

Article

Development and Validation of a Machine Learning-Based Prediction Model for Illness Uncertainty in Patients with Malignant Tumors

Yan Lu ^{1,†}, Hui Zeng ^{1,2,†}, Li Mao ¹, Jingjing Gong ¹, Yuxuan Wei ^{1,2} and Xiaodan Li ^{1,*}¹ Department of Obstetrics and Gynecology, Peking University People's Hospital, Beijing 100044, China; luyan@pkuph.edu.cn (Y.L.)² College of Nursing, Hebei University, Baoding 071002, China

* Correspondence: lixiaodan@pkuph.edu.cn; Tel.: +86-150-1030-5099

† These authors contributed equally to this work.

Highlights

What are the main findings?

- Anxiety was the only predictor shared by LASSO, Random Forest, and XGBoost. Seven consensus predictors supported by at least two algorithms were used to construct the nomogram.
- The nomogram showed good internal discrimination (AUC = 0.724) but exhibited calibration bias in time-stratified temporal validation ($p < 0.05$), revealing temporal heterogeneity.

What are the implications of the main findings?

- The model shifts illness uncertainty identification from post hoc assessment to admission screening, enabling proactive risk stratification in clinical practice.
- Modifiable predictors (anxiety, depression, social support, coping styles) offer clear targets for early psychosocial interventions.

Abstract

Objectives: To develop and validate an individualized prediction model for assessing the risk of illness uncertainty in patients with malignant tumors, using a cross-sectional design. **Methods:** Patients with malignant tumors treated at Peking University People's Hospital from August 2024 to January 2025 were enrolled. The Mishel Uncertainty in Illness Scale (MUIS) was used to classify patients into high-risk and low-to-moderate risk groups. Patients were divided into a model development set and a time-stratified validation set. The development set was further randomly split into a training set and an internal validation set at a 7:3 ratio. Three machine learning algorithms Least Absolute Shrinkage and Selection Operator (LASSO), Random Forest (RF), and eXtreme Gradient Boosting (XGBoost) were employed to screen for common predictors, and a nomogram was constructed based on logistic regression. The model's performance was evaluated using the Area Under the Receiver Operating Characteristic Curve, calibration curves, and the Hosmer–Lemeshow test. **Results:** A total of 966 patients were included, with 676 in the development set and 290 in the time-stratified validation set. Seven predictors were ultimately identified for the nomogram: age, education level, diagnosis, depression, anxiety, medical coping modes, and social support. Notably, anxiety was the only variable jointly identified by all three algorithms, while the other six were commonly selected by both Random Forest and XGBoost. The nomogram showed good discrimination in training (AUC = 0.763,



Academic Editors: Philip R. Debruyne, Patricia Schofield, Laura Tack and Daniele Giansanti

Received: 30 May 2026

Revised: 9 July 2026

Accepted: 14 July 2026

Published: 17 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

95% CI: 0.717–0.808) and internal validation (AUC = 0.724, 95% CI: 0.645–0.803), with well-calibrated probabilities ($p > 0.05$). In time-stratified validation, discrimination was acceptable (AUC = 0.663, 95% CI: 0.58–0.739), but calibration revealed risk underestimation for low-to-moderate risk group ($p < 0.05$). **Conclusions:** The nomogram shows acceptable internal performance as an exploratory concurrent classifier rather than a genuine predictor, but calibration bias and modest temporal discrimination (AUC = 0.663) indicate it is not ready for clinical use. Further optimization and multi-center prospective validation are required.

Keywords: illness uncertainty; malignant tumors; nomogram; anxiety; depression

1. Introduction

Illness uncertainty refers to a patients' inability to determine the meaning of illness-related events and accurately predict the disease outcome [1]. This phenomenon is prevalent among patients with malignant tumors, with an incidence rate as high as 30–50% [2]. With continuous improvements in early diagnosis and treatment methods, the survival period of these patients has significantly prolonged [3]. However, as a persistent psychological stressor, illness uncertainty severely impacts their treatment compliance and quality of life [4]. Extensive research shows that illness uncertainty not only has a reciprocal causal relationship with negative emotions like anxiety and depression but may also affect tumor progression and long-term survival through neuroendocrine and immune pathways [5,6]. Therefore, early identification of patients with a high risk of illness uncertainty and the implementation of stratified psychological interventions are of great clinical significance for improving their overall prognosis.

Currently, clinical assessment of illness uncertainty primarily relies on the Mishel Uncertainty in Illness Scale [7], which comprises 33 items and requires approximately 10–15 min to complete. While brief versions such as the 22-item MUIS-C have been developed to reduce respondent burden [8], these tools still serve as direct measures of how patients perceive uncertainty, rather than as predictors of its occurrence [9]. Our model is not intended to replace the MUIS as a diagnostic tool. Instead, it offers a complementary approach to risk stratification, shifting the focus from direct measurement to predictive screening. The model draws on data that are either routinely available in clinical records or readily collectible via brief, validated questionnaires, thereby moving illness uncertainty identification from a reactive, post hoc exercise to a proactive screening process at admission. Of note, this does not mean that questionnaire-based data collection can be completely avoided; rather, it repurposes such data for early risk prediction. The predictive model developed in this study, through further external validation and implementation research, could be assessed for its feasibility and performance in real-world clinical practice.

In recent years, risk prediction models have been widely applied in the management of psychological complications in patients with tumors [10,11]. Based on common risk factors, these models use statistical methods to establish quantitative relationships between predictors and outcomes, thereby individualizing the assessment of event probability [12]. Compared to traditional single-variable analyses, machine learning can handle high-dimensional data and capture non-linear interactions between variables. By integrating multiple algorithms, it can screen for more robust and clinically meaningful predictors [13,14]. Among them, a nomogram serves as a visual representation of a risk prediction model. By integrating the scores of multiple independent predictive indicators, it can intuitively predict individual outcome risks and has been widely used in oncology for

prognosis judgment and complication risk prediction [15,16]. Compared to pure machine learning models that require complex computer knowledge, a nomogram only requires a ruler or simple calculations to derive a risk value, demonstrating better operability and scalability in clinical practice.

To date, most researchers have explored the influencing factors of illness uncertainty. A cross-sectional study on breast cancer patients found that social support levels and coping styles were significantly negatively correlated with illness uncertainty [17]. Another survey including 586 patients with lung cancer showed that anxiety, depression, and disease stage were independent predictors of illness uncertainty [18]. Additionally, demographic and clinical characteristics such as age, educational level, and treatment stage have also been confirmed to be associated with illness uncertainty [19,20]. However, although several systematic reviews and meta-analyses have synthesized these correlates, the majority of primary studies remain cross-sectional in design, focusing on correlation analyses of single or limited psychosocial variables; consequently, they do not provide predictive models that utilize multi-algorithm joint screening and can quantify individual risk. More critically, existing models severely lack consideration of the temporal dimension, such as differences in patient characteristics and treatment modes during different admission periods, and their stability and extrapolability have not been rigorously verified through time-stratified validation [21]. Because illness uncertainty is comprehensively affected by multi-dimensional factors such as a patient's cognition, emotions, and social support, predictors screened by a single machine learning algorithm may have biases. Integrating multiple algorithms to screen core variables jointly can significantly improve the robustness and clinical interpretability of the predictors [22].

Therefore, this study uses the MUIS score as the outcome and employs LASSO, Random Forest, and XGBoost to jointly screen predictors. Each algorithm offers complementary strengths: LASSO shrinks irrelevant coefficients, Random Forest captures non-linear interactions, and XGBoost provides gradient boosted importance. Taking the intersection of variables selected by all three reduces overfitting risk from single algorithm choice. The final nomogram is deliberately based on logistic regression rather than a black box ensemble, ensuring transparent, bedside usable scoring while still benefiting from machine learning-guided variable preselection. We construct an individualized nomogram and evaluate it via internal and time-stratified validation, explicitly testing temporal stability, a dimension largely overlooked previously. The aim is a simple, objective tool for early risk stratification, potentially shifting uncertainty identification from post hoc assessment to admission screening.

