RESEARCH

Open Access

Peritoneal cytology predicting distant metastasis in uterine carcinosarcoma: machine learning model development and validation

Lin Qiaoming^{1,2}, Guan Qi^{1,2}, Chen Danru^{1,2}, Li Lilan^{1,2} and Lin Yibin^{1,2*}

Abstract

Objective This study develops and validates a machine learning model using peritoneal cytology to predict distant metastasis in uterine carcinosarcoma, aiding clinical decision-making.

Methods This study utilized detailed clinical data and peritoneal cytology findings from uterine carcinosarcoma patients in the SEER database. Eight machine learning algorithms—Logistic Regression, SVM, GBM, Neural Network, RandomForest, KNN, AdaBoost, and LightGBM—were applied to predict distant metastasis. Model performance was assessed using AUC, calibration curves, DCA, confusion matrices, sensitivity, and specificity. The Logistic Regression model was visualized with a nomogram, and its results were analyzed. SHAP values were used to interpret the best-performing machine learning model.

Results Peritoneal cytology, T stage, age, and tumor size were key factors influencing distant metastasis in uterine carcinosarcoma patients. Peritoneal cytology had significant weight in the prediction models. The logistic regression model demonstrated excellent predictive performance with an AUC of 0.882 in the training set and 0.881 in the internal test set. The model was visualized and interpreted using a nomogram. In comprehensive evaluations, GBM was identified as the best-performing model and was explained using SHAP values. Additionally, calibration and DCA curves indicated that both models have significant potential clinical utility.

Conclusion This study introduces the first effective tool for predicting distant metastasis in uterine carcinosarcoma patients by integrating peritoneal cytology features into model construction. It aids in early identification of high-risk patients, enhancing follow-up and monitoring during tumor development, and supports the optimization of personalized treatment strategies.

Keywords Peritoneal Cytology, Distant Metastasis, Uterine Carcinosarcoma, Machine Learning Model

*Correspondence:

Lin Yibin

linyibin@fjmu.edu.cn

¹ Department of Gynecology, Clinical Oncology School of Fujian Medical University, Fujian Cancer Hospital, N0.420 Fuma Road, Fuzhou, Fujian 350014, China

² Fujian Medical University, Fuzhou, Fujian 350122, China

Introduction

Uterine Carcinosarcoma (UCS), also known as malignant mixed Müllerian tumor, is a rare gynecological malignancy with a poor prognosis, accounting for approximately 5% of all uterine tumors [1]. UCS is characterized by a high rate of lymphatic diffusion and significant tendencies for peritoneal and hematogenous metastases [2]. Literature reports that up to 30–40% of UCS patients present with lymph node metastasis at initial diagnosis, while about 10% exhibit visceral metastases, particularly pulmonary involvement [3]. Consequently, the five-year



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

survival rate for patients with locally advanced or metastatic disease typically does not exceed 10%–30% [4].

Due to its rarity, specific treatment guidelines for UCS are limited. The prevailing theory is the "conversion hypothesis," suggesting that UCS may originate from an endometrial tumor clone and subsequently undergo metaplastic differentiation [5]. Therefore, current management largely follows the guidelines for endometrial cancer. Despite constituting only a small fraction of endometrial cancers, UCS exhibits a higher risk of distant metastasis and recurrence, leading to poorer patient outcomes. Early identification of high-risk patients with distant metastases and implementation of targeted comprehensive treatment strategies are crucial for improving prognosis.

Cytoreductive surgery is the primary treatment for UCS patients [6, 7]. Comprehensive surgical staging, including abdominal lavage, hysterectomy, salpingooophorectomy, and lymphadenectomy, is recommended for all operable patients [8]. Peritoneal cytology, which involves analyzing exfoliated cancer cells from intraoperative peritoneal lavage fluid or aspiration samples, helps detect free cancer cells in the peritoneal cavity. Preoperative fine needle aspiration can also provide quick and safe cytological assessments with minimal patient discomfort. This technique identifies minimal metastatic lesions not yet visible as masses or nodules, enabling early detection of potential peritoneal metastases beyond the uterus. Early detection of subclinical metastases, often overlooked in imaging studies, is particularly valuable.

Initially, positive peritoneal cytology was classified as stage IIIA under the International Federation of Gynecology and Obstetrics(FIGO) 1988 staging criteria for endometrial cancer. However, in 2009, FIGO revised its guidelines [9] to exclude peritoneal cytology from the staging system due to controversies regarding its prognostic significance [10, 11]. This change resulted in a decline in peritoneal cytology sampling during hysterectomies between 2010 and 2017 [12]. Despite this, several international authorities, including the European Society of Medical Oncology (ESMO), European Society of Gynaecological Oncology (ESGO), European Society for Radiotherapy & Oncology (ESTRO), Japanese Society of Gynecologic Oncology (JSGO), National Comprehensive Cancer Network (NCCN), and American Joint Committee on Cancer (AJCC), continue to support collecting peritoneal cytology samples during surgery and including them in pathology reports [13–15]. Similarly, the FIGO Gynecologic Oncology Committee recommends collecting peritoneal cytology samples, even though it has been removed from formal staging criteria, emphasizing that "positive cytology must be reported separately without affecting staging" [16]. Additionally, the 2021 ESGO/

ESTRO/ESP guidelines highlight that malignant peritoneal cytology is associated with lower survival rates [17]. Recent studies have further shown that positive peritoneal cytology has significant prognostic implications, particularly for non-endometrioid types of endometrial cancer [18–20]. Thus, although peritoneal cytology is not used for formal staging, its presence provides important information for evaluating disease progression and patient prognosis, especially in specific types of endometrial cancer, serving as a valuable supplement to staging.

This study aims to develop and validate a predictive model that integrates peritoneal cytology findings to forecast the development of distant metastasis in UCS patients. By combining peritoneal cytology with other clinicopathological features, we seek to provide clinicians with an effective tool for identifying high-risk populations, optimizing medical resource allocation, and supporting personalized treatment strategies.Additionally, given the current lack of international consensus on the role of peritoneal cytology in UCS, our work may provide valuable data to support future updates to diagnostic and treatment guidelines.

