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.

Factors associated with mental health literacy among residents of Fengtai District, Beijing: an interpretable machine learning analysis

Jiaonan Liu
Fengtai District Mental Health Center (Fengtai District Mental Health Prevention and Treatment Hospital)
Jingqian Wang
Fengtai District Mental Health Center (Fengtai District Mental Health Prevention and Treatment Hospital)
Yujun Lang
Fengtai District Mental Health Center (Fengtai District Mental Health Prevention and Treatment Hospital)
Shuangyi Qi
Fengtai District Mental Health Center (Fengtai District Mental Health Prevention and Treatment Hospital)
Chen Zhang
Fengtai District Mental Health Center (Fengtai District Mental Health Prevention and Treatment Hospital)
Zeyu Tian
Fengtai District Mental Health Center (Fengtai District Mental Health Prevention and Treatment Hospital)
Zhiwu Li
Fengtai District Mental Health Center (Fengtai District Mental Health Prevention and Treatment Hospital)
Qi Tian
tianqi09190126.com

Fengtai District Mental Health Center (Fengtai District Mental Health Prevention and Treatment Hospital)

Research Article

Keywords: mental health literacy, machine learning, Beijing

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11212048/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

Mental health literacy (MHL) remains moderate to low in China. Using an interpretable machine learning (ML) approach, we aimed to identify factors associated with MHL attainment among adults in Fengtai District, Beijing, and to quantify their relative contributions.

Methods

We conducted a secondary analysis of the 2022 Fengtai District Mental Health Literacy Survey, a multi-stage cross-sectional survey. MHL attainment was assessed with the Beijing Resident Mental Health Literacy Monitoring Questionnaire, requiring all three dimension-specific thresholds. Six ML models (logistic regression [LR], random forest [RF], *k*-nearest neighbors [KNN], decision tree [DT], naive Bayes [NB], and support vector machine [SVM]) were trained on 15 biopsychosocial predictors. RF was prespecified as the primary model. Internal validation used 1000 bootstrap replicates with pooled out-of-bag predictions. Discrimination was assessed primarily by the area under the