2. Materials and Methods

2.1. Study Design

This was a single-center, cross-sectional study conducted at the Department of Oncology, Peking University People's Hospital, Beijing, between August 2024 and January 2025. Patients were consecutively recruited from inpatient wards and outpatient clinics. Eligibility was assessed through a two-step process: initial screening of electronic medical records, followed by confirmation during routine consultations with the attending physician. The study was performed in accordance with the Declaration of Helsinki. The protocol was approved by the hospital's Institutional Review Board (Ethics Approval Number: 2024PHB227-001), and written informed consent was obtained from all participants prior to enrollment.

2.2. Participants

Inclusion criteria were: (1) age ≥ 18 years; (2) histopathologically or cytologically confirmed diagnosis of a malignant tumor, all clinical stages, including locally advanced and metastatic disease were eligible, primary tumor types covered digestive system, respiratory system, gynecological, breast, hematologic, urologic, bone tumors, and other miscellaneous malignancies; (3) alert, oriented, with no communication barriers, and capable of independently completing self-administered questionnaires; (4) provision of written informed consent.

Exclusion criteria comprised: (1) admission to the intensive care unit during the study period; (2) terminal illness, operationally defined as an estimated life expectancy of less than 3 months or an Eastern Cooperative Oncology Group (ECOG) performance status of 4, as determined by the primary oncology team; (3) pre-existing severe psychiatric disorders (e.g., schizophrenia, bipolar disorder, or major depressive disorder with psychotic features) or documented dementia/cognitive impairment that could potentially confound the assessment of anxiety and depressive symptoms.

2.3. Sample Size

The sample size was planned to support subsequent prediction model development. With 18 candidate predictors, we adopted the rule of 10 events per predictor variable [23]. Based on an estimated prevalence of approximately 30% for high illness uncertainty in our target population, as suggested by previous literature and our local pilot data [2,18], the minimum required sample was calculated as 600 patients.

2.4. Data Collection

Data were collected using an electronic self-administered questionnaire created and distributed via the Wenjuanxing platform. A standardized instruction script was read to all participants, explaining the study purpose, confidentiality, and voluntariness, as well as basic operational steps. After providing written informed consent, patients scanned a QR code with their mobile phones to access the questionnaire. Researchers remained available to offer technical assistance only and did not interpret, paraphrase, or otherwise influence responses to any item, in order to minimize social desirability bias. For missing item-level data, we applied the following rule: if fewer than 10% of items on a given scale were missing, the missing values were imputed using the subscale mean; otherwise, the entire questionnaire was considered invalid and excluded.

Based on literature review and clinical experience [2,24], 18 candidate predictors were selected, covering three domains: (1) demographic characteristics (age, gender, education, residence, monthly household income, marital status, occupation, payment method, religious belief, and primary caregiver); (2) disease-related clinical features (disease duration, tumor type, smoking, and alcohol use); (3) psychosocial factors. The psychosocial instruments were administered in their validated Chinese versions, as follows:

Depression assessed using the 9-item Patient Health Questionnaire (PHQ-9) [25]. This study referred to the Chinese version adapted by Bian et al. [26]. Each item is scored from 0 to 3, with total scores ranging from 0 to 27. Scores of 0–4 indicate no depression, 5–9 mild, 10–14 moderate, 15–19 moderately severe, and 20–27 severe depression. In the present study, the Cronbach's α for the PHQ-9 was 0.825.

Anxiety evaluated using the 7-item Generalized Anxiety Disorder Scale (GAD-7) [27]. The study used the version translated by He et al. [28]. Each item is scored from 0 to 3, with total scores ranging from 0 to 21. Scores of 0–4 indicate no anxiety, 5–9 mild, 10–14 moderate, and 15–21 severe anxiety. The GAD-7 showed a Cronbach's α of 0.840 in this sample.

Medical coping measured using the Medical Coping Modes Questionnaire (MCMQ) [29]. This study referred to the Chinese version adapted by Shen et al. [30], which consists of 20 items across three subscales: confrontation, avoidance, and resignation. Each item is rated on a 1-to-4 Likert scale; higher subscale scores indicate a greater tendency to adopt that particular coping style. In the present study, the Cronbach's α for the MCMQ was 0.901.

Social support assessed using the Social Support Rating Scale (SSRS). The SSRS was originally developed by Xiao et al. [31], which comprises 10 items covering three dimensions: objective support, subjective support, and utilization of support. Scoring follows the established manual, with higher total scores reflecting greater perceived social support. In the present study, the Cronbach's α for the SSRS was 0.768.

2.5. Patient Illness Uncertainty Assessment and Grouping

The Mishel Uncertainty in Illness Scale (MUIS) was first developed by Mishel [32] in 1981. This study used a revised Chinese version that has undergone reliability and validity testing [7]. The scale consists of 33 items across 4 dimensions: ambiguity (13 items), complexity (7 items), lack of information (7 items), and unpredictability (5 items). Each item is scored on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Notably, item 15 is not included in the total score; therefore, the effective total score ranges from 32 to 160. In the current sample, the Cronbach's α coefficient for the total MUIS was 0.828, with subscale values ranging from 0.73 to 0.89, indicating good internal consistency.

Based on the score distribution, patients were initially classified into low (32–74.7), moderate (74.8–117.4), and high (117.5–160) uncertainty levels. Clinically, the moderate group represents a gray zone with considerable distress, yet these patients remain below the high uncertainty threshold. For a binary screening tool designed to identify those most in need of intensive intervention, combining the low and moderate groups into a single low-to-moderate risk category is pragmatic and consistent with prior studies that have used the upper tertile or mean plus one standard deviation as the cut-off for defining high risk [33]. Given the sample distribution characteristics and the need for balanced groups in the subsequent binary logistic regression analyses, the low and moderate groups were merged into a single “low-to-moderate illness uncertainty group,” while the high group was retained as the “high illness uncertainty group.” Accordingly, patients with MUIS scores ≤ 117.4 were assigned to the low-to-moderate group, and those with scores ≥ 117.5 were assigned to the high group.

2.6. Data Analysis

Data analysis was performed using R 4.5.1 software. All tests were two-sided, with a significance level of $\alpha = 0.05$. Normally distributed quantitative data were expressed as mean $\pm$ standard deviation, and comparisons between groups were made using independent sample *t*-tests. Skewed data were expressed as M (P25, P75), and comparisons were made using the Mann–Whitney *U* test. Categorical data were described as frequencies (percentages) [*n* (%)], and comparisons were made using the χ^2 test.

Three machine learning algorithms including LASSO regression, Random Forest, and XGBoost were used to screen for common predictors. All models were evaluated using 10-fold cross-validation. The RF model was trained using 500 decision trees, sorting variable importance using the Mean Decrease Gini index and retaining the top 10 most important variables, as their cumulative importance accounted for the majority of predictive contribution. For LASSO regression, non-informative variables were compressed to zero coefficients through a binomial penalty, using the minimum criteria and the 1-standard-error rule to select the optimal regularization parameter; variables with non-zero coefficients under the λ_{1se} criterion were considered significant predictors. In XGBoost, Shapley

Additive Explanations (SHAP) were calculated to quantify each variable's contribution to the model's prediction, selecting the top 10 variables sorted by gain. Default tuning parameters were used for both RF and XGBoost models, as optimized through the 10-fold cross-validation procedure. Ultimately, the common variables identified by the three methods were visualized using a Venn diagram. Based on these common variables, a logistic regression model was built to construct the nomogram. The model's discrimination was evaluated using the Receiver Operating Characteristic curve and Area Under the Curve, while consistency was assessed using calibration curves, followed by model validation.

3. Results

3.1. Patient Characteristics

Of the 1050 questionnaires distributed, 1006 were returned, and the response rate was 95.8%. After excluding 40 invalid responses, 966 valid questionnaires were retained for the final statistical analysis, yielding an effective response rate of 92.0%. Of these, 676 patients were assigned to the model development set and 290 to the time-stratified validation set. The development set was randomly divided into a training set ($n = 474$) and an internal validation set ($n = 202$) at a 7:3 ratio. Comparison of baseline data between the training and internal validation sets showed that, except for depression scores ($p = 0.0321$), all other variables were evenly distributed with no statistically significant differences ($p > 0.05$), indicating good random grouping and internal consistency within the development set. The clinical characteristics of the study sample are presented in Table 1. In the training set ($n = 474$), there were 328 cases (69.2%) in the high illness uncertainty group and 146 cases (30.8%) in the low-to-moderate group. Univariate analysis showed significant differences between the two groups in depression ($p < 0.001$), anxiety ($p < 0.001$), and social support ($p < 0.001$). The depression and anxiety scores in the high group were significantly higher than those in the low-to-moderate group, while social support scores were significantly lower, suggesting that psychological state and social support levels may be important influencing factors. Other variables such as age, gender, educational level, residence, disease duration, and diagnosis showed no statistically significant differences ($p > 0.05$) (Table 2).

