

Machine learning and SHAP-based prediction of depression among residents in Gansu Province

Yan Geng

Lanzhou University

Ningjing Zhang

Gansu Provincial Center for Disease Control and Prevention

Lulu Xie

Gansu Provincial Center for Disease Control and Prevention

Weimin Pan

Gansu Provincial Center for Disease Control and Prevention

Xiaofeng Luo

luoxiaof@lzu.edu.cn

Lanzhou University

Research Article

Keywords:

Posted Date: October 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10607452/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

Depression is a leading cause of disability worldwide, yet community-based screening and early identification remain inadequate. Conventional statistical approaches assume linear relationships and are limited in capturing the complex interactions between factors.

Methods

We evaluated five supervised classifiers via nested 10-fold cross-validation on data from the 2024–2025 Gansu Provincial Mental Health Literacy and Common Mental Disorders Survey, and interpreted the final model using SHapley Additive exPlanations (SHAP). Restricted cubic splines were used to model nonlinear dose-response relationships, and additive and multiplicative interactions were quantified using RERI, AP, synergy index, and 3D probability surface plots.

Results

XGBoost achieved satisfactory discriminative performance (AUROC = 0.884; AUPRC = 0.328; sensitivity = 0.690). SHAP analysis identified insomnia as the most influential predictor, followed by mental health literacy (MHL) attitude and age. Restricted cubic spline analysis demonstrated significant nonlinear trends for insomnia, MHL attitude, and age. Insomnia and poor MHL attitude exhibited synergistic interaction, while insomnia and age showed antagonistic interaction.

Conclusion

Integrating XGBoost with SHAP explainability yields both accurate prediction and mechanistic transparency. The findings challenge conventional linear assumptions and support a dual-target prevention strategy combining insomnia management with MHL attitude modification, alongside age-stratified screening in community settings.

Introduction

Depression is a common and debilitating mental disorder characterized by persistent sadness, loss of interest or pleasure, and associated cognitive and somatic symptoms, affecting approximately 332 million people worldwide ^[1, 2]. Beyond severely diminishing quality of life, depression is closely linked to substantially elevated suicide risk—approximately 20% higher than that of people without depression ^[3]. Depression imposes the heaviest burden among mental disorders, with annual DALYs exceeding 56 million, and this burden continues to escalate amid growing global mental health challenges ^[4].

In community and clinical settings, the identification of depressive symptoms relies predominantly on subjective self-report, which gives rise to two recognition gaps: genuine patients often fail to recognize their condition, attributing persistent dysphoria and somatic discomfort to mundane stress or personality traits, thereby delaying or foregoing intervention^[5], whereas some without clinically significant distress may misinterpret transient psychological fluctuations as pathological states, leading to unnecessary help-seeking and misallocation of healthcare resources^[6, 7]. Furthermore, current screening and diagnosis depend heavily on self-administered scales that are susceptible to recall bias and social desirability effects, compromising diagnostic accuracy^[8, 9]. Against this backdrop, the objective delineation of population characteristics associated with heightened depression risk is essential for optimizing the allocation of mental health resources and improving the efficiency of early intervention.

Depression arises from a complex interplay of biological, psychological, and social determinants. Among the modifiable risk factors, insomnia has been consistently identified as a robust precursor and correlate of depression in epidemiological studies^[10]; yet the precise dose-response relationship between insomnia severity and depression risk remains incompletely characterized, with emerging evidence suggesting that this association may not follow a simple monotonic trajectory. Concurrently, mental health literacy (MHL)—encompassing the ability to recognize mental disorders, knowledge about risk factors and treatments, and attitudes toward help-seeking and stigma—has garnered increasing research attention as a potentially critical psychological determinant of mental health outcomes^[11]. Nevertheless, existing evidence examining the association between MHL and depression has predominantly adopted a linear perspective, while the nuanced, potentially non-monotonic role of specific MHL subdomains—particularly the attitudinal dimension—remains insufficiently elucidated.

Elucidating these nuanced, potentially non-monotonic associations requires methodological approaches capable of capturing nonlinearity and complex interactions. Traditional models, predicated on linear or additive assumptions between predictors and outcomes, are inherently limited in capturing nonlinear dose-response relationships and high-order interactions underlying depression risk^[12]. Therefore, a superior modeling approach capable of addressing this complexity is required. In recent years, machine learning has been increasingly applied to develop mental health prediction tools, demonstrating superior performance in modeling intricate, high-dimensional patterns without stringent distributional assumptions^[13]. However, the "black-box" nature of many algorithms constrains their interpretability and practical utility^[14]. Accordingly, a transparent and interpretable framework is essential for elucidating the mechanisms through which individual features operate and for identifying heterogeneity in risk factor profiles across diverse populations.

In response, we first evaluated five supervised classifiers—logistic regression (LR)^[15], support vector machine (SVM)^[16], random forest (RF)^[17], gradient boosting machine (GBM)^[18], and XGBoost (XGB)—to identify the optimal predictive model^[19], and subsequently integrated SHAP-based interpretability analysis to construct an explainable depression prediction framework. The SHAP (SHapley Additive

exPlanations) framework provides a principled approach to decompose model predictions into marginal contributions from individual features, thereby furnishing both global summaries of variable importance and directional effects, and local, case-specific explanations of how risk and protective factors shape individualized predictions ^[20]. Building upon this framework, the present study aims to identify key determinants of depression in a community population, and to elucidate the pairwise interactions and complex nonlinear relationships between critical features—particularly insomnia and mental health literacy attitude—and depressive symptoms, thereby providing an empirical evidence base for the formulation of community-based screening strategies and depression prevention policies.

Methods

Study Design and Participants

This study used data from the 2024–2025 Gansu Provincial Residents' Mental Health Literacy and Common Mental Disorders Survey. A multistage stratified cluster sampling strategy was employed. Participants aged 18 years or older who had resided locally for at least 6 months were included. Individuals registered in the Severe Mental Disorder Management Information System or who refused participation or were unreachable were excluded. Written informed consent was obtained from all participants prior to enrollment. The study protocol was approved by the Ethics Committee of the School of Public Health, Lanzhou University (Approval No. IRB25010201). All procedures were performed in accordance with the Declaration of Helsinki and relevant institutional guidelines.

Variables

Depressive symptoms were assessed using the Patient Health Questionnaire-9 (PHQ-9) ^[21], a validated 9-item self-report instrument measuring the severity and frequency of depressive symptoms over the preceding 2 weeks. Each item is scored on a 4-point Likert scale (0 = not at all, 1 = several days, 2 = more than half the days, 3 = nearly every day), yielding a total score from 0 to 27. For predictive modeling, depression (PHQ-9 $\geq$ 10) was operationalized as the binary outcome, with participants scoring < 10 as the reference group. The Cronbach's alpha of the PHQ-9 was 0.853.

Mental Health Literacy (MHL) was measured using the National Mental Health Literacy Questionnaire (NMHLQ) ^[22], developed by the Chinese Academy of Sciences. The instrument comprises three subdomains: (1) MHL knowledge (20 true/false items, scored 0–100); (2) MHL attitude (8 self-assessment items, scored 8–32); and (3) MHL skill (2 case-analysis items, scored 0–40). Adequate MHL was defined as simultaneously meeting subdomain thresholds (knowledge $\geq$ 80, attitude $\geq$ 24, skill $\geq$ 28); however, for the prediction model, the continuous subdomain scores were retained as separate features to maximize granular predictive information. Cronbach's alpha was 0.755 for the MHL knowledge subscale, 0.894 for the MHL attitude subscale, and 0.789 for the MHL skill subscale.

