

Development of a Machine Learning-Based Predictive Model for Social Frailty Risk in Cancer Patients Undergoing Chemotherapy

Yao Chen

HanZhong Central Hospital

Wang Xinyi

HanZhong Central Hospital

Liu Qingchun

HanZhong Central Hospital

Zhang Wei

HanZhong Central Hospital

Ma Xiuqin

HanZhong Central Hospital

Shen Jinlian

HanZhong Central Hospital

Wang Rong

2807442653@qq.com

HanZhong Central Hospital

Research Article

Keywords: Machine learning, Cancer, Chemotherapy, Social frailty, Prediction model

Posted Date: October 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10837343/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Development of a Machine Learning-Based Predictive Model for Social Frailty Risk in Cancer Patients Undergoing Chemotherapy

Author names: Yao Chen¹, Wang Xinyi¹, Liu Qingchun¹, Zhang Wei¹, Ma Xiuqin¹,
Shen Jinlian¹, Wang Rong^{1*} (corresponding author)

Affiliation: ¹HanZhong Central Hospital, HanZhong, 723000, China.

Email Address of the Corresponding Author: 2807442653@qq.com

Objective: This study aims to build a social frailty risk prediction model for cancer patients undergoing chemotherapy based on machine learning.

Methods: From June 2025 to June 2026, 253 cancer patients who received their first chemotherapy at a certain tertiary grade A hospital in Shaanxi Province were selected as the research subjects. The relevant clinical data of the patients were collected, and the patients were randomly divided into a training set (178 cases) and a validation set (75 cases) in a ratio of 7:3. The outcome variable was whether social frailty occurred after the end of the chemotherapy cycle. Five machine learning algorithms, including logistic regression, random forest, XGBoost, decision tree, and naive Bayes, were used to build the prediction model. The performance of the model was evaluated using indicators such as the area under the receiver operating characteristic curve (AUC), Brier score, etc., and the optimal model was selected. The SHAP algorithm was used to conduct interpretive analysis of the optimal model.

Results: Among the 253 cancer patients who received their first chemotherapy, 83 cases (32.81%) experienced social frailty after the end of the chemotherapy cycle. Among the five models, XGBoost exhibited high AUC (0.889), accuracy (0.867), sensitivity (0.708), and specificity (0.941), with the lowest Brier score (0.110). The SHAP values of XGBoost were calculated, and it was found that PHQ-9, PSSS, and Sleep

disorder were the top 3 key characteristic variables for predicting social frailty in cancer patients undergoing chemotherapy.

Conclusion: The social frailty risk prediction model for cancer patients undergoing chemotherapy based on XGBoost has good predictive ability and can provide a reference for optimizing the management of social frailty in cancer patients undergoing chemotherapy.

Keywords: Machine learning; Cancer; Chemotherapy; Social frailty; Prediction model

Cancer remains a major global public health concern, with persistently high incidence and mortality rates that pose a significant threat to population health and socioeconomic development. The International Agency for Research on Cancer reported approximately 19.96 million new cancer cases and 9.7 million cancer-related deaths worldwide in 2022^[1]. In China, cancer also exhibits a high prevalence, with approximately 4.8247 million new cases annually and a noticeable trend toward younger age at onset^[2]. Chemotherapy is one of the core treatment modalities for malignant tumors, playing an irreplaceable role in neoadjuvant therapy, postoperative elimination of residual lesions, and palliative care for advanced disease^[3]. However, chemotherapeutic agents lack cellular specificity, and while exerting cytotoxic effects on tumor cells, they can also damage normal tissues, leading to a range of toxicities including nausea, vomiting, bone marrow suppression, and alopecia. These adverse effects not only reduce patient adherence to treatment but also exacerbate the multidimensional physical, psychological, and social burdens^[4], thereby inducing or aggravating frailty. As a critical dimension of frailty, social frailty specifically refers to a state of persistent deficiency in basic social resources, characterized by reduced social behaviors, lack of social engagement, and diminished self-management capacity^[5]. Chemotherapy-induced decline in physical function, increased

financial burden, and restricted social roles can further disrupt patients' social support systems and exacerbate feelings of social isolation. Studies have demonstrated that the prevalence of social frailty among cancer patients is higher than that observed in patients with chronic heart failure^[6]. Moreover, preoperative social frailty has been identified as an independent predictor of postoperative malnutrition and is closely associated with postoperative complications in patients with gastrointestinal cancer^[7]. For patients undergoing chemotherapy, social frailty may create a vicious cycle of "physiological impairment-insufficient social support-psychological distress-worsened prognosis," underscoring the urgent need for effective tools for early identification and intervention. In recent years, machine learning (ML) has been increasingly applied in the field of medical prediction. By uncovering complex nonlinear interactions among variables, ML can significantly enhance model predictive performance^[8] and may offer a more precise approach for social frailty risk assessment. This study aims to develop a risk prediction model for social frailty in cancer patients undergoing chemotherapy using machine learning algorithms, and to visualize and interpret the optimal model through the SHAPley Additive exPlanations (SHAP) method, thereby providing a reference for healthcare professionals to identify high-risk populations early and formulate individualized intervention strategies.

1 Subjects and Methods

1.1 Research Object

From June 2025 to February 2026, a prospective cohort study was conducted at a tertiary hospital in Shaanxi Province, enrolling cancer patients who met the inclusion and exclusion criteria and were receiving their first course of chemotherapy. The inclusion criteria were as follows: (1) age $\geq$ 18 years; (2) pathologically or cytologically confirmed diagnosis

of malignancy; (3) first-time chemotherapy; (4) conscious state with basic communication and comprehension abilities; and (5) provision of informed consent and voluntary participation. The exclusion criteria included: (1) severe cognitive or psychiatric disorders; (2) expected survival < 6 months; (3) presence of severe complications or comorbidities; and (4) concurrent participation in other conflicting studies. The sample size was determined using the events per variable (EPV) method in logistic regression. Based on an estimated inclusion of 8–10 predictive factors, an EPV of 10, and a preliminary pilot study indicating a social frailty incidence of 37.14% among cancer patients undergoing chemotherapy, the required sample size was calculated to be 237, accounting for a 10% rate of invalid questionnaires. Ultimately, 260 patients were enrolled. During the study, 4 patients withdrew due to failure to complete the full chemotherapy cycle because of side effects, and 3 patients were lost to follow-up, resulting in a final analytical sample of 253 patients. This study was approved by the Ethics Committee of Hanzhong Central Hospital (Approval No. 2026089).

1.2 Research Methods

Potential influencing factors of social frailty in cancer patients undergoing chemotherapy were identified through a comprehensive literature review across multiple databases. Guided by the health ecology theory, data collection was structured according to the following dimensions. The personal biological layer included age, sex, body mass index (BMI), chronic comorbidities, nutritional status, and tumor site. The behavioral layer encompassed sleep status, psychological condition, and cancer-related fatigue. The interpersonal layer comprised marital status and social support. The living and working conditions layer included place of residence, personal income, financial toxicity, and utilization of chronic disease resources. The policy and environmental

layer covered type of medical insurance and cultural factors. All these data were collected prior to the initiation of the first chemotherapy cycle (baseline period) to construct a multidimensional predictor dataset. Upon completion of all chemotherapy cycles, the social frailty status of patients was assessed using the Social Frailty Scale, which served as the endpoint outcome variable for model prediction.

1.2.1 Social Frailty

Social frailty was assessed using the scale developed by Ma et al.^[9], which comprises five items: (1) Help: whether the patient was able to help friends or family members during the previous year; (2) Participation: whether the patient participated in social or leisure activities during the previous year; (3) Loneliness: whether the patient felt lonely during the past week; (4) Financial: whether the patient's income during the previous year was sufficient to sustain living expenses for the past year; and (5) Talk: whether the patient had someone to talk to every day. For the Loneliness item (item 3), a response of "no" was scored as 0, while "yes" was scored as 1. For the remaining items, a response of "no" was scored as 1, and "yes" as 0. The total score ranged from 0 to 5, with 0 indicating no social frailty, 1–2 indicating pre-social frailty, and ≥ 3 indicating social frailty. The Cronbach's α coefficient of the scale was 0.602.

1.2.2 General Information Survey Form

A self-designed questionnaire was developed by the research team, covering demographic and clinical characteristics including age, sex, smoking status, alcohol consumption, marital status, educational level, monthly per-capita household income, type of medical insurance, presence of chronic diseases, body mass index (BMI), and tumor site.