Table 1. Characteristics of the participants in the training and internal validation sets ($n = 676$).

Variable	Training Set (n)	Internal Validation Set ($n = 202$)	χ^2/Z	p Value
Age M[P ₂₅ ,P ₇₅]	50.00 (36.2, 61.0)	47 (38.0, 59.0)	0.477	0.634
Gender n (%)			0.704	0.401
Male	345 (72.8)	154 (76.2)		
Female	129 (27.2)	48 (23.8)		
Education n (%)			1.743	0.783
Primary school or below	44 (9.3)	23 (11.4)		
Junior high school	103 (21.7)	49 (24.3)		
High school or junior college	170 (35.9)	65 (32.2)		
Bachelor's degree	131 (27.6)	53 (26.2)		
Master's degree or above	26 (5.5)	12 (5.9)		
Residence n (%)			2.131	0.775
Rural	126 (26.6)	54 (26.7)		
Town	118 (24.9)	53 (26.2)		
Urban	230 (48.5)	95 (47.0)		
Monthly household income per capita n (%)			2.886	0.577
<2000 RMB	91 (19.2)	35 (17.3)		
2001–5000 RMB	139 (29.3)	51 (25.2)		
5001–10,000 RMB	123 (25.9)	55 (27.2)		
10,001–20,000 RMB	78 (16.5)	36 (17.8)		
>20,000 RMB	43 (9.1)	25 (12.4)		

Table 1. Cont.

Variable	Training Set (n)	Internal Validation Set (n = 202)	χ^2/Z	p Value
Marital status n (%)			5.306	0.169
Unmarried	77 (16.2)	29 (14.4)		
Married	358 (75.5)	162 (80.2)		
Divorced	23 (4.9)	3 (1.5)		
Widowed	16 (3.4)	8 (4.0)		
Occupation n (%)			3.298	0.115
Manual worker	113 (23.9)	43 (21.3)		
Mental worker	65 (13.7)	23 (11.3)		
Retired	118 (24.9)	47 (23.3)		
Others	178 (37.6)	89 (44.1)		
Method of medical payment n (%)			3.201	0.798
Urban employee medical insurance	322 (67.9)	147 (72.8)		
New rural cooperative medical scheme	119 (25.1)	41 (20.3)		
Government-funded medical care	11 (2.3)	4 (2.0)		
Self-pay	11 (2.3)	6 (3.0)		
Commercial insurance	4 (0.8)	1 (0.5)		
Others	7 (1.5)	3 (1.5)		
Religious belief n (%)			0.187	0.665
Yes	455 (96.0)	196 (97.0)		
No	19 (4.0)	6 (3.0)		
Primary caregiver n (%)			7.943	0.159
Spouse	207 (43.7)	99 (49.0)		
Parent	84 (17.7)	34 (16.8)		
Child	111 (23.4)	38 (18.8)		
Sibling	24 (5.1)	6 (3.0)		
Nurse	26 (5.5)	19 (9.4)		
Others	22 (4.6)	6 (3.0)		
Disease duration n (%)			1.884	0.596
<1 month	147 (31.0)	67 (33.2)		
2–5 months	177 (37.3)	78 (38.6)		
6–12 months	64 (13.5)	29 (14.4)		
>12 months	86 (18.1)	28 (13.9)		
Cancer type n (%)			0.987	0.117
Digestive system cancer	43 (9.1)	22 (11.1)		
Respiratory system cancer	2 (0.4)	2 (1.0)		
Gynecological cancer	107 (22.6)	51 (25.2)		
Breast cancer	140 (29.5)	70 (34.7)		
Hematologic malignancy	13 (2.7)	8 (4.0)		
Urologic cancer	2 (0.4)	1 (0.5)		
Bone tumor	146 (30.8)	43 (21.3)		
Others	21 (4.4)	5 (2.5)		
Smoking history n (%)			0.291	0.589
Yes	80 (16.9)	30 (14.9)		
No	394 (83.1)	172 (85.1)		
Alcohol consumption n (%)			0.251	0.616
Yes	89 (18.8)	42 (20.8)		
No	385 (81.2)	160 (79.2)		
Depression M[P ₂₅ ,P ₇₅]	3.00 (0.00, 7.00)	4.00 (1.00, 8.70)	2.147	0.032
Anxiety M[P ₂₅ ,P ₇₅]	2.00 (0.00, 7.00)	2.00 (0.00, 7.00)	0.369	0.712
Medical coping M[P ₂₅ ,P ₇₅]	52.00 (49.00, 54.00)	51.00 (49.00, 54.70)	0.094	0.925
Social support M[P ₂₅ ,P ₇₅]	43.00 (37.00, 48.00)	43.00 (37.00, 49.00)	0.281	0.779

Notes: Data are presented as median (interquartile range, IQR) for continuous variables and number (%) for categorical variables. Group comparisons were performed using the Mann–Whitney *U* test (*Z*) for continuous variables and the chi-square test (χ^2) for categorical variables. Abbreviations: RMB, Renminbi (Chinese Yuan; 1 USD $\approx$ 7.2 RMB at the time of data collection). Occupational categories were defined as follows: “manual worker” included individuals engaged in blue collar occupations, agricultural production, and service industry personnel; “mental worker” included professionals such as teachers, physicians, lawyers, and government/enterprise administrative staff; “retired” referred to individuals who had formally withdrawn from the workforce due to age or other reasons; “others” included students, military personnel, unemployed individuals, and those not classifiable into the above categories.

Table 2. Comparison of characteristics between the high and low-to-moderate illness uncertainty groups in training set.

Variable	High Illness Uncertainty Group (<i>n</i> = 328)	Low-to-Moderate Illness Uncertainty Group (<i>n</i> = 146)	χ^2/Z	<i>p</i> Value
Age M[P ₂₅ ,P ₇₅]	50.00 (39.00, 61.00)	49.00 (34.25, 60.75)	2.572	0.198
Gender <i>n</i> (%)			0.156	0.693
Male	241 (73.5)	104 (71.2)		
Female	87 (26.5)	42 (28.8)		
Education <i>n</i> (%)			6.134	0.189
Primary school or below	32 (9.8)	12 (8.2)		
Junior high school	74 (22.6)	29 (19.9)		
High school or junior college	114 (34.8)	56 (38.4)		
Bachelor's degree	95 (29.0)	36 (24.7)		
Master's degree or above	13 (4.0)	13 (8.9)		
Residence <i>n</i> (%)			4.801	0.063
Rural	93 (28.4)	33 (22.6)		
Town	71 (21.6)	47 (32.2)		
Urban	164 (50.0)	66 (45.2)		
Monthly household income per capita <i>n</i> (%)			6.209	0.184
≤2000 RMB	68 (20.7)	23 (15.8)		
2001–5000 RMB	101 (30.8)	38 (26.0)		
5001–10,000 RMB	79 (24.1)	44 (30.1)		
10,001–20,000 RMB	48 (14.6)	30 (20.5)		
>20,000 RMB	32 (9.8)	11 (7.5)		
Marital status <i>n</i> (%)			5.013	0.953
Unmarried	53 (16.2)	24 (16.4)		
Married	248 (75.6)	110 (75.3)		
Divorced	15 (4.6)	8 (5.5)		
Widowed	12 (3.7)	4 (2.7)		
Occupation <i>n</i> (%)			6.661	0.753
Manual worker	76 (23.1)	37 (25.4)		
Mental worker	47 (14.3)	18 (12.3)		
Retired	83 (25.3)	35 (24.0)		
Others	122 (37.2)	56 (38.4)		
Method of medical payment <i>n</i> (%)			7.801	0.056
Urban employee medical insurance	221 (67.4)	101 (69.2)		
New rural cooperative medical scheme	86 (26.2)	33 (22.6)		
Government-funded medical care	7 (2.1)	4 (2.7)		
Self-pay	10 (3.0)	1 (0.7)		
Commercial insurance	3 (0.9)	1 (0.7)		
Others	1 (0.3)	6 (4.1)		
Religious belief <i>n</i> (%)			0.032	0.858
Yes	314 (95.7)	141 (96.6)		
No	14 (4.3)	5 (3.4)		
Primary caregiver <i>n</i> (%)			6.840	0.232
Spouse	145 (44.2)	62 (42.5)		
Parent	58 (17.7)	26 (17.8)		
Child	81 (24.7)	30 (20.5)		
Sibling	17 (5.2)	7 (4.8)		
Nurse	17 (5.2)	9 (6.2)		
Others	10 (3.0)	12 (8.2)		
Disease duration <i>n</i> (%)			0.266	0.966
<1 month	101 (30.8)	46 (31.5)		
2–5 months	123 (37.5)	54 (37.0)		
6–12 months	43 (13.1)	21 (14.4)		
>12 months	61 (18.6)	25 (17.1)		
Cancer type <i>n</i> (%)			7.891	0.084
Digestive system cancer	30 (9.1)	13 (8.9)		
Respiratory system cancer	1 (0.3)	1 (0.7)		
Gynecological cancer	66 (20.1)	41 (28.1)		

