BMC Psychology

https://doi.org/10.1186/s40359-025-03809-3

Article in Press

Enhancing happiness and life satisfaction in university students: analysis with a machine learning approach

Received: 16 January 2025

Accepted: 3 December 2025

Published online: 09 December 2025

Cite this article as: Long Q., Wang Y., Axelin A. *et al.* Enhancing happiness and life satisfaction in university students: analysis with a machine learning approach. *BMC Psychol* (2025). https://doi.org/10.1186/s40359-025-03809-3

Qing Long, Yuning Wang, Anna Axelin, Feng Zheng, Zeng Cao, Xiaotian Li & Jia Guo

Enhancing Happiness and Life Satisfaction in University Students: Analysis with a Machine Learning Approach

Qing Long[1]; Yuning Wang[2]; Anna Axelin[3]; Feng Zheng[4]; Zeng Cao[5]; Xiaotian Li[6]; Jia Guo[7*]

1. Xiangya School of Nursing, Central South University, Changsha, China; longqing226@163.com
2. Department of Computing, University of Turku, Turku, Finland; yuning.y.wang@utu.fi
3. Department of Nursing Science, Faculty of Medicine, University of Turku, Turku, Finland; anmaax@utu.fi
4. Department of Physical Education and Research, Central South University, Changsha, China zhengfeng100@126.com
5. Department of Physical Medicine & Rehabilitation, Xiangya Hospital, Central South University, Changsha, China; caozengxyyy@csu.edu.cn
6. Department of Physical Education and Research, Central South University, Changsha, China; xiaotianli@csu.edu.com
7. Xiangya School of Nursing, Central South University, Changsha, China; guojia621@163.com

Corresponding author: Jia Guo* guojia621@163.com

ABSTRACT

Background: Subjective well-being (SWB) is vital for the personal growth of university students. Machine learning approach have been increasingly used in identifying SWB predictors for their ability to capture complex and multidimensional predictors. Still, the feature selection is not often justified from a theoretical perspective.

Objective: Under the guidance of the conceptual model of psychology and public health, this study aims to apply machine learning to identify the top predictors of happiness and life satisfaction (LS) as the two components of SWB among a sample of university students.

Methods: This cross-sectional study analyzed university students from the China Family Panel Studies, including 816 participants from the 2022 wave for model development and 724 from the 2020 wave for external validation. The development set was randomly split into a training set (70%) and a test set (30%). Forty-two variables across the conceptual model of psychology and public health were included. Missing values were imputed using multiple imputation, LASSO regression was used for feature selection, and SMOTE-IPF addressed class imbalance. Five tree-based machine learning models (Random Forest, AdaBoost, Gradient Boosting, XGBoost, and LightGBM) were trained with 10-fold cross-validation, and the best model was chosen according to cross-validated AUC. Performance was further evaluated in the internal test and external validation sets using ROC and PR curves, accuracy, sensitivity, specificity, F1-score, and other metrics. The model explanation was enhanced with SHAP values to assess the detailed contribution of each predictor and Venn diagrams to evaluate shared predictors of happiness and LS.

Results: Among 816 university students, 15.9% reported low happiness and 28.9% reported low LS in the development set, with similar proportions observed in the external validation set. The Random Forest model achieved the best performance for happiness prediction (AUC = 0.831 in the test set and 0.741 in the external validation set), while XGBoost performed best for LS (AUC = 0.730 and 0.748, respectively). SHAP analysis revealed interpersonal relationships were the strongest predictor of happiness, while future confidence was the top predictor of LS. Shared predictors across both outcomes included future confidence, interpersonal relationships, depressive symptoms, and the relationship with mother.

Conclusions: The machine learning approach demonstrates good predictive performance, thus may offer new thoughts for supporting SWB among university students, such as strengthening interpersonal relationships and fostering future confidence.

Keywords: subjective well-being; happiness; life satisfaction; university students; machine learning; feature importance

1. INTRODUCTION

Subjective well-being (SWB) is defined as the overall evaluation of an individual's affective feelings and life [1]. It is commonly conceptualized as comprising happiness and life satisfaction (LS) [2]. For university students, SWB is associated with academic growth, self-efficacy, and physical and mental health [3, 4], laying a solid foundation for long-term well-being. SWB varies across the life span following a U-shaped pattern, with the university stage being a critical period characterized by potential fluctuations and declines [5, 6]. It remains a challenge for university students globally [7]. For instance, 19.1% of Vietnamese university students reported low happiness [8], and the proportion of European university students with low LS ranged from 27.1% in Czechia to 71.9% in Turkey [9]. In China, the situation appears similarly concerning; about one-third of graduate students report low happiness [10], and 22.7-55.0% of university students report low LS [11]. Consequently, understanding the predictors of happiness and LS of university students has become a priority.

The predictors of happiness and LS should be considered separately, because they are independent but complementary dimensions of SWB [12]. Happiness represents the affective dimension of SWB, often involving immediate, accessible positive feelings without complex cognitive evaluation [5]. LS represents the cognitive evaluative dimension of self-rated overall quality of life, which may require demanding thinking and a complex comparison between desired and actual life situations [13]. Evidence suggests different key predictors influence these two dimensions. In a global adult sample across 132 countries, material prosperity (e.g., income) more strongly predicted LS, whereas social psychological prosperity (e.g., autonomy, respect, and social relationships) more strongly predicted happiness [13]. However, little is understood about the predictors of happiness and LS among university students.

Most existing studies that explored the predictors of SWB rely on traditional statistical approaches [14-17]. These methods typically rely on predefined hypotheses and linear assumptions, which limit their ability to capture the complex and potentially nonlinear relationships between diverse influencing factors and the two dimensions of SWB [18]. In contrast, machine learning (ML) offers notable advantages in modelling high-dimensional and heterogeneous data. It is less constrained by statistical assumptions and can effectively capture nonlinear patterns [19]. Furthermore, explainable ML tools such as SHapley Additive Explanations (SHAP) enhance interpretability by identifying and ranking the top predictors [20] and revealing complex underlying relationships between the predictors and SWB [21]. These capabilities make ML a promising approach for investigating the predictors of happiness and LS.

Existing studies have explored predictors of SWB among university students using ML, reporting predictive accuracies ranging from 69% to 90% [10, 22, 23]. These studies often selected candidate predictors based on literature review, but with little justification, including psychological factors (e.g., emotion regulation strategies and coping styles) [10] or behavioral indicators (e.g., daily steps and sleep duration) [23]. For example, one study uses algorithm-based feature selection in a pool of 298 psychosocial variables, then identifies the top 20 predictors, such as depressive symptoms and personality traits [17]. Feature selection methods in previous studies are not often grounded in a theoretical perspective [10, 22, 23], which can lead to inconsistent predictive performance across studies and may overlook theoretically important variables [24]. This challenge might become even more pronounced when dealing with high-dimensional datasets, which often contain irrelevant features that compromise model performance and reduce interpretability for stakeholders [24].

Consequently, theoretical relevance is an important consideration in feature selection of ML models on predictors of SWB [25]. Psychological theories have advanced understanding of individual-level processes that influence SWB, such as personality traits and emotions, but they often overlap considerably and give limited attention to contextual influences [25]. In contrast, public health research has emphasized population-level factors of SWB, such as health, disease burden, and social conditions. However, it has often been criticized for lacking a clear theoretical foundation, and has primarily focused on empirically identifying relevant determinants and correlates of SWB [25]. Thus, when understanding SWB's predictors, an integrated conceptual model that combines insights from psychology and public health could provide a holistic conceptual basis for feature selection.

This study aimed to apply a combined ML and feature selection strategy based on psychology and public health perspectives to identify the top predictors of happiness and LS among university students. The findings can warrant greater attention to prioritize the key predictors and inform targeted strategies accordingly.

2. METHOD

2.1 Study design

This study used a cross-sectional design and was reported following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist, as is shown in supplementary table S1 [26].

2.2 Setting and data sources

This study used data from the China Family Panel Studies (CFPS, http://www.isss.edu.cn/cfps/ EN/), a nationwide longitudinal survey developed by the Institute of Social Science Survey (ISSS) at Peking University in 2010 and conducted biennially. The CFPS aims to track individual, family, and community data, reflecting changes in China's social, economic, demographic, educational, and health landscapes. To ensure the representativeness and validity of the data, the baseline survey of CFPS in 2010 used a three-stage sampling method with implicit stratification to obtain an equal probability sample [27]. Firstly, counties/administrative equivalents were drawn from 25 selected provinces (CFPS does not cover Tibet, Qinghai, Xinjiang, Ningxia, Inner Mongolia, Hainan, Hong Kong, Macau, or Taiwan). Secondly, communities were drawn from selected counties/administrative equivalents. The socioeconomic level was used to indicate implicit stratification at these two stages. Thirdly, 25 households were randomly drawn from each sampled community based on the onsite sampling frame, and members of every household were asked to participate in the survey. Thus, it represents 95% of the total population in the Chinese mainland.

We selected the cross-sectional data from the CFPS 2022 and 2020 surveys for two reasons. First, it provides the most recent, nationally representative dataset that aligns with the present study's aims from both psychological and public health perspectives. Several indicators relevant to the conceptual model of psychology and public health, such as dietary habits, online gaming and shopping, watching short videos, and family support, were first introduced in the 2020 survey and retained in the 2022 survey [28]. These indicators have also recently attracted growing research interest among university students [29-31]. Second, the CFPS survey's proportion of university students was 3.02% in 2022 and 2.53% in 2020, which is closer to the Ministry of Education's estimate of 2.97% [32].

