Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Joint Phenotypes of Depressive Symptoms and Learning Weariness in Adolescents

Jianli Chen
Southeast University
Jianqian Chao#
m15961396557_2@163.com

Southeast University
Xuedong Liu
Nanjing Audit University
Minlin Li
Southeast University
Gangrui Tan
Southeast University
Shengxuan Jin
Southeast University

Research Article

Keywords: adolescents, depressive symptoms, learning weariness, joint phenotype, machine learning

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11204268/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background: Depressive symptoms and learning weariness represent overlapping dimensions of adolescent well-being. Characterizing their joint distribution may inform more differentiated assessment and support. This study examined joint screening phenotypes, identified socioecological correlates of co-occurrence, and compared three models for classifying the co-occurring phenotype.

Methods: This school-based cross-sectional study included 57,753 students aged 10–18 years in Jiangsu Province, China. Screening results defined four mutually exclusive phenotypes: double negative, learning weariness only, depressive symptoms only, and co-occurring depressive symptoms and learning weariness. Multinomial logistic regression examined adjusted associations with co-occurrence relative to the double-negative group. Logistic regression, random forest (RF), and gradient boosting machine (GBM) models classified co-occurrence versus the remaining phenotypes using stratified training, validation, and test sets (70%, 15%, and 15%). Performance was assessed using areas under the receiver operating characteristic curve (AUROC) and precision–recall curve (AUPRC), Brier scores, sensitivity, and specificity. SHapley Additive exPlanations (SHAP) quantified predictor contributions to GBM outputs.

Results: The double-negative, learning-weariness-only, depressive-symptoms-only, and co-occurring phenotypes accounted for 61.3%, 21.8%, 8.7%, and 8.2% of students, respectively. Bullying victimization, sleep problems, and internet addiction were associated with higher odds of co-occurrence, with adjusted odds ratios of 3.20, 2.68, and 2.51, respectively. Residential setting and family characteristics were also associated with co-occurrence. GBM achieved the highest test-set AUROC (0.886) and AUPRC (0.628), followed by RF (0.852 and 0.556) and logistic regression (0.811 and 0.441). Corresponding Brier scores were 0.096, 0.108, and 0.123. Sleep problems, internet addiction, and bullying victimization contributed most to GBM classification.

Conclusions: Joint screening differentiated domain-specific and co-occurring emotional and academic difficulties, identifying a substantial learning-weariness-only group and an 8.2% co-occurring group. GBM demonstrated the strongest discrimination among the evaluated models. These findings highlight the potential value of integrated emotional and academic assessment and provide an empirical basis for further research on differentiated school-based support.

Background

Depressive symptoms constitute a substantial mental health burden among young people in China. A 2025 meta-analysis of 53 studies involving 314,867 children and adolescents estimated a pooled prevalence of 23.3% (95% confidence interval [CI], 21.4%–25.1%) [1]. A broader 2024 synthesis of 439 studies encompassing 1,497,524 participants across educational stages, including university students, reported a pooled prevalence of 26.17% (95% CI, 25.00%–27.41%) [2]. Although differences in populations, instruments, and screening thresholds preclude direct comparison, these estimates indicate a considerable burden of depressive symptoms. They also highlight the importance of

examining how emotional difficulties intersect with students' experiences of learning, rather than treating mental health and educational functioning as separate concerns.

Learning weariness encompasses negative cognitions and emotions about learning and disengagement from learning activities [3]. Related research has examined burnout within a demands–resources framework [4]. Academic pressure provides one possible connection between learning weariness and mental health: a systematic review found that 48 of 52 studies reported an association between academic pressure or the timing of the school year and at least one adverse mental health outcome [5]. More directly, a two-wave study of 1,092 Chinese adolescents identified reciprocal longitudinal associations between depressive symptoms and learning weariness [6]. Nevertheless, correlations between continuous measures do not establish how many students meet both screening thresholds, nor do they distinguish students experiencing only one difficulty. Classifying students into four mutually exclusive groups—neither outcome, depressive symptoms alone, learning weariness alone, and co-occurring symptoms—can reveal patterns that remain obscured when the outcomes are analysed separately.

Evidence also supports examining these joint patterns across multiple socioecological domains. A German retrospective cohort study involving 62,796 children and adolescents found that diagnosed sleep disorders were associated with subsequent depression (hazard ratio, 2.34; 95% CI, 2.09–2.63) [7]. A 2025 meta-analysis of 33 studies involving 303,243 adolescents reported a positive association between internet addiction and depression (pooled r = 0.318) [8]. At the peer level, a 2024 meta-analysis of 27 studies found that victimization was associated with higher odds of depression (odds ratio, 2.79; 95% CI, 2.43–3.21) [9]. Family and school relationships provide additional context. A systematic review of 37 studies linked better parent–child communication with more favourable adolescent mental health [10], while a 2026 meta-analysis involving 34,875 participants identified an association between interparental conflict and problematic internet use (r = 0.239), illustrating connections between family conditions and individual behaviour [11]. A review of 34 longitudinal studies and two intervention studies further supported the relevance of school connectedness to youth depression and anxiety [12]. These findings motivate an integrated analysis of individual, family, peer, and school characteristics; however, associations with a single outcome cannot be assumed to characterize students with co-occurring depressive symptoms and burnout.

A complementary methodological question is whether these characteristics can help identify the co-occurring phenotype. Machine learning has expanded across paediatric research, with an earlier systematic review documenting 363 studies [13], but methodological quality remains consequential. A systematic review of 129 applications in mental health research identified recurring weaknesses in preprocessing, model evaluation, and transparent reporting [14]. Accordingly, comparisons between regression and machine-learning models should assess more than discrimination alone. For an imbalanced binary outcome, precision–recall performance complements the area under the receiver operating characteristic curve [15], while probability accuracy and calibration require separate attention [16]. SHapley Additive exPlanations (SHAP) can characterize how predictors contribute to fitted tree-

model outputs [17], providing a model-based perspective alongside adjusted regression associations. Contemporary TRIPOD + AI and PROBAST + AI guidance further emphasizes transparent reporting, appropriate validation, and careful assessment of bias and applicability [18, 19].