Table 2. Cont.

Variable	High Illness Uncertainty Group (n = 328)	Low-to-Moderate Illness Uncertainty Group (n = 146)	χ^2/Z	p Value
Breast cancer	107 (32.6)	33 (22.6)		
Hematologic malignancy	11 (3.4)	2 (1.4)		
Urologic cancer	1 (0.3)	1 (0.7)		
Bone tumor	101 (30.8)	45 (30.8)		
Others	11 (3.4)	10 (6.8)		
Smoking history n (%)			0.039	0.981
Yes	53 (16.8)	25 (17.1)		
No	273 (83.2)	121 (82.9)		
Alcohol consumption n (%)			0.283	0.595
Yes	59 (18.0)	30 (20.5)		
No	269 (82.0)	116 (79.5)		
Depression M[P ₂₅ ,P ₇₅]	4.50 (1.00, 8.25)	1.00 (0.00, 4.00)	3.340	<0.001
Anxiety M[P ₂₅ ,P ₇₅]	4.00 (0.00, 7.00)	2.00 (1.00, 4.00)	3.386	<0.001
Medical coping M[P ₂₅ ,P ₇₅]	52.00 (49.00, 54.00)	52.00 (49.00, 55.00)	2.289	0.445
Social support M[P ₂₅ ,P ₇₅]	41.00 (37.00, 47.00)	46.00 (41.00, 50.00)	6.882	<0.001

Notes: Data are presented as median (interquartile range, IQR) for continuous variables and number (%) for categorical variables. Group comparisons were performed using the Mann–Whitney *U* test (*Z*) for continuous variables and the chi-square test (χ^2) for categorical variables.

3.2. Machine Learning Results

3.2.1. LASSO Regression Preliminary Screening

LASSO regression selected anxiety as the only non-zero predictor under the λ_{1se} criterion.

3.2.2. Random Forest Importance Ranking

The Out-of-Bag error rate plot of the RF model showed a rapid decline as the number of decision trees increased, stabilizing after reaching about 200 trees. The error rate curves for both high and low-to-moderate groups also remained stable with no upward trend, indicating good model stability. By calculating the Mean Decrease Gini index, the top 10 important variables were: social support, anxiety, depression, age, medical coping modes, educational level, monthly household income, time since diagnosis, disease duration, and disease diagnosis (Figure 1A).

3.2.3. XGBoost Importance Ranking

Variable importance ranking (Figure 1B) showed the top 10 variables were: anxiety, social support, depression, medical coping modes, age, disease diagnosis, caregiver, educational level, time since diagnosis, and medical payment method.

3.2.4. Common Factor Screening

The three algorithms showed different degrees of overlap in variable selection. LASSO regression selected anxiety as the only non-zero predictor. Random Forest and XGBoost identified several overlapping predictors. The Venn diagram showed that anxiety was the only variable shared by all three algorithms. Using the predefined consensus rule of selection by at least two algorithms, seven predictors were retained for the final nomogram: age, education level, diagnosis, depression, anxiety, medical coping, and social support. Collinearity diagnostics showed that the Variance Inflation Factor (VIF) of each variable was less than 4, indicating no severe multicollinearity (Figure 1C). In predictive modeling, variables that are not statistically significant in univariate analysis can still contribute to multivariable prediction by capturing joint effects or interactions. Therefore, we retained all seven variables identified by the ensemble algorithms.

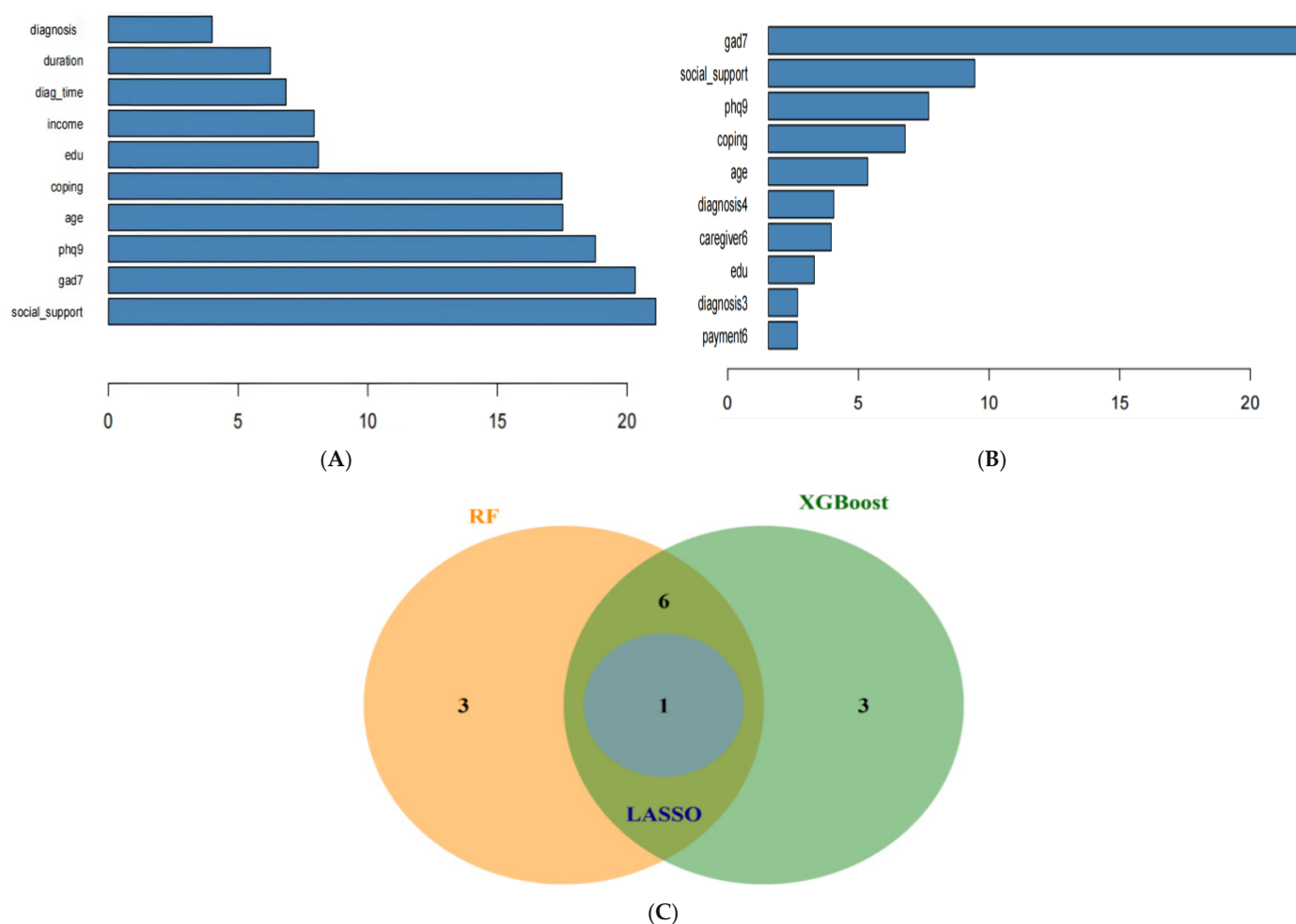


Figure 1. Core predictors identified by integrated machine learning algorithms. (A) Random Forest variable importance ranking. (B) XGBoost variable importance ranking. (C) Venn diagram of the three algorithms.