1.2.3 PHQ-9

Developed by the American Psychiatric Association based on the DSM-IV diagnostic criteria and depressive symptomatology standards^[10],

this scale is widely used for self-assessment of depressive symptoms over the past two weeks. It comprises 9 items without specific subdimensions. Each item is scored on a 4-point Likert scale, where 0=never, 1=1–3 days, 2=4–6 days, and 3=nearly every day. The total score is the sum of all item scores, with higher scores indicating more severe depressive symptoms. The Cronbach's α coefficient of this scale is 0.941.

1.2.4 Cancer-Related Fatigue Scale

Developed by Okuyama in 2000^[11], this scale has been widely used to assess cancer-related fatigue symptoms in patients. The scale consists of 15 items encompassing three dimensions: physical fatigue, affective fatigue, and cognitive fatigue. Each item is rated on a Likert 5-point scale: 1 = not at all, 2 = very little, 3 = a little, 4 = quite a bit, and 5 = very much. Specifically, the scoring for each dimension is as follows: physical fatigue score = (items 1+2+3+5+6+9+15) – 7, with a possible range of 0 to 28; affective fatigue score = 20 – (items 8+11+12+14), with a range of 0 to 16; and cognitive fatigue score = (items 4+7+10+13) – 4, with a range of 0 to 16. The sum of the three dimension scores yields the total scale score, ranging from 0 to 60, with a higher total score indicating a more severe level of cancer-related fatigue. The Cronbach's α coefficient of this scale was 0.929.

1.2.5 Pittsburgh Sleep Quality Index (PSQI)

Developed by Buysse in 1989^[12], this scale is designed to assess patients' sleep quality over the past month, comprising seven dimensions: sleep quality, sleep latency, total sleep time, sleep efficiency, sleep disturbances, use of hypnotics, and daytime dysfunction. Each dimension is scored on a 0–3 scale, with a total score ranging from 0 to 21. Higher scores indicate more severe sleep disturbances. In this study, a PSQI score exceeding 7 was used to identify sleep disorders. The Cronbach's α coefficient of this scale was 0.84.

1.2.6 Comprehensive Score for Financial Toxicity-Patient Reported Outcome Measure (COST-PROM)

Developed by de Souza et al. in 2014^[13], this scale has been widely used to assess cancer patients' perceived subjective and objective financial and work-related distress over the past week. The scale comprises 11 items, including 4 positively worded and 7 negatively worded items. The total score ranges from 0 to 44, with lower scores indicating greater financial toxicity. The Cronbach's α coefficient of this scale is 0.889.

1.2.7 Chinese Version of the Chronic Illness Resource Survey Scale (CV-CIRS)

Revised by Glasgow et al. in 2005 based on the short form of the Chronic Illness Resource Survey^[14], this scale is used to assess patients' utilization of resources. It comprises 6 dimensions and 19 items. The total score ranges from 10 to 95, with higher scores indicating more favorable resource utilization. The Cronbach's α coefficient of this scale is 0.860.

1.2.8 Perceived Social Support Scale (PSSS)

Developed by Zimet et al.^[15], this scale is used to assess the level of social support perceived by individuals. It comprises 3 dimensions and 12 items. The total score ranges from 12 to 84, with higher scores indicating greater perceived social support. The Cronbach's α coefficient of this scale is 0.894.

1.3 Data Collection and Quality Control Methods

The survey was conducted using paper-based questionnaires. Three charge nurses were recruited to assist with the investigation and received standardized training prior to the study. After obtaining informed consent from cancer patients undergoing chemotherapy, uniform instructions were provided to explain the questionnaire completion methods and precautions. For patients who had difficulty completing the questionnaire,

researchers provided assistance by reading each item aloud and recording their responses. Questionnaires were collected and verified on site; any missing or omitted items were supplemented immediately. Upon completion of the chemotherapy cycle, patients were again invited to complete the Social Frailty Scale (as the outcome assessment) to ensure the integrity of the endpoint data. After questionnaire retrieval, all responses were double-checked by two independent reviewers. Invalid questionnaires—defined as those with a completion time of less than 5 minutes, more than 50% of consecutive identical responses, or showing an obvious patterned answering tendency—were excluded to ensure data validity.

1.4 Statistical Methods

Two people are responsible for data entry and verification. No missing data were present in the dataset. The random seed is set to 926, and the data is divided in a 7:3 ratio to obtain the training set and the validation set. The training set is used for model construction, and the validation set is used for model validation. To avoid the instability of model estimation caused by rare levels in categorical variables, in the data preprocessing step, the frequency of levels for all categorical variables in the training set is statistically analyzed. Levels with a frequency of less than 5 are merged into the 'Other' category. This merging threshold is determined based on the training set, and the validation set follows the merging rules of the training set for the same processing to ensure the consistency of the modeling data. Data analysis and model construction are accomplished using R 4.5.2. Baseline characteristics were compared between the social frailty and non-social frailty groups using independent-sample t-tests for continuous variables and Chi-square tests for categorical variables. Variables with $P < 0.05$ in these comparisons were then entered into univariate logistic regression

models, followed by LASSO regression using the glmnet package. The optimal penalty coefficient λ is determined through 10-fold cross-validation. Based on the LASSO selection results, a multi-factor Logistic regression model is further constructed, and the variables with $P < 0.05$ are retained as the final modeling variables. Correlation heatmap analysis is performed on the final included modeling variables, and the variance inflation factor (VIF) is calculated to comprehensively check for potential multicollinearity issues. Five different machine learning models are constructed: logistic regression, random forest, XGBoost, decision tree, and naive Bayes. In terms of model evaluation, first, 10-fold cross-validation is conducted within the training set to calculate the AUC, accuracy, sensitivity, specificity, and Brier Score of each model to evaluate the internal stability and generalization ability of the model. Then, the trained model is applied to the independent validation set for internal validation. The 95% confidence interval of the validation set AUC is estimated using the Bootstrap resampling method (repeated 2000 times), and the receiver operating characteristic (ROC) curve is drawn to visually compare the discrimination of each model. In the assessment of model calibration and clinical practicality, the calibration curve is used to evaluate the consistency between predicted probabilities and actual observations, and the decision curve analysis (DCA) is used to evaluate the net clinical benefit of each model at different threshold probabilities to measure its practical application value. The best-performing model based on the AUC and clinical applicability of the validation set is selected as the final recommended model. For the model with the best performance on the validation set, further SHAP analysis is conducted for interpretability, and the contribution direction and size of each variable to the prediction outcome are visualized through swarm plots and feature importance bar charts to enhance the interpretability of the model.

2 Results

2.1 Univariate analysis of factors influencing social frailty in cancer patients undergoing chemotherapy

This study included a total of 253 cancer patients who received chemotherapy for the first time. After the end of the chemotherapy cycle, 83 cases (32.81%) experienced social frailty, and 170 cases (67.19%) did not experience social frailty. Randomly divide the population into a training set (178 cases) and a validation set (75 cases) in a 7:3 ratio. In the training set, there were 59 cases (33.15%) of socially frail individuals and 119 cases (66.85%) of non socially frail individuals; In the validation set, there were 24 cases (32.00%) of socially frail individuals and 51 cases (68.00%) of non socially frail individuals. The baseline characteristics of the training set are shown in Table 1. Univariate logistic regression analyses for the nine variables that showed significant differences in Table 1 (Education, Income, malnutrition, chronic disease, sleep disorder, PHQ-9, CRF, PSSS, and COST-PROM) are presented in Table 2, along with their OR and 95% CI.

2.2 Feature screening

Based on the training set (n=178 cases) data, taking the occurrence of social frailty after the end of the chemotherapy cycle as the outcome variable, the 9 influencing factors with statistical significance ($P < 0.05$) in univariate logistic regression analysis were included in LASSO regression. The optimal penalty coefficient $\lambda_{1se} = 0.0516$ was determined through 10 fold cross validation. Under the λ_{1se} criterion, 7 core predictor variables with non-zero coefficients were further selected from the 9 variables initially screened for single factor analysis (see Figure 1 for variable selection path and Figure 2 for cross validation error curve). Incorporate the 7 variables selected by LASSO into the multivariate logistic regression model. The correlation heatmap analysis (see Figure 3)

shows that the absolute values of the correlation coefficients between variables are all below 0.5; Further calculation of variance inflation factor (VIF) showed that the VIFs of all seven variables were far below the critical value of 5 (see Table 3), indicating that there was no serious multicollinearity issue among the included variables. The results of multiple logistic regression showed that PHQ-9, COST-PROM, Malnutrition, PSSS, Sleep disorder, and CRF were independent influencing factors for social frailty in cancer chemotherapy patients ($P<0.05$), as shown in Table 4.