Against this background, the present study examined 57,753 students aged 10–18 years in Jiangsu, China, spanning primary, junior secondary, general senior secondary, and vocational education. Its contribution lies in combining a four-category joint screening framework with socioecological association analyses and interpretable classification modelling within the same large school-based sample. Specifically, we aimed to characterize the distribution of the four screening phenotypes, examine adjusted correlates of the co-occurring phenotype, and compare the internally evaluated classification performance of logistic regression, random forest, and gradient boosting machine models. By distinguishing associations with co-occurrence from the ability to classify individual students, this approach provides a more differentiated account of overlapping emotional and academic difficulties and an empirical basis for subsequent longitudinal research and external validation.

Methods

Study design and participants

This study was a secondary analysis of a school-based cross-sectional survey of family education and mental health in Jiangsu Province, China. The analytic sample comprised 57,753 students aged 10–18 years after checks of age eligibility, data completeness, and logical consistency. Students attended primary schools, middle schools, general high schools, and technical high schools. The unit of analysis was the individual student.

Depressive symptoms

Depressive symptoms were assessed using the Center for Epidemiologic Studies Depression Scale (CES-D), which has demonstrated validity and measurement stability among Chinese adolescents [20, 21]. Responses were scored from 0 to 3, with positively worded items reverse-scored. Higher total scores indicated greater symptom severity. A score of at least 16 defined a positive screen for depressive symptoms [22]. Internal consistency in this sample was good (Cronbach's $\alpha = 0.845$). Screening results indicated depressive symptomatology rather than a clinical diagnosis.

Learning Weariness

Learning weariness was assessed using the 17-item Learning Weariness Scale for Junior High School Students developed by Zhao [3]. Items covered cognitive, behavioural, and emotional aspects of learning and were rated from 1 to 5, with higher scores indicating greater learning weariness. Following a previous application of the same scale, a mean item score of at least 3 was used as the operational

threshold for learning weariness [23]. This threshold defined the research classification and was not interpreted as a clinical diagnostic cutoff.

Joint screening phenotypes

Longitudinal research has documented associations between depressive symptoms and learning weariness in adolescents [24]. Distinguishing single-domain from co-occurring symptoms also has precedent in adolescent epidemiological research [25]. In this study, the two binary screening results defined four mutually exclusive groups: double negative, learning weariness only, depressive symptoms only, and co-occurring depressive symptoms and learning weariness. Double negative indicated scores below both thresholds. The four-category outcome was used for descriptive analysis and multinomial logistic regression, with the double-negative group as the reference. For machine learning, co-occurrence was the positive class, and the other three groups constituted the negative class.

Socioecological variables

The survey assessed characteristics across individual, family, peer, and school domains. Demographic and family variables included age, sex, school stage, region, residential setting, only-child status, body mass index (BMI), parental education, parental marital status, living arrangement, educational expenditure, parental involvement, and parental companionship. Behavioural and school-related measures included sleep quality, academic stress, screen time, internet addiction, early romantic involvement, and bullying victimization.

Bullying victimization was assessed using the frequency categories of the Olweus Bully/Victim Questionnaire [26]. Students reporting victimization at least two or three times per month were classified as positive. Academic stress was measured using the 16-item Educational Stress Scale for Adolescents (ESSA), covering pressure from study, workload, worry about grades, self-expectation, and despondency [27]. Items were rated from 1 to 5, yielding a total score of 16–80, with higher scores indicating greater academic stress. Recent research among Chinese students has also linked academic stress with depressive and anxiety symptoms [28].

Sleep quality was assessed using the survey sleep score, with scores above 5 classified as poor sleep and scores of 5 or below as good sleep. The survey documentation referenced the Pittsburgh Sleep Quality Index (PSQI)[29], for which a global score above 5 was originally used to distinguish poor from good sleepers. BMI was calculated as weight in kilograms divided by height in metres squared and categorized using Chinese age- and sex-specific criteria [30]. Early romantic involvement was recorded as a yes/no item. Daily screen time was categorized as no more than 2 hours versus more than 2 hours. Weekly parental companionship was categorized as no more than 10 hours versus more than 10 hours for descriptive analyses and expressed per 10-hour increase in regression.

Data cleaning and missing values

Data cleaning included checks of duplicate records, permissible ranges, skip patterns, consistency between date of birth and survey date, scale completeness, and logical relationships among variables. Students with missing primary outcomes were excluded from the joint-phenotype analysis. For machine learning, categorical predictors underwent mode imputation and one-hot encoding, and continuous predictors underwent median imputation.

Statistical analysis

Categorical characteristics were summarized as counts and percentages. Table 1 presents characteristics by depressive-symptom screening status to show the component outcome underlying the joint classification. Percentages use the corresponding column total. Unadjusted Pearson chi-square tests were calculated from the displayed contingency counts, without a continuity correction. These descriptive comparisons did not incorporate sampling weights or school-level clustering and were interpreted as exploratory.