3.3. Nomogram Construction and Validation

The aforementioned seven variables were included in a logistic regression to build a nomogram model. Each factor corresponds to a specific score, and the total score is calculated by summing these seven scores and locating the sum on the total points axis. Because variables were standardized, the scales for continuous variables in the nomogram reflect standardized scores (Figure 2A). In practical application, patients' raw values must first be converted into standardized values based on the training set's mean and standard deviation before reading the individual scores and predicted probabilities on the nomogram (means and standard deviations in Table 3). Discrimination evaluation: the AUC for the training set was 0.763 (95% CI: 0.717–0.808), 0.724 for the internal validation set (95% CI: 0.645–0.803), and 0.663 for the time-stratified validation set (95% CI: 0.587–0.739) (Figure 2B). Calibration evaluation: both the training set ($\chi^2 = 9.438$, $p = 0.307$, Brier = 0.174) and the internal validation set ($\chi^2 = 14.169$, $p = 0.078$, Brier = 0.168) demonstrated good calibration, with highly consistent predicted and actual observed risks. In the time-stratified validation set, the HL test indicated calibration bias ($\chi^2 = 29.831$, $p < 0.001$, Brier = 0.177), and the calibration curve showed that the model underestimated the risk for low-to-moderate group patients in this population (Figure 2C). Full classification metrics and decision curve analysis are available in the Supplementary Materials.

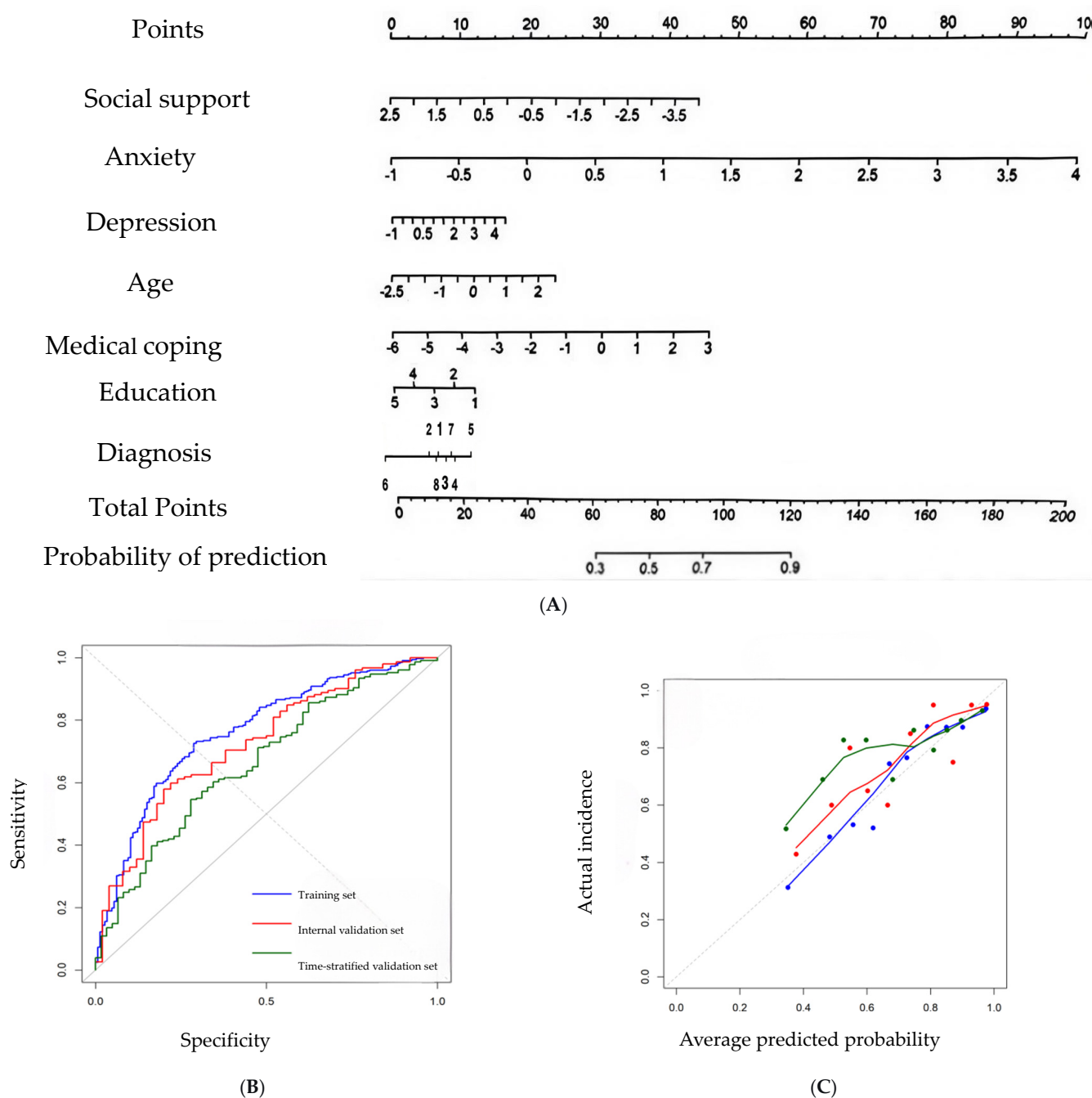


Figure 2. Construction and validation of the risk prediction nomogram. **Notes:** In the nomogram, education level is coded as 1 = primary school or below, 2 = junior high school, 3 = high school or junior college, 4 = bachelor's degree, and 5 = master's degree or above. Diagnosis is represented by the dummy variable diagnosis 3, where 1 indicates gynecological cancer and 0 indicates other cancer types. (A) Nomogram. (B) ROC curves. (C) Calibration curves.

Table 3. Coding and distribution of predictors included in the nomogram in the training set.

Predictor	Variable Type/Coding	Training Set Value
Age	Continuous, years	52.3 ± 15.2
Anxiety	Continuous, GAD-7	6.5 ± 5.0
Depression	Continuous, PHQ-9	5.9 ± 5.2
Medical coping	Continuous score	51.8 ± 4.9
Social support	Continuous score	43.2 ± 8.1

Table 3. Cont.

Predictor	Variable Type/Coding	Training Set Value
Education level	Ordinal categorical	
Primary school or below	Code = 1	44 (9.3%)
Junior high school	Code = 2	103 (21.7%)
High school or junior college	Code = 3	170 (35.9%)
Bachelor's degree code	Code = 4	131 (27.6%)
Master's degree or above	Code = 5	26 (5.5%)
Diagnosis 3	Binary dummy variable	
Gynecological cancer	Code = 1	107 (22.6%)
Other cancer types	Code = 2	367 (77.4%)

Notes: Continuous variables are presented as mean $\pm$ standard deviation. Categorical variables are presented as number and percentage. Education level was entered as an ordinal categorical variable coded from 1 to 5. Diagnosis was entered as a binary dummy variable, diagnosis 3, where 1 indicates gynecological cancer and 0 indicates all other cancer types. Continuous predictors were standardized when constructing the nomogram using the mean and standard deviation from the training set.