2.3 Construction of a social frailty risk prediction model for cancer chemotherapy patients

Six predictive factors (PHQ-9, COST-PROM, Malnutrition, PSSS, Sleep disorder, CRF) selected through multiple factor analysis were used as independent variables, and five machine learning algorithms—logistic regression, random forest, XGBoost, decision tree, and naive Bayes—were used to construct predictive models. The ROC curve of the training set tumor chemotherapy patient social frailty risk prediction model is shown in Figure 4a, the DCA curve is shown in Figure 4b, the calibration curve is shown in Figure 4c, and the predictive performance is shown in Table 5.

2.4 Validation of the Social Frailty Risk Prediction Model for Cancer Chemotherapy Patients

In the validation set, the predictive ability of five models was evaluated, and the results showed that XGBoost had the highest discrimination ($AUC=0.889$), while Decision Tree had the worst discrimination ($AUC=0.780$). XGBoost has high accuracy, sensitivity, and specificity, indicating its superior predictive accuracy for samples. The Brier score is used to evaluate the model calibration, with lower values indicating better model calibration. XGBoost has the lowest Brier

score and the best improvement in calibration. Overall, XGBoost has the best predictive performance, so further SHAP analysis is needed to visualize its decision-making process. The ROC curve of the social frailty risk prediction model for cancer patients undergoing chemotherapy is shown in Figure 5a, the DCA curve is shown in Figure 5b, the calibration curve is shown in Figure 5c, and the predictive performance is shown in Table 6.

2.5 Interpretability analysis based on SHAP algorithm

2.5.1 Interpretation and analysis of global sample characteristics

XGBoost is interpreted and analyzed by the SHAP algorithm. The order of the importance of the SHAP is shown in Figure 6A. PHQ-9 plays a key role in the risk of social frailty in cancer chemotherapy patients. PSSS is one of the important variables affecting the occurrence of social frailty, followed by sleep disorders and COST-PROM. In order to further clarify the positive/negative relationship between each feature and the target variable, a XGBoost SHAP summary chart is generated, as shown in Figure 6B. Each line represents one feature, each point represents one sample, and the abscissa displays the SHAP value of each sample. The larger the value, the greater the impact of the feature on the model output (positive value represents positive impact, negative value represents negative impact), and the color of the point represents the height of the feature value (light color represents high feature value, and dark color represents low feature value). The results showed that factors such as PHQ-9, Malnutrition, Sleep disorder, and CRF had varying degrees of positive impact (risk factors) on the prediction outcomes, while PSSS and COST-PROM had negative impacts (protective factors).

2.5.2 Attribution analysis of personalized characteristics

On the basis of the SHAP summary chart, the top three clinical characteristics of the model were further output to explain the specific

impact of predictive variables on the prediction results. The horizontal axis of the SHAP dependence chart is the variation range of the clinical feature, and the vertical axis is the SHAP value of the clinical feature. The SHAP value >0 indicates that the risk of social frailty in patients is increased. This study defined the early warning range of PHQ-9, PSSS and sleep disorder according to the SHAP dependence diagram and clinical opinions. The results showed that the risk of social frailty of patients with the three characteristics increased when they reached a specific threshold (PHQ-9 >11.146 , PSSS <44.408 , sleep disorder=yes), as shown in Fig. 7a-c.

3 Discussion

3.1 Current situation of social frailty of cancer patients undergoing chemotherapy

A total of 253 cancer patients who received chemotherapy for the first time were included in this study. The incidence of social frailty after the end of chemotherapy cycle was 32.81%, which was lower than the 39.81% reported in a longitudinal study of prostate cancer patients^[16]. The reason may be that the cancer patients who received chemotherapy for the first time in this study are in the initial stage of tumor treatment. At this time, the patient's family and social support system are relatively centralized, which may buffer the sharp decline of social function to a certain extent^[17]. Wang et al's systematic review of cancer patients' frailty during chemotherapy shows that the overall prevalence of frailty during chemotherapy is 34.50%^[4], which is similar to the prevalence of social frailty in this study, reflecting that the impairment of social function accounts for an important proportion of chemotherapy-related frailty. The results of this study showed that the incidence of social frailty in patients with cancer chemotherapy was higher than that in patients with osteosarcopenia^[18](12.3%) and diabetes^[6] (26.3%). Compared with

patients with other chronic diseases, cancer patients are not only affected by the primary disease, but also need to go to the hospital for periodic chemotherapy many times, which not only leads to the decline of patients' physical function, but also affects patients' social resources and social participation ability through the reduction of social interaction and the limitation of social roles^[19]. In addition, chemotherapy related complications (peripheral neuritis, alopecia, nausea and vomiting, etc.) can easily lead to patients' image disorder and limited physical function^[20], induce stigma and social anxiety, increase psychological pressure, and then induce social frailty. Therefore, screening cancer patients for social frailty before chemotherapy, and psychological counseling and strengthening social support for high-risk groups in the early stage can prevent or delay the development process of social frailty.

3.2 Influencing factors of social frailty of patients with tumor chemotherapy

This study selected influencing factors based on previous research and Health Ecology Theory, and conducted risk prediction for the social frailty of cancer chemotherapy patients. After 10-fold cross validation and internal validation of independent test set, the calibration, discrimination and clinical applicability of the model are good, which provides a reliable risk assessment tool for clinical medical staff. The model constructed in this study includes 6 predictive factors: PHQ-9, PSSS, Sleep disorder, COST-PROM, CRF, and Malnutrition.

3.2.1 PHQ-9

In the SHAP feature importance analysis of this study, PHQ-9 ranked first, proving that depression is an important factor contributing to the social frailty of cancer chemotherapy patients. Relevant research has found that depression causes patients to lose interest in life, avoid social behaviors, resulting in a decline in social participation^[21], and patients

gradually withdraw from social interactions, forming a negative cycle of "depression-social withdrawal-social frailty"^[22]. Secondly, depression and chronic inflammation promote each other, further damaging cognitive functions and social functions by releasing pro-inflammatory cytokines^[23]. Therefore, medical staff should not only pay attention to the physiological functions of chemotherapy patients but also pay attention to mental health. They should provide early psychological intervention for patients with negative emotions and be vigilant against the occurrence of social frailty, achieving dual-level care of both the body and the mind.

3.2.2 PSSS

The results of this study show that improving social support can help patients resist social frailty. The convoy model of social relations points out that^[24], a high level of social support can alleviate the psychological pressure of patients in the treatment process and enhance their psychological resilience in the face of disease; At the same time, it will also provide patients with practical material and care help to reduce the economic and living burden. It is suggested that the medical staff should actively promote the communication between the patients and their families and friends, and understand the needs and feelings of the patients while providing help for the patients; In addition, the community can also regularly organize community activities, guide patients to understand and use community resources, and enhance patients' sense of social participation, so as to reduce the risk of social frailty.

3.2.3 Sleep disorder

The results of this study confirmed that patients with sleep disorders are prone to social frailty. Qi and other studies have also proved that^[25], insomnia is closely related to the limitation of social activities, and insomnia is closely related to anxiety and depression, which can also lead to social frailty. It is suggested that medical staff should incorporate sleep

assessment into the routine nursing of chemotherapy patients, and provide targeted sleep hygiene guidance and necessary drug treatment for patients with sleep disorders as soon as possible.

3.2.4 COST-PROM

Economic toxicity is a unique dilemma faced by cancer patients. The direct medical expenditure, income reduction and indirect economic losses related to chemotherapy constitute the economic burden of patients^[26]. Relevant studies have found that economic toxicity is closely related to patients' health-related social needs^[27]. When health-related needs cannot be met, patients will reduce unnecessary social activities to save money, thereby weakening social functions. Medical staff should take the initiative to screen patients with high economic toxicity, help them understand the medical policy or select eligible people to participate in clinical trials, so as to reduce the adverse impact of economic burden on social function.