Insomnia was evaluated with the Insomnia Severity Index (ISI), a 7-item instrument rated on a 0–4 scale (total range: 0–28). A score ≥ 8 denoted clinically significant insomnia symptoms^[23]. The Cronbach's alpha of the ISI was 0.923.

Self-rated health status was measured by a single-item question with five response options: good, relatively good, moderate, relatively poor, and poor.

Alcohol consumption was defined as drinking at least once per week on average during the past 12 months; otherwise classified as non-drinker (**yes/no**). **Smoking status** was defined as having a continuous or cumulative smoking history of 6 months or more; otherwise classified as non-smoker (**yes/no**).

Regular physical exercise was defined as engaging in moderate- or higher-intensity physical activity for at least 3 sessions per week, with each session lasting ≥ 30 minutes (**yes/no**). **Weekly social activity participation** was defined as engaging in any social activity at least once per week—such as visiting friends or relatives, playing chess or cards, working, or volunteering (**yes/no**).

Machine Learning and Explainability Analysis

Five supervised classifiers—logistic regression, support vector machine, random forest, gradient boosting machine, and XGBoost—were evaluated using nested 10-fold cross-validation with a 5-fold inner loop. To address class imbalance, SMOTE was applied exclusively to the training folds. Model performance was quantified by classification error, accuracy, F1-score, ROC-AUC, sensitivity, specificity, and PR-AUC, reported as mean $\pm$ standard deviation across the outer folds. Additionally, 95% confidence intervals for ROC-AUC and PR-AUC were estimated via 1,000 bootstrap resamples. The final model was selected based on the highest AUROC among models achieving a minimum sensitivity of 0.60, with AUPRC as a secondary consideration.

To identify key predictors and their directional effects, SHAP values were computed from the final model. A beeswarm plot was generated to visualize the distribution and direction of SHAP values across samples, and predictors were ranked globally by mean absolute SHAP values. Representative cases—true positive, true negative, and false negative—were selected from the hold-out test set to generate SHAP force plots, illustrating how each feature drove the predicted probability from the base value to the final output.

For predictors flagged by SHAP analysis as potentially having nonlinear associations with depression or participating in pairwise interactions, three supplementary analyses were conducted. First, multivariable logistic regression with 4-knot restricted cubic splines was used to model potential nonlinear relationships, with the median of each predictor as the reference point and all other covariates adjusted. Nonlinearity was formally tested using the likelihood ratio test by comparing the RCS model against a linear logistic model. Second, multiplicative interactions were visualized using three-dimensional

probability surface plots and contour plots, depicting the joint predicted probability of depression across the interacting variables. Third, additive interactions were quantified using the relative excess risk due to interaction (RERI), attributable proportion (AP), and synergy index (S), with 95% confidence intervals derived from bootstrap resampling. A positive RERI indicates a super-additive (synergistic) effect, whereas a negative RERI indicates a sub-additive (antagonistic) effect.

Result

A total of 14,335 participants were included (mean age 56.38 ± 17.04 years; 48.8% male). The prevalence of depression symptoms was 4.98%; adequate mental health literacy and insomnia were observed in 11.08% and 21.25%. Depression severity was significantly associated with all characteristics except alcohol consumption and ethnicity. (Table 1)

Table 1
Prevalence of Depression Symptoms by Participant Characteristics (N = 14335)

Variables	No depression (n, %)	depression (n, %)	χ^2	P-value
Overall	13621 (95.02)	714 (4.98)		
Mental health literacy			73.58	< 0.001
Inadequate	12041 (94.47)	705 (5.53)		
Adequate	1580 (99.43)	9 (0.57)		
Insomnia			1418.42	< 0.001
Yes	2493 (81.85)	553 (18.15)		
No	11128 (98.57)	161 (1.43)		
Gender			16.11	< 0.001
Male	6696 (95.77)	296 (4.23)		
Female	6925 (94.31)	418 (5.69)		
Residence			16.51	< 0.001
Urban	5819 (95.88)	250 (4.12)		
Rural	7802 (94.39)	464 (5.61)		
Age (years)			45.88	< 0.001
18-	1081 (92.95)	82 (7.05)		
30-	2629 (95.36)	128 (4.64)		
45-	3305 (96.47)	121 (3.53)		
60-	4737 (95.18)	240 (4.82)		
75-	1869 (92.89)	143 (7.11)		
Ethnicity			0.03	0.866
Han	12548 (95.01)	659 (4.99)		
Other	1073 (95.12)	55 (4.88)		
Education level			84.59	< 0.001
Primary or below	2963 (92.22)	250 (7.78)		
Junior high	3394 (96.39)	127 (3.61)		
Senior high/vocational	3252 (96.56)	116 (3.44)		

Variables	No depression (n, %)	depression (n, %)	χ^2	<i>P</i>-value
Undergraduate or above	4012 (94.78)	221 (5.22)		
Occupation			13.84	0.003
Farmer	8143 (94.77)	449 (5.23)		
Service/manual worker	2771 (95.45)	132 (4.55)		
Professional or Technical worker	1526 (94.14)	95 (5.86)		
Other	1181 (96.88)	38 (3.12)		
Current marital status			121.83	< 0.001
Single	919 (92.55)	74 (7.45)		
Married	11401 (95.89)	489 (4.11)		
Other	1301 (89.60)	151 (10.40)		
Living arrangement			63.36	< 0.001
Living with family	12847 (95.38)	622 (4.62)		
Living alone	540 (88.82)	68 (11.18)		
Other	234 (90.70)	24 (9.30)		
Personal monthly income,CNY			50.38	< 0.001
<1,500	6742 (93.78)	447 (6.22)		
1,500-	3667 (96.65)	127 (3.35)		
3,000-	2434 (96.05)	100 (3.95)		
5,000-	635 (94.92)	34 (5.08)		
Alcohol consumption			0.29	0.591
No	11965 (94.98)	632 (5.02)		
Yes	1656 (95.28)	82 (4.72)		
Smoking status			0.02	0.889
No	10447 (95.03)	546 (4.97)		
Yes	3174 (94.97)	168 (5.03)		
Weekly social activity participation			200.15	< 0.001
Yes	11312 (96.22)	444 (3.78)		

Variables	No depression (n, %)	depression (n, %)	χ^2	<i>P</i>-value
No	2309 (89.53)	270 (10.47)		
Regular physical exercise			122.04	< 0.001
Yes	8076 (96.72)	274 (3.28)		
No	5545 (92.65)	440 (7.35)		
Self-rated health status			1302.87	< 0.001
Good	5892 (98.71)	77 (1.29)		
Relatively good	4619 (97.26)	130 (2.74)		
Fair	2667 (89.44)	315 (10.56)		
Relatively poor	376 (71.89)	147 (28.11)		
Poor	67 (59.82)	45 (40.18)		

Random forest and XGBoost exhibited comparable discriminative performance, with AUROC values of 0.886 and 0.884, respectively, and AUPRC values of 0.322 and 0.328. However, random forest yielded a sensitivity of only 0.380, whereas XGBoost achieved a substantially higher sensitivity of 0.690 while maintaining a satisfactory specificity of 0.883. Given the low depression prevalence (4.98%) in this sample, sensitivity is a critical metric for minimizing false negatives in community screening. The discriminative performance of the five classifiers is illustrated in Figure 1. Therefore, XGBoost was selected as the final model for subsequent depression prediction analysis. The detailed predictive performance metrics are summarized in Table 2.

