Liu et al. BMC Public Health (2025) 25:264
https://doi.org/10.1186/s12889-025-21506-z

BMC Public Health

RESEARCH

Open Access

Identification of depressive symptoms in adolescents using machine learning combining childhood and adolescence features

Xinzhu Liu[1†], Rui Cang[1†], Zihe Zhang[2†], Ping Li[3], Hui Wu[4], Wei Liu[1*] and Shu Li[1*]

Abstract

Background Depressive symptoms in adolescents can significantly affect their daily lives and pose risks to their future development. These symptoms may be linked to various factors experienced during both childhood and adolescence. Machine learning (ML) has attracted substantial attention in the field of adolescent depression; however, studies establishing prediction models have primarily considered childhood or adolescent features separately, resulting in a lack of analyses that incorporate factors from both stages.

Methods We collected 39 features related to childhood and adolescence. Using the maximum relevance-minimum redundancy method and four ML algorithms, we determined the optimal feature subset for identifying depressive symptoms and constructed child-adolescent models. Stepwise logistic regression and four ML methods were employed to create demographic and combined models, respectively. The performance of each model was evaluated using a test set, and SHapley Additive exPlanations (SHAP) were utilized to interpret the models' prediction results.

Results The proposed child-adolescent models exhibited superior performance on the test set than the demographic and combined models (AUC: 0.835–0.879 versus 0.530 and 0.840–0.876, respectively). The optimal feature subset included two childhood features (relationship quality with peers and parental absence) and four adolescence features (social trust, academic pressure, importance of the internet for entertainment, and positive parenting behaviour). These features were found to be more effective than demographic characteristics in distinguishing depressive symptoms in adolescents.

Conclusions This study demonstrates the correlation between adolescence depressive symptoms and specific factors from both childhood and adolescence, as well as the potential of ML to predict it. These findings may serve as a reference for future intervention studies.

†Xinzhu Liu, Rui Cang and Zihe Zhang contributed equally to this work and should be regarded as co-first authors.

*Correspondence:
Wei Liu
wliu@cmu.edu.cn
Shu Li
lishu@cmu.edu.cn

Full list of author information is available at the end of the article

Keywords Child-adolescent model, Depression, Adolescence, Childhood

Introduction

Patients with depressive symptoms typically exhibit a low mood, loss of interest and pleasure, and lack of energy or fatigue [1], which can seriously affect their daily lives. According to data from the World Health Organization, approximately one-seventh (14%) of the global adolescent population suffers from mental health problems, many of which remain undetected or untreated [2]. In China, approximately 14.8% of adolescents are at varying degrees of risk for depression [3]. Adolescence is a critical period for rapid social, emotional, and cognitive development and is essential for an effective transition to adulthood [4]. If adolescent depression symptoms are not appropriately addressed, these individuals are at higher risk of developing adult depression and suicidal behaviour, which can have life-threatening consequences [5]. Identifying the factors influencing depressive symptoms in adolescents is crucial for developing effective intervention measures and treatment plans.

Social interpersonal relationships, family dynamics, and school environments may play a significant role in the development of adolescent depression. Factors impacting adolescents' mental health include family economic conditions [6], parents' physical and mental health [7], adolescents' relationships with their parents [8], academic achievements and pressures [9], social support, and internet use [10, 11, 12]. Additionally, certain personal traits and childhood environments can favour depressive symptoms in adolescents. Some studies have indicated that childhood loneliness, the quality of family relationships, early trauma, and experiences of loss [13, 14, 15, 16] can predict psychological problems in adolescence. Clearly, childhood and adolescence are critical periods of psychological development, and early intervention can effectively prevent the onset of depression, reduce long-term effects, improve quality of life, and lower treatment costs. However, the causes of adolescent depressive symptoms have not been fully elucidated. Therefore, identifying and addressing relevant factors from both adolescence and childhood is essential to prevent potential depressive symptoms.

Previous studies have predominantly utilised traditional statistical methods to analyse the relationship between individual or specific features and depression, resulting in relatively simple findings. However, the factors influencing adolescent depressive symptoms may be multifaceted and complex. Machine learning (ML) offers the ability to identify potential influencing factors by analysing large datasets, effectively addressing multicollinearity issues, and providing more objective, reliable, and repeatable predictive models [17]. Research applying ML methods in the field of adolescent depression has generally focused on features related to either childhood or adolescence separately to develop prediction models [6, 18]. There is a notable lack of research analysing features from both childhood and adolescence concurrently. A comprehensive analysis of these two stages can enhance understanding of how depressive symptoms evolve and accumulate over time and help identify potential risk factors earlier.

This study employs ML techniques to concurrently analyse features from both childhood and adolescence, providing a more holistic understanding of depressive symptom development. It aims to identify the key factors contributing to the onset of depressive symptoms in adolescents and offers empirical evidence to inform strategies for mitigating depressive psychological outcomes among such population.

Methods

Data and samples

We derived a dataset from the China Family Panel Studies (CFPS), a nationwide, large-scale, multidisciplinary social survey organised and implemented by the China Social Science Survey Center of Peking University [19]. The CFPS surveyed sample households using the multistage probability proportional sampling method, covering 25 provinces.

The present study utilised the most recent 2020 adolescent survey data, which offers comprehensive information on the features of 13–15-year-olds. To identify relevant childhood features, we traced participants back to 2014, when they were 7–9 years old. For cases with missing data from 2014, we used available data from 2012, when participants were 6–7 years old.

We excluded respondents based on the following criteria: (1) missing data on adolescents' depression scores, (2) missing data on demographic characteristics, (3) missing data on features related to adolescence, (4) unmatched samples between 2020 and 2014 or 2012, (5) missing data on features related to childhood and (6) diagnosed with a history of depression.

After screening, we included data concerning 505 adolescents for analysis. Figure 1 presents a detailed flowchart of this study.