2.3 Participant selection

This study focused on university students in China. Individuals were eligible for inclusion if they were enrolled in diploma-level or higher education programs at the time of the survey, while those not pursuing such programs were excluded. The 2022 CFPS survey database was used for model development, and the 2020 wave for external validation. The final analytic samples comprised 816 participants in the development group and 724 in the validation group, as illustrated in Figure 1.

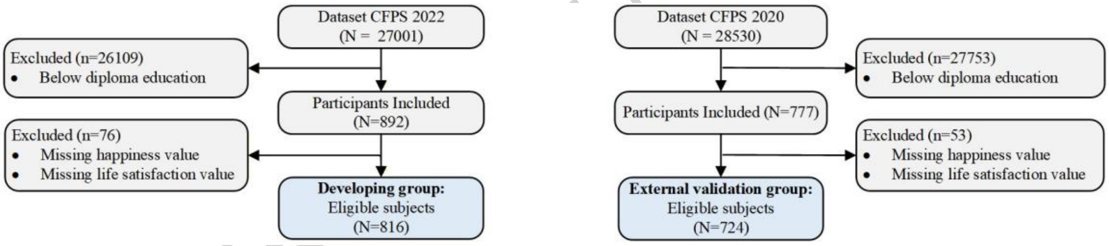

Figure 1. Participants' selection process for the study dataset

2.4. Measures

The CFPS questionnaire was developed by ISSS through a systematic process involving multiple rounds of expert consultation and culturally adapted questionnaires from widely recognized international surveys (e.g., the National Longitudinal Surveys of Youth) to ensure measurement reliability and validity among the Chinese population [27]. The CFPS 2020 and 2022 questionnaires are publicly available at http://isss.pku.edu.cn/cfps/en/documentation/questionnaires/index.htm.

2.4.1 Assessment of SWB

In the CFPS 2022 and 2020 surveys, SWB was assessed in the "subjective attitude" module through two dimensions: happiness and LS. Happiness was measured by the question, "How happy do you feel about yourself?", with responses ranging from 0 to 10, where higher scores indicate greater happiness. Scores of 0-6 were classified as low happiness, and 7-10 as high happiness [33]. LS was measured by the question, "How satisfied are you with your life?", with responses ranging from 1 to 5, where higher scores indicate greater satisfaction. Scores of 1-3 were classified as not satisfied, and 4-5 as satisfied [33, 34]. Single-item measures of happiness and LS have been widely used in large-scale international surveys and have demonstrated substantial correlations with multi-item scales and acceptable reliability across diverse populations [35].

2.4.2 Features considered for SWB analysis

Guided by the integrated conceptual model combining psychology and public health perspectives, we selected potential features for SWB analysis based on prior evidence of their association with SWB, the availability of corresponding variables in the CFPS 2022 and 2020 datasets, and multiple rounds of expert consultation. This research initially considered 47 common features across CFPS 2022 and CFPS 2020 for SWB prediction model development. Of these, five variables with more than 30% missing values, including nap duration, sleep duration, daily online learning, daily online gaming, and daily online shopping, were excluded from the analysis to ensure validity [36]. Finally, 42 variables were included.

The 42 variables included were organized according to the conceptual model of psychology and public health [25] and grouped into seven broad categories: 1) Basic demographics: age, gender, and ethnicity; 2) Socioeconomic status: average household income, family size, education level; 3) Health and functioning: health, health improvement [37], chronic disease [38], Body Mass Index (BMI) [39], depression symptoms [37], academic self-efficacy [40], academic stress [41], poor sleep frequency [42], nap habit [43], frequency of physical activity (PA) [44], PA duration [45], protein consumption [46], fruit and vegetables consumption [46], smoking, and alcohol use; 4) Personality: nuanced traits, such as future confidence [47]; 5) Social support: medical insurance, school satisfaction [48], interpersonal relationships [48], meeting frequency with father; contact frequency with father; meeting frequency with mother; contact frequency with mother; relationship with father; relationship with mother [49]; mobile internet use, computer internet use [50], online learning, online shopping, online gaming, weekly short video, daily short video [51], WeChat use, WeChat moments sharing frequency [37]; 6) Religion and culture: religious belief [52]; 7) Geography and infrastructure: residence. A detailed description of 42 variable definitions and assignments is provided in Supplementary Table S2.

2.5 Bias

Firstly, this study is based on cross-sectional data. Cross-sectional data are subject to omitted variable bias, where individual, unobserved effects may be correlated with observed variables. Estimated effects would then include the effects from these unobserved factors, which could either magnify or diminish the measurement of the true effect [53]. Secondly, some measures used in this study contained meaningful amounts of missing data, which could introduce bias if missingness does not occur completely at random.

2.6 Data analytic plan

The overall methodological framework employed in the data analysis is illustrated in Figure 2. The key steps include data processing, feature selection, model selection, performance evaluation, and model explanation.

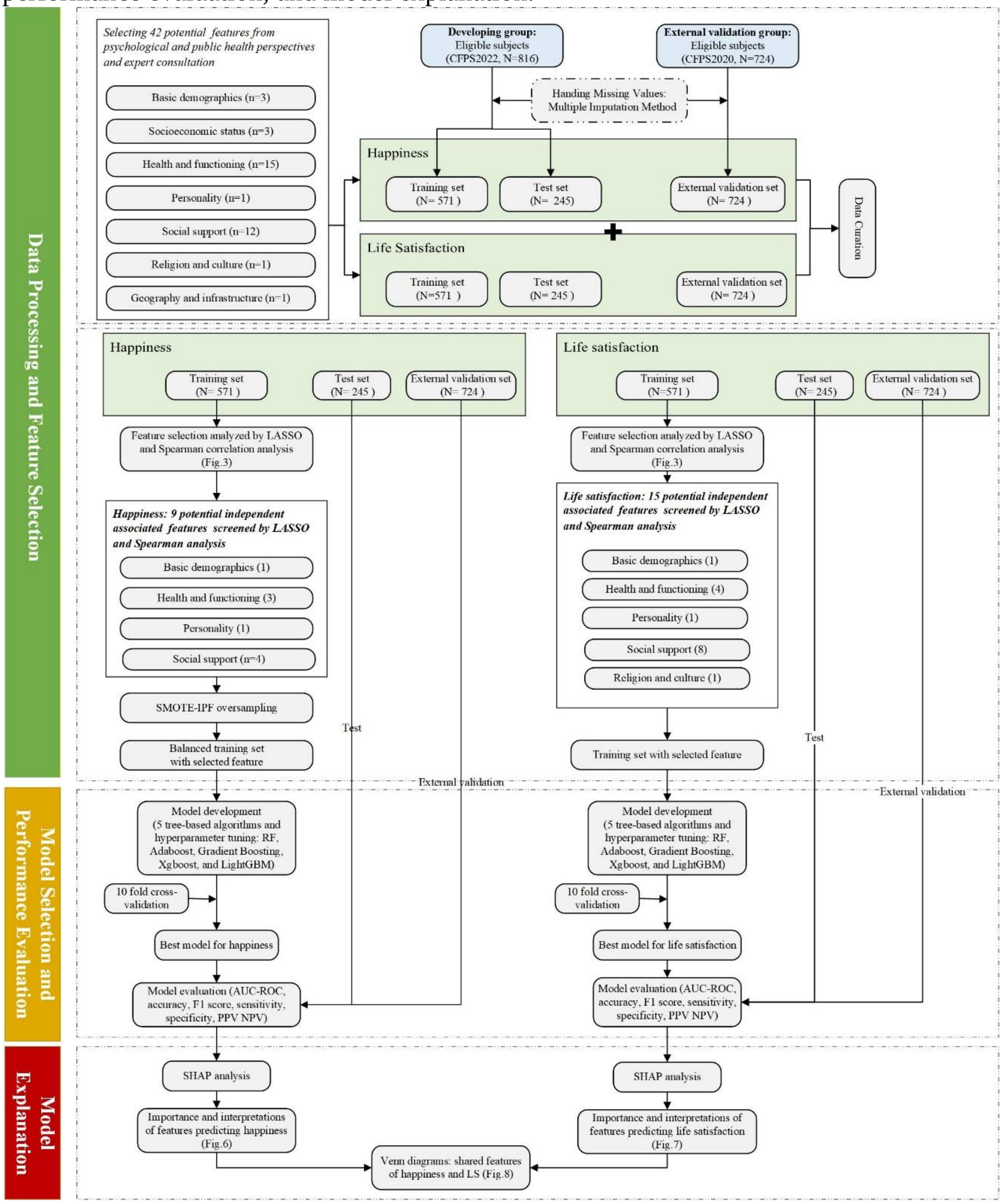

Figure 2. Overview of the study design. *LASSO:* least absolute shrinkage and selection operator; SMOTE-IPF: Synthetic Minority Oversampling Technique with Iterative Partitioning Filter; *RF:* random forest; *AdaBoost:* adaptive boosting; *XGBoost:* extreme gradient boosting; *LightGBM:* light gradient boosting machine; *AUC:* area under the receiver operating characteristic curve; *ROC:* receiver operating characteristic; *PPV:* positive predictive value; *NPV:* negative predictive value

2.6.1 Data processing and feature selection

For the features considered for SWB analysis, missing values were observed, with a mean non-response rate of 2.04% (ranging from 0.00% to 17.65%) in the development group from CFPS 2022 and 3.19% (range from 0.00% to 24.17%) in the external validation group from CFPS 2020. Missing rates of study variables in the development and external validation groups are summarized in Supplementary Table S3. Continuous variables with normal distribution were summarized as mean ± standard deviation (SD), while those with skewed distributions were presented as median (interquartile range, IQR). Categorical variables were expressed as frequencies and percentages.