Table 2
Predictive performance of five machine learning classifiers for depression.

Machine learner	Classification error rate	Accuracy	F-beta	Area under the ROC curve	Sensitivity	Specificity	Area under the PR curve
LR	0.251	0.749	0.229	0.829	0.748	0.749	0.241
SVM	0.256	0.744	0.230	0.833	0.765	0.743	0.249
RF	0.070	0.931	0.353	0.886	0.380	0.959	0.322
GB	0.154	0.845	0.315	0.875	0.708	0.852	0.313
XGB	0.127	0.873	0.352	0.884	0.690	0.883	0.328

Figure 2 presents the SHAP-based feature importance and directional effects from the optimal XGBoost model. Insomnia was the most influential predictor, followed by MHL attitude, age, sex, social activity regularity, and occupation. For insomnia, lower severity scores clustered predominantly in the negative SHAP region, whereas higher severity scores shifted rightward toward zero and the positive region; notably, low-score points were scattered across both negative and positive SHAP values, suggesting potential nonlinear associations and interactions with other variables. For MHL attitude, higher scores were associated with increased depression risk, with a similarly dispersed distribution of low-score points. Older age and female sex were linked to higher predicted probability, while regular exercise, social participation, and better self-rated health showed negative SHAP values. Socioeconomic indicators (education, personal and household income) and other demographic factors (smoking, drinking, marital status, ethnicity, and living arrangement) exhibited relatively lower importance.

Figure 3 shows the SHAP interaction matrix for the top seven predictors. The diagonal panels display the distributions of main-effect SHAP values for each feature. Wide dispersions were observed for insomnia, MHL attitude and age, suggesting nonlinear main effects on depression risk. Most off-diagonal panels showed interaction values clustered tightly around zero, indicating limited pairwise interactions overall. However, several off-diagonal cells—particularly those involving insomnia with MHL attitude and age—exhibited somewhat broader distributions, suggesting that the effect of insomnia may be modified by these factors.

Figure 5 shows the multiplicative and additive interaction effects between insomnia and poor MHL attitude (reverse-scored), and between insomnia and age. The 3D response surface for insomnia and poor MHL attitude (Fig. 5A) exhibited a convex upward pattern, with the predicted probability of depression rising sharply to nearly 0.9 when both scores increased simultaneously. Multiplicative interaction testing was significant ($P = 0.015$). For insomnia and age, the 3D response surface (Fig. 5B) showed a sub-additive pattern below the simple additive plane. Multiplicative interaction testing was also significant ($P = 0.011$). Additive interaction analysis for both pairs (Fig. 5C) revealed a significant synergistic effect for insomnia and poor MHL attitude ($\text{RERI} = 0.030$, 95% CI: 0.016–0.044; $\text{AP} = 0.118$, 95% CI: 0.065–0.160; $S = 1.185$, 95% CI: 1.106–1.242) and a significant sub-additive effect for insomnia and age ($\text{RERI} = -0.006$, 95% CI: -0.008 to -0.004; $\text{AP} = -0.296$, 95% CI: -0.464 to -0.167; $S = 0.713$, 95% CI: 0.603–0.821).

Discussion

This study developed a machine learning-based prediction model for depression in a large community sample from Northwest China. SHAP analysis identified insomnia as the most influential predictor, followed by mental health literacy attitude and age. Restricted cubic spline modeling further revealed non-linear relationships between insomnia, MHL attitude, and depression risk. Additionally, significant synergistic and antagonistic interactions were observed between insomnia and MHL attitude, and between insomnia and age, respectively. These findings illustrate how integrating flexible predictive

modeling with post-hoc interpretability can uncover complex predictor–outcome relationships that may inform targeted community-based prevention strategies.

The discriminative performance of the final XGBoost model—excellent by conventional criteria—places it in the upper tier of comparable depression prediction studies ^[24, 25]. This improved performance may be attributed to two methodological strengths. First, we applied SMOTE exclusively to the training set to mitigate class imbalance—critical given the low prevalence of severe depression in this study—thereby enhancing the model’s sensitivity to minority-class patterns without distorting the validation distribution^[26]. SMOTE has been increasingly adopted in mental health prediction models to address skewed outcome distributions and has demonstrated superior performance over simple random oversampling by generating synthetic minority samples rather than duplicating existing cases ^[27]. Second, XGBoost captures complex, high-order feature interactions without prespecified functional forms, enabling it to model nonlinear relationships and variable dependencies that traditional logistic regression cannot accommodate ^[19].

It is noteworthy that the AUPRC of the final model remained relatively modest despite the robust AUROC. This is expected under extreme class imbalance: with a positive class proportion of only 4.98%, the baseline AUPRC approximates the prevalence itself, and even well-performing classifiers yield values far below their AUROC counterparts ^[28]. Empirical studies with similarly imbalanced outcomes have further demonstrated that AUPRC is directly governed by event rate and sample size, with values inherently constrained in rare-outcome settings regardless of modeling strategy ^[29]. Beyond this mathematical property, the limited predictive granularity of questionnaire-based features further constrains the precision–recall trade-off. The predictors in this study, spanning sociodemographic, lifestyle, and psychological domains, capture only a fraction of depressive symptom variance. Unmeasured biological and contextual factors known to influence depression risk (e.g., inflammatory markers, recent life events) were unavailable in this cross-sectional survey ^[30, 31]. Consequently, the modest AUPRC reflects both class imbalance and the inherent ceiling of questionnaire-based features, yet this does not compromise the interpretability or clinical utility of our SHAP-based framework ^[20]. Thus, the AUPRC primarily signals the need for richer feature sets in future models rather than a limitation of the modeling strategy itself.

SHAP analysis identified insomnia as the dominant predictor, followed by MHL attitude and age. Closer inspection revealed anomalous distributional patterns for these two variables that defied linear assumptions: some low-severity insomnia cases exhibited unexpectedly high-risk contributions, while MHL attitude showed a reversed, heterogeneous gradient. Such patterns—entirely invisible to conventional logistic regression—directly motivated the subsequent restricted cubic spline analysis and interaction testing, illustrating how explainable machine learning can uncover anomalies that generate novel, testable hypotheses.

The most striking finding was the non-monotonic insomnia–depression relationship: moderate severity conferred the highest risk, whereas severe insomnia was paradoxically associated with low risk. This

counterintuitive pattern likely reflects two broad processes. First, severe cases may be underrepresented or confounded in cross-sectional data: those receiving treatment show attenuated concurrent depression, while the most functionally impaired may be unable to participate in community surveys^[32, 33]. Second, severe chronic insomnia may become biologically or temporally decoupled from depressive states through neuroadaptive changes (e.g., upregulated sleep homeostasis, blunted subjective complaint) or through persistence as a residual symptom after depression remission^[34–36]. Irrespective of mechanism, this nonlinear pattern challenges the simplistic "more insomnia, more depression" assumption. The moderate-severity window—where sleep disturbance is distressing yet not medically managed—may represent a critical intervention target. Longitudinal designs with objective sleep monitoring and treatment history are warranted to disentangle these hypotheses.

