873-3.430) | 65.5  |
| Myometrial invasion | >50% | 14 | 2.581(2.151-3.097) | 19.7  |
| Cervical stromal invasion | Yes | 6 | 2.411(1.715-3.388) | 9.7  |
| LVSI (Lymphovascular invasion) | Positive | 12 | 5.011(3.610-6.956) | 62.0  |
| CA125 (Carbohydrate antigen 125) | 　 | 　 | 　 | 　 |
| 　 | >30 | 1 | 3.967(0.478-32.893) | NA |
| 　 | >35 | 8 | 2.990(2.215-4.036) | 10.0  |
| 　 | >40 | 2 | 8.131(4.451-14.853) | 0.0  |
| 　 | >50 | 2 | 6.674(4.331-10.283) | 0.0  |
| CA153 (Carbohydrate antigen 153) | >16.85 | 1 | 6.108(2.697-18.333) | NA |
| CA199 (Carbohydrate antigen 199) | >18.88 | 1 | 3.765(1.505-9.418) | NA |
| Ki-67 | 　 | 　 | 　 | 　 |
| 　 | Per unit | 3 | 1.028(1.014-1.041) | 54.8  |
| 　 | >50% | 3 | 2.397(1.385-4.151) | 0.0  |
| P53 | Aberrant | 5 | 2.402(1.143-5.049) | 35.3 |
| PR (Progesterone receptor) | 　 | 　 | 　 | 　 |
| 　 | 1-unit decrease | 2 | 0.989(0.974-1.005) | 79.6  |
| 　 | Negative | 2 | 1.234(0.576-2.642) | 0.0  |
| Tumor size | 　 | 　 | 　 | 　 |
| 　 | Per 1 cm | 3 | 1.348(1.128-1.611) | 18.3  |
| 　 | >4 cm | 2 | 2.065(1.317-3.237) | 0.0  |
| 　 | 2-5 cm | 1 | 1.510(1.340-1.700) | NA |
| 　 | 5-10 cm | 1 | 2.710(2.390-3.060) | NA |
| 　 | 10 cm- | 1 | 3.380(2.900-3.950) | NA |
| ER (Estrogen receptor) | 　 | 　 | 　 | 　 |
| 　 | Negative | 3 | 3.388(1.894-6.061) | 0.0  |
| 　 | 1-unit decrease | 2 | 0.979(0.962-0.995) | 81.1  |
| Enlarged lymph nodes | Positive | 1 | 3.590(1.400-9.170) | NA |
| HGB (Hemoglobin) | 　 | 1 | 0.983(0.967-0.999) | NA |
| MELF pattern | Present | 1 | 1.977(0.508-7.695) | NA |
| Mitosis | 　 | 1 | 3.202(1.650-6.214) | NA |
| SII (Systemic immune-inflammatory index) | >636.74 | 1 | 3.996(1.808-8.833) | NA |

## Discussion

**Summary of the main findings**

This study demonstrated that the ML modeling variables for predicting LNM status in EC patients included clinical features, radiomic features, and radiomics plus clinical features. Among models based on various modeling variables, the nomogram constructed based on logistic regression was dominant, and its C-index appeared to be not inferior to other models in the training and validation sets. In addition, the C-index of ML models based on clinical features was not inferior to that of radiomics-based ML models. Meanwhile, among the nomograms constructed based on logistic regression, the C-index of nomograms based on clinical features alone was close to that of nomograms based on radiomics alone.ML models based on radiomics plus clinical features exhibited the best predictive performance, which was also applied to the nomograms.