Materials and methods

Data preparation

Data for this study were sourced from Surveillance, Epidemiology, and End Results Program(SEER) public database, utilizing SEER*Stat software version 8.4.4 for data extraction. Our analysis centered on patients diagnosed with UCS across 17 Registries between 2000 and 2021, with data submission in November 2023 [21]. Cases were screened using the "site code ICD-O-3 / WHO 2008", specifying the uterus as the site of origin (codes C54.0-C54.9, C55.9), and identified by the malignant tissue type defined as carcinosarcoma according to the "ICD-O-3 Histology/Behavior" codes (8950/3, 8951/3, 8980/3, 8981/3).For each patient, demographic and clinical variables were extracted, encompassing age, race, marital status, median family income, rural-urban continuum code, time to diagnosis, interval from diagnosis to treatment initiation, tumor dimensions, peritoneal cytology results, histological grade and TNM staging (AJCC 7th edition). Assessment of distant metastasis was conducted using the combinatorial staging data within the SEER database. Inclusion criteria comprised: histopathologically confirmed disease, single primary tumors, and diagnoses made between 2010 and 2021. Patients were excluded if they had missing data regarding peritoneal cytology and distant metastasis. Ultimately, a cohort of 3,434 endometrial cancer patients met the criteria for detailed analysis. This study adheres to the Declaration of Helsinki principles. Given that the SEER data are de-identified and

available for research purposes, local ethics committee approval was not required.

Data processing

In this study, any data entries with more than 35% missing parameters were excluded from the analysis. The remaining features underwent preprocessing through multiple imputation (MI) facilitated by a multi-classification regression model [22, 23]. Patient data were randomly partitioned into a training set and an internal test set at a ratio of 7:3, where the former was utilized for model development and the latter served for validation and evaluation. Continuous variables are reported as mean ± standard deviation (SD) if normally distributed, as assessed by the Kolmogorov-Smirnov test, or as median (interquartile range) if not, with comparisons made using the t-test or Mann-Whitney U test, respectively. Categorical variables are summarized as counts and frequencies, with comparisons conducted via chisquare or Fisher's exact tests. All statistical analyses were two-tailed, with P < 0.05 denoting statistical significance.To address the imbalance in the dataset due to the low incidence of distant metastases, we applied two techniques during the machine learning phase: resampling and weighted processing. To address class imbalance, we used the SMOTENC method [24] to increase minority class samples in the training set and retained original data for the test set to assess model generalization. For weighting, we assigned higher weights to minority class samples and lower weights to majority ones, based on the inverse of their proportions, [25] ensuring the model's attention to the minority class during training while allowing unbiased evaluation on an untouched test set. Both approaches were compared against untreated data post-modeling to evaluate their effectiveness.

Factor screening

To evaluate the correlations among features in the training set, we applied Spearman correlation analysis, with a correlation coefficient threshold set at 0.7. A coefficient below this threshold suggests an absence of significant multicollinearity among characteristic variables. Initially, we generated a correlation heatmap based on the Spearman coefficients to visualize the degree of association between each pair of variables. To delve deeper into the structural relationships among these variables, we conducted cluster analysis using[1-abs(spearman_ cor)] as the distance measure. This approach allows for an equal emphasis on both negative and positive correlations, ensuring a balanced evaluation of variable similarity. The clustering results were represented as a dendrogram, highlighting the hierarchical structure of feature relationships.When two or more features exhibit high correlation, they tend to provide similar information. Including all such highly correlated features in a predictive model can unnecessarily increase model complexity without significantly enhancing its performance.Following this, we used the occurrence of distant metastasis as the outcome variable and performed univariate analysis to identify predictors significantly associated with distant metastasis (P < 0.05) within the training set. The variables that met this significance threshold were subsequently included in multivariate logistic regression analysis. The final feature set for the machine learning model was then determined based on the results of this multivariate logistic regression analysis (P < 0.05), ensuring only the most relevant predictors were selected.

Model construction and evaluation

Based on the feature selection methodology described above, we constructed eight distinct machine learning algorithms to develop predictive models for distant metastasis in UCS, including Logistic Regression, Support Vector Machine (SVM), Gradient Boosting Machine (GBM), Neural Network (NeuralNet), RandomForest, K-Nearest Neighbors (KNN), AdaBoost, and LightGBM. Logistic Regression, a linear model suitable for linearly separable features, offers simplicity and interpretability; SVM maximizes the margin between classes by identifying the optimal hyperplane, providing robust classification especially in highdimensional spaces; [26] GBM iteratively builds weak classifiers to enhance predictive power, capturing nonlinear relationships and interaction effects while optimizing the loss function via gradient descent; [27] Neural Networks emulate the human brain's structure through multiple layers of neurons, enabling the learning of complex features and modeling of nonlinear relationships; Random Forest integrates multiple decision trees to improve stability and accuracy, reducing overfitting and enhancing generalization; [28] KNN, an instance-based learning method, predicts categories by calculating distances between new samples and existing ones; [29] AdaBoost improves predictive ability by iteratively adjusting sample weights, focusing more on misclassified instances; [30] and LightGBM, an efficient gradient boosting framework, accelerates model training using histograms and feature parallelization [31]. During the training process, these algorithms underwent tenfold cross-validation on the training set data to obtain a robust estimate of model performance. In evaluating the performance of the predictive models, we adopted a comprehensive assessment system designed to thoroughly measure

both the classification ability and clinical utility of the models. The primary evaluation tool was the area under the receiver operating characteristic curve (ROC-AUC), which assessed the model's overall discrimination capability. Additionally, calibration curves were used to verify the accuracy of the predicted probabilities, ensuring that the model's predictions were well-calibrated. Decision Curve Analysis (DCA) evaluated the practical application value of the model from the perspective of clinical benefit, providing insights into its real-world applicability. Moreover, we utilized confusion matrices to transparently present the model prediction results, determining the optimal threshold for the test set based on model accuracy. Various performance metrics were computed, including accuracy, sensitivity, specificity, precision, and F1 score, to offer detailed insights into the model's classification performance. This comprehensive approach ensured a thorough understanding of the model's effectiveness and reliability in clinical settings.

Accuracy =
$$\frac{TP + TN}{TP + TN + FP + FN}$$

Sensitivity = $\frac{TP}{TP + FN}$
Specificity = $\frac{TN}{TN + FP}$
Precision = $\frac{TP}{TP + FP}$

F1 Score =
$$2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity}$$

These are based on four basic values: true positives (True Positive, TP), false positives (False Positive, FP), true negatives (True Negative, TN), and false negatives (False Negative, FN).