Rather than a linear association, the relationship between MHL attitude and depression risk followed an inverted U-shaped trajectory, with peak vulnerability at moderate attitudinal levels. This pattern reflects an **awareness–action gap** analogous to the **insight paradox** observed in psychiatric populations^[37, 38], wherein individuals with moderate scores recognize their emotional distress yet lack the self-efficacy or help-seeking intention to translate awareness into effective action. This state of heightened insight without adequate coping resources may amplify helplessness and entrapment,

paradoxically increasing depression risk. Conversely, low attitudinal scores may reflect **defensive pessimism**—a protective strategy in which individuals lower expectations and disengage from mental health concerns to shield themselves from anticipated psychological distress^[39]. In the absence of severe chronic stress, this low-expectancy stance may be adaptive, preventing the affective dysregulation and rumination that accompany high awareness without resolution. This interpretation cautions against regarding low attitudes as inherently pathological; without concurrent stress exposure, such attitudes may signal resilience rather than vulnerability.

In contrast, MHL knowledge showed a straightforward monotonic protective gradient. The divergence between attitude and knowledge carries two intervention implications. First, the highest-risk profile may be defined by the intersection of moderate attitude and low knowledge—individuals who possess insufficient literacy to access resources while simultaneously occupying the attitudinal "danger zone" where awareness exceeds coping capacity. Second, public health campaigns focused solely on knowledge dissemination may fail if attitudinal barriers remain unaddressed. Consequently, attitude-targeted interventions (e.g., contact-based stigma reduction, peer support promotion) should be integrated with, or potentially prioritized over, pure knowledge dissemination in community prevention programs.

The 3D predicted probability surface revealed a convex joint effect of insomnia and poor MHL attitude, with depression risk rising sharply to nearly 0.9 when both conditions co-occurred. Additive interaction measures confirmed significant positive synergy. Self-stigma and other negative attitudes toward mental health, together with insomnia, operate as interconnected nodes within a broader psychopathology network, amplifying affective dysregulation through shared bridge symptoms such as daytime

dysfunction and emotional instability^[40], suggesting that attitudinal barriers may function similarly to potentiate insomnia-related depression risk. Clinically, patients with chronic insomnia experience significant stigma, which acts as a barrier to treatment-seeking and reinforces the tendency to attribute sleep disturbances to benign causes rather than clinical conditions^[41]; when coupled with negative mental health attitudes that impede help-seeking and prolong untreated insomnia^[42], these attitudinal barriers may perpetuate a self-sustaining cycle. Neurobiologically, chronic insomnia and depression share common pathophysiological substrates^[43]; when these biological changes coincide with prolonged untreated insomnia, they likely produce synergistic effects beyond simple additivity.

Conversely, the 3D surface for insomnia and age showed a concave joint effect below the simple additive plane, and additive interaction tests revealed significant antagonism. This suggests that the depressogenic impact of insomnia is attenuated in older adults. One mechanism is age-related normalization: older adults may perceive sleep disturbances as an expected, normative concomitant of aging rather than a pathological stressor warranting clinical concern, thereby reducing its psychological salience and downstream affective consequences.^[44] This is supported by epidemiological data showing that the association between insomnia and depression is substantially stronger in younger and middle-aged adults than in those aged 50 years or older.^[46] Consequently, insomnia screening should be prioritized in younger and middle-aged populations, where the condition exerts its strongest relative effect on depression risk, whereas older adults may require broader psychosocial interventions targeting bereavement, social isolation, and physical comorbidity.

From a public health standpoint, our findings support a **dual-target strategy**: insomnia management should be coupled with MHL attitude modification (stigma reduction, help-seeking promotion), as the synergistic interaction indicates that isolated sleep interventions leave substantial residual risk unaddressed. The moderate-insomnia window warrants prioritized screening, as this group is most amenable to early prevention yet least likely to seek care. Finally, the antagonistic interaction with age suggests **age-stratified prevention**: younger adults with insomnia require urgent, targeted intervention, whereas older adults may benefit more from broader psychosocial support than from sleep-centric programs alone.

Conclusion

This study developed an interpretable machine learning framework for depression prediction in a large community sample from Northwest China. The XGBoost classifier achieved satisfactory discriminative performance (AUROC = 0.884) with clinically meaningful sensitivity (0.690), demonstrating its utility for community-based screening. SHAP analysis identified insomnia as the dominant predictor, followed by mental health literacy attitude and age. Restricted cubic spline modeling further revealed significant nonlinear relationships: moderate insomnia severity conferred the highest depression risk, while MHL attitude exhibited an inverted U-shaped association, peaking at moderate attitudinal levels. Additionally, insomnia and poor MHL attitude demonstrated a synergistic interaction, whereas insomnia and age

showed antagonism, suggesting that the depressogenic effect of sleep disturbance is attenuated in older adults.

These findings challenge conventional linear assumptions and support a dual-target prevention strategy that integrates insomnia management with MHL attitude modification, particularly among younger and middle-aged adults where insomnia exerts its strongest effect. The moderate-insomnia window warrants prioritized screening, as this group is most amenable to early intervention yet least likely to seek care. Future longitudinal studies incorporating objective sleep measures, biomarkers, and richer contextual features are needed to validate these cross-sectional associations and refine the precision of community-based prediction models.

Declarations

Funding

This study received no specific funding.

Author Contribution

Y.G. conceived the study, performed the machine learning modeling and SHAP interpretability analysis, generated all figures and tables, and wrote the original draft. X.L. supervised the study design, coordinated data acquisition and ethical approval, provided critical intellectual input, and led the manuscript revision. N.Z. contributed to data curation, literature review, and preliminary statistical analysis. L.X. participated in data collection, questionnaire administration, and verification of statistical results. W.P. assisted with data collection, quality control, and manuscript review. All authors read and approved the final manuscript.

Data Availability

The datasets used and/or analysed during the current study are not publicly available due to participant privacy and ethical restrictions, but are available from the corresponding author on reasonable request. The analysis code used for machine learning modeling, SHAP interpretability, and statistical visualization is available from the corresponding author upon reasonable request.

References

1. WHO, Depression. 2026. <https://www.who.int/news-room/fact-sheets/detail/depression>. Accessed 10 Jul 2026.
2. Santomauro DF, Miller PA, Shadid J, et al. Updated trends in the global prevalence and burden of mental disorders, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023.