Dependent variables

This study used the eight-item Center for Epidemiological Studies Depression Scale (CES-D) to calculate depression scores and screen for depressive symptoms [20]. The CES-D assesses the frequency of the following situations occurring in the past week: (1) I feel emotionally low, (2)

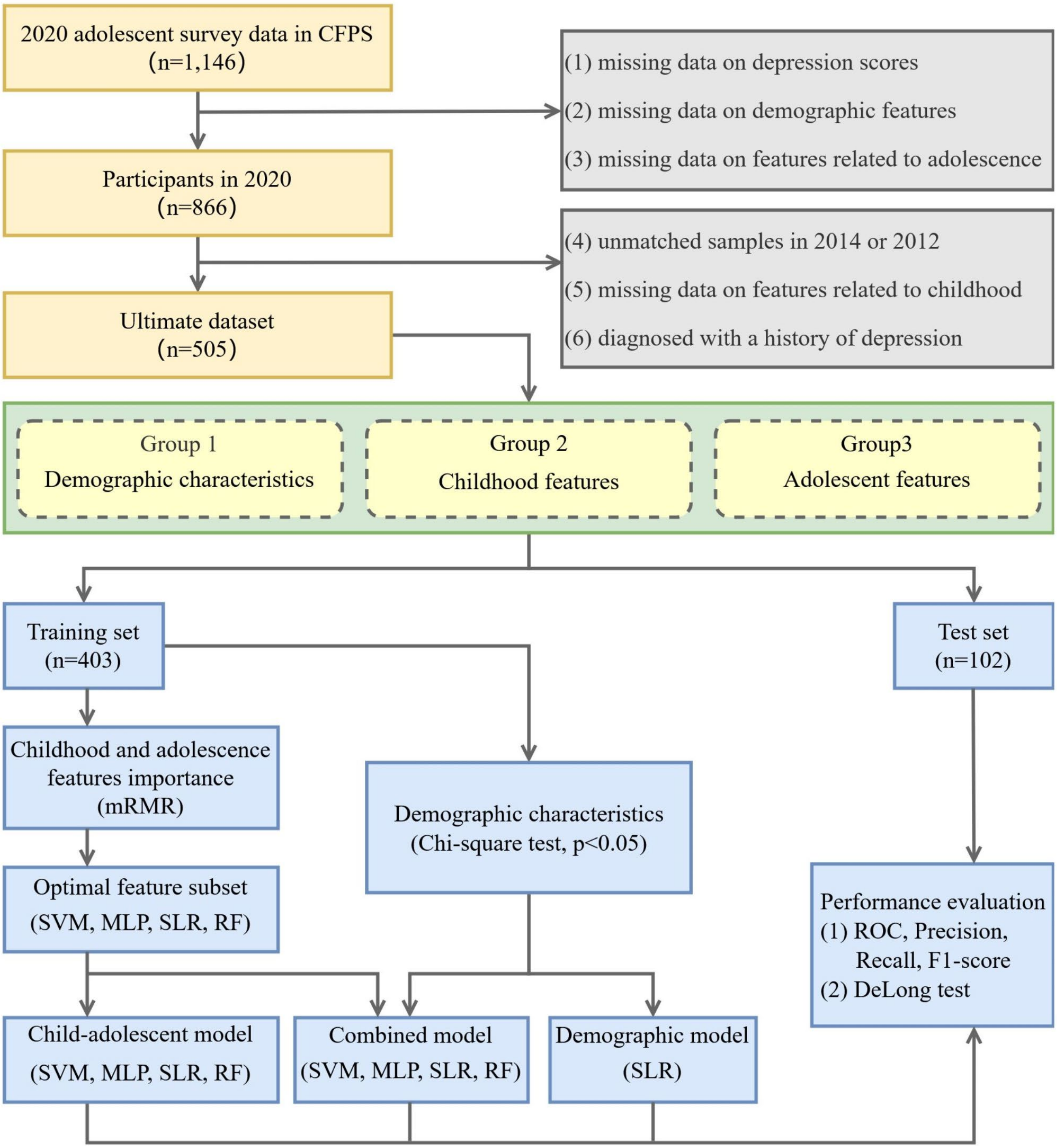

Fig. 1 Flowchart of this study

I find it difficult to do anything, (3) I have poor sleep, (4) I feel happy, (5) I feel lonely, (6) I live happily, (7) I feel sad and find it difficult to get through, and (8) I feel that life cannot continue. The coded response options were as follows: 0 = rarely or never (< 1 day), 1 = sometimes or a little (1–2 days), 2 = occasionally or a moderate amount of time (3–4 days), and 3 = most or all of the time (5–7 days), with items (4) and (6) being reverse-scored. The total score ranged between 0 and 24, with a cut-off point of 9 used to generate variables for depressive symptoms (non-depressive symptoms, total score < 9; depressive symptoms, total score ≥ 9) [21]. The Cronbach's α value for the sample was 0.852 (see Additional file 1).

Independent variables

Demographic characteristics

These comprised the number of family members, gender, age, self-assessment of health status, and residence.

Childhood features

These consisted of three groups, encompassing a total of nine specific features. (1) Family relationships: parental absence during childhood; (2) child behaviour and lifestyle: positive personality, relationship quality with peers, and independence; and (3) parent behaviour: care for children's education, punish children for poor academic work during childhood, parental care, restriction on watching TV, and educational expectations during childhood.

Some of them were obtained through calculations from multiple perspectives. Parental absence during childhood feature included "How long do you live with your father in a year" and "How long do you live with your mother?". Positive personality feature was calculated based on three questions: "Children are naturally optimistic", "They will think carefully before acting, not impulsive", and "Children are easy to overcome irritability". Relationship quality with peers was calculated based on three questions: "Children get along well with peers", "They enjoy helping others", and "Children are liked by peers". Parental care feature was based on the frequency of occurrence of the following four questions: "Proactively communicate with children", "Discuss school matters with children", "Require children to complete homework", "Check children's homework". Restriction on watching TV feature was based on the frequency of occurrence of the following two questions: "prevent children from watching TV", "restrict the programs that children watch".

Adolescent features

These consisted of five groups, encompassing a total of thirty specific features. (1) Child behaviour and lifestyle: internet use, importance of the internet for entertainment, importance of the internet for social engagement, importance of the internet for learning, physical exercise, midday napping, nighttime sleep duration, quarreling with parents, heart-to-heart talks with parents, sharing troubles with others, pocket money, and positive behaviour; (2) subjective perception: social support, subjective well-being, social trust, neighbourhood trust, and self-acceptance; (3) family relationships: family connectedness, and parental absence during adolescence; (4) education and learning: non-weekend study hours, weekend study hours, class leader, academic performance, academic pressure, school satisfaction, skill cultivation, and extracurricular tutoring class; and (5) parental behaviour: positive parenting behaviour, punish children for poor academic work during adolescence, and educational expectations during adolescence.

