의학박사 학위논문

Self-management Strategies' Clustering and Prediction of Health-Related Quality of Life in Cancer Survivors and Survival in Advanced Cancer Patients Using Machine Learning Techniques

자가건강전략의 클러스터링과
머신러닝 기법을 사용한 암생존자의 삶의 질 및
진행성 암환자의 생존 예측

2023년 2월

서울대학교 대학원
의과학과 (시스템정보맞춤의학) 전공
정 주 연

Ph.D. Dissertation of Philosophy in Medical Science

Self-management Strategies' Clustering and Prediction of Health-Related Quality of Life in Cancer Survivors and Survival in Advanced Cancer Patients Using Machine Learning Techniques

자가건강전략의 클러스터링과
머신러닝 기법을 사용한 암생존자의 삶의 질 및
진행성 암환자의 생존 예측

February 2023

Graduate School of Biomedical Sciences
Seoul National University

Ju Youn Jung

Self-management Strategies' Clustering and Prediction of Health-Related Quality of Life in Cancer Survivors and Survival in Advanced Cancer Patients Using Machine Learning Techniques

Submitting a Ph.D. Dissertation of
Medicine

October 2022

Graduate School of Biomedical Sciences
Seoul National University

Ju Youn Jung

Confirming the Ph.D. Dissertation written by
Ju Youn Jung
January 2023

Chair	김 주 한	(Seal)
Vice Chair	윤 영 호	(Seal)
Examiner	박 현 애	(Seal)
Examiner	박 상 민	(Seal)
Examiner	오 상 우	(Seal)

Abstract

Background: In cancer-care, self-management strategies can help cancer patients improve their health-related quality of life (HRQoL) or survival, irrespective of the cancer stage or their treatment plan. However, there is insufficient research on the clustering of self-management strategies considering cancer stages in natural clinical settings; the prediction model of HRQoL or survival in cancer patients also lacks research. In addition, research that has comprehensively identified the relationship between self-management strategies, HRQoL, and survival still needs to be completed. Hence, we investigated their relationship using clustering methods, machine learning techniques (MLT), and path analysis of structural equation modeling (SEM).

Methods: In cancer survivors, cluster analyses using principal component analyses in varimax rotation and clustering of the k-means method were conducted to examine the interrelationship among self-management strategies in smart management strategies for health assessment tool (SAT). Multivariate-adjusted analyses were performed to identify the association of self-management strategies with HRQoL after 6 months. We constructed the HRQoL prediction model and compared the performance of the model with ensemble algorithms including decision tree, random forest, gradient boosting, eXtreme Gradient Boost (XGBoost), and LightGBM. Next, we selected the XGBoost model for further analysis. We demonstrated critical features of HRQoL and extracted the individual prediction result in the XGBoost model using SHAP. In advanced cancer patients, self-management clustering and multivariate-adjusted analyses for examining the association of the strategies with the HRQoL were conducted the same way as in cancer survivors. We performed dimensional multiple Cox proportional hazard regression analyses to determine critical predictors for 1-year survival. We established a survival prediction model with the XGBoost method using MLT with the critical predictors in the Cox regression model. To examine the causal relationship among SAT strategies, HRQoL, and survival, we used a subgroup analysis and a path analysis of structural equation modeling.

Results: All cancer survivors and advanced cancer patients experienced two clusters in the self-management strategies concurrently. However, the strategy clusters differed by cancer stage. Advanced-stage cancer patients used core strategies along with preparation and implementation strategies to overcome their crisis. Among all cancer patients, the self-management strategies had a positive association with improved HRQoL, even in advanced cancer patients. In the prediction model development, the XGBoost model for HRQoL showed high performance in cancer survivors. The important variables for each HRQoL factor were different. Moreover, there was a specific method to provide customized healthcare services by employing the individual prediction method with SHAP with a web-based survey study for cancer survivors. In advanced cancer patients, the univariate dimensional Cox model showed that ECOG performance status, marital status, sex, global QoL, dyspnea, pain, appetite loss, constipation, depression at baseline, and clinically meaningful change of emotional functioning were predictive factors with worse survival. In the prediction model using MLT, the XGBoost model of survival showed high performance. The performance was optimum when the model was constructed by combining variables selected by the Cox model and MLT methods: depression, pain, appetite loss, constipation, sex, ECOG performance status, and clinically meaningful change in emotional functioning. We also revealed a causal relationship among SAT strategies, depression, and survival in advanced cancer patients using path analysis.

Conclusions: This study is the first to examine the self−management strategy clusters considering cancer stages and different groups of cancer patients, such as cancer survivors and advanced cancer patients. To our knowledge, this is first study to have developed and validated HRQoL prediction models, interpreted the models, and suggested utilization of these results in a clinical setting for cancer survivors. Additionally, we revealed an association of self−management strategies with HRQoL and survival in advanced cancer patients using MLT methods and path analysis. These study results can increase the understanding of self−management strategies and help healthcare providers with healthcare services for cancer patients in the cancer−care continuum.

Keyword: Cancer survivors, Advanced cancer patients, HRQoL, Survival, Self−management strategies, Clustering, Machine learning, Cancer−care continuum

Student Number: 2015−30610

Table of Contents

List of Figures

List of Tables

List of Abbreviation

Abbreviation	Meaning
aOR	adjusted Odds Ratio
AUROC	area under the receiver−operating−characteristic curve
AUPRC	area under the precision−recall curve
CI	Confidence Interval
DT	Decision Trees
GB	Gradient Boosting
HRQoL	Health Related Quality of Life
HS	Health status
ML	Machine Learning
MLT	Machine Learning Technique
NS	Non−Significant
QoL	Quality of Life
RF	Random Forest
SAT	Smart Management Strategies for Health Assessment Tool
SAT−C	SAT Core strategy
SAT−I	SAT Implementation strategy
SAT−P	SAT Preparation strategy
SD	Standard deviation
SEM	Structural equation modeling
SHAP	Shapley Additive exPlanations
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

Chapter 1. Introduction

1.1. Study Background

Cancer is the first leading cause of death globally and in South Korea [1, 2]. Owing to the advancement in early cancer detection and cancer treatment technology, the rate of cancer survival has significantly improved in the previous 10 years [3]. The 5-year survival rate of cancer was 70.7% in 2015–2019 and the number of cancer prevalent cases was 2 million in 2019 in South Korea [2]. Globally, there were 19.3 million new cancer cases and 10 million cancer-related deaths in 2020 [4]. Of the 50 million patients globally, who have cancer in 5 years, many have advanced cancer, requiring palliative care [4]. In 2040, the global cancer burden is predicted to be 28.4 million cases and increase further [5].

With the increasing incidence of cancer, there has been more attention on improving health-related quality of life (HRQoL) among cancer survivors and advanced cancer patients [6-10]. Cancer patients often have various health-related problems including sleep disturbances, fatigue, pain, or constipation; depression or low emotional functioning; loss of relationship or employment; and low meaning of life or low will to live, considering physical, emotional, social, and spiritual HRQoL, respectively. The increasing interest in HRQoL is rather obvious among cancer patients for sustaining their daily lives and survival [6-9]. Similar to chronic diseases, cancer requires continuous HRQoL management because of the continued difficulties described above and the management of cancer itself [11]. HRQoL management is required for both cancer survivors and advanced cancer patients. Advanced cancer patients need palliative care services for improving their symptom management, reducing psychological distress, and enhancing their quality of life [12]. To manage cancer patients' HRQoL, effective strategic approaches are necessary. Engaging cancer patients from diagnosis to end-of-life and managing their HRQoL with self-management strategies can improve their autonomy, enhance their overall QoL, and maintain their dignity until the end of life [13].

In the cancer-care continuum, self-management strategies affected cancer patients' symptom management positively, reduced their psychological distress, and improved the QoL during treatment, after the end of treatment, and at the end of life [11]. Most cancer patients use self-management strategies actively. However, some researchers reported slight differences in critical self-management strategies or the impact of self-management strategies on HRQoL according to treatment period or cancer type [11, 14-16].

Cancer stages, which are classified as early or advanced, are important in the self-management for cancer patients [17, 18]. The early-stage generally refers to stages I and II, while the advanced-stage refers to stages III and IV at time of diagnosis. Irrespective of cancer types, advanced-stage cancer patients have poorer HRQoL and more severe symptoms than early-stage cancer patients [19]. Thus, when applying self-management strategies to cancer patients, we must consider the cancer stage to optimize and improve their HRQoL [20].

Self-management strategies are variously defined [11, 16]. Previous studies on self-management strategies in cancer patients mainly focused on the use of complementary and alternative medicine (CAM) and health behaviors [15]. Physical, emotional, social, and spiritual health behaviors, including healthy diet, exercise, conscious living, turning to family, and religious life, are self-management strategies [14, 15]. Traditionally, the Transtheoretical Model (TTM) was used for health management, which integrates processes and principles of health behavior change across major intervention theories [15]. The TTM had a significant advantage in identifying the transition of health behavior practice steps in healthcare and was an effective strategy for changing smoking cessation management [21-23]. Commonly used self-management strategies in psychology and business administration include leadership strategies that are also applied to health management recently [24-26]. The improved leadership was positively associated with health behavioral practice, positive growth, and physical and mental quality of life in cancer patients [25]. Self-leadership could help patients practice healthy behaviors and cope with psychological difficulties. The improvement of self-leadership may help patients control their behavior and increase self-efficacy [11, 27, 28]. Brief COPE is a crisis and stress coping strategy developed by a psychologist; it includes 14 adaptive and maladaptive strategies [29]. In the study of patients with incurable cancer, emotional support and acceptance strategies were positively related to QoL and mood. Self−blame and denial strategies were negatively associated with QoL and mood [30]. However, CAM, TTM, leadership, and brief COPE have uncertainty around continuous health management as they do not include the strategies to support self−management practice continuously. Thus, we newly defined self−management strategies in cancer patients' healthcare for promoting physical, emotional, social, and spiritual health behaviors, active participation in treatment decision−making, improvement of quality of life, and even positive growth after cancer crisis.

We developed a new assessment tool of self−management strategies called the smart management strategies for health assessment tool (SAT) as per the above definition [31]. The SAT can support cancer patients to practice self−management strategies continuously and overcome cancer−related difficulties. We developed SAT by extracting cancer patients−tailored crisis overcoming strategies according to the interviews with cancer patients and experts and multidisciplinary literature reviews on crisis overcoming strategies and behavioral changes in psychology, business administration, education, healthcare, and medicine. The SAT strategies were aimed at overcoming health crisis and facilitating positive growth. They included three sets of 15 sub−strategies in SAT−C (Core strategy), SAT−P (Preparation strategy), and SAT−I (Implementation strategy): 1) proactive problem−solving, 2) positive−reframing, 3) creating empowered relationships, 4) experience sharing strategies in SAT−C; 5) priority−based planning, 6) goal and action setting, 7) healthy environment development, 8) life value pursuing, and 9) rational decision−making strategies in SAT−P; 1) self−implementation, 2) self−sustaining, 3) activity−coping, 4) self−motivation, 5) reflecting, and 6) energy−conserving strategies in SAT−I [31]. Cancer patients' responses to the three sets were as follows: 1) SAT−C strategies are effective regardless of the treatment time or stage of health behavior practice, 2) SAT−P strategies support the treatment or self−management practice preparation process, 3) SAT−I strategies assist with the treatment or self−management practice. In a previous study, increased use of SAT strategies was associated with better QoL and more health behavior practices [31]. However, there are no studies examining how these 15 sub−strategies could be used in actual cancer patients. The 15 sub−strategies in SAT are interrelated with each other; thus, they are expected not to be used alone. Cancer patients under crisis can adopt many strategies concurrently to overcome the crisis instead of utilizing the strategies individually. Thus, the clustering method can identify how cancer patients use the self−management strategies under crisis.

This study first attempted to identify how the SAT strategies bundled in a cluster differently according to cancer patient groups with different cancer stages. It also examined the relationship of SAT clusters with HRQoL in cancer patients. We included cancer survivors at different cancer stages (early [I, II] and advanced [III, IV]) who can practice health behaviors for improving their HRQoL and survival, and advanced cancer patients diagnosed with stage IV who require palliative care; we examined and compared the pattern of using SAT strategies and their association with HRQoL among cancer survivors with early−stage and advanced−stage cancer and those with a remaining life expectancy of less than 1 year.

To our knowledge, no previous study has been conducted using machine learning models predicting HRQoL, although it is important in the prediction of cancer survivorship. Thus, we also developed and validated HRQoL prediction model for cancer survivors and survival prediction model for advanced cancer patients to determine important features for predicting HRQoL or survival. To facilitate actual application in clinical settings, we developed simple HRQoL prediction model with cancer survivors' socio−demographic and clinical variables and SAT strategies that are easily obtained by a survey. We also used a web−based survey to deliver survey results with visually organized information for health to facilitate easy application in clinical settings. For advanced cancer patients, we developed a survival prediction model with socio−demographic and clinical variables, SAT strategies, and HRQoL features. Our study results would help in the development of related self−management interventions to improve HRQoL or survival among patients in the cancer−care continuum. Moreover, these results can broaden the understanding regarding the importance of self−management strategies for cancer patients to improve their HRQoL and survival and help healthcare providers provide personalized healthcare service in the cancer−care continuum.

1.2. Literature Review

1.2.1. Clustering research in cancer patients

Clustering studies conducted on cancer patients mainly focused on symptom cluster studies for symptom management. The symptom cluster research is based on empirical and theoretical evidence regarding the symptoms that cancer patients experience after diagnosis; these symptoms manifest simultaneously [18]. The symptom cluster studies mainly examined the relationship between various physical and mental symptom clusters experienced by cancer patients and HRQoL [32].

Self-management strategies' clustering studies of cancer patients have not yet been actively conducted. Cancer patients use self-management strategies to overcome crises after cancer diagnosis, manage their health, and improve their quality of life [11]. Psychological, empirical, and theoretical studies revealed that self-management strategies are used simultaneously, as the symptom cluster described above. The studies examined the relationship between QoL of cancer patients by clustering coping strategies through Brief COPE, a questionnaire developed by psychologists; coping strategies of Brief COPE were divided into adaptive (planning, positive reframing, acceptance, religion, use of emotional support, humor, use of instrumental support, and active coping) and maladaptive strategies (denial, self-distraction, substance use, venting, behavioral disengagement, and self-blame) to understand their relationship with QoL [29, 33]. Adaptive coping strategies are effective in improving QoL of cancer patients over time [34].

In addition to Brief COPE, a study using the coping strategy questionnaire and the Cognitive Emotional Regulation Questionnaire (CERQ) showed the relationship between strategy use patterns and QoL of cancer patients who may experience severe emotional distress in special circumstances (pregnant cancer patients) using the clustering method [35]. The nine subscales of CERQ were divided into three clusters (positive, internalizing, blaming coping) using the K-means clustering method; those practicing the internalizing coping strategy could handle the highest level of distress [35].

However, research on self-management strategies' clustering does not compare symptom cluster research. Moreover, no study has identified cancer patients' actual strategy patterns by clustering the newly developed SAT. The author examined different strategy use patterns considering different cancer stages, such as early stages or advanced stages, among 15 sub-strategies of SAT. The author also tried to study whether there is a difference in the association between each SAT cluster and HRQoL among patients at different cancer stages.

1.2.2. Prediction of HRQoL in cancer survivors

Considering HRQoL of cancer survivals, general QoL and patients' subjective health statuses were the prognostic indicators predicting survival [36, 37]. In one literature review study on the relationship between various HRQoL measurements and survival time, global QoL was a critical indicator for predicting the survival of cancer patients [36]. Several studies that used the European Organization for Research and Treatment of Cancer Quality of Life Core Questionnaire (EORCT QLQ-C30) to measure heterogeneous cancer patients' HRQoL reported that global QoL predicted survival [36, 38-40]. In one study, subjective awareness of health status was also a significant predictor of survival in cancer patients [41].

According to previous studies, many factors predict cancer survivors' QoL. Among these, coping strategies, self-management such as health behavior practice, and emotional management were positively related to the QoL [25, 42-45]. Many studies showed that self-management of health behaviors is strongly associated with QoL [25, 42, 43]. Mental health, including posttraumatic growth, subjective life satisfaction, depression, or anxiety, was also associated with cancer patients' QoL [25, 45]. Previous studies showed that cancer patients and survivors with greater mental health had higher QoL [46]. Moreover, coping strategies such as emotional support and acceptance strategies were positively related to QoL and mood, and self-blame and denial strategies were negatively related to QoL and mood [47].

In a systematic review study that investigated personal self−management strategies that affect HRQoL in colorectal cancer patients, low positive reframing, low positive growth, low belief, low meaning−seeking, and low sense of coherence were associated with lower HRQoL in colorectal cancer patients up to 5 years after diagnosis [48]. In a systematic review study examining the relationship between psychological intervention and fatigue in breast cancer survivors, it was found that breast cancer survivors experiencing difficulties with coping had higher levels of fatigue and more negative views about the future [49]. Additionally, various psychoeducational interventional studies improved the HRQoL of cancer survivors. These studies also showed that self−management through coping skills training was helpful in improving HRQoL of cancer survivors [50]. Despite these intervention programs, no study identified self−management strategies important for HRQoL of cancer survivors.

In another study conducted by the author, SAT strategies affected health behavior practice and mental health (satisfaction of life, depression, or anxiety) significantly, which are associated with the HRQoL of cancer survivors. Therefore, there could be a significant association between SAT strategies and the HRQoL of cancer survivors. However, critical variables among SAT strategies that significantly affect HRQoL (global QoL and physical, mental, social, and spiritual health status in this study) of cancer survivors have not yet been identified. Additionally, no study has examined the HRQoL predictive power of cancer survivors using machine learning techniques only employing self−management strategies.

Therefore, in this study, the author found self-management strategies important for predicting HRQoL variables of cancer survivors; HRQoL predictive performance of self-management strategies was also determined. Some essential self-management strategies were expected for customized self-management intervention to improve the QoL of cancer survivors.

1.2.3. Prediction of survival in advanced cancer patients

Previous studies that identified survival predictors for advanced cancer patients revealed that HRQoL had an independent effect on survival [12, 51, 52]. In a study conducted in South Korea with advanced cancer patients, physical functioning, fatigue, vomiting, loss of appetite, and constipation were independent prognostic factors predicting survival [51]. In a secondary analysis of a randomized controlled trial (RCT) study on early palliative care interventions for patients with advanced cancer, depression at diagnosis and significant worsening of depressive symptoms for 3 months were significant prognostic factors predicting survival [12]. Additionally, various HRQoL variables are reported as predictive factors of survival of patients with advanced cancer [51]. Therefore, improving HRQoL can be strongly related to improving survival in advanced cancer patients.

Self-management strategies are strongly related to factors affecting HRQoL that have a significant correlation with the survival of patients with advanced cancer, as described above [53]. From the diagnosis of later−stage cancer, advanced cancer patients face various physical, mental, social, and spiritual challenges related to cancer treatment and prognosis [53]. To manage these challenges, coping strategies can be used for self−management, and this can lead to better QoL and improved mood in advanced cancer patients [54]. In one study, emotional support and acceptance strategies were associated with a higher quality of life and mood, while coping strategies such as rejection and self−blame were associated with lower quality of life and greater depressive and anxiety symptoms [54]. A longitudinal study also revealed that coping capacity and self−efficacy are important variables predicting the long−term QoL of advanced cancer patients [55]. In another study, positive reframing and active coping strategies were associated with better QoL and improved depressive symptoms in advanced cancer patients with accurate prognostic awareness of their condition [34]. This study explained that coping strategies play a role in buffering the accurate perception of the prognosis and the lower quality of life and mood [34]. However, there are inconsistent results regarding the coping strategy can reduce the survival of advanced cancer patients [53]. Despite individual studies considering that HRQoL predicts survival, and self−management strategies (coping strategies) that predict HRQoL in patients with advanced cancer, there is no study that simultaneously examined the relevance of survival with self−management strategies and HRQoL. Therefore, this study aimed to comprehensively reveal how self−management strategies and HRQoL prognostic factors help in predicting the survival of advanced cancer patients through a secondary analysis of an integrated early palliative care intervention RCT study targeting patients with advanced cancer.

1.2.4. K−means clustering analysis, structural equation modeling, and explainable machine learning techniques

For an evidence−based healthcare service, it is crucial to select an appropriate research methodology. For cluster analysis, there are hierarchical and non−hierarchical methods of clustering [56]. In hierarchical clustering analysis, one cluster can be involved within other clusters entirely. However, no overlap between clusters can be allowed [57]. The hierarchical clustering analysis is susceptible to outliers; thus, it is not appropriate for identifying robust patterns.

K−means clustering analysis, which is the most popular non−hierarchical method, can find non−overlapping clusters [58]. Patients can be assigned to clusters once the number of clusters is determined [57]. K−means clustering analysis is less susceptible to outliers and the inclusion/exclusion of irrelevant variables. For identifying the empirical pattern of using SAT strategies among cancer patients, we chose the K−means clustering method to examine 15 sub−strategies of SAT through dimensionality reduction considering different cancer stages on the non−hierarchical position.

Structural equation modeling (SEM) is a research methodology that can compensate for the shortcoming [59]. There are two advantages of SEM over traditional multivariate statistical analysis. First, a series of dependence relationships between variables can be estimated at the same time. Second, measurement errors can be reflected [60]. SEM includes both path analysis that uses bivariate correlations in estimating the relationship between constructs in a path diagram that visually shows all variable relationships and the relationship analysis between construct and measured variables [60]. When a construct is measured in multiple items and combined, averaged, or measured as a single item, if the researcher believes that the reliability cannot be estimated or does not need to be reflected, it is assumed that the measured variable ultimately measures the construct. Only the path analysis is performed, excluding the analysis of the relationship between the construct and the measured variables [60]. SEM is widely used in social sciences and business management administration, but there have also been attempts regarding its application in medical and health sciences [61]. This is because SEM has the statistical strength to analyze complicated relationships among variables and can test linear causal relationships even with non-experimental data [61]. SEM enables researchers to explain the process or result of phenomena such as health behaviors and disease.

Machine learning techniques (MLTs) have gained attention in medical and health sciences. ML models were increasingly adopted in clinical oncology to explain the nonlinear relationships and predict clinical outcomes [62]. The most recent MLT, called explainable MLT, can make individual predictions [63]. Explainable MLTs can analyze the factors that positively or negatively affect the patient’ s specific health outcome using an algorithm such as Shapley Additive Explanation (SHAP).