- Lancet. 2026;407(10543):2040–64.
3. Li S, Pan W, Yip PSF, et al. Uncovering the heterogeneous effects of depression on suicide risk conditioned by linguistic features: a double machine learning approach. *Comput Hum Behav*. 2024;152:10.
 4. Institute for Health Metrics and Evaluation. Global Burden of Disease Study 2023 Results Tool. 2026. <http://ghdx.healthdata.org/gbd-results-tool>. Accessed 10 Jul 2026.
 5. Schomerus G, Stolzenburg S, Freitag S, et al. Stigma as a barrier to recognizing personal mental illness and seeking help: a prospective study among untreated persons with mental illness. *Eur Arch Psychiatry Clin Neurosci*. 2019;269(4):469–79.
 6. Durà-Vilà G, Littlewood R, Leavey G. Depression and the medicalization of sadness: conceptualization and recommended help-seeking. *Int J Soc Psychiatry*. 2013;59(2):165–75.
 7. Nielsen TH. Disorder or distress? The hermeneutical injustices of overdiagnosis within psychiatry. *Synthese*. 2025;205(2):65.
 8. Thombs B, Kwakkenbos L, Levis A, et al. Addressing overestimation of the prevalence of depression based on self-report screening questionnaires. *CMAJ*. 2018;190:E44–9.
 9. Levis B, Benedetti A, Ioannidis JPA, et al. Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: individual participant data meta-analysis. *J Clin Epidemiol*. 2020;122:115–e128111.
 10. Baglioni C, Battagliese G, Feige B, et al. Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. *J Affect Disord*. 2011;135(1):10–9.
 11. Yang Y, Miao L, Liu X, et al. The impact of mental health literacy on depression, anxiety and well-being among vocational nursing students: mediating roles of resilience. *Front Psychiatry*. 2025;16:1585642.
 12. Takahashi Y, Ueki M, Yamada M, et al. Improved metabolomic data-based prediction of depressive symptoms using nonlinear machine learning with feature selection. *Transl Psychiatry*. 2020;10(1):157.
 13. Lu C, Wan S, Liu Z. Determinants of depressive symptoms in multinational middle-aged and older adults. *NPJ Digit Med*. 2025;8(1):501.
 14. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1(5):206–15.
 15. Cox DR. The regression analysis of binary sequences. *J R Stat Soc Ser B*. 1958;20(2):215–32.
 16. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–97.
 17. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
 18. Friedman J. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2000;29:1189–232.
 19. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco: ACM;

2016. pp. 785–794.
20. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. arXiv. 2017. arXiv:1705.07874. <https://arxiv.org/abs/1705.07874>. Accessed 15 Jul 2026.
 21. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–13.
 22. Institute of Psychology, Chinese Academy of Sciences. National Mental Health Literacy Questionnaire. 2026. <https://cmhr.psych.ac.cn/MentalHealthLiteracy.html>. Accessed 15 Jul 2026.
 23. Morin CM. *Insomnia: psychological assessment and management*. New York: Guilford Press; 1993.
 24. Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Assessing the fit of the model*. Applied Logistic Regression. 3rd ed. Hoboken: Wiley; 2013. pp. 153–225.
 25. Nickson D, Meyer C, Walasek L, et al. Prediction and diagnosis of depression using machine learning with electronic health records data: a systematic review. *BMC Med Inf Decis Mak*. 2023;23(1):271.
 26. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–57. 10.1613/jair.953.
 27. Mohseni-Takaloo S, Mohseni H, Mozaffari-Khosravi H, et al. The effect of data balancing approaches on the prediction of metabolic syndrome using non-invasive parameters based on random forest. *BMC Bioinformatics*. 2024;25(1):18.
 28. 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.
 29. Jacobucci R, Li X. Does minority case sampling improve performance with imbalanced outcomes in psychological research? *J Behav Data Sci*. 2022;2(1):59–74.
 30. Hakulinen C, Elovainio M, Pulkki-Råback L, et al. Personality and depressive symptoms: individual participant meta-analysis of 10 cohort studies. *Depress Anxiety*. 2015;32(7):461–70.
 31. Abdulla H, Maalouf M, Jelinek HF. Machine learning for the prediction of depression progression from inflammation markers. In: *Proceedings of the 45th Annual International Conference of the IEEE Engineering in Medicine and Biology Society: IEEE*; 2023. pp. 1–4. 10.1109/embc40787.2023.10340436
 32. Bhattacharya S, Goicoechea C, Heshmati S, et al. Efficacy of cognitive behavioral therapy for anxiety-related disorders: a meta-analysis of recent literature. *Curr Psychiatry Rep*. 2023;25(1):19–30.
 33. Czeisler MÉ, Wiley JF, Czeisler CA, et al. Uncovering survivorship bias in longitudinal mental health surveys during the COVID-19 pandemic. *Epidemiol Psychiatr Sci*. 2021;30:e45.
 34. Alkadhi K, Zagaar M, Alhaider I, et al. Neurobiological consequences of sleep deprivation. *Curr Neuropharmacol*. 2013;11(3):231–49.
 35. Simpson NS, DiIombi M, Scott-Sutherland J, et al. Repeating patterns of sleep restriction and recovery: do we get used to it? *Brain Behav Immun*. 2016;58:142–51.
 36. Iovieno N, Nieuwenhuizen A, Clain A, et al. Residual symptoms after remission of major depressive disorder with fluoxetine and risk of relapse. *Depress Anxiety*. 2011;28:137–44.

37. Lopez-Morinigo JD, David AS. Is too much insight bad for you? *Br J Psychiatry*. 2024;225(4):454–7.
38. Jorm AF. Mental health literacy: empowering the community to take action for better mental health. *Am Psychol*. 2012;67(3):231–43.
39. Thomas SP. Defensive pessimism. In: Zeigler-Hill V, Shackelford TK, editors. *Encyclopedia of Personality and Individual Differences*. Cham: Springer; 2020. pp. 1036–8. 10.1007/978-3-319-24612-3_1061.
40. Zhang X, Lin R, Zhang Z, et al. Comorbidity network of self-stigma, insomnia, and mental health in chronic disease patients: a network analysis. *Psychol Res Behav Manag*. 2025;18:2333–45.
41. He S, Ke XJ, Wu Y, Kong XY, Wang Y, Sun HQ, Xia DZ, Chen GH. The stigma of patients with chronic insomnia: a clinical study. *BMC Psychiatry*. 2022;22(1):449. 10.1186/s12888-022-04091-y.
42. Rauch L, Schneider T, Wendt C. Seeking professional help for sleep-related complaints. *Front Public Health*. 2024;12:1430574. 10.3389/fpubh.2024.1430574.
43. Freeman D, Sheaves B, Waite F, et al. Sleep disturbance and psychiatric disorders. *Lancet Psychiatry*. 2020;7(7):628–37.
44. Ghazi SN, Behrens A, Berglund JS, et al. Examining sleep health and its associations with technology use among older adults in Sweden: insights from a population-based study. *BMC Public Health*. 2025;25(1):2896.
45. Sawadogo W, Njoku A, Jegede J. Insomnia symptoms, mental health diagnosis, mental health care utilization, and perceived barriers in U.S. males and females. *J Clin Med*. 2025;14(9):2989.