To ensure the data quality for subsequent ML analysis, multiple imputation (MI) was performed for the remaining variables with missing values, using the *miceforest* package in Python (version 3.12). This simulation-based approach provides more accurate estimates of missing values than single-value imputation methods and was adopted to enhance the analytical robustness of the study [54]. Distribution of study variables in the development and validation group before and after imputation is illustrated in Supplementary Figure S1 and Figure S2, showing that the imputation preserved the original data structure. After imputation, the development group was further divided into a training set (70%) and a test set (30%) to mitigate overfitting.

To select features for predictive modelling, we initially examined potential multicollinearity using Spearman's correlation coefficients among the initial 42 features. The analysis indicated that none of the variable pairs exceeded the commonly used threshold of 0.8, suggesting an absence of problematic collinearity. Thereafter, we applied the least absolute shrinkage and selection operator (LASSO) regression to the training dataset [55]. This procedure was used to mitigate overfitting by shrinking the coefficients of less informative variables and further addressing the multicollinearity. In the LASSO regression, the optimal regularization parameter "λ" was determined via cross-validation to minimize model error [56]. In addition, the resulting set of variables was then used for predictive model development.

The constructed happiness dataset is severely imbalanced, with only 15.9% of participants reporting low happiness and the remaining 84.1% reporting otherwise. To address this imbalance, we applied the Synthetic Minority Oversampling Technique with Iterative Partitioning Filter (SMOTE-IPF) [57]. SMOTE-IPF extends the traditional SMOTE by combining oversampling with an ensemble-based noise filtering procedure [57]. The SMOTE-IPF reduces the potential risk of introducing synthetic samples into noisy regions [58]. The SMOTE-IPF was only applied to the training set to avoid any possibility of information leakage or overfitting.

2.6.2 Model selection and performance evaluation

To identify the best model for predicting happiness and LS from the selected input features in each dataset, five tree-based classifiers underwent training were used to construct the model: Random Forest (RF), Adaptive boosting (Adaboost), GradientBoosting, Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM). These algorithms were chosen for their widespread use and robustness in handling complex data structures and analyzing feature importance [59, 60]. To validate the robustness of the optimal model and mitigate potential overfitting from the oversampled data, we employed 10-fold cross-validation on the training data. During the cross-validation, hyperparameters were tuned using the grid search strategy, with the corresponding hyperparameters outlined in Supplementary Table S4. The computed average area under the receiver operating characteristic curve (AUC) scores over tenfold were regarded as the basis for selecting the optimal model in the training set.

The selected optimal model was then evaluated by the test set (i.e., 2022 CFPS) and an external validation set (i.e., 2020 CFPS). Model performance was assessed using receiver operating characteristic (ROC) curves and the area under the precision-recall (PR) curves, along with several key performance metrics, including AUC-ROC score, accuracy, sensitivity/recall, specificity, negative predictive value (NPV), positive predictive value (PPV), F1 score, and kappa value. In addition, confusion matrices were also generated to provide a direct visualization of classification outcomes, with a default cutoff of 0.5. The AUC-ROC score was regarded as the primary criterion for selecting the best-performing model, as it reflects the model's overall discriminative ability across all possible classification thresholds [61].

2.6.3 Model explanation: feature importance and model transparency

To enhance the model transparency, the final selected model was then tested on the test dataset, and the SHAP method was used to quantify the contribution of each predictive feature for happiness and LS [62]. SHAP values measure each feature's contribution to a model's prediction by analyzing how the prediction changes when the feature is added to or removed from different combinations of other features. This analysis ensures an evaluation of each feature's impact. By using SHAP, we can understand how individual features influence the model's predictions and their relative importance.

Data cleaning and descriptive statistics were performed using STATA version 17. ML methods were implemented using Python software (version 3.12). A detailed list of the libraries, their versions, and their specific applications in this study is provided in Supplementary Table S5.

3. RESULTS

3.1 Data processing results

In the development dataset (CFPS 2022), 130 participants (15.9%) reported low levels of happiness, and 236 (28.9%) were not satisfied with their lives. In the external validation dataset (CFPS 2020), 113 participants (15.6%) reported low levels of happiness, and 239 (33.0%) were not satisfied with their lives. The selected features for ML were categorized into seven domains. Descriptive statistics of the 42 variables across the development and validation datasets after MI are summarized in Table S6.

Across the seven domains, the two datasets demonstrated both similarities and differences. For demographics, both datasets had a median age of 21 years, and the gender distribution was nearly similar. For socioeconomic status, the family size was similar across datasets, with a median of four members in both 2022 and 2020. Regarding health and functioning, depressive symptoms were more frequent in 2022 (23.8% vs 10.8%). The median score of academic self-efficacy and academic stress was 3.00 in both datasets. For personality, both datasets reported consistent levels of future confidence, with a median score of 4.00 [3- 4]. Regarding social support, interpersonal relationships had a median score of 7.00 [6.00- 8.00] in both datasets, and the relationship with mother was reported as very close by 52.1% in 2022 vs 55.0% in 2020. Daily short video use was also more prevalent in 2022 (76.1% vs 61.2%). For religion and culture, only 1.1% of students reported religious beliefs in both datasets. Finally, for geography and infrastructure, urban residence was slightly more common in 2020 (59.1% vs 56.4%).

3.2 Identification of informative features

To identify the most informative variables associated with happiness and LS. Spearman correlation analysis (see Figure 3A) initially assessed the potential multicollinearity. It confirmed that no problematic correlations existed between the selected features, with the highest observed correlation coefficient being 0.78, well below the conventional 0.8 threshold. LASSO regression with 10-fold cross-validation was employed on the training dataset to mitigate overfitting and multicollinearity among the 42 potential variables. The LASSO regression cross-validation process (Figure 3B) identified an optimal regularization parameter (λ) of 0.0255. The coefficient paths (Figure 3C) illustrate the sequential variable elimination process across different λ values, demonstrating the stability and importance of selected features. For happiness, the variables retained by LASSO included depression symptoms, frequency of PA, fruit and vegetable consumption, future confidence, interpersonal relationships, relationship with father, and relationship with mother.

For LS, Spearman correlation analysis was consistent with the happiness results (Figure 3D), with the highest observed correlation being 0.78, and the optimal λ was 0.0233 (Figure 3E). The coefficient paths (Figure 3F) emphasized the variables selected in the final model. The predictors retained for LS included age, health status, depression symptoms, academic self-efficacy, academic stress, future confidence, interpersonal relationships, relationship with mother, online gaming, weekly short video use, WeChat use, WeChat moments sharing frequency, and religious belief.