4. Discussion

4.1. Illness Uncertainty Predictors

4.1.1. The Core Role of Anxiety and Depression

In this study, anxiety and depression were strongly associated with illness uncertainty. Univariate analysis showed significantly higher scores in the high-risk group ($p < 0.001$). Notably, anxiety was the only variable jointly identified by all three algorithms (LASSO, RF, XGBoost), suggesting it may be a key psychological correlate rather than a causal driver [34]. This aligns closely with Mishel's uncertainty in illness theory [35]: when patients are anxious, their information processing abilities may decline, potentially making them more prone to negative interpretations and catastrophic thinking regarding disease information, which could exacerbate uncertainty [36]. Depression is linked to negative cognitive biases, which may direct attention to negative cues and potentially amplify perceived uncertainty [37]. Studies suggest mood disorders are associated with altered appraisal processes, possibly contributing to confusion and worry [38]. Both anxiety and depression are modifiable factors; early referral to psycho-oncology or interventions may improve mood and might indirectly reduce uncertainty.

4.1.2. The Protective Role of Social Support and Coping Styles

Social support showed significant group differences (high group 41.0 vs. low-to-moderate group 46.0, $p < 0.001$) and was identified by both RF and XGBoost as a key predictor. As an external resource, its protective effects operate on multiple levels: emotional support alleviates fear, informational support aids understanding, and tangible support reduces life burdens [7]. Previous studies confirm that social support is negatively correlated with uncertainty and acts as a crucial protective factor for psychological adaptation [24]. Furthermore, stress coping theory indicates individuals adopt different strategies when facing stress [39]. In this model, medical coping styles were included, suggesting patients leaning towards "confrontation" strategies exhibit lower uncertainty, while those using "avoidance" or "resignation" show higher levels. This matches Lazarus' stress coping theory [40]. Research highlights an interaction between social support and coping: high support provides resources that foster positive coping, while positive coping helps patients better utilize available support [41]. This virtuous cycle appears to be an important pathway associated with reduced uncertainty.

4.1.3. The Impact of Demographic and Clinical Characteristics

Age, educational level, and disease diagnosis were identified as predictors. Older patients may experience declines in information processing and cognitive flexibility, struggling with complex medical information, which aligns with Wang, T. et al.'s findings [42]. Patients with lower educational levels have limited ability to acquire and comprehend medical information. This implies that communication with older or less-educated patients should use plain language, with increased repetition and confirmation [43]. Disease diagnosis reflects the differing psychological impacts of various tumors, which possess unique trajectories, treatment complexities, and prognoses [44]. For instance, patients with better prognoses might focus on recurrence risks and long-term quality of life, whereas those with poorer prognoses focus on survival time and symptom control. This highlights the need for specialized, individualized psychological assessments and support tailored to specific diagnostic types.

4.2. Model Predictive Performance and Validation

The constructed nomogram demonstrated good discrimination in the training set (AUC = 0.763) and internal validation set (AUC = 0.724). Calibration tests showed high consistency between predicted and actual risks ($p > 0.05$), indicating stable predictive performance within the development and concurrent random populations. Compared to previous studies, this model's predictive efficacy is modest. In the stricter time-stratified validation, discrimination decreased (AUC = 0.663), and significant calibration bias emerged ($p < 0.001$), primarily underestimating risk in low-to-moderate risk individuals. This phenomenon is similar to the "temporal drift" observed by Bevilacqua et al. [45] in building a BCRL predictive model. This may be due to subtle shifts over time in patient characteristics, treatment modalities, or nursing focuses at our center. For example, changes in patients' disease awareness, access to medical information, and psychosocial resources over time can alter uncertainty mechanisms [9,46]. Brier score (0.177) remains acceptable, but calibration bias urges caution. Notably, the time-stratified cohort came from the same institution but a later window, showing that even single-center temporal heterogeneity can degrade performance. This implies that cross-sectionally developed models may have limited time transportability; routine updating or recalibration may be needed for longitudinal use. Independent multi-center validation with longer follow-up is essential to assess whether drift is random or systematic, and this temporal decay should be given greater weight when considering the model's readiness for practice.

4.3. Strengths and Limitations

This is one of the few studies integrating three machine learning algorithms, LASSO regression, Random Forest and XGBoost, to jointly screen the predictors of illness uncertainty in patients with malignant tumors, so as to enhance the robustness and clinical interpretability of the selected variables. In addition, time-stratified validation was used to evaluate the stability of the model in patients at different time points of diagnosis, which was more rigorous than traditional internal validation. The final nomogram included seven predictors, including potentially modifiable psychological and social factors, such as anxiety, depression, social support, and coping style, which provided a clear target for clinical psychological intervention. At the same time, its visual form is helpful for early risk stratification and doctor–patient communication.

However, several limitations should be acknowledged. The single-center cross-sectional design may have introduced selection bias and limits generalizability, and the temporal ordering of predictors and outcome cannot be established. Thus, the model effectively classifies patients according to their current profiles rather than predicting future

psychological states. The dichotomization of MUIS scores into high versus low-to-moderate risk may have lost information, and the chosen cut-off of 117.5 lacks established clinical validation. The time-stratified validation, while more stringent than internal validation, still originates from the same institution and does not constitute independent external validation from another center, limiting transportability. All psychosocial predictors were derived from self-reported questionnaires, subject to recall and social desirability bias. The model's overall discriminative performance with an internal AUC of 0.724 is only moderate, and the calibration deviation in the time-stratified set ($p < 0.05$) indicates temporal heterogeneity and poor transportability over time, necessitating caution in clinical application. Therefore, the model should be used only as a preliminary screening aid, not as a definitive diagnostic tool. Future multi-center prospective studies are needed to further validate and optimize this model, and subsequent research could consider constructing dynamic prediction models that capture longitudinal changes in illness uncertainty.

4.4. Implications for Practice

Despite its limitations, this model offers potential clinical value as a preliminary screening tool. It shifts the identification window for illness uncertainty from “post hoc assessment” to “initial admission screening,” providing a quantitative basis for stratified psychological nursing, although the approach still requires collection of several standardized questionnaires. By collecting seven routine variables upon admission, medical staff can rapidly conduct risk assessments, transitioning from “passive response” to “proactive screening. Secondly, among the seven predictors, anxiety, depression, social support, and coping modes are modifiable, offering clear intervention targets: (1) for patients with high anxiety/depression scores, early referrals to psycho-oncology or pharmacological interventions may be beneficial, and clinical decisions should consider individual patient circumstances; (2) for those with low social support, family meetings and peer support groups can strengthen social resource links; (3) for patients with negative coping styles, health education can help build positive disease-coping strategies. Finally, the nomogram's visual nature makes it an effective auxiliary tool for doctor–patient communication. Staff can intuitively display risk probabilities and explain related factors to patients, enhancing their understanding and acceptance of their psychological state and improving compliance with subsequent psychological interventions.

5. Conclusions

Based on three machine learning algorithms, this study constructed and internally validated a nomogram for concurrent risk stratification of illness uncertainty in patients with malignant tumors. Seven core predictors were identified: age, education level, disease diagnosis, depression, anxiety, medical coping style, and social support. Anxiety, being the only variable selected by all three algorithms, suggests it may be a prominent psychological correlate of illness uncertainty rather than a causal mechanism. The model showed good discrimination in the training and internal validation sets, but calibration bias was observed in the time-stratified validation set, indicating temporal heterogeneity. Given the single-center cross-sectional design, dichotomization of MUIS, lack of multi-center temporal validation, and self-report biases, this nomogram should be viewed as an exploratory tool that requires further external validation and optimization before any clinical implementation can be considered. Modifiable predictors provide clear targets for psychosocial interventions. Future multi-center prospective studies are needed to further refine and externally validate the model.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/healthcare14142160/s1>, Table S1. Discrimination, calibration, and classification performance of the nomogram. Supplementary File S1: Informed Consent.

Author Contributions: Conceptualization, Y.L., J.G. and X.L.; methodology, Y.L.; software, Y.L., L.M. and H.Z.; validation, Y.L., L.M. and H.Z.; formal analysis, Y.L.; investigation, Y.L., H.Z. and Y.W.; resources, L.M.; data curation, Y.L., H.Z., J.G. and Y.W.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L. and X.L.; visualization, Y.L. and H.Z.; supervision, L.M., J.G. and X.L.; project administration, J.G. and X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Project No. U25A2029).

Institutional Review Board Statement: This study was approved on 20 June 2024 by the Institutional Review Board (IRB) at Ethics Review Committee of Peking University People's Hospital (approval no. 2024PHB227-001).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Due to patient privacy restrictions, the data are not publicly available. De-identified data can be requested from the corresponding author.