Figures

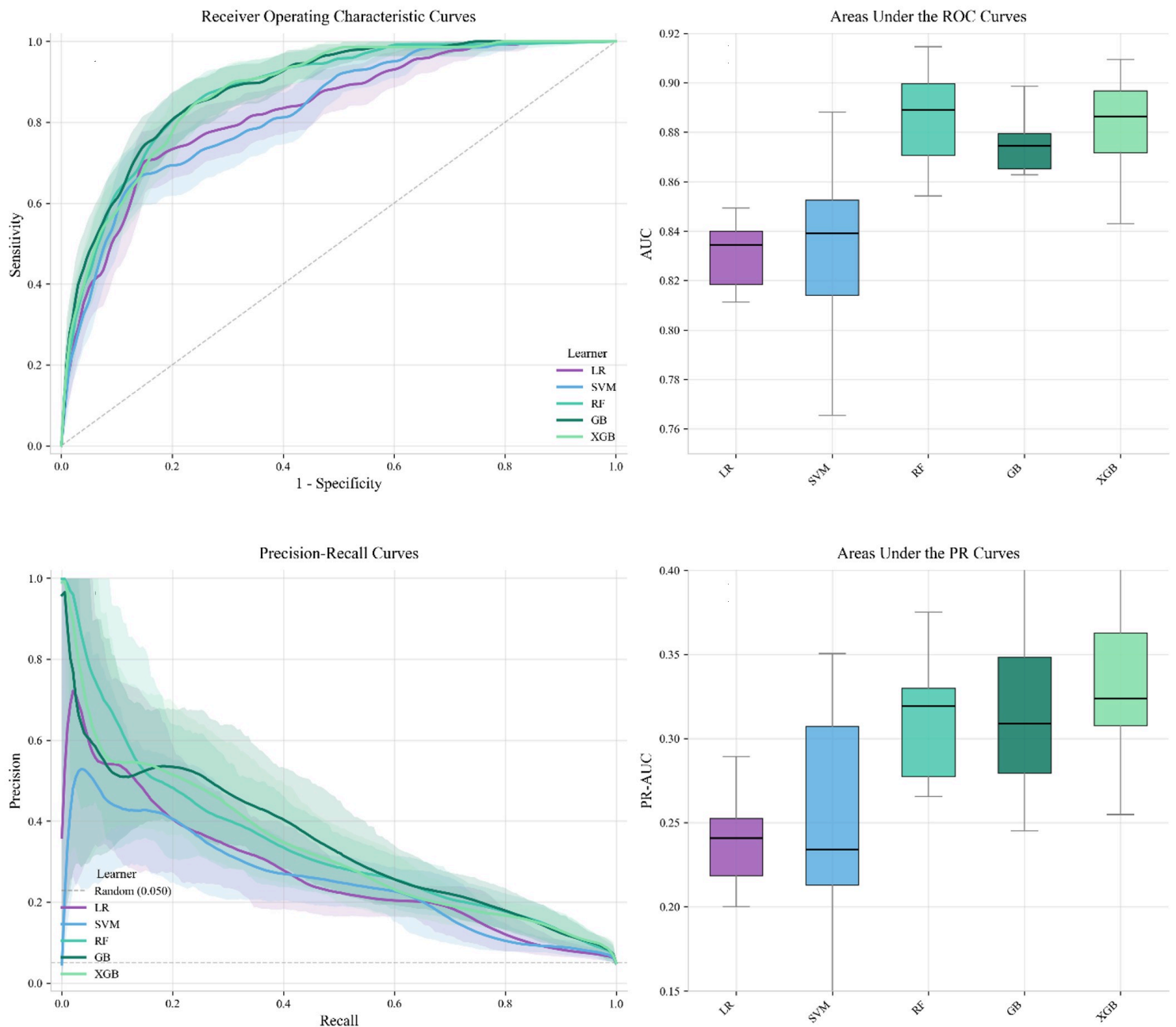


Figure 1

Performance comparison of five machine learning classifiers for depression prediction. (A) Receiver operating characteristic (ROC) curves. (B) Areas under the ROC curves (AUC) across the outer folds of nested cross-validation. (C) Precision–recall curves. (D) Areas under the precision–recall curves (PR-AUC) across the outer folds. LR, logistic regression; SVM, support vector machine; RF, random forest; GB, gradient boosting; XGB, XGBoost.

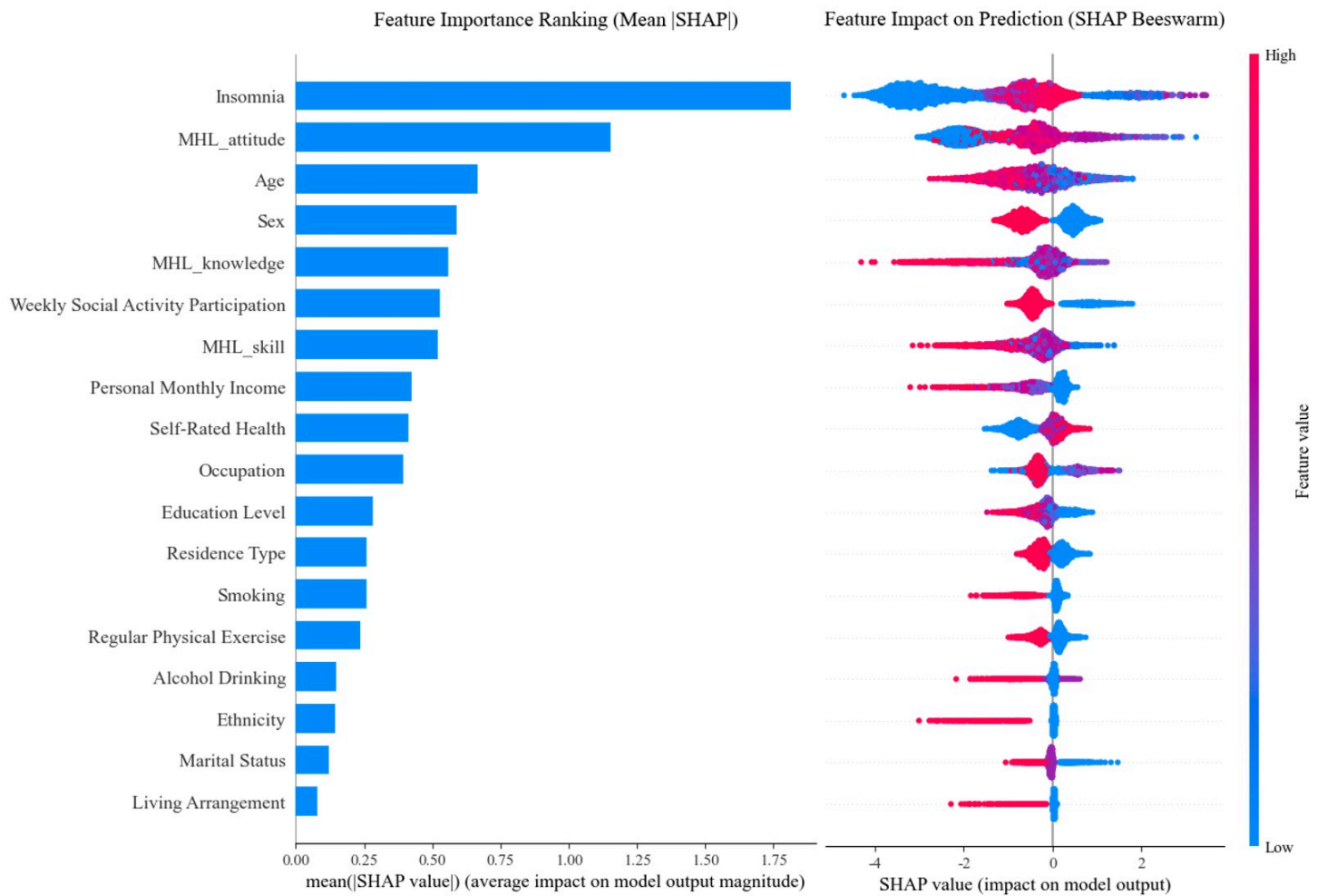


Figure 2

SHAP summary plot illustrating feature importance and directional effects from the optimal XGBoost model.

