

Machine learning–based dynamic prediction model for postoperative malnutrition risk in esophageal cancer patients A retrospective study

Mengyao Zhu

the Air Force Medical University

Fei'e Jiang

the Second Affiliated Hospital of Air Force Medical University

Wenchen Wang

the Second Affiliated Hospital of Air Force Medical University

Yidan bai

the Air Force Medical University

Xiaoying Shi

the Air Force Hospital of Western Theater Command

Xiangni Su

suxiangni001@163.com

the Air Force Medical University

Research Article

Keywords: Machine learning, Prediction model, Esophageal cancer, Malnutrition

Posted Date: September 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10947928/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Machine learning–based dynamic prediction model for postoperative malnutrition risk in esophageal cancer patients A retrospective study

*Zhu Mengyao^{1,2†}, Jiang Fei'e^{1,3}, Wang Wenchen³, Bai Yidan¹, Shi Xiaoying⁵, Su Xiangni^{1,4} **

1. Department of Nursing, the Air Force Medical University, Xi'an 710032, China, 2. School of Nursing, Shaanxi University of Chinese Medicine, Xianyang, 712046, China, 3. Department of Thoracic Surgery, the Second Affiliated Hospital of Air Force Medical University, Xi'an 710032, China, 4. Department of Military Medical Innovation Center, the Air Force Medical University, Xi'an 710032, China, 5. Department of Aircrew, the Air Force Hospital of Western Theater Command, Chengdu 500643, China.

*Corresponding author: Su Xiangni; Email: suxiangni001@163.com;

Abstract

Background: Malnutrition is the most common postoperative complication associated with esophageal cancer (EC), which severely impacts patient survival outcomes. However, early predictive strategies for postoperative malnutrition in EC patients remain underdeveloped.

Objective: This study aimed to develop a dynamic predictive model for postoperative malnutrition in EC patients, to facilitate early clinical identification and timely intervention of at-risk individuals.

Methods: In this study, 805 EC patients were enrolled as the training and validation set. Additionally, 101 EC patients were enrolled as the external validation set. Baseline data, pathological features, and biochemical indicators were collected at 1, 3, and 6 months postoperatively. LASSO regression was used to screen baseline characteristics at 1 month postoperatively. Four machine learning models (Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest) were constructed and compared via 10-fold cross-validation to select the optimal one, which was then externally validated at 3 and 6 months. Model performance was evaluated using accuracy, sensitivity, specificity, and AUC, and SHAP was applied to interpret feature importance.

Results: The prediction model constructed using Random Forest demonstrated favorable

1. The Project Supported by Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2026JC-YBMS-0894)

predictive performance. The AUC value of the model at 1 month postoperatively was 0.974 (0.9491-1), and the AUC values with 95% confidence intervals (CI) for external validation at 3 and 6 months postoperatively were 0.814 (0.726-0.900) and 0.801 (0.714-0.889). Additionally, feature importance analysis revealed that BMI below 18.5 kg/m², weight loss >3 kg within 3 months, and age >70 years were associated with a higher risk of malnutrition.

Conclusions: This study developed a postoperative malnutrition risk prediction model for EC patients based on Random Forest, which accurately predicts malnutrition risk at 3 and 6 months postoperatively, thereby providing auxiliary decision-making support for early clinical identification.

Keywords: Machine learning; Prediction model; Esophageal cancer; Malnutrition

Introduction

Based on GLOBOCAN 2022 data, esophageal cancer (EC) is the 11th most frequently diagnosed malignancy and the 7th deadliest cancer globally, responsible for 2.6% of incident cases and 4.6% of cancer fatalities [1-2]. China account of 2.6% of all new 48.3% of the global incidence and mortality of EC, making it one of the highest-risk regions for disease [3]. This is closely associated with the unique anatomical structure of the esophagus and the nature of its treatment. On the one hand, the esophageal anatomy predisposes tumors to grow outward along the wall, causing luminal compression, on the other hand, radical esophagectomy as the standard curative approach entails subsequent resection and reconstruction that inevitably inflict additional structural damage. The combined effects of these factors contribute to long-term stenosis and functional impairment, which in turn compromise oral intake and nutrient absorption, ultimately leading to malnutrition [4-5].

EC is associated with the most severe malnutrition among all cancers, ranking first in the incidence of cancer-related malnutrition. Studies show that 60-80% of EC patients suffer from malnutrition, with approximately 22% of deaths attributable to malnutrition and related complications [6-8]. Malnutrition has a severe adverse effect on disease-related economic burden, short-term clinical outcomes, postoperative recovery, and long-term prognosis among patients with EC. Hence, a dynamic prediction model for continuous risk monitoring is urgently needed.

The NRS 2002 [9] is a common nutritional screening tool, but its simplified indicators limit its capacity to capture the dynamic nutritional changes and interactions in postoperative EC patients. In contrast, machine learning (ML) can learn complex nonlinear relationships from high-dimensional data, uncovering risk factors and their interactions [10], thereby improving prediction and enabling early risk detection [11]. Several studies have confirmed that ML-based algorithms achieve high predictive accuracy for malnutrition in cancer patients, providing a reliable basis for assisted nutritional assessment [12,13]. Moreover, SHAP quantifies feature contributions, reinforcing both model credibility and clinical utility [14]. However, ML-based prediction models for postoperative malnutrition in EC patients are still scarce.

Therefore, the study aims to develop a dynamic prediction model for malnutrition risk in EC patients using ML algorithms.

Methods

Study design and participants

In this study, 805 EC patients who met the inclusion and exclusion criteria were retrospectively enrolled from a tertiary hospital in Xi'an between January 2020 and December 2025 as the training and validation set (n=805). Additionally, 101 EC patients who met the same criteria were retrospectively enrolled between January 2026 and May 2026 as the external validation set (n=101).

Inclusion criteria:

(1) Meets the diagnostic criteria for esophageal cancer according to the 2020 edition of the Guidelines for the Diagnosis and Treatment of Esophageal Cancer; (2) Patients who have undergone radical resection for esophageal cancer; (3) Age $\geq$ 18 years; (4) No prior history of chemotherapy or surgery for esophageal cancer; (5) No systemic infectious or autoimmune diseases before surgery.

Exclusion criteria:

(1) Patients with treatment interruption greater than 6 months; (2) Patients with other concomitant malignancies; (3) Patients with metastasis; (4) Patients with major organ damage involving the heart, liver, kidneys, or other vital organs; (5) Patients with incomplete clinical case data.

Predictors

Data collected in this study included general patient information (age, sex, marital status, number of children, smoking, drinking, comorbidities, and body weight), pathological data (tumor location, type, stage, size, surgical type, nutritional support, adjuvant therapy, neoadjuvant chemotherapy dose, postoperative bleeding volume, and complications), and laboratory indicators (hemoglobin, red blood cells, platelets, neutrophils, lymphocytes, white blood cells, albumin, total protein, creatinine, and blood urea nitrogen).

Clinical Outcome

The total score of NRS2002 consists of three components: nutritional status score (0–3 points), disease severity score (0–3 points), and age score (1 point for age $\geq$ 70). The nutritional status score is determined based on the degree of weight loss and recent food intake, taking the highest score: 0 points (normal), 1 point (weight loss $>5\%$ within 3 months or food intake 50%–75% of normal), 2 points (weight loss $>5\%$ within 2 months, or BMI 18.5–20.5 with poor general condition, or food intake 25%–50%), 3 points (weight loss $>5\%$ within 1 month or $>15\%$ within 3 months, or BMI <18.5 with poor general condition, or food intake $<25\%$). The disease severity score similarly takes the highest score: 0 points (normal requirements), 1 point (chronic diseases such as tumors, fractures, diabetes, etc.), 2 points (severe diseases requiring bed rest

such as stroke, severe pneumonia, hematologic malignancies), 3 points (ICU with mechanical ventilation such as brain injury or bone marrow transplantation). Age ≥ 70 is scored as 1 point. A total score ≥ 3 indicates nutritional risk [11]. In the present study, we used the internationally validated NRS2002 with a cutoff of 3 points to define nutritional risk, and all outcomes were evaluated using the same criteria in both the internal and external validation datasets to ensure comparability and reliability.

Sample Size

To ensure an adequate sample size for prediction model development, we estimated the required sample size using the `pmsampsize` package. The calculation was based on an expected C-statistic of 0.8, 14 candidate predictor variables, and an estimated malnutrition prevalence of 86% in the target population. The minimum required sample size was 801 cases. After adjusting for a 15% invalid response rate (dividing by 0.85), the final estimated sample size was 942 cases. The total number of eligible cases enrolled ($N = 906$) met the EPV criterion (≥ 10 events per variable), ensuring sufficient events for reliable model development.