Model interpretation

To enhance the usability and interpretability of our logistic regression model, we utilized a nomogram for intuitive visualization and result interpretation. This tool not only simplifies model application but also clarifies the contribution of each covariate to the overall prediction score, providing an easy method to estimate individual patient probabilities of distant metastasis. The odds ratio (Odds = P / (1-P)) derived from the nomogram reflects the likelihood of distant metastasis relative to no occurrence.Based on comprehensive

Table 1	Peritoneal	cytology	and	distant	metastasis	grouped
chi-squa	re test					

Distant	Peritoneal	X2	Р		
metastases	Negative	Suspicious	Positive		
Yes	203	33	441	123.45	0.0001
No	2248	94	415		

performance evaluations, the GBM model was identified as the optimal predictor for distant metastasis in UCS patients, with features ranked by their importance. SHapley Additive exPlanations(SHAP) plots were used to visualize feature contributions, allowing for quantitative analysis of each variable's impact on distant metastasis risk [32]. Positive SHAP values indicate risk factors, while negative values suggest protective factors.We provided personalized explanations for two randomly selected patients regarding their likelihood of developing distant metastases based on model predictions.

All statistical analyses were conducted using R software (version 4.4.1), and the corresponding analysis code is available upon request from the authors.

Result

Baseline information and correlation analysis

In this study, we analyzed data from a total of 3,434 UCS patients to investigate the relationship between peritoneal cytology findings and distant metastasis, revealing a significant association ($\chi^2 = 123.45$, p < 0.0001) (Table 1)., suggesting that peritoneal cytology may serve as an Important reference indicators for primary clinical screening. Then, patients were divided into a training set and an internal test set in a 7:3 ratio, with baseline characteristics summarized in Table 2; in the training set, statistically significant differences (p < 0.05) were observed between patients with and without distant metastasis regarding peritoneal cytology, differentiation grade, T stage, N stage, time from diagnosis to treatment, tumor size, and age (Table 3). Figure 1 presents the Spearman correlation analysis of various feature metrics in the training set, where a darker color indicates a higher correlation, with a threshold of 0.7 suggesting strong associations (Fig. 1A),

The results of the hierarchical clustering indicate that there is no significant multicollinearity among the feature variables in the training set (Fig. 1B).

Univariate and multivariable logistic regression

In univariate logistic regression (LR) analysis, positive peritoneal cytology, differentiation grade G3, T stage, N stage, and tumor size were identified as risk factors

Features	Training set(N=2	2404)	Test set(<i>N</i> =1030)		
	0 (<i>N</i> =1930)	1 (<i>N</i> =474)	0 (N=827)	1 (<i>N</i> =203)	
Peritoneal cytology					
Negative	1562 (80.9%)	143 (30.2%)	686 (83.0)	60 (29.6)	
Suspicious	75 (3.9%)	21 (4.4%)	19 (2.3)	12 (5.9)	
Positive	293 (15.2%)	310 (65.4%)	122 (14.8)	131 (64.5)	
Year of diagnosis					
2010-2013	595 (30.8%)	145 (30.6%)	253 (30.6)	73 (36.0)	
2014-2017	642 (33.3%)	171 (36.1%)	289 (34.9)	55 (27.1)	
2018-1021	693 (35.9%)	158 (33.3%)	285 (34.5)	75 (36.9)	
Race					
Whites	1284 (66.5%)	308 (65%)	545 (65.9)	121 (59.6)	
Blacks	455 (23.6%)	117 (24.7%)	193 (23.3)	59 (29.1)	
Asian or Pacific Islander	178 (9.2%)	46 (9.7%)	79 (9.6)	21 (10.3)	
Native American/Alaska Native	13 (0.7%)	3 (0.6%)	10 (1.2)	2 (1.0)	
Grade					
G1	76 (3.9%)	4 (0.8%)	22 (2.7)	3 (1.5)	
G2	91 (4.7%)	10 (2.1%)	36 (4.4)	6 (3.0)	
G3	1763 (91.3%)	460 (97%)	769 (93.0)	194 (95.6)	
T-7th					
T1/T2	1517 (78.6%)	72 (15.2%)	657 (79.4%)	38 (18.7%)	
ТЗ/Т4	413 (21.4%)	402 (84.8%)	170 (20.6%)	165 (81.3%)	
N-7th					
NO	1473 (76.3%)	255 (53.8%)	650 (78.6)	114 (56.2)	
N1	281 (14.6%)	120 (25.3%)	114 (13.8)	47 (23.2)	
N2	176 (9.1%)	99 (20.9%)	63 (7.6)	42 (20.7)	
Marital status					
married	1002 (51.9%)	269 (56.8%)	403 (48.7%)	115 (56.7%)	
others	928 (48.1%)	205 (43.2%)	424 (51.3%)	88 (43.3%)	
Median household income					
<50000	126 (6.5%)	39 (8.2%)	48 (5.8)	15 (7.4)	
50000-90000	1169 (60.6%)	291 (61.4%)	505 (61.1)	125 (61.6)	
>90000	635 (32.9%)	144 (30.4%)	274 (33.1)	63 (31.0)	
Urban-rural coding					
An urban area with a population of 1 million or more	1137 (58.9%)	290 (61.2%)	506 (61.2)	119 (58.6)	
Metropolitan areas with a population of 25-1 million	444 (23%)	102 (21.5%)	203 (24.5)	48 (23.6)	
Metropolitan areas with a population of less than 250,000	144 (7.5%)	31 (6.5%)	45 (5.4)	14 (6.9)	
Non-metropolitan area and adjacent	120 (6.2%)	31 (6.5%)	47 (5.7)	15 (7.4)	
Non-metropolitan area and not adjacent	85 (4.4%)	20 (4.2%)	26 (3.1)	7 (3.4)	
Time from diagnosis to treatment (day)	28 (15-42)	23 (8-36)	28(16-42)	18(4-33.5)	
Tumor size	57 (40-80)	75 (55-100)	60(40-81.5)	75(50-110)	
Age	67 (61-73)	65 (59-72)	67 (61-74)	65 (58-72)	

for distant metastasis in UCS patients (all odds ratios [OR] > 1, 95% confidence intervals [CI] > 1, p < 0.05). Conversely, age and time from diagnosis to treatment were found to be protective factors (OR and 95% CI < 1, p < 0.05). Multivariate LR analysis further revealed that

positive peritoneal cytology, T stage, and tumor size remained independent risk factors for distant metastasis (OR and 95% CI>1, p < 0.05), while age was confirmed as an independent protective factor (OR and 95% CI<1, p < 0.05) (Table 4).