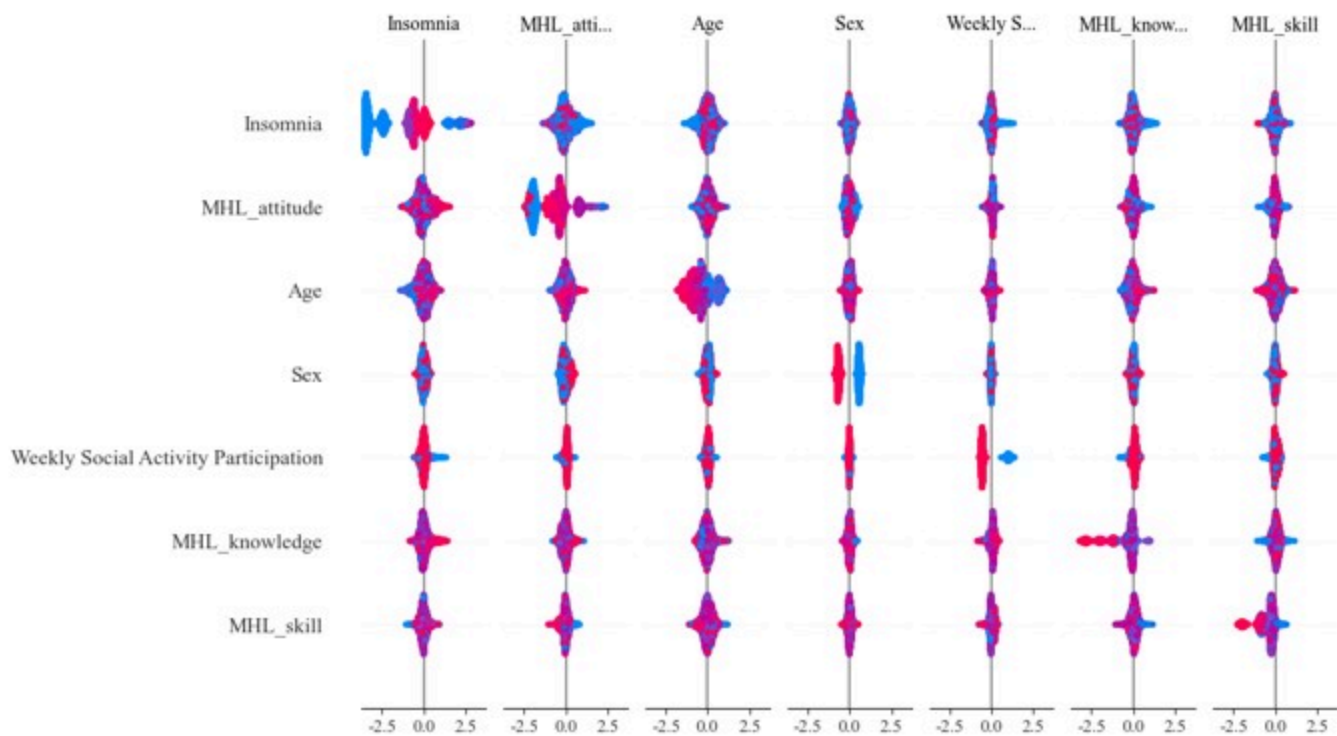


Figure 3

SHAP interaction value matrix of predictors. The diagonal panels display the main-effect SHAP value distribution for each variable, and the off-diagonal panels illustrate pairwise interaction effects.

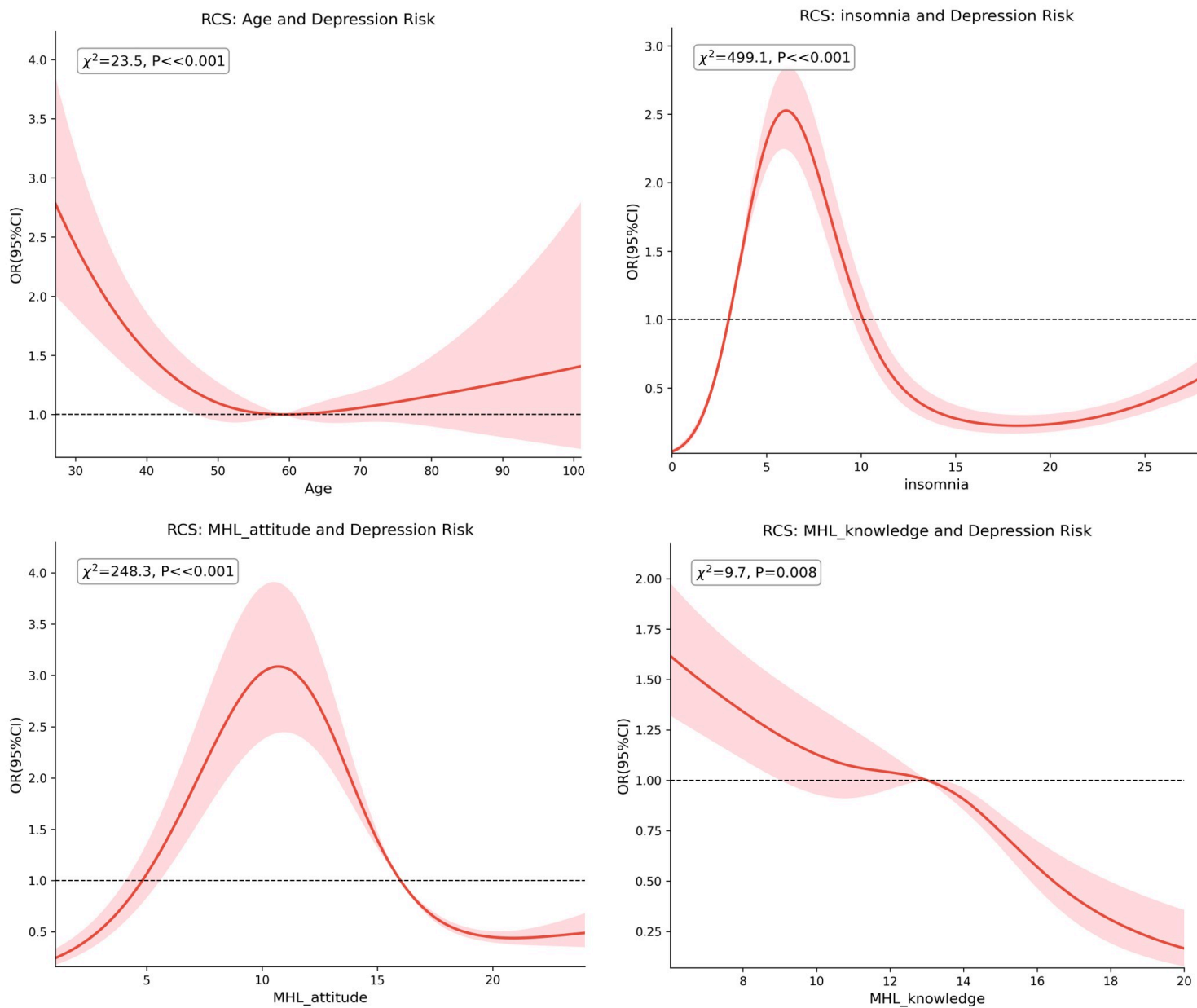


Figure 4

Nonlinear relationships of age, insomnia, MHL attitude, and MHL knowledge with depression risk