Acknowledgments: The authors thank all the patients with malignant tumors who participated in this study for their time and cooperation. We also express our gratitude to the medical and nursing staff of Peking University People's Hospital for their assistance with data collection.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AUC	Area Under the Curve
BCRL	Breast Cancer-Related Lymphedema
CBT	Cognitive-Behavioral Therapy
CI	Confidence Interval
HL	Hosmer–Lemeshow
LASSO	Least Absolute Shrinkage and Selection Operator
MUIS	Mishel Uncertainty in Illness Scale
RF	Random Forest
ROC	Receiver Operating Characteristic
SHAP	Shapley Additive Explanations
VIF	Variance Inflation Factor
XGBoost	eXtreme Gradient Boosting

References

1. Almeida Filho, R.F.; Trezza, M.; Comassetto, I.; Silva, L.; Lopes, M.P.; Santana, K.G.S.; Praxedes, F.A.S. Spirituality in the uncertainty of illness: The perspective of oncology patients. *Rev. Bras. Enferm.* **2023**, *76*, e20220712. [[CrossRef](#)] [[PubMed](#)]
2. Guan, T.; Chapman, M.V.; de Saxe Zerden, L.; Sharma, A.; Chen, D.G.; Song, L. Correlates of illness uncertainty in cancer survivors and family caregivers: A systematic review and meta-analysis. *Support. Care Cancer* **2023**, *31*, 242. [[CrossRef](#)] [[PubMed](#)]
3. Wagle, N.S.; Nogueira, L.; Devasia, T.P.; Mariotto, A.B.; Yabroff, K.R.; Islami, F.; Jemal, A.; Alteri, R.; Ganz, P.A.; Siegel, R.L. Cancer treatment and survivorship statistics, 2025. *CA Cancer J. Clin.* **2025**, *75*, 308–340. [[CrossRef](#)] [[PubMed](#)]
4. Hong, S.; Lee, J.; Lee, J.; Chang, J.H.; Park, C.G.; Kim, T.H. The Mediating Effect of Uncertainty in Illness on Cancer Coping in Patients with Primary Malignant Brain Tumors. *Cancer Nurs.* **2023**, *46*, 467–476. [[CrossRef](#)] [[PubMed](#)]

5. Guan, T.; Chapman, M.V.; de Saxe Zerden, L.; Zimmer, C.; Braden, C.J.; Rizo, C.F.; Northouse, L.; Song, L. Illness uncertainty and quality of life in patients with advanced cancer and their family caregivers: An actor-partner interdependence model analysis. *Psychooncology* **2023**, *32*, 1744–1751. [[CrossRef](#)] [[PubMed](#)]
6. Wolynec, K.; Sharp, J.; Fisher, K.; Tothill, R.W.; Bowtell, D.; Mileschkin, L.; Schofield, P. Psychological distress, understanding of cancer and illness uncertainty in patients with Cancer of Unknown Primary. *Psychooncology* **2022**, *31*, 1869–1876. [[CrossRef](#)] [[PubMed](#)]
7. Yang, E. Study on Reliability and Validity of Chinese Version of Mishel Illness Uncertainty Scale. Master's Thesis, National Taiwan University, Taiwan, China, 2008.
8. Lemaire, G.S.; Lenz, E.R. Perceived uncertainty about menopause in women attending an educational program. *Int. J. Nurs. Stud.* **1995**, *32*, 39–48. [[CrossRef](#)] [[PubMed](#)]
9. Li, J.; Sun, D.; Zhang, X.; Zhao, L.; Zhang, Y.; Wang, H.; Ni, N.; Jiang, G. The relationship between anticipatory grief and illness uncertainty among Chinese family caregivers of patients with advanced lung cancer: A cross-sectional study. *BMC Palliat. Care* **2022**, *21*, 30. [[CrossRef](#)] [[PubMed](#)]
10. Pezzolato, M.; Spada, G.E.; Fragale, E.; Cutica, I.; Masiero, M.; Marzorati, C.; Pravettoni, G. Predictive Models of Psychological Distress, Quality of Life, and Adherence to Medication in Breast Cancer Patients: A Scoping Review. *Patient Prefer. Adherence* **2023**, *17*, 3461–3473. [[CrossRef](#)] [[PubMed](#)]
11. Zhou, W.; Wang, Y.; Long, H.; Wang, J. Construction of a Nomogram Prediction Model for Individualized Prediction of Cervical Cancer Risk: A Retrospective Study. *Int. J. Women's Health* **2025**, *17*, 4435–4446. [[CrossRef](#)] [[PubMed](#)]
12. Shen, A.; Wei, X.; Zhu, F.; Sun, M.; Ke, S.; Qiang, W.; Lu, Q. Risk prediction models for breast cancer-related lymphedema: A systematic review and meta-analysis. *Eur. J. Oncol. Nurs.* **2023**, *64*, 102326. [[CrossRef](#)] [[PubMed](#)]
13. Xiao, Y.; Zhao, Z.; Su, C.G.; Li, J.; Liu, J. An interpretable machine learning model for predicting depression in middle-aged and elderly cancer patients in China: A study based on the CHARLS cohort. *BMC Psychiatry* **2025**, *25*, 610. [[CrossRef](#)] [[PubMed](#)]
14. Li, M.M.; Wu, P.P.; Qiang, W.M.; Li, J.Q.; Zhu, M.Y.; Yang, X.L.; Wang, Y. Development and validation of a risk prediction model for breast cancer-related lymphedema in postoperative patients with breast cancer. *Eur. J. Oncol. Nurs.* **2023**, *63*, 102258. [[CrossRef](#)] [[PubMed](#)]
15. O'Connor, S.; Vercell, A.; Wong, D.; Yorke, J.; Fallatah, F.A.; Cave, L.; Anny Chen, L.Y. The application and use of artificial intelligence in cancer nursing: A systematic review. *Eur. J. Oncol. Nurs.* **2024**, *68*, 102510. [[CrossRef](#)] [[PubMed](#)]
16. Ren, H.; Yang, T.; Yin, X.; Tong, L.; Shi, J.; Yang, J.; Zhu, Z.; Li, H. Prediction of high-level fear of cancer recurrence in breast cancer survivors: An integrative approach utilizing random forest algorithm and visual nomogram. *Eur. J. Oncol. Nurs.* **2024**, *70*, 102579. [[CrossRef](#)] [[PubMed](#)]
17. Eche-Ugwu, I.J.; Mazzola, E.; Stevens, S.E.; Wolfe, J.; Feraco, A.M. Illness Uncertainty in Parents of Children with Cancer-An Exploratory Analysis to Inform Intervention Development. *Pediatr. Blood Cancer* **2025**, *72*, e31726. [[CrossRef](#)] [[PubMed](#)]
18. Rodriguez-Gonzalez, A.; Velasco-Durantez, V.; Martin-Abreu, C.; Cruz-Castellanos, P.; Hernandez, R.; Gil-Raga, M.; Garcia-Torralba, E.; Garcia-Garcia, T.; Jimenez-Fonseca, P.; Calderon, C. Fatigue, Emotional Distress, and Illness Uncertainty in Patients with Metastatic Cancer: Results from the Prospective NEOETIC_SEOM Study. *Curr. Oncol.* **2022**, *29*, 9722–9732. [[CrossRef](#)] [[PubMed](#)]
19. Redondo-Sáenz, D.; Solano-López, A.L.; Vílchez-Barboza, V. Body image, illness uncertainty and symptom clusters in surgically treated breast cancer survivors: An exploratory factor analysis and correlational study. *Eur. J. Oncol. Nurs.* **2024**, *72*, 102662. [[CrossRef](#)] [[PubMed](#)]
20. Langmuir, T.; Chu, A.K.; Sehabi, G.; Giguère, L.; Lamarche, J.; Boudjatat, W.A.; Lebel, S. A new landscape in illness uncertainty: A systematic review and thematic synthesis of the experience of uncertainty in patients with advanced cancer receiving immunotherapy or targeted therapy. *Psycho-Oncology* **2023**, *32*, 356–367. [[CrossRef](#)]
21. Wang, F.; Shang, X.; Li, W.; Wang, M.; Li, F. Time-stratified modeling of cognitive impairment risk in rural aging populations: Nomogram development (2011) and external validation (2013) using CHARLS. *Arch. Gerontol. Geriatr.* **2025**, *140*, 106033. [[CrossRef](#)] [[PubMed](#)]
22. Malik, S.; Naqvi, A.; Tenorio, B.G.; Farrukh, F.; Tariq, R.; Adler, D.G. Machine Learning for Predicting Colorectal Cancer and Advanced Colorectal Polyps: A Systematic Review and Meta-Analysis. *J. Clin. Gastroenterol.* **2026**, *60*, 237–246. [[CrossRef](#)] [[PubMed](#)]
23. Peduzzi, P.; Concato, J.; Kemper, E.; Holford, T.R.; Feinstein, A.R. A simulation study of the number of events per variable in logistic regression analysis. *J. Clin. Epidemiol.* **1996**, *49*, 1373–1379. [[CrossRef](#)] [[PubMed](#)]
24. Wu, S.; Guo, X.; Tang, H.; Li, Y.; Dong, W.; Lu, G.; Hou, C.; Chen, C. The Relationship Between Illness Uncertainty and Social Support Among Cancer Patients: A Meta-analysis. *Cancer Nurs.* **2025**, *48*, 416–423. [[CrossRef](#)] [[PubMed](#)]
25. Kroenke, K.; Spitzer, R.L.; Williams, J.B. The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med.* **2001**, *16*, 606–613. [[CrossRef](#)] [[PubMed](#)]

26. Bian, C.D.; He, X.Y.; Qian, J.; Wu, W.Y.; Li, C.B. Application of the Patient Health Questionnaire-9 in general hospitals. *J. Tongji Univ. (Med. Sci.)* **2009**, *30*, 136–140.
27. Spitzer, R.L.; Kroenke, K.; Williams, J.B.W.; Löwe, B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch. Intern. Med.* **2006**, *166*, 1092–1097. [[CrossRef](#)] [[PubMed](#)]
28. He, X.Y.; Li, C.B.; Qian, J.; Cui, H.S.; Wu, W.Y. Reliability and validity of the Generalized Anxiety Disorder Scale in general hospitals. *Shanghai Arch. Psychiatry* **2010**, *22*, 200–203.
29. Feifel, H.; Strack, S.; Nagy, V.T. Degree of life-threat and differential use of coping modes. *J. Psychosom. Res.* **1987**, *31*, 91–99. [[CrossRef](#)] [[PubMed](#)]
30. Shen, X.H.; Jiang, Q.J. Report on 701 cases of the Chinese version of the Medical Coping Modes Questionnaire. *Chin. J. Behav. Med. Sci.* **2000**, *8*, 22–24.
31. Xiao, S.Y. Theoretical basis and research application of the Social Support Rating Scale. *Chin. J. Clin. Psychiatry* **1994**, *4*, 3.
32. Mishel, M.H. The measurement of uncertainty in illness. *Nurs. Res.* **1981**, *30*, 258–263. [[CrossRef](#)] [[PubMed](#)]
33. Ye, Z.J.; She, Y.; Liang, M.Z.; Tish, K.; Jane, D.; Hu, Q.; Zeng, Z.; Hu, G.Y.; Zhu, Y.F.; Qiu, H.Z. Re-revision of the Chinese version of the Mishel Uncertainty in Illness Scale and its reliability and validity test in Chinese cancer patients. *Chin. Gen. Pract.* **2018**, *21*, 1091–1097. [[CrossRef](#)]
34. Cruz-Castellanos, P.; Gil-Raga, M.; Jiménez-Fonseca, P.; Ghanem, I.; Hernández, R.; Piera-Molons, N.; Cano, J.M.; Gallego-Martinez, A.; Garcia-Torralba, E.; Calderon, C. Uncertainty and hope in relation to anxiety and depression in advanced lung cancer. *BMJ Support. Palliat. Care* **2024**, *13*, e847–e850. [[CrossRef](#)] [[PubMed](#)]
35. Guan, T.; Zhang, Y.; Northouse, L.L.; Song, L. Psychometric properties of the Brief Mishel Uncertainty in illness scales for patients with advanced cancer and their family caregivers. *BMC Psychol.* **2025**, *13*, 317. [[CrossRef](#)] [[PubMed](#)]
36. Guan, T.; Qan'ir, Y.; Song, L. Systematic review of illness uncertainty management interventions for cancer patients and their family caregivers. *Support. Care Cancer* **2021**, *29*, 4623–4640. [[CrossRef](#)] [[PubMed](#)]
37. Stone, L.; Olsen, A. Illness uncertainty and risk management for people with cancer. *Aust. J. Gen. Pract.* **2022**, *51*, 321–326. [[CrossRef](#)] [[PubMed](#)]
38. Longcoy, L.H.; Liang, S.Y.; Doorenbos, A.Z. The Role of Resilience in Coping with Future Uncertainty Among People with Brain Tumors: Cross-Sectional Study. *JMIR Cancer* **2026**, *12*, e71674. [[CrossRef](#)] [[PubMed](#)]
39. Obispo, B.; Cruz-Castellanos, P.; Fernández-Montes, A.; Coca-Membribes, S.; Rogado, J.; Antoñanzas, M.; Castillo-Trujillo, O.A.; López-Ceballos, H.; Rodríguez-González, A.; Jiménez-Fonseca, P.; et al. Coping strategies as mediators of uncertainty and psychological distress in patients with advanced cancer. *Psychooncology* **2023**, *32*, 1694–1701. [[CrossRef](#)] [[PubMed](#)]
40. Yang, Y.; Qian, X.; Tang, X.; Shen, C.; Zhou, Y.; Pan, X.; Li, Y. The links between symptom burden, illness perception, psychological resilience, social support, coping modes, and cancer-related worry in Chinese early-stage lung cancer patients after surgery: A cross-sectional study. *BMC Psychol.* **2024**, *12*, 463. [[CrossRef](#)]
41. Zhang, Y.; Mohile, S.; Culakova, E.; Norton, S.; Loh, K.P.; Kadambi, S.; Kehoe, L.; Aarne Grossman, V.; Vogel, V.G.; Burnette, B.L.; et al. The shared uncertainty experience of older adults with advanced cancer and their caregivers. *Psychooncology* **2022**, *31*, 1041–1049. [[CrossRef](#)]
42. Wang, T.; Sun, J.; Gu, D.; Shen, S.; Zhou, Y.; Wang, Z. Dyadic effects of social support, illness uncertainty on anxiety and depression among lung cancer patients and their caregivers: A cross-sectional study. *Support. Care Cancer* **2023**, *31*, 402. [[CrossRef](#)] [[PubMed](#)]
43. Tang, W.W.; Han, M.L.; Xu, S.H.; Deng, Y.X.; Shen, Q. Analysis of sleep quality, disease uncertainty, and psychological tolerance in patients undergoing chemotherapy for digestive tract malignancies. *World J. Clin. Cases* **2024**, *12*, 4247–4255. [[CrossRef](#)] [[PubMed](#)]
44. Liu, X.; Qi, X.; Fang, D.; Feng, S.; Chen, J.; Li, X.; Dang, J.; Wu, L. The impact of cognitive behavioral therapy on disease uncertainty, stressful life events, quality of life, anxiety, and depression in glioma patients undergoing chemotherapy: A quasi-experimental study. *BMC Psychiatry* **2025**, *25*, 272. [[CrossRef](#)] [[PubMed](#)]
45. Bevilacqua, J.L.B.; Kattan, M.W.; Changhong, Y.; Koifman, S.; Mattos, I.E.; Koifman, R.J.; Bergmann, A. Nomograms for Predicting the Risk of Arm Lymphedema after Axillary Dissection in Breast Cancer. *Ann. Surg. Oncol.* **2012**, *19*, 2580–2589. [[CrossRef](#)]
46. Poshtan, M.M.; Aflakseir, A.; Witthöft, M.; Cherry, M.G.; Rahimi Taghanaki, C.; Ramzi, M. The role of intolerance of uncertainty in health anxiety in cancer patients: Exploring demographic and cancer-related variations. *Eur. J. Oncol. Nurs.* **2025**, *74*, 102793. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.