SHAP values reveal individual prediction by predicting a particular data point (e.g., for one patient) and explaining the reason behind specific data and their outcome (e.g., why one patient had low QoL). Shapley value is obtained by calculating the net contribution for possible combinations of all variables [64]. To calculate the Shapley Value, the marginal contribution is measured while changing each variable considering the combination of all variables. The average marginal contribution implies the shapley value [64]. SHAP can measure contribution to the model by comprehensively considering the interactions between variables [64]. As Shapley values consider all permutations, SHAP is a unified approach with global and local interpretability and consistency [65]. Additionally, SHAP can visualize results intuitively and effectively, which helps in interpreting the results of medical research. Thus, many medical studies of individual prediction and global prediction used SHAP to interpret and visualize the results [65]. SHAP has been used in medical and clinical oncology [66−68]. However, to the best of our knowledge, the SHAP framework has not been used in the health management field. Thus, we applied the SHAP framework to visualize individual predictions for providing customized healthcare services to cancer patients.

1.3. Research Objectives and Hypothesis

1.3.1. Research objectives

This study aimed to examine the relationship among the self-management strategies (SAT), HRQoL, and survival in cancer survivors and advanced cancer patients in the cancer-care continuum (Figure 1).

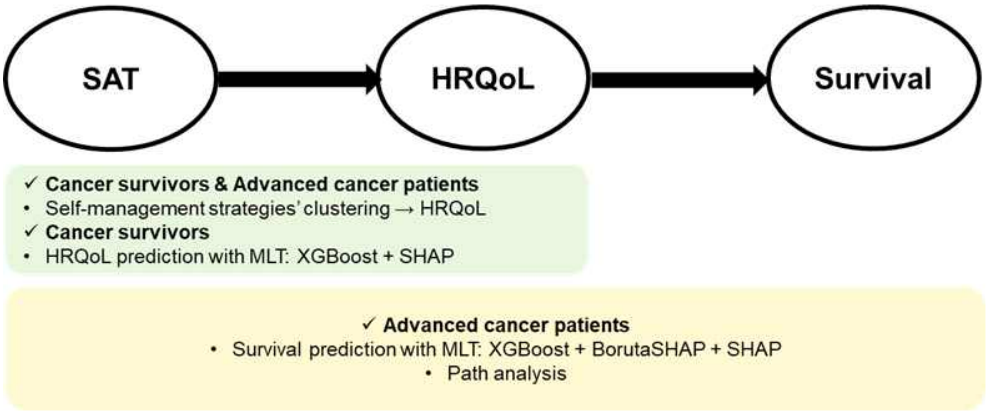


Figure 1. The relationship between SAT, HRQoL, and survival

This study assumed that SAT, HRQoL, and survival would have a causal relationship as shown in Figure 1; the following research objectives were considered for identifying the relationship:

1. To find and compare the different patterns of using self-management strategies (SAT) according to cancer stages among cancer survivors with early and advanced stages and advanced cancer patients in palliative care and to determine the effect of SAT strategies on HRQoL.

2. To develop and validate the HRQoL prediction model in cancer survivors with machine learning techniques.

3. To develop and validate the 1-year survival prediction model in advanced cancer patients with machine learning techniques.

4. To find the causal relationship of self-management strategies, HRQoL, and survival comprehensively using a path analysis.

1.3.2 Research hypothesis

First, the SAT-C (Core) strategies that aid cancer patients to overcome various difficulties are used proactively with SAT-P (Preparation); SAT-I (Implementation) is used primarily in advanced-stage cancer patients than in early-stage cancer patients. As advanced-stage cancer patients commonly have more physical, emotional, social, and spiritual difficulties than early-stage cancer patients, the former will use the core strategies more actively to overcome severe difficulties [30, 69]. The usage pattern of the strategy is the same for advanced-stage cancer survivors as well as advanced cancer patients in palliative care. Additionally, SAT strategies are related to HRQoL in both cancer survivors and advanced cancer patients.

Second, the simple model with socio-demographic and clinical variables and SAT strategies will predict HRQoL with high performance. In this model, the SAT strategies will be more critical to predict HRQoL than socio-demographic and clinical variables.

Third, HRQoL variables will be more critical in predicting

1-year survival in advanced cancer patients than SAT strategies.

Finally, although the SAT strategies have a strong relationship with HRQoL, they are not significantly related to survival in advanced cancer patients. However, HRQoL can act as a mediating variable between SAT and survival.

1.4. Definition of cancer survivors and advanced cancer patients in this study

1.4.1 Definition of Cancer survivors

The National Coalition for Cancer Survivorship (NCCS), a non-profit organization of cancer-related experts in the United States, proposed a new concept to define cancer survivors [70]. The survival of cancer was defined as the entire process from cancer diagnosis until treatment [71]. There are many other ways to define cancer survivors. According to Van Leeuwen et al. (2018), cancer survivors imply those who have completed treatment and are disease-free, i.e., those diagnosed with cancer but without any evidence of active disease following treatment [72]. Historically, the concept of cancer survivors has also meant surviving for several years without recurrence or metastasis after cancer treatment [73].

In this study, cancer survivors were defined as those who survive after the diagnosis of cancer, excluding those with terminal cancer.

1.4.2. Definition of Advanced cancer patients

The American Cancer Society (ACS) suggests that advanced cancer is incurable and usually spreads from its original site; however, all advanced cancers are not necessarily metastatic [74]. The National Cancer Institute (NCI) also presents that advanced cancer can spread to other sites of the body and cannot be cured or managed by treatment [75].

In this study, advanced cancer survivors were defined by the physician (oncologist) as progressive cancer patients who were histologically or cytologically diagnosed with a solid tumor with a life expectancy of less than 12 months.

Chapter 2. Methods

2.1. Study Design

2.1.1. Study Design for Cancer Survivors

This was a prospective cohort study with a web-based survey. The cancer survivors completed baseline and post-baseline surveys after 6 months on "HealthingU" website. HealthingU was developed to deliver visually organized health information immediately after participants completed the surveys (Figure 2). Participants were encouraged to finish each survey on HealthingU website; however, those who could not access online surveys received the survey and the result by mail.

2.1.2. Study design for Advanced cancer patients

This was a randomized controlled trial including early palliative care to identify efficacy of the newly developed integrated early palliative care program and to compare it with the existing standard palliative care. However, we explored the hypotheses described above in a secondary analysis using this clinical trial data.

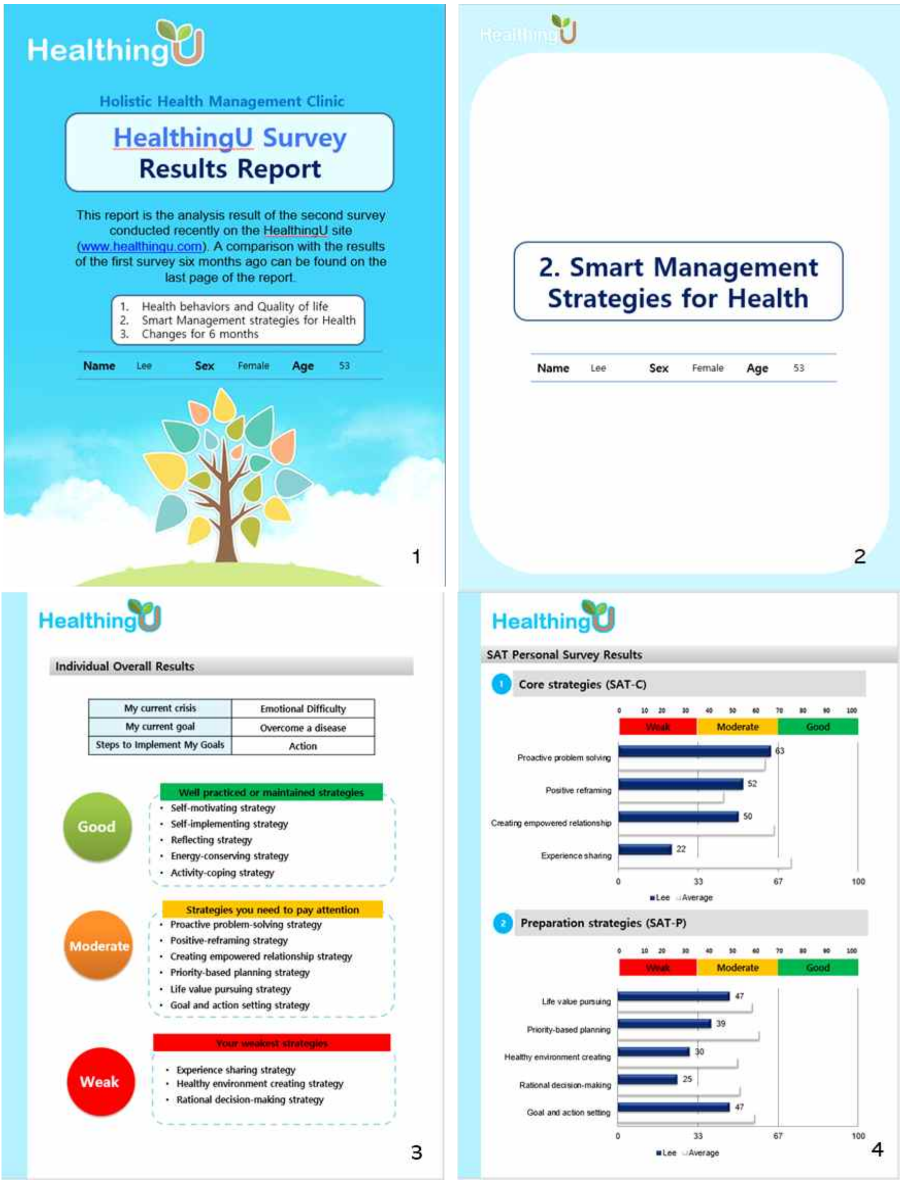


Figure 2. HealthingU survey report sample provided to cancer survivors in the study

2.2. Study Participants

2.2.1. Cancer Survivors

In total, 540 cancer survivors who provided informed consent and completed the baseline questionnaire participated in the study from four hospitals in Korea. Cancer survivors visited outpatient clinics and were recommended by the physician to participated in this study. A coordinator from each hospital explained this study in detail, obtained the written informed consent, and assisted in conducting a web-based surveys. Ultimately, 256 cancer survivors were enrolled in the study after completing both the baseline and post-baseline surveys for 6 months. The inclusion criteria were as follows: 1) diagnosed with cancer, 2) over 18 years old, 3) able to read and write Korean, 4) able to use the internet and have an email address, 5) submitted the written informed consent. This study was approved by the institutional review board from four hospitals (IRB No. 1308-087-514).

2.2.2. Advanced Cancer Patients

Among 151 advanced cancer patients who submitted informed consent and were subsequently screened, 144 participated in the study. After 12 weeks, 64 advanced cancer patients were excluded due to death of 47 and withdrawal or censorship of 17 of them; finally, 80 advanced cancer patients remained. We analyzed the data of 144 advanced cancer patients using intention to treat. The inclusion criteria were as follows: 1) over 20 years old, 2) progressive cancer patients with histologically or cytologically diagnosed solid tumor, 3) those with ECOG performance status of 0–2, 4) oncologist-determined prognosis within 12 months, and 5) those who wish to participate in the research. The exclusion criteria were as follows: 1) cannot speak, listen, and read Korean, 2) those who are unable to participate due to poor health conditions (e.g., shortness of breath), 3) those who stopped chemotherapy, and 4) those that previously received palliative care or are presently receiving it. This trial was registered in ClinicalTrials.gov, ID NCT03181854.

2.3. Measurements

2.3.1. Measurements for Cancer Survivors

I. Socio−demographic and clinical variables

The participants' age, gender, income levels, educational levels, residence, marital status, religion, cancer type, cancer stage, and treatment stage were recorded. Considering distribution of the participants, the socio−demographic and clinical variables' cut−off points were converted to binary variables for subsequent analysis. The cut−off point was set where the number of participants is about half distributed or based on a similar status (e.g., 'religion' has various religions, but it was divided as 'having religion' or 'no religion').

II. Self−management strategies

Self−management strategies in cancer survivors were assessed by the SAT (Supplementary Information 1). The SAT comprises the following three sets for assessing self−management strategies to overcome health crises: SAT−C (core strategies) included 28 items and 4 subscales; SAT−P (preparation strategies) included 30 items and 5 subscales; SAT−I (implementation strategies) included 31 items and 6 subscales [31]. SAT is measured as 1-4 points (1 point = not used at all, 4 points = very good). The score ranges of SAT were from 0 to 100. Higher SAT scores indicated better self−management strategies. Following conversion of continuous SAT scores into binary variables, they were classified based on 66.66. The SAT scores higher than 66.66 mean higher strategy group and were less than that of 66.66 mean lower strategy group.

The change scores was calculated using different scores between the baseline and post-baseline SAT scores. When we converted the continuous SAT change scores into binary variables, SAT scores were classified based on clinically meaningful change scores defined as those with an effect size larger than 0.5 in comparison to baseline at post-baseline (after 6 months) [76]. When categorical features were not required for analysis, we used continuous scores.

III. Health Related Quality of Life

HRQoL for cancer survivors was measured by the European Organization for Research and Treatment of Cancer quality of life questionnaire C-30 (EORTC QLQ-C-30) and Health Status Questionnaire (HSQ). The EORTC QLQ-C-30 has 30 items with five functional scales, nine symptom scales, and one global QoL [77]. We chose the global QoL score to represent QoL measurement because it was verified as the survival predictor in cancer survivors, and its meaning was the closest to general QoL for healthcare. The score ranges were from 0 to 100; higher scores implied better QoL.

Health Status Questionnaire (HSQ) assesses health status in cancer survivors holistically considering physical, emotional, social, and spiritual health statuses [78, 79]. The scale ranges were as follows: 1 = Excellent, 2= Very Good, 3=Good, 4=Bad, 5 = Very Bad. We inverted the scale scores as 1= Very Bad and 5=Excellent to match the score direction between HSQ and the global QoL. Higher scores indicated better health status.

2.3.2. Measurements for Advanced Cancer Patients

I. Socio-demographic and clinical variables

The participants' age, sex, income levels, educational levels, residence, marital status, religion, tumor site, Eastern Cooperative Oncology Group (ECOG) performance status, early palliative care (EPC) groups, and death after 1 year of study were recorded. Categorical variables such as sex, income, or educational status were converted to binary variables in the same way as cancer survivors' data.

II. Self-management strategies

The measurement of self-management strategies in advanced cancer patients was the same as that in cancer survivors. However, SAT Short-Form (SAT-SF) was used to reduce fatigue in advanced cancer patients [80].

III. Health-Related Quality of Life

HRQoL for advanced cancer patient was measured by the

EORTC QLQ-C15-PAL, which is a short form of the EORTC QLQ-C30 for use in palliative care [81], the McGill QoL Questionnaire (MQOL), and the Patient Health Questionnaire-9 (PHQ-9). The EORTC QLQ-C15 PAL has two functional scales (physical and emotional functioning), seven symptom scales (fatigue, nausea and vomiting, pain, dyspnea, insomnia, appetite loss, constipation, diarrhea, financial difficulties), and one global QoL. The score ranges were the same as that of EORTC QLQ-C30. Higher functional scales and global QoL' scores implied better QoL and higher symptom scales implied worse QoL.

MQOL measures subjective well-being of patients with a life-threatening disease such as cancer considering four domains: physical, mental, existential well-being, and social support [82, 83]. We only used existential well-being and social support domains with eight items in this study. The scale ranges were from 0 to 10; higher score implied better QoL.

PHQ-9 measures depression symptoms developed in compliance with the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition(DSM-IV); it also corresponds to the DSM-V criteria [84, 85]. The PHQ-9 comprises nine items with scale ranges from 0 to 3. Depression symptom was calculated by adding all nine items' scores (ranges, 0 to 27). The severity of depressive symptom was as follows: none-minimal (score, 0-4), mild (score, 5-9), moderate (score, 10-14), moderately severe (score, 15-19), and severe (score, 20-27).

The change in scores in HRQoL was calculated using different scores between the baseline and 12 weeks. When we converted the continuous change scores into binary variables, the change scores were classified based on clinically meaningful change scores defined as those with an effect size larger than 0.5 compared to baseline scores at 12 weeks [76].

2.4. Statistical Methods

2.4.1. Self-management strategy clustering

For clustering analyses, a principal component analysis (PCA) with varimax rotation was performed to examine the interrelationship among 15 SAT self-management strategies and to reduce their dimensionality [18]. To ensure suitability of the data for the PCA, Kaiser-Meyer-Olkin (KMO), inter-item correlational coefficient, and Bartlett's test of sphericity were first confirmed [18]. Cancer survivors and advanced cancer patients were assigned to a cluster subgroup as per the nearest mean between cluster groups and the squared Euclidean distance implying similarity among the participants' factor loadings [86].

2.4.2. The relationship of self-management strategy clustering with HRQoL

Adjusted least squares means (LS means) of HRQoL (the global QoL and physical, emotional, social, and spiritual health statuses) were compared between two subgroups in each cluster with general linear modeling. Subgroup analyses sorting the participants into early-stage or advanced-stage groups especially for cancer survivors were performed to identify different effects of the self-management strategy's cluster on HRQoL. A two-tailed p-value < 0.05 was considered statistically significant. IBM SPSS version 23 (IBM Corp., Armonk, NY, USA) and SAS version 9.3 (SAS Institute, Cary, NC, USA) were used for statistical analyses.

2.4.3. HRQoL Prediction model development and validation with machine learning techniques

I. Data preprocessing

Before developing prediction model for HRQoL among cancer survivors, data pre-processing was first performed as follows [87]:

First, we performed missing imputation of primary data acquired by Likert scales before scoring. We used a Markov Chain Monte Carlo (MCMC) multiple imputation method because MCMC can predict categorical and continuous features together while handling arbitrary pattern of data missing [88]. For uniformity, we eliminated the features with >20% missing values before applying MCMC to impute all remaining missing data [87].

Second, all categorical variables were converted into binary features to form prediction models using machine learning techniques. As "cancer type" was difficult to divide into binary features, a one-hot encoding method was conducted to convert "cancer type" into dummy variable. Variables of continuous scores were used in prediction models to suggest a personalized healthcare plan based on cancer survivors' scores of survey results presenting the influence of SAT scores on HRQoL or survival. Thus, other pre-processing steps including normalization and scalarization were not performed.

Third, outcome features (HRQoL) were encoded as binary features. The global QoL was encoded based on the scoring report classifying the problematic group (≤33.33) or non-problematic group (>33.33). Four health statuses (physical, emotional, social, and spiritual health statuses) were encoded as "Good" (1 = Excellent, 2 = Very Good, 3 = Good) or "Bad" (4 = Bad, and 5 = Very Bad) [87]. The global QoL was the primary outcome, and four health statuses were the secondary outcomes for cancer survivors.

Finally, we constructed prediction models by removing one variable in SAT (the self-sustaining strategy) because it showed high correlation with other variables (Pearson correlation > 0.7) to avoid multicollinearity and to improve stability of the prediction model in cancer survivors [67, 89].

II. Machine Learning Techniques

To develop and validate the predictive model' s performance, various machine learning algorithms were used including decision trees (DT), random forest (RF), gradient boosting (GB), eXtreme Gradient Boost (XGBoost), and LightGBM; these ensemble algorithms did not require pre-processing, i.e., normalization and scalarization of each feature. We chose these ensemble methods to use cancer survivors' survey scores without any processing. We finally selected XGBoost algorithm to develop and validate prediction models and to extract important features and individual prediction results [87]. For advanced cancer patients, only XGBoost algorithm was used to predict performance of survival prediction models.

III. Extreme gradient boosting (XGBoost)

XGBoost is a widely used tree-based ensemble learning algorithm

[90]. XGBoost generally showed greater prediction performance than other algorithms [91]. Although XGBoost is fundamentally based on GBM, it is used more frequently than GBM because it addresses some limitations of GBM such as lack of regularization and slow execution time [90, 92]. Both XGBoost and LightGBM are relatively new ensemble algorithms based on GBM with no significant difference in prediction performance. However, LightGBM could easily show overfitting problems when applied to a relatively small dataset [93]. Thus, XGBoost was finally selected to develop and validate the prediction models in this study.

IV. Feature importance and individual prediction with Shapley additive explanations

Recently, the techniques of explainable artificial intelligence (XAI) have emerged to improve the interpretability of complex machine learning (ML) models [64]. In medicine and healthcare, the interpretability is critical to derive a conclusion depending on the obtained MLT results [94].

SHAP algorithm is a popular XAI technique [65, 95]. SHAP values can depict important features with the overall model performance indicating positive or negative values that are visually intuitive. SHAP values can predict a specific data point and explain the reason behind a particular health outcome [65]. For example, SHAP values can predict one cancer survivor’s global QoL outcome and explain why the survivor had low global

QoL with specific variables affecting the QoL positively or negatively. The individual prediction of SHAP is a new function of XAI that was not implemented in conventional statistical methods [87].

SHAP values are obtained by evaluating and adding the net contribution for feasible combinations of all variables [95]. The marginal contribution of each feature was calculated by changing it while considering the combination of all features [95]. SHAP values are the average of the marginal contributions obtained by this approach; they approve measurement of the contribution to the prediction model by considering interaction between features comprehensively [95].