**Comparison with previous studies**

The accuracy of CT, MRI, PET/CT, ultrasound, and other imaging means for preoperative detection of LNM in EC patients was explored previously, with a focus on MRI and PET/CT. Bollineni VR et al. (69) conducted a systematic review of 13 original studies and found that the sensitivity and specificity of 18F-FDG PET/CT were 0.72 (95% CI: 0.55-0.98) and 0.92 (95% CI: 0.84-0.97) for preoperative detection of LNM in EC patients, respectively. A recent study that the sensitivity was 0.68 (95% CI: 0.63-0.73) for 18F-FDG PET and 0.96 (95% CI: 0.96-0.97) for PET/CT in the preoperative detection of LNM in EC patients (70). Qiu et al. (71) conducted a systematic review of 14 studies and confirmed that MRI had a sensitivity of 0.59 (95% CI: 0.48-0.69) and a specificity of 0.95 (95% CI: 0.93-0.96) for preoperative prediction of pelvic or/and para-aortic LNM in EC patients, whereas its sensitivity and specificity for preoperative prediction of pelvic LNM were 0.65 (95% CI: 0.51-0.77) and 0.95 (95% CI: 0.93-0.96), respectively. In a systematic review by Luomaranta A et al. (72), MRI exhibited similar sensitivity and specificity as the findings by Qiu et al. for preoperative detection in EC patients. The detection rate of sentinel LNM in EC patients by ultrasound also appears to be less satisfactory (73). It can be concluded that imaging means have a favorable specificity but a seriously insufficient sensitivity for preoperative detection of LNM in EC patients. The current study revealed that ML models had a higher sensitivity (>0.8), and ML models based on clinical features showed a higher sensitivity but a lower specificity to some extent.

In addition, this study showed that despite a high sensitivity in clinical practice, the Mayo criteria have a worrying specificity, which, however, was inferred based on a small number of studies. Therefore, the accuracy of the Mayo criteria for the identification of LNM in EC patients still requires further validation.

Some diversified ML methods, such as convolutional neural network, support vector machine, and XGBoost with high predictive performance, have been developed (74, 75), but logistic regression was still the most popular ML method in clinical practice. The main reason is that a nomogram can be constructed based on logistic regression. Nomograms, easy-to-use and highly visualized, are important for predicting LNM in tumors, such as the Briganti nomogram for prostate cancer (76). In this study, logistic regression appeared to have a satisfactory predictive performance. Subsequent studies, therefore, can be conducted to develop more general predictive nomograms for LNM in EC patients.

Modeling variables are key to improving the accuracy of ML. However, only a small number of studies summarized relevant evidence. Reijnen C et al. (77) conducted a systematic review and found an association of CA-125 and thrombocytosis with the risk of LNM in EC patients. A systematic review by Fu et al. (78) found that tumor diameter was associated with LNM. Therefore, the lack of independent predictors for LNM in EC patients has posed a challenge to the early identification of LNM status in EC patients. This study summarized the modeling variables incorporated in ML. Given that the sensitivity of models based on clinical features is desirable (>0.8), it is feasible to construct risk equations or nomograms for preoperative prediction of LNM in EC patients based on this study.

**Strengths and limitations**

In terms of strengths, this is the first systematic review that explores the performance of ML methods for preoperative diagnosis of LNM in EC patients, and summarizes main modeling variables (clinical features, and radiomic features), providing guidance and references for the development of clinical risk tools in the future. There are several limitations in this study. First, the included studies mainly focused on logistic regression, with less attention to other models, thus making it hard to summarize the application value of other models. Second, internal validation with random sampling was mainly adopted for model validation in the included studies, restricting the interpretation and generalizability of models. Especially for radiomics-based models, it poses a serious challenge since the radiomic features are seriously influenced by the experience of the radiologists and the configuration of radiation devices. Finally, the included studies were mainly case-control studies, some of which had small sample sizes, raising some concerns about the stability of the models.

**Conclusions**

ML is a feasible option for preoperative prediction of LNM status in EC patients, and the nomograms of logistic regression models based on clinical features exhibit high sensitivity and specificity. In addition, the predictive performance of models based on radiomics plus clinical features is higher. In the future, large-sample studies covering different races are expected to develop predictive nomograms based on clinical features, which can be popularized and applied in clinical practice. Given the excellent predictive performance of ML models based on radiomics plus clinical features, we also look forward to accelerating the development and application of radiomics and proposing standardized criteria for their application, thereby promoting the intelligent diagnosis of complex diseases and intelligent prediction of prognosis based on radiomics.

## References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249.

2. Makker V, MacKay H, Ray-Coquard I, et al. Endometrial cancer. Nat Rev Dis Primers. 2021;7(1):88.

3. Erratum: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2020;70(4):313.

4. Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer. 2019;144(8):1941-1953.

5. Koh WJ, Abu-Rustum NR, Bean S, et al. Uterine Neoplasms, Version 1.2018, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2018;16(2):170-199.