Table 1
Participant characteristics by depressive-symptom screening status

Variable	Category	Total (N = 57,753)	Depressive symptoms absent (n = 47,992)	Depressive symptoms present (n = 9,761)	P value
Age group (years)	≤ 12	28,203 (48.8)	24,236 (50.5)	3,967 (40.6)	< 0.001
	13–15	22,082 (38.2)	17,949 (37.4)	4,133 (42.3)	
	≥ 16	7,468 (12.9)	5,807 (12.1)	1,661 (17.0)	
Sex	Male	30,103 (52.1)	24,764 (51.6)	5,339 (54.7)	< 0.001
	Female	27,650 (47.9)	23,228 (48.4)	4,422 (45.3)	
Residential setting	Urban	23,618 (40.9)	19,821 (41.3)	3,797 (38.9)	< 0.001
	Town	17,468 (30.2)	14,589 (30.4)	2,879 (29.5)	
	Rural	15,780 (27.3)	12,910 (26.9)	2,870 (29.4)	
	Suburban	887 (1.5)	672 (1.4)	215 (2.2)	
Region	Northern Jiangsu	18,173 (31.5)	15,069 (31.4)	3,104 (31.8)	0.014
	Central Jiangsu	18,127 (31.4)	14,974 (31.2)	3,153 (32.3)	
	Southern Jiangsu	21,453 (37.1)	17,949 (37.4)	3,504 (35.9)	
learning weariness	No burnout (mean < 3)	40,455 (70.0)	35,418 (73.8)	5,037 (51.6)	< 0.001
	Burnout (mean ≥ 3)	17,298 (30.0)	12,574 (26.2)	4,724 (48.4)	
Sleep quality	Good (≤ 5)	44,403 (76.9)	37,434 (78.0)	6,969 (71.4)	< 0.001
	Poor (> 5)	13,350 (23.1)	10,558 (22.0)	2,792 (28.6)	

Values are n (%); CNY, Chinese yuan; BMI, body mass index.

Variable	Category	Total (N = 57,753)	Depressive symptoms absent (n = 47,992)	Depressive symptoms present (n = 9,761)	P value
Daily screen time	≤ 2 h	40,376 (69.9)	34,842 (72.6)	5,534 (56.7)	< 0.001
	> 2 h	17,377 (30.1)	13,150 (27.4)	4,227 (43.3)	
Early romantic involvement	No	50,033 (86.6)	42,185 (87.9)	7,848 (80.4)	< 0.001
	Yes	7,720 (13.4)	5,807 (12.1)	1,913 (19.6)	
Bullying victimization	No	45,172 (78.2)	38,911 (81.1)	6,261 (64.1)	< 0.001
	Yes	12,581 (21.8)	9,081 (18.9)	3,500 (35.9)	
School stage	Primary school	30,734 (53.2)	26,156 (54.5)	4,578 (46.9)	< 0.001
	Middle school	20,450 (35.4)	16,653 (34.7)	3,797 (38.9)	
	High school	5,248 (9.1)	4,223 (8.8)	1,025 (10.5)	
	Technical high school	1,321 (2.3)	960 (2.0)	361 (3.7)	
Only-child status	Yes	19,709 (34.1)	16,605 (34.6)	3,104 (31.8)	< 0.001
	No	38,044 (65.9)	31,387 (65.4)	6,657 (68.2)	
Parental education	Middle school or below	29,557 (51.2)	24,452 (51.0)	5,105 (52.3)	< 0.001
	High school	15,494 (26.8)	12,849 (26.8)	2,645 (27.1)	
	Bachelor's degree	11,990 (20.8)	10,116 (21.1)	1,874 (19.2)	
	Graduate degree	712 (1.2)	575 (1.2)	137 (1.4)	
Parental marital status	Married	52,961 (91.7)	44,293 (92.3)	8,668 (88.8)	< 0.001

Values are n (%); CNY, Chinese yuan; BMI, body mass index.

Variable	Category	Total (N = 57,753)	Depressive symptoms absent (n = 47,992)	Depressive symptoms present (n = 9,761)	P value
	Divorced	3,172 (5.5)	2,450 (5.1)	722 (7.4)	
	Remarried after divorce	1,148 (2.0)	865 (1.8)	283 (2.9)	
	Widowed	472 (0.8)	384 (0.8)	88 (0.9)	
Parental educational expenditure, previous year (CNY)	None	3,351 (5.8)	2,625 (5.5)	726 (7.4)	< 0.001
	1–4,999	16,867 (29.2)	13,788 (28.7)	3,079 (31.5)	
	5,000–19,999	24,084 (41.7)	20,066 (41.8)	4,018 (41.2)	
	20,000–49,999	9,476 (16.4)	8,129 (16.9)	1,347 (13.8)	
	≥ 50,000	3,975 (6.9)	3,384 (7.1)	591 (6.1)	
Weekly parental companionship	≤ 10 h	20,117 (34.8)	15,768 (32.9)	4,349 (44.6)	< 0.001
	> 10 h	37,636 (65.2)	32,224 (67.1)	5,412 (55.4)	
Living arrangement	Both parents	36,327 (62.9)	30,482 (63.5)	5,845 (59.9)	< 0.001
	Mother only	5,486 (9.5)	4,293 (8.9)	1,193 (12.2)	
	Father only	1,328 (2.3)	1,047 (2.2)	281 (2.9)	
	Parents and grandparent(s)	10,800 (18.7)	9,210 (19.2)	1,590 (16.3)	
	Grandparent(s) only	3,465 (6.0)	2,691 (5.6)	774 (7.9)	
	Living alone	58 (0.1)	41 (0.1)	17 (0.2)	
	Other	289 (0.5)	228 (0.5)	61 (0.6)	

Values are n (%); CNY, Chinese yuan; BMI, body mass index.

Variable	Category	Total (N = 57,753)	Depressive symptoms absent (n = 47,992)	Depressive symptoms present (n = 9,761)	P value
BMI category	Underweight	3,754 (6.5)	3,052 (6.4)	702 (7.2)	< 0.001
	Normal weight	40,427 (70.0)	33,920 (70.7)	6,507 (66.7)	
	Overweight	8,374 (14.5)	6,860 (14.3)	1,514 (15.5)	
	Obesity	5,198 (9.0)	4,160 (8.7)	1,038 (10.6)	
Values are n (%); CNY, Chinese yuan; BMI, body mass index.					