Positive parenting behaviour was calculated based on the following 11 questions: (1) When you do something wrong, parents will ask you the reasons clearly. (2) Encourage you to work hard. (3) Be friendly when talking to you. (4) Encourage you to think independently. (5) When parents ask you to do something, they will tell you the reasons for doing it. (6) Parents like to talk to you. (7) Parents will ask you about the school situation. (8) Parents tell you stories. (9) Parents play with you. (10) Parents praise you. (11) Parents attend parent teacher conferences held by the school. Extracurricular tutoring class is composed of two questions: "Did you attend the tutoring class" and "Did you hire a tutor?". Parental absence during adolescence included "How long do you live with your father in a year" and "How long do you live with your mother?". Positive behaviour was calculated the total score for the following four questions: (1) The child is focused, (2) The child follows rules and regulations, (3) Once started, it must be completed, and (4) The child likes things to be neatly arranged.

For detailed descriptions of all features used in this study, see Additional File 2.

Model establishment

We constructed demographic, child-adolescent, and combined models. The training set ($n = 403$) and test set ($n = 102$) were obtained by randomly dividing all data in an 8:2 ratio. A chi-square test was used to analyse the demographic characteristics, with $p < 0.05$ considered statistically significant. Borderline-SMOTE, an oversampling method that focuses on boundary samples and generates more discriminative composite samples than other algorithms, was used to address the imbalance in the proportion of samples with and without depressive symptoms. Additionally, features were normalized using z-scores before model construction.

The analysis of childhood and adolescence features involved three steps. First, in the training set, we applied the maximum relevance minimum redundancy (mRMR) algorithm for feature ranking and preliminary screening. The mRMR algorithm maximizes the correlation between independent variables and the target variable while minimizing redundancy among independent variables [22]. We used the F-statistic to assess the correlation between features and adolescent depressive symptoms and the Pearson correlation coefficient to evaluate feature redundancy. The top 20 features were then selected. Second, based on the ranking, features were included one by one in four ML methods: Support Vector Machine (SVM), Multilayer Perceptron (MLP), Random Forest (RF), and Stepwise Logistic Regression (SLR). A 5-fold cross-validation strategy was employed to further screen features and construct models. Finally, we evaluated the performance of the four ML models. The optimal feature subset was determined as the one with the maximum area under the curve (AUC) and the fewest number of features for predicting depressive symptoms. Models constructed based on this optimal feature subset were identified as child-adolescent models.

Using the important demographic characteristics in the training set and the optimal feature subset, we constructed combined models with the four ML algorithms mentioned above. Additionally, we used SLR to develop a demographic model.

We adjusted the hyperparameters of the RF, MLP, and SVM algorithms to enhance predictive performance. For the RF model, the random search method was used to optimise parameters and prevent overfitting, including the number of estimators (80 to 120, in increments of 5), maximum tree depth (2, 5, or 8), minimum samples required to split a node (3, 5, or 10), and minimum samples per leaf node (5, 10, or 15). The cross-validation grid search method (GridSearchCV) was utilized to identify the optimal hyperparameters for MLP models, including regularization strength (1e-05, 1e-03, 1e-01, 1e + 01, 1e + 03, 1e + 05), number of layers, and number of weights per layer ([50,100,50], [10,20,50,100], [100,50,25,10]). GridSearchCV was also applied to optimise SVM model parameters, including γ (16 values between 1e-05 and 1e + 05), C (16 values between 1e-06 and 1e + 06), and the kernel function (radial basis, linear, and sigmoid functions). The optimised hyperparameter values for each model are detailed in Additional File 3.

Model evaluation and interpretation

In the test set, we assessed the predictive performance of all models using precision, recall, F1 score, and the AUC value. We compared the receiver operating characteristic (ROC) curves of each model using the DeLong test, a widely used hypothesis-testing method for comparing differences in AUC values and evaluating classifier performance [23].

To explain the prediction results of the optimal models, we utilised the SHapley Additive exPlanations (SHAP) method. SHAP offers an intuitive visualization of the complex relationships between features and prediction outcomes [24]. We generated SHAP summary and feature importance plots to gain a deeper understanding of the model and explain its decision-making process.

Data analysis

Data processing and analyses were conducted using Python 3.7. The mrmr package was used for the mRMR algorithm, the sklearn package was used for the other algorithms, and the Statsmodels package was used for SLR. Additionally, the Scipy package was used for chi-square tests. SHAP summary and feature importance plots were generated using SHAP package and explainer, and Stata 17 was employed for survey data cleaning. The statistical significance level was set at $p < 0.05$.

Results

Demographic characteristics

Table 1 presents the detailed demographic characteristics analysis results. Amongst the 505 participants, 63 (12.9%)

Table 1 Analysis of with and without depressive symptoms, according to demographic characteristics

Demographics characteristics	Total (%)	Training set($n = 403$)			Test set($n = 102$)		
		Without depressive symptoms (%)	With depressive symptoms (%)	*p*-value	Without depressive symptoms (%)	With depressive symptoms (%)	*p*-value
Age (years)				0.022*			0.254
≤ 14	272(53.8)	197(56.8)	22(39.3)		45(50.6)	8(72.7)	
> 14	233(46.2)	150(43.2)	34(60.7)		46(49.4)	3(27.3)	
Gender				0.165			> 0.999
Man	261(51.6)	187(53.9)	24(42.9)		45(49.5)	5(45.4)	
Woman	244(48.4)	160(46.1)	32(57.1)		46(50.5)	6(54.6)	
Residence				0.460			0.677
City	218(43.2)	152(43.8)	21(37.5)		39(42.9)	6(54.6)	
Village	287(56.8)	195(56.2)	35(62.5)		52(57.1)	5(45.4)	
Number of family members				> 0.999			0.606
≤ 5	337(66.7)	228(65.7)	37(66.1)		63(69.2)	9(81.8)	
> 5	168(33.3)	119(34.3)	19(33.9)		28(30.8)	2(18.2)	
Self-assessment of health status				0.002*			0.006*
Unhealth	3(0.5)	2(0.6)	0(0.0)		1(1.1)	0(0.0)	
General health	5(0.9)	0(0.0)	2(3.6)		1(1.1)	2(18.2)	
Health	497(98.6)	345(99.4)	54(96.4)		89(97.8)	9(81.8)	

*$P < 0.05$

exhibit depressive symptoms. There are significant differences in depressive symptoms in adolescents of different ages and health conditions within the training set, indicating that older adolescents or those with poorer health conditions are more likely to experience depressive symptoms. No significant differences are observed across variables when comparing the baseline characteristics of the test and training groups [see Additional file 4].