Table 3 Difference analysis of the training set

Features	0 (<i>N</i> =1930)	1 (<i>N</i> =474)	Р
Peritoneal cytology			<0.001
Negative	1562 (80.9%)	143 (30.2%)	
Suspicious	75 (3.9%)	21 (4.4%)	
Positive	293 (15.2%)	310 (65.4%)	
Year of diagnosis			0.447
2010-2013	595 (30.8%)	145 (30.6%)	
2014-2017	642 (33.3%)	171 (36.1%)	
2018-1021	693 (35.9%)	158 (33.3%)	
Race			0.932
Whites	1284 (66.5%)	308 (65%)	
Blacks	455 (23.6%)	117 (24.7%)	
Asian or Pacific Islander	178 (9.2%)	46 (9.7%)	
Native American/Alaska Native	13 (0.7%)	3 (0.6%)	
Grade			< 0.001
G1	76 (3.9%)	4 (0.8%)	
G2	91 (4.7%)	10 (2.1%)	
G3	1763 (91.3%)	460 (97%)	
T-7th			< 0.001
Т1/Т2	1517 (78.6%)	72 (15.2%)	
Т3/Т4	413 (21.4%)	402 (84.8%)	
N-7th			< 0.001
NO	1473 (76.3%)	255 (53.8%)	
N1	281 (14.6%)	120 (25.3%)	
N2	176 (9.1%)	99 (20.9%)	
Marital status			0.066
married	1002 (51.9%)	269 (56.8%)	
others	928 (48.1%)	205 (43.2%)	
Median household income			0.303
<50000	126 (6.5%)	39 (8.2%)	
50000-90000	1169 (60.6%)	291 (61.4%)	
>90000	635 (32.9%)	144 (30.4%)	
Urban-rural coding			0.872
An urban area with a population of 1 million or more	1137 (58.9%)	290 (61.2%)	
Metropolitan areas with a population of 25-1 million	444 (23%)	102 (21.5%)	
Metropolitan areas with a population of less than 250,000	144 (7.5%)	31 (6.5%)	
Non-metropolitan area and adjacent	120 (6.2%)	31 (6.5%)	
Non-metropolitan area and not adjacent	85 (4.4%)	20 (4.2%)	
Time from diagnosis to treatment (day)	28 (15-42)	23 (8-36)	< 0.001
Tumor size	57 (40-80)	75 (55-100)	< 0.001
Age	67 (61-73)	65 (59-72)	< 0.001

Model building and performance evaluation

In constructing a machine learning model to predict distant metastasis, we selected key features based on multifactor logistic regression (LR), including peritoneal cytology status, T stage, tumor size, and age. We enhanced the performance of multiple classification algorithms using tenfold cross-validation. Results demonstrated excellent predictive power across all models, with area under the receiver operating characteristic curve (ROC-AUC) values exceeding 0.7 (Fig. 2). Calibration curves (Fig. 3), decision curve analysis (DCA) (Fig. 4), accuracy, recall (sensitivity) (Table 5), and other metrics also performed well, indicating the effectiveness of these models for predicting



Fig. 1 Multicollinearity: (A) The relationship between the feature indicators in the training set through Spearman correlation analysis, with a strong correlation between variables greater than 0.7. B Hierarchical cluster analysis of [1 - abs(spearman_cor)] as a distance metric, and a strong correlation between variables greater than 0.3

distant metastasis.Notably, while the Random Forest model showed strong performance on the training set, significant overfitting was observed in the internal test set, leading us to exclude it as the best candidate. Using AUC as the primary evaluation criterion, Logistic Regression achieved the highest AUC values of 0.882 on the training set and 0.881 on the internal test set, demonstrating robust discrimination. In handling imbalanced datasets, the F1 score is more critical than ROC-AUC. A "baseline model" that always predicts distant metastasis achieved an F1 score of 0.332. Our models significantly outperformed this baseline, with GBM and AdaBoost showing superior F1 scores compared to Logistic Regression. However, due to AdaBoost's lower recall rate, which could lead to missed diagnoses, we selected the Gradient Boosting Machine (GBM) model as optimal. GBM achieved an F1 score of 0.630, demonstrating superior generalization and clinical applicability. In conclusion, the GBM model was chosen as the optimal model. Feature importance rankings (Fig. 5) highlighted peritoneal cytology as a critical feature for improving model performance in both LR and GBM models. To further validate the model's performance and ensure transparency, we constructed confusion

matrices for both models (Fig. 6). Comparisons revealed that the GBM model performed better at distinguishing cases of distant metastases.

In a "baseline model" that always predicts distant metastasis, the precision is approximately 0.197. Although our developed models significantly outperform this baseline model, to further optimize precision, we thoroughly investigated the impact of data imbalance on model performance and adopted two main strategies: "SMOTE-NC for Synthetic Sampling" and "Adjusting Sample Weights". Comparing these approaches using the GBM model, our results are summarized in Table 6. While these techniques improved model accuracy on the internal test set, they did so at the expense of sensitivity (recall), which is crucial in clinical settings. For a preliminary screening tool aimed at identifying UCS patients at risk for distant metastasis, missing actual cases of distant metastasis is clinically more serious than reducing overall precision. Therefore, we opted to use the GBM model trained on unbalanced data as the optimal model for practical application, ensuring higher sensitivity and minimizing missed diagnoses. This decision was partly due to our sufficiently large sample size, which helped ensure the accuracy of our results.