Missing Data

Records with $>30\%$ missing values were excluded from the analysis. Missing data among the remaining records were imputed using the `mice` package (Multivariate Imputation by Chained Equations) in R software (R Foundation for Statistical Computing, R4.5.2). The predictive mean matching (PMM) method was applied for imputation, generating five imputed datasets with a maximum of 10 iterations. The ID and time point variables were not included as predictors in the imputation model. All analytical variables were incorporated into the imputation model to ensure data consistency and minimize bias.

Modeling and Interpretation

Based on univariate analyses, we initially selected 13 candidate variables ($P < 0.05$). LASSO regression was then applied to further screen these variables. The optimal penalty parameter λ was determined using 10-fold cross-validation based on the minimum criterion ($\lambda_{\min} = 0.0321$). Ultimately, 10 variables with non-zero regression coefficients were identified as important predictors closely associated with the outcome variable, aiming to enhance the model's predictive performance and generalizability [15].

Machine Learning Models

Four machine learning algorithms—logistic regression (LR) [16], support vector machine (SVM) [17], decision tree (DT) [18], and random forest (RF) [19]—were selected to construct binary classification prediction models. These algorithms each offer distinct advantages in handling imbalanced data, model interpretability, and robustness. These four models each possess distinct advantages in machine learning applications. LR yields predictions endowed

with probabilistic meaning and offers a high degree of interpretability that renders it particularly suitable for fields such as medicine and finance wherein explicit quantification of variable effects is required. SVM leverages kernel functions to handle nonlinear problems with elegance and demonstrates strong generalization capability on high-dimensional small-sample datasets and exhibits distinct advantages in scenarios where feature dimensionality substantially surpasses sample size while concurrently mitigating overfitting effectively. DT is capable of processing both numerical and categorical features simultaneously and its decision pathways are intuitive and transparent and amenable to visualization which facilitates review by domain experts and it also possesses a certain degree of tolerance for missing data. RF employs bagging and random feature subspace selection to markedly reduce overfitting and exhibits commendable robustness to noise and outliers and automatically ranks feature importance and sustains stable high-accuracy performance in the context of high-dimensional large-scale data.

Model Performance Evaluation

First, a comprehensive set of evaluation metrics was used to assess the performance of the four machine learning models, including accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), Kappa statistic, F1 score, and the area under the receiver operating characteristic curve (AUC). Second, hyperparameters were optimized using grid search combined with 10-fold cross-validation, and model performance was comprehensively evaluated using the average AUC, accuracy, sensitivity, and specificity [20]. The model development process is shown in **Figure 1**.

Model Output

Shapley Additive Explanations (SHAP) [21] values were used to measure feature importance by attributing the model's output to individual feature factors, thereby explaining the impact of each feature on the model's predictions (**Equation 1**).

$$g(x) = \phi_0 + \sum_{j=1}^M \phi_j$$

Among them, ϕ_0 is the average predicted value over all samples, referred to as the baseline value; ϕ_j is the SHAP value of feature j ; and M is the total number of features. When ϕ_j is greater than zero, the feature increases the predicted value and exerts a positive effect. Based on the SHAP values, the direction and magnitude of each feature's influence can be obtained. Furthermore, we used SHAP to provide individualized explanations for the predictions, quantifying feature contributions and directions of influence. This was intended to enhance model interpretability and facilitate the understanding of results in a clinical context.

Ethical Considerations

This study has been approved by the Ethics Committee of Tangdu Hospital, Air Force Medical

University, and the ethical approval document (approval number: K-HG-202510-08) has been obtained.

Statistical methods

Statistical analyses were performed using R4.5.2. Continuous variables were first tested for normality using the Shapiro–Wilk test. Data conforming to a normal distribution were expressed as mean $\pm$ standard deviation ($\bar{x} \pm s$); data not conforming to a normal distribution were expressed as median (interquartile range) $M[P_{25}, P_{75}]$. Categorical variables were expressed as frequencies (percentages). All tests were two-sided, and $P < 0.05$ was deemed statistically significant. Variable screening and regression analyses were conducted using R. Candidate variables were first evaluated by univariate logistic regression, and those with $P < 0.05$ were entered into LASSO regression, wherein the optimal penalty parameter λ was selected via 10-fold cross-validation based on the minimum mean squared error criterion, retaining variables with non-zero coefficients. The selected variables were subsequently incorporated into four models using the Enter method. Model performance evaluation, including receiver operating characteristic (ROC) curve analysis and area under the curve (AUC) calculation, as well as graphical visualizations (including four-quadrant plots and line charts), were all performed in RStudio. All model development and internal/external validation procedures were carried out in RStudio.

Software and Tools

Statistical analyses were performed using R software (version 4.5.2) with RStudio as the integrated development environment. The main R packages employed covered multiple stages of the analytical pipeline. For data preprocessing, **tidyr** was used for data tidying and reshaping. Four classification algorithms—logistic regression, support vector machine, decision tree, and random forest—were implemented via **glm** (stats package), **svm** (e1071), **rpart**, and **randomForest** (or ranger), respectively. The **caret** package facilitated end-to-end machine learning workflows, including data splitting, resampling, hyperparameter tuning, and performance evaluation. Model discrimination was assessed using **pROC** to generate ROC curves and compute AUC. Visualization was performed with **ggplot2**, supplemented by **gridExtra** and **scales** for advanced plotting and layout. For model interpretation, SHAP values were computed and visualized using **kernelshap** and **shapviz** to quantify feature contributions. All code is provided in **Supplementary Table 1**.

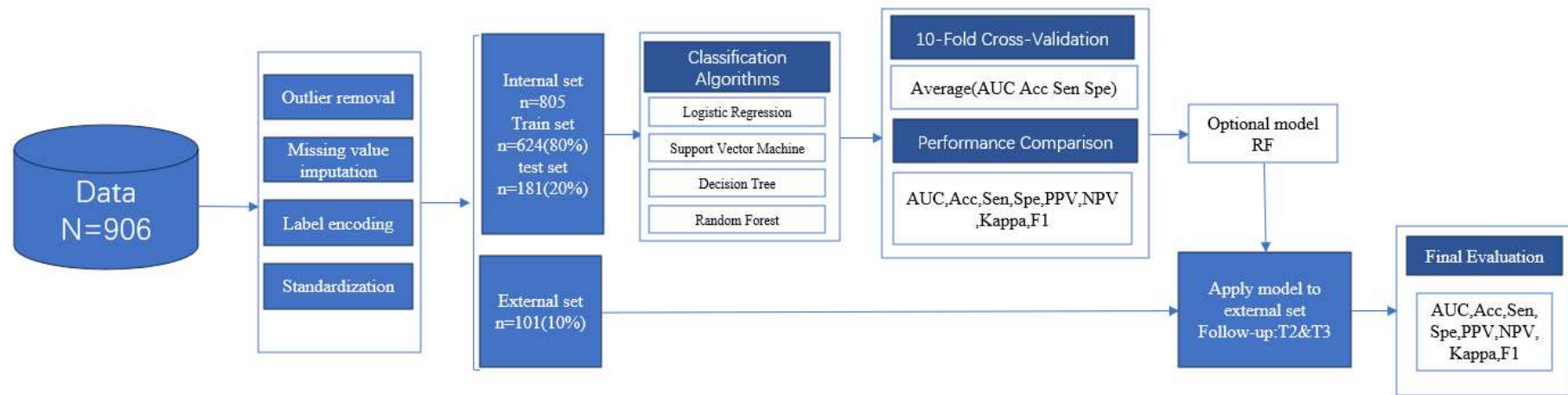


Figure 1. Study flowchart. Flowchart of machine learning model development, internal validation, and external validation. Data were cleaned and standardized, then split into training, internal test, and external validation sets. Four classifiers were compared; Random Forest (RF) was chosen by 10-fold CV and finally assessed on the internal test set and external follow-up (T2, T3).

Results

Participant characteristics

This study enrolled 906 patients, who were allocated to an internal development cohort (n=805) and an external validation cohort (n=101) based on time point or study center. A comparison of baseline characteristics between the two datasets is presented in Table 1.

Table 1. Baseline clinical characteristics of the training/validation and external validation sets.