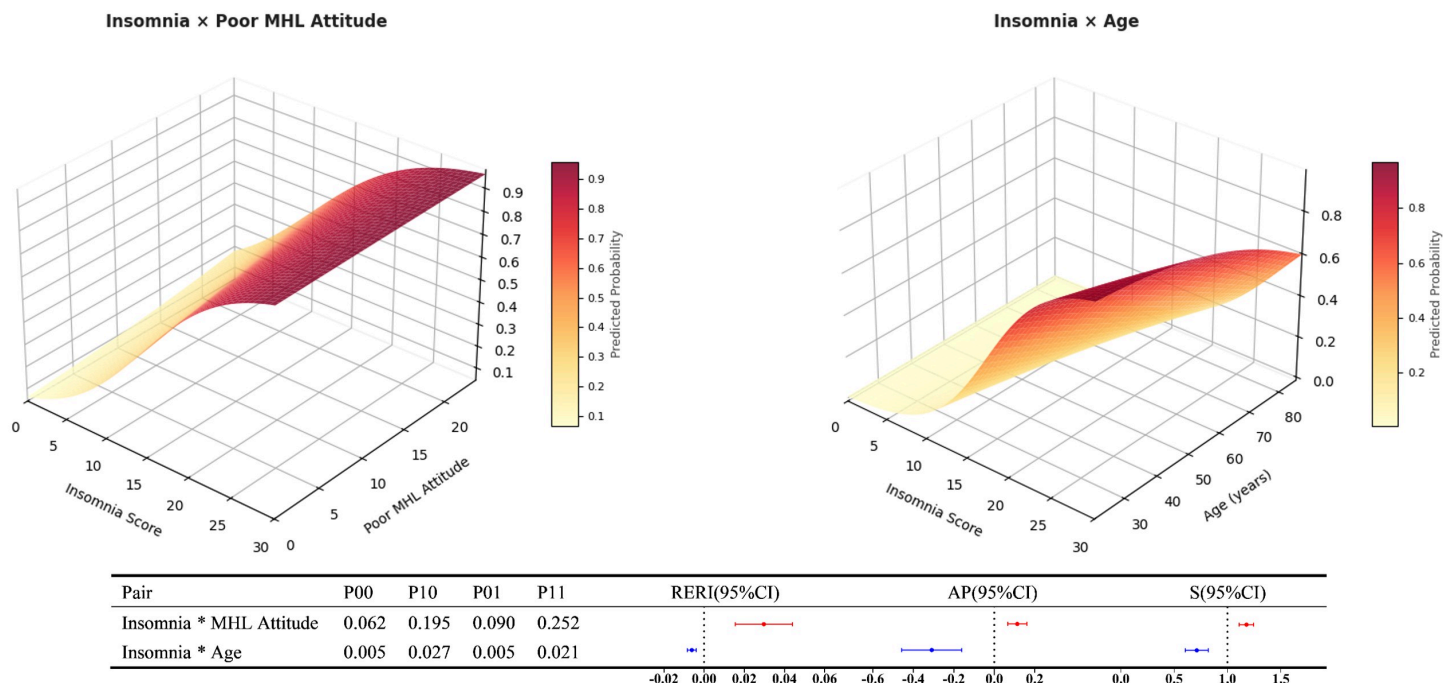


Figure 5

Multiplicative and additive interaction effects between insomnia and poor MHL attitude, and between insomnia and age, on depression risk. Upper panels: three-dimensional predicted probability response surface plots showing the joint effects of insomnia and poor MHL attitude (A) and insomnia and age (B). Lower panel: forest plots of additive interaction measures (RERI, relative excess risk due to interaction; AP, attributable proportion; S, synergy index) with 95% confidence intervals.

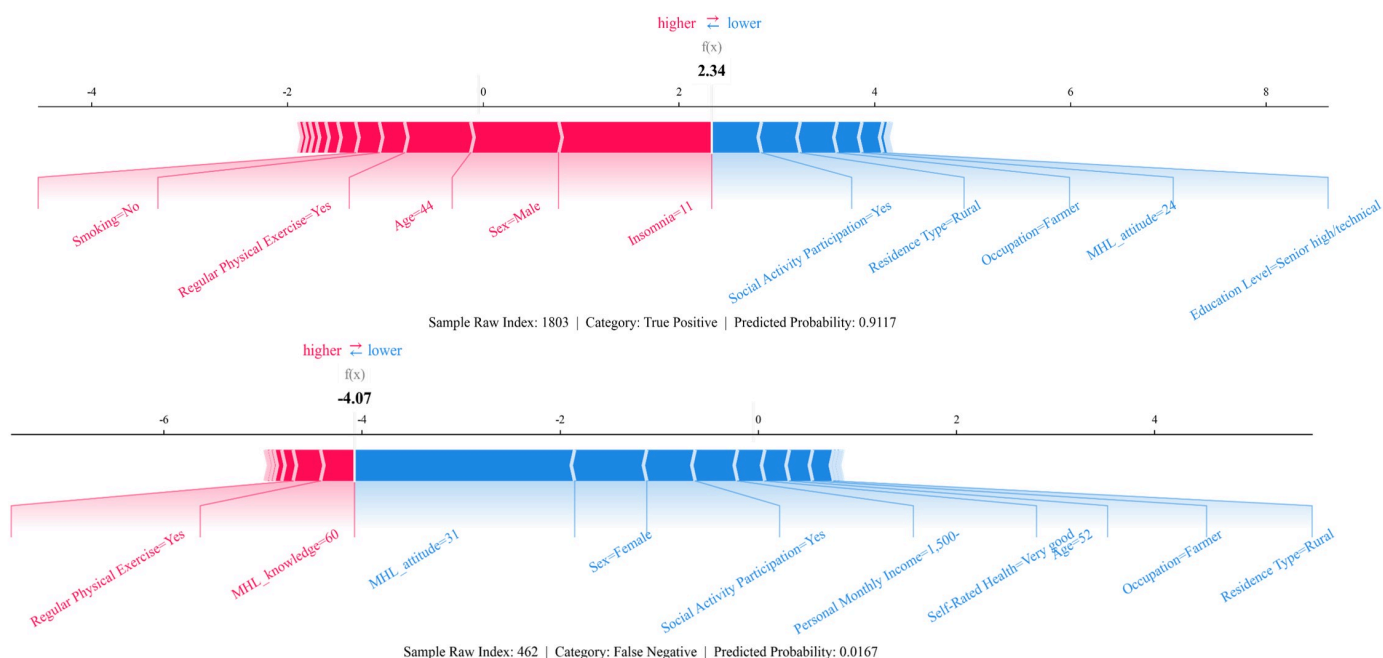


Figure 6

Force plots of feature contributions for individual prediction.