Table 4 Univariate logistic regression analysis and multivariate logistic regression analysis

			Univariate analysis	sis Multivariate analy		alysis
Features	0 (<i>N</i> =1930)	1 (<i>N</i> =474)	OR (95%CI)	Р	OR (95%CI)	Р
Peritoneal cytology						
Negative	1562 (80.9%)	143 (30.2%)				
Suspicious	75 (3.9%)	21 (4.4%)	3.06 (1.83-5.11)	<i>P<</i> .001	2.27 (1.25-4.15)	P=.007
Positive	293 (15.2%)	310 (65.4%)	11.56 (9.14-14.61)	<i>P</i> <.001	5.25 (4.00-6.91)	P<.001
Year of diagnosis						
2010-2013	595 (30.8%)	145 (30.6%)				
2014-2017	642 (33.3%)	171 (36.1%)	1.09 (0.85-1.40)	P=.482		
2018-1021	693 (35.9%)	158 (33.3%)	0.94 (0.73-1.20)	P=.602		
Race						
Whites	1284 (66.5%)	308 (65%)				
Blacks	455 (23.6%)	117 (24.7%)	1.07 (0.84-1.36)	P=.567		
Asian or Pacific Islander	178 (9.2%)	46 (9.7%)	1.08 (0.76-1.52)	P=.674		
Native American/Alaska Native	13 (0.7%)	3 (0.6%)	0.96 (0.27-3.40)	P=.952		
Grade						
G1	76 (3.9%)	4 (0.8%)				
G2	91 (4.7%)	10 (2.1%)	2.09 (0.63-6.92)	P=.229	1.22 (0.30-4.90)	P=.783
G3	1763 (91.3%)	460 (97%)	4.96 (1.80-13.62)	P=.002	3.08 (0.97-9.78)	P=.057
T-7th						
T1/T2	1517 (78.6%)	72 (15.2%)				
Т3/Т4	413 (21.4%)	402 (84.8%)	20.51 (15.60-26.96)	<i>P</i> <.001	9.77 (7.24-13.19)	P<.001
N-7th						
NO	1473 (76.3%)	255 (53.8%)				
N1	281 (14.6%)	120 (25.3%)	2.47 (1.92-3.17)	<i>P</i> <.001	1.19 (0.86-1.64)	P=.288
N2	176 (9.1%)	99 (20.9%)	3.25 (2.46-4.30)	<i>P</i> <.001	1.20 (0.84-1.71)	P=.306
Marital status						
married	1002 (51.9%)	269 (56.8%)				
others	928 (48.1%)	205 (43.2%)	0.82 (0.67-1.01)	P=.059		
Median household income						
<50000	126 (6.5%)	39 (8.2%)				
50000-90000	1169 (60.6%)	291 (61.4%)	0.80 (0.55-1.18)	P=.263		
>90000	635 (32.9%)	144 (30.4%)	0.73 (0.49-1.10)	P=.130		
Urban-rural coding						
An urban area with a population of 1 million or more	1137 (58.9%)	290 (61.2%)				
Metropolitan areas with a population of 25-1 million	444 (23%)	102 (21.5%)	0.90 (0.70-1.16)	P=.414		
Metropolitan areas with a population of less than 250,000	144 (7.5%)	31 (6.5%)	0.84 (0.56-1.27)	<i>P</i> =.416		
Non-metropolitan area and adjacent	120 (6.2%)	31 (6.5%)	1.01 (0.67-1.53)	P=.952		
Non-metropolitan area and not adjacent	85 (4.4%)	20 (4.2%)	0.92 (0.56-1.53)	P=.754		
Time from diagnosis to treatment (day)	28 (15-42)	23 (8-36)	0.99 (0.99-1.00)	P=.001	1.00 (0.99-1.00)	P=.087
Tumor size	57 (40-80)	75 (55-100)	1.01 (1.01-1.02)	<i>P</i> <.001	1.01 (1.00-1.01)	P<.001
Age	67 (61-73)	65 (59-72)	0.98 (0.97-0.99)	<i>P</i> <.001	0.98 (0.96-0.99)	P<.001

Model interpretability

During the model interpretability processing phase, the logistic regression model was interpreted and applied using Nomogram plots (Fig. 7).The optimal GBM model,

we constructed a feature importance ranking based on SHAP values, revealing that in UCS patients, the key factors influencing distant metastasis were T stage, peritoneal cytology, tumor size, and age; in the visualization

Evaluation of the perf	formance of the train	ing set				
Model	Threshold	Accuracy	Sensitivity	Specificity	Precision	F1
Logistic	0.113	0.768	0.914	0.733	0.456	0.609
SVM	0.146	0.779	0.882	0.754	0.468	0.612
GBM	0.189	0.814	0.890	0.796	0.517	0.654
NeuralNetwork	0.239	0.817	0.831	0.813	0.522	0.641
RandomForest	0.500	0.972	0.867	0.998	0.990	0.925
KNN	0.202	0.831	0.975	0.796	0.540	0.695
Adaboost	0.500	0.852	0.622	0.908	0.625	0.624
LightGBM	0.166	0.791	0.905	0.763	0.484	0.631
Test set performance	evaluation					
Model	Threshold	Accuracy	Sensitivity	Specificity	Precision	F1
Logistic	0.113	0.758	0.872	0.730	0.442	0.587
SVM	0.146	0.779	0.862	0.758	0.467	0.606
GBM	0.189	0.805	0.842	0.796	0.503	0.630
NeuralNetwork	0.239	0.806	0.788	0.810	0.505	0.615
RandomForest	0.500	0.816	0.458	0.903	0.538	0.495
KNN	0.202	0.793	0.833	0.784	0.486	0.613
Adaboost	0.500	0.860	0.606	0.923	0.658	0.631
LightGBM	0.166	0.769	0.808	0.759	0.452	0.580







(Fig. 8B), yellow indicates risk factors for distant metastasis, and purple highlights protective factors. To further elucidate the model's predictions, we randomly selected two groups of UCS patients, one at high risk and another at low risk of distant metastasis, with Fig. 8A illustrating

the low-risk group, exemplified by a patient aged 67 years with a tumor size of 55 mm, T stage beyond the uterus (T3/T4), and negative peritoneal cytology, where negative peritoneal cytology served as an important protective factor against distant metastasis, and Fig. 8C



Fig. 3 Calibration curves for the training set (A) and the test set (B)







depicting the high-risk group, characterized by a patient aged 62 years with a tumor size of 70 mm, T stage beyond the uterus (T3/T4), and positive peritoneal cytology, which was identified as a critical risk factor for distant metastasis.

Discussion

Uterine carcinosarcoma (UCS) is a highly malignant gynecological tumor characterized by complex and aggressive biology, with a propensity for early and distant metastases. Currently, there is a lack of reliable



Fig. 5 Ranking of feature importance for logistic regression model (A) and GBM model (B)

indicators or prediction models to assess the risk of distant metastasis in UCS patients. This study successfully developed and validated a predictive model that incorporates peritoneal cytology features to evaluate distant metastasis risk in UCS patients. The model exhibited strong discrimination and calibration capabilities, highlighting its potential as a valuable clinical tool.