3.2.5 CRF

Cancer related fatigue is one of the most common symptoms of cancer patients, which can reduce the physical, psychological and social functions of patients, and bring significant pain to patients and their families^[28]. Relevant studies have pointed out that^[29, 30], cancer-related fatigue will significantly interrupt the social and work role functions of patients, and affect the integration of patients into society. Many studies have pointed out that^[31-33], exercise can reduce systemic inflammation and improve CRF, so that patients can better complete daily life activities, and then improve the quality of life and social participation.

3.2.6 Malnutrition

The results of this study confirm that cancer patients undergoing chemotherapy with malnutrition are more prone to social frailty. The research shows that malnutrition leads to decreased muscle strength,

weakened immune function, and lack of energy in patients, which directly limits their daily activity ability and social participation^[34, 35]. Moreover, chemotherapy-induced gastrointestinal adverse reactions and enhanced catabolism caused by tumors can further exacerbate the patient's malnutrition status^[36]. Malnutrition not only directly weakens physical functions but also reduces patients' tolerance to chemotherapy and increases the risk of treatment-related toxicities^[37], forming a vicious cycle of "chemotherapy adverse reactions → malnutrition → decline in physical functions → reduction in social participation". Therefore, medical staff should dynamically monitor the nutritional status of patients and implement individualized nutritional support plans for patients at risk of malnutrition as soon as possible to improve the nutritional status and thereby reduce the risk of social frailty.

3.3 Comparison of prediction models for social frailty of cancer patients undergoing chemotherapy

This study systematically evaluated the performance of five machine learning algorithms, namely logistic regression, random forest, XGBoost, decision tree, and naive Bayes, in predicting the social frailty risk of cancer chemotherapy patients. The results showed that the XGBoost prediction model achieved the best predictive performance (training set: AUC 0.991, F1 score 0.895; validation set: AUC 0.889, F1 score 0.773). This is related to its ensemble learning mechanism and gradient boosting framework, which can effectively capture the complex nonlinear relationships and interactions between variables^[38]. XGBoost introduced regularization strategies, which improved the model's generalization ability while controlling the risk of overfitting, and still had a high degree of discrimination on the validation set^[39]. The logistic regression model showed good stability on both the training set and validation set, and could directly quantify the direction and intensity of the impact of each

factor on the social frailty risk, but it tended to ignore the complex interactions between variables. Naive Bayes and random forest also achieved good predictive results on the validation set, but the AUC gap between the training set and validation set of random forest was large, suggesting a certain tendency of overfitting; naive Bayes is based on the assumption of feature conditional independence, has higher computational efficiency, and in this study, the variables did not have correlations, so the prediction accuracy was good. The decision tree had the lowest AUC on the validation set, possibly because the sample size was small, which reduced the model's generalization ability. In summary, XGBoost performed outstandingly in terms of discrimination, calibration, and clinical practicability, and was suitable for predicting the social frailty risk of cancer chemotherapy patients. It can be used as an auxiliary tool for early clinical screening, helping medical staff identify high-risk individuals and providing reference for formulating individualized intervention plans.

3.4 Clinical application value of SHAP interpretability analysis

In machine learning models, the "black-box" nature renders the internal decision-making process of the model lacking in interpretability. To enhance the transparency of the model's decision-making process, this study conducts an interpretive analysis of the optimal model, XGBoost, using the SHAP method, making the model's decision-making process traceable and understandable. At the global level, the SHAP feature importance ranking and summary plots are used to systematically present the contribution degree and direction of each variable. PHQ-9, Malnutrition, Sleep disorder, and CRF have varying degrees of positive impacts on the prediction results (risk factors), while PSSS and COST-PROM have negative impacts (protective factors). This is highly consistent with the results of the multivariate regression, indirectly

reflecting the reliability of the core predictors. At the individual level, by conducting SHAP dependence plot analysis on the top three clinical features influencing the model, the early-warning thresholds of the key features are further defined: when PHQ-9>11.146, PSSS<44.408, and Sleep disorder=Yes, the risk of social frailty in patients significantly increases. It is suggested that medical staff can evaluate these quantitative indicators upon admission to screen high-risk populations.

4 Limitations and future directions

This study has the following limitations: (1) The sample was sourced from a single center, and the sample size was only 253 cases. The extrapolability of the model still needs further verification. It is recommended that future research conduct prospective cohort studies with multi-center and large-scale samples to further enhance the predictive ability of the model. (2) This study only collected relevant data before admission chemotherapy and after the completion of chemotherapy, without conducting dynamic evaluations of psychological and social factors during the chemotherapy process. In the future, repeated measurement designs can be adopted to construct dynamic prediction models to reflect the internal relationship between social frailty and chemotherapy cycles. (3) This study only conducted internal validation and lacked independent data for external validation. It is suggested that future research be carried out in multiple centers to evaluate the applicability of the model in hospitals of different levels. (4) All the predictive factors in this study were obtained from patients' subjective reports, lacking objective indicators. Future research can consider combining objective indicators from wearable devices and medical records to enrich the dimensions of predictive factors. (5) The Cronbach's α coefficient of the social frailty scale used in this study was 0.602, indicating that it is acceptable but has low internal consistency,

which may affect the stability of social frailty measurement. It is recommended that future research use social frailty assessment tools with higher reliability and validity, or combine multiple tools for comprehensive evaluation.

5 Conclusion

The results of this study indicate that the incidence of social frailty among patients undergoing cancer chemotherapy is at a relatively high level. The five predictive models constructed based on six factors (PHQ-9, PSSS, Sleep disorder, COST-PROM, CRF, Malnutrition), namely logistic regression, random forest, XGBoost, decision tree and naive Bayes, have good discrimination, calibration and clinical practicability. Among them, XGBoost demonstrated the best predictive performance, highlighting the significant advantages of machine learning in disease prediction. By interpreting XGBoost based on the SHAP algorithm, the transparency and reliability of the model's decision-making process were enhanced, facilitating healthcare professionals to better understand the model's prediction process. Although the model performed well in identifying social frailty, it still needs to expand the sample size for verification and enhance its applicability in different environments.

Table 1 Clinical data of social frailty of tumor chemotherapy patients in the training set

Variable		level	Non-social frailty (119)	Social frailty (59)	statistical value	p
Age			61.53±9.08	59.25±8.77	t = 1.613	0.113
Gender	Male		67 (56.3)	31 (52.5)	$\chi^2 = 0.225$	0.753
	Female		52 (43.7)	28 (47.5)		
Marital status	Unmarried		9 (7.6)	8 (13.6)	$\chi^2 = 1.566$	0.312
	Married		110 (92.4)	51 (86.4)		
Education	junior high school		46 (38.7)	19 (32.2)	$\chi^2 = 7.984$	0.019
	high school and above		37 (31.1)	10 (16.9)		
	Primary school and below		36 (30.3)	30 (50.8)		
Income (In CNY)	CNY 2,000–4,000		48 (40.3)	19 (32.2)	$\chi^2 = 7.029$	0.031
	below CNY 2,000		41 (34.5)	32 (54.2)		
	above CNY 4,000		30 (25.2)	8 (13.6)		
malnutrition	No		78 (65.5)	23 (39.0)	$\chi^2 = 11.339$	0.001
	Yes		41 (34.5)	36 (61.0)		
Insurance	Resident medical insurance		26 (21.8)	11 (18.6)	$\chi^2 = 0.883$	0.646
	Rural Cooperative Medical Insurance		68 (57.1)	38 (64.4)		
	Employee Medical Insurance		25 (21.0)	10 (16.9)		
Chronic disease	No		80 (67.2)	24 (40.7)	$\chi^2 = 11.402$	0.001
	Yes		39 (32.8)	35 (59.3)		
Lesion location	Respiratory system tumors		50 (42.0)	21 (35.6)	$\chi^2 = 7.187$	0.243
	Digestive system tumors		36 (30.3)	22 (37.3)		
	Female reproductive system tumors		14 (11.8)	2 (3.4)		
	Head and neck tumors		4 (3.4)	1 (1.7)		
	Breast tumors		9 (7.6)	8 (13.6)		
	Other		6 (5.0)	5 (8.5)		
Sleep disorder	No		82 (68.9)	13 (22.0)	$\chi^2 = 36.204$	<0.001
	Yes		37 (31.1)	46 (78.0)		
BMI			23.14±3.38	22.15±3.63	t=1.752	0.074
PHQ.9			9.72±4.75	13.51±4.11	t=5.494	<0.001
CRF			26.59±7.57	32.19±7.31	t=4.755	<0.001
PSSS			56.34±10.99	47.32±12.80	t=4.632	<0.001
COST PROM			17.09±7.47	12.75±4.62	t=4.762	<0.001
CV.CIRS			57.36±10.35	54.22±10.47	t=1.891	0.059