Optimal feature subset

We identified the optimal feature subset when the number of features was six, at which point the AUC values of all four ML models reached their maximum. This subset included two childhood features, namely, relationship quality with peers and parental absence; and four adolescent features, namely, social trust, academic pressure, importance of the internet for entertainment, and positive parenting behaviour. The rankings of childhood and adolescent features obtained from the mRMR analysis are provided in Additional File 5.

Model evaluation

Table 2; Fig. 2 present the evaluation indicators and ROCs for the demographic, child-adolescent, and combined models, respectively.

In the test set, all four child-adolescent models achieve good predictive performance, far superior to the demographic model. Amongst these, the SLR model is the most stable, with an AUC of 0.879. Incorporating two demographic characteristics (age and self-assessment of health status) into the optimal feature subset results in a decrease in performance for the combined models (AUC: 0.835–0.879 vs. 0.840–0.876). These findings suggest that the optimal feature subset plays a more critical role in predicting adolescent depressive symptoms than demographic characteristics.

The DeLong test indicates significant differences in ROC values between the child-adolescent and demographic models, as well as between the combined and demographic models ($p < 0.001$). However, we observe no significant difference between the child-adolescent and combined models ($p = 0.112–0.375$).

Model interpretation

Owing to the optimal predictive performance of the child-adolescent models, we further analysed them. Figure 3 illustrates that the models generated by the four ML algorithms show some differences in feature importance rankings and the impact of each feature on predictions. Despite these differences, we observe commonalities amongst them. The three most influential features in each model are social trust, academic pressure, and positive parenting behaviour. Additionally, social trust, academic pressure, and importance of the internet for entertainment are positively correlated with depressive symptoms. In contrast, strong relationship quality with peers, parental absence during childhood, and positive parenting behaviour are negatively correlated with depressive symptoms.

Discussion

This study is innovative in being the first to employ multiple ML methods to identify depressive symptoms in adolescents using features from both childhood and adolescence. By integrating feature selection with model construction, we identified key influencing factors. Our

Table 2 Performance of all models predicting depressive symptoms on the training and test sets

Data sets	Models	Precision	Recall	F1-score	AUC
Training set	Demographic	0.614 ± 0.03	0.613 ± 0.03	0.612 ± 0.03	0.628 ± 0.03
	Child-adolescent_RF	0.806 ± 0.06	0.762 ± 0.09	0.750 ± 0.09	0.884 ± 0.02
	Child-adolescent_SLR	0.838 ± 0.01	0.834 ± 0.00	0.833 ± 0.00	0.879 ± 0.01
	Child-adolescent_SVM	0.857 ± 0.01	0.852 ± 0.01	0.851 ± 0.01	0.885 ± 0.01
	Child-adolescent_MLP	0.855 ± 0.02	0.851 ± 0.01	0.850 ± 0.01	0.907 ± 0.02
	Combined_RF	0.795 ± 0.04	0.792 ± 0.04	0.792 ± 0.04	0.861 ± 0.01
	Combined_SLR	0.834 ± 0.01	0.830 ± 0.01	0.829 ± 0.01	0.882 ± 0.01
	Combined_SVM	0.841 ± 0.02	0.837 ± 0.02	0.836 ± 0.02	0.886 ± 0.01
	Combined_MLP	0.850 ± 0.03	0.847 ± 0.03	0.847 ± 0.03	0.901 ± 0.02
Test set	Demographic	0.798	0.495	0.591	0.530
	Child-adolescent_RF	0.872	0.763	0.803	0.835
	Child-adolescent_SLR	0.913	0.845	0.868	0.879
	Child-adolescent_SVM	0.920	0.814	0.846	0.876
	Child-adolescent_MLP	0.889	0.784	0.820	0.866
	Combined_RF	0.879	0.804	0.832	0.840
	Combined_SLR	0.900	0.845	0.865	0.864
	Combined_SVM	0.907	0.814	0.845	0.876
	Combined_MLP	0.875	0.784	0.817	0.857