V. Model Development and Selection

For selecting final prediction model variables, three XGBoost models predicting the global QoL as the primary outcome were trained as follows (Figure 3) [87]:

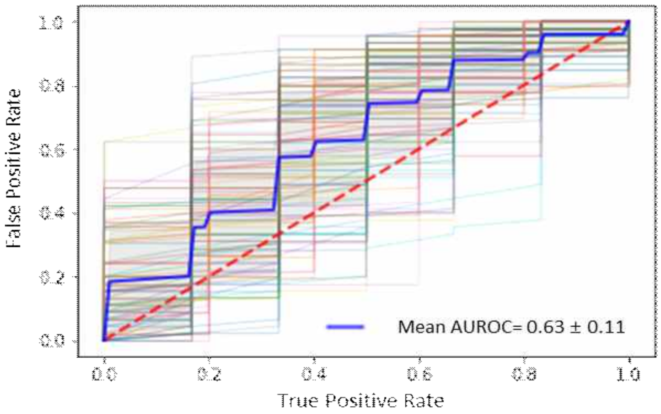
1. Model 1: The global QoL prediction model based on socio-demographic and clinical variables

2. Model 2: The global QoL prediction model based on socio-demographic and clinical variables + SAT scores

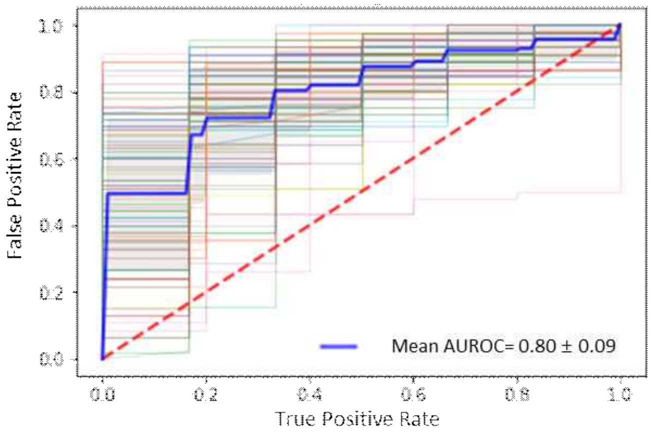
3. Model 3: The global QoL prediction model based on only SAT scores

We compared prediction performance of the above three models to identify the impact of SAT scores on QoL prediction. After identifying the high power of SAT scores on predicting the QoL, we selected Model 2 including all socio-demographic, clinical variables, and SAT scores.

(A) Model 1



(B) Model 2



(C) Model 3

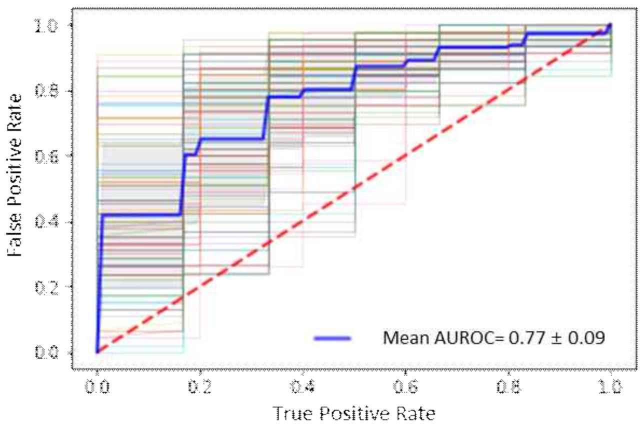


Figure 3. Comparison of prediction performance among models 1, 2, and 3 for final model selection

VI. Final Prediction Model Training and Evaluation

For obtaining the prediction models for HRQoL, XGBoost models were trained using the input features in Model 2 described above. The prediction performance of these XGBoost models was compared with the following four ensemble algorithms in MLT: decision tree, random forest, gradient boosting, and LightGBM. DT and RF are relatively conventional classifiers and GB and LightGBM are recent MLT classifiers.

All analyses were performed using open-source libraries called scikit-learn in Python 3.8. We conducted five-fold cross-validation to determine optimal model hyperparameters for the training dataset using grid search. The obtained optimal hyperparameters were applied to test area under the receiver-operating-characteristic curve (AUROC) on the test dataset. After tuning the hyperparameters for each model, validation of repeated stratified K-fold was conducted to evaluate the final models' performance with fine-tuned parameters; the models were trained in four folds iteratively and tested in the fifth fold (Figure 4). As the dataset was imbalanced, the repeated stratified K-fold method was selected for evaluation.

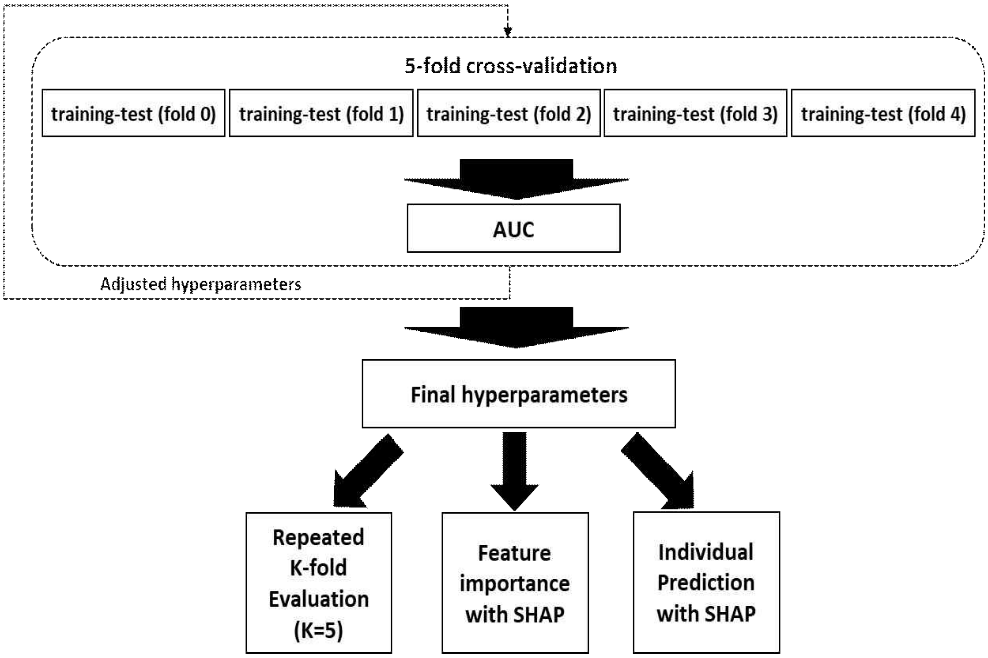


Figure 4. Flowchart of finding model parameters, model evaluation, important features, and individual prediction [94]

Feature importance assessments, individual prediction, and visualized interpretations were also conducted using open-source libraries as scikit learn, SHAP, and XGBoost in Python 3.8. Using the open-source XGBoost package (version 0.81.), gradient-boosted regression tree (GBRT) models were trained. SHAP values were analyzed using the open source SHAP package (version 0.29.1). Visualized results and graphs were generated using the SHAP package. The area under the precision-recall curve (AUPRC), AUROC, accuracy, and F1 scores were utilized for prediction model assessment [94].

2.4.4. Survival Prediction model development and validation with machine learning techniques

I. Critical feature selection with survival analyses

A multidimensional approach with three steps was performed to examine critical predictors of survival in advanced cancer patients [96]. In the first step, we conducted univariate analyses using the Cox proportional hazard regression models to screen associated predictors with survival within each dimension. Next, we conducted multiple Cox proportional hazard regression with one-dimensional approaches to examine potentially critical predictors. Then, we selected significantly associated variables for the subsequent one-dimensional multiple Cox proportional hazard regression analysis with a backward selection method [97]. Finally, we conducted multiple Cox proportional hazard regression analysis with a multidimensional approach including selected variables from the second step to examine ultimate critical predictors of survival. SAS version 9.3 (SAS Institute, Cary, NC, USA) was used and the statistical significance was defined as $P < 0.05$. The change scores of variables were analyzed by intention to treat, i.e., all participants assigned randomly were included in the statistical analyses with patients who died after enrolment [98]. Sensitivity analyses were also performed to analyze the impact of handling missing data using average imputation and compete case analysis [98].

II. Developing survival prediction model and validation with machine learning techniques

The development and validation of survival prediction models with MLT was the same as that for cancer survivors overall. However, the predictive variable (outcome) was survival, not HRQoL. Additionally, the predictive model was composed of all selected features, which were selected in univariate analyses using Cox regression analyses. Only selected features were used in prediction model 1 to reduce multi−collinearity because HRQoL variables are highly correlated with each other.

For selecting important variables predicting the survival in MLT, we used BorutaSHAP selection method to select critical features for survival [99]. BorutaSHAP selects all variables related to a given ensemble model (XGBoost in this study) along with all relevant features [99]. During the implementation of BorutaSHAP in Python, BorutaSHAP produces more consistent result than other metrics because the result is obtained by comparing averages of features and shadow features' SHAP importance values [95, 100]. Thus, we chose BorutaSHAP as a feature selection method of MLT and developed model 2 using selected variables with BorutaSHAP.

Lastly, we developed model 3 including the selected variables with BorutaSHAP and dimensional multiple Cox proportional regression analysis. Then, we compared the predictive performances of three models as follows: 1) the predictive model of all selected variables, 2) the predictive survival model of selected variables by BorutaSHAP, and 3) the predictive survival model of combined variables selected by BorutaSHAP in MLT and by dimensional multiple Cox proportional regression model in the conventional statistical method.

2.4.5. Causal relationship among SAT, HRQoL, and Survival

For examining the causal relationship of SAT with HRQoL and survival, we conducted subgroup analyses and path analyses of SEM. After finding critical predictors for predicting survival in advanced cancer patients in the Cox proportional regression model, we performed a path analysis to identify the causal relationship among SAT, the critical HRQoL variable, and survival.

Chapter 3. Results

3.1. Study Participants' characteristics

3.1.1. Cancer survivors' characteristics

The participants' characteristics are summarized in Table 1. Approximately 50% of survivors were over 50 years of age. Approximately 23% of survivors were at cancer stages III and IV. The treatment of approximately 66% of survivors was 5 years after treatment termination (Table 1).

Table 1. Cancer survivors' socio-demographic and clinical characteristics [101]

Total (N=256)	No (%)
Age(years)	
<50	115(44.9)
≥50	141(55.1)
Income (won)	
<4,000	125(48.8)
≥4,000	131(51.2)
Education	
≤High school graduates	97(37.9)
≥University graduates	159(62.1)
Residence	
Other areas	102(39.8)
Metropolitan area	154(60.2)
Sex	
Male	91(35.5)
Female	165(64.5)
Marriage	
Single	37(14.5)
Married	219(85.5)
Religion	
No	94(36.7)
Yes	162(63.3)
Cancer type	
Breast cancer	93(36.3)
Lung cancer	69(27)
Colon cancer	57(22.3)
Gastric cancer	17(6.6)
etc.	20(7.8)
Cancer stage	
Ⅰ, Ⅱ	196(76.6)
Ⅲ,Ⅳ	60(23.4)
Treatment stage	
Diagnosis	0(0)
Treatment progress	1(0.4)
Withing 5 years of treatment termination	85(33.2)
5 years after treatment termination	170(66.4)

3.1.2. Advanced cancer patients' characteristics

The characteristics of advanced cancer patients are summarized in Table 2. The average age of advanced cancer patients was 60.68 years and 57.6% of the patients were men; 81.3% of the patients showed ECOG performance status 1 and 50.7% patients participated in the holistic care of early palliative care (Table 2).

Table 2. Advanced cancer patients' socio-demographic and clinical characteristics

Features	Category	N (%) or Mean ±SD	Min-Max
Age		60.68 (8.9)	38-82
Sex	Male	83 (57.6)	
	Female	61 (42.4)	
Monthly income (won)	<3,000,000	111 (77.1)	
	≥3,000,000	32 (22.2)	
Educational Status	<High school graduates	68 (47.2)	
	≥High school graduates	76 (52.8)	
Residence	Other areas	86 (59.7)	
	Metropolitan area	58 (40.3)	
Marital status	Single	35 (24.3)	
	Married	109 (75.7)	
Religion	No	61 (42.4)	
	Yes	83 (75.6)	
Tumor site	Breast & Gastrointestinal cancer	43 (29.9)	
	Lung, Hepatobiliary & other cancer	101 (70.1)	
ECOG performance status	0	7 (4.9)	
	1	117 (81.3)	
	2	20 (13.9)	
Palliative care	Integrated Early palliative care	73 (50.7)	
	Standard palliative care	71 (49.3)	

3.2. Self-management clustering results

3.2.1. Self-management clustering results in cancer survivors

The adequacy of the data for PCA was examined as KMO=0.947, Bartlett's test of sphericity ≤0.0001, and the correlation coefficients of all SAT subscales ≥0.3 [101]. The results of the PCA in cancer survivors showed the following two strategy clusters: 1) strategy cluster 1 including nine SAT strategies, i.e., self-implementing strategy (SAT-I), self-sustaining strategy (SAT-I), priority-based planning strategy (SAT-P), goal and action setting strategy (SAT-P), activity-coping strategy (SAT-I), healthy environment creating strategy (SAT-P), self-motivation strategy (SAT-I), reflecting strategy (SAT-I), and energy-conserving strategy (SAT-I); and 2) strategy cluster 2 including six SAT strategies, i.e., creating empowered relationship strategy (SAT-C), experience sharing strategy (SAT-C), positive-reframing strategy (SAT-C), proactive problem-solving strategy (SAT-C), life value pursuing strategy (SAT-P), and rational decision-making strategy (SAT-P). The strategy cluster 1 and strategy cluster 2 explained 67.44% of the whole variance (cluster 1, 39.87%; cluster 2, 27.57%) [101] (Table 3).

In survivors with early-stage cancer, there were also two clusters as follows: 1) strategy cluster 1 including 12 SAT strategies, i.e., rational decision-making strategy (SAT-P), proactive problem-solving strategy (SAT-C), goal and action setting strategy (SAT-P), self-sustaining strategy (SAT-I), self-implementing strategy (SAT-I), self-motivation strategy (SAT-I), positive-reframing strategy (SAT-C), activity-coping strategy (SAT-I), healthy environment creating strategy (SATP), life value pursuing strategy (SAT-P), creating empowered relationship strategy (SAT-C), and priority-based planning strategy (SAT-P); and 2) strategy cluster 2 including three SAT strategies, i.e., experience sharing strategy (SAT-C), energy-conserving strategy (SAT-I), and reflecting strategy (SAT-I) [67]. The strategy cluster 1 and 2 explained 66.33% of the whole variance (cluster 1, 41.53%; cluster 2, 27.8%).

In survivors with advanced-stage cancer, there were also two clusters as follows: 1) strategy cluster 1 including 11 SAT strategies, i.e., self-implementing strategy (SAT-I), self-sustaining strategy (SAT-I), goal and action setting strategy (SAT-P), priority-based planning strategy (SAT-P), activity-coping strategy (SAT-I), self-motivation strategy (SAT-I), proactive problem-solving strategy (SAT-C), healthy environment creating strategy (SAT-P), reflecting strategy (SAT-I), life value pursuing strategy (SAT-P), and energy-conserving strategy (SAT-I); and 2) strategy cluster 2 including four SAT strategies, i.e., experience sharing strategy (SAT-C), creating empowered relationship strategy (SAT-C), positive-reframing strategy (SAT-C), and rational decision-making strategy (SAT-P).

Table 3. Factor loadings from the PCA result of SAT strategies in cancer survivors [101]

Total (256)	Factors		Advanced stage 3–4	Factors		Early stage 1–2	Factors	
SAT strategies	Cluster 1	Cluster 2	SAT strategies	Cluster 1	Cluster 2	SAT strategies	Cluster 1	Cluster 2
Self–implementing strategy (SAT–I)	.856		Rational decision–making strategy (SAT–P)	.910		Self–implementing strategy (SAT–I)	.872	
Self–sustaining strategy (SAT–I)	.826		Proactive problem–solving strategy (SAT–C)	.869		Self–sustaining strategy (SAT–I)	.852	
Priority–based planning strategy (SAT–P)	.804		Goal and action setting strategy (SAT–P)	.847		Goal and action setting strategy (SAT–P)	.796	
Goal and action setting strategy (SAT–P)	.771		Self–sustaining strategy (SAT–I)	.795		Priority–based planning strategy (SAT–P)	.791	
Activity–coping strategy (SAT–I)	.735		Self–implementing strategy (SAT–I)	.777		Activity–coping strategy (SAT–I)	.755	
Healthy environment creating strategy (SAT–P)	.690		Self–motivation strategy (SAT–I)	.772		Self–motivation strategy (SAT–I)	.714	
Self–motivation strategy (SAT–I)	.668		Positive–reframing strategy (SAT–C)	.758		Proactive problem–solving strategy (SAT–C)	.669	
Reflecting strategy (SAT–I)	.636		Activity–coping strategy (SAT–I)	.756		Healthy environment creating strategy (SAT–P)	.661	
Energy–conserving strategy (SAT–I)	.427		Healthy environment creating strategy (SAT–P)	.731		Reflecting strategy (SAT–I)	.611	
Creating empowered relationship strategy (SAT–C)		.793	Life value pursuing strategy (SAT–P)	.718		Life value pursuing strategy (SAT–P)	.604	
Experience sharing strategy (SAT–C)		.775	Creating empowered relationship strategy (SAT–C)	.717		Energy–conserving strategy (SAT–I)	.455	
Positive–reframing strategy (SAT–C)		.714	Priority–based planning strategy (SAT–P)	.622		Experience sharing strategy (SAT–C)		.824
Proactive		.631	Experience sharing		.845	Creating empowered		.778

problem−solving strategy (SAT−C)			strategy (SAT−C)			relationship strategy (SAT−C)		
Life value pursuing strategy (SAT−P)		.612	Energy−conserving strategy (SAT−I)		.783	Positive−reframing strategy (SAT−C)		.644
Rational decision−making strategy (SAT−P)		.565	Reflecting strategy (SAT−I)		.588	Rational decision−making strategy (SAT−P)		.608
Cronbach's α	0.929	0.887	Cronbach's α	0.954	0.645	Cronbach's α	0.942	0.805
Eigenvalues	9.082	1.033	Eigenvalues	8.798	1.151	Eigenvalues	9.754	1.003
Explained variance	5.981	4.135	Explained variance	6.23	3.719	Explained variance	7.515	3.242
Explained %	39.87	27.57	Explained %	41.53	24.80	Explained %	50.1	21.61
Cumulative %	39.87	67.44	Cumulative %	41.53	66.33	Cumulative %	50.1	71.71

3.2.2. Self-management clustering results in advanced cancer patients

The adequacy of data for the PCA was examined as, KMO=0.945, Bartlett's test of sphericity ≤0.0001, and the correlation coefficients of all SAT subscales ≥0.3. The results of the PCA in advanced cancer patients yielded the following two strategy clusters: 1) strategy cluster 1 including nine SAT strategies, i.e., positive-reframing strategy (SAT-C), proactive problem-solving strategy (SAT-C), experience sharing strategy (SAT-C), creating empowered relationship strategy (SAT-C), life value pursuing strategy (SAT-P), activity-coping strategy (SAT-I), goal and action setting strategy (SAT-P), self-sustaining strategy (SAT-I), and energy-conserving strategy (SAT-I); and 2) strategy cluster 2 including six SAT strategies, i.e., healthy environment creating strategy (SAT-P), self-implementing strategy (SAT-I), priority-based planning strategy (SAT-P), rational decision-making strategy (SAT-P), self-motivation strategy (SAT-I), and reflecting strategy (SAT-I). The strategy cluster 1 and 2 explained 72.18% of the whole variance (cluster 1, 39.93%; cluster 2, 32.25%) (Table 4).

Table 4. Factor loadings from the results of principal component analysis of SAT strategies in advanced cancer patients

Total (144) SAT strategies	Factors Cluster 1	Cluster 2
Positive−reframing strategy (SAT−C)	.827	
Proactive problem−solving strategy (SAT−C)	.815	
Experience sharing strategy (SAT−C)	.797	
Creating empowered relationship strategy (SAT−C)	.788	
Life value pursuing strategy (SAT−P)	.746	
Activity−coping strategy (SAT−I)	.712	
Goal and action setting strategy (SAT−P)	.673	
Self−sustaining strategy (SAT−I)	.646	
Energy−conserving strategy (SAT−I)	.553	
Healthy environment creating strategy (SAT−P)		.879
Self−implementing strategy (SAT−I)		.799
Priority−based planning strategy (SAT−P)		.749
Rational decision−making strategy (SAT−P)		.700
Self−motivation strategy (SAT−I)		.618
Reflecting strategy (SAT−I)		.617
Cronbach's α	0.910	0.865
Eigenvalues	9.779	1.048
Explained variance	5.990	4.838
Explained %	39.93	32.25
Cumulative %	39.93	72.18

3.3. The association of self-management clustering with HRQoL

3.3.1. The association of self-management clustering with HRQoL in cancer survivors

Among cancer survivors, higher-strategy group showed significantly improved global QoL and all health statuses (physical, emotional, social, and spiritual) after 6 months (Table 4). The strategy cluster 1 and 2 showed the same results. Higher SAT strategies at baseline predicted better global QoL and overall health statuses after 6 months [101].

In early-stage cancer survivors, higher-strategy group in cluster 1 showed improved global QoL and spiritual health status after 6 months. However, higher-strategy group in cluster 2 showed greater differences in global QoL and all health statuses after 6 months.

In advanced-stage cancer survivors, only higher-strategy group in cluster 1 showed greater differences in overall health statuses after 6 months. There was no difference in the global QoL between subgroups in strategy cluster 1 and 2. Only higher SAT strategies in cluster 1 predicted better health statuses after 6 months (Table 5).