6. Abu-Rustum N, Yashar C, Arend R, et al. Uterine Neoplasms, Version 1.2023, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2023;21(2):181-209.

7. Mariani A, Webb MJ, Keeney GL, Haddock MG, Calori G, Podratz KC. Low-risk corpus cancer: is lymphadenectomy or radiotherapy necessary? Am J Obstet Gynecol. 2000;182(6):1506-1519.

8. Cho SJ, Sunwoo L, Baik SH, Bae YJ, Choi BS, Kim JH. Brain metastasis detection using machine learning: a systematic review and meta-analysis. Neuro Oncol. 2021;23(2):214-225.

9. Wu JH, Liu TYA, Hsu WT, Ho JH, Lee CC. Performance and Limitation of Machine Learning Algorithms for Diabetic Retinopathy Screening: Meta-analysis. J Med Internet Res. 2021;23(7):e23863.

10. Fleuren LM, Klausch TLT, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Med. 2020;46(3):383-400.

11. Aimo A, Pelliccia F, Panichella G, et al. Indications of beta-adrenoceptor blockers in Takotsubo syndrome and theoretical reasons to prefer agents with vasodilating activity. Int J Cardiol. 2021;333:45-50.

12. Tang R, Luo R, Tang S, Song H, Chen X. Machine learning in predicting antimicrobial resistance: a systematic review and meta-analysis. Int J Antimicrob Agents. 2022;60(5-6):106684.

13. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. Bmj. 2019;367:l5358.

14. Hickman SE, Woitek R, Le EPV, et al. Machine Learning for Workflow Applications in Screening Mammography: Systematic Review and Meta-Analysis. Radiology. 2022;302(1):88-104.

15. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Health. 2019;1(6):e271-e297.

16. Akazawa M, Hashimoto K. Artificial intelligence in gynecologic cancers: Current status and future challenges - A systematic review. Artif Intell Med. 2021;120:102164.

17. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019;170(1):51-58.

18. Debray TP, Damen JA, Riley RD, et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res. 2019;28(9):2768-2786.

19. Huang Y, Jiang P, Kong W, et al. Comprehensive Assessment of ERα, PR, Ki67, P53 to Predict the Risk of Lymph Node Metastasis in Low-Risk Endometrial Cancer. J Invest Surg. 2023;36(1):2152508.

20. Feng M, Zhao Y, Chen J, et al. A deep learning model for lymph node metastasis prediction based on digital histopathological images of primary endometrial cancer. Quant Imaging Med Surg. 2023;13(3):1899-1913.

21. Vinklerová P, Ovesná P, Hausnerová J, et al. External validation study of endometrial cancer preoperative risk stratification model (ENDORISK). Front Oncol. 2022;12:939226.

22. Soydal Ç, Varlı B, Araz M, Bakırarar B, Taşkın S, Ortaç UF. Radiomics analysis of uterine tumors in 18F-fluorodeoxyglucose positron emission tomography for prediction of lymph node metastases in endometrial carcinoma. Turk J Med Sci. 2022;52(3):762-769.

23. Otani S, Himoto Y, Nishio M, et al. Radiomic machine learning for pretreatment assessment of prognostic risk factors for endometrial cancer and its effects on radiologists' decisions of deep myometrial invasion. Magn Reson Imaging. 2022;85:161-167.

24. Lu W, Chen X, Ni J, et al. A Model to Identify Candidates for Lymph Node Dissection Among Patients With High-Risk Endometrial Endometrioid Carcinoma According to Mayo Criteria. Front Oncol. 2022;12:895834.

25. Liu XF, Yan BC, Li Y, Ma FH, Qiang JW. Radiomics Nomogram in Assisting Lymphadenectomy Decisions by Predicting Lymph Node Metastasis in Early-Stage Endometrial Cancer. Front Oncol. 2022;12:894918.

26. Liu XF, Yan BC, Li Y, Ma FH, Qiang JW. Radiomics feature as a preoperative predictive of lymphovascular invasion in early-stage endometrial cancer: A multicenter study. Front Oncol. 2022;12:966529.

27. Liro M, Śniadecki M, Wycinka E, et al. Incorporation of Tumor-Free Distance and Other Alternative Ultrasound Biomarkers into a Myometrial Invasion-Based Model Better Predicts Lymph Node Metastasis in Endometrial Cancer: Evidence and Future Prospects. Diagnostics (Basel). 2022;12(11).