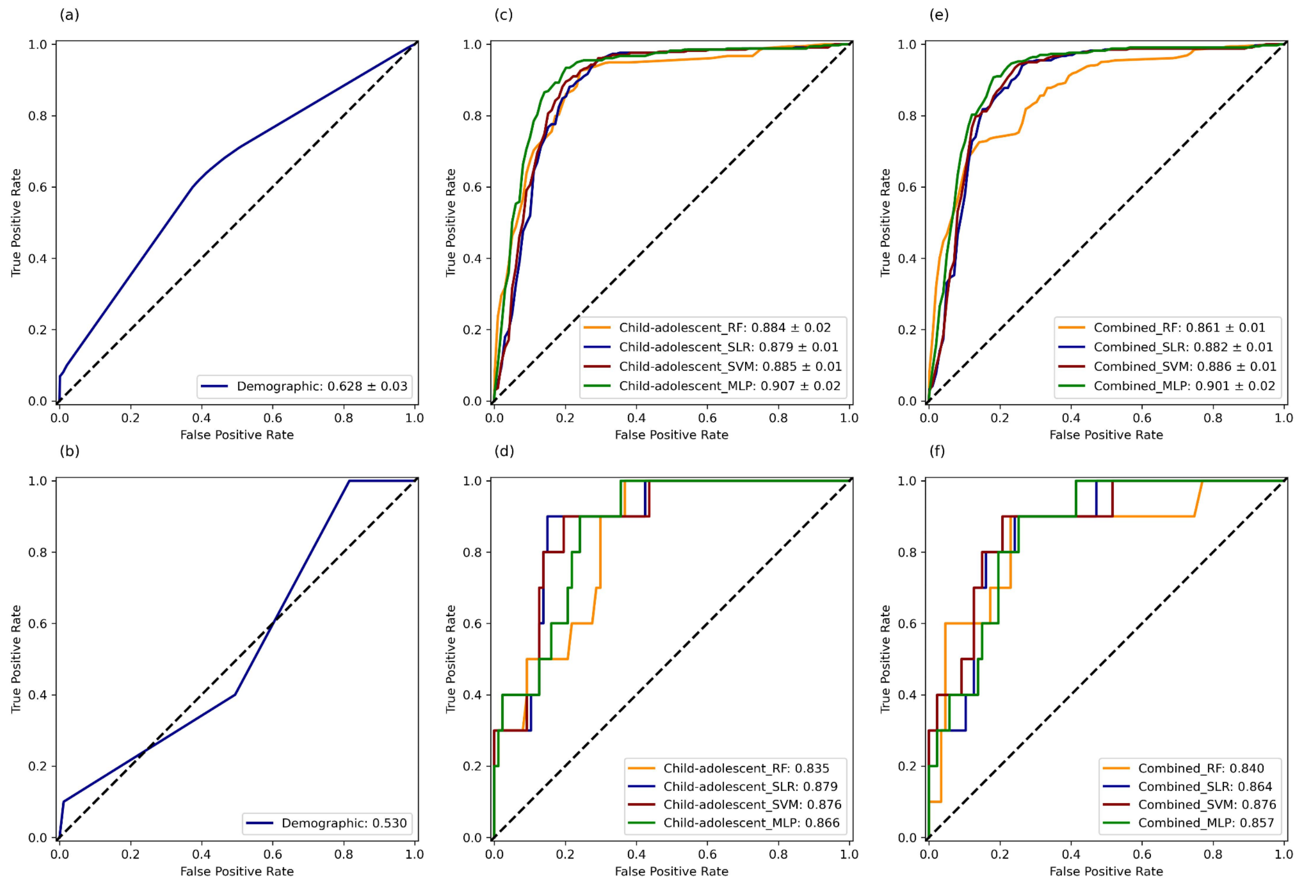

Fig. 2 ROC curves for all models. (**a**), (**b**): Demographic models on the training and test sets; (**c**), (**d**): Child-adolescent models on the training and test sets; (**e**), (**f**): Combined models on the training and test sets

findings underscore the robust performance of ML in predicting adolescent depressive symptoms and provide valuable insights for developing targeted intervention and prevention strategies. Additionally, this study fills a gap in the combined prediction of adolescent depression based on two-stage characteristics.

In this study, the prevalence of depressive symptoms in adolescents was 12.9%. Adolescence is a period of significant changes in emotional behaviour and related neural circuits [5], which enhances the risk of depression. The findings also suggest that adolescents with poor health are more likely to experience depressive symptoms, consistent with previous studies [25]. Additionally, this study found that older adolescents were more likely to experience depressive symptoms.

Using traditional logistic regression methods, one study reported that negative childhood experiences could adversely impact adolescent mental health [26]. Some studies have utilised ML methods to identify factors influencing adolescent depression during the outbreak of the new crown pneumonia, with prediction models achieving AUC values of 0.770–0.857 [27, 28]. Muhammad et al. developed four ML models to predict depressive symptoms in international students, yielding AUC values of 0.494–0.842 [29]. These studies confirmed that the home–school environment, sleep quality, and social relationships during adolescence can affect adolescents' psychological health. However, none of these studies considered features from both childhood and adolescence. This study utilised four ML algorithms and two-stage variables to identify predictive factors for adolescent depressive symptoms. Previous research has validated the effectiveness of ML models in detecting mental disorders in children and adolescents [30]. According to a widely used empirical rule, ideally each predictor variable requires at least 10 samples to support it [31]. Our sample size ($n = 505$) and the 39 features related to childhood and adolescence satisfy this requirement. Notably, we applied the mRMR method for dimensionality reduction before model construction, which enhanced the proportion between independent variables in the input model and the sample size. The mRMR method is effective in situations with a small sample size but many independent variables, and it can screen large datasets, such as genetic and omics data [32, 33]. The child-adolescent models developed using the four ML methods