Table 5. Comparison of adjusted LS means between higher and lower strategy groups according to the strategy clusters among total and cancer stage' s subgroups in cancer survivors[101]

Groups	Global Quality of Life (after 6 months) lsmeans(SE)	*P*	Physical Health Status (after 6 months) lsmeans(SE)	*P*	Emotional Health Status (after 6 months) lsmeans(SE)	*P*	Social Health Status (after 6 months) lsmeans(SE)	*P*	Spiritual Health Status (after 6 months) lsmeans(SE)	*P*
Total (N=256) *										
Strategy Cluster 1										
Lower strategy group	50.86(7.19)	0.0014	2.50(0.32)	0.014	2.85(0.31)	0.006	2.89(0.27)	0.024	2.76(0.32)	<.0001
Higher strategy group	58.76(7.12)		2.77(0.32)		3.14(0.31)		3.10(0.27)		3.21(0.32)	
Strategy Cluster 2										
Lower strategy group	50.66(7.12)	0.0003	2.40(0.31)	<.0001	2.79(0.30)	<.0001	2.81(0.27)	<.0001	2.70(0.31)	<.0001
Higher strategy group	60.12(7.11)		2.91(0.31)		3.24(0.30)		3.21(0.27)		3.33(0.31)	
Early stage 1-2 (N=196)**										
Strategy Cluster 1										
Lower strategy group	50.44(7.53)	0.015	2.62(0.34)	0.104	2.76(0.33)	0.116	2.94(0.28)	0.18	2.68(0.33)	0.003
Higher strategy group	57.37(7.53)		2.83(0.34)		2.95(0.33)		3.09(0.28)		3.06(0.33)	
Strategy Cluster 2										
Lower strategy group	49.45(7.32)	0.0002	2.57(0.33)	0.005	2.66(0.31)	0.0001	2.87(0.27)	0.001	2.62(0.32)	<.0001
Higher strategy group	61.21(7.48)		2.97(0.34)		3.18(0.32)		3.24(0.28)		3.28(0.33)	
Advanced stage 3-4 (N=60)**										
Strategy Cluster 1										
Lower strategy group	58 (5.09)	0.616	2.38(0.22)	0.003	2.91(0.21)	0.0024	2.91(0.20)	0.006	2.89(0.21)	0.005
Higher strategy group	60.57(5.07)		3.06(0.21)		3.59(0.21)		3.49(0.20)		3.52(0.21)	
Strategy Cluster 2										
Lower strategy group	56.86(5.18)	0.388	2.60(0.24)	0.343	3.15(0.24)	0.425	3.18(0.22)	0.887	3.07(0.23)	0.27
Higher strategy group	60.80(4.70)		2.80(0.22)		3.32(0.21)		3.21(0.20)		3.30(0.21)	

Abbreviations: LS, least squares; SE, standard error

*adjusted for age, gender, education, income, religion residence, marriage, cancer type, cancer stage, treatment stage

** adjusted for age, gender, education, income, religion residence, marriage, cancer type, treatment stage

3.3.2. The association of self-management clustering with HRQoL in advanced cancer patients

In advanced cancer patients, higher-strategy group showed significantly greater global QoL, existential well-being, and social support in McGill QoL in both strategy clusters 1 and 2. Only higher-strategy group in strategy cluster 2 showed significantly lower depression. Emotional and physical functioning did not differ between higher-strategy and lower-strategy groups in both strategy clusters 1 and 2 (Table 6).

Table 6. Comparison of adjusted LS mean scores between higher-strategy groups and lower-strategy groups according to strategy clusters in advanced cancer patients

Groups	Global Quality of Life lsmean (SE)	*P*	Emotional Functioning lsmean (SE)	*P*	Physical Functioning lsmean (SE)	*P*
Strategy Cluster 1*						
Lower strategy group	49.39 (3.73)	0.04	72.72 (3.65)	0.64	75.56 (3.10)	0.77
Higher strategy group	57.36 (4.07)		74.64 (3.42)		76.60 (2.90)	
Strategy Cluster 2*						
Lower strategy group	49.81 (3.57)	0.02	72.40 (3.29)	0.39	74.63 (2.79)	0.27
Higher strategy group	59.52 (4.47)		76.15 (3.99)		78.74 (3.39)	

Groups	McGill QoL Existential Well-being lsmean (SE)	*P*	McGill QoL Social Support lsmean (SE)	*P*	Depression (PHQ-9) lsmean (SE)	*P*
Strategy Cluster 1*						
Lower strategy group	5.30 (0.32)	0.01	6.41 (0.34)	0.03	6.86 (1.12)	0.63
Higher strategy group	6.14 (0.35)		7.20 (0.37)		7.21 (1.00)	
Strategy Cluster 2*						
Lower strategy group	5.23 (0.30)	0.0001	6.19 (0.31)	<.0001	7.76 (0.98)	0.04
Higher strategy group	6.62 (0.37)		8.04 (0.38)		5.70 (1.19)	

Abbreviations: LS, least squares; SE, standard error; *P*, p-value
*Adjusted for age, sex, education, income, religion, residence, marriage, ECOG performance status, tumor site

3.4. HRQoL prediction model development and validation

3.4.1. Participants and Features

In total, 256 cancer survivors were included in the dataset after the pre-processing. Table 7-9 show the average or prevalence of participants' socio-demographic and clinical and SAT features. The average age of the participants was 51.28 years; 62.1% had an undergraduate or higher level of degree; 60.2% lived in a metropolitan area; 85.5% were married; 63.3% were religious; 36.3% had breast cancer, 27% had lung cancer, 22.3% had colon cancer, 6.6% had gastric cancer, and 7.8% had other cancers; 53.9% had stage Ⅲ or Ⅳ cancers; and 66.4% completed treatment more than 5 years ago [101].

Table 7. Socio−demographic and clinical features in the dataset

Features	Category	N (%) or Mean ± SD	Min−Max
Age		51.28 ± 9.64	23−77
Sex	Male	91 (35.5)	
	Female	165 (64.5)	
Income (won)	< 4,000,000	125 (48.8)	
	≥4,000,000	131 (51.2)	
Education	≤High school graduates	97 (37.9)	
	≥University graduates	159 (62.1)	
Residence	Other areas	102 (39.8)	
	Metropolitan area	154 (60.2)	
Marriage	Single	37 (14.5)	
	Married	219 (85.5)	
Religion	No	94 (36.7)	
	Yes	162 (63.3)	
Cancer type	Breast cancer	93 (36.3)	
	Lung cancer	69 (27)	
	Colon cancer	57 (22.3)	
	Gastric cancer	17 (6.6)	
	etc.	20 (7.8)	
Cancer stage	I, II	196(76.6)	
	III, IV	60(23.4)	
Treatment stage	≤5 years after treatment	86 (33.6)	
	>5 years after treatment	170 (66.4)	

SD, standard deviation

Table 8. Smart management strategies for health Assessment Tool (SAT) features in the dataset

Features	Mean ± SD	Min−Max
Baseline		
SAT−C		
Proactive problem−solving strategy	64.1 ± 18.93	10−100
Positive−reframing strategy	68.42 ± 21.44	14.8−100
Creating empowered relationship strategy	77.06 ± 18.73	11.1−100
Experience−sharing strategy	48.78 ± 25.19	0−100
SAT−P		
Goal and action setting	48.68 ± 21.19	0−100
Rational decision−making strategy	63.54 ± 17.74	22.2−100
Healthy environment−creating strategy	57.32 ± 19.81	0−100
Priority−based planning strategy	54.69 ± 21.22	0−100
Life value−pursuing strategy	61.12 ± 20.68	6.7−100
SAT−I		
Self−motivating strategy	57.63 ± 19.49	4.8−100
Self−implementing strategy	54.13 ± 21.58	0−100
Reflecting strategy	44.79 ± 27.16	0−100
Energy−conserving strategy	58.42 ± 19.27	0−100
Activity−coping strategy	57.58 ± 22.70	0−100
Change		
SAT−C		
Proactive problem−solving strategy	−1.18 ± 16.84	−100−36.7
Positive−reframing strategy	−1.35 ± 18.03	−91.6−51.9
Creating empowered relationship strategy	−4.32 ± 17.38	−94.4−38.9
Experience−sharing strategy	−1.74 ± 24.28	−88.9−66.7
SAT−P		
Goal and action setting	0.13 ± 19.62	−90−53.3
Rational decision−making strategy	−3.08 ± 17.10	−88.9−55.6
Healthy environment−creating strategy	−0.62 ±18.09	−60−46.7
Priority−based planning strategy	−1.04 ± 19.83	−75−58.3
Life value−pursuing strategy	−3.33 ± 18.82	−73.3−46.7
SAT−I		
Self−motivating strategy	−1.10 ± 17.99	−76.2−61.9
Self−implementing strategy	−0.42 ± 19.79	−75−58.3
Reflecting strategy	−3.32 ± 27.98	−100−100
Energy−conserving strategy	−0.95 ± 20.78	−77.8−20.78
Activity−coping strategy	−0.31 ± 21.38	−86.7−73.3

SD, standard deviation

The outcome features are described in Table 6. Considering the primary outcome, 88.67% showed high global QoL. Regarding secondary outcomes, 66.41% showed good physical health status, 82.03% showed good emotional health status, 88.67% showed good social health status, and 75.39% showed good spiritual health status [101].

Table 9. Outcome features in the dataset

Features	Category	N (%)
Primary outcome		
Global Quality of Life	High	227 (88.67)
	Low	29 (11.33)
Secondary outcomes		
Physical health status	Good	170 (66.41)
	Bad	86 (33.59)
Mental health status	Good	210 (82.03)
	Bad	46 (17.97)
Social health status	Good	227 (88.67)
	Bad	29 (11.33)
Spiritual health status	Good	193 (75.39)
	Bad	63 (24.61)

3.4.2. Prediction Model Performance

Table 10 showed the prediction performance comparison results among XGBoost and other ensemble algorithms for the global QoL (primary outcome). XGBoost showed the best performance for the main performance measures in this study such as AUROC and AUPRC. XGBoost' s accuracy and F1 scores showed the second-best results. The AUROC of XGBoost yielded 0.80 (95% CI, 0.78 to 0.82) and the AUPRC was 0.96 (95% CI, 0.95 to 0.97). The accuracy was 0.88 (95% CI, 0.84 to 0.92) and the F1 scores were 0.93 (95% CI, 0.91 to 0.95). The final tuned hyperparameters of selected XGBoost model for the primary outcome were explained in Table 11.

Table 10. Comparison of the prediction performance for primary outcomes in different methods

Methods	AUROC (mean, 95% CI)	AUPRC (mean, 95% CI)	Accuracy (mean, 95% CI)	F1 score (mean, 95% CI)
DT	0.63 (0.61–0.65)	0.91 (0.90–0.92)	0.88 (0.83–0.90)	0.94 (0.92–0.95)
RF	0.78 (0.76–0.80)	0.96 (0.95–0.97)	0.88 (0.87–0.90)	0.94 (0.93–0.95)
GB	0.78 (0.76–0.80)	0.95 (0.94–0.96)	0.87 (0.81–0.92)	0.93 (0.89–0.95)
XGBoost	0.80 (0.78–0.82)	0.96 (0.95–0.97)	0.88 (0.84–0.92)	0.93 (0.91–0.95)
LightGBM	0.79 (0.77–0.81)	0.96 (0.95–0.97)	0.89 (0.88–0.90)	0.94 (0.94–0.95)

AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision–recall curve; CI, confidential interval.

Table 11. The final tuned hyperparameters of the XGBoost model for predicting global QoL

Hyperparameters	Values
Number of trees (n_estimator)	50
Maxinum tree depth (max_depth)	3
Learning rate (learning_rate)	0.06
Subsample proportion (subsample)	0.89
Minimum sum of instance weight needed in a child node (min_child_weight)	2
Minimal loss to expand on a leaf node (gamma)	1.5

Table 12 showed the prediction performance comparison results for overall health statuses in XGBoost. Overall, the prediction performance for health statuses was slightly lower than the performance for the global QoL. The prediction performance for physical health status was AUROC of 0.66 (95% CI, 0.65 to 0.67), AUPRC of 0.77 (95% CI, 0.75 to 0.79), accuracy of 0.69 (95% CI, 0.57 to 0.78), and F1 score of 0.79 (95% CI, 0.72 to 0.84). The prediction performance for emotional health status was AUROC of 0.71 (95% CI, 0.69 to 0.73), AUPRC of 0.90 (95% CI, 0.89 to 0.92), accuracy of 0.83 (95% CI, 0.78–0.86), and F1 score of 0.90 (95% CI, 0.88 to 0.93). The prediction performance for social health status was AUROC of 0.77 (95% CI, 0.75–0.79), AUPRC of 0.96 (95% CI, 0.95–0.97), accuracy of 0.88 (95% CI, 0.84 to 0.92), and F1 score of 0.94 (95% CI, 0.92 to 0.96). The prediction performance for spiritual health status was AUROC of 0.75 (95% CI, 0.74 to 0.76), AUPRC of 0.87 (95% CI, 0.85 to 0.89), accuracy of 0.79 (95% CI, 0.71 to 0.88), and F1 score of 0.87 (95% CI, 0.81 to 0.92).

Table 12. The prediction performance for health statuses' outcomes in the XGBoost model

Outcomes	AUROC (mean, 95% CI)	AUPRC (mean, 95% CI)	Accuracy (mean, 95% CI)	F1 score (mean, 95% CI)
Physical HS	0.66 (0.65−0.67)	0.77 (0.75−0.79)	0.69 (0.57−0.78)	0.79 (0.72−0.84)
Mental HS	0.71 (0.69−0.73)	0.90 (0.89−0.92)	0.83 (0.78−0.86)	0.90 (0.88−0.93)
Social HS	0.77 (0.75−0.79)	0.96 (0.95−0.97)	0.88 (0.84−0.92)	0.94 (0.92−0.96)
Spiritual HS	0.75 (0.74−0.76)	0.87 (0.85−0.89)	0.79 (0.71−0.88)	0.87 (0.81−0.92)

AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision−recall curve; CI, confidential interval.

Figure 5-7 showed the predictive performance of AUROC and AUPRC. Figure 5 showed the AUROC and AUPRC for predicting the global QoL in the XGBoost model. Figure 6 showed the AUROC and AUPRC for predicting overall health statuses in the XGBoost models. Figure 7 showed different algorithms' AUROC and AUPRC for predicting the global QoL as a primary outcome in XGBoost models.

(A) AUROC (B) AUPRC

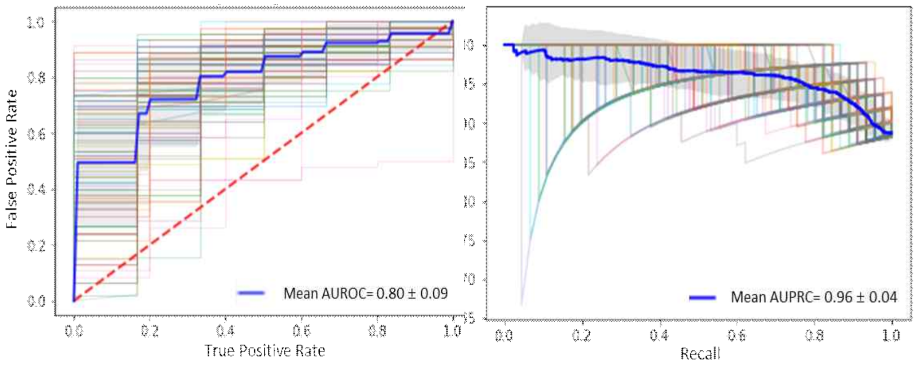


Figure 5. (A) Area under the receiver operator characteristic curve (AUROC) and (B) area under the precision-recall curve (AUPRC) in the XGBoost model for global QoL

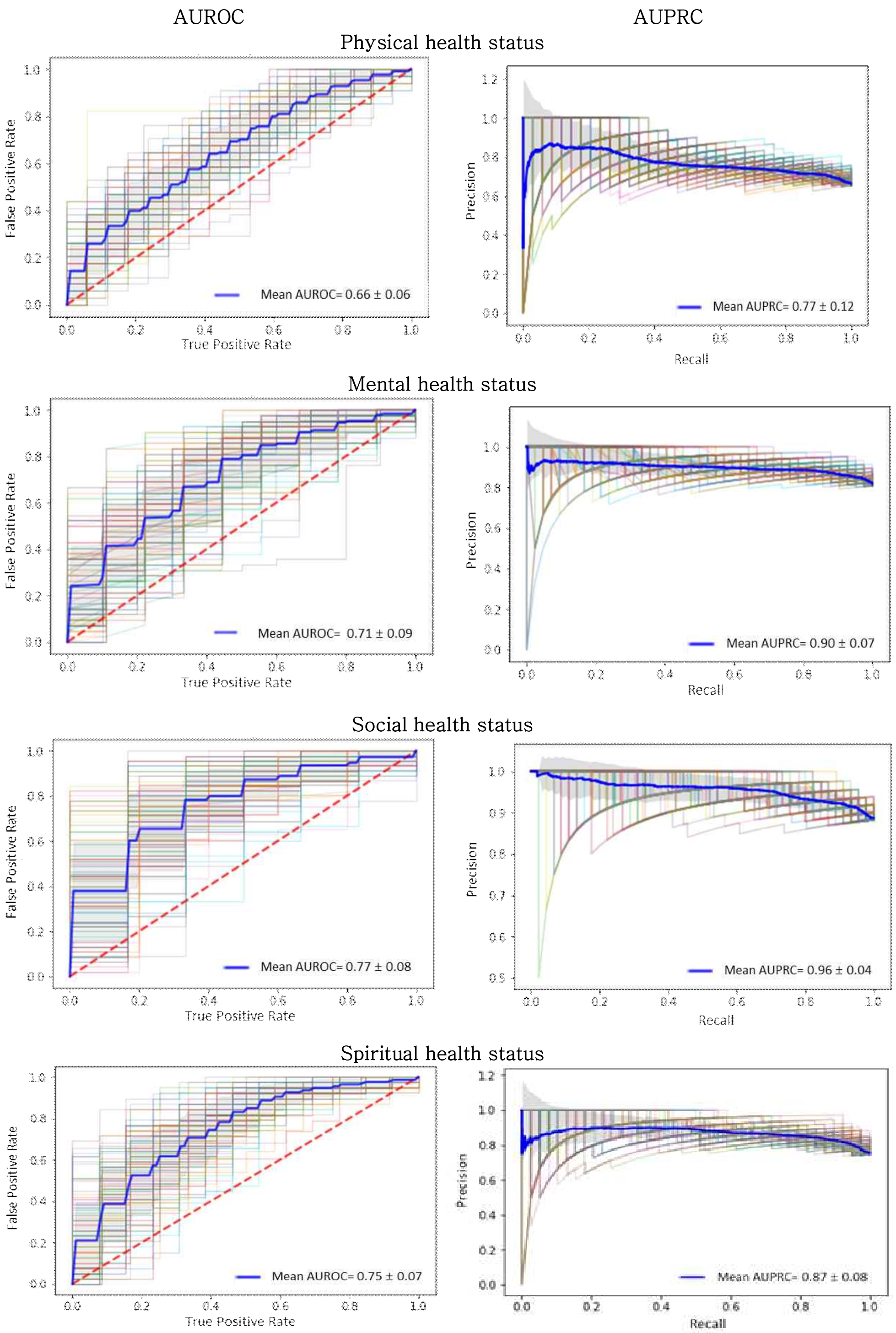


Figure 6. Area under the receiver operator characteristic curve (AUROC and area under the precision-recall curve (AUPRC) in the XGBoost model for health statuses

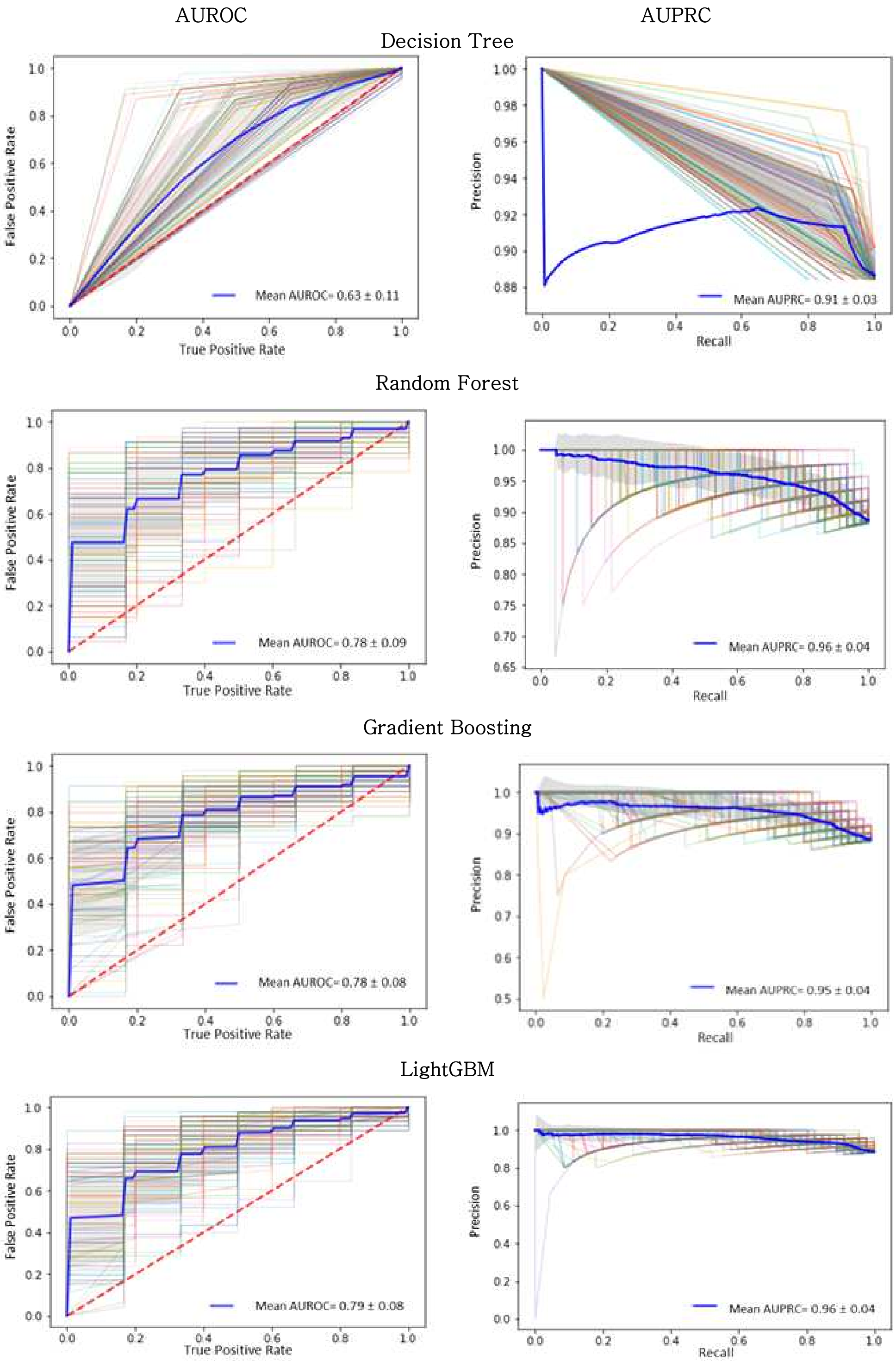
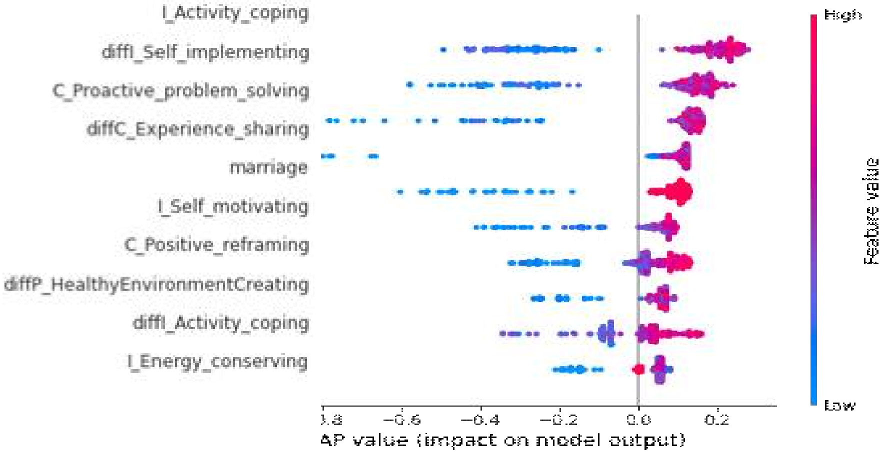


Figure 7. Area under the receiver operator characteristic curve (AUROC and area under the precision–recall curve (AUPRC) of different algorithms for predicting global QoL in the XGBoost model

3.4.3. Feature Importance

The final tuned XGBoost models were applied to extract SHAP values in the global QoL and each health status prediction model. Figure 8–9 showed the bar plots of feature importance and the beeswarm plots of the top ten critical features in each model. The top three important features for predicting the global QoL were the activity–coping strategy, change in values of the self–implementing strategy for 6 months, and the proactive problem–solving strategy [101]. The top three critical features for predicting the physical HS were the activity–coping strategy, the healthy environment–creating strategy, and age. The top three critical features for predicting the emotional HS were proactive problem–solving, positive–reframing, and rational decision–making strategies. The top three critical features for predicting the social HS were activity–coping, healthy environment–creating, and the self–motivating strategies. The top three critical features for predicting the spiritual HS were positive–reframing strategy, religion, and income. When comparing the results, the top three most important features for each outcome were different. The baseline scores of SAT were more critical than the change in scores of SAT strategy for predicting HRQoL.