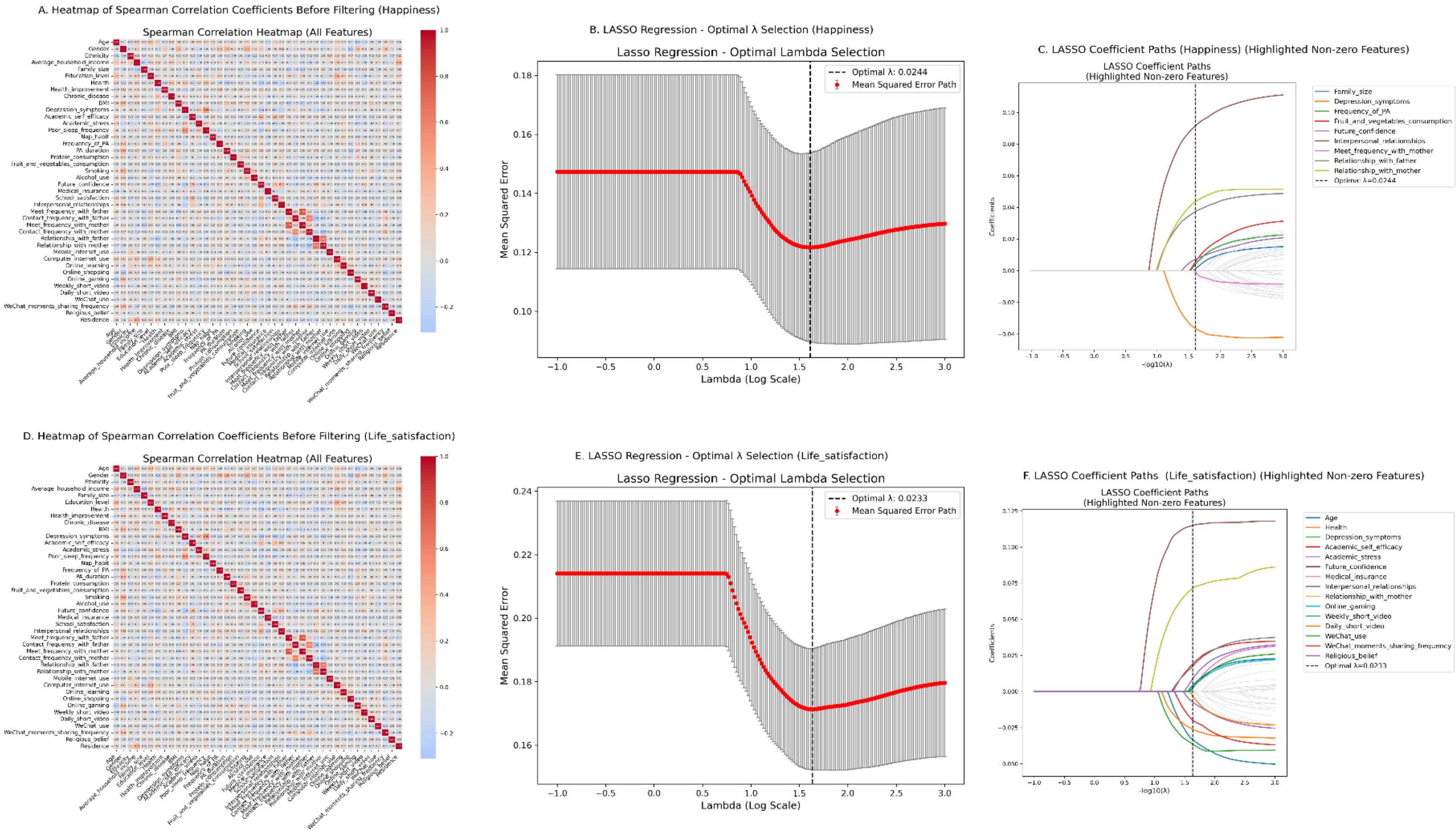

Figure 3. Feature selection and performance evaluation for predicting happiness and life satisfaction. PA: Physical Activity; BMI: Body Mass Index; LASSO: Least absolute shrinkage and selection operator. A) Heatmap of Spearman correlation coefficients before filtering for happiness. B) Lasso regression cross-validation process to determine the optimal regularization parameter (λ) for happiness. The mean squared error path is shown for different λ values, with optimal λ highlighted. C) Coefficient paths of all features for happiness as a function of log-transformed λ, showing the shrinkage process and the selection of important predictors. D) Heatmap of Spearman correlation coefficients before filtering for life satisfaction. E) Cross-validation process for life satisfaction, with optimal λ highlighted. F) Coefficient paths of all features for life satisfaction.

3.3 Robustness analysis: oversampling and class imbalance

To evaluate the impact of class imbalance and the effectiveness of our oversampling strategy, we trained Random Forest classifiers for happiness on the original imbalanced training set (102 low-happiness vs. 469 high-happiness cases) and the oversampled dataset. On the original training set, the model achieved an AUC of 0.832, an accuracy of 0.890, a sensitivity of 0.982, and a specificity of 0.179 in the test set. After applying SMOTE-IPF oversampling with a balanced target ratio of 1:1 (469 low-happiness vs. 469 high-happiness cases), the model's performance improved, yielding an AUC of 0.831, accuracy of 0.837, sensitivity of 0.885, and specificity of 0.464, as reported in Supplementary Table S7.

3.4 Model selection and performance evaluation

Based on the identified features, five ML models were developed to identify the optimal algorithms for predicting happiness and LS.

For happiness, nine variables were included in the multivariate model. Of the models evaluated, the Random Forest demonstrated the best performance in the training set, with an overall cross-validated AUC of 0.965 (Figure 4A), ranging from 0.953 to 0.990 across folds (Figure 4B). In the test set, the model achieved an AUC of 0.831 and an AP of 0.975, with an accuracy of 0.837, sensitivity of 0.885, specificity of 0.464, PPV of 0.928, NPV of 0.342, F1 score of 0.906, and Cohen's kappa of 0.302 (Table 1). In the external validation set, the Random Forest yielded an AUC of 0.741 and an AP of 0.923, with an accuracy of 0.834, sensitivity of 0.921, specificity of 0.363, PPV of 0.887, NPV of 0.461, F1 score of 0.904, and kappa of 0.311 (Figure 4C-D; Table 1). The corresponding confusion matrices for the test and external validation sets are provided in Supplementary Figure S3.

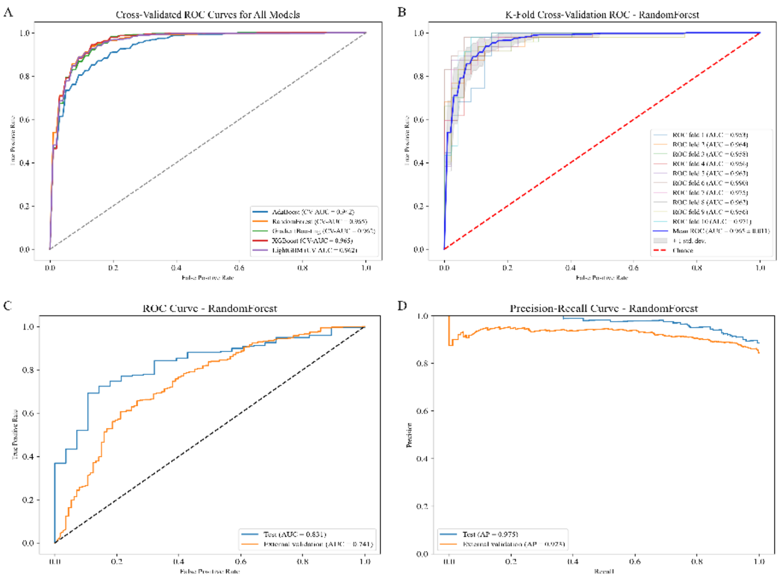

Figure 4. Model performance of the Random Forest for predicting happiness. AUC: Area under the receiver operating characteristic curve; ROC: Receiver operating characteristic. (A) ROC curves in the test set, with cv-AUC values reported for each model; (B) K-fold cross-validation ROC curves of the best-performing model (Random Forest); (C) ROC curves in the test and external validation sets; (D) Precision-Recall curves in the test and external validation sets.

For LS, 15 variables were included in the multivariate model. The XGBoost model performed best, with an overall cross-validated AUC of 0.813 (Figure 5A), ranging from 0.754 to 0.869 across folds (Figure 5B). In the test set, the model achieved an AUC of 0.730 and an AP of 0.881, with an accuracy of 0.767, sensitivity of 0.897, specificity of 0.367, PPV of 0.814, NPV of 0.537, F1 score of 0.853, and a kappa of 0.296. In the external validation set, the XGBoost model yielded an AUC of 0.748 and an AP of 0.847, with an accuracy of 0.715, sensitivity of 0.897, specificity of 0.347, PPV of 0.736, NPV of 0.624, F1 score of 0.809, and kappa of 0.275 (Figure 5C-D; Table 1). The corresponding confusion matrices for the test and external validation sets are provided in Supplementary Figure S4.

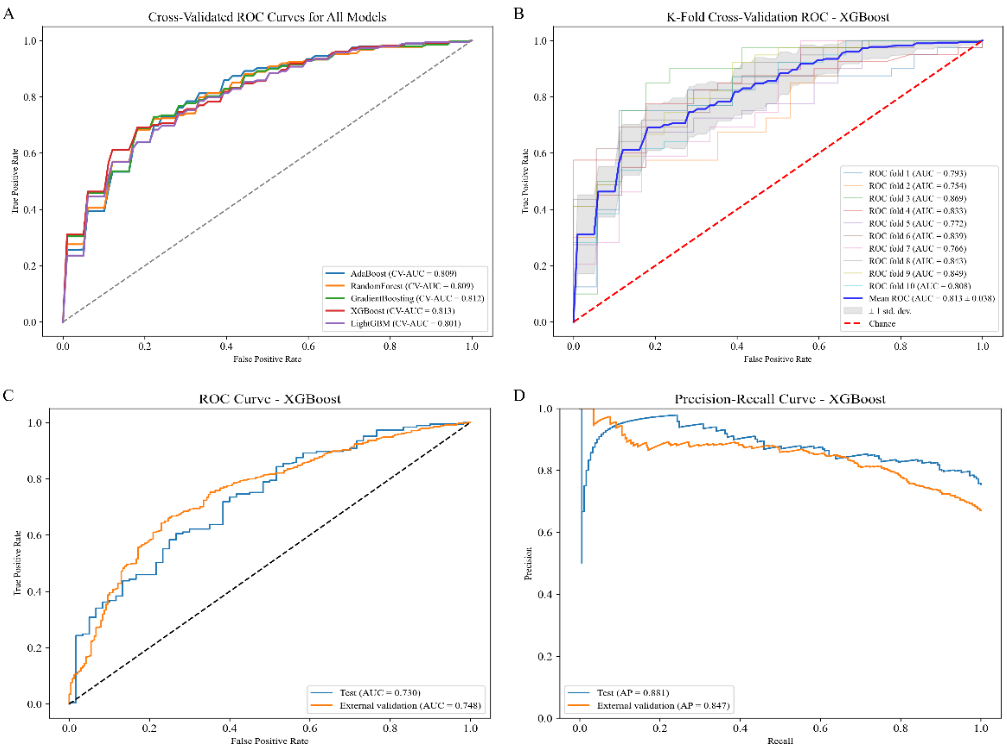

Figure 5. Model performance of the XGBoost model for predicting life satisfaction. AUC: Area under the receiver operating characteristic curve; ROC: Receiver operating characteristic. (A) ROC curves in the test set, with cv-AUC values reported for each model; (B) K-fold cross-validation ROC curves of the best-performing model (XGBoost); (C) ROC curves in the test and external validation sets; (D) Precision-recall curves in the test and external validation sets.

Table 1. Performance metrics for the optimal model in the test and external validation datasets.