Table 2 Univariate logistic regression analysis results of influencing factors of social frailty in cancer patients

Variable	Level	Coefficient	Std Error	OR	95%CI	Overall_P
Education	junior high school	-0.424	0.449	0.654	(0.272~1.577)	0.018
	Primary school and below	0.702	0.368	2.018	(0.981~4.151)	0.018
	high school and above					
Income (In CNY)	below CNY 2,000	0.679	0.359	1.972	(0.975~3.988)	0.03
	above CNY 4,000	-0.395	0.481	0.674	(0.262~1.731)	0.03
	CNY 2,000~4,000					
malnutrition	Yes	1.091	0.329	2.978	(1.562~5.678)	0.001
	No					
Chronic disease	Yes	1.096	0.329	2.991	(1.569~5.703)	0.001
	No					
Sleep disorder	No	2.059	0.371	7.842	(3.787~16.237)	<0.001
	Yes					
PHQ-9		0.174	0.038	1.19	(1.105~1.281)	<0.001
CRF		0.099	0.023	1.104	(1.055~1.156)	<0.001
PSSS		-0.064	0.015	0.938	(0.911~0.965)	<0.001
COST-PROM		-0.1	0.027	0.904	(0.859~0.953)	<0.001

Table 3 VIF values of the seven variables selected by LASSO regression

Variable	VIF
malnutrition	1.136
Chronic disease	1.100
Sleep disorder	1.125
PHQ.9	1.295
CRF	1.045
PSSS	1.096
COST PROM	1.070

Table 4 Results of multivariable logistic regression analysis of factors influencing social frailty in cancer patients

Variable	Level	Coefficient	Std Error	Wald	OR	95%CI	P
PHQ.9		0.158	0.054	8.460	1.171	1.053~1.302	0.004
COST PROM		-0.099	0.036	7.484	0.906	0.844~0.972	0.006
Malnutrition	Yes	1.246	0.456	7.456	3.475	1.421~8.495	0.006
	No						
PSSS		-0.049	0.019	6.662	0.952	0.917~0.988	0.010
Sleep disorder	Yes	1.136	0.467	5.921	3.115	1.247~7.781	0.015
	No						

CRF		0.064	0.028	5.302	1.067	1.010~1.127	0.021
Chronic disease	Yes	0.808	0.449	3.236	2.244	0.930~5.411	0.072
	No						

Table 5 Prediction performance of the social frailty prediction model for cancer patients undergoing chemotherapy in the training set

Model	AUC	95%AUC	Accuracy	Sensitivity	Specificity	Precision	F1	Brier
XGBoost	0.991	0.980~0.999	0.853	0.922	0.708	0.870	0.895	0.051
Random Forest	0.973	0.951~0.990	0.840	0.941	0.625	0.842	0.889	0.083
Decision Tree	0.907	0.861~0.950	0.800	0.784	0.833	0.909	0.842	0.101
Naive Bayes	0.896	0.843~0.942	0.773	0.824	0.667	0.840	0.832	0.120
Logistic	0.873	0.817~0.925	0.760	0.941	0.375	0.762	0.842	0.148

Table 6 Prediction Performance of the social frailty Prediction Model for Cancer Chemotherapy Patients in the Validation Set

Model	AUC	95%AUC	Accuracy	Sensitivity	Specificity	Precision	F1	Brier
XGBoost	0.889	0.790~0.966	0.867	0.708	0.941	0.850	0.773	0.110
Naive Bayes	0.885	0.801~0.953	0.773	0.667	0.824	0.640	0.653	0.126
Random Forest	0.879	0.776~0.963	0.853	0.708	0.922	0.810	0.756	0.122
Logistic	0.862	0.767~0.940	0.760	0.375	0.941	0.750	0.500	0.148
Decision Tree	0.780	0.648~0.896	0.827	0.667	0.902	0.762	0.711	0.159

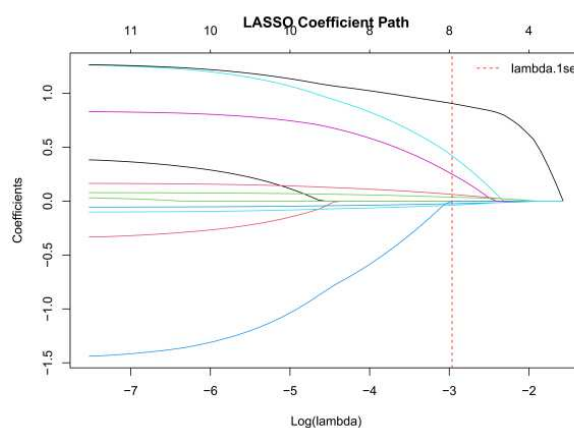


Figure 1 LASSO regression coefficient path

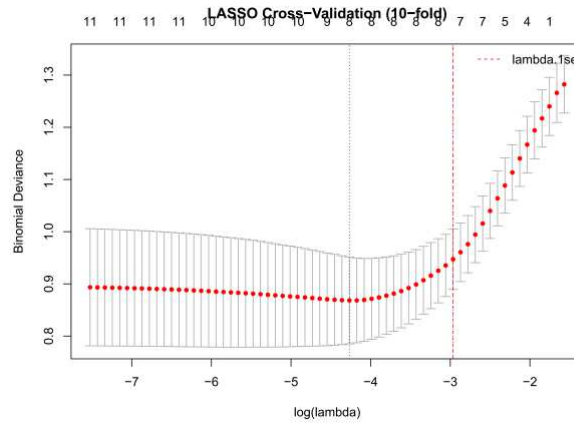


Figure 2 LASSO regression cross validation curve

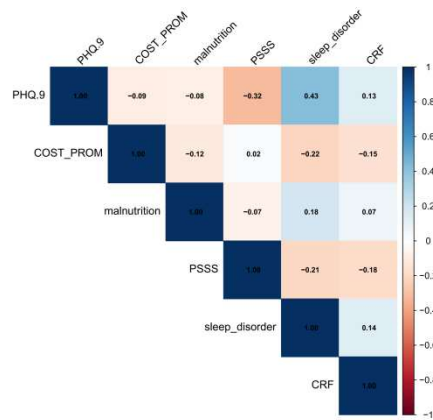


Figure 3 Correlation heatmap analysis

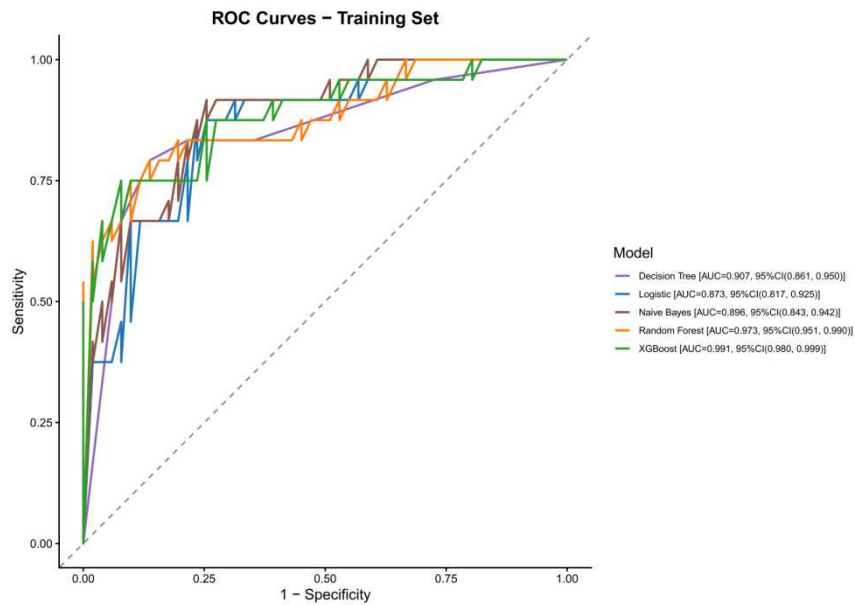


Figure 4a ROC curve of the social frailty risk prediction model for cancer patients undergoing chemotherapy in the training set