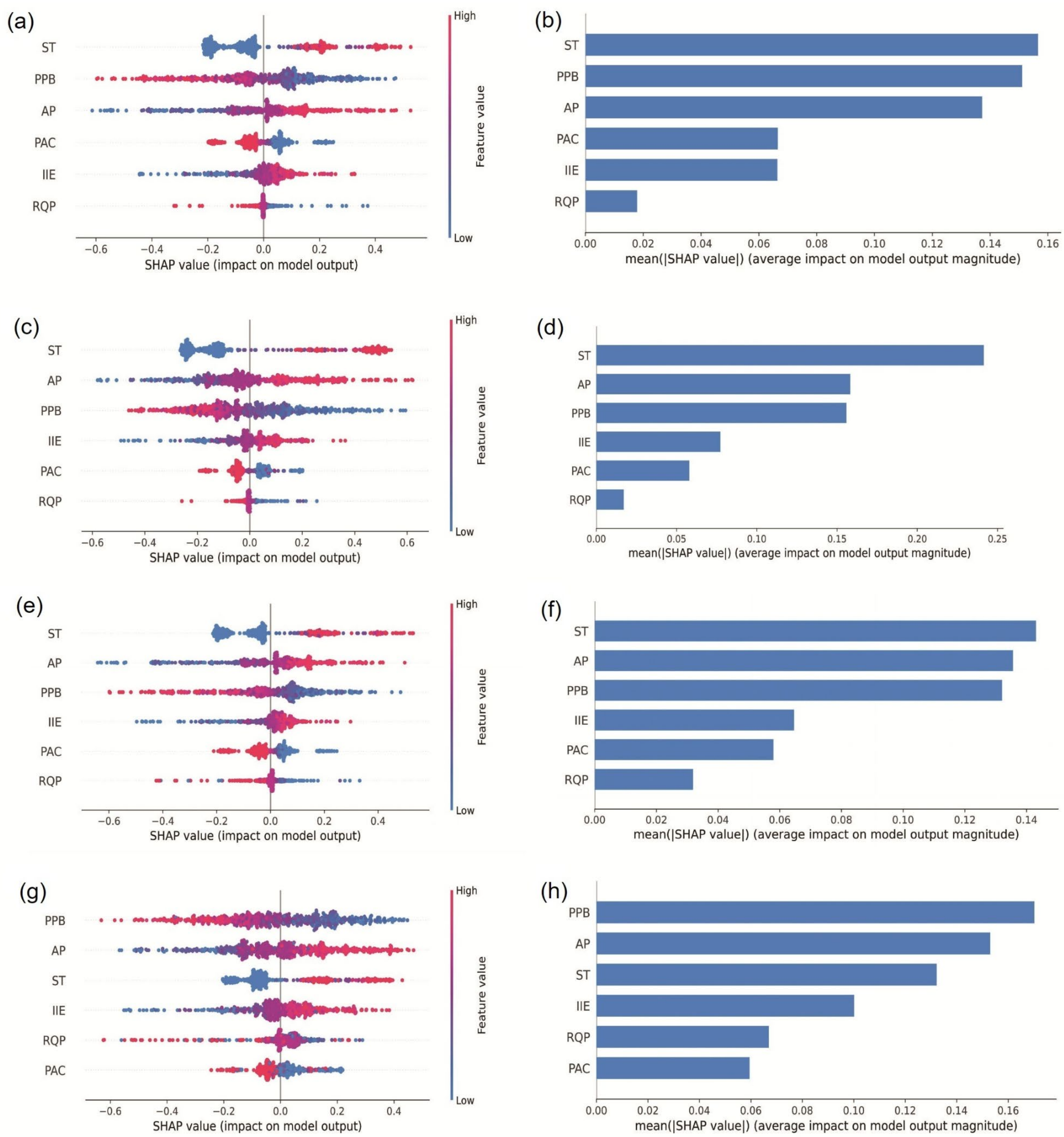

Fig. 3 SHAP summary plots and feature importance visualizations for the child-adolescent models. (**a**), (**b**): SHAP summary and feature importance plots for the RF model; (**c**), (**d**): SHAP summary and feature importance plots for the SLR model; (**e**), (**f**): SHAP summary and feature importance plots for the SVM model; (**g**), (**h**): SHAP summary and feature importance plots for the MLP model. Abbreviations: ST: social trust; PPB: positive parenting behaviour; AP: academic pressure; PAC: parental absence during childhood; IIE: importance of the internet for entertainment; RQP: relationship quality with peers

demonstrated stable and robust predictive performance, achieving a maximum AUC of 0.879 in the test set. However, incorporating demographic characteristics into the models resulted in a decreased AUC, suggesting that these demographic features were not crucial for predicting adolescent depressive symptoms. The key influencing factors identified for adolescent depression include two childhood features (relationship quality with peers and parental absence) and four adolescent features (social trust, academic pressure, importance of the internet for entertainment, and positive parenting behaviour).

Certain childhood behaviours can influence mental health during adolescence and even into adulthood [6, 34]. This study found that positive relationship quality with peers during childhood and living with parents reduced the likelihood of developing depressive symptoms during adolescence. This study measured relationship quality with peers in childhood across three dimensions: getting along well with peers, being helpful, and being liked by peers. These factors contribute positively to the development of children's personality and cognition. Positive interactions with peers enhance social confidence, while children who are helpful gain a sense of achievement by assisting others [35]. Additionally, receiving positive social feedback during childhood positively impacts social and mental health in both adolescents and adults. A lack of peer interaction can increase loneliness [36] and reduce feelings of social belonging amongst adolescents [11]. This deficiency not only limits the development of social skills but may also lead to feelings of isolation and social disconnection, thereby increasing emotional stress when facing challenges. Without supportive peer networks, teenagers may be more susceptible to depression due to their limited experience and skills in managing social and emotional stress.

In this study, we investigated the impact of parental absence on children and adolescents. Our findings indicate that parental absence during childhood is closely related to adolescent depression. A previous study identified early parent–child separation as a favourable indicator for predicting depression in adolescents and adults; it increases the risk of depressive symptoms by 1.6 times, which is consistent with our research findings [8]. Separation from parents is a major source of stress for children, which can lead to insecure parent–child attachment, and depression is significantly correlated with the quality of parent–child attachment. Children who lack parental attachment may exhibit negative emotions, such as fear and anxiety, as a response to separation from their parents, thereby increasing the risk of depression [14]. A meta-analysis of the literature suggests that parental absence can affect the nutritional status, medication use, experiences of violence and abuse, and mental disorders of children and adolescents, with mental disorders being the most common outcome, particularly depression [15]. However, as adolescents become more independent and autonomous, their ability to cope and adapt may strengthen, potentially mitigating the impact of parental absence on depression, which may be why the feature appears more important during childhood.

Therefore, families should focus on cultivating positive social attitudes and skills in their children and encourage participation in social and team activities. Additionally, families can set an example by sharing stories and leading by example, thereby fostering children's empathy. Parents who do not live with their children should maintain regular contact through phone calls, video chats, and other means to stay informed about their lives and emotional well-being, providing timely care and support.

Our study also shows that adolescents who consider the internet to be important for leisure and entertainment are more likely to experience depression. The extent to which adolescents use the internet for entertainment varies, with long-term participation in online shopping and gaming leading to easy addiction [10]. Such addiction can lead to problems, such as insufficient sleep, long-term fatigue, and the deterioration of social skills, thus increasing the risk of depression [37]. Through a detailed analysis of internet usage, a study revealed that adolescents participating in online gaming, shopping, and entertainment are prone to more severe depressive symptoms, but the time they spend on online learning is not significantly related to their depression level [38].

How parents treat their adolescent children influences the likelihood of depression. More amiable treatment, the adoption of encouraging educational methods, and active participation in their children's growth all reduce the likelihood of adolescents experiencing depression. Adolescents strive for autonomy and space [39], and conflicts with their parents are normal. When parents are too dominant and controlling, this may inhibit the child's psychological development, potentially triggering various psychological problems. In contrast, adolescents experiencing positive parenting behaviour tend to have increased confidence and self-efficacy, leading to more effective goal-oriented behaviours [5]. This type of adolescent receives correct guidance during a critical period of mood maturation, greatly reducing the probability of psychological problems such as depression.

Academic pressure is a potential factor influencing the onset of depressive symptoms in adolescents [9, 31, 40], which aligns with the findings of this study. Adolescents face increased academic workloads, often involving key examinations closely related to future life development. The resulting pressure may affect their physical and mental health, negatively impacting their academic performance and career prospects [9], which can readily lead to negative emotions and depression.

Social trust refers to whether the respondents believed that most people were selfish or helpful. Those adolescents who believed that most people were selfish were more likely to experience symptoms of depression. On the one hand, adolescents who believed that most people were helpful showed a willingness to interact with others with a positive attitude and an optimistic and kind mindset, which greatly reduced the risk of depression. On the other hand, when adolescents receive care and help from others in society, they are likely to perceive others as helpful, their sense of social belonging will be stronger, and they will be less prone to depression [11].

This study has several limitations. First, to include as many features from childhood and adolescence as possible, samples with missing values were excluded, resulting in a relatively small sample size. Future research should aim to include a larger sample to validate the model. Second, while 16–17-year-olds are also adolescents, the dataset did not consider the characteristics of such adolescents, therefore, they were not included. In future research, we hope to obtain a wider range of population data for a more comprehensive analysis of adolescent depression. Third, this study classified adolescents as depressed based on a CES-D threshold of 9. The accuracy of classifying samples with scores near this threshold may be limited. Subsequent studies could refine depression scoring through regression analysis for more precise assessments. Finally, although a wide range of factors was analysed, they were confined to social and psychological risk factors. Other potential contributors to depressive symptoms, such as environmental factors and diet, were not included in the publicly available dataset. Future research should develop more comprehensive scales to investigate relevant populations and gather more diverse data for effective analysis.