In our study, peritoneal cytology emerged as a critical feature influencing distant metastasis in UCS patients. This is likely because UCS, being a highly malignant tumor, exhibits strong invasive and early metastatic tendencies. When peritoneal cytology results are positive, it indicates that tumor cells have acquired the ability to breach the basement membrane and enter the abdominal cavity, suggesting their invasive biological behavior and potential for distant metastasis [33, 34]. Compared to other invasive diagnostic methods such as laparoscopic biopsy, peritoneal cytology offers a relatively non-invasive approach to rapidly and safely obtain cytological results, thereby minimizing patient discomfort and complications. Unlike imaging examinations, which can be subjective and dependent on the expertise of the imaging physician, equipment quality, and the size of metastatic tumors. Peritoneal cytology results are typically included in routine pathology reports, providing easily accessible and highly standardized data. This reliable data foundation is crucial for constructing and validating predictive models. Additionally, the application of genetic analysis in peritoneal cytology holds significant promise. Advanced technologies such as high-throughput sequencing, liquid biopsy, and multiomics integration offer more detailed and comprehensive information compared to traditional pathology [35]. These innovations are expected to enhance the prediction of distant metastasis and improve clinical management outcomes, while also opening new avenues for genetic research and personalized medicine.

In our current study, the results also demonstrate that T stage significantly contributes to distant metastasis in UCS patients. Specifically, diagnoses of T3 or T4 stages indicate extensive local invasion beyond the uterus, often involving lymphatic and blood vessels [36]. This facilitates tumor cells entering the circulation, thereby promoting distant metastasis. Additionally, our results show a positive correlation between tumor size and distant metastasis risk in UCS patients [37]. Larger tumors tend to exhibit higher cell proliferation rates and greater aggressiveness, [38] increasing the likelihood of local and distant spread. The hypoxic environment within growing tumors may activate pro-metastatic signaling pathways, [39] further enhancing metastatic potential. These findings underscore the need for more aggressive evaluation and management strategies for high-risk patients, including wider surgical resection and postoperative adjuvant treatments.Interestingly, we observed an inverse but weak relationship between age and distant metastasis in UCS patients. In certain specific tumor types, younger patients' tumors exhibited unique biological properties [40-42] such as higher proliferation rates, greater invasiveness, and metastatic ability, [43, 44] potentially due to specific gene mutations or



Fig. 6 Confusion matrices for logistic regression models (A: training set, B: test set) and GBM models (C: training set, D: test set)

GBM		ROC-AUC	Threshold	Accuracy	Sensitivity	Specificity	Precision	F1
No treatment	Train	0.906	0.189	0.814	0.89	0.796	0.517	0.654
	Test	0.874	0.189	0.805	0.842	0.796	0.503	0.63
Resampling	Train	0.927	0.529	0.859	0.875	0.844	0.848	0.861
	Test	0.872	0.529	0.825	0.768	0.839	0.54	0.634
Weighted	Train	0.911	0.525	0.829	0.88	0.816	0.54	0.669
	Test	0.871	0.525	0.81	0.818	0.808	0.511	0.629

 Table 6
 Comparison of before and after processing of data imbalances



regulatory mechanisms [45]. This suggests that molecular typing distribution and associated biological behaviors may vary with age, highlighting the importance of understanding age-related differences in UCS biology. The 2023 FIGO update on endometrial cancer staging and molecular typing reflects deeper insights into the complexity and potential biological behavior of these tumors [46]. Further research is needed to explore agerelated biological differences and their implications for clinical management.

In our current study, some limitations must be acknowledged. First, this is a retrospective analysis that poses challenges such as data quality issues, information bias and selection bias, so future studies should adopt a prospective design to overcome these limitations and provide more reliable data support. Second, the performance of our machine learning model may be affected by geographical and hospital differences, and patient characteristics and treatment patterns may vary significantly between regions and medical institutions. To ensure the robustness and generalization ability of the model, we need more institutions to participate in the external validation. To this end, in follow-up studies, we plan to incorporate a more diverse and broader multicenter dataset for validation and testing of models, aiming to address potential variations in model performance across different populations and healthcare settings. In addition, it is an important direction for future research to explore the mechanism of occurrence and development through peritoneal cytology examination combined with genomics and proteomics, and to look for more specific and sensitive predictors.

In conclusion, based on the large-scale multicenter data set, our prediction model provides new ideas and technical support for distant metastasis prediction of USC patients. In practice, the prediction results of this model can help develop personalized follow-up plan, especially for those patients predicted to be at high risk of metastasis, suggesting more frequent and targeted monitoring measures to ensure early detection of potential problems and timely intervention treatment. This not only improves the quality of life of patients, but also enables the more reasonable allocation of medical resources, and realizes the maximum efficiency of medical services.



Fig. 8 SHAP plot of the GBM model: Figure B uses the GBM machine learning model to rank the importance of the characteristic variables of distant metastasis in UCS patients. In these graphs, yellow represents a variable as a risk factor for distant transfer, while purple represents a variable as a protective factor. Figure A shows the distribution of SHAP values in the low-risk group and C

Conclusion

This study introduces the first effective tool for predicting distant metastasis in uterine carcinosarcoma patients by integrating peritoneal cytology features into model construction. It aids in early identification of high-risk patients, enhancing follow-up and monitoring during tumor development, and supports the optimization of personalized treatment strategies.

Abbreviations

UCS	Uterine Carcinosarcoma
SEER	Surveillance, Epidemiology, and End Results Program
SMOTENC	Synthetic Minority Over-sampling Technique for Nominal and
	Continuous
SVM	Support Vector Machine
GBM	Gradient Boosting Machine
KNN	K-Nearest Neighbors
LR	Logistic regression
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
DCA	Decision Curve Analysis
SHAP	SHapley Additive exPlanations
FIGO	International Federation of Gynecology and Obstetrics
ESMO	European Society of Medical Oncology

European Society of Gynaecological Oncology
European Society for Radiotherapy & Oncology
Japanese Society of Gynecologic Oncology
National Comprehensive Cancer Network
American Joint Committee on Cancer
Standard deviation

Acknowledgements

We extend our gratitude to the participants and researchers involved in the SEER database and all related open data and studies, whose contributions have been instrumental to our research.

Conflict of interest

The authors declare no competing interests.