Characteristic	Training set(n=805) n (%); Median [Q1, Q3]	External set(n=101) n (%) [Q1, Q3]	validation n (%); Median	<i>P</i>
Malnutrition				
No	491(61.0)	46(45.5)		
Yes	314(39.0)	55(54.5)		
Ganer				0.630
Male	699(83.1%)	82(81.2%)		
Female	136(16.9%)	19(18.8%)		
Age	63.0 [57.0, 68.0]	63.0 [60.0, 69.0]		0.119
Smoking				0.335
No	342 (42.5%)	48 (47.5%)		
Yes	463 (57.5%)	53 (52.5%)		
Drinking				0.979
No	509 (63.2%)	64 (63.4%)		
Yes	296 (36.8%)	37 (36.6%)		
Histological type				0.123
Squamous Cell Carcinoma (SCC)	631 (78.4%)	88 (87.1%)		
Adenocarcinoma (AC)	160 (19.9%)	12 (11.9%)		
Adenosquamous	14 (1.7%)	1 (1.0%)		
Tumor location				0.018
upper	49 (6.1%)	13 (12.9%)		
middle	232 (28.8%)	33 (32.7%)		
lower	524 (65.1%)	55 (54.5%)		
T				<0.001
0	120 (14.9%)	37 (36.6%)		
1	158 (19.6%)	20 (19.8%)		
2	83 (10.3%)	9 (8.9%)		
3	341 (42.4%)	31 (30.7%)		
4	103 (12.8%)	4 (4.0%)		

N			0.023
0	504 (62.6%)	66 (65.3%)	
1	186 (23.1%)	19 (18.8%)	
2	73 (9.1%)	15 (14.9%)	
3	42 (5.2%)	1 (1.0%)	
Neoadjuvant chemoradiotherapy			0.283
No	30 (3.7%)	6 (5.9%)	
yes	775 (96.3%)	95 (94.1%)	
Type of surgery			<0.001
Minimally invasive	146 (18.1%)	60 (59.4%)	
Open thoracotomy	659 (81.9%)	41 (40.6%)	
Feeding tube			0.098
Nasogastric tube	384 (47.7%)	57 (56.4%)	
Nasoduodenal tube	421 (52.3%)	44 (43.6%)	
Comorbidities			0.857
No	591 (73.4%)	75 (74.3%)	
Yes	214 (26.6%)	26 (25.7%)	
Complications			0.005
No	500 (62.1%)	77 (76.2%)	
Yes	305 (37.9%)	24 (23.8%)	
Length of hospital stay	15.0 [12.0, 21.0]	14.0 [11.0, 20.0]	0.017
Postoperative month1(T1)			
No	491(61.0)	46(45.5)	0.062
Yes	314(39.0)	55(54.5)	

Selection of Predictive Factors

Univariate Analysis

Univariate logistic regression was performed to screen all candidate variables using a threshold of $P < 0.05$, ultimately identifying 13 significant variables (**Table 2**).

Table 2. Univariable analysis

Characteristic	Univariable analysis
	<i>P</i>
Enteral Nutrition	0.0003
Parenteral Nutrition	0.0005
Chemotherapeutic Agents	0.8096
Gender(M/F)	0.3906
Age	<0.0001

Weight loss within 3 months	<0.0001
Tumor Type (SCC/AC)	0.1740
Tumor Site	0.6516
T	0.3015
N	0.9768
Adjuvant Therapy	0.4927
Type of Surgery (Open/Mini)	0.0020
Feeding tube (Nasal nutrition/ Nasogastric)	0.9174
Comorbidities (N/Y)	0.0338
Complications (N/Y)	0.2619
Intraoperative Blood Loss	0.0565
Hemoglobin	<0.0001
Red Blood Cell	<0.0001
Blood Urea Nitrogen	0.3802
Creatinine	0.3337
PNI	0.0002
PLR	0.9380
NLR	0.0224
SII	0.0082
GNRI	<0.0001
BMI	<0.0001

M: male; F: Female; SCC: Squamous Cell Carcinoma; AC: Adenocarcinoma; Open: Open thoracotomy; Mini: Minimally invasive

LASSO Variable Selection

Through 10-fold cross-validation optimization, the LASSO regression identified 10 predictors for postoperative malnutrition at a lambda (λ) value of 0.0321. These predictors included enteral nutrition, parenteral nutrition, age, weight loss within 3 months, type of surgery, hemoglobin, neutrophil count, albumin, GNRI, and BMI. Of these, BMI, age, weight loss within 3 months, surgical approach, and neutrophil count demonstrated statistically significant differences between the malnutrition risk group and the non-malnutrition risk group ($P < 0.05$) (Figure 2; Table 3).

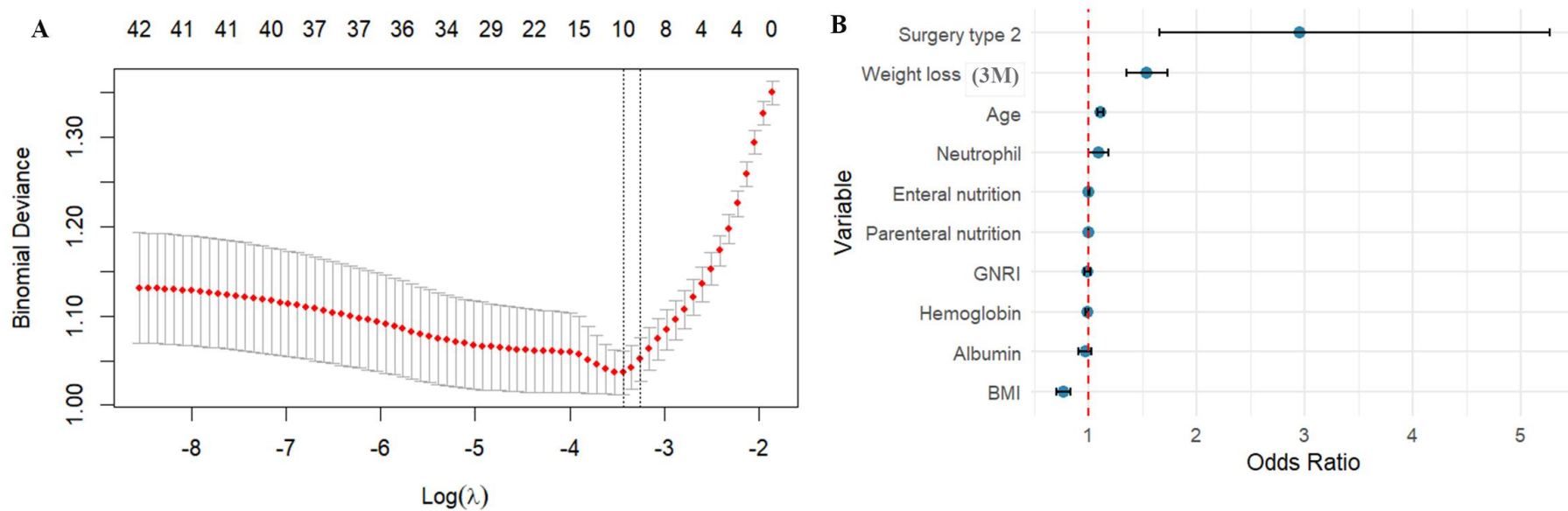


Figure 2. **LASSO Variable Selection.** (A) The optimal penalty parameter $\lambda = 0.0321$; (B) OR of LASSO-Selected Variables (95%CI)

Table 3. OR of LASSO-Selected Variables(95%CI)

Variable	OR	95%CI	<i>P</i>
Enteral Nutrition	1.0034	(0.9996-1.0072)	0.0831
Parenteral Nutrition	1.0005	(1.0001-1.0008)	0.0054
Age	1.1123	(1.0850-1.1402)	0.0000
Weight loss within 3 months	1.5340	(1.3557-1.7356)	0.0000
Open thoracotomy	2.9548	(1.6553-5.2744)	0.0002
Hemoglobin	0.9886	(0.9752-1.0022)	0.0993
Neutrophil count	1.0935	(1.0051-1.1896)	0.0376
Albumin	0.9691	(0.9104-1.0315)	0.3239
GNRI	0.9894	(0.9604-1.0193)	0.4827
BMI	0.7687	(0.7047-0.8384)	0.0000

Evaluation of Model Performance

Initially, four models (LR, SVM, DT, and RF) were built using the training set ($n = 624$) and test set ($n = 181$). Then, the performance of these models was assessed and compared by means of test-set evaluation metrics and 10-fold cross-validation. The results showed that RF achieved the highest accuracy (0.961) and the highest AUC (0.974), and was located at the outermost ring of the radar chart, indicating not only higher predictive accuracy but also better overall predictive performance. DT had accuracy (0.939) and AUC (0.943), lower than those of RF, ranking second. SVM and LR had accuracy (0.762, 0.751) and AUC (0.810, 0.815), which were lower than those of RF and DT (**Figure 3 and Table 4**). Furthermore, the study also presents the mean AUC and standard deviation (derived from fold-to-fold variation in cross-validation) of the four models. Among them, RF achieved the highest mean AUC (0.97) and the smallest standard deviation (0.018), located in the upper-right quadrant of the four-quadrant plot, indicating the best predictive performance and greatest stability. DT ranked second in mean AUC (0.92) but had a larger standard deviation (0.06), showing notable fluctuation and inferior robustness to RF. SVM (0.884 ± 0.060) and LR (0.868 ± 0.085) had lower means and moderate variability, with overall performance worse than the RF and DT. Overall, RF substantially outperformed the other models in terms of AUC and was therefore selected as the preferred model for this dataset. (**Figure 4; Table 5**).