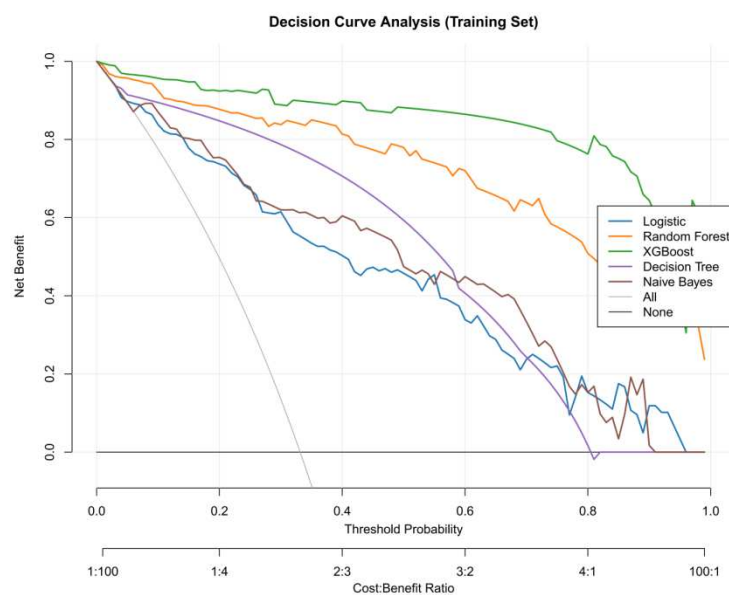


Figure 4b DCA curve of the social frailty risk prediction model for cancer patients undergoing chemotherapy in the training set

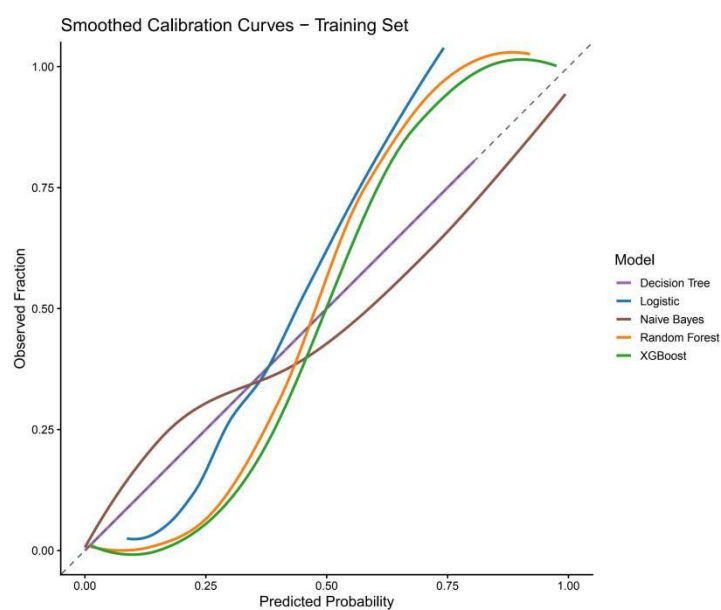


Figure 4c Calibration curve of social frailty risk prediction model for cancer patients undergoing chemotherapy in the training set

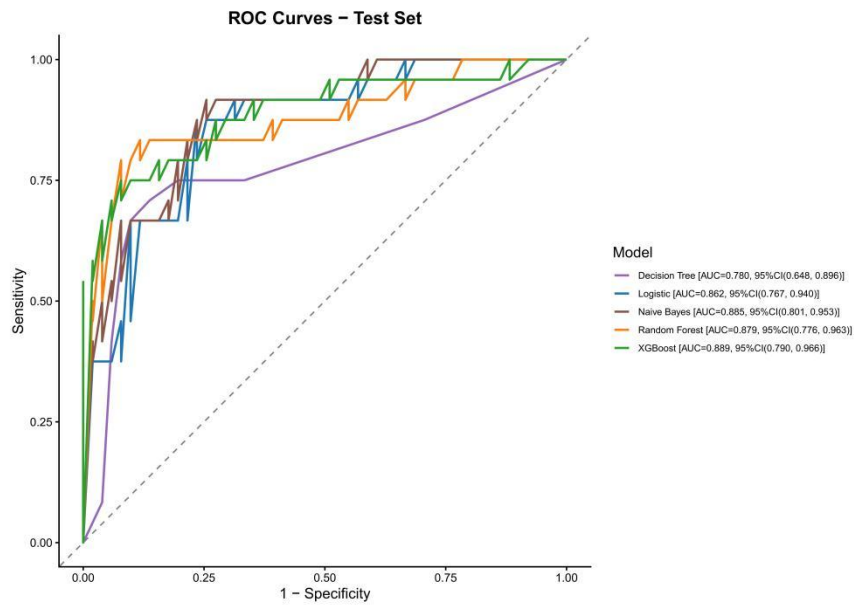


Figure 5a: ROC curve of the social frailty risk prediction model for cancer patients undergoing chemotherapy in the validation set

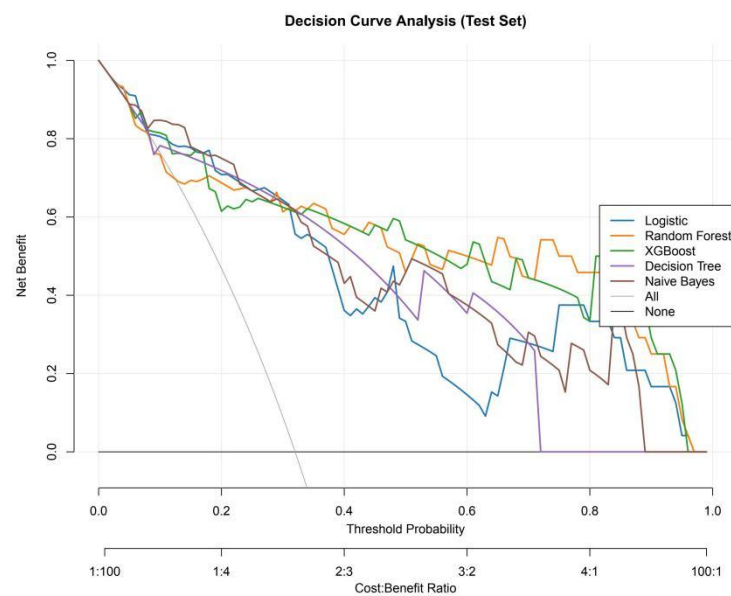


Figure 5b Validation set DCA curve of social frailty risk prediction model for cancer patients undergoing chemotherapy

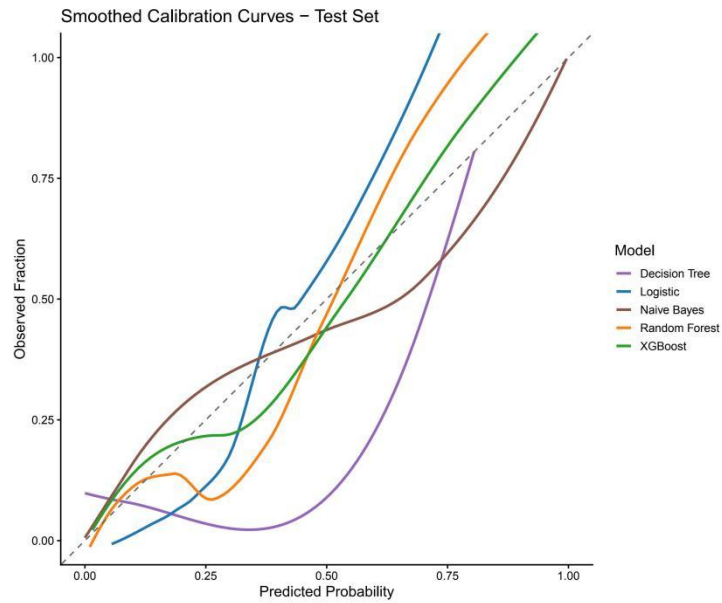


Figure 5c Calibration curve of social frailty risk prediction model for cancer patients undergoing chemotherapy in the validation set