Conclusion

This study used ML to identify significant influencing factors of depression, confirming its effectiveness in accurately predicting adolescent depression symptoms. These findings may enhance understanding of adolescent depression by society, schools, and families, guiding efforts to promote mental health within such population. During childhood, parents should encourage their children to actively interact with peers and avoid feelings of loneliness. In adolescence, parents should discourage excessive internet use, while schools and families should implement diversified educational approaches to reduce students' learning pressure.

Abbreviations

ML	Machine Learning
CFPS	China Family Panel Studies
CES-D	Center for Epidemiological Studies Depression Scale
mRMR	Maximum Relevance Minimum Redundancy
RF	Random Forest
SVM	Support Vector Machine
MLP	Multilayer Perceptron
SLR	Stepwise Logistic Regression
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
SHAP	SHapley Additive exPlanations

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-025-21506-z.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Acknowledgements

We thank all the authors who participated in this study. We would like to thank all the reviewers who participated in the review.

Author contributions

Xinzhu Liu, Rui Cang and Zihe Zhang contributed to the acquisition and analysis of data, drafting and the revision of the manuscript. Ping Li contributed to revision of the manuscript and analyze the psychology of children and adolescents. Hui Wu contributed to the analysis of data and the revision of the manuscript. Wei Liu and Shu Li were responsible for the conception, design, drafting and the revision of the manuscript.

Funding

This work was supported by the Education Department of Liaoning Province [grant numbers: LJKZ0785].

Data availability

The datasets generated and analysed during the current study are available in the CFPS repository, www.isss.pku.edu.cn/cfps/.

Declarations

Ethics approval and consent to participate

The CFPS received ethical approval from the Biomedical Ethics Review Committee at Peking University, identifed by the approval code IRB00001052-14010, and obtained informed consent from all participants included in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Health and Intelligent Engineering, College of Health Management, China Medical University, 110122 Shenyang, Liaoning Province, China
[2]The First Hospital of China Medical University, 110001 Shenyang, Liaoning Province, China
[3]Department of Pediatrics, the First Hospital of China Medical Univerisity, 110001 Shenyang, Liaoning Province, China
[4]Department of Social Medicine, College of Health Management, China Medical University, 110122 Shenyang, Liaoning Province, China

Received: 15 October 2024 / Accepted: 16 January 2025
Published online: 22 January 2025

References

1. Uher R, Pavlova B. Long-term effects of depression treatment. Lancet Psychiatry. 2016;3(2):95–6. https://doi.org/10.1016/S2215-0366(15)00578-7.
2. WHO. World Health Organization. (2021). Mental health of adolescents. Retrieved from [WHO] [Internet]. https://www.who.int/zh/news-room/fact-sheets/detail/adolescent-mental-health
3. Report on national mental health development in china.In:Fu XL, Zhang K, Chen. XF,Chen ZY,editors.The Social Sciences Academic Press. Beijing:Academic;2023.p. 30-69
4. Thapar A, Eyre O, Patel V, Brent D. Depression in young people. Lancet. 2022;400(10352):617–31. https://doi.org/10.1016/S0140-6736(22)01012-1.