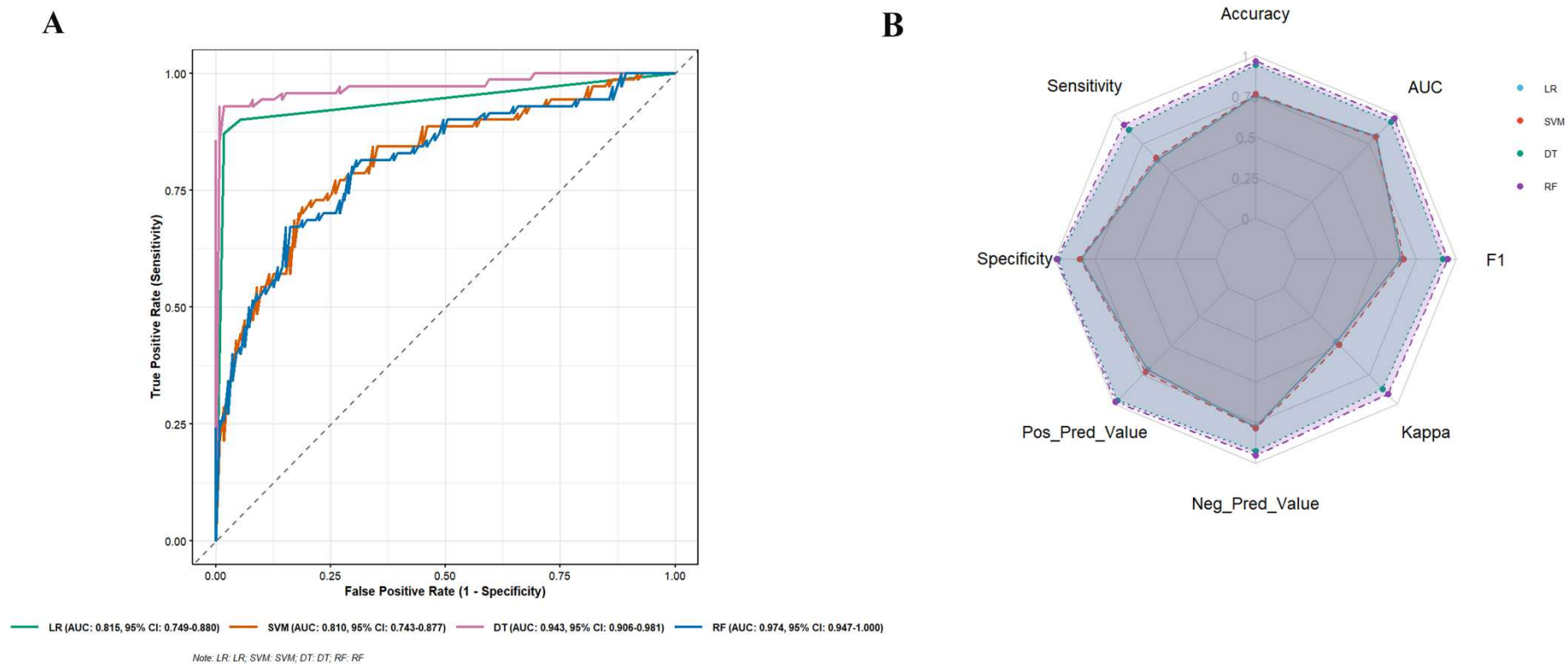


Figure 3. Evaluation of Model Performance. (A) ROC curves of the four models: LR (blue), SVM (red), DT (green), and RF (purple). (B) Radar chart of evaluation metrics: accuracy, sensitivity, specificity, PPV, NPV, kappa, and F1-score. Logistic Regression: LR, Support Vector Machine: SVM, Decision Tree: DT, Random Forest: RF

Table 4.: Evaluation Metrics of the Four Models

Evaluation metrics	LR	SVM	DT	RF
Accuracy	0.7514(0.6818,0.8125)	0.7624(0.6936-0.8524)	0.9392(0.8939-0.9693)	0.9613(0.9219-0.9843)
Sensitivity	0.6143	0.6286	0.8714	0.9143
Specificity	0.8378	0.8468	0.9820	0.9910
Pos Pred Value	0.7049	0.7213	0.9683	0.9846
Neg Pred Value	0.7750	0.7833	0.9237	0.9483
Kappa	0.4631	0.4870	0.8695	0.9174
F1	0.6565	0.6718	0.9173	0.9481
AUC	0.815(0.749-0.880)	0.810(0.743-0.877)	0.943(0.906-0.981)	0.974(0.947-1.000)

Logistic Regression: LR, Support Vector Machine: SVM, Decision Tree: DT, Random Forest: RF

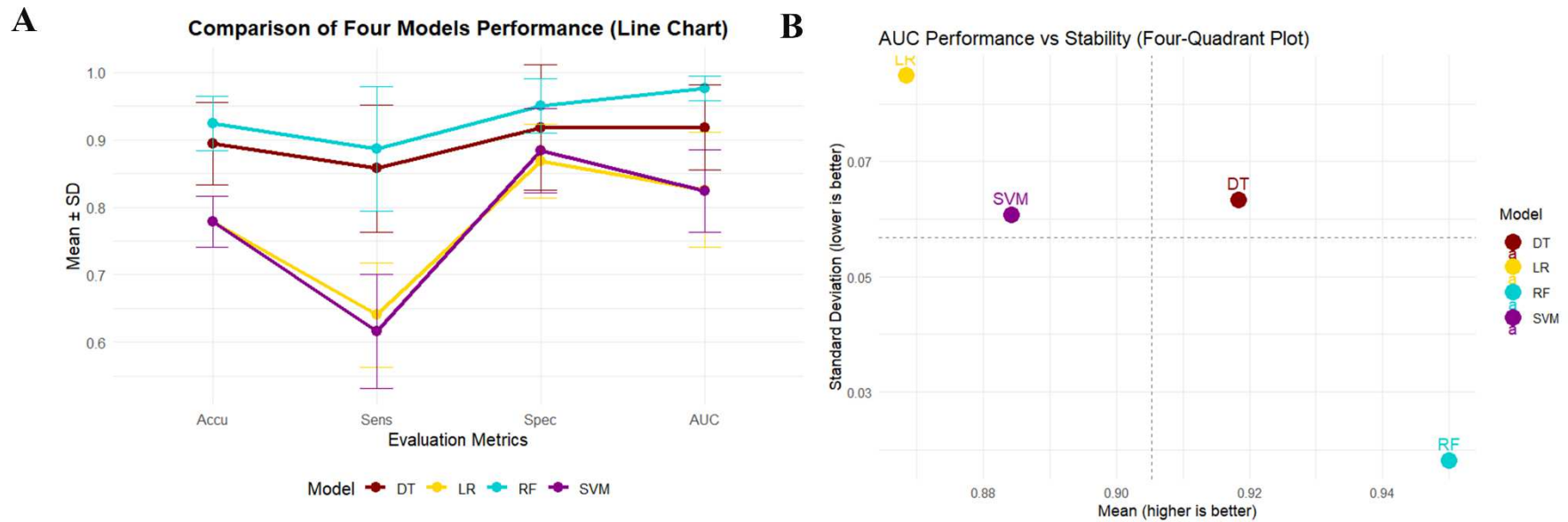


Figure 4. Evaluation of Model Performance. (A) **Line chart:** The mean $\pm$ standard deviation of accuracy, sensitivity, specificity, and AUC for the four models; (B) **Four-quadrant plot:** The mean $\pm$ standard deviation of AUC for the four models. Logistic Regression: LR, Support Vector Machine: SVM, Decision Tree: DT, Random Forest: RF

Table 5. Evaluation Metrics of the Four Models

Evaluation metrics	LR	SVM	DT	RF
Accu Mean	0.7789	0.7789	0.8943	0.9248
Sens Mean	0.6402	0.6155	0.8575	0.8860
Spec Mean	0.8684	0.8842	0.9184	0.9500
ROC Mean	0.826	0.8243	0.9238	0.9761
Accu SD	0.0379	0.0377	0.0611	0.0403
Sens SD	0.0772	0.0847	0.0944	0.0924
Spec SD	0.0541	0.0623	0.0932	0.0401
ROC SD	0.085	0.0607	0.0633	0.0182

Logistic Regression: LR, Support Vector Machine: SVM, Decision Tree: DT, Random Forest: RF

External validation of the random forest model at 3 and 6 months postoperatively

Based on baseline data collected at 1 month postoperatively, we used RF to construct a prediction model for postoperative malnutrition risk in EC patients, and then externally validated this model using an external validation set ($n = 101$) at 3 and 6 months after surgery, with performance metrics at both time points presented in **Figure 5** and **Table 6**. The results showed that the 3-month model achieved an accuracy of 0.782 and an AUC of 0.814, while the 6-month model achieved an accuracy of 0.719 and an AUC of 0.801, indicating that both models demonstrated good predictive performance (**Figure 5; Table 6**). These results indicate that the RF-based prediction model achieves satisfactory predictive performance.