Multinomial logistic regression used the double-negative phenotype as the reference category. Adjusted odds ratios (aORs) and 95% confidence intervals (CIs) are reported for selected associations with the co-occurring phenotype, which was the inferential focus of this report. Adjustment covered demographic, family, sleep, digital-behaviour, school, and peer domains. Regression estimates describe cross-sectional associations with phenotype membership and were not interpreted as changes in absolute risk or as causal effects.

Machine learning and model interpretation

Students were randomly divided into training (70%), validation (15%), and test (15%) sets, stratified by the co-occurring outcome. The models were class-weighted logistic regression, RF, and GBM. Classification thresholds were selected by maximizing the Youden index in the validation set and were fixed before test-set assessment. Test-set performance was used for the principal model comparison; training-set performance was reported separately[31].

Outcome-defining information was not used as predictor input. Preprocessing was confined to the training data, and the validation and test sets were kept separate from model fitting to prevent information leakage.

Performance measures were AUROC, AUPRC, Brier score, sensitivity, and specificity. AUPRC summarized precision–recall performance for the positive class and was interpreted in relation to the frequency of co-occurrence. The Brier score represented mean squared probability error; a lower score indicated better probability accuracy. Brier scores were not treated as a substitute for a calibration curve or calibration parameters [32]. Comparisons between models were descriptive because uncertainty estimates for performance differences were not presented.

Mean absolute SHAP values were calculated using Tree SHAP to summarize predictor contributions in GBM. These values describe the magnitude of contributions to the model output, not the direction of an association or the effect of an intervention. In keeping with methodological concerns in mental-health prediction research, model discrimination and explanation were interpreted separately from readiness for clinical or school use .

Software and ethics

Data cleaning and model analyses were performed in Python 3.9 using pandas, statsmodels, scikit-learn, and SHAP. The study protocol was approved by the Human Research Committee of the Jiangsu Research and Guidance Center for Mental Health Education in Primary and Secondary Schools (approval no. JS-MH2024.001). Informed consent was obtained for participation. The study was conducted in accordance with the principles of the Declaration of Helsinki [33].

Results

Participant characteristics and phenotype distribution

The analytic sample comprised 57,753 students with a mean age of 12.88 years (SD, 1.99), including 27,650 females (47.9%). Participants attended four school stages and resided in urban, town, rural, and suburban settings (Table 1). Overall, 9,761 students (16.9%) screened positive for depressive symptoms and 17,298 (30.0%) for learning weariness.The joint screening classification identified 35,418 students (61.3%) in the double-negative group, 12,574 (21.8%) in the learning weariness-only group, 5,037 (8.7%) in the depressive-symptoms-only group, and 4,724 (8.2%) in the co-occurring group (Table 2). learning weariness alone was approximately 2.5 times as prevalent as depressive symptoms alone.

Table 2
Distribution of joint screening phenotypes

Joint phenotype	Screening combination	n	Percentage (%)
Double negative	Depressive symptoms negative; learning weariness negative	35,418	61.3
learning weariness only	Depressive symptoms negative; learning weariness positive	12,574	21.8
Depressive symptoms only	Depressive symptoms positive; learning weariness negative	5,037	8.7
Co-occurring phenotype	Depressive symptoms positive; learning weariness positive	4,724	8.2
Total	—	57,753	100.0

Correlates of the co-occurring phenotype

Adjusted associations with the co-occurring phenotype, relative to the double-negative group, are presented in Table 3. Bullying victimization (aOR = 3.20; 95% CI, 2.90–3.52), sleep problems (aOR = 2.68; 95% CI, 2.47–2.91), and internet addiction (aOR = 2.51; 95% CI, 2.26–2.79) were associated with higher odds of co-occurrence.

Table 3
Selected adjusted associations with the co-occurring phenotype relative to the double-negative group

Variable (comparison)	aOR	95% CI
Female (vs male)	1.16	1.07–1.25
Town (vs urban)	1.14	1.02–1.27
Rural (vs urban)	1.39	1.25–1.54
Suburban (vs urban)	1.52	1.09–2.12
Parental divorce	1.37	1.18–1.59
Parental remarriage after divorce	1.69	1.37–2.08
Sleep problems	2.68	2.47–2.91
Internet addiction	2.51	2.26–2.79
Bullying victimization	3.20	2.90–3.52
Parental noninvolvement (vs active cooperation)	2.59	1.78–3.75
Weekly parental companionship (per 10-h increase)	0.975	0.965–0.985

aOR, adjusted odds ratio; CI, confidence interval.

Family characteristics were also associated with phenotype membership. Parental noninvolvement, compared with active cooperation in education, was associated with higher odds of co-occurrence (aOR = 2.59; 95% CI, 1.78–3.75). The corresponding aORs for parental divorce and remarriage after divorce were 1.37 (95% CI, 1.18–1.59) and 1.69 (95% CI, 1.37–2.08), respectively. Each additional 10 hours of weekly parental companionship was associated with 2.5% lower odds of co-occurrence (aOR = 0.975; 95% CI, 0.965–0.985).

Females had higher odds of co-occurrence than males (aOR = 1.16; 95% CI, 1.07–1.25). Relative to urban residence, the adjusted odds were higher for town (aOR = 1.14; 95% CI, 1.02–1.27), rural (aOR = 1.39; 95% CI, 1.25–1.54), and suburban residence (aOR = 1.52; 95% CI, 1.09–2.12).