5. Jiang MM, Wang DW, Wu ZY, Gao K, Guo PP, Kong Y. The influence of internet use frequency, family atmosphere, and academic performance on adolescent depression: based on the chain mediating effect of self-adjustment and campus deviant behavior. Front Psychol 13, 992053. https://doi.org/10.3389/fpsyg.2022.992053
6. Mills A. Depression in young adulthood is strongly linked to childhood and adolescent social environment. Evid Based Nurs. 2023;26:15–15. https://doi.org/10.1136/ebnurs-2022-103559.
7. Coles DC, Cage J, Mothers, Children T. Matern Child Health J. 2022;26:1015–21. https://doi.org/10.1007/s10995-022-03400-x.: An Exploration of the Relationship Between Maternal Mental Health and Child Well-Being.
8. Zhao X, Wang S, Yu Y, Zhang A, Tao F, Liu D, et al. Association of early parent-child separation with multi-domain implications in adolescence and emerging adulthood: a prospective cohort study in China. J Adolesc Health. 2023;73(1):53–60. https://doi.org/10.1016/j.jadohealth.2023.02.018.
9. Zhou T, Cheng G, Wu X, Li R, Li C, Tian G, et al. The associations between Sleep Duration, Academic pressure, and depressive symptoms among Chinese adolescents: results from China Family Panel studies. Int J Environ Res Public Health. 2021;18(11):6134. https://doi.org/10.3390/ijerph18116134.
10. Rabadi L, Ajlouni M, Masannat S, Bataineh S, Batarseh G, Yessin A et al. The relationship between depression and internet addiction among university students in Jordan. Journal of Addiction Research & Therapy. 2017. https://doi.10.4172/2155-6105.1000349. https://doi.org/10.4172/2155-6105.1000349
11. Ruan Q-N, Shen G-H, Yang J-S, Yan W-J. The interplay of self-acceptance, social comparison and attributional style in adolescent mental health: cross-sectional study. Bjpsych Open. 2023;9(6):e202. https://doi.org/10.1192/bjo.2023.594.
12. Lam LT, Peng Z-W. Effect of pathological use of the internet on adolescent Mental Health. Arch Pediatr Adolesc Med. 2010;164(10):901–6. https://doi.org/10.1001/archpediatrics.2010.159.
13. Qualter P, Brown SL, Munn P, Rotenberg KJ. Childhood loneliness as a predictor of adolescent depressive symptoms: an 8-year longitudinal study. Eur Child Adolesc Psychiatry. 2010;19(6):493–501. https://doi.org/10.1007/s00787-009-0059-y.
14. Shuang M, Yiqing W, Ling J, Guanzhen O, Jing G, Zhiyong Q, Xiaohua W. Relationship between parent-child attachment and depression among migrant children and left-behind children in China. Public Health. 2011;204:1–8. https://doi.org/10.1016/j.puhe.2021.12.015.
15. Fellmeth G, Rose-Clarke K, Zhao C, Busert LK, Zheng Y, Massazza. A,.Devakumar, D. Health impacts of parental migration on left-behind children and adolescents: a systematic review and meta-analysis. Lancet. 2018;392(10164):2567–82. https://doi.org/10.1016/s0140-6736(18)32558-3.
16. Herrenkohl TI, Kosterman R, Mason WA, Hawkins JD, McCarty CA, McCauley E. Effects of childhood conduct problems and family adversity on health, health behaviors, and service use in early adulthood: tests of developmental pathways involving adolescent risk taking and depression. Dev Psychopathol. 2010;22(3):655–65. https://doi.org/10.1017/S0954579410000349.
17. Lee Y, Ragguett RM, Mansur RB, Boutilier JJ, Rosenblat JD, Trevizol A, et al. Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. J Affect Disord. 2018;241:519–32. https://doi.org/10.1016/j.jad.2018.08.073.
18. Gee DG, Caballero C. Predicting Mental Health in Adolescence: Frontoinsular Circuitry, emotion in Daily Life, and risk for Depression. Biol Psychiatry-Cognitive Neurosci Neuroimaging. 2019;4(8):684–5. https://doi.org/10.1016/j.bpsc.2019.06.004.
19. Xie Y, Hu J. An introduction to the China Family Panel Studies (CFPS). Chin Sociol Rev. 2014;47(1):3–29.
20. Zivin K, Llewellyn DJ, Lang IA, Vijan S, Kabeto MU, Miller EM, Langa KM. Depression among older adults in the United States and England. Am J Geriatric Psychiatry: Official J Am Association Geriatric Psychiatry. 2010;18(11):1036–44. https://doi.org/10.1097/JGP.0b013e3181dba6d2.
21. Briggs R, Carey D, O'Halloran AM, Kenny RA, Kennelly SP. Validation of the 8-item centre for Epidemiological Studies Depression Scale in a cohort of community-dwelling older people: data from the Irish longitudinal study on Ageing (TILDA). Eur Geriatr Med. 2018;9(1):121–6. https://doi.org/10.1007/s41999-017-0016-0.
22. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 2005;27(8):1226–38. https://doi.org/10.1109/tpami.2005.159.
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837–45.
24. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems; Long Beach, California, USA: Curran Associates Inc.; 2017. p. 4768–77.
25. Förtsch K, Viermann R, Reinauer C, Baumeister H, Warschburger P, Holl RW, et al. The impact of COVID-19 pandemic on Mental Health of adolescents with Chronic Medical conditions: findings from a German Pediatric Outpatient Clinic. J Adolesc Health: Official Publication Soc Adolesc Med. 2024;74(4):847–9. https://doi.org/10.1016/j.jadohealth.2023.11.397.
26. Koechlin H, Donado C, Berde CB, Kossowsky J. Effects of Childhood Life events on Adjustment problems in adolescence: a longitudinal study. J Dev Behav Pediatr. 2018;39(8):629–41. https://doi.org/10.1097/DBP.0000000000000596.
27. Ren Z, Xin Y, Wang Z, Liu D, Ho RCM, Ho CSH. What factors are most closely Associated with Mood disorders in adolescents during the COVID-19 Pandemic?A cross-sectional study based on 1,771 adolescents in Shandong Province, China. Front Psychiatry. 2021;12:728278. https://doi.org/10.3389/fpsyt.2021.728278.
28. Kiss O, Qu Z, Muller-Oehring EM, Baker FC, Mirzasoleiman B. Sleep, brain systems, and persistent stress in early adolescents during COVID-19: insights from the ABCD study. J Affect Disord. 2024;346:234–41. https://doi.org/10.1016/j.jad.2023.10.158.
29. Rahman MA, Kohli T. Mental health analysis of international students using machine learning techniques. PLoS ONE. 2024;19:e0304132–0304132. https://doi.org/10.1371/journal.pone.0304132.
30. Khare SK, March S, Barua PD, Vikram M, Gadre, Acharya UR. Application of data fusion for automated detection of children with developmental and mental disorders: a systematic review of the last decade. Inform Fusion. 2023;99:101898. https://doi.org/10.1016/j.inffus.2023.101898.
31. Baeza-Delgado C, Cerdá Alberich L, Carot-Sierra JM, Veiga-Canuto D, Martínez de Las Heras B, Raza B, et al. A practical solution to estimate the sample size required for clinical prediction models generated from observational research on data. Eur Radiol Exp. 2022;6(1):22. https://doi.org/10.1186/s41747-022-00276-y.
32. Chen Z, Gao S, Ding C, Luo T, Xu J, Xu S, et al. CT-based non-invasive identification of the most common gene mutation status in patients with non-small cell lung cancer. Med Phys. 2024;51(3):1872–82. https://doi.org/10.1002/mp.16744.
33. Li S, Liu S, Sun X, Hao L, Gao Q. Identification of endocrine-disrupting chemicals targeting key DCM-associated genes via bioinformatics and machine learning. Ecotoxicol Environ Saf. 2024. https://doi.org/10.1016/j.ecoenv.2024.116168.
34. Weeks M, Ploubidis GB, Cairney J, Wild TC, Naicker K, Colman I. Developmental pathways linking childhood and adolescent internalizing, externalizing, academic competence, and adolescent depression. J Adolesc. 2016;51:30–40. https://doi.org/10.1016/j.adolescence.2016.05.009.
35. Xin Z, Li S, Jia Y, Yuan H. Analysis of self-healing of depression by helping others in adolescents from the perspective of constructivism. Front Psychiatry. 2023;14:1201923. https://doi.org/10.3389/fpsyt.2023.1201923.
36. Pearce E, Myles-Hooton P, Johnson S, Hards E, Olsen S, Clisu D, et al. Loneliness as an active ingredient in preventing or alleviating youth anxiety and depression: a critical interpretative synthesis incorporating principles from rapid realist reviews. Translational Psychiatry. 2021;11(1):628. https://doi.org/10.1038/s41398-021-01740-w.
37. Alonzo R, Hussain J, Stranges S, Anderson KK. Interplay between social media use, sleep quality, and mental health in youth: a systematic review. Sleep Med Rev. 2021;56:101414. https://doi.org/10.1016/j.smrv.2020.101414.
38. Ma Y, Gu J. Internet and depression in adolescents: evidence from China. Front Psychol. 2023;14:1026920. https://doi.org/10.3389/fpsyg.2023.1026920.
39. Sanders RA. Adolescent psychosocial, social, and cognitive development. Pediatr Rev. 2013;34(8):354–9. https://doi.org/10.1542/pir.34-8-354.

40. Quach AS, Epstein NB, Riley PJ, Falconier MK, Fang X. Effects of parental warmth and academic pressure on anxiety and depression symptoms in Chinese adolescents. J Child Fam stud. 2015;24:106–16. https://doi.org/10.1007/s10826-013-9818-y.

Publisher's note