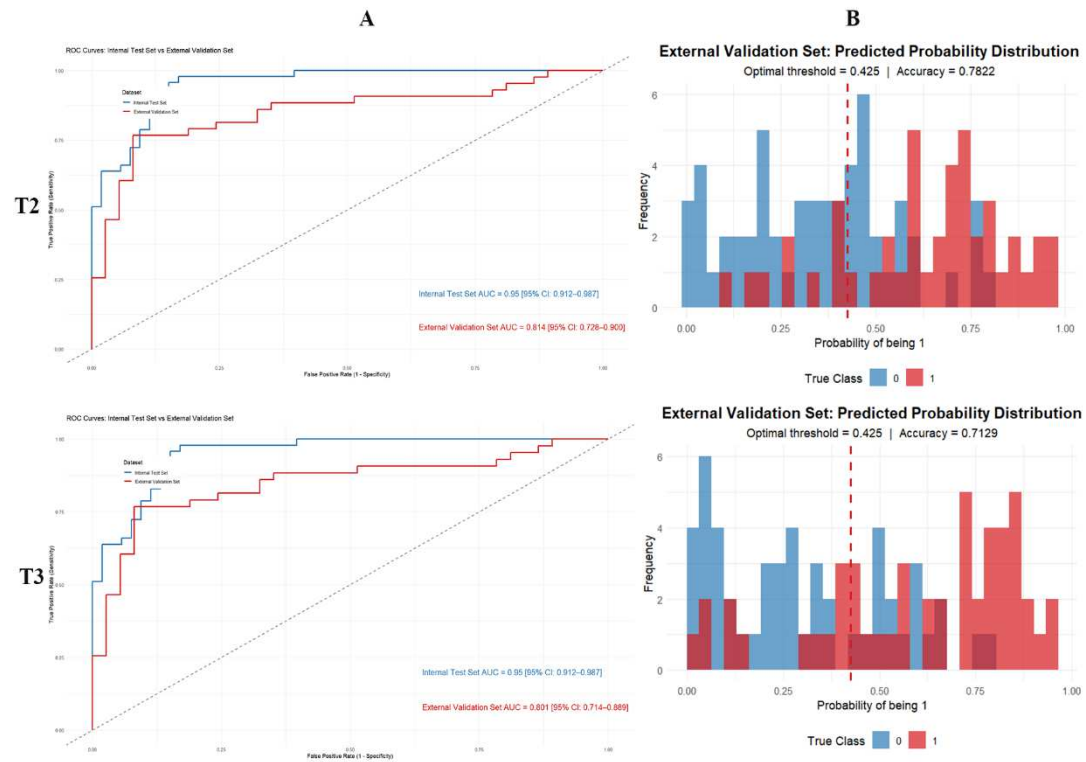


Figure 5. External validation of the random forest model. (A) ROC curves: blue curve for the 1-month postoperative model (training set) and red curve for the external validation set. (B) Predicted probability distribution plots for the external validation set at T2 (3 months postoperatively) and T3 (6 months postoperatively).

Table 6. Evaluation metrics of the model.

Performance metrics	T1Baseline model (optimal threshold: 0.425)	T2(Postoperative month 3)	T3(Postoperative month 6)
Accuracy	0.9613	0.7822	0.7129
Sensitivity	0.9286	0.7895	0.7451
Specificity	0.991	0.7727	0.6800
PPV	0.9483	0.8182	0.7037
NPV	0.9846	0.7391	0.7234
Kappa	0.9174	0.5593	0.4253
F1	0.9692	0.8036	0.7238
AUC	0.9745(0.9491-1)	0.8144(0.7266-0.9002)	0.8014(0.7139-0.8889)

Identification of Important Risk Factors Contributing to the Model

The SHAP value analysis identified the most influential predictors of malnutrition risk (**Figure 6**). Among the 10 selected features, BMI (SHAP = 0.167), age (0.137), and weight loss within 3 months (0.107) were the top three contributors, with substantially higher SHAP values than the other variables. Surgical approach (0.040) and parenteral nutrition (0.038) formed the second tier, indicating that disease severity and nutritional support are also important contributing factors. Furthermore, laboratory parameters including hemoglobin (0.020), GNRI (0.019), and albumin (0.019) demonstrated notable predictive contributions.

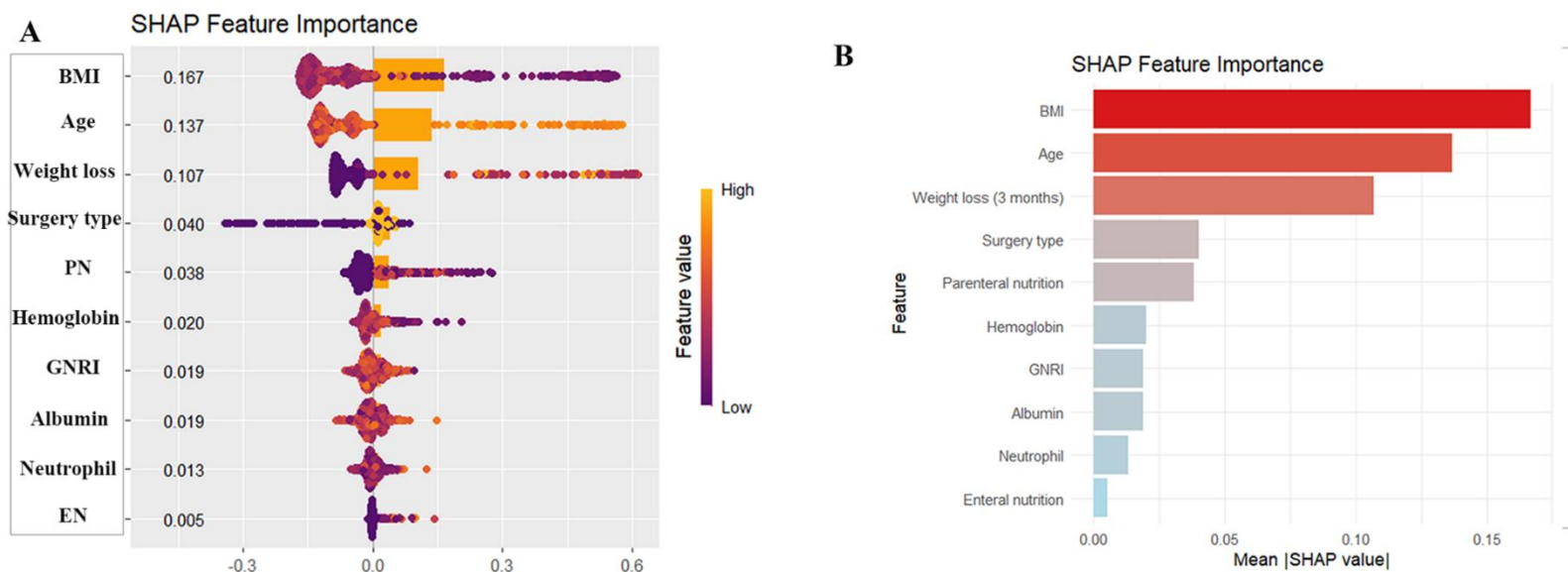


Figure 6. Identification of Important Risk Factors Contributing to the Model. A: SHAP Dependence Plot; B: Feature Importance Bar Plot.

In **Figure 7**, To elucidate the marginal effects of key continuous predictors, we present SHAP dependence plots for the model's predicted malnutrition risk. The y-axis represents the SHAP value, with positive values indicating an increased risk of malnutrition and negative values indicating a decreased risk. Each dot corresponds to an individual patient, and the color gradient from blue to red denotes increasing values of the respective feature.

(a) SHAP dependence plot for BMI (kg/m^2). A clear threshold effect was observed at approximately 18.5 kg/m^2 . Patients with BMI below this value exhibited substantially elevated SHAP values, indicating increased malnutrition risk, while those above this threshold showed negative SHAP values, suggesting reduced risk.

(b) SHAP dependence plot for age (years). SHAP values showed a modest positive correlation with age, with patients aged >70 years tending to have higher values, suggesting that advanced age is an independent risk factor for malnutrition.

(c) SHAP dependence plot for weight loss within 3 months. SHAP values increased substantially with greater weight loss, with a clear upward trend observed beyond approximately 2–3kg of weight loss. The predominantly positive SHAP values suggest that even modest weight loss contributes to elevated malnutrition risk.

(d) SHAP dependence plot for surgery type (minimally invasive surgery=1, thoracotomy=2). Patients undergoing thoracotomy exhibited substantially higher SHAP values compared with those receiving minimally invasive surgery, indicating that the greater inflammatory stress and catabolic response associated with open chest surgery significantly increased the risk of malnutrition.

(e) SHAP dependence plot for hemoglobin (g/L). Hemoglobin levels showed a clear negative correlation with SHAP values, with a threshold effect observed at approximately 100 g/L. Patients with hemoglobin below this level exhibited predominantly positive SHAP values, indicating increased malnutrition risk, while those above this threshold showed negative SHAP values, suggesting reduced risk.