Internal model performance

GBM achieved the highest test-set AUROC (0.886) and AUPRC (0.628), followed by RF (0.852 and 0.556, respectively) and logistic regression (0.811 and 0.441, respectively) (Table 4; Fig. 1). The absolute

differences favouring GBM over logistic regression were 0.075 for AUROC and 0.187 for AUPRC; the corresponding differences over RF were 0.034 and 0.072.

Table 4
Internal classification performance of the three models

Dataset / model	AUROC	AUPRC	Brier score	Sensitivity	Specificity
Training / GBM	0.906	0.663	0.087	0.842	0.821
Training / RF	0.881	0.604	0.099	0.815	0.794
Training / Logistic regression	0.828	0.482	0.116	0.768	0.749
Test / GBM	0.886	0.628	0.096	0.826	0.807
Test / RF	0.852	0.556	0.108	0.798	0.781
Test / Logistic regression	0.811	0.441	0.123	0.752	0.738

AUROC, area under the receiver operating characteristic curve; AUPRC, area under the precision–recall curve; GBM, gradient boosting machine; RF, random forest.

Panel A shows the training set and panel B the test set. The positive class is the co-occurring phenotype. The diagonal represents chance discrimination. GBM, gradient boosting machine; RF, random forest; AUROC, area under the receiver operating characteristic curve.

Bars summarize the magnitude of contributions to the GBM output. Larger values indicate greater model contribution, not stronger causal effects. SHAP, SHapley Additive exPlanations; GBM, gradient boosting machine.

The ranking was consistent for the Brier score, which was lowest for GBM (0.096), followed by RF (0.108) and logistic regression (0.123). At the classification threshold selected in the validation set, GBM achieved a test-set sensitivity of 0.826 and specificity of 0.807.

GBM interpretation

SHAP analysis identified sleep problems, internet addiction, and bullying victimization as the leading contributors to GBM predictions, as measured by mean absolute SHAP values (Fig. 2). Other prominent contributors included early romantic involvement, residential setting, overweight or obesity, parental companionship, screen time, school stage, and parental marital status. The SHAP rankings highlighted behavioural, peer, and family characteristics contributing to classification of the co-occurring phenotype.

Discussion

Principal findings

Joint screening distinguished domain-specific and co-occurring depressive symptoms and learning weariness, with learning weariness alone substantially more prevalent than depressive symptoms alone. Co-occurrence was associated with behavioural, peer, and family characteristics. GBM achieved the strongest internal classification performance, supporting the value of combining joint phenotyping with socioecological analysis and interpretable modelling.

Socioecological associations

The prominence of sleep and digital behaviour is consistent with longitudinal evidence linking sleep to school engagement, performance, and learning weariness [34], and problematic device use to sleep and mental health [35]. Associations with bullying and family circumstances further situate co-occurring difficulties within students' interpersonal environments. Recent evidence on school connectedness and school-based interventions supports investigating coordinated support across educational and psychological domains [36, 37].

Interpretation of model performance

GBM's comparative advantage is consistent with its capacity to capture nonlinear relationships and predictor combinations. Consistent rankings across discrimination, precision–recall performance, and probability error strengthen the internal comparison. These complementary dimensions provide a broader assessment than any single metric, consistent with contemporary evaluation guidance .Convergence between regression findings and SHAP contributions links adjusted associations with information used for classification. This enhances interpretability and identifies domains for further investigation, while SHAP contributions describe model behaviour rather than causal effects [38].

Implications for school assessment

The joint framework may help inform focused follow-up by distinguishing emotional, academic, and co-occurring difficulties. It may also provide a useful framework for communication among teachers, mental-health professionals, and families when considering coordinated assessment. Research on school-based help-seeking further highlights the importance of connecting identification with accessible support [39].Phenotype distributions could inform assessment resource planning, while socioecological findings identify relevant contexts for follow-up. The model offers a foundation for evaluating whether contextual information collected before comprehensive screening can improve initial case finding.

Strengths and limitations

Key strengths of this study include the large school-based sample spanning multiple educational stages and the integration of joint phenotyping, adjusted socioecological analyses, and interpretable classification modelling. Separate training, validation, and test partitions, together with complementary performance measures, strengthened the internal model comparison. Several limitations should nevertheless be considered. First, the cross-sectional design precludes determination of temporal or causal relationships; therefore, the observed associations should not be interpreted as evidence that the identified factors cause the co-occurring phenotype. Second, most measures were based on self-

reported information and may be subject to recall, reporting, and common-method biases. In addition, the screening phenotypes represent symptom-based classifications rather than clinical diagnoses. Third, although separate data partitions were used for model development and evaluation, all partitions originated from the same survey population. The reported performance therefore represents internal rather than external validation and should not be assumed to generalize to other schools, regions, or time periods. School-level clustering was not incorporated into the present analyses, and future studies should examine its potential influence. Further external and prospective validation, including assessment of calibration, performance uncertainty, subgroup consistency, and measurement comparability, is needed before considering broader school-based application.

Conclusions

Joint screening distinguished domain-specific and co-occurring depressive symptoms and learning weariness, providing a more differentiated account of adolescent emotional and academic difficulties. Socioecological analyses highlighted behavioural, peer, and family correlates of co-occurrence, while GBM achieved the strongest internal classification performance among the evaluated models. Together, these findings highlight the potential value of considering emotional and academic difficulties jointly and provide an empirical foundation for further research on differentiated school support pathways.

Abbreviations

aOR, adjusted odds ratio; AUPRC, area under the precision–recall curve; AUROC, area under the receiver operating characteristic curve; BMI, body mass index; CES-D, Center for Epidemiologic Studies Depression Scale; CI, confidence interval; ESSA, Educational Stress Scale for Adolescents; GBM, gradient boosting machine; RF, random forest; SHAP, SHapley Additive exPlanations.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Human Research Committee of the Jiangsu Research and Guidance Center for Mental Health Education in Primary and Secondary Schools (approval no. JS-MH2024.001). Informed consent to participate was obtained from the parents or legal guardians of all participants through a school-organized consent process, including all participants younger than 16 years. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Clinical trial number:

not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The data include sensitive family and mental-health information about minors and are not publicly available. Requests for access may be directed to the data management institution and are subject to ethical approval and applicable data-governance requirements.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Key Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (2022SJZD141) and the National Natural Science Foundation of China (81872711).