28. Hsiao YY, Fu HC, Wu CH, et al. Quantitative Measurement of Progesterone Receptor Immunohistochemical Expression to Predict Lymph Node Metastasis in Endometrial Cancer. Diagnostics (Basel). 2022;12(4).

29. Guo X, Lin C, Zhao J, Tang M. Development of a novel predictive model for lymph node metastasis in patients with endometrial endometrioid carcinoma. BMC Cancer. 2022;22(1):1333.

30. Guani B, Gaillard T, Teo-Fortin LA, et al. Estimation risk of lymph nodal invasion in patients with early-stage cervical cancer: Cervical cancer application. Front Oncol. 2022;12:935628.

31. Grube M, Reijnen C, Lucas PJF, et al. Improved preoperative risk stratification in endometrial carcinoma patients: external validation of the ENDORISK Bayesian network model in a large population-based case series. J Cancer Res Clin Oncol. 2023;149(7):3361-3369.

32. Bo J, Jia H, Zhang Y, et al. Preoperative Prediction Value of Pelvic Lymph Node Metastasis of Endometrial Cancer: Combining of ADC Value and Radiomics Features of the Primary Lesion and Clinical Parameters. J Oncol. 2022;2022:3335048.

33. Asami Y, Hiranuma K, Takayanagi D, et al. Predictive model for the preoperative assessment and prognostic modeling of lymph node metastasis in endometrial cancer. Sci Rep. 2022;12(1):19004.

34. Zhang K, Zhang Y, Fang X, et al. Nomograms of Combining Apparent Diffusion Coefficient Value and Radiomics for Preoperative Risk Evaluation in Endometrial Carcinoma. Front Oncol. 2021;11:705456.

35. Yang LY, Siow TY, Lin YC, et al. Computer-Aided Segmentation and Machine Learning of Integrated Clinical and Diffusion-Weighted Imaging Parameters for Predicting Lymph Node Metastasis in Endometrial Cancer. Cancers (Basel). 2021;13(6).

36. Yan BC, Li Y, Ma FH, et al. Radiologists with MRI-based radiomics aids to predict the pelvic lymph node metastasis in endometrial cancer: a multicenter study. Eur Radiol. 2021;31(1):411-422.

37. Xu Y, Zhao R. A Prediction Model of Endometrial Cancer Lesion Metastasis under Region of Interest Target Detection Algorithm. Scientific Programming. 2021;2021(1):9928842.

38. Wang Z, Zhang S, Ma Y, Li W, Tian J, Liu T. A nomogram prediction model for lymph node metastasis in endometrial cancer patients. BMC Cancer. 2021;21(1):748.

39. Liro M, Śniadecki M, Wycinka E, et al. Ultrasound Measurement of Tumor-Free Distance from the Serosal Surface as the Alternative to Measuring the Depth of Myometrial Invasion in Predicting Lymph Node Metastases in Endometrial Cancer. Diagnostics (Basel). 2021;11(8).

40. Li X, Cheng Y, Dong Y, et al. Development and validation of predictive model for lymph node metastasis in endometrial cancer: a SEER analysis. Ann Transl Med. 2021;9(7):538.

41. Lei H, Xu S, Mao X, et al. Systemic Immune-Inflammatory Index as a Predictor of Lymph Node Metastasis in Endometrial Cancer. J Inflamm Res. 2021;14:7131-7142.

42. Jiang P, Yuan R. Analysis of Factors Related to Lymph Node Metastasis in Early-Stage Type 1 Endometrial Cancer: Verifying the Clinical Value of Positive Threshold of the Immunohistochemical Parameter Ki67. Cancer Manag Res. 2021;13:6319-6328.

43. Jiang P, Huang Y, Tu Y, et al. Combining Clinicopathological Parameters and Molecular Indicators to Predict Lymph Node Metastasis in Endometrioid Type Endometrial Adenocarcinoma. Front Oncol. 2021;11:682925.

44. Zhang Y, Zhao W, Chen Z, Zhao X, Ren P, Zhu M. Establishment and evaluation of a risk-scoring system for lymph node metastasis in early-stage endometrial carcinoma: Achieving preoperative risk stratification. J Obstet Gynaecol Res. 2020;46(11):2305-2313.