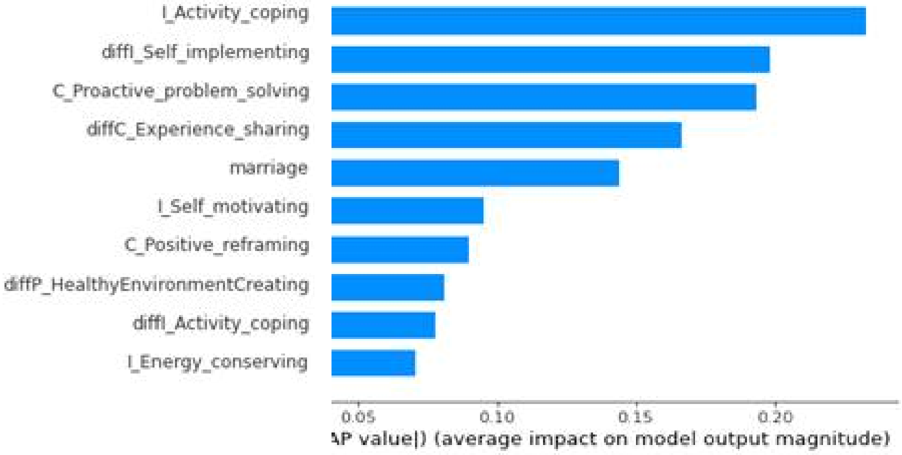
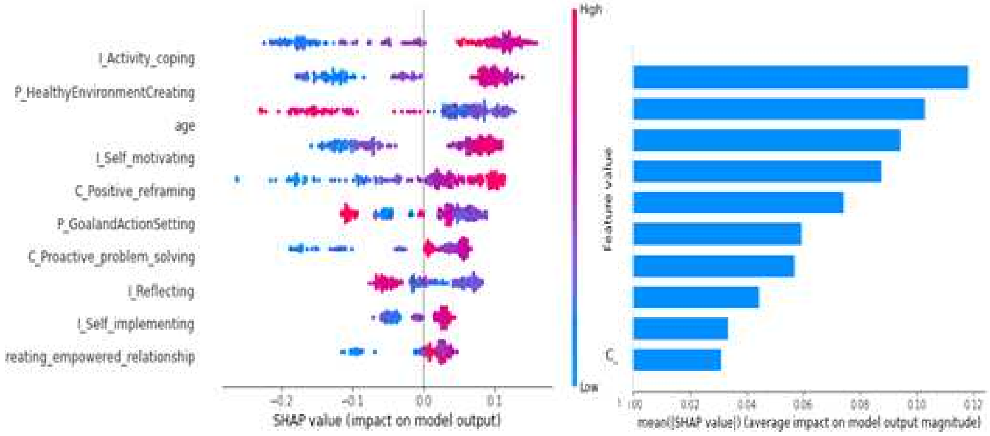
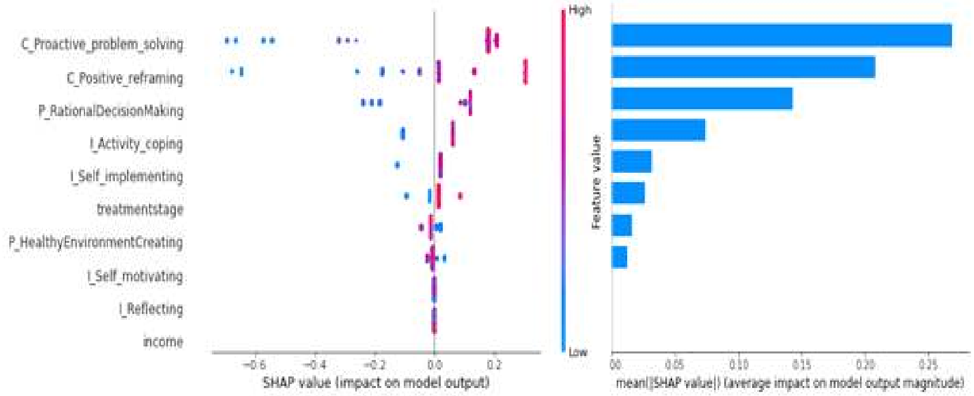


Figure 8. The feature importance beeswarm and bar plots for primary and secondary outcomes in the XGboost model for global QoL

Physical health status



Mental health status



Social health status

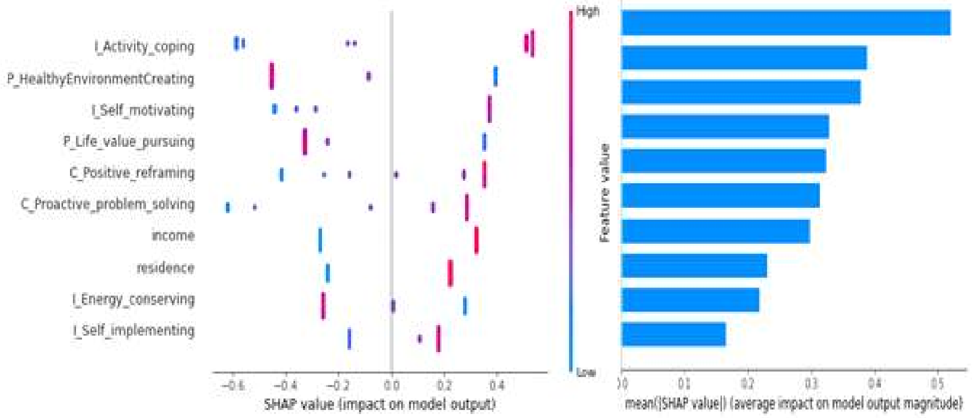


Figure 9. The feature importance beeswarm and bar plots for primary and secondary outcomes in the XGboost model for overall health statuses (continued)

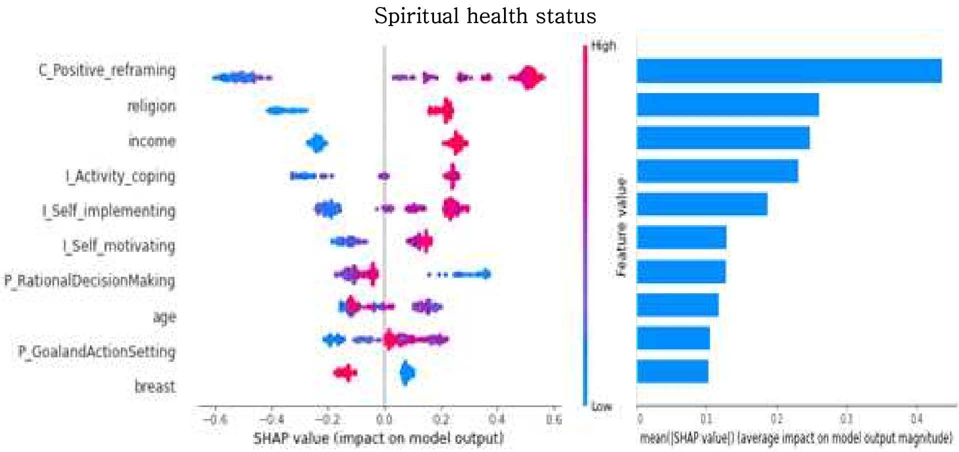


Figure 9. The feature importance beeswarm and bar plots for primary and secondary outcomes in the XGboost model for overall health statuses

3.4.3. Individual Prediction

For the individual prediction results, we selected two samples using the XGBoost model for the global QoL and identified the composition of the results. Blue arrows indicated that the features contributed to decreasing the outcome, whereas red arrows indicated that the features contributed to increasing the outcome. In figure 9, the prediction result for the global QoL was 2.07, i.e., the global QoL was high (true) for the survivor, although this survivor was living alone; there was a negative impact of low baseline scores of the rational decision-making strategy in SAT on the QoL. In figure 10, the prediction result for the global QoL was -0.56, i.e., the global QoL was low (false) for this survivor. Although there were some positive variables on the global QoL (red arrows), the low scores of the proactive problem-solving strategy in SAT-C, the activity-coping strategy in SAT-I, and the self-motivating strategy in SAT-I at baseline, and the decrease of using the activity-coping strategy in SAT-I for 6 months appeared to have more negative effects on the global QoL. Therefore, this survivor' s global QoL was low (false).

(A)

(B)

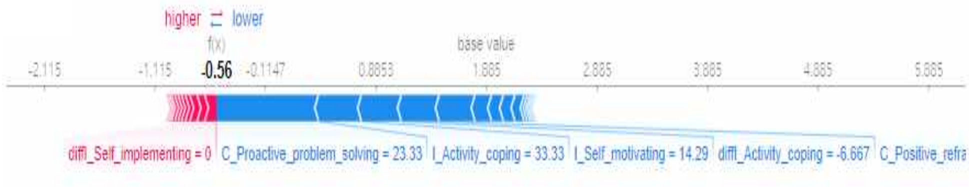


Figure 10. (A) Positive and (B) negative global QoL compositions of the individual prediction from one patient sample in cancer survivors

3.5. Survival prediction model development and validation

3.5.1. Finding critical predictors for 1−year survival in advanced cancer patients

I. Univariate and multiple proportional hazard regression analyses of socio−demographic and clinical variables

The results of univariate and multiple proportional hazard regression of socio−demographic and clinical variables with a backward selection are summarized in Table 13. The multiple proportional hazard regression comprised variables that showed statistical significance in the univariate analyses at the 0.05 level to select variables for subsequent dimensional multiple proportional hazard regression modeling. Sex, marital status, and ECOG performance status were finally selected in this multiple Cox proportional hazard model. Male sex (HR 1.762, 95% CI 1.15 to 2.70), living alone (HR 1.666, 95% CI 1.045 to 2.654), and ECOG performance status of 2 (HR 1.895, 95% CI 1.042 to 3.447) were associated with worse survival.

Table 13. Univariate and multiple proportional hazard regression analyses of socio-demographic and clinical variables with backward selection

Variables	Category	Crude HR (95% CI)	*P*	Multiple HR (95% CI)	*P*
Age (median)	≥61	1	0.64	~	
	<61	0.907 (0.599−1.373)			
Sex	Female	1	0.02	1	0.01
	Male	1.65 (1.08−2.524)		1.762 (1.150−2.700)	
Monthly income (Won)	≥3,000,000	1	0.56	~	
	<3,000,000	1.16(0.700−1.922)			
Educational Status	≥High school graduates	1	0.93	~	
	<High school graduates	0.98 (0.652−1.481)			
Residence	Metropolitan area	1	0.07	~	
	Other areas	1.49 (0.966−2.299)			
Marital status	Married	1	0.02	1	0.03
	Single	1.736 (1.112−2.712)		1.666 (1.045−2.654)	
Religion	No	1	0.12	~	
	Yes	0.710 (0.463−1.089)			
Tumor site	Breast & Gastrointestinal cancer	1	0.57	~	
	Lung, Hepatobiliary & other cancer	1.137 (0.729−1.774)			
ECOG performance status	0, 1	1	0.01	1	0.04
	2	2.132 (1.202−3.783)		1.895 (1.042−3.447)	
Treatment	Holistic Care	1	0.08	~	
	General Care	0.695 (0.460−1.049)			

HR, hazard ratio; CI, confidence interval

II. Univariate and multiple proportional hazard regression analyses of HRQoL

The multiple proportional hazard regression analyses of HRQoL showed that only depression (PHQ-9) was selected. Moderate or more severe depression was associated with poor survival (HR 2.786, 95% CI 1.818 to 4.268) (Table 14).

Table 14. Univariate and multiple proportional hazard regression analyses of Health-related Quality of Life variables with backward selection

Variable	Category		Crude HR (95% CI)	P	Multiple HR (95% CI)	P
Quality of Life (EORTC QLQ-C15 PAL)	Global quality of life	>33.33	1	0.01	NS*	
		≤33.33	1.807 (1.152-2.834)			
	Physical functioning	>33.33	1	0.53	~**	
		≤33.33	1.281 (0.593-2.771)			
	Emotional functioning	>33.33	1	0.45	~	
		≤33.33	1.276 (0.679-2.396)			
	Dyspnea	<66.66	1	0.01	NS	
		≥66.66	2.299 (1.188-4.447)			
	Pain	<66.66	1	0.01	NS	
		≥66.66	1.754 (1.132-2.717)			
	Insomnia	<66.66	1	0.2	~	
		≥66.66	1.367 (0.846-2.208)			
	Appetite loss	<66.66	1	0.003	NS	
		≥66.66	1.943 (1.257-3.003)			
	Constipation	<66.66	1	0.002	NS	
		≥66.66	2.113 (1.321-3.378)			
	Fatigue	<66.66	1	0.05	~	
		≥66.66	1.693 (0.999-2.870)			
	Nausea and vomiting	<66.66	1	0.07	~	
		≥66.66	1.717 (0.953-3.097)			
McGill Quality of life	Emotional Well-Being (33.3%)	>5	1	0.27	~	
		≤5	1.279 (0.827-1.979)			
	Social Support (33.3%)	>5.5	1	0.13	~	
		≤5.5	1.410 (0.899-2.212)			
PHQ-9	Depression	<10	1	<.0001	1	<.0001
		≥10	2.786 (1.818-4.268)		2.786 (1.818-4.268)	

HR, Hazard ratio; CI, confidence interval

* "NS" means that there was no statistically significant association in multivariate analyses but there was correlation in the univariate analyses

** "~" means that there was no correlation in univariate analyses

III. Univariate and multiple proportional hazard regression analyses of Self−Management Strategies

The univariate proportional hazard regression analyses of self−management strategies showed no statistically significant difference in survival in SAT variables (Table 15).

Table 15. Univariate proportional hazard regression analyses of Self−management strategies

Variable	Baseline Category		Crude HR (95% CI)	P−value
SAT Cluster	Strategy Cluster 1			
	Lower strategy group	≥66.66	1	0.91
	Higher strategy group	<66.66	0.972 (0.602−1.571)	
	Strategy Cluster 2			
	Lower strategy group	≥66.66	1	0.63
	Higher strategy group	<66.66	0.863 (0.470−1.584)	
Self−management strategy (SAT)	SAT−C Total	≥66.66	1	0.41
		<66.66	1.195 (0.784−1.822)	
	SAT−P Total	≥66.66	1	0.88
		<66.66	1.041 (0.607−1.785)	
	SAT−I Total	≥66.66	1	0.84
		<66.66	0.934 (0.484−1.802)	
	SAT Total	≥66.66	1	0.94
		<66.66	0.938 (0.554−1.590)	

HR, Hazard ratio; CI, confidence interval

IV. Univariate and multiple proportional hazard regression analyses of HRQOL's Clinically meaningful change

The multiple proportional hazard regression analyses of HRQoL’s clinically meaningful change showed that only “Emotional functioning change” was selected. No clinically meaningful change in emotional functioning was associated with poor survival (HR 1.880, 95% CI 1.045 to 3.384) (Table 16).

Table 16. Univariate and multiple proportional hazard regression analyses of HRQoL variables' clinically meaningful changes with backward selection

Variable	Change Category		Crude HR (95% CI)	*P*	Multiple HR (95% CI)	*P*
Quality of Life (EORTC QLQ-C 15 PAL)	Global quality of life	≥10.42	1	0.16	~**	
		<10.42	1.484 (0.853-2.582)			
	Physical functioning	≥10.89	1	0.31	~	
		<10.89	1.359 (0.756-2.444)			
	Emotional functioning	≥11.79	1	0.04	1	0.04
		<11.79	1.880 (1.045-3.384)		1.880 (1.045-3.384)	
	Dyspnea	<12.63	1	0.44	~	
		≥12.63	1.272 (0.693-2.335)			
	Pain	<14.34	1	0.48	~	
		≥14.34	1.222 (0.703-2.125)			
	Insomnia	<14.40	1	0.36	~	
		≥14.40	1.438 (0.665-3.110)			
	Appetite loss	<14.27	1	0.13	~	
		≥14.27	1.693 (0.850-3.371)			
	Constipation	<14.40	1	0.43	~	
		≥14.40	1.360 (0.629-2.940)			
	Fatigue	<11.03	1	0.68	~	
		≥11.03	0.892 (0.513-1.551)			
	Nausea and vomiting	<12.25	1	0.48	~	
		≥12.25	1.280 (0.643-2.546)			
McGill Quality of life	Emotional Well-Being (33.3%)	≥0.91	1	0.06	~	
		<0.91	1.753 (0.974-3.154)			
	Social Support (33.3%)	≥0.99	1	0.04	NS*	
		<0.99	1.825 (1.015-3.284)			
PHQ-9	Depression	<2.79	1	0.93	~	
		≥2.79	1.028 (0.547-1.932)			

HR, Hazard ratio; CI, confidence interval
* "NS" means that there was no statistically significant association in multivariate analyses but there was correlation in the univariate analyses
** "~" means that there was no correlation in univariate analyses

V. Univariate and multiple proportional hazard regression analyses of Self-management strategies' clinically meaningful change

The univariate proportional hazard regression analyses of self-management strategies' clinically meaningful change showed no statistically significant difference survival according to the SAT's clinically meaningful change (Table 17).

Table 17. Univariate proportional hazard regression analyses of Self-management strategies' clinically meaningful change

Clinically Meaningful Change Category		Crude HR (95% CI)	*P*
SAT Cluster	Strategy Cluster 1		
	Lower change group	1	0.78
	Higher change group	1.096 (0.583-2.058)	
	Strategy Cluster 2		
	Lower change group	1	0.63
	Higher change group	1.159 (0.631-2.128)	
Self-management strategies (SAT)	SAT-C Total	1	0.7
		0.878 (0.455-1.693)	
	SAT-P Total	1	0.43
		0.767 (0.397-1.479)	
	SAT-I Total	1	0.27
		0.692 (0.359-1.334)	
	SAT Total	1	0.75
		0.897 (0.465-1.732)	

HR, Hazard ratio; CI, confidence interval

VI. Final Dimensional multiple proportional hazard regression analyses

The final dimensional multiple proportional hazard regression analyses of the selected variables in multiple proportional hazard regression analysis of socio-demographic and clinical, HRQoL, SAT strategy, and the clinically meaningful change of HRQoL and SAT strategy with backward selection by intention to treat showed that "ECOG performance status," "Emotional functioning change," and "Depression" were the selected variables predicting survival. ECOG performance status of 2 (HR 1.894, 95% CI 1.030 to 3.481), no clinically meaningful positive change in emotional functioning (HR 2.322, 95% CI 1.278 to 4.219), and moderate or more severe depression (HR 3.042, 95% CI 1.956-4.729) were associated with worse survival (Table 18).

Table 18. Dimensional Multiple proportional hazard regression analyses of socio-demographic, clinical, HRQoL, and self-management strategies with backward selection

Measure	Variables	Category	Adjusted HR (95% CI)	P-value
Socio-demographic and clinical variables	ECOG performance status	0, 1	1	0.04
		2	1.894 (1.030-3.481)	
Quality of Life Change	Emotional Functioning Change	≥11.79	1	0.01
		<11.79	2.322 (1.278-4.219)	
PHQ-9	Depression	<10	1	<.0001
		≥10	3.042 (1.956-4.729)	

HR, Hazard ratio; CI, confidence interval

VII. Sensitivity analyses of dimensional multiple proportional hazard regression analyses

In the first sensitivity analysis using an average imputation method, the final dimensional multiple proportional hazard regression result showed that "ECOG performance status," "Emotional functioning change," and "Depression" were the finally selected variables predicting survival; this result was the same as that of intention to treat analysis. ECOG performance status of 2 (HR 1.894, 95% CI 1.030 to 3.481), no clinically meaningful positive change in emotional functioning (HR 2.322, 95% CI 1.278 to 4.219), and moderate or more severe depression (HR 3.042, 95% CI 1.956 to 4.729) were associated with poor survival (Table 19).

Table 19. Dimensional Multiple proportional hazard regression analyses of socio-demographic, clinical, HRQoL, and self-management strategies with backward selection (in average imputation)

Measure	Variables	Category	Adjusted HR (95% CI)	P-value
Socio-demographic and clinical variables	ECOG performance status	0, 1	1	0.04
		2	1.894 (1.030-3.481)	
Quality of Life Change	Emotional Functioning Change	≥11.79	1	0.01
		<11.79	2.322 (1.278-4.219)	
PHQ-9	Depression	<10	1	<.0001
		≥10	3.042 (1.956-4.729)	

HR, Hazard ratio; CI, confidence interval

In the second sensitivity analysis using complete case analysis, the final dimensional multiple proportional hazard regression result showed that "ECOG performance status" and "Depression" were the selected variables for predicting survival. ECOG performance status of 2 (HR 1.870, 95% CI 1.026 to 3.409) and moderate or more severe depression (HR 2.735, 95% CI 1.779 to 4.204) were associated with poor survival (Table 20).