Model		Test set	External validation set
AUC-ROC	Happiness [a]	0.831	0.741
	LS [b]	0.730	0.748
Accuracy	Happiness [a]	0.837	0.834
	LS [b]	0.767	0.715
Sensitivity/ Recall	Happiness [a]	0.885	0.921
	LS [b]	0.897	0.897
Specificity	Happiness [a]	0.464	0.363
	LS [b]	0.367	0.347
PPV	Happiness [a]	0.928	0.887
	LS [b]	0.814	0.736
NPV	Happiness [a]	0.342	0.461
	LS [b]	0.537	0.624
F1 score	Happiness [a]	0.906	0.904
	LS [b]	0.853	0.809
Kappa	Happiness [a]	0.302	0.311
	LS [b]	0.296	0.275

Note: a: performance of the Random Forest (best model for happiness); b: performance of the XGBoost (best model for life satisfaction); RF: Random Forest; AdaBoost: Adaptive Boosting; XGBoost: Extreme Gradient Boosting; LightGBM: Light Gradient Boosting Machine; AUC: area under the receiver operating characteristic curve; ROC: Receiver operating characteristic; PPV: positive predictive value; NPV: negative predictive value.

3.5 Model explanation: feature importance and model transparency

This study evaluated the relative importance of various factors in predicting happiness and LS using the selected ML models. We applied SHAP to interpret the contribution of each feature to the predictive model by averaging its absolute mean SHAP values, analyzed separately for the test and external validation sets. First, we used feature importance plots to describe the hierarchical ranking of predictors, where the y-axis lists features in descending order of importance and the x-axis displays their absolute mean SHAP values. Second, we used SHAP summary plots to visually illustrate how each variable impacts the model's prediction output. The SHAP summary plots use the x-axis to display SHAP values, which reflect the influence of each feature on the model's predictions: positive SHAP values (i.e., feature value influence on the right side) indicate that a feature increases the likelihood of feeling happy, while negative values (i.e., feature value influence on the left side) decrease it. The y-axis lists the features in descending order of importance, and each point is color-coded to represent the feature's actual value (i.e., blue indicates the lowest value, and red indicates the highest).

3.5.1 Key predictors of happiness

The key predictors of happiness identified by the Random Forest model are shown in Figure 6A-B. Interpersonal relationships were the most important contributor, with mean SHAP values of 0.123 in the test set and 0.118 in the external validation set. The consistency across datasets indicates the robustness of this finding. Other influential variables included meeting frequency with mother, relationships with father and mother, depression symptoms, future confidence, and the frequency of PA. SHAP summary plots (Figure 6C-D) further revealed that higher interpersonal relationship scores, more frequent meetings with mother, greater future confidence, and higher PA frequency increased the probability of reporting higher happiness, whereas depressive symptoms decreased this probability.

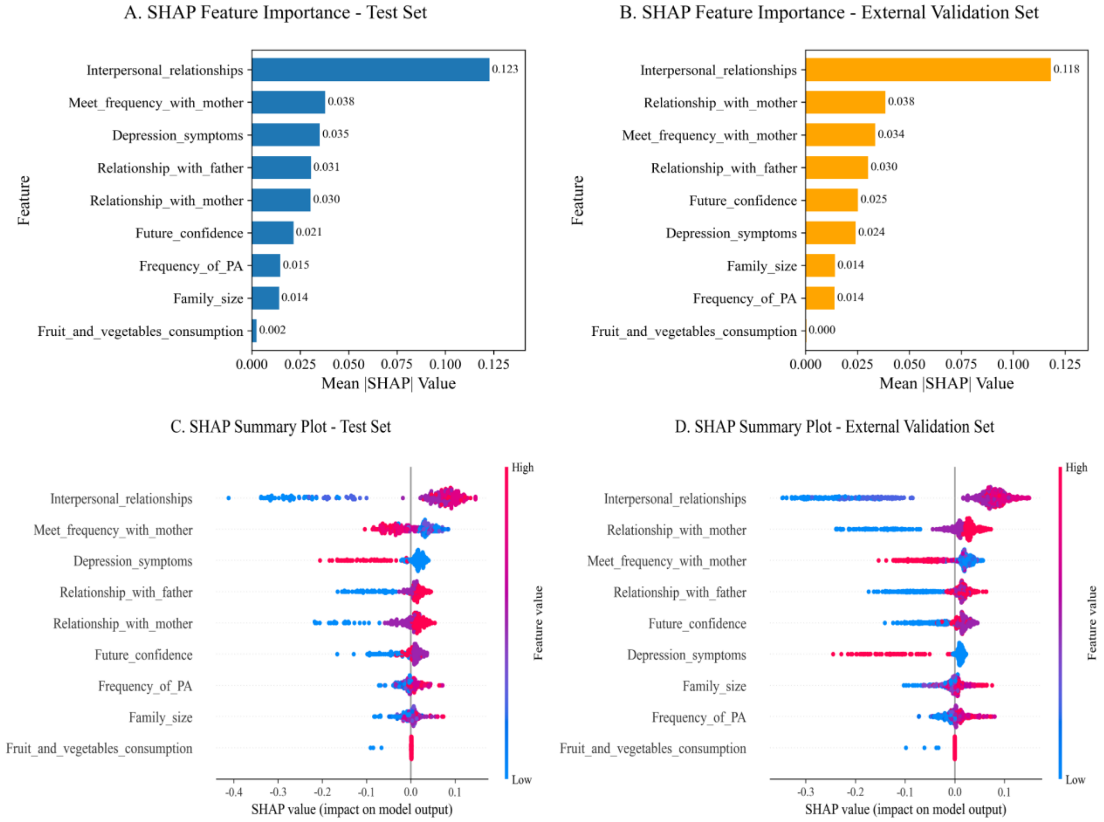

Figure 6. SHAP interpretation of Random Forest for happiness. SHAP: SHapley Additive exPlanations; PA: Physical Activity. Feature importance plots (A-B) for the test and external validation sets. SHAP summary plots (C-D) for the test set and external validation set.

3.5.2 Key predictors of LS

The variables sorted by SHAP values in the XGBoost model for LS are illustrated in Figure 7A-B and in both the test and external validation sets, future confidence, academic self-efficacy, and health emerged as the strongest contributors (mean SHAP values = 0.578, 0.406, and 0.390 in the test set; 0.561, 0.408, and 0.363 in the external validation set, respectively). Other important variables included relationship with mother, WeChat moments sharing frequency, academic stress, and age. The SHAP summary plots (Figure 7C-D) further revealed that higher future confidence, higher academic self-efficacy, better health, a closer relationship with mother, and a lower frequency of WeChat moments sharing were associated with an increased likelihood of reporting higher LS. The consistency between the test and external validation sets confirms the robustness of these findings.

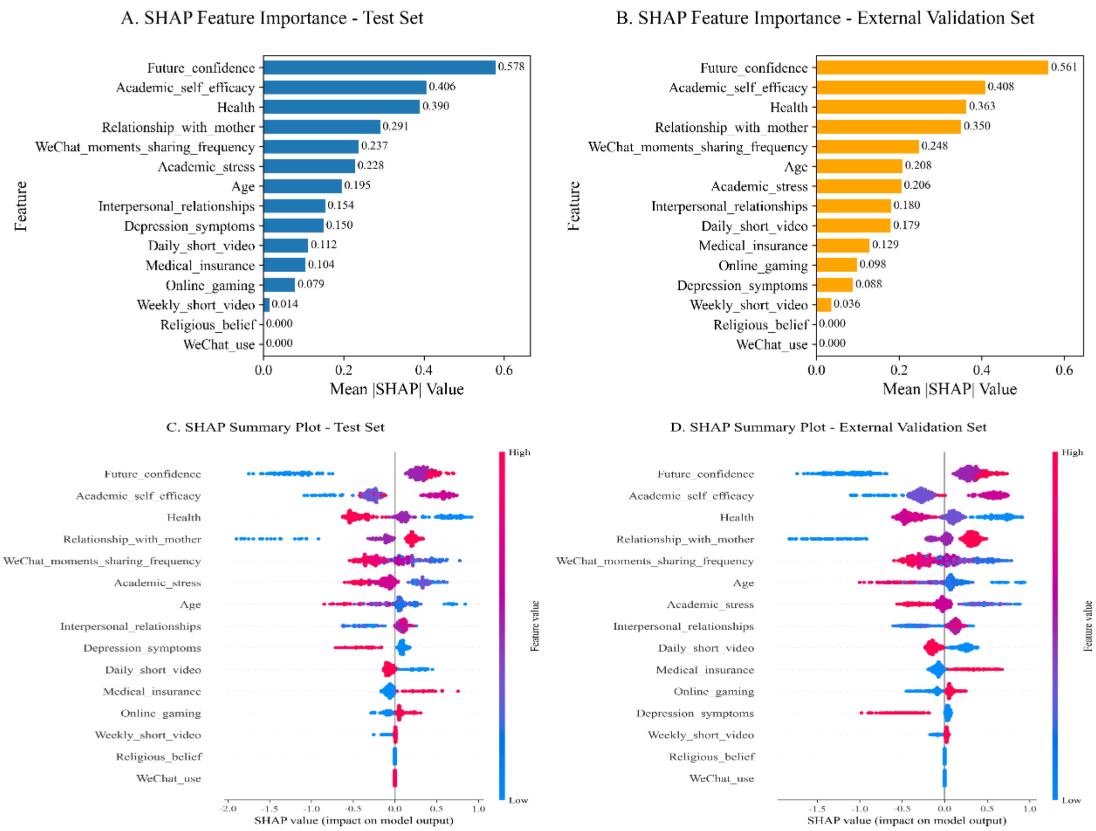

Figure 7. SHAP interpretation of XGBoost for life satisfaction. SHAP: SHapley Additive exPlanations. Feature importance plots (A-B) for the test and external validation sets. SHAP summary plots (C-D) for the test set and external validation set.