45. Reijnen C, Gogou E, Visser NCM, et al. Preoperative risk stratification in endometrial cancer (ENDORISK) by a Bayesian network model: A development and validation study. PLoS Med. 2020;17(5):e1003111.

46. Eriksson LSE, Epstein E, Testa AC, et al. Ultrasound-based risk model for preoperative prediction of lymph-node metastases in women with endometrial cancer: model-development study. Ultrasound Obstet Gynecol. 2020;56(3):443-452.

47. Crivellaro C, Landoni C, Elisei F, et al. Combining positron emission tomography/computed tomography, radiomics, and sentinel lymph node mapping for nodal staging of endometrial cancer patients. Int J Gynecol Cancer. 2020;30(3):378-382.

48. Chen J, He B, Dong D, et al. Noninvasive CT radiomic model for preoperative prediction of lymph node metastasis in early cervical carcinoma. Br J Radiol. 2020;93(1108):20190558.

49. Berg HF, Ju Z, Myrvold M, et al. Development of prediction models for lymph node metastasis in endometrioid endometrial carcinoma. Br J Cancer. 2020;122(7):1014-1022.

50. Xu X, Li H, Wang S, et al. Multiplanar MRI-Based Predictive Model for Preoperative Assessment of Lymph Node Metastasis in Endometrial Cancer. Front Oncol. 2019;9:1007.

51. Meydanli MM, Aslan K, Oz M, Muftuoglu KH, Yalcin I, Engin-Ustun Y. A novel multivariable prediction model for lymphatic dissemination in endometrioid endometrial cancer: The lymph node Metastasis Risk Index. Eur J Obstet Gynecol Reprod Biol. 2019;240:310-315.

52. Kan Y, Dong D, Zhang Y, et al. Radiomic signature as a predictive factor for lymph node metastasis in early-stage cervical cancer. J Magn Reson Imaging. 2019;49(1):304-310.

53. Günakan E, Atan S, Haberal AN, Küçükyıldız İ A, Gökçe E, Ayhan A. A novel prediction method for lymph node involvement in endometrial cancer: machine learning. Int J Gynecol Cancer. 2019;29(2):320-324.

54. Dong Y, Cheng Y, Tian W, et al. An Externally Validated Nomogram for Predicting Lymph Node Metastasis of Presumed Stage I and II Endometrial Cancer. Front Oncol. 2019;9:1218.

55. De Bernardi E, Buda A, Guerra L, et al. Radiomics of the primary tumour as a tool to improve (18)F-FDG-PET sensitivity in detecting nodal metastases in endometrial cancer. EJNMMI Res. 2018;8(1):86.

56. Teixeira AM, Ribeiro R, Schmeler KM, Herzog TJ, Nicolau SM, Marques RM. A preoperative and intraoperative scoring system to predict nodal metastasis in endometrial cancer. Int J Gynaecol Obstet. 2017;137(1):78-85.

57. Taşkın S, Şükür YE, Varlı B, et al. Nomogram with potential clinical use to predict lymph node metastasis in endometrial cancer patients diagnosed incidentally by postoperative pathological assessment. Arch Gynecol Obstet. 2017;296(4):803-809.

58. Yang B, Shan B, Xue X, et al. Predicting Lymph Node Metastasis in Endometrial Cancer Using Serum CA125 Combined with Immunohistochemical Markers PR and Ki67, and a Comparison with Other Prediction Models. PLoS One. 2016;11(5):e0155145.

59. Pollom EL, Conklin CM, von Eyben R, Folkins AK, Kidd EA. Nomogram to Predict Risk of Lymph Node Metastases in Patients With Endometrioid Endometrial Cancer. Int J Gynecol Pathol. 2016;35(5):395-401.

60. Koskas M, Fournier M, Vanderstraeten A, et al. Evaluation of models to predict lymph node metastasis in endometrial cancer: A multicentre study. Eur J Cancer. 2016;61:52-60.

61. Koskas M, Luton D, Graesslin O, et al. Direct Comparison of Logistic Regression and Recursive Partitioning to Predict Lymph Node Metastasis in Endometrial Cancer. Int J Gynecol Cancer. 2015;25(6):1037-1043.