(f) SHAP dependence plot for parenteral nutrition volume (mL). SHAP values increased with higher nutrition volume, particularly above 1500–2000 mL, suggesting that greater parenteral nutrition requirements may reflect more severe underlying disease and poorer baseline nutritional status.

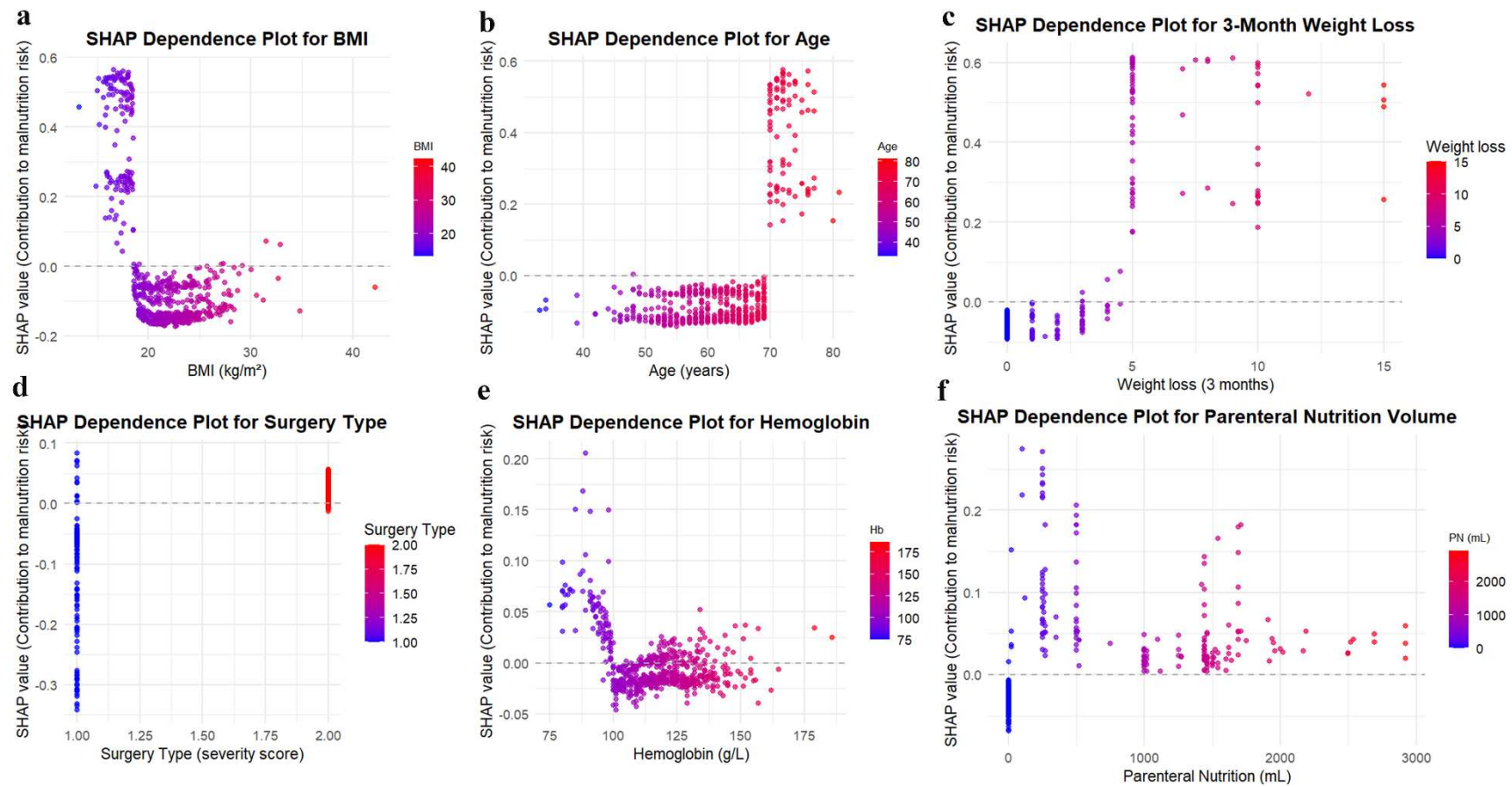


Figure 7. Local explanation for individual diagnostic interpretation. The Shapley additive explanations (SHAP) independence plot of body measurement features for (a) BMI, (b) Age, (c) weight loss within 3-months postoperatively, (d) Surgery type, (e) Hemoglobin, and (f) Parenteral Nutrition.

The SHAP method was employed to provide individualized interpretations of the model predictions, thereby enhancing model explainability and clinical applicability (**Figure 8**).

Figure 8A shows the results of the personalized prediction breakdown after the model identified this EC patient as being at risk for malnutrition. The patient's final predicted probability was 0.99, which is significantly higher than the baseline risk (mean predicted probability = 0.43). The positive factors contributing most significantly to the predicted probability included low BMI (19.77 kg/m²), low albumin (32.7 g/L), low GNRI (90.33), elevated neutrophil count (8.2×10^9 /L), advanced age (72 years), and open-chest surgery. This chart clearly illustrates the key drivers contributing to this patient's high risk of malnutrition.

Figure 8B presents the feature contribution bar plot for a nutritionally normal individual. The final predicted probability for this patient was below 0.5 (below the baseline risk). Most features showed negative SHAP contributions (blue bars), indicating that these features pulled the prediction toward lower risk. For this nutritionally normal patient, normal BMI, albumin, and hemoglobin were the main protective factors (pushing risk down). Although neutrophil count slightly increased the risk, the three protective factors had a stronger combined effect, keeping the final prediction within the normal range.

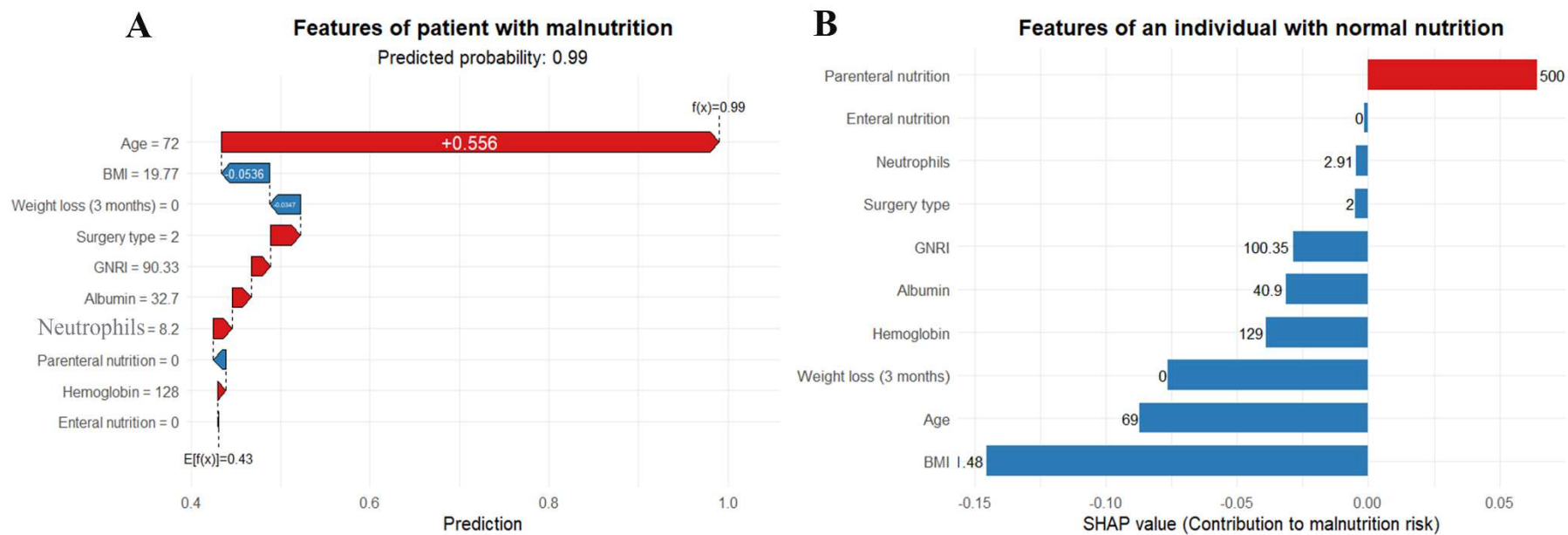


Figure 8. Local explanation for individual diagnostic interpretation. A: features of patient with malnutrition; B: features of an individual with normal nutrition.

Discussion

In this retrospective cohort study, we enrolled 906 EC patients to develop and validate a ML-based prediction model for the early identification of postoperative malnutrition risk. The results showed that the RF model exhibited the best performance and provided accurate dynamic predictions of malnutrition risk at 3 and 6 months postoperatively among EC patients. Finally, we used SHAP values to quantify the importance of the major predictors and visualize them at the global level, and further conducted individual-level interpretable visualizations separately for malnourished and well-nourished patients.