Authors' contributions

All authors made substantial contributions to the interpretation of data and critically revised the manuscript. All authors approved the final version submitted and agreed to be accountable for their contributions. Specific contributions are as follows: Qiaoming Lin: Writing—original draft, review & editing, data analysis, visualization, methodology, conceptualization. Qi Guan: Writing—original draft, methodology, investigation, Lilan Li:formal analysis, original draft.Yibin Lin: Writing—review & editing, validation, supervision, project administration, data curation, conceptualization. All authors contributed to the revision of the paper, interpreted the results, read, and approved the final manuscript. The submitting author reports no conflicts of interest.

Funding

This study was funded by Fujian Provincial Science and Technology Innovation Joint Fund Project(No. 2023Y9454), Fujian Provincial Natural Science Foundation Project(No. 2024J011087) and Fujian Provincial Health Commission Science and Technology Plan Project(No. 2024CXA032).

Data availability

Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate

This study utilized publicly available abstracted data, and therefore did not require additional ethical approval.

Competing interests

The authors declare no competing interests.

Received: 29 December 2024 Accepted: 23 March 2025 Published online: 26 April 2025

References

- 1. Matsuo K, Ross MS, Machida H, Blake EA, Roman LD. Trends of uterine carcinosarcoma in the United States. J Gynecol Oncol. 2018;29(2):e22.
- Pradhan TS, Stevens EE, Ablavsky M, Salame G, Lee YC, Abulafia O. FIGO staging for carcinosarcoma: can the revised staging system predict overall survival? Gynecol Oncol. 2011;123(2):221–4.
- Ravishankar P, Smith DA, Avril S, Kikano E, Ramaiya NH. Uterine carcinosarcoma: a primer for radiologists. Abdom Radiol (NY). 2019;44(8):2874–85.
- Bansal N, Herzog TJ, Seshan VE, Schiff PB, Burke WM, Cohen CJ, Wright JD. Uterine carcinosarcomas and grade 3 endometrioid cancers: evidence for distinct tumor behavior. Obstet Gynecol. 2008;112(1):64–70.
- de Jong RA, Nijman HW, Wijbrandi TF, Reyners AK, Boezen HM, Hollema H. Molecular markers and clinical behavior of uterine carcinosarcomas: focus on the epithelial tumor component. Mod Pathol. 2011;24(10):1368–79.
- Gracia M, Yildirim Y, Macuks R, Mancari R, Achimas-Cadariu P, Polterauer S, Iacoponi S, Zapardiel I. SARCUT Study Group Influence of Clinical and Surgical Factors on Uterine Carcinosarcoma Survival. Cancers (Basel). 2023;15(5):1463.
- Koh WJ, Abu-Rustum NR, Bean S, Bradley K, Campos SM, Cho KR, Chon HS, Chu C, Cohn D, Crispens MA, Damast S, Dorigo O, Eifel PJ, Fisher CM, Frederick P, Gaffney DK, George S, Han E, Higgins S, Huh WK, Lurain JR 3rd, Mariani A, Mutch D, Nagel C, Nekhlyudov L, Fader AN, Remmenga SW, Reynolds RK, Tillmanns T, Ueda S, Wyse E, Yashar CM, McMillian NR, Scavone JL. Uterine Neoplasms, Version 1.2018, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2018;16(2):170–99.
- 8. Rungruang B, Olawaiye AB. Comprehensive surgical staging for endometrial cancer. Rev Obstet Gynecol. 2012;5(1):28–34.
- 9. Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. Int J Gynaecol Obstet. 2009;105(2):103–4.
- Tebeu PM, Popowski Y, Verkooijen HM, Bouchardy C, Ludicke F, Usel M, Major AL. Positive peritoneal cytology in early-stage endometrial cancer does not influence prognosis. Br J Cancer. 2004;91(4):720–4.
- Saga Y, Imai M, Jobo T, Kuramoto H, Takahashi K, Konno R, Ohwada M, Suzuki M. Is peritoneal cytology a prognostic factor of endometrial cancer confined to the uterus? Gynecol Oncol. 2006;103(1):277–80.
- Matsuo K, Klar M, Harter P, Miller H, Nusbaum DJ, Matsuzaki S, Roman LD, Wright JD. Trends in peritoneal cytology evaluation at hysterectomy for endometrial cancer in the United States. Gynecol Oncol. 2021;161(3):710–9.
- National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guidelines in Oncology. Uterine Neoplasms Version 1 2022. Fort Washington, PA: NCCN; 2021. Available from: https://www.nccn.org/professionals/ physician_gls/pdf/uterine.pdf.