Table 20. Dimensional Multiple proportional hazard regression analyses of socio-demographic, clinical, HRQoL, and self-management strategies with backward selection in complete case analysis

Measure	Variables	Category	Adjusted HR (95% CI)	P-value
Sociodemographic and clinical variables	ECOG performance status	0, 1	1	0.04
		2	1.870 (1.026-3.409)	
PHQ-9	Depression	<10	1	<.0001
		≥10	2.735 (1.779-4.204)	

HR, Hazard ratio; CI, confidence interval

3.5.2. Prediction model development and validation for survival in advanced cancer patients

I. Participants and Features

In total, 132 advanced cancer patients were involved in the dataset after the pre-processing. We removed censored data (N=12) for developing survival prediction model because confirming survival prediction performance at 1 year was the main purpose for developing the survival predictive model. Table 21 shows the average input or prevalence of input and outcome features in the dataset. The input features are those selected from the univariate analyses of conventional methods with statistical significance level ≤ 0.05. Regarding outcome, 69.7% died and 30.3% survived. Considering input features, 57.6% were male; 75.7% were married; and 86.1% had ECOG status of 2. Regarding input variables of HRQoL, the average scores of the global QoL were 54.55; those of dyspnea were 19.19; pain were 25.25; appetite loss were 30.56; constipation were 27.02; depression were 6.69; clinically meaningful emotional functioning change were 2.27; and clinically meaningful social support change in McGill QoL were -0.04.

Table 21. "Input" and "Outcome" features in the dataset

Features	Category	N (%) or Mean ± SD	Min-Max
Outcome Feature			
Outcome (target)	Survival	40 (30.3)	
	Death	92 (69.7)	
Input Features*			
Sex	Male	83 (57.6)	
	Female	61 (42.4)	
Marital status	Single	35 (24.3)	
	Married	109 (75.7)	
ECOG performance status	0, 1	20 (13.9)	
	2	124 (86.1)	
Quality of Life (EORTC QLQ-C15 PAL)	Global quality of life	54.55 (20.91)	11.11-100
	Dyspnea	19.19 (25.43)	0-100
	Pain	25.25 (27.74)	0-100
	Appetite loss	30.56 (28.55)	0-100
	Constipation	27.02 (29.45)	0-100
PHQ-9	Depression	6.69 (5.51)	0-24
Clinically meaningful change	Emotional Functioning	2.27 (13.93)	-50-66.66
	Social support	-0.04 (1.57)	-5-5.5

*Input features are "features" that were selected from the univariate analyses of conventional methods with statistical significance level ≤0.05.

II. Prediction Model Performance

A. Model 1: Survival prediction results of all selected features

Table 22 shows the survival prediction performance of XGBoost model including all selected features. The AUROC was 0.74 (0.54-0.93), AUPRC was 0.84 (0.71-0.97), accuracy was 0.72 (0.58-0.85), and F1 was 0.81 (0.73-0.90). The final tuned hyperparameters are explained in Table 23.

Table 22. Survival prediction results of selected features including features that were selected in the univariate analysis of conventional methods

Method	AUROC (mean, 95% CI)	AUPRC (mean, 95% CI)	Accuracy (mean, 95% CI)	F1 score (mean, 95% CI)
XGBoost	0.74(0.54−0.93)	0.84(0.71−0.97)	0.72(0.58−0.85)	0.81(0.73−0.90)

AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision−recall curve; CI, confidential interval.

Table 23. The final tuned hyperparameters of the XGBoost model for predicting survival in cancer patients

Hyperparameters	Values
Number of trees (n_estimator)	150
Maxinum tree depth (max_depth)	2
Learning rate (learning_rate)	0.03
Subsample proportion (subsample)	0.6
Minimum sum of instance weight needed in a child node (min_child_weight)	5
Minimal loss to expand on a leaf node (gamma)	0

B. Model 2: Survival prediction results of selected features by BorutaSHAP

Table 24 showed the survival prediction performance of XGBoost model constructing; only five features were selected by BorutaSHAP methods including depression, appetite loss, pain, constipation, and sex (Figure 11). The survival prediction performance was slightly improved compared to that in Table 22. Only five selected features predicted the survival of advanced cancer patients with good performance. The final tuned hyperparameters of BorutaSHAP' s model are explained in Table 25.

Table 24. Survival prediction results of features that were selected by BorutaSHAP method

Method	AUROC (mean, 95% CI)	AUPRC (mean, 95% CI)	Accuracy (mean, 95% CI)	F1 score (mean, 95% CI)
XGBoost	0.75(0.54-0.92)	0.87(0.75-0.97)	0.72(0.56-0.85)	0.82(0.71-0.90)

AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; CI, confidential interval.

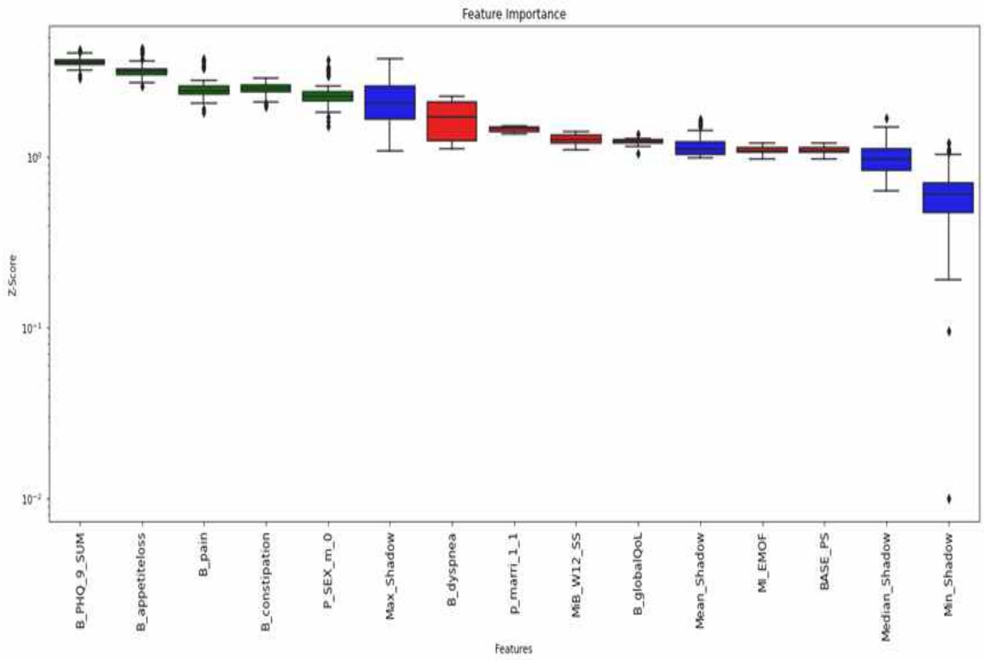


Figure 11. BorutaSHAP's selection of potential features. Green means the selected features.

Table 25. The final tuned hyperparameters of the XGBoost model for predicting survival in cancer patients

Hyperparameters	Values
Number of trees (n_estimator)	150
Maxinum tree depth (max_depth)	2
Learning rate (learning_rate)	0.03
Subsample proportion (subsample)	0.6
Minimum sum of instance weight needed in a child node (min_child_weight)	5
Minimal loss to expand on a leaf node (gamma)	0

C. Model 3: Survival prediction results of combined selected features by BorutaSHAP and Dimensional multiple Cox proportional hazard regression model

Table 26 shows the survival prediction performance of XGBoost model constructing; only seven features were selected by BorutaSHAP methods and Dimensional multiple Cox proportional hazard regression model including depression, appetite loss, pain, constipation, sex, emotional functioning change, and ECOG performance status. The survival prediction performance was slightly improved in the AUROC and the accuracy and was the same as that in the AUPRC; F1 score was compared to the predictive model with selected feature by only BorutaSHAP. The final tuned hyperparameters of the combined model are explained in Table 27. Comprehensively, the survival prediction results were the best in the model combined with features selected by both MLT and the conventional method.

Table 26. Survival prediction results of combined selected features by BorutaSHAP in MLT and the conventional method

Method	AUROC (mean, 95% CI)	AUPRC (mean, 95% CI)	Accuracy (mean, 95% CI)	F1 score (mean, 95% CI)
XGBoost	0.76(0.60-0.93)	0.87(0.74-0.97)	0.73(0.58-0.85)	0.82(0.72-0.90)

AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; CI, confidential interval.

Table 27. The final tuned hyperparameters of the combined model for predicting survival in cancer patients

Hyperparameters	Values
Number of trees (n_estimator)	100
Maxinum tree depth (max_depth)	2
Learning rate (learning_rate)	0.04
Subsample proportion (subsample)	0.5
Minimum sum of instance weight needed in a child node (min_child_weight)	5
Minimal loss to expand on a leaf node (gamma)	3

3.5.3. Individual predictions

For individual prediction results in advanced cancer patients, we selected two samples using model 3 and identified composition of the results. In figure 12 (A), the prediction result for 1−year survival was 2.12, i.e., death was high (true) for the advanced cancer patients because of sex (male=0), pain, constipation, appetite loss, and moderate depression. In figure 12 (B), the prediction result for 1−year survival was -0.23, i.e., death was low (false) for this patient. Although sex (male=0) increased the incidence of death, no appetite loss, depression, constipation, and pain ultimately reduced death occurrence. Therefore, this advanced cancer patient survived.

(A)

(B)

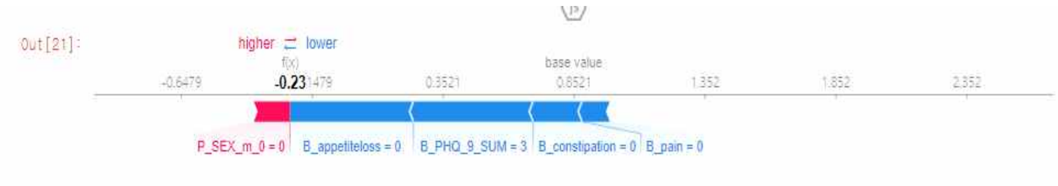


Figure 12. (A) Positive and (B) negative survival compositions of the individual prediction from one patient sample in advanced cancer patients

3.6. Causal relationship among SAT, HRQoL, and Survival

3.6.1. The relationship between SAT, HRQoL (depression or emotional functioning change), and survival in the subgroup analyses

Table 28-29 shows the results of adjusted proportional hazard regression of SAT among the higher and lower depression group or higher clinically meaningful emotional functioning change group and lower change group adjusting sex, marital status, and ECOG performance status that were significantly correlated with survival in the adjusted proportional hazard regression model with backward selection. In Table 28, there was a statistically significant difference between the higher and lower SAT group only in the higher depression group. Among advanced cancer patients with higher depression, the risk of death was reduced with increasing use of the SAT strategy. However, there was no difference in the subgroup analysis of the emotional functioning change groups. Additionally, there was no difference in the subgroup analyses of clinically meaningful SAT change between higher and lower depression groups and higher and lower emotional functioning change groups (Table 29). We found only an association between SAT, depression, and survival in advanced cancer patients.

Table 28. Subgroup analyses of SAT between higher vs. lower Depression and Emotional Functioning Change groups

Depression Subgroup		Variables		Crude HR (95% CI)	*P*
Depression	Higher group[a] (n=46)	SAT Total	<66.66	1	0.047
			≥66.66	0.21 (0.04−0.98)	
	Lower group (n=98)		<66.66	1	0.48
			≥66.66	0.81 (0.44−1.48)	
Emotional Functioning Change Subgroup		**Variables**		**Crude HR (95% CI)**	***P***
Emotional Functioning Change	Higher change group(n=24)	SAT Total	<66.66	1	0.42
			≥66.66	2.31 (0.30−17.77)	
	Lower change group (n=120)		<66.66	1	0.7
			≥66.66	0.90 (0.52−1.56)	

Abbreviation: HR, Hazard ratio; CI, confidence interval; *P*, p-value; SAT, Smart management strategies for health Assessment Tool.
[a] Higher group was categorized based on the baseline scores of PHQ-9 greater than 10.

Table 29. Subgroup analyses of clinically meaningful SAT change between higher vs. lower Depression and Emotional Functioning Change groups

Depression Subgroup		SAT clinically meaningful change group		Crude HR (95% CI)	*P*
Depression	Higher group (n=46)	SAT Total	Clinically meaningful change group [a]	1	0.72
				1.30 (0.31−5.48)	
	Lower group (n=98)		No clinically meaningful change group	1	0.77
				0.90 (0.42−1.89)	
Emotional Functioning Change Subgroup		**SAT clinically meaningful change group**		**Crude HR (95% CI)**	***P***
Emotional Functioning Change	Higher change group (n=24)	SAT Total	Clinically meaningful change group [a]	1	0.19
				0.45 (0.14−1.54)	
	Lower change group (n=120		No clinically meaningful change group	1	0.51
				1.32 (0.57−3.03)	

Abbreviation: HR, Hazard ratio; CI, confidence interval; *P*, p-value; SAT, Smart management strategies for health Assessment Tool.
[a] Clinically meaningful change group is the group showing clinically meaningful change based on an effect size more than 0.5 for 12 weeks. The other was categorized into no clinically meaningful change group.

3.6.2. The causal relationship among SAT, Depression, and Survival in a path analysis

To examine the causal relationship among SAT, depression, and survival, as we hypothesized, we conducted a path analysis (Figure 13). SAT had a significant effect on depression (B=−0.21, p<0.05) but not on survival, and depression had a significant effect on survival (B=0.31, p<0.001) (Table 30). Table 31 shows results of verifying the significance of indirect effects through bootstrapping to investigate the mediating effect of depression regarding the relationship among SAT strategy, depression, and survival in advanced cancer patients. From the analysis, a causal relationship showing that SAT affects the survival of advanced cancer patients through depression.

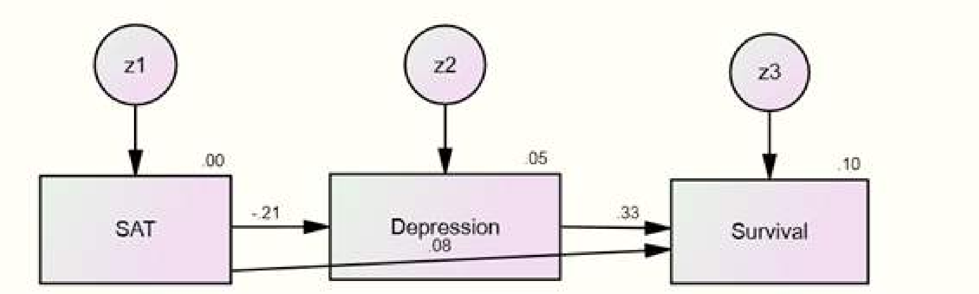


Figure 13. The results of path analysis

Table 30. The path relationship between SAT, depression, and survival

Regressions	B	ß	S.E.	C.R.
Depression ← SAT	−0.21	−0.25	0.10	−2.50*
Death (Survival) ← SAT	0.08	0.10	0.10	0.96
Death (Survival) ← Depression	0.31	0.32	0.08	3.73***

* p<.05, ** p<.01, *** p<.001

Table 31. Bootstrapping results for the mediating effect of depression in the relationship between SAT, depression, and survival

Independent Variable	Mediation Variable	Dependent Variable	B	P
SAT	Depression	Death (Survival)	−0.066	0.005**

* p<.05, ** p<.01, *** p<.001

Chapter 4. Discussion

This study is the first to analyze and verify differences in the actual usage patterns of self-management strategies, including cancer survivors and advanced cancer patients in the cancer care continuum. This study also first found simple predictive models to predict HRQoL using only self-management strategies and personal characteristics for cancer survivors and only HRQoL and personal characteristics for advanced cancer patients with good performance, interpreted the predictive model, and suggested usage of the results using machine learning techniques in clinical practice. Moreover, this study first identified the causal relationship between self-management strategies, HRQoL (depression in this study), and survival in advanced cancer patients using path analysis in structural equation modeling.

The strategy clusters of early-stage cancer patients were different from those of advanced-stage cancer patients, and this pattern of using self-management strategies was like that of advanced cancer patients. Interestingly, the core strategies (SAT-C) showed a tendency to be bundled together with the preparation (SAT-P) and the implementation (SAT-I) strategies, compatible with the study hypotheses. This tendency became clearer as all the sub-strategies of the SAT-C were tied to SAT-P and I in advanced cancer patients. Advanced-stage cancer patients may experience more physical, emotional, social, and spiritual hardships when managing their health in comparison with early-stage cancer patients [19, 102, 103]. Thus, to overcome the crisis and improve their HRQoL, advanced-stage cancer patients may use SAT-C strategies more proactively together with the SAT-P and SAT-I strategies, because the SAT-C strategies are effective to overcome crisis regardless of the treatment or health behavior stages, which suit the name of "Core" [31]. In a previous study, patients with cancer showed highly individualized approaches to integrate self-management strategies into their lives for overcoming a crisis [16]. This study indicated that cancer patients' self-management strategies may differ according to the cancer stage. Advanced cancer patients in our study might use more SAT-C strategies to integrate self-management strategies for overcoming a crisis.

Although cancer patients' HRQoL was better with increasing use of self-management strategies, the results differed slightly considering the cancer stages. For early-stage cancer patients (survivors in this study), the QoL and physical, emotional, social, and spiritual health statuses were better after using the cluster including most SAT-C strategies than the other cluster, which included most strategies of SAT-P and SAT-I. The cluster with most SAT-C strategies better differentiated all health statuses of early-stage cancer patients after 6 months than the other cluster including most strategies of supporting preparation for action or the implementation of health behaviors. Thus, the core strategies can contribute to increasing HRQoL such as the general quality of life and even health statuses (physical, emotional, social, and spiritual) for cancer survivors. For advanced-stage cancer survivors, the self-management strategies have no impact on the global QoL after

6 months regardless of cluster types. Most advanced-stage cancer survivors may experience more severe symptoms and poorer QoL than early-stage cancer survivors [16, 19]. Thus, advanced-stage cancer survivors may have poor QoL regardless of whether the self-management strategies are used more often. However, for advanced cancer patients, the strategy clusters have a positive impact on global QoL as well as existential well-being, social support, and depression. The difference in results may be due to the timing of QoL measurement: HRQoL after 6 months was measured for cancer survivors but that at baseline was measured for advanced cancer patients. There was a strong relationship of the strategy clusters with HRQoL in advanced cancer patients except for physical and emotional functioning. These results suggest that even advanced cancer patients can maintain high QoL if they use self-management strategies more proactively.

This study developed and validated various prediction models for cancer survivors' HRQoL and advanced cancer patients' survival. In cancer survivors, the HRQoL was highly predictable only with personal characteristics such as socio-demographic and clinical variables and the self-management strategies in SAT. The development of these simple models for the prediction of HRQoL can enhance the ease of application for increasing HRQoL of cancer survivors in real clinical settings [87]. The important features for the global QoL and health statuses were different. Thus, a customized self-management strategy is necessary for cancer survivors for healthcare services considering their global QoL or overall health statuses separately. By developing a web-based survey targeting cancer survivors, this approach could provide cancer survivors with useful health information, while allowing research [87]. We developed HealthingU website to achieve this, which allowed cancer survivors to view customized survey results regarding their health immediately after they completed the web-based survey. We also found a method to provide cancer survivors with truly personalized healthcare information using a newly developed XAI technology named SHAP. The individual prediction of SHAP is a new function, which is unavailable in the conventional statistical methods [95]. This technology can be expected to open a new horizon in supplying cancer survivors with personalized healthcare services.

The predictive modeling results of survival in advanced cancer patients were slightly different between conventional statistical methods and machine learning techniques. This might be reasonable result due to the difference between input variables (traditional statistical methods used binary variables; MLT used continuous variables) and calculation methods. However, depression at baseline was the most important predictor of survival for advanced cancer patients in both conventional and ML methods, which is consistent with the previous studies [12, 104, 105]. Although previous studies reported that depression was related with worse survival in patients with cancer, depression treatment in advanced cancer patients is unclear [12]. This study's result suggested a possible way to treat depression effectively by performing subgroup analyses, showing the causal relationship among SAT, depression, and survival in advanced cancer patients, and determining the critical SAT strategies associated with depression. These results consistently showed that SAT strategies affected depression directly, which significantly affected survival. Among SAT strategies, the self-implementing and self-sustaining strategies in SAT-I significantly affected depression, consistent with previous studies' results showing that adjusting the life's goal, taking action, and finding acceptance were related to less depression and better QoL for advanced cancer patients [34, 55, 106] (Supplementary Information 2). The self-implementing and the self-sustaining strategies included the item-level strategies as follows: "I used a concrete method (memo, alarm function) to remind myself of my to-do list (the self-implementing strategy)," "I used my given time carefully," "I adjusted my work and life according to my body rhythm," "I tried to do align my actions with my plan," "I sometimes set my work aside and took time to re-examine the purpose of life." According to the result, treating depression for advanced cancer patients nearing death and individual behavioral implementing strategies may be more helpful than cognitive, emotional, and social strategies as the core strategies in SAT.

Previous studies reported both benefits and harmful effects of emotional functioning for survival in cancer patients [107-109]. However, this study's result suggested that we can expect improved survival if there are clinically meaningful changes in emotional functioning even in advanced cancer patients. However, there was no causal association among SAT, emotional functioning change, and survival. Thus, emotional functioning change cannot mediate SAT and survival like depression; however, it significantly affects survival among advanced cancer patients; thus, it is necessary to monitor any noteworthy change in emotional functioning after advanced cancer diagnosis in cancer patients.

HRQoL variables predicted survival more effectively than the SAT strategies, consistent with the following study hypothesis: higher depression, pain, appetite loss, and constipation, male in sex, worse ECOG performance, and no clinically meaningful change of emotional functioning were the critical predictors of worse survival considering conventional statistical method and MLT comprehensively. In previous studies, physical symptoms, depression, and male sex were prognostic factors of survival in cancer patients [12, 104, 105, 110, 111]. This study suggests that both conventional statistical methods and MLT should be considered to find critical prognostic factors of cancer survivorship because the factors for survival of cancer patients may differ slightly depending on whether variables are processed or analyzed. In this study, the prediction power of survival was the highest when including selected variables using both conventional statistical methods and MLT.