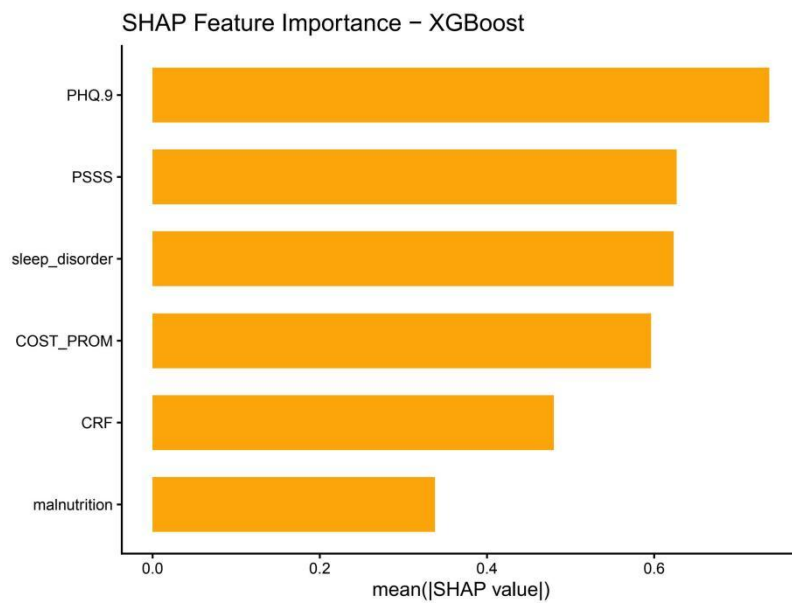


Figure 6a SHAP feature importance ranking chart

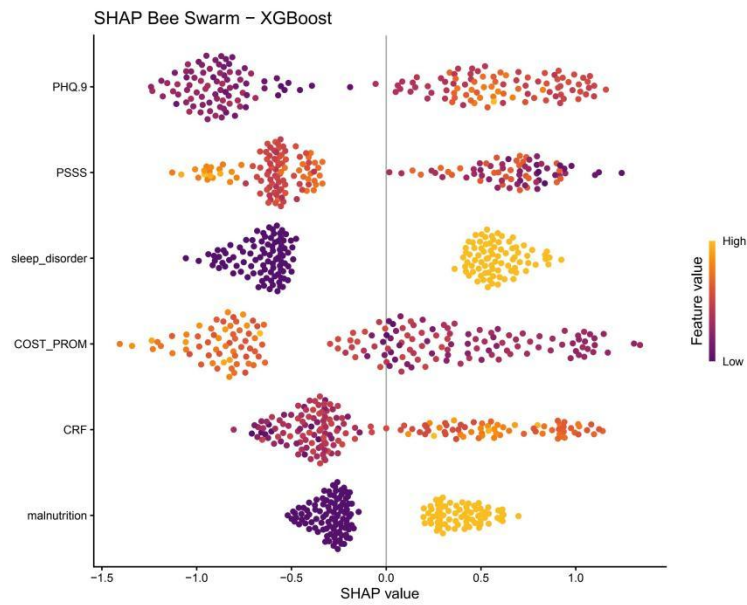


Figure 6b SHAP summary

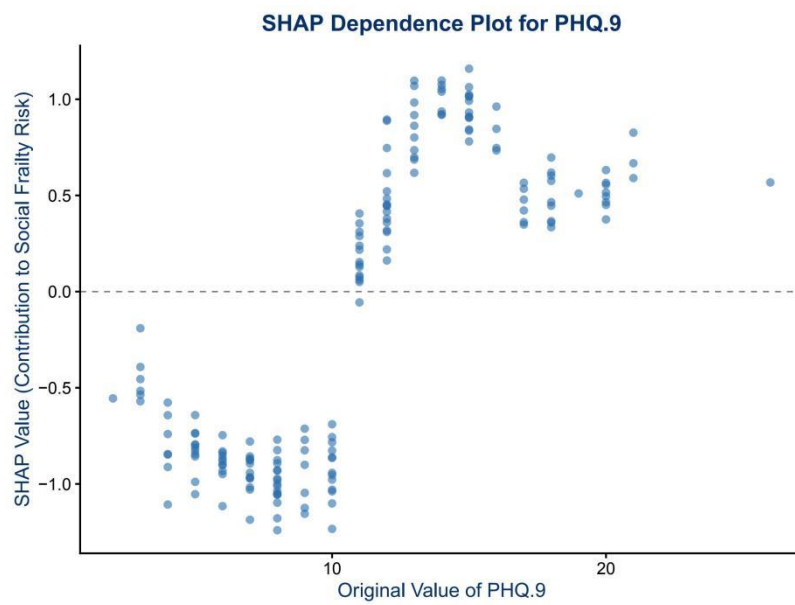


Figure 7a SHAP dependency diagram of PHQ-9

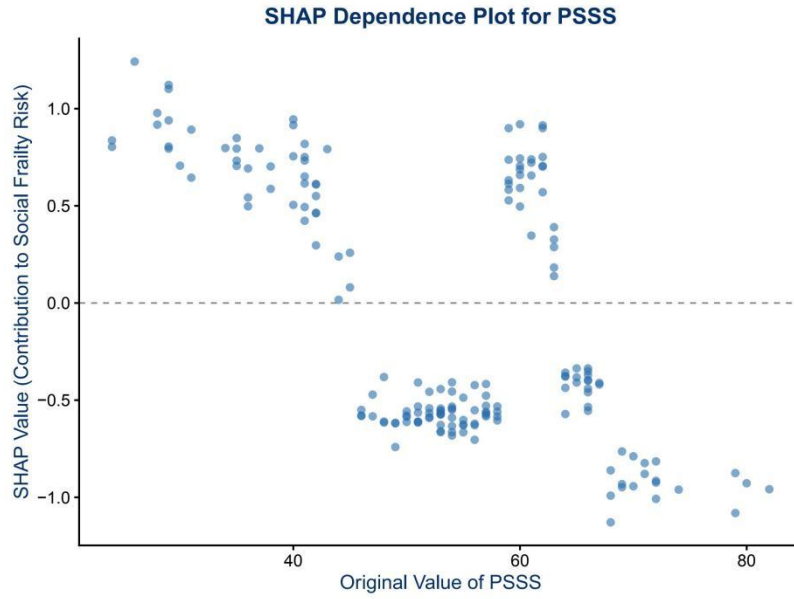


Figure 7b SHAP dependency diagram of PSSS

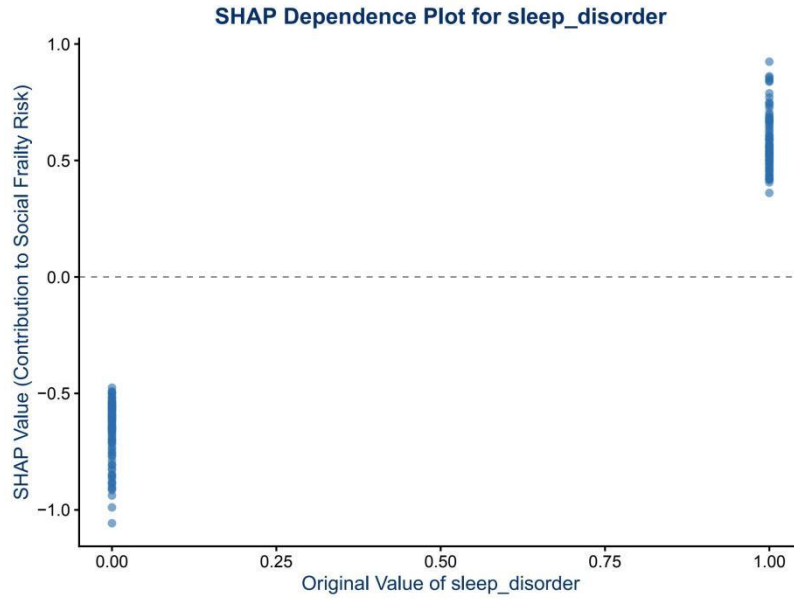


Figure 7c SHAP dependency diagram of sleep-disorder

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[J]. CA: a cancer journal for clinicians, 2024, 74(3): 229-263.
- [2] Zheng R S, Chen R, Han B F, et al. Cancer incidence and mortality in China, 2022[J]. Zhonghua zhong liu za zhi [Chinese journal of oncology], 2024, 46(3): 221-231.
- [3] van der Elzen C, Aires F, Rodrigues E D, et al. Clinical Outcomes and Toxicity Profile of Chemoradiotherapy in Older Versus Younger Patients With Anal Cancer: A Retrospective Cohort Analysis[J]. Clinical oncology (Royal College of Radiologists (Great Britain)), 2025, 47: 103929.