Malnutrition is the most common complication among patients with EC. In the present study, with an NRS2002 score of ≥ 3 used as the criterion for malnutrition risk, the prevalence of malnutrition risk was 39.01% at 1 month after surgery, and increased to 65.47% and 60.12% at 3 and 6 months postoperatively. These findings indicate that the risk of postoperative malnutrition in EC patients is extremely high and represents a dynamic process, which is consistent with previous studies based on the NRS 2002 scoring system. One study [22] of 1,482 EC patients reported a 50% malnutrition risk, whereas another [23] reported a prevalence of 41.25%, collectively suggesting that EC patients face a substantial malnutrition burden.

LASSO regression was initially employed to screen for major predictors of malnutrition in EC patients, identifying ten variables: BMI, age, weight loss within 3 months, surgical approach, neutrophil count, enteral nutrition, parenteral nutrition, GNRI, albumin, and hemoglobin. We then performed a SHAP-based visualization to assess the contribution of each factor and to determine optimal cutoff values for the key factors. Among these, BMI, age, and weight loss ranked as the top three contributors. Notably, this ranking corroborates previous findings [24], in which similar predictors were identified in a malnutrition model for critically ill patients, suggesting that these factors are robust across diverse patient populations.

BMI was identified as playing a critical role in predicting malnutrition risk. Since BMI is derived from body weight and weight loss directly reflects nutritional depletion, both serve as straightforward indicators of a patient's nutritional status. In our analysis, a clear threshold effect emerged for BMI at approximately 18.5 kg/m²; patients below this cutoff exhibited a sharply elevated risk, effectively distinguishing a high-risk subgroup. Notably, this threshold aligns closely with the NRS 2002 criteria, reinforcing its clinical utility. This finding is also consistent with a previous study [25], which used the NRS 2002 to investigate weight loss trajectories as predictors of malnutrition in gastric cancer patients and identified BMI > 18.5 kg/m² as an independent protective factor for postoperative malnutrition (OR = 0.564, 95% CI: 0.490–0.649). Similarly, another study [26] focusing on malnutrition prediction in gastric cancer reported results consistent with our study.

Weight loss is widely regarded as one of the most robust and sensitive objective markers for predicting malnutrition risk. In our study, weight loss >3 kg within three months was identified as an independent risk factor, which is consistent with previous study [27] showing that significant weight loss (mean 5.9%) occurs around week 5 of chemoradiotherapy and is accompanied by progressive declines in Mini Nutritional Assessment (MNA) scores. This finding was also confirmed in another study [28], which developed a malnutrition prediction model for elderly inpatients and also ranked weight loss among the most important predictors of GLIM-defined malnutrition. In parallel, longitudinal trajectory modelling in postoperative gastric cancer patients uncovered distinct weight-loss patterns; nearly half of the cohort exhibited moderate-to-severe loss, and malnutrition prevalence reached a peak of 51.6% at 6 months, underscoring the utility of dynamic weight profiles in stratifying high-risk individuals [24]. Collectively, these findings suggest that weight loss is closely correlated with nutritional decline across the cancer care continuum, underscoring its value for early malnutrition risk identification.

Furthermore, advancing age was positively associated with malnutrition risk, which may stem from the interplay between physiological decline and chronic disease-related malnutrition, collectively forming a detrimental cycle [29]. More specifically, our analysis identified a categorical threshold of age ≥ 70 years and a progressive risk increase per additional year, corroborating earlier findings [26] which reported age ≥ 65 years as a significant predictor of malnutrition (OR = 1.653, 95% CI: 1.023–2.671). This discrepancy may be attributed to differences in disease biology and surgical trauma between esophageal and gastric cancer. EC patients often present with prolonged dysphagia and poorer preoperative nutritional reserves, and esophagectomy involves a more extensive surgical resection with higher anastomotic complication rates, further reducing the physiological tolerance to aging [30–31]. Notably, the ≥ 70 years cutoff corresponds to the age-scoring component of the NRS 2002, further supporting its clinical applicability.

Inflammatory and nutritional parameters are recognized as important indicators for evaluating the nutritional status and prognosis of EC patients [32]. Our study confirmed a nonlinear hemoglobin-malnutrition relationship with a threshold at 100 g/L. In agreement with earlier study [26], which reported hemoglobin also was a significant predictor. However, we extended their work by providing a specific cutoff. This divergence likely arises from differing aims: they tested statistical significance, while we sought a clinically actionable threshold for a prediction model, necessitating SHAP-based optimization. This finding provides a practical quantitative marker for early risk stratification in clinical practice. Although the identified biomarker indicators are closely associated with malnutrition, they may also reflect systemic inflammation or disease severity. Therefore, such indicators should be interpreted in the context of the

patient's overall health status and disease progression in clinical interpretation. Collectively, these findings validate the phenotypic and etiologic constructs of the NRS2002 criteria and provide quantitative risk thresholds, offering clinicians actionable references for precision nutritional screening and individualized intervention in EC patients.

To visually illustrate the contribution of individual features, we used SHAP to provide a personalized interpretation of a typical case of malnutrition and compared it with a nutritionally normal individual. This patient's baseline risk was 0.43, which increased to 0.99 after incorporating all features. The primary positive risk factors included low BMI, low albumin, low GNRI, elevated neutrophil count, advanced age, and a high surgical severity score. Notably, neither the absence of recent weight loss nor the lack of nutritional support significantly reduced the risk, suggesting that assessing weight changes or nutritional history alone is insufficient in the context of inflammation and hypoproteinemia. In contrast, most features in well-nourished individuals made a negative contribution; normal BMI, albumin, and hemoglobin levels kept the risk within a reasonable range. This comparison validates the model's discriminatory power and provides an intuitive basis for decision-making, enabling clinicians to quickly identify key factors.

In the study, Employing four machine-learning classifiers (LR, SVM, DT, and RF) to respectively construct malnutrition-risk prediction models, and assessing their performance via 10-fold CV and test-set evaluation measures, the present study demonstrated that RF outperformed the others, a superiority that was further confirmed in external validation cohorts at postoperative months 3 and 6 with AUC values of 0.814(0.726-0.900) and 0.801(0.714-0.889) for the respective time points, wherein RF consistently maintained outstanding accuracy and precision.

Limitations

Several limitations of the present study merit consideration. First, the retrospective design of this cohort study may introduce inherent bias. Second, NRS2002 was employed as the sole outcome measure for malnutrition risk, and thus the generalizability of our model to other nutritional assessment instruments remains to be established. Third, given that both the training and internal validation cohorts and the external validation cohort were derived from a single institution. Therefore, future multi-center prospective external validation is still warranted to further demonstrate the robustness and generalizability of our model.

Conclusions

This study developed a postoperative malnutrition risk prediction model for EC patients based on Random Forest, which accurately predicts malnutrition risk at 3 and 6 months postoperatively. In addition, SHAP analysis identified the optimal cutoff values of the three most influential predictors: BMI ≤ 18.5 kg/m², weight loss >3 kg within 3 months, and age >70

years. These thresholds can serve as references for malnutrition risk screening, aiding early clinical identification and contributing to early prevention. Further prospective studies with large cohorts are needed to externally validate and refine this model, ultimately facilitating its integration into routine clinical practice.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. The research protocol was approved by the Ethics Committee of [the Second Affiliated Hospital of Air Force Medical University] (Approval No. [K-HG-202510-08]). Due to the retrospective nature of this study, which involved the analysis of existing medical records and/or historical data, and the fact that no additional interventions were performed on the participants, the requirement for individual informed consent was waived by the above-mentioned Ethics Committee. All patient data were anonymized and de-identified prior to analysis to ensure privacy protection.

Consent to participate

Informed consent was waived by the Ethics Committee due to the retrospective design of the study, as it posed no more than minimal risk to the subjects and it was impractical to obtain consent from all individuals whose data were used.

Funding

This work was supported by [The Project Supported by Natural Science Basic Research Plan in Shaanxi Province of China] under Grant No. [No. 2026JC-YBMS-0894].

Human Ethics and Consent to Participate declarations

Included in the "Ethics approval and consent to participate" section above.

Availability of data and materials

The clinical data used in this study were derived from [the Second Affiliated Hospital of Air Force Medical University] and are not publicly available due to patient privacy protection and institutional data governance policies. De-identified aggregated data may be made available from the corresponding author upon reasonable request and with the approval of the relevant ethics committee. The analytical code and machine learning models developed in this study are available from the corresponding author upon reasonable request.