Despite new findings, this study had several limitations. First, the modest sample size of cancer survivors and advanced cancer patients prevented us from generalizing our results. These study results should be evaluated with a larger sample of cancer survivors or advanced cancer patients to examine whether they are reproducible. To prevent the problems related to the small sample size, we performed hyper-parameter tuning and repeated K-fold methods to alleviate overfitting when constructing prediction modeling [68, 112]. Moreover, we conducted the bootstrapping validation to evaluate XGBoost models for cancer survivors and advanced cancer patients [94]. The predictive performances of the XGBoost models after bootstrapping showed similar results to those of the repeated K-fold method; the primary outcome's performance of the XGBoost model was AUROC = 0.77, AUPRC = 0.97 in cancer survivors, and the survival performance of XGBoost model was AUROC = 0.81, AUPRC = 0.91 in advanced cancer patients (Supplementary Information 3-4). The performance of the survival prediction model after bootstrapping in advanced cancer patients was improved compared with the performance of the repeated K-fold method. Based on the bootstrapping results, it is expected that similar results will be derived when they are verified with more extensive cancer patient data. However, to increase the verification power of these predictive models, securing larger number of samples or verifying the test set with different cancer survivors or advanced cancer patients' data will be necessary. Second, we statistically identified the causal relationship of SAT strategies with HRQoL and survival by path analysis. However, future studies should examine the effect of SAT strategies on improving depression for improved survival in advanced cancer patients by using a verified method like a randomized controlled trial (RCT). Third, we developed and validated 1-year survival prediction model in advanced cancer patients without censored data (n=12). In this study, we developed survival prediction model to identify the prediction power of SAT or HRQoL for 1-year survival in advanced cancer patients. However, with the recent development in technology, survival analysis including censored data has become possible with the XGBoost algorithm used in this study. In future, the survival analysis results of MLT could be compared with those of conventional statistical methods such as Cox regression.

We aimed to evaluate the association among the self-management strategies (SAT), HRQoL, and survival for cancer survivors and advanced cancer patients considering the cancer-care continuum using clustering methods and MLT. Our study results increased the understanding of using SAT strategies and the effects of the SAT strategies on HRQoL and survival considering cancer stages. Our study results presented that the SAT strategies can be effective in improving HRQoL for cancer survivors and critical predictors of survival for even advanced cancer patients undergoing depression. These results can improve the effectiveness of self-managed healthcare for patients with cancer irrespective of the cancer stages in clinical settings.

Chapter 5. Conclusion

Regarding the cancer−care continuum, we examined the self−management strategies' clustering considering cancer stages and developed the prediction models for HRQoL or survival in cancer survivors or advanced cancer patients using a clustering method and machine learning techniques.

The results of the study found that the usage pattern of the self−management strategies may differ as per cancer stages. The self−management strategies' clusters affected HRQoL positively in both cancer survivors and advanced cancer patients. The prediction model performance of HRQoL for cancer survivors or survival for advanced cancer patients showed high performance. Comprehensively, this study revealed the association of self−management strategies using SAT with HRQoL and survival in the cancer−care continuum. The SAT strategies had a strong relationship with HRQoL but were not significantly related to survival. However, this study found that the SAT strategies had indirect effect on survival though a critical HRQoL variable such as depression predicting the survival.

Thus, the SAT strategies could be an effective intervention tool to support cancer patients regardless of cancer stages to improve their HRQoL and survival though improving critical HRQoL in clinical settings.

Bibliography

1. Miller, K.D., et al., *Cancer treatment and survivorship statistics, 2016.* CA: a cancer journal for clinicians, 2016. **66**(4): p. 271-289.
2. Kang, M.J., et al., *Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2019.* Cancer Research and Treatment: Official Journal of Korean Cancer Association, 2022. **54**(2): p. 330.
3. Miller, K.D., et al., *Cancer treatment and survivorship statistics, 2019.* CA: a cancer journal for clinicians, 2019. **69**(5): p. 363-385.
4. Ferlay, J., et al., *Global cancer observatory: cancer today. International Agency for Research on Cancer.* Lyon, France, 2020.
5. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: a cancer journal for clinicians, 2021. **71**(3): p. 209-249.
6. Kang, D., et al., *Impact of objective financial burden and subjective financial distress on spiritual well-being and quality of life among working-age cancer survivors.* Supportive Care in Cancer, 2022. **30**(6): p. 4917-4926.
7. Lee, E.S., et al., *Health-related quality of life in survivors with breast cancer 1 year after diagnosis compared with the general population: a prospective cohort study.* Annals of surgery, 2011. **253**(1): p. 101-108.
8. Shin, H.W., et al., *Correlates of existential well-being and their association with health-related quality of life in breast cancer survivors compared with the general population.* Breast cancer research and treatment, 2009. **118**(1): p. 139-150.
9. Chochinov, H.M., *Depression in cancer patients.* The lancet oncology, 2001. **2**(8): p. 499-505.
10. Stark, D.P.H. and A. House, *Anxiety in cancer patients.* British journal of cancer, 2000. **83**(10): p. 1261-1267.
11. McCorkle, R., et al., *Self-management: Enabling and empowering patients living with cancer as a chronic illness.* CA: a cancer journal for clinicians, 2011. **61**(1): p. 50-62.
12. Pirl, W.F., et al., *Depression and survival in metastatic non-small-cell lung cancer: Effects of early palliative care.* Journal of Clinical Oncology, 2012. **30**(12): p. 1310.
13. Hammer, M.J., et al., *Self-management for adult patients with cancer: an integrative review.* Cancer Nursing, 2015. **38**(2): p. E10-E26.
14. Shneerson, C., et al., *Patterns of self-management practices undertaken by cancer survivors: Variations in demographic factors.* European journal of cancer care, 2015. **24**(5): p. 683-694.
15. Sieverding, M., et al., *Prevalence and frequency of*

self-management strategies among female cancer survivors: The neglected roles of social relations and conscious living. Supportive Care in Cancer, 2020. **28**(2): p. 607-616.
16. Dunne, S., et al., *Investigating the impact of self-management behaviours on quality of life and fear of recurrence in head and neck cancer survivors: A population-based survey.* Psycho-Oncology, 2019. **28**(4): p. 742-749.
17. Gotay, C.C., *The experience of cancer during early and advanced stages: The views of patients and their mates.* Social Science & Medicine, 1984. **18**(7): p. 605-613.
18. Nho, J.-H., S.R. Kim, and J.-H. Nam, *Symptom clustering and quality of life in patients with ovarian cancer undergoing chemotherapy.* European Journal of Oncology Nursing, 2017. **30**: p. 8-14.
19. Miovic, M. and S. Block, *Psychiatric disorders in advanced cancer.* Cancer, 2007. **110**(8): p. 1665-1676.
20. Bail, J.R., et al. *Psychological symptoms in advanced cancer*. in *Seminars in oncology nursing*. 2018. Elsevier.
21. Hall, S.M., et al., *Treatment for cigarette smoking among depressed mental health outpatients: a randomized clinical trial.* American journal of public health, 2006. **96**(10): p. 1808-1814.
22. Dijkstra, A., B. Conijn, and H. De Vries, *A match–mismatch test of a stage model of behaviour change in tobacco smoking.* Addiction, 2006. **101**(7): p. 1035-1043.
23. Hollis, J.F., et al., *Teen reach: outcomes from a randomized, controlled trial of a tobacco reduction program for teens seen in primary medical care.* Pediatrics, 2005. **115**(4): p. 981-989.
24. Yun, Y.H., et al., *A randomized controlled trial of physical activity, dietary habit, and distress management with the Leadership and Coaching for Health (LEACH) program for disease-free cancer survivors.* BMC cancer, 2017. **17**(1): p. 1-12.
25. Yun, Y.H., et al., *The association of self-leadership, health behaviors, and posttraumatic growth with health-related quality of life in patients with cancer.* Psycho-Oncology, 2014. **23**(12): p. 1423-1430.
26. Lee, M.K., S.Y. Park, and G.-S. Choi, *Association of self-leadership and planning with performing an exercise in patients with colorectal cancer: cross-sectional study.* Cancer Nursing, 2020. **43**(1): p. E1-E9.
27. Baumeister, R.F., K.D. Vohs, and D.M. Tice, *The strength model of self-control.* Current directions in psychological science, 2007. **16**(6): p. 351-355.
28. Bandura, A. and R. Wood, *Effect of perceived controllability and performance standards on self-regulation of complex decision making.* Journal of personality and social psychology, 1989. **56**(5): p. 805.
29. Carver, C.S., *You want to measure coping but your protocol'too long: Consider the brief cope.* International journal of behavioral

medicine, 1997. **4**(1): p. 92-100.
30. Hagan, T.L., et al., *Coping in patients with incurable lung and gastrointestinal cancers: a validation study of the brief COPE.* Journal of pain and symptom management, 2017. **53**(1): p. 131-138.
31. Yun, Y.H., et al., *Patient-reported assessment of self-management strategies of health in cancer patients: development and validation of the Smart Management Strategy for Health Assessment Tool (SAT).* Psycho-Oncology, 2015. **24**(12): p. 1723-1730.
32. Kwekkeboom, K.L. *Cancer symptom cluster management*. in *Seminars in oncology nursing*. 2016. Elsevier.
33. Kilbane, K.S., et al., *Adaptive and maladaptive coping mechanisms used by patients with esophageal Cancer after Esophagectomy.* Journal of Surgical Research, 2021. **258**: p. 1-7.
34. Nipp, R.D., et al., *Coping and prognostic awareness in patients with advanced cancer.* Journal of Clinical Oncology, 2017. **35**(22): p. 2551.
35. Vandenbroucke, T., et al., *Psychological distress and cognitive coping in pregnant women diagnosed with cancer and their partners.* Psycho-Oncology, 2017. **26**(8): p. 1215-1221.
36. Montazeri, A., *Quality of life data as prognostic indicators of survival in cancer patients: an overview of the literature from 1982 to 2008*. Health and quality of life outcomes, 2009. **7**(1): p. 1-21.
37. Kim, M.-K., et al., *Health-related quality of life and sociodemographic characteristics as prognostic indicators of long-term survival in disease-free cervical cancer survivors.* International Journal of Gynecologic Cancer, 2016. **26**(4).
38. Coates, A., F. Porzsolt, and D. Osoba, *Quality of life in oncology practice: prognostic value of EORTC QLQ-C30 scores in patients with advanced malignancy.* European journal of cancer, 1997. **33**(7): p. 1025-1030.
39. Dancey, J., et al., *Quality of life scores: an independent prognostic variable in a general population of cancer patients receiving chemotherapy.* Quality of life research, 1997. **6**(2): p. 0-0.
40. Langendijk, H., et al., *The prognostic impact of quality of life assessed with the EORTC QLQ-C30 in inoperable non-small cell lung carcinoma treated with radiotherapy.* Radiotherapy and oncology, 2000. **55**(1): p. 19-25.
41. Greer, J.A., et al., *Perceptions of health status and survival in patients with metastatic lung cancer.* Journal of pain and symptom management, 2014. **48**(4): p. 548-557.
42. Oberguggenberger, A., et al., *Health behavior and quality of life outcome in breast cancer survivors: prevalence rates and predictors.* Clinical breast cancer, 2018. **18**(1): p. 38-44.
43. Hawkes, A.L., et al., *Effects of a multiple health behavior change intervention for colorectal cancer survivors on psychosocial outcomes and quality of life: a randomized controlled trial.* Annals of Behavioral Medicine, 2014. **48**(3): p. 359-370.
44. Kates, J.M., *Treatment-related Decisional Conflict, Quality of Life, and Comorbidity in Older Adults with Cancer.* Asia-Pacific Journal of

Oncology Nursing, 2018. 5(4): p. 421-429.
45. Morrill, E.F., et al., *The interaction of post-traumatic growth and post-traumatic stress symptoms in predicting depressive symptoms and quality of life.* Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer, 2008. 17(9): p. 948-953.
46. Yang, S.-K. and Y. Ha, *Exploring the relationships between posttraumatic growth, wisdom, and quality of life in older cancer survivors.* Asian Pacific Journal of Cancer Prevention: APJCP, 2019. 20(9): p. 2667.
47. Stewart, G.L., S.H. Courtright, and C.C. Manz, *Self-leadership: A paradoxical core of organizational behavior.* Annual Review of Organizational Psychology and Organizational Behavior, 2019. 6: p. 47-67.
48. Bours, M.J., et al., *Candidate predictors of health-related quality of life of colorectal cancer survivors: a systematic review.* The oncologist, 2016. 21(4): p. 433-452.
49. Abrahams, H., et al., *The relationship of fatigue in breast cancer survivors with quality of life and factors to address in psychological interventions: a systematic review.* Clinical psychology review, 2018. 63: p. 1-11.
50. Cuthbert, C.A., et al., *Self-management interventions for cancer survivors: a systematic review and evaluation of intervention content and theories.* Psycho-Oncology, 2019. 28(11): p. 2119-2140.
51. Lee, Y.J., et al., *EORTC QLQ-C15-PAL quality of life score as a prognostic indicator of survival in patients with far advanced cancer.* Supportive Care in Cancer, 2014. 22(7): p. 1941-1948.
52. Giese-Davis, J., et al., *Decrease in depression symptoms is associated with longer survival in patients with metastatic breast cancer: a secondary analysis.* Journal of clinical oncology, 2011. 29(4): p. 413.
53. Greer, J.A., et al., *Understanding and addressing the role of coping in palliative care for patients with advanced cancer.* Journal of clinical oncology, 2020. 38(9): p. 915.
54. Nipp, R.D., et al., *The relationship between coping strategies, quality of life, and mood in patients with incurable cancer.* Cancer, 2016. 122(13): p. 2110-2116.
55. Lam, W.W., et al., *Goal adjustment influence on psychological well-being following advanced breast cancer diagnosis.* Psycho-Oncology, 2016. 25(1): p. 58-65.
56. Everitt, B.S., et al., *Cluster analysis: Wiley series in probability and statistics.* Southern Gate, Chichester, West SussexUnited Kingdom: John Wiley & Sons, 2011.
57. Violán, C., et al., *Multimorbidity patterns with K-means nonhierarchical cluster analysis.* BMC family practice, 2018. 19(1): p. 1-11.
58. Zeebaree, D.Q., et al., *Combination of K-means clustering with Genetic Algorithm: A review.* International Journal of Applied

Engineering Research, 2017. 12(24): p. 14238-14245.
59. Gallagher, D., L. Ting, and A. Palmer, *A journey into the unknown; taking the fear out of structural equation modeling with AMOS for the first-time user.* The marketing review, 2008. 8(3): p. 255-275.
60. Collier, J.E., *Applied structural equation modeling using AMOS: Basic to advanced techniques*. 2020: Routledge.
61. Beran, T.N. and C. Violato, *Structural equation modeling in medical research: a primer.* BMC research notes, 2010. 3(1): p. 1-10.
62. Kang, J., et al., *Machine learning approaches for predicting radiation therapy outcomes: a clinician's perspective.* International Journal of Radiation Oncology* Biology* Physics, 2015. 93(5): p. 1127-1135.
63. Rudin, C., *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead.* Nature Machine Intelligence, 2019. 1(5): p. 206-215.
64. Gunning, D., *Explainable artificial intelligence (XAI), Defense Advanced Research Projects Agency (DARPA), 2017*. 2017.
65. Lundberg, S.M. and S.-I. Lee, *A unified approach to interpreting model predictions.* Advances in neural information processing systems, 2017. 30.
66. Wong, C.W., et al., *Explainable tree-based predictions for unplanned 30-day readmission of patients with cancer using clinical embeddings.* JCO Clinical Cancer Informatics, 2021. 5: p. 155-167.
67. Kocak, B., et al., *Machine learning-based unenhanced CT texture analysis for predicting BAP1 mutation status of clear cell renal cell carcinomas.* Acta Radiologica, 2020. 61(6): p. 856-864.
68. Hu, C.-A., et al., *Using a machine learning approach to predict mortality in critically ill influenza patients: a cross-sectional retrospective multicentre study in Taiwan.* BMJ open, 2020. 10(2): p. e033898.
69. Kershaw, T., et al., *Coping strategies and quality of life in women with advanced breast cancer and their family caregivers.* Psychology & Health, 2004. 19(2): p. 139-155.
70. Survivorship, N.C.f.C. *Defining Cancer Survivorship: Living With, Through and Beyond a Cancer Diagnosis*. [cited September 19, 2018; Available from: https://www.canceradvocacy.org/news/defining—cancer—survivorship.
71. Park, E.R., J. Peppercorn, and A. El-Jawahri, *Shades of survivorship.* Journal of the National Comprehensive Cancer Network, 2018. 16(10): p. 1163-1165.
72. Van Leeuwen, M., et al., *Understanding the quality of life (QOL) issues in survivors of cancer: towards the development of an EORTC QOL cancer survivorship questionnaire.* Health and Quality of life Outcomes, 2018. 16(1): p. 1-15.
73. Feuerstein, M., *Defining cancer survivorship.* Journal of cancer survivorship: research and practice, 2007. 1(1): p. 5-7.
74. Society, A.C. *Advanced cancer*. Accessed 28 June 2016 2014; Available from: http://www.cancer.org/acs/groups/cid/documents/webcontent/003082—p

df.pdf
75. Institute, N.C., *Coping with advanced cancer.*, N.I.o. Department of Health and Human Services and Health., Editors. 2014.
76. Yun, Y.H., et al., *Web-based tailored education program for disease-free cancer survivors with cancer-related fatigue: a randomized controlled trial.* Journal of Clinical Oncology, 2012. **30**(12): p. 1296-1303.
77. Yun, Y.H., et al., *Validation of the Korean version of the EORTC QLQ-C30.* Quality of life research, 2004. **13**(4): p. 863-868.
78. Yun, Y.H., et al., *Employee health behaviors, self-reported health status, and association with absenteeism.* Journal of occupational and environmental medicine, 2016. **58**(9): p. 932-939.
79. Yun, Y.H., et al., *Attitudes of the general population, cancer patients, their family caregivers, and physicians toward dying and death: a nationwide survey.* Global Journal of Health Science, 2017. **9**(10): p. 201-201.
80. Yun, Y.H., et al., *Development and validation of the smart management strategy for health assessment tool-short form (SAT-SF) in cancer survivors.* Quality of Life Research, 2018. **27**(2): p. 347-354.
81. Arraras, J.I., et al., *The EORTC QLQ-C15-PAL questionnaire: validation study for Spanish bone metastases patients.* Quality of Life Research, 2014. **23**(3): p. 849-855.
82. Kim, S.H., et al., *Validation study of the Korean version of the McGill Quality of Life Questionnaire.* Palliative Medicine, 2007. **21**(5): p. 441-447.
83. Cohen, S.R., et al., *The McGill Quality of Life Questionnaire: a measure of quality of life appropriate for people with advanced disease. A preliminary study of validity and acceptability.* Palliative medicine, 1995. **9**(3): p. 207-219.
84. Kroenke, K., R.L. Spitzer, and J.B. Williams, *The PHQ-9: validity of a brief depression severity measure.* Journal of general internal medicine, 2001. **16**(9): p. 606-613.
85. Spitzer, R.L., et al., *Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study.* Jama, 1999. **282**(18): p. 1737-1744.
86. Okagbue, H.I., et al., *Unique clusters of patterns of breast cancer survivorship.* Health and Technology, 2022. **12**(2): p. 365-384.
87. Jung, J.Y. and Y.H. Yun, *Prediction of Self-management Strategies for Future Quality of life and Health Statuses in Cancer Survivors Using Machine Learning Techniques: A Prospective Cohort Study with a Web-Based Survey.* 2022.
88. Kim, S. and D. Kim, *Imputation method for missing data based on clustering and measure of property.* The Korean Journal of Applied Statistics, 2018. **31**(1): p. 29-40.
89. Zeynalova, A., et al., *Preoperative evaluation of tumour consistency in pituitary macroadenomas: a machine learning-based histogram analysis on conventional T2-weighted MRI.* Neuroradiology, 2019.

61(7): p. 767-774.
90. Chen, T. and C. Guestrin. *Xgboost: A scalable tree boosting system*. in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.
91. Putatunda, S. and K. Rama. *A comparative analysis of hyperopt as against other approaches for hyper-parameter optimization of XGBoost*. in *Proceedings of the 2018 International Conference on Signal Processing and Machine Learning*. 2018.
92. Li, B., et al., *Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods.* Frontiers in genetics, 2018. **9**: p. 237.
93. Kinnander, M., *Predicting profitability of new customers using gradient boosting tree models: Evaluating the predictive capabilities of the XGBoost, LightGBM and CatBoost algorithms*. 2020.
94. Domínguez-Olmedo, J.L., et al., *Machine learning applied to clinical laboratory data in Spain for COVID-19 outcome prediction: model development and validation.* Journal of medical Internet research, 2021. **23**(4): p. e26211.
95. Lundberg, S.M., G.G. Erion, and S.-I. Lee, *Consistent individualized feature attribution for tree ensembles.* arXiv preprint arXiv:1802.03888, 2018.
96. Yun, Y., et al., *Predictors of prescription of morphine for severe cancer pain by physicians in Korea.* Annals of oncology, 2005. **16**(6): p. 966-971.
97. Yun, Y.H., et al., *Prognostic value of quality of life score in disease-free survivors of surgically-treated lung cancer.* BMC cancer, 2016. **16**(1): p. 1-10.
98. Vanbutsele, G., et al., *Effect of early and systematic integration of palliative care in patients with advanced cancer: a randomised controlled trial.* The Lancet Oncology, 2018. **19**(3): p. 394-404.
99. Kursa, M.B., A. Jankowski, and W.R. Rudnicki, *Boruta–a system for feature selection.* Fundamenta Informaticae, 2010. **101**(4): p. 271-285.
100. Hooker, G. and L. Mentch, *Please stop permuting features: An explanation and alternatives.* arXiv e-prints, 2019: p. arXiv: 1905.03151.
101. Jung, J.Y. and Y.H. Yun, *Self-Management Strategy Clustering, Quality of Life, and Health Status in Cancer Patients Considering Cancer Stages.* International Journal of Behavioral Medicine, 2022: p. 1-8.
102. Walsh, D., S. Donnelly, and L. Rybicki, *The symptoms of advanced cancer: relationship to age, gender, and performance status in 1,000 patients.* Supportive care in cancer, 2000. **8**(3): p. 175-179.
103. Morita, T., et al., *Contributing factors to physical symptoms in terminally-ill cancer patients.* Journal of pain and symptom management, 1999. **18**(5): p. 338-346.
104. Satin, J.R., W. Linden, and M.J. Phillips, *Depression as a predictor of disease progression and mortality in cancer patients: a meta-analysis.*

Cancer, 2009. **115**(22): p. 5349-5361.
105. Pinquart, M. and P. Duberstein, *Depression and cancer mortality: a meta-analysis.* Psychological medicine, 2010. **40**(11): p. 1797-1810.
106. Beesley, V.L., et al., *Coping strategies, trajectories, and their associations with patient-reported outcomes among women with ovarian cancer.* Supportive Care in Cancer, 2018. **26**(12): p. 4133-4142.
107. Coyne, J.C., et al., *Emotional well-being does not predict survival in head and neck cancer patients: a Radiation Therapy Oncology Group study.* Cancer: Interdisciplinary International Journal of the American Cancer Society, 2007. **110**(11): p. 2568-2575.
108. Carey, M., et al., *The prognostic effects of performance status and quality of life scores on progression-free survival and overall survival in advanced ovarian cancer.* Gynecologic oncology, 2008. **108**(1): p. 100-105.
109. Blazeby, J., S. Brookes, and D. Alderson, *The prognostic value of quality of life scores during treatment for oesophageal cancer.* Gut, 2001. **49**(2): p. 227-230.
110. Efficace, F., et al., *Is a patient's self-reported health-related quality of life a prognostic factor for survival in non-small-cell lung cancer patients? A multivariate analysis of prognostic factors of EORTC study 08975.* Annals of oncology, 2006. **17**(11): p. 1698-1704.
111. Chang, V.T., et al., *Quality of life and survival: the role of multidimensional symptom assessment.* Cancer, 1998. **83**(1): p. 173-179.
112. Brownlee, J., *XGBoost With python: Gradient boosted trees with XGBoost and scikit-learn*. 2016: Machine Learning Mastery.