3.5.3 Shared predictors of happiness and LS

Furthermore, the Venn diagram (Figure 8) was used to reveal four shared predictors of the optimal prediction model, including future confidence (personality domain), interpersonal relationships and relationship with mother (social support domain), and depressive symptoms (health and functioning domain).

Figure 8. The Venn diagram shows the shared predictors of happiness and life satisfaction

4. DISCUSSION

Guided by an integrated conceptual model of psychology and public health, and using ML techniques, the study simultaneously identified a set of top predictors for happiness and LS. Interpersonal relationships were identified as the top predictor for happiness, and future confidence was the top predictor for LS. Future confidence, depressive symptoms, interpersonal relationships, and the relationship with mother were consistently important across both outcomes. These findings may inform a comprehensive understanding of targeted intervention strategies for SWB among university students.

This finding revealed that interpersonal relationships, representing the social support domain of the conceptual model in our study, were the top predictor of happiness among university students. This result differs from findings from a survey of global adult populations, where broader social-psychological prosperity, such as autonomy and respect, is more influential [13]. Since the predictors of happiness vary across the life course, university students, who are in a transitional life stage, may rely more on peer interactions and social integration for their happiness [63-65]. As university students' social networks expand, their expectations for harmonious interactions grow [66]. Adults, in contrast, may place greater value on autonomy and social respect, as their sense of happiness is often grounded in self-realization and societal recognition [13]. Evidence from Chinese university students further supports this interpretation, showing that supportive relationships could provide emotional resources and foster positive emotions [64, 67], while inadequate relationships might increase vulnerability to mental health problems [68]. Taken together, interpersonal relationships play a particularly important role during university years.

Future confidence, representing the personality domain of the conceptual model in our study, emerged as the top predictor of LS. This finding also differs from research in global adult populations, where LS has been shown to depend more on material prosperity, such as income [13]. These differences also revealed the transitional nature of the university stage, during which academic demands and career uncertainties place particular salience on expectations about the future [69], whereas for adults, with more established roles in work and family [70], their LS is more likely to connect with material conditions [13]. In this context, future confidence may function as a key psychological trait that supports adaptive responses to uncertainty and fosters a more optimistic evaluation of life circumstances [71]. However, the positive role of future confidence may be undermined by external pressures, including intense job competition and high societal expectations, distorting students' perceptions of their future opportunities [69]. Our findings suggest that future confidence might be pivotal in enhancing university students' LS.

Happiness and LS shared four key predictors, including future confidence, interpersonal relationships, depressive symptoms, and the relationship with mother, which fall into personality, health and functioning, and social support of the conceptual model. On the one hand, these findings also contribute to the empirical evidence supporting the conceptual model of psychology and public health. On the other hand, these findings extend previous studies, which have typically reported similar results for these variables concerning either happiness or LS of SWB [72, 73]. This cross-dimensional effect may be understood because these variables shape affective experiences and cognitive evaluations simultaneously. Regarding interpersonal relationships, previous studies have reported that social connections not only generate momentary joy but also help individuals achieve interpersonal goals and gain recognition [73, 74]. Future confidence, while it is important to long-term satisfaction with life, also fosters positive emotions, reduces stress, and promotes a hopeful perspective in the present [72]. Reasonably, less depressed students were happier and more satisfied, as depression may lower positive affect and foster negative cognitive appraisals [75]. Lastly, the enduring influence of mothers reflects the primary identification figure and source of support during the transition from late adolescence to emerging adulthood. This relationship can offer emotional security and stability, making it a source of both daily happiness and LS. Such patterns have been observed in both Eastern and Western university populations [76], showing the cross-cultural relevance of maternal bonds in shaping students' SWB.

4.1 Limitations

Although our study utilized a large-scale, nationally representative dataset of Chinese university students and applied ML methods to identify the top predictors of happiness and LS, based on the conceptual model of psychology and public health perspective, it presents some points for improvement and limitations. First, causal inference among the identified features of our findings is limited because of the cross-sectional data analysis of CFPS. Although CFPS offers rich psychological and public-health indicators, some variables of interest (e.g., dietary habits, online gaming and shopping, watching short videos, and family support) were unavailable in earlier waves, making longitudinal analyses infeasible [28]. Second, the 2020 and 2022 waves were collected during the COVID-19 pandemic, when remote study, social isolation, and stress may have influenced students' happiness and LS [77]. However, we could not explicitly assess pandemic-related factors due to data limitations, the relative importance of these predictors may vary across contexts. Third, happiness and LS were assessed using single-item measures, which may increase measurement error and limit the ability to capture the multidimensional nature of the constructs [35]. Multi-item scales might provide a more detailed understanding of the constructs underlying. Fourth, to address class imbalance, we applied SMOTE-IPF to the happiness model [78], which modestly increased specificity (from 0.179 to 0.464) and slightly changed AUC (from 0.832 to 0.831) in the RF model (Table S7). Although we used a balanced oversampling ratio and ten-fold cross-validation to reduce overfitting, synthetic sampling can still amplify noise.

4.2 Implications

The findings provide two main directions for future research and practice. For research implications, first, future studies could extend these insights by examining related constructs in greater depth, such as identifying which type of personality traits or aspects of interpersonal relationships (e.g., intimate relationships, supervisory relationships, and parental relationships). Second, to enhance the validity and robustness, future research should also employ larger and more balanced datasets and adopt longitudinal, experimental, or causal modelling approaches (e.g., structural equation modelling or modern causal ML techniques). Cross-cultural comparisons would further clarify whether the relative importance of these predictors differs across cultural contexts. Third, our study using ML within a conceptual model may provide an example methodological approach to identify the predictors of happiness and LS across different populations or life stages.

For practice implications, this study provides evidence-based guidance for universities and public health authorities to promote happiness and LS of Chinese university students. For example, interventions should simultaneously cultivate students' confidence in navigating future challenges and strengthen their social connections. In addition, depressive symptoms and the mother-child relationship further reveal the need for regular mental health screening, support services, and family engagement where appropriate.

CONCLUSION

This study provides valuable insights into the top predictors of happiness and LS among Chinese university students based on the conceptual model of psychology and public health. Perhaps decision makers can take cues from the predictors and the importance weight from this study to develop targeted strategies for supporting university students' SWB.

List of abbreviations

SWB: Subjective well-being

LS: Life satisfaction

ML: Machine learning

SHAP: SHapley Additive exPlanations

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

CFPS: China Family Panel Studies

ISSS: Institute of Social Science Survey

PA: Physical activity

LASSO: Least absolute shrinkage and selection operator

SMOTE-IPF: Synthetic Minority Oversampling Technique with Iterative Partitioning Filter

RF: Random Forest

AdaBoost: Adaptive boosting

XGBoost: Extreme Gradient Boosting

LightGBM: Light Gradient Boosting Machine

AUC: Area under the receiver operating characteristic curve

ROC: Receiver operating characteristic

PR: Precision-recall

PPV: Positive predictive value

NPV: Negative predictive value

SD: Standard deviation

IQR: Interquartile range

MI: Multiple imputation

BMI: Body Mass Index

Ethics approval and consent to participate

The studies involving human participants received approval from the Peking University Biomedical Ethics Review Committee (IRB00001052-14010). Written informed consent was obtained from all participants or their legal guardians per ethical guidelines and regulations.

Consent for publication

Not applicable.

Availability of data and materials

The dataset utilized in this paper is publicly accessible via the Peking University Open Research Data Platform. It can be downloaded for research purposes at the following link: https://opendata.pku.edu.cn/dataverse/CFPS?language=en.

Competing interests

The authors declare no conflict of interest.

Funding

This work was supported by the Hunan Province science and technology innovation leading talent project, China (2024RC1003), and the Central South University frontier cross project, China (2023QYJC041).

Author Contributions

Conceptualization, J.G., Q. L., A.A., Y.N.W.; methodology, Y.N.W.; Q. L.; investigation, Q. L.; writing—original draft preparation, Q. L., J.G., A.A., Y.N.W.; F.Z., Z.C., X.T.L.; writing—review and editing, Q. L., J.G., A.A., Y.N.W.; project administration, J.G.; formal analysis, Y.N.W.; Q. L., F.Z., Z.C., X.T.L.; visualization, Q. L., A.A., J.G., Y.N.W.; supervision, J.G.; funding acquisition, J.G., All authors have read and agreed to the published version of the manuscript.

Acknowledgments

We extend our gratitude to Ms Xiaoyu Tang, who provided invaluable assistance with machine learning analysis.

REFERENCES

1. Diener E, Oishi S, Tay L. Advances in subjective well-being research. Nat Hum Behav. 2018;2:253-60. https://doi.org/10.1038/s41562-018-0307-6.