References

- [1]. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024 May-Jun;74(3):229-263.
- [2]. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, 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):209-49.
- [3]. Morgan E, Soerjomataram I, Rungay H, Coleman HG, Thrift AP, Vignat J, et al. The Global Landscape of Esophageal Squamous Cell Carcinoma and Esophageal Adenocarcinoma Incidence and Mortality in 2020 and Projections to 2040: New Estimates From GLOBOCAN 2020. *Gastroenterology.* 2022;163(3):649-58 e2.
- [4]. Jordan T, Mastnak DM, Palamar N, Kozjek NR. Nutritional Therapy for Patients with Esophageal Cancer. *Nutr Cancer.* 2018 Jan;70(1):23-29.
- [5]. SONG C, CAO J, ZHANG F, et al. Nutritional Risk Assessment by Scored Patient-Generated Subjective Global Assessment Associated with Demographic Characteristics in 23,904 Common Malignant Tumors Patients. *Nutr Cancer.* 2019;71(1):50-60.
- [6]. Zhang L, Lu Y, Fang Y. Nutritional status and related factors of patients with advanced gastrointestinal cancer. *Br J Nutr.* 2014 Apr 14;111(7):1239-44.
- [7]. Zheng F, Chen K, Zhang X, Wang Q, Zhang Z, Niu W. Prevalence of malnutrition and associated factors in Chinese children and adolescents aged 3-14 years using machine learning algorithms. *J Glob Health.* 2025 Jul 21;15:04204. doi: 10.7189/jogh.15.04204. PMID: 40689479; PMCID: PMC12278689.
- [8]. Wang X, Li C, Xu L, Jiang S, Guan C, Che L, Wang Y, Man X, Xu Y. Construction and validation of prognostic models for acute kidney disease and mortality in patients at risk of malnutrition: an interpretable machine learning approach. *Clin Kidney J.* 2025 Mar 13;18(4):sfaf080. doi: 10.1093/ckj/sfaf080. PMID: 40236512; PMCID: PMC11997651.
- [9]. Sguanci M, Palomares SM, Cangelosi G, Petrelli F, Sandri E, Ferrara G, Mancin S. Artificial Intelligence in the Management of Malnutrition in Cancer Patients: A Systematic Review. *Adv Nutr.* 2025 Jul;16(7):100438. doi: 10.1016/j.advnut.2025.100438. Epub 2025 May 5. PMID: 40334987; PMCID: PMC12281439.

- [10]. Zhang Z, Wan Z, Zhu Y, Zhang L, Zhang L, Wan H. Prevalence of malnutrition comparing NRS2002, MUST, and PG-SGA with the GLIM criteria in adults with cancer: A multi-center study. *Nutrition*. 2021 Mar;83:111072. doi: 10.1016/j.nut.2020.111072. Epub 2020 Nov 19. PMID: 33360034
- [11]. Robert Tibshirani, Regression Shrinkage and Selection Via the Lasso, *Journal of the Royal Statistical Society: Series B (Methodological)*, Volume 58, Issue 1, January 1996, Pages 267 – 288, <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [12]. Shi H, Yang D, Tang K, Hu C, Li L, Zhang L, Gong T, Cui Y. Explainable machine learning model for predicting the occurrence of postoperative malnutrition in children with congenital heart disease. *Clin Nutr*. 2022 Jan;41(1):202-210. doi: 10.1016/j.clnu.2021.11.006. Epub 2021 Nov 10. PMID: 34906845.
- [13]. Yin L, Lin X, Liu J, Li N, He X, Zhang M, Guo J, Yang J, Deng L, Wang Y, Liang T, Wang C, Jiang H, Fu Z, Li S, Wang K, Guo Z, Ba Y, Li W, Song C, Cui J, Shi H, Xu H; Investigation on Nutrition Status and Clinical Outcome of Common Cancers (INSCOC) Group. Classification Tree-Based Machine Learning to Visualize and Validate a Decision Tool for Identifying Malnutrition in Cancer Patients. *JPEN J Parenter Enteral Nutr*. 2021 Nov;45(8):1736-1748. doi: 10.1002/jpen.2070. Epub 2021 Mar 15. PMID: 33415743.
- [14]. Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med*. 2011 Oct;18(10):1099-104. doi: 10.1111/j.1553-2712.2011.01185.x. PMID: 21996075.
- [15]. Huang, Shujun et al. “Applications of Support Vector Machine (SVM) Learning in Cancer Genomics.” *Cancer genomics & proteomics* vol. 15,1 (2018): 41-51. doi:10.21873/cgp.20063.
- [16]. Che D, Liu Q, Rasheed K, Tao X. Decision tree and ensemble learning algorithms with their applications in bioinformatics. *Adv Exp Med Biol*. 2011;696:191-9. doi: 10.1007/978-1-4419-7046-6_19. PMID: 21431559.
- [17]. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Brief Bioinform*. 2023;24(2):bbad002. doi:10.1093/bib/bbad002
- [18]. Zeyauddin, M., Abidin, S., Khan, I. *et al*. Clinical predictive fusion network for accurate disease prediction in patient cohorts. *Sci Rep* **16**, 3626 (2026). <https://doi.org/10.1038/s41598-025-33645-9>
- [19]. Cao J, Xu H, Li W, Guo Z, Lin Y, Shi Y, Hu W, Ba Y, Li S, Li Z, Wang K, Wu J, He Y, Yang J, Xie C, Zhou F, Song X, Chen G, Ma W, Luo S, Chen Z, Cong M,

- Ma H, Zhou C, Wang W, Qi Luo, Shi Y, Qi Y, Jiang H, Guan W, Chen J, Chen J, Fang Y, Zhou L, Feng Y, Tan R, Ou J, Zhao Q, Wu J, Xin Lin, Yang L, Fu Z, Wang C, Deng L, Li T, Song C, Shi H; Investigation on Nutrition Status and Clinical Outcome of Common Cancers (INSCOC) Group, Chinese Society of Nutritional Oncology. Nutritional assessment and risk factors associated to malnutrition in patients with esophageal cancer. *Curr Probl Cancer*. 2021 Feb;45(1):100638. doi: 10.1016/j.cupr.2020.100638. Epub 2020 Aug 14. PMID: 32829957.
- [20]. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011-2022). *Comput Methods Programs Biomed*. 2022 Nov;226:107161. doi: 10.1016/j.cmpb.2022.107161. Epub 2022 Sep 27.
- [21]. Kong X, Liu P, Wang G, Sun S, Li L. Methods for diagnosing malnutrition in patients with esophageal cancer, and the association with nutritional and inflammatory indices: A cross-sectional study. *Oncol Lett*. 2025 Mar 5;29(5):223. doi: 10.3892/ol.2025.14969. PMID: 40110582; PMCID: PMC11921283.
- [22]. Liu Y, Xu Y, Guo L, Chen Z, Xia X, Chen F, Tang L, Jiang H, Xie C. Development and external validation of machine learning models for the early prediction of malnutrition in critically ill patients: a prospective observational study. *BMC Med Inform Decis Mak*. 2025 Jul 3;25(1):248. doi: 10.1186/s12911-025-03082-9. PMID: 40611154; PMCID: PMC12225150.
- [23]. Huang W, Wang C, Wang Y, et al. Predicting malnutrition in gastric cancer patients using computed tomography(CT) deep learning features and clinical data. *Clin Nutr*. 2024;43(3):881-891. doi:10.1016/j.clnu.2024.02.005
- [24]. Yan M, Lin Z, Chen R, Liu Y, Jian J and Zhuo C (2025) Identification of postoperative weight loss trajectories and development of a machine learning-based tool for predicting malnutrition in gastric cancer patients. *Front. Nutr*. 12:1678879. doi: 10.3389/fnut.2025.1678879
- [25]. Lin HY, Feng HC, Wang YJ. Trajectory of weight loss and nutritional decline in head and neck cancer patients undergoing concurrent chemoradiotherapy: a longitudinal observational study. *Support Care Cancer*. 2025. (published online Oct 29).
- [26]. Wang X, Yang F, Zhu M, et al. Development and assessment of assisted diagnosis models using machine learning for identifying elderly patients with malnutrition: cohort study. *J Med Internet Res*. 2023;25:e42435.
- [27]. Norman K, Haß U, Pirlich M. Malnutrition in Older Adults-Recent Advances and

Remaining Challenges. *Nutrients*. 2021 Aug 12;13(8):2764. doi: 10.3390/nu13082764. PMID: 34444924; PMCID: PMC8399049

- [28]. Jiang W, Zhang B, Xu J, Xue L, Wang L. Current status and perspectives of esophageal cancer: a comprehensive review. *Cancer Commun (Lond)*. 2025 Mar;45(3):281-331.
- [29]. Waters JK, Reznik SI. Update on Management of Squamous Cell Esophageal Cancer. *Curr Oncol Rep*. 2022 Mar;24(3):375-385.
- [30]. Aoyama T, Kazama K, Maezawa Y, Hara K. Usefulness of Nutrition and Inflammation Assessment Tools in Esophageal Cancer Treatment. *In Vivo (Athens, Greece)*. 2023 Jan-Feb;37(1):22-35. DOI: 10.21873/invivo.13051. PMID: 36593006; PMCID: PMC9843752.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.zip](#)