Author contributions

Conceptualization: Jianli Chen and Jianqian Chao. Methodology: Jianli Chen and Gangrui Tan. Formal analysis: Minlin Li. Investigation: Jianqian Chao and Xuedong Liu. Data curation: Shengxuan Jin. Writing—original draft: Jianli Chen. Writing—review and editing: all authors. Funding acquisition: Jianqian Chao and Xuedong Liu. All authors have read and approved the manuscript.

Acknowledgements

We thank the participating students and their parents or guardians, as well as the teachers, research staff, medical personnel, and volunteers who supported data collection and study coordination.

References

1. Feng Y, Wang F. [Meta analysis of the prevalence of depressive symptoms among children and adolescents in China from 2015 to 2024]. Zhongguo Dang Dai Er Ke Za Zhi. 2025;27(5):529–39. 10.7499/j.issn.1008-8830.2409109. Chinese.
2. Zhou J, Liu Y, Ma J, Feng Z, Hu J, Hu J, et al. Prevalence of depressive symptoms among children and adolescents in china: a systematic review and meta-analysis. Child Adolesc Psychiatry Ment Health. 2024;18(1):150. 10.1186/s13034-024-00841-w.
3. Zhao Y. Establishment and application of junior middle school students' learning weariness scale. Journal of Shanghai Educational Research. 2019;(10):27–30. Chinese. 10.16194/j.cnki.31-1059/g4.2019.10.006
4. Salmela-Aro K, Upadyaya K. School burnout and engagement in the context of demands–resources model. Br J Educ Psychol. 2014;84(Pt 1):137–51. 10.1111/bjep.12018.

5. Steare T, Gutiérrez Muñoz C, Sullivan A, Lewis G. The association between academic pressure and adolescent mental health problems: A systematic review. J Affect Disord. 2023;339:302–17. 10.1016/j.jad.2023.07.028.
6. Zhang R, Chen H, Liu L, Wang F, Yang Z. Unraveling the longitudinal relationships between connectedness to nature, depressive symptoms, and learning burnout in adolescents. J Adolesc. 2024;96(6):1212–23. 10.1002/jad.12330.
7. Kaur N, Vetter C, Konrad M, Kostev K. Investigation of the Association between Sleep Disorders with Subsequent Depression in Children and Adolescents—A Retrospective Cohort Study with 62,796 Patients. Child (Basel). 2024;11(7):758. 10.3390/children11070758.
8. Soriano-Molina E, Limiñana-Gras RM, Patró-Hernández RM, Rubio-Aparicio M. The Association Between Internet Addiction and Adolescents' Mental Health: A Meta-Analytic Review. Behav Sci (Basel). 2025;15(2):116. 10.3390/bs15020116.
9. Song Q, Yuan T, Hu Y, Liu X, Fei J, Zhao X, et al. The Effect of Peer Victimization During Adolescence on Depression and Gender Differences: A Systematic Review and Meta-analysis. Trauma Violence Abuse. 2024;25(4):2862–76. 10.1177/15248380241227538.
10. Zapf H, Boettcher J, Haukeland Y, Orm S, Coslar S, Fjermestad K. A systematic review of the association between parent-child communication and adolescent mental health. JCPP Adv. 2024;4(2):e12205. 10.1002/jcv2.12205.
11. Wang Z, Zhang Q, Ran G, Wang Y, Lu L. The Association Between Interparental Conflict and Problematic Internet Use: A Three-Level Meta-Analysis. J Adolesc. 2026;98(3):807–19. 10.1002/jad.70093.
12. Raniti M, Rakesh D, Patton GC, Sawyer SM. The role of school connectedness in the prevention of youth depression and anxiety: a systematic review with youth consultation. BMC Public Health. 2022;22(1):2152. 10.1186/s12889-022-14364-6.
13. Hoodbhoy Z, Masroor Jeelani S, Aziz A, Habib MI, Iqbal B, Akmal W, et al. Machine Learning for Child and Adolescent Health: A Systematic Review. Pediatrics. 2021;147(1):e2020011833. 10.1542/peds.2020-011833.
14. Tornero-Costa R, Martinez-Millana A, Azzopardi-Muscat N, Lazeri L, Traver V, Novillo-Ortiz D. Methodological and Quality Flaws in the Use of Artificial Intelligence in Mental Health Research: Systematic Review. JMIR Ment Health. 2023;10:e42045. 10.2196/42045.
15. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE. 2015;10(3):e0118432. 10.1371/journal.pone.0118432.
16. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230. 10.1186/s12916-019-1466-7.
17. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell. 2020;2(1):56–67.

10.1038/s42256-019-0138-9.

18. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. 10.1136/bmj-2023-078378.
19. Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, Dhiman P, et al. PROBAST + AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. 2025;388:e082505. 10.1136/bmj-2024-082505.
20. Yang W, Xiong G, Garrido LE, Zhang JX, Wang MC, Wang C. Factor structure and criterion validity across the full scale and ten short forms of the CES-D among Chinese adolescents. Psychol Assess. 2018;30(9):1186–98. 10.1037/pas0000559.
21. Zhu X, Shek DTL, Dou D. Factor structure of the Chinese CES-D and invariance analyses across gender and over time among Chinese adolescents. J Affect Disord. 2021;295:639–46. 10.1016/j.jad.2021.08.122.
22. Vilagut G, Forero CG, Barbaglia G, Alonso J. Screening for depression in the general population with the Center for Epidemiologic Studies Depression (CES-D): a systematic review with meta-analysis. PLoS ONE. 2016;11(5):e0155431. 10.1371/journal.pone.0155431.
23. Luo AQ, Cheng K, Yang YM, Renzhen, Mido. Ciren Zhima. [Effectiveness of solution-focused group counseling for learning weariness among junior high school students in Daofu County]. Psychol China. 2024;6(2):236–45. 10.35534/pc.0602026. Chinese.
24. Song S, Guo R, Chen X, Li C. How school burnout affects depression symptoms among Chinese adolescents: evidence from individual and peer clique level. J Youth Adolesc. 2025;54(1):92–102. 10.1007/s10964-024-02044-0.
25. de Castro F, Cappa C, Madans J. Anxiety and depression signs among adolescents in 26 low- and middle-income countries: prevalence and association with functional difficulties. J Adolesc Health. 2023;72(1 Suppl):S79–87. 10.1016/j.jadohealth.2022.03.022.
26. Gaete J, Valenzuela D, Godoy MI, Rojas-Barahona CA, Salmivalli C, Araya R. Validation of the revised Olweus Bully/Victim Questionnaire (OBVQ-R) among adolescents in Chile. Front Psychol. 2021;12:578661. 10.3389/fpsyg.2021.578661.
27. Sun J, Dunne MP, Hou XY, Xu AQ. Educational Stress Scale for Adolescents: Development, Validity, and Reliability With Chinese Students. J Psychoeduc Assess. 2011;29(6):534–46. 10.1177/0734282910394976.
28. Wang D, Li H, Zhang X, Wang X, Qiu Y, Ma Z, et al. Changes and relevant factors in depressive and anxiety symptoms among Chinese adolescents after the Double Reduction policy: a repeated cross-sectional survey with a nested longitudinal subsample. J Affect Disord. 2024;367:88–95. 10.1016/j.jad.2024.08.191.
29. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193–213. 10.1016/0165-1781(89)90047-4.

30. Ji CY, Working Group on Obesity in China. Report on childhood obesity in China (1)—body mass index reference for screening overweight and obesity in Chinese school-age children. Biomed Environ Sci. 2005;18(6):390–400.
31. Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ. 2024;384:e074819. 10.1136/bmj-2023-074819.
32. Van Calster B, Collins GS, Vickers AJ, Wynants L, Kerr KF, Barreñada L, et al. Evaluation of performance measures in predictive artificial intelligence models to support medical decisions: overview and guidance. Lancet Digit Health. 2025;7(12):100916. 10.1016/j.landig.2025.100916.
33. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants. JAMA. 2025;333(1):71–4. 10.1001/jama.2024.21972.
34. Bacaro V, Carpentier L, Crocetti E. Sleep well, study well: a systematic review of longitudinal studies on the interplay between sleep and school experience in adolescence. Int J Environ Res Public Health. 2023;20(6):4829. 10.3390/ijerph20064829.
35. Dibben GO, Martin A, Shore CB, Johnstone A, McMellon C, Palmer V, et al. Adolescents' interactive electronic device use, sleep and mental health: a systematic review of prospective studies. J Sleep Res. 2023;32(5):e13899. 10.1111/jsr.13899.
36. Meng X, Chen Y, Rudasill KM, Xu L, Yu D, Harrell DR. The reciprocal relationship between school connectedness and adolescent depressive symptoms: a meta-analytic cross-lagged panel analysis. J Youth Adolesc. 2025;54(10):2574–92. 10.1007/s10964-025-02212-w.
37. Cohen KA, Ito S, Ahuvia IL, Yang Y, Zhang Y, Renshaw TL, et al. Brief school-based interventions targeting student mental health or well-being: a systematic review and meta-analysis. Clin Child Fam Psychol Rev. 2024;27(3):732–806. 10.1007/s10567-024-00487-2.
38. Vowels MJ. Trying to outrun causality with machine learning: limitations of model explainability techniques for exploratory research. Psychol Methods. 2024 Sep;9. 10.1037/met0000699. Online ahead of print.
39. McPhail L, Thornicroft G, Gronholm PC. Help-seeking processes related to targeted school-based mental health services: systematic review. BMC Public Health. 2024;24:1217. 10.1186/s12889-024-18714-4.