62. Bendifallah S, Canlorbe G, Raimond E, et al. External validation of nomograms designed to predict lymphatic dissemination in patients with early-stage endometrioid endometrial cancer: a multicenter study. Am J Obstet Gynecol. 2015;212(1):56.e51-57.

63. Bendifallah S, Canlorbe G, Laas E, et al. A Predictive Model Using Histopathologic Characteristics of Early-Stage Type 1 Endometrial Cancer to Identify Patients at High Risk for Lymph Node Metastasis. Ann Surg Oncol. 2015;22(13):4224-4232.

64. Koskas M, Genin AS, Graesslin O, et al. Evaluation of a method of predicting lymph node metastasis in endometrial cancer based on five pre-operative characteristics. Eur J Obstet Gynecol Reprod Biol. 2014;172:115-119.

65. Kang S, Lee JM, Lee JK, et al. A Web-based nomogram predicting para-aortic nodal metastasis in incompletely staged patients with endometrial cancer: a Korean Multicenter Study. Int J Gynecol Cancer. 2014;24(3):513-519.

66. Luomaranta A, Leminen A, Loukovaara M. Prediction of lymph node and distant metastasis in patients with endometrial carcinoma: a new model based on demographics, biochemical factors, and tumor histology. Gynecol Oncol. 2013;129(1):28-32.

67. Koskas M, Chereau E, Ballester M, et al. Accuracy of a nomogram for prediction of lymph-node metastasis detected with conventional histopathology and ultrastaging in endometrial cancer. Br J Cancer. 2013;108(6):1267-1272.

68. Benoit MF, Ward KK. Uterine Cancer Normogram to Predict Lymph Node Metastasis: Comparison to the Mayo Algorithm and an External Vali ation of a Model in a North American Population. European Journal of Gynaecological Oncology. 2020;41(5):681-684.

69. Bollineni VR, Ytre-Hauge S, Bollineni-Balabay O, Salvesen HB, Haldorsen IS. High Diagnostic Value of 18F-FDG PET/CT in Endometrial Cancer: Systematic Review and Meta-Analysis of the Literature. J Nucl Med. 2016;57(6):879-885.

70. Hu J, Zhang K, Yan Y, Zang Y, Wang Y, Xue F. Diagnostic accuracy of preoperative (18)F-FDG PET or PET/CT in detecting pelvic and para-aortic lymph node metastasis in patients with endometrial cancer: a systematic review and meta-analysis. Arch Gynecol Obstet. 2019;300(3):519-529.

71. Bi Q, Chen Y, Wu K, et al. The Diagnostic Value of MRI for Preoperative Staging in Patients with Endometrial Cancer: A Meta-Analysis. Acad Radiol. 2020;27(7):960-968.

72. Luomaranta A, Leminen A, Loukovaara M. Magnetic resonance imaging in the assessment of high-risk features of endometrial carcinoma: a meta-analysis. Int J Gynecol Cancer. 2015;25(5):837-842.

73. Burg LC, Hengeveld EM, In 't Hout J, Bulten J, Bult P, Zusterzeel PLM. Ultrastaging methods of sentinel lymph nodes in endometrial cancer - a systematic review. Int J Gynecol Cancer. 2021;31(5):744-753.

74. Mukherjee J, Sharma R, Dutta P, Bhunia B. Artificial intelligence in healthcare: a mastery. Biotechnol Genet Eng Rev. 2024;40(3):1659-1708.

75. Zhang A, Wu Z, Wu E, et al. Leveraging physiology and artificial intelligence to deliver advancements in health care. Physiol Rev. 2023;103(4):2423-2450.

76. Gandaglia G, Martini A, Ploussard G, et al. External Validation of the 2019 Briganti Nomogram for the Identification of Prostate Cancer Patients Who Should Be Considered for an Extended Pelvic Lymph Node Dissection. Eur Urol. 2020;78(2):138-142.

77. Reijnen C, IntHout J, Massuger L, 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(9):e880-e890.

78. Fu R, Zhang D, Yu X, Zhang H. The association of tumor diameter with lymph node metastasis and recurrence in patients with endometrial cancer: a systematic review and meta-analysis. Transl Cancer Res. 2022;11(11):4159-4177.