2. Diener E, Oishi S, Lucas RE. Subjective well-being: The science of happiness and life satisfaction. In: Oxford handbook of positive psychology, 2nd ed. New York, NY, US: Oxford University Press; 2009. p. 187-94.
3. Lew B, Huen J, Yu P, Yuan L, Wang D-F, Ping F, et al. Associations between depression, anxiety, stress, hopelessness, subjective well-being, coping styles and suicide in Chinese university students. PLOS ONE. 2019;14:e0217372. https://doi.org/10.1371/journal.pone.0217372.
4. Brett CE, Mathieson ML, Rowley AM. Determinants of well-being in university students: The role of residential status, stress, loneliness, resilience, and sense of coherence. Curr Psychol. 2023;42:19699-708. https://doi.org/10.1007/s12144-022-03125-8.
5. Steptoe A, Deaton A, Stone AA. Psychological well-being, health and ageing. Lancet. 2015;385:640-8. https://doi.org/10.1016/S0140-6736(13)61489-0.
6. Blanchflower DG. Is happiness U-shaped everywhere? Age and subjective well-being in 145 countries. J Popul Econ. 2021;34:575-624. https://doi.org/10.1007/s00148-020-00797-z.
7. Handa S, Pereira A, Holmqvist G. The Rapid Decline of Happiness: Exploring Life Satisfaction among Young People across the World. Applied Research Quality Life. 2023;18:1549-79. https://doi.org/10.1007/s11482-023-10153-4.
8. Tien Nam P, Thanh Tung P, Phuong Linh B, Hanh Dung N, Van Minh H. Happiness among university students and associated factors: A cross-sectional study in Vietnam. J Public Health Res. 2024;13:22799036241272402. https://doi.org/10.1177/22799036241272402.
9. Rogowska AM, Ochnik D, Kuśnierz C, Jakubiak M, Schütz A, Held MJ, et al. Satisfaction with life among university students from nine countries: Cross-national study during the first wave of COVID-19 pandemic. BMC Public Health. 2021;21:2262. https://doi.org/10.1186/s12889-021-12288-1.
10. Jiang X, Ji L, Chen Y, Zhou C, Ge C, Zhang X. How to Improve the Well-Being of Youths: An Exploratory Study of the Relationships Among Coping Style, Emotion Regulation, and Subjective Well-Being Using the Random Forest Classification and Structural Equation Modeling. Front Psychol. 2021;12:637712. https://doi.org/10.3389/fpsyg.2021.637712.
11. Zhang J, Zhao S, Lester D, Zhou C. Life satisfaction and its correlates among college students in China: A test of social reference theory. Asian Journal of Psychiatry. 2014;10:17-20. https://doi.org/10.1016/j.ajp.2013.06.014.
12. Bardo AR. THE COMPARABILITY OF HAPPINESS AND LIFE SATISFACTION: A LIFE COURSE APPROACH.
13. Diener E, Ng W, Harter J, Arora R. Wealth and happiness across the world: material prosperity predicts life evaluation, whereas psychosocial prosperity predicts positive feeling. J Pers Soc Psychol. 2010;99:52-61. https://doi.org/10.1037/a0018066.
14. Barbayannis G, Bandari M, Zheng X, Baquerizo H, Pecor KW, Ming X. Academic Stress and Mental Well-Being in College Students: Correlations, Affected Groups, and COVID-19. Front Psychol. 2022;13:886344. https://doi.org/10.3389/fpsyg.2022.886344.
15. Zalazar-Jaime MF, Moretti LS, Medrano LA. Contribution of Academic Satisfaction Judgments to Subjective Well-Being. Front Psychol. 2022;13. https://doi.org/10.3389/fpsyg.2022.772346.
16. Ding L-L, Ren X-H, Zhu L-J, He L-P, Chen Y, Yao Y-S. Life Satisfaction and Its Relationship With Personality Traits Among Medical College Students in China. Cureus. 2024;16:e57503. https://doi.org/10.7759/cureus.57503.
17. Zhang Z, Wang K, He Z, Qi X. The Relationship Between Physical Activity and Subjective Well-Being in College Students Over the Course of a Semester: A Short-Term Longitudinal Study. Psych J. 2025. https://doi.org/10.1002/pchj.70049.
18. Poldrack RA, Huckins G, Varoquaux G. Establishment of Best Practices for Evidence for Prediction A Review. JAMA Psychiatry. 2020;77:534-40. https://doi.org/10.1001/jamapsychiatry.2019.3671.
19. Crown WH. Potential Application of Machine Learning in Health Outcomes Research and Some Statistical Cautions. Value in Health. 2015;18:137-40. https://doi.org/10.1016/j.jval.2014.12.005.
20. Ali S, Akhlaq F, Imran AS, Kastrati Z, Daudpota SM, Moosa M. The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. Comput Biol Med. 2023;166:107555. https://doi.org/10.1016/j.compbiomed.2023.107555.

21. Osawa I, Goto T, Tabuchi T, Koga HK, Tsugawa Y. Machine-learning approaches to identify determining factors of happiness during the COVID-19 pandemic: retrospective cohort study. BMJ Open. 2022;12:e054862. https://doi.org/10.1136/bmjopen-2021-054862.
22. Zhang N, Liu C, Chen Z, An L, Ren D, Yuan F, et al. Prediction of adolescent subjective well-being: A machine learning approach. Gen Psychiatr. 2019;32:e100096. https://doi.org/10.1136/gpsych-2019-100096.
23. Kılıç AC, Karakuş A, Alptekin E. Prediction of University Students' Subjective Well-Being with Sleep and Physical Activity Data using Classification Algorithms. Procedia Computer Science. 2022;207:2648-57. https://doi.org/10.1016/j.procs.2022.09.323.
24. Kornowicz J, Thommes K. Algorithm, expert, or both? Evaluating the role of feature selection methods on user preferences and reliance. PLOS ONE. 2025;20:e0318874. https://doi.org/10.1371/journal.pone.0318874.
25. Das KV, Jones-Harrell C, Fan Y, Ramaswami A, Orlove B, Botchwey N. Understanding subjective well-being: perspectives from psychology and public health. Public Health Rev. 2020;41:25. https://doi.org/10.1186/s40985-020-00142-5.
26. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Int J Surg. 2014;12:1500-24. https://doi.org/10.1016/j.ijsu.2014.07.014.
27. Xie Y, Hu J. An Introduction to the China Family Panel Studies (CFPS). Chinese Sociological Review. 2014;47:3-29. https://doi.org/10.2753/CSA2162-0555470101.2014.11082908.
28. Zhang L, Lu C, Yi C, Liu Z, Zeng Y. Causal effects and functional mechanisms of the Internet on residents' physical fitness-An empirical analysis based on China family panel survey. Front Public Health. 2023;10:1111987. https://doi.org/10.3389/fpubh.2022.1111987.
29. Zhao Y, Soh KG, Saad HBA, Rong W, Liu C, Wang X. Effects of active video games on mental health among college students: a systematic review. BMC Public Health. 2024;24:3482. https://doi.org/10.1186/s12889-024-21011-9.
30. Lee DC, O'Brien KM, McCrabb S, Wolfenden L, Tzelepis F, Barnes C, et al. Strategies for enhancing the implementation of school-based policies or practices targeting diet, physical activity, obesity, tobacco or alcohol use. Cochrane Database Syst Rev. 2024;12:CD011677. https://doi.org/10.1002/14651858.CD011677.pub4.
31. Li J, Zhou Z, Hao S, Zang L. Optimal intensity and dose of exercise to improve university students' mental health: a systematic review and network meta-analysis of 48 randomized controlled trials. Eur J Appl Physiol. 2025;125:1395-410. https://doi.org/10.1007/s00421-024-05688-9.
32. Li C, Kang L, Miles TP, Khan MM. Factors affecting academic performance of college students in China during COVID-19 pandemic: a cross-sectional analysis. Front Psychol. 2023;14:1268480. https://doi.org/10.3389/fpsyg.2023.1268480.
33. Fan X, Guo X, Ren Z, Li X, He M, Shi H, et al. The prevalence of depressive symptoms and associated factors in middle-aged and elderly Chinese people. Journal of Affective Disorders. 2021;293:222-8. https://doi.org/10.1016/j.jad.2021.06.044.
34. Lei X, Shen Y, Smith JP, Zhou G. Life satisfaction in China and consumption and income inequalities. Rev Econ Household. 2018;16:75-95. https://doi.org/10.1007/s11150-017-9386-9.
35. Raudenská P. Single-item measures of happiness and life satisfaction: the issue of cross-country invariance of popular general well-being measures. Humanit Soc Sci Commun. 2023;10:1-18. https://doi.org/10.1057/s41599-023-02299-1.
36. Bjegovic-Mikanovic V, Wenzel H, Laaser U. Data Mining Approach: What Determines the Well-being of Women in Montenegro, North Macedonia, and Serbia? Front Public Health. 2022;10:873845. https://doi.org/10.3389/fpubh.2022.873845.
37. Zhang C, Liang X. Association between WeChat use and mental health among middle-aged and older adults: a secondary data analysis of the 2020 China Family Panel Studies database. BMJ Open. 2023;13:e073553. https://doi.org/10.1136/bmjopen-2023-073553.
38. Zhou M, Sun X, Huang L. Chronic disease and medical spending of Chinese elderly in rural region. Int J Qual Health Care. 2021;33:mzaa142. https://doi.org/10.1093/intqhc/mzaa142.
39. Wang L, Ren J, Chen J, Gao R, Bai B, An H, et al. Lifestyle choices mediate the association between educational attainment and BMI in older adults in China: A cross-sectional study. Front Public Health. 2022;10:1000953. https://doi.org/10.3389/fpubh.2022.1000953.