Figures

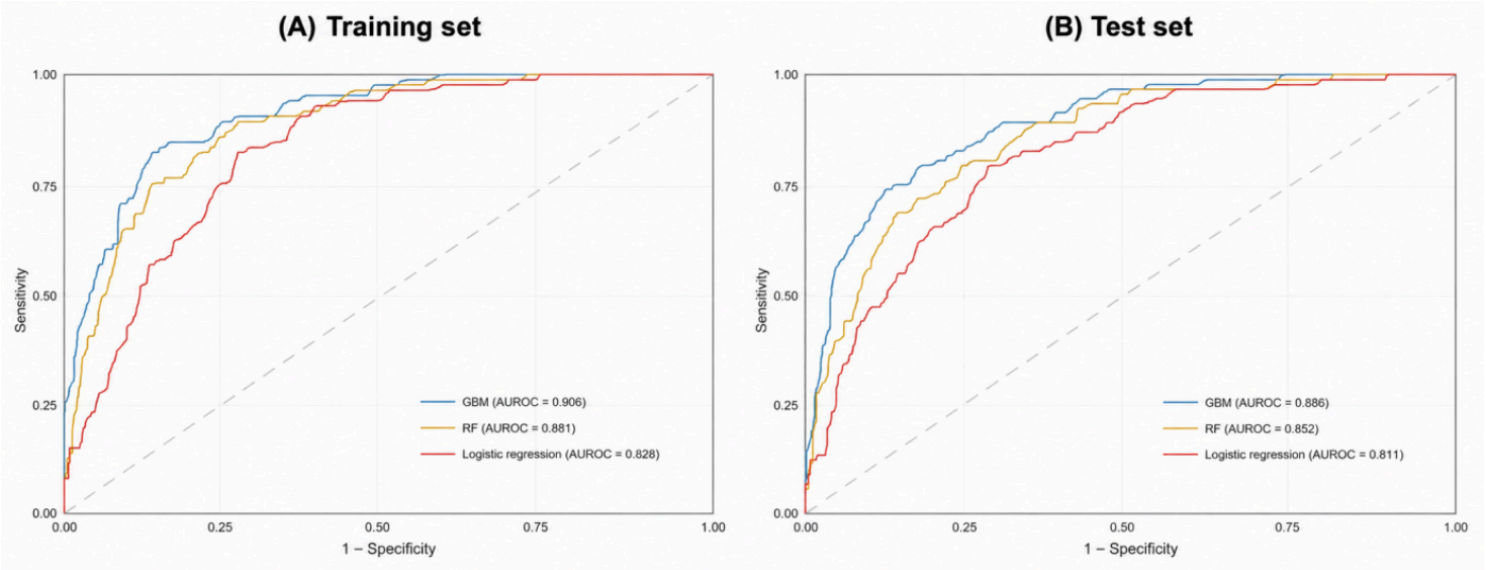


Figure 1

Receiver operating characteristic curves in the training and test sets

Panel A shows the training set and panel B the test set. The positive class is the co-occurring phenotype. The diagonal represents chance discrimination. GBM, gradient boosting machine; RF, random forest; AUROC, area under the receiver operating characteristic curve.

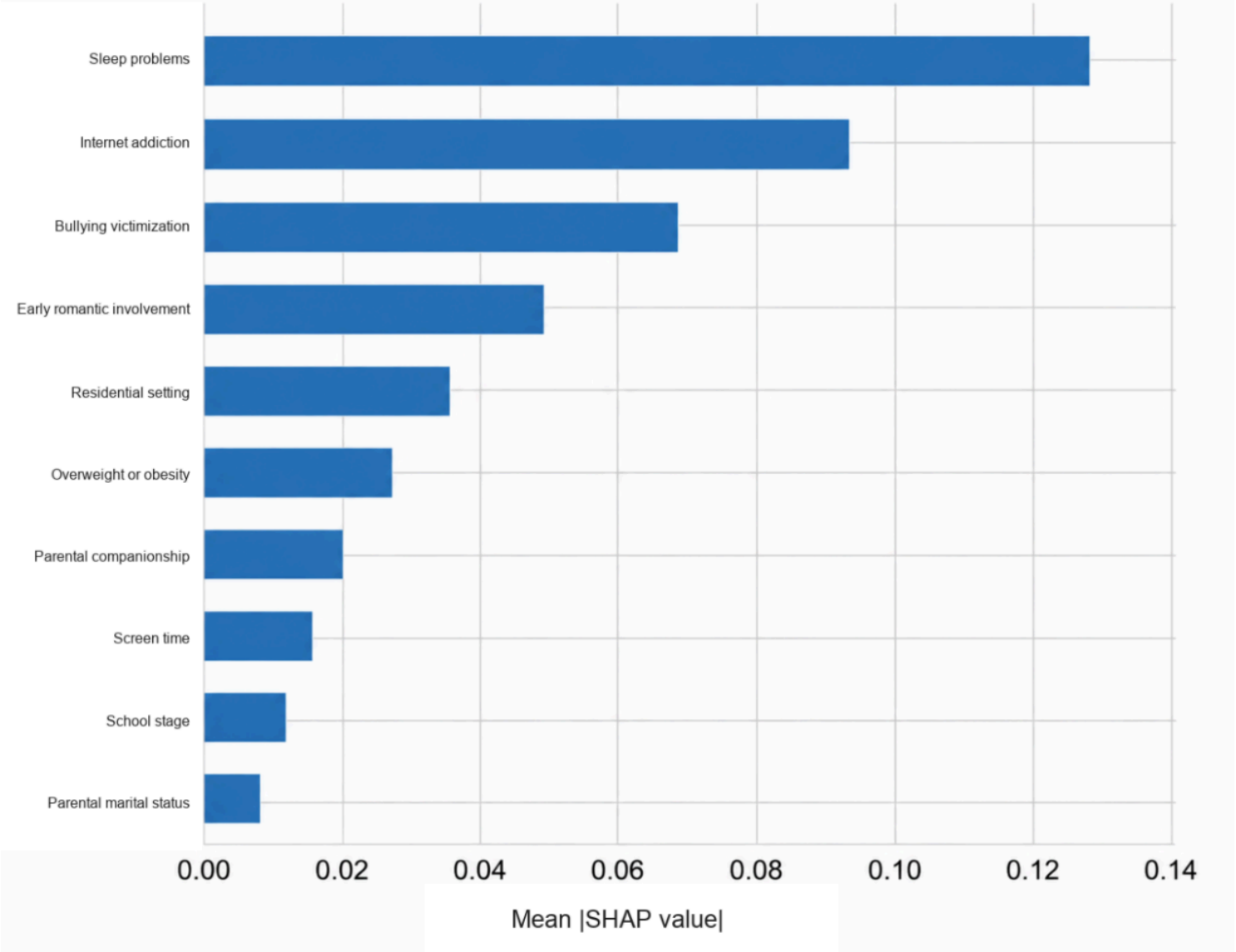


Figure 2

Mean absolute SHAP values for leading predictors in GBM

Bars summarize the magnitude of contributions to the GBM output. Larger values indicate greater model contribution, not stronger causal effects. SHAP, SHapley Additive exPlanations; GBM, gradient boosting machine.