- [4] Wang T, Jiang J, Song Z, et al. Prevalence of Frailty and Its Predictors Among Patients With Cancer at the Chemotherapy Stage: Systematic Review[J]. JMIR cancer, 2025, 11: e69936.
- [5] Bunt S, Steverink N, Olthof J, et al. Social frailty in older adults: a scoping review[J]. European journal of ageing, 2017, 14(3): 323-334.
- [6] Liang B, Liang W, Xu D, et al. Current status and factors associated with social frailty in older adults with type 2 diabetes: a cross-sectional study[J]. Frontiers in public health, 2026, 14: 1891728.
- [7] Mizuta M, Kondo S, Hibi A, et al. Association between preoperative social frailty and malnutrition six months post-surgery in older patients with gastrointestinal cancer: A prospective cohort study[J]. Journal of geriatric oncology, 2024, 15(6): 101782.
- [8] Petch J, Di S, Nelson W. Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology[J]. The Canadian journal of cardiology, 2022, 38(2): 204-213.
- [9] Ma L, Sun F, Tang Z. Social Frailty Is Associated with Physical Functioning, Cognition, and Depression, and Predicts Mortality[J]. J Nutr Health Aging, 2018, 22(8): 989-995.
- [10] Kroenke K, Spitzer R L, Williams J B. The PHQ-9: validity of a brief depression severity measure[J]. Journal of general internal medicine, 2001, 16(9): 606-613.
- [11] Okuyama T, Akechi T, Kugaya A, et al. Development and validation of the cancer fatigue scale: a brief, three-dimensional, self-rating scale for assessment of fatigue in cancer patients[J]. Journal of pain and symptom management, 2000, 19(1): 5-14.
- [12] Buysse D J, Reynolds C F R, Monk T H, et al. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research[J]. Psychiatry research, 1989, 28(2): 193-213.
- [13] de Souza J A, Yap B J, Hlubocky F J, et al. The development of a financial toxicity patient-reported outcome in cancer: The COST measure[J]. Cancer, 2014, 120(20): 3245-3253.
- [14] Glasgow R E, Toobert D J, Barrera M J, et al. The Chronic Illness Resources Survey: cross-validation and sensitivity to intervention[J]. Health education research, 2005, 20(4): 402-409.
- [15] Zimet G D, Powell S S, Farley G K, et al. Psychometric characteristics of the Multidimensional Scale of Perceived Social Support[J]. J Pers Assess, 1990, 55(3-4): 610-617.
- [16] Wan Y, Li N, Zhuang S, et al. How Does Social Frailty Evolve Among Patients with Prostate Cancer? Evidence from Regression Models versus Fuzzy Set Qualitative Comparative Analysis[J]. Risk management and healthcare policy, 2026, 19: 597629.
- [17] Wei H, He W. Social Support and Chemotherapy-Induced Peripheral Neuropathy in Breast Cancer: A Full Mediation Dual-Pathway Model Through Physical Activity and Psychological Distress[J]. Psycho-oncology, 2026, 35(4): e70470.
- [18] Inoue T, Maeda K, Satake S, et al. Osteosarcopenia, the co-existence of osteoporosis and sarcopenia, is associated with social frailty in older adults[J]. Aging clinical and experimental research, 2022, 34(3): 535-543.
- [19] Wei J, Bai Z, Bu M, et al. Influence of symptom burden on social alienation in lung cancer patients receiving chemotherapy: the chain mediating effect of perceived social support and positive psychological capital[J]. BMC psychology, 2025, 13(1): 99.
- [20] Kim S, Kim S, Park H. Mediating roles of cancer stigma and resilience in the relationship between type D personality and quality of life among patients with breast cancer undergoing chemotherapy[J]. European journal of oncology nursing : the official journal of European Oncology Nursing Society, 2026, 81: 103122.

- [21] Tang J, Shen C, Wu L, et al. Prevalence and potential influencing factors of social frailty among community-dwelling older adults in China: systematic review and meta-analysis[J]. BMC geriatrics, 2026: 10-1186.
- [22] Ostir G V, Ottenbacher K J, Fried L P, et al. The Effect of Depressive Symptoms on the Association Between Functional Status and Social Participation[J]. Social indicators research, 2007, 80(2): 379-392.
- [23] Leonard B E. Inflammation, depression and dementia: are they connected?[J]. Neurochemical research, 2007, 32(10): 1749-1756.
- [24] Antonucci T C, Ajrouch K J, Birditt K S. The convoy model: explaining social relations from a multidisciplinary perspective[J]. The Gerontologist, 2014, 54(1): 82-92.
- [25] Qi X, Kong H, Li J, et al. The relationship between insomnia and multidimensional frailty in community-dwelling older adults: A systematic review[J]. Archives of gerontology and geriatrics, 2025, 129: 105661.
- [26] Smith G L, Banegas M P, Acquati C, et al. Navigating financial toxicity in patients with cancer: A multidisciplinary management approach[J]. CA: a cancer journal for clinicians, 2022, 72(5): 437-453.
- [27] Zheng Z, Hu X, Banegas M P, et al. Health-related social needs, medical financial hardship, and mortality risk among cancer survivors[J]. Cancer, 2024, 130(17): 2938-2947.
- [28] Portenoy R K, Itri L M. Cancer-related fatigue: guidelines for evaluation and management[J]. The oncologist, 1999, 4(1): 1-10.
- [29] Jones J M, Howell D, Longo C, et al. The association of cancer-related fatigue on the social, vocational and healthcare-related dimensions of cancer survivorship[J]. Journal of cancer survivorship : research and practice, 2025, 19(1): 97-108.
- [30] Behringer K, Goergen H, Müller H, et al. Cancer-Related Fatigue in Patients With and Survivors of Hodgkin Lymphoma: The Impact on Treatment Outcome and Social Reintegration[J]. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2016, 34(36): 4329-4337.
- [31] Hilfiker R, Meichtry A, Eicher M, et al. Exercise and other non-pharmaceutical interventions for cancer-related fatigue in patients during or after cancer treatment: a systematic review incorporating an indirect-comparisons meta-analysis[J]. British journal of sports medicine, 2018, 52(10): 651-658.
- [32] Hu Y, Gu S, Bu Z, et al. Effect of exercise for patients with advanced lung cancer and cancer-related fatigue: A systematic review and meta-analysis[J]. Journal of sport and health science, 2025, 14: 101017.
- [33] Fabi A, Bhargava R, Fatigoni S, et al. Cancer-related fatigue: ESMO Clinical Practice Guidelines for diagnosis and treatment[J]. Annals of oncology : official journal of the European Society for Medical Oncology, 2020, 31(6): 713-723.
- [34] Kobayashi N, Tokuno S, Nozoe M, et al. Relationship between social frailty and malnutrition in community-dwelling older adults undergoing rehabilitation[J]. Clinical nutrition ESPEN, 2025, 69: 318-322.
- [35] Makizako H, Kubozono T, Kiyama R, et al. Associations of social frailty with loss of muscle mass and muscle weakness among community-dwelling older adults[J]. Geriatrics & gerontology international, 2019, 19(1): 76-80.
- [36] Arends J, Bachmann P, Baracos V, et al. ESPEN guidelines on nutrition in cancer patients[J]. Clinical nutrition (Edinburgh, Scotland), 2017, 36(1): 11-48.

[37] Aprile G, Basile D, Giaretta R, et al. The Clinical Value of Nutritional Care before and during Active Cancer Treatment[J]. *Nutrients*, 2021, 13(4): 1196.

[38] Lin S, Lu W, Wang T, et al. Predictive model of acute kidney injury in critically ill patients with acute pancreatitis: a machine learning approach using the MIMIC-IV database[J]. *Renal failure*, 2024, 46(1): 2303395.

[39] Hou N, Li M, He L, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost[J]. *Journal of translational medicine*, 2020, 18(1): 462.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by [Wang Xinyi], [Shen Jinlian] and [Wang Rong]. The first draft of the manuscript was written by [Yao Chen] and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics approval

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Hanzhong Central Hospital (Approval No. 2026089).

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Clinical trial number: not applicable.