- 14. Colombo N, Creutzberg C, Amant F, et al. ESMO-ESGO-ESTRO Consensus Conference On Endometrial Cancer: diagnosis, treatment and follow-up. Ann Oncol. 2016;27(1):16–41.
- Yamagami W, Mikami M, Nagase S, et al. Japan Society of Gynecologic Oncology 2018 guidelines for treatment of uterine body neoplasms. J Gynecol Oncol. 2020;31(1):e18.
- Amant F, Mirza MR, Koskas M, Creutzberg CL. Cancer of the corpus uteri. Int J Gynaecol Obstet. 2018;143(Suppl 2):37–50.
- Concin N, Matias-Guiu X, Vergote I, et al. ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma. Int J Gynecol Cancer. 2021;31(1):12–39.
- Matsuo K, Matsuzaki S, Nusbaum DJ, et al. Significance of Malignant Peritoneal Cytology on Survival of Women with Uterine Sarcoma. Ann Surg Oncol. 2021;28(3):1740–8.
- Matsuo K, Nusbaum DJ, Matsuzaki S, et al. Malignant peritoneal cytology and increased mortality risk in stage I non-endometrioid endometrial cancer. Gynecol Oncol. 2020;159(1):43–51.
- 20. Sakai K, Yamagami W, Takahashi F, et al. Prognostic impact of peritoneal cytology on treating endometrial cancer using data from the Japan Society of Obstetrics and Gynecology cancer registry. J Gynecol Oncol. 2024;36:e41.
- 21. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer. cancer.gov) SEER*Stat Database: Incidence - SEER Research Data, 17 Registries, Nov 2023 Sub (2000-2021) - Linked To County Attributes - Time Dependent (1990-2022) Income/Rurality, 1969-2022 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2024, based on the November 2023 submission.
- 22. Rui F, Yeo YH, Xu L, Zheng Q, Xu X, Ni W, Tan Y, Zeng QL, He Z, Tian X, Xue Q, Qiu Y, Zhu C, Ding W, Wang J, Huang R, Xu Y, Chen Y, Fan J, Fan Z, Qi X, Huang DQ, Xie Q, Shi J, Wu C, Li J. Development of a machine learning-based model to predict hepatic inflammation in chronic hepatitis B patients with concurrent hepatic steatosis: a cohort study. EClinicalMedicine. 2024;16(68):102419.
- Barakat MS, Field M, Ghose A, et al. The effect of imputing missing clinical attribute values on training lung cancer survival prediction model performance. Health Inf Sci Syst. 2017;5(1):16.
- 24. Fonseca J, Bacao F. Geometric SMOTE for imbalanced datasets with nominal and continuous features. Expert Systems with Applications. 2023;234:121053.
- Yang K, Yu Z, Chen CLP, Cao W, You J, Wong H-S. Incremental Weighted Ensemble Broad Learning System for Imbalanced Data. IEEE Transactions on Knowledge and Data Engineering. 2022;34(12):5809–24.
- Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics Proteomics. 2018;15(1):41–51.
- Dash TK, Chakraborty C, Mahapatra S, Panda G. Gradient boosting machine and efficient combination of features for speech-based detection of COVID-19. IEEE J Biomed Health Inform. 2022;26(11):5364–71.
- Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
 Lu X, Chen Y, Zhang G, Zeng X, Lai L, Qu C. Application of interpret-
- able machine learning algorithms to predict acute kidney injury in patients with cerebral infarction in ICU. J Stroke Cerebrovasc Dis. 2024;33(7):107729.
- Natras R, Soja B, Schmidt M. Ensemble Machine Learning of Random Forest, AdaBoost and XGBoost for Vertical Total Electron Content Forecasting. Remote Sens. 2022;14:3547.
- Chae M, Yoon H, Lee H, Choi J. Hearing Recovery Prediction for Patients with Chronic Otitis Media Who Underwent Canal-Wall-Down Mastoidectomy. J Clin Med. 2024;13(6):1557.
- Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput Methods Progr Biomed. 2022;214:106584.
- Takenaka M, Kamii M, Iida Y, Yanaihara N, Suzuki J, Takahashi K, Yanagida S, Saito M, Takano H, Yamada K, Okamoto A. Re-thinking the prognostic significance of positive peritoneal cytology in endometrial cancer. Gynecol Oncol. 2021;161(1):135–42.
- Matsuo K, Matsuzaki S, Miller H, et al. Clinico-pathological significance of suspicious peritoneal cytology in endometrial cancer. J Surg Oncol. 2021;124(4):687–98.
- Villiger AS, Zurbriggen S, Imboden S, Solass W, Christe L, Saner FAM, Gmür A, Rau TT, Mueller MD, Siegenthaler F. Reviving peritoneal cytology:

Exploring its role in endometrial cancer molecular classification. Gynecol Oncol. 2024;182:148–55.

- 36. Jónsdóttir B, Marcickiewicz J, Borgfeldt C, Bjurberg M, Dahm-Kähler P, Flöter-Rådestad A, Hellman K, Holmberg E, Kjølhede P, Rosenberg P, Tholander B, Åvall-Lundqvist E, Stålberg K, Högberg T. Preoperative and intraoperative assessment of myometrial invasion in endometrial cancer-A Swedish Gynecologic Cancer Group (SweGCG) study. Acta Obstet Gynecol Scand. 2021;100(8):1526–33.
- Gracia M, Yildirim Y, Macuks R, Mancari R, Achimas-Cadariu P, Polterauer S, Iacoponi S, Zapardiel I. SARCUT Study Group Influence of Clinical and Surgical Factors on Uterine Carcinosarcoma Survival. Cancers (Basel). 2023;15(5):1463.
- Gao J, Ao Y, Wang S, Chen Z, Zhang Y, Ding J, Jiang J. WHO histological classification and tumor size are predictors of the locally aggressive behavior of thymic epithelial tumors. Lung Cancer. 2024;187:107446.
- Baek S, Yu SE, Deng YH, Lee YJ, Lee DG, Kim S, Yoon S, Kim HS, Park J, Lee CH, Lee JB, Kong HJ, Kang SG, Shin YM, Sung HJ. Quenching Epigenetic Drug Resistance Using Antihypoxic Microparticles in Glioblastoma Patient-Derived Chips. Adv Healthc Mater. 2022;11(8):e2102226.
- Chen MT, Sun HF, Zhao Y, Fu WY, Yang LP, Gao SP, Li LD, Jiang HL, Jin W. Comparison of patterns and prognosis among distant metastatic breast cancer patients by age groups: a SEER population-based analysis. Sci Rep. 2017;7(1):9254.
- Huang H, Xu S, Wang X, Liu S, Liu J. Patient Age Is Significantly Related to Distant Metastasis of Papillary Thyroid Microcarcinoma. Front Endocrinol (Lausanne). 2021;12:748238.
- Purushotham A, Shamil E, Cariati M, Agbaje O, Muhidin A, Gillett C, Mera A, Sivanadiyan K, Harries M, Sullivan R, Pinder SE, Garmo H, Holmberg L. Age at diagnosis and distant metastasis in breast cancer – A surprising inverse relationship. European Journal of Cancer. 2014;50(10):1697–705.
- Azim HA Jr, Nguyen B, Brohée S, Zoppoli G, Sotiriou C. Genomic aberrations in young and elderly breast cancer patients. BMC Med. 2015;13:266.
- Buza N, Baine I, Hui P. Precision genotyping diagnosis of lung tumors with trophoblastic morphology in young women. Mod Pathol. 2019;32(9):1271–80.
- Cuicui ZHAO, Hong LIU. Molecular Biological Pathogenesis of Young Breast Cancer[J]. Cancer Research on Prevention and Treatment. 2020;47(3):213–7.
- 46. Matias-Guiu X, Lax S, Raspollini MR, Palacios J, Zheng W, Liu C, de Brot L, Lordello L, Hardisson D, Gaffney D, Mutch D, Scambia G, Creutzberg CL, Fotopoulou C, Berek JS, Concin N. FIGO 2023 staging for endometrial cancer, when, if it is not now? Eur J Cancer. 2024;213:115115.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.