40. Man X, Liu J, Bai Y. The Influence of Discrepancies between Parents' Educational Aspirations and Children's Educational Expectations on Depressive Symptoms of Left-Behind Children in Rural China: The Mediating Role of Self-Efficacy. Int J Environ Res Public Health. 2021;18:11713. https://doi.org/10.3390/ijerph182111713.
41. Jiang M, Gao K, Wu Z, Guo P. The influence of academic pressure on adolescents' problem behavior: Chain mediating effects of self-control, parent-child conflict, and subjective well-being. Front Psychol. 2022;13:954330. https://doi.org/10.3389/fpsyg.2022.954330.
42. Chu Y, Aune D, Yu C, Wu Y, Ferrari G, Rezende LFM, et al. Temporal trends in sleep pattern among Chinese adults between 2010 and 2018: findings from five consecutive nationally representative surveys. Public Health. 2023;225:360-8. https://doi.org/10.1016/j.puhe.2023.10.004.
43. Liu X, Wei X, Zhang L. Exploring the Influence of Napping Habits on Job Satisfaction: A Quasi-Natural Experimental Study Based on Longitudinal Data from China. Behav Sci (Basel). 2025;15:770. https://doi.org/10.3390/bs15060770.
44. Ni RJ, Yu Y. Relationship between physical activity and risk of depression in a married group. BMC Public Health. 2024;24:829. https://doi.org/10.1186/s12889-024-18339-7.
45. Zhang X, Zhang Y, Guo B, Chen G, Zhang R, Jing Q, et al. The impact of physical activity on household out-of-pocket medical expenditure among adults aged 45 and over in urban China: The mediating role of spousal health behaviour. SSM Popul Health. 2024;25:101643. https://doi.org/10.1016/j.ssmph.2024.101643.
46. Liu W, Ren Y, Liu J, Loy J-P. The effect of Internet use on adolescent nutritional outcomes: evidence from China. J Health Popul Nutr. 2025;44:138. https://doi.org/10.1186/s41043-025-00856-9.
47. Fang G, Tang T, Zhao F, Zhu Y. The social scar of the pandemic: Impacts of COVID-19 exposure on interpersonal trust. J Asian Econ. 2023;86:101609. https://doi.org/10.1016/j.asieco.2023.101609.
48. Shi H, Zhao H, Ren Z, He M, Li Y, Pu Y, et al. Factors Associated with Subjective Well-Being of Chinese Adolescents Aged 10-15: Based on China Family Panel Studies. International Journal of Environmental Research and Public Health. 2022;19:6962. https://doi.org/10.3390/ijerph19126962.
49. Lu Y, Lin Y-Y, Qu J-Q, Zeng Y, Wu W-Z. Children's internal migration and subjective well-being of older parents left behind: Spiritual or financial support? Frontiers in Public Health. 2023;11.
50. Zhong J, Wu W, Zhao F. The impact of Internet use on the subjective well-being of Chinese residents: From a multidimensional perspective. Front Psychol. 2022;13:950287. https://doi.org/10.3389/fpsyg.2022.950287.
51. Zhang C, Zhu B. Digital gratification: short video consumption and mental health in rural China. Front Public Health. 2025;13:1536191. https://doi.org/10.3389/fpubh.2025.1536191.
52. Chen Y, Zhao Y, Wang Z. The effect of religious belief on Chinese elderly health. BMC Public Health. 2020;20:627. https://doi.org/10.1186/s12889-020-08774-7.
53. Xue X, Cheng M. Social capital and health in China: exploring the mediating role of lifestyle. BMC Public Health. 2017;17:863. https://doi.org/10.1186/s12889-017-4883-6.
54. Blazek K, Van Zwieten A, Saglimbene V, Teixeira-Pinto A. A practical guide to multiple imputation of missing data in nephrology. Kidney International. 2021;99:68-74. https://doi.org/10.1016/j.kint.2020.07.035.
55. Wolfe A. Regulation Techniques for Multicollinearity: Lasso, Ridge, and Elastic Nets.
56. Sun H, Zhang C, Ouyang A, Dai Z, Song P, Yao J. Multi-classification model incorporating radiomics and clinic-radiological features for predicting invasiveness and differentiation of pulmonary adenocarcinoma nodules. BioMedical Engineering OnLine. 2023;22:112. https://doi.org/10.1186/s12938-023-01180-1.
57. Sáez JA, Luengo J, Stefanowski J, Herrera F. SMOTE-IPF: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering. Information Sciences. 2015;291:184-203. https://doi.org/10.1016/j.ins.2014.08.051.
58. Fernandez A, Garcia S, Herrera F, Chawla NV. SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. Journal of Artificial Intelligence Research. 2018;61:863-905. https://doi.org/10.1613/jair.1.11192.

59. Holgado-Apaza LA, Ulloa-Gallardo NJ, Aragon-Navarrete RN, Riva-Ruiz R, Odagawa-Aragon NK, Castellon-Apaza DD, et al. The Exploration of Predictors for Peruvian Teachers' Life Satisfaction through an Ensemble of Feature Selection Methods and Machine Learning. Sustainability. 2024;16:7532. https://doi.org/10.3390/su16177532.
60. Li X, Li C, Wang H, Jiang L, Chen M. Comparison of radiomics-based machine-learning classifiers for the pretreatment prediction of pathologic complete response to neoadjuvant therapy in breast cancer. PeerJ. 2024;12:e17683. https://doi.org/10.7717/peerj.17683.
61. Verbakel JY, Steyerberg EW, Uno H, Cock BD, Wynants L, Collins GS, et al. ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models. J Clin Epidemiol. 2020;126:207-16. https://doi.org/10.1016/j.jclinepi.2020.01.028.
62. Sun J, Sun C-K, Tang Y-X, Liu T-C, Lu C-J. Application of SHAP for Explainable Machine Learning on Age-Based Subgrouping Mammography Questionnaire Data for Positive Mammography Prediction and Risk Factor Identification. Healthcare. 2023;11:2000. https://doi.org/10.3390/healthcare11142000.
63. Chui WH, Wong MYH. Gender Differences in Happiness and Life Satisfaction Among Adolescents in Hong Kong: Relationships and Self-Concept. Soc Indic Res. 2016;125:1035-51. https://doi.org/10.1007/s11205-015-0867-z.
64. Jiang Y, Lu C, Chen J, Miao Y, Li Y, Deng Q. Happiness in University Students: Personal, Familial, and Social Factors: A Cross-Sectional Questionnaire Survey. Int J Environ Res Public Health. 2022;19:4713. https://doi.org/10.3390/ijerph19084713.
65. Zhang J, Zhao S, Deng H, Yuan C, Yang Z. Influence of interpersonal relationship on subjective well-being of college students: The mediating role of psychological capital. PLOS ONE. 2024;19:e0293198. https://doi.org/10.1371/journal.pone.0293198.
66. Sun J, Zhang X, Wang Y, Wang J, Li J, Cao F. The associations of interpersonal sensitivity with mental distress and trait aggression in early adulthood: a prospective cohort study. Journal of Affective Disorders. 2020;272:50-7. https://doi.org/10.1016/j.jad.2020.03.161.
67. Uchida Y, Ogihara Y. Personal or Interpersonal Construal of Happiness: A Cultural Psychological Perspective. International Journal of Well-being. 2012;2.
68. Rudolph KD, Lansford JE, Rodkin PC. Interpersonal Theories of Developmental Psychopathology. In: Developmental Psychopathology. John Wiley & Sons, Ltd; 2016. p. 1-69. https://doi.org/10.1002/9781119125556.devpsy307.
69. Liu K, Liang L, Zheng C, Fei J, Zhang J, Xu J, et al. Future confidence trends in Chinese youth transitioning to adulthood: Role of subjective social status and academic performance. Asian Journal of Social Psychology. 2025;28:e70007. https://doi.org/10.1111/ajsp.70007.
70. Jing S, Li Z, Stanley DMJJ, Guo X, Wenjing W. Work-Family Enrichment: Influence of Job Autonomy on Job Satisfaction of Knowledge Employees. Front Psychol. 2021;12. https://doi.org/10.3389/fpsyg.2021.726550.
71. Stankov L, Kleitman S. Processes on the borderline between cognitive abilities and personality: Confidence and its realism. In: The SAGE handbook of personality theory and assessment, Vol 1: Personality theories and models. Thousand Oaks, CA, US: Sage Publications, Inc; 2008. p. 545-59. https://doi.org/10.4135/9781849200462.n26.
72. Mamani-Benito O, Carranza Esteban RF, Caycho-Rodríguez T, Castillo-Blanco R, Tito-Betancur M, Alfaro Vásquez R, et al. The influence of self-esteem, depression, and life satisfaction on the future expectations of Peruvian university students. Front Educ. 2023;8. https://doi.org/10.3389/feduc.2023.976906.
73. Segrin C, Taylor M. Positive interpersonal relationships mediate the association between social skills and psychological well-being. Personality and Individual Differences. 2007;43:637-46. https://doi.org/10.1016/j.paid.2007.01.017.
74. Okita DO. State of Interpersonal Relationships of Freshmen at Universities. In: Aloka PJ, editor. Utilising Positive Psychology for the Transition into University Life. Cham: Springer Nature Switzerland; 2024. p. 123-43. https://doi.org/10.1007/978-3-031-72520-3_8.
75. Seo EH, Kim S-G, Kim SH, Kim JH, Park JH, Yoon H-J. Life satisfaction and happiness associated with depressive symptoms among university students: a cross-sectional study in Korea. Annals of General Psychiatry. 2018;17:52. https://doi.org/10.1186/s12991-018-0223-1.

76. De Coninck D, Matthijs K, Luyten P. Subjective well-being among first-year university students: A two-wave prospective study in Flanders, Belgium. SUCCESS. 2019;10:33-45. https://doi.org/10.5204/ssj.v10i1.642.
77. Song Y, Wang L, Liu Y. Excessive or reduced rest-day sleep compensation linked to depression in Chinese adults. Front Psychiatry. 2025;16:1601613. https://doi.org/10.3389/fpsyt.2025.1601613.
78. Lee YW, Choi JW, Shin E-H. Machine learning model for predicting malaria using clinical information. Computers in Biology and Medicine. 2021;129:104151. https://doi.org/10.1016/j.compbiomed.2020.104151.