Abstract in Korean

연구 배경: 암 케어 연속선상에서 자가관리전략은 암 병기 또는 치료 계획과 관계없이 암환자의 건강관련 삶의 질 또는 생존을 개선하는데 도움이 될 수 있다. 그러나 실제 임상 현장에서 암 병기를 고려한 자가관리전략이 어떻게 클러스터링 되는지에 대한 연구와 암환자의 건강관련 삶의 질 또는 생존 예측 모델은 부족한 실정이다. 또한 암환자의 자가관리전략과 건강관련 삶의 질, 생존 간의 관계를 종합적으로 살펴본 연구는 아직까지 없는 실정이다. 따라서 본 연구는 클러스터링 통계 방법, 머신러닝 기술 및 구조방정식 모델의 경로분석을 활용하여 암환자의 자가관리전략, 건강관련 삶의 질 및 생존 간의 관계를 규명하고자 하였다.

연구 방법: 암생존자의 경우, 새롭게 개발한 건강경영전략(Smart Management Strategies for Health Assessment Tool, SAT)으로 자가관리전략을 측정하여 SAT 전략들 간의 상호관계를 조사하기 위해 주성분 분석과 K-mean 클러스터링 방법을 사용한 군집 분석을 수행하였다. 또한 SAT 전략과 6개월 후의 HRQoL 간의 연관성을 확인하기 위해 다변량 분석을 수행하였다. 암생존자의 HRQoL 예측 모델 개발 및 검증을 위해서는 예측 모델을 구성하고, 결정 트리, 랜덤 포레스트, 경사 부스팅(Gradient boosting), XGBoost, and LightGBM의 앙상블 알고리즘을 사용하여 모델의 성능을 비교하였다. 모델 비교 후, 추가 분석을 위해 최종적으로 XGBoost 모델이 선택되었고, XGBoost의 HRQoL 예측 모델의 중요한 변수를 찾고자 SHAP을 사용하여 특성 중요도 (Feature importance) 및 개별 예측 (Individual prediction) 분석을 수행하였다.

진행성 암환자에서 HRQoL과 SAT 전략의 연관성을 조사하기 위한 클러스터링 및 다변량 분석 방법은 암생존자에서 수행했던 방법과 동일하였다. 생존 예측 모델 개발을 위해 기존의 통계분석을 사용하여 차원 다중 Cox 비례 위험 회귀 분석을 수행하였고, 머신러닝 기법의

XGBoost방법으로 생존 예측 모델을 개발하였다. 본 연구에서는 전통적 통계 방법에 의해 선택된 변수와 머신러닝 기법에 의해 선택된 변수 및 두 방법에 의해 선택된 변수를 결합하여 예측모델을 개별적으로 구성하였고, 성능을 비교하였다. 또한 구조방정식 모델을 활용한 경로분석을 통해 SAT 전략과 HRQoL, 생존 간의 인과관계를 규명하고자 하였다.

연구 결과: 암생존자 및 진행성 암환자의 SAT 전략 클러스터링은 암병기에 따라 다르게 나타났다. 중기-말기 단계 암 환자들은 초기 단계 암 환자들에 비해 위기를 극복하기 위해 자가관리전략에서 치료 시기 및 암병기에 관계없이 모든 단계에서 중요한 핵심 전략을 준비 및 실행전략과 함께 사용하는 것으로 나타났다. 또한 이러한 SAT 전략은 진행성 암환자를 포함하여 모든 암환자에게서 개선된 HRQoL과 긍정적인 연관성을 보여주었다. 머신러닝을 활용한 HRQoL의 예측 모델은 암생존자에서 높은 예측 성능을 보여주었다. 그러나, 각 HRQoL 요인에 대한 중요 변수는 서로 다르게 나타났다. 또한 본 연구는 암생존자를 대상으로 한 웹기반 설문 조사 연구와 새롭게 찾아낸 SHAP을 통한 개인 예측 방법을 접목함으로써 암생존자를 대상으로 한 개인 맞춤형 의료 서비스 제공 방안을 구체적으로 제시하였다. 진행성 암환자에서 차원별 단변량 Cox 모델에서는 ECOG 수행 상태, 성별, 결혼상태, 진단시점에서의 일반적 삶의 질 저하, 호흡곤란, 통증, 식욕감퇴, 변비, 우울, 12주 동안의 임상적으로 의미 있는 정서적 기능 및 사회적 지지의 변화가 최종적으로 더 저하된 생존과 관련이 있는 요인으로 나타났다. 머신러닝방법을 활용한 예측 모형에서도 높은 생존 예측 성능이 나타났고, BorutaSHAP을 통해서는 우울, 통증, 식욕감퇴, 변비, 성별이 생존과 연관된 중요한 요인으로 선별되었다. 기존의 전통적 통계방법과 머신러닝 기법으로 선정된 변수를 결합하여 모델을 구성하였을 때, 생존 예측 모형에서 가장 높은 성능이 발견되었다. 경로분석에서는 SAT전략, 우울, 생존 간의 인과관계를 밝혔으며, 우울 변수를 완전 매개로 SAT 전략의 생존에 대한 간접효과가 있는 것이 발견되었다.

연구 결론: 본 연구는 처음으로 암생존자 및 진행성 암환자를 모두 포함하여 암병기를 고려한 자가관리전략 사용 군집 분석을 시도하였다. 또한 본 연구는 처음으로 암생존자에게 중요한 건강관련 삶의 질을 예측하는 단순한 모델을 개발 및 검증하였고, 설명 가능한 인공지능 알고리즘을 활용하여 모델을 해석하고, 암생존자를 위해 임상환경에서 본 연구의 결과 활용할 수 있는 방안을 제안하였다. 또한 본 연구에서는 머신러닝 기법과 경로분석을 사용하여 진행성 암환자의 자가관리전략과 건강관련 삶의 질 및 생존 간에 직·간접적으로 긍정적인 연관성이 있음을 발견하였다. 이러한 연구결과는 새롭게 개발한 SAT 자가관리전략이 임상장면에서 암환자에게 유용한 개입 도구로 사용될 수 있음을 보여준다. 종합적으로 본 연구는 암환자의 자가관리전략 사용 및 그 효과성에 대한 이해의 폭을 넓혔고, 의료제공자가 암 케어 연속선상에서 암환자에게 도움이 되는 의료 서비스를 제공하는데 자가관리전략을 어떻게 활용할 수 있을지 종합적인 결과 및 임상적 활용방안을 제시하였다는데 의의가 있다.

Supplementary Information (SI) 1

SAT (Smart Management Strategy for Health Assessment Tool) Questionnaire

In this survey, you will evaluate strategies that you might use to proactively overcome a crisis and grow.

Please answer the following questions.

1) What is the current biggest **crisis** you are currently facing?

① Diagnosis of the disease ② Suffering in the treatment process ③ Relapse of the disease ④ Difficulty in personal relationship ⑤ Financial difficulty

① Emotional difficulty ⑦ Difficulty in family relationship ⑧ Loss of self-esteem ⑨ Other (______________________________)

2) What is your best **goal**?

① Getting a treatment ② Overcoming the illness ③ Health recovery ④ Positive growth (Inner maturity) ⑤ Living a life of giving (such as volunteering)

① To return to a former job ⑦ Recovery of personal relationship ⑧ Recovery of self-confidence/ self-esteem ⑨ Other (______________________________)

3) The following items relate to **the current action step** for the above goal. Please put a check mark on the answer that best describes your current level.

① I don't plan on acting in the near future	② I am not active at the moment but I am thinking about acting within a month.	③ I am not active at the moment but I am thinking about acting within a week.	④ I have been active for less than 6 months	⑤ I have been active for more than 6 months

The following items are the core competencies that you might use to overcome **the above crisis** proactively and grow positively in your life. Please read each question carefully and circle the number which best describes your activity **<u>during the past month</u>**.

SAT: Core Strategy		Never	Some times	Quite often	Always
1	I put into action first what is valuable in my life.	1	2	3	4
2	I tried to solve the problem rather than blaming it.	1	2	3	4
3	I concentrated on what I could do rather than what I could not do.	1	2	3	4
4	I led my life proactively.	1	2	3	4
5	I changed the way I lived when I thought it was necessary.	1	2	3	4
6	I made the best choice in any given situation.	1	2	3	4
7	I put my best effort into everyday life.	1	2	3	4
8	I have been careful not to be dragged into negative emotions.	1	2	3	4
9	I often said “I can do it well.” to myself.	1	2	3	4
10	I forgave those who hurt my feeling.	1	2	3	4
11	I tried to see the brighter side more than the dark side.	1	2	3	4
12	I discarded negative perceptions in crisis situations.	1	2	3	4
13	I tried to live happily even in difficult situations.	1	2	3	4
14	I opened all possibilities and thought positively even in despairing situations.	1	2	3	4
15	I thought of my future in a positive way.	1	2	3	4
16	I had faith in myself and my life.	1	2	3	4
17	I overcame difficulties by reminding myself of the purpose of my life.	1	2	3	4
18	I tried to change the way I think and live accordingly.	1	2	3	4
19	I regarded the crisis I am facing as a chance for personal growth.	1	2	3	4
20	I have managed to pull through the crisis thinking of those who love me.	1	2	3	4
21	I strengthened my will to live remembering the ones I love.	1	2	3	4
22	I realized that the people around me would be happy if I am happy with myself.	1	2	3	4
23	I kept good relationships with my family, the people around me, and the medical staff.	1	2	3	4

24	I trusted the hospital and the medical staff.	1	2	3	4
25	I appreciated the preciousness of my life.	1	2	3	4
26	I shared my experiences with people who had gone through a similar crisis.	1	2	3	4
27	I have been encouraged by stories from people who overcame a crisis.	1	2	3	4
28	I had someone to share my fear and worries with.	1	2	3	4

The following items are **preparation strategies** that you might use to establish **the above goals** and a detailed life plan with which to overcome a crisis proactively and grow. Please read each question carefully and circle the number which best describes your activity **<u>during the past month</u>.**

SAT: Preparation Strategy		Never	Some times	Quite often	Always
1	I have thought of the meaning of the problems I am facing in my life.	1	2	3	4
2	I have newly found a reason and meaning for life.	1	2	3	4
3	I have kept telling myself that the direction is more important than the pace of life.	1	2	3	4
4	I have written down what I really want to do in the future.	1	2	3	4
5	I have realized clearly what kind of life I want to live.	1	2	3	4
6	I had a solid understanding of who I am and what I am good at.	1	2	3	4
7	I established a specific goal of my life and started to act on it.	1	2	3	4
8	I set up a life goal which would maximize my strengths.	1	2	3	4
9	I set up detailed action plan to accomplish my life goal.	1	2	3	4
10	I established simple and definite goals before acting.	1	2	3	4
11	I drew up balanced plans according to my diverse roles.	1	2	3	4
12	I clearly knew the purpose of what I had to do.	1	2	3	4
13	I had my own way to deal with stresses.	1	2	3	4
14	I had concretely imagined what I want to be in the future.	1	2	3	4
15	I devised a checklist to examine my behavior.	1	2	3	4
16	I have examined my health habits and identified those to be fixed.	1	2	3	4
17	I figured out what part of my surrounding environment required a change.	1	2	3	4
18	I tried to view the situation objectively when making a decision.	1	2	3	4
19	I had a clear understanding of the pros and cons before making an important decision.	1	2	3	4
20	I consulted fully with the medical staff and my family before making a decision.	1	2	3	4
21	I engaged myself proactively in making a decision.	1	2	3	4
22	I prioritized “to do” list before I drew up plans.	1	2	3	4
23	I made plans in such a way that the most urgent and important ones would be done first.	1	2	3	4

24	I wrote down my goal and posted it where it could be seen to remind myself.	1	2	3	4
25	I carried out my plans in order of ease.	1	2	3	4
26	I actively referred to better methods, if any, to carry out my plans.	1	2	3	4
27	I collected information from various sources, such as family members and the medical staff.	1	2	3	4
28	I actively asked questions or looked for necessary information to solve problems.	1	2	3	4
29	I have done my best to complete the tasks that I could and asked others to do the ones I could not do.	1	2	3	4
30	I repeated good behaviors so that they become a habit.	1	2	3	4

The following items are the **implementation strategies** that you might carry out to overcome **the above crisis** proactively and growth positively in your life. Please read each question carefully and circle the number which best describes your activity **<u>during the past month</u>.**

	SAT: Implementation Strategy	Never	Some-times	Quite often	Always
1	I carried out the easiest plans first and therefore often felt sense of achievement.	1	2	3	4
2	I did not postpone my work, but complete it on time.	1	2	3	4
3	I used a concrete method (memo, alarm function, etc.) to remind myself of my to-do list.	1	2	3	4
4	I **conscientiously** put my plans into action.	1	2	3	4
5	I primarily engaged in the activities that I could enjoy.	1	2	3	4
6	I have found my own way that I can enjoy in any kind of activities.	1	2	3	4
7	I created activities that I can enjoy by myself and concentrate on.	1	2	3	4
8	I have found my own way to energize myself. (Ex: listening to music, exercising)	1	2	3	4
9	I maintained physical and mental strength to handle stresses.	1	2	3	4
10	I carried out my plan in order of priority.	1	2	3	4
11	I used my given time carefully.	1	2	3	4
12	I developed my competency by using my resources and time effectively.	1	2	3	4
13	I tried to keep balance among my diverse roles.	1	2	3	4
14	I adjusted my work and life according to my body rhythm.	1	2	3	4
15	I had a strong will to learn.	1	2	3	4
16	I tried to fix my old bad habits.	1	2	3	4
17	I tried to keep my values straight everyday.	1	2	3	4
18	I tried to do align my actions with my plan.	1	2	3	4
19	I have checked persistently if my plans were carried out well.	1	2	3	4
20	I sometimes set my work aside and took time to reexamine the purpose of life.	1	2	3	4
21	I regularly had time for **examining my attitude.**	1	2	3	4
22	I did my best to carry out my plans but not to the point of exhaustion.	1	2	3	4

23	I skipped unnecessary works or procedures as much as possible.	1	2	3	4
24	I paid less attention to the things that I did not have to be concerned with.	1	2	3	4
25	I rewarded myself when I accomplished my plan or overcame difficulty.	1	2	3	4
26	I felt joy in every experience I had.	1	2	3	4
27	I looked upon the present from the bigger picture.	1	2	3	4
28	I tried to gain peace of mind by calming myself down.	1	2	3	4
29	I believed that I could grow after overcoming a crisis successfully.	1	2	3	4
30	I found meaning in trying rather than being obsessed with the results.	1	2	3	4
31	I tried my best in everything and went with the flow for the outcome.	1	2	3	4
32	I listened to the objective opinions of the medical staff and others around me about whether I carry out my plans well.	1	2	3	4
33	I reflected on the feedback from the medical staff and people around me.	1	2	3	4

Supplementary Information (SI) 2

SI 2. The univariate and adjusted logistic regression analyses of each SAT sub−strategy for depression with backward selection

Measure	Category		OR(95%CI)	P	Adjusted OR (95%CI)	P
SAT−C	Positive reframing	≤66.66	1	0.58		
		>66.66	0.82 (0.40−1.68)			
	Proactive problem solving	≤66.66	1	0.38		
		>66.66	0.71 (0.34−1.51)			
	Creating empowered relationship	≤66.66	1	0.28		
		>66.66	1.50 (0.72−3.14)			
	Experience sharing	≤66.66	1	0.81		
		>66.66	1.09 (0.54−2.21)			
SAT−P	Life value pursuing	≤66.66	1	0.63		
		>66.66	0.84 (0.40−1.74)			
	Goal and action setting	≤66.66	1	0.15		
		>66.66	0.54 (0.23−1.26)			
	Rational decision−making	≤66.66	1	0.12		
		>66.66	0.52 (0.22−1.20)			
	Priority−based planning	≤66.66	1	0.4		
		>66.66	0.72 (0.33−1.57)			
	Healthy environment creating	≤66.66	1	0.11		
		>66.66	0.43 (0.15−1.23)			
SAT−I	Self−implementing	≤66.66	1	0.04	1	0.02
		>66.66	0.35 (0.12−0.97)		0.24 (0.07−0.78)	
	Activity−coping	≤66.66	1	0.32		
		>66.66	0.68 (0.32−1.45)			
	Self−sustaining	≤66.66	1	0.02	1	0.03
		>66.66	0.26 (0.08−0.89)		0.27 (0.08−0.90)	
	Energy−conserving	≤66.66	1	0.29		
		>66.66	0.68 (0.33−1.40)			
	Self−motivating	≤66.66	1	0.71		
		>66.66	0.84 (0.34−2.09)			

	Reflecting	≤66.66	1	0.95
		>66.66	0.98 (0.46-2.09)	

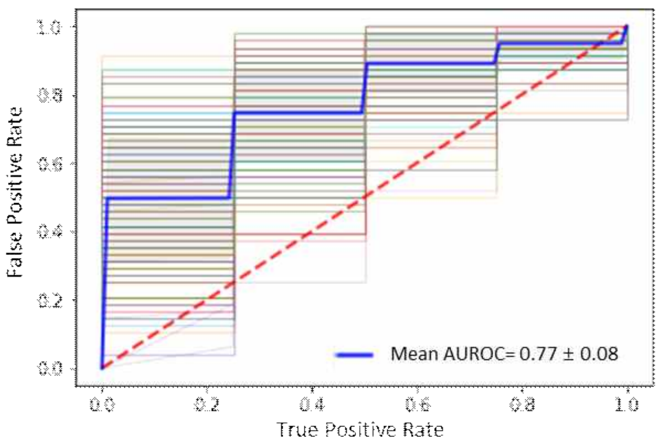
aOR, adjusted odd ratio; CI, confidence interval; P, p-value

*Adjusted for age, sex, education, income, religion, residence, marriage, ECOG performance status, tumor site

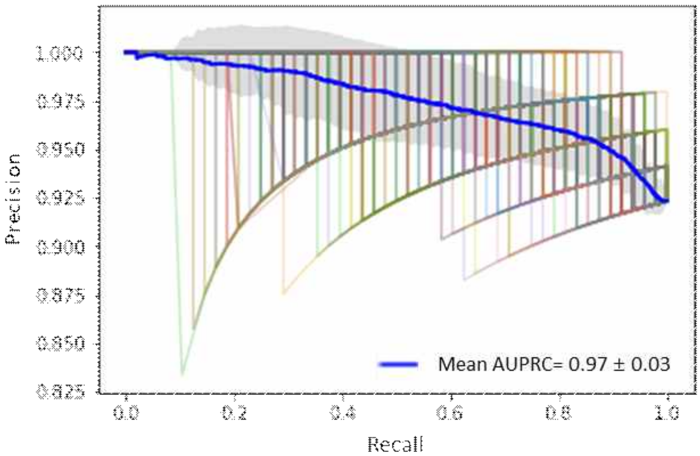
Supplementary Information (SI) 3

SI 2. Area under the receiver operator characteristic curve (AUROC) and area under the precision-recall curve (AUPRC) in the XGBoost model for predicting global QoL after the bootstrap validation for cancer survivors.

AUROC



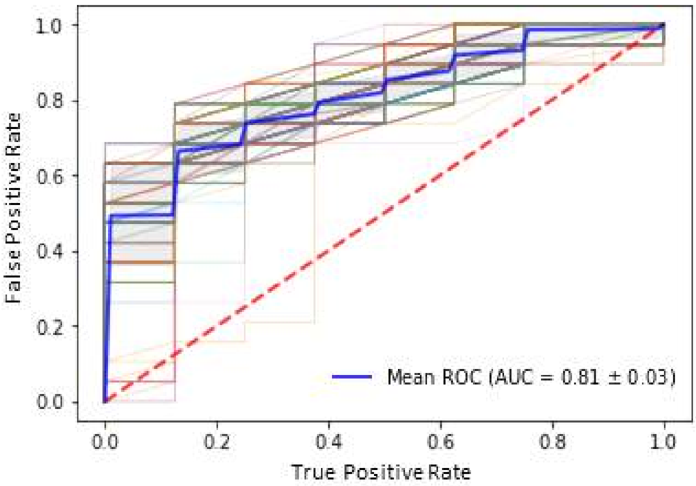
AUPRC



Supplementary Information (SI) 4

SI 2. Area under the receiver operator characteristic curve (AUROC) and area under the precision-recall curve (AUPRC) in the combined variables' XGBoost model for predicting survival after the bootstrap validation for advanced cancer patients.

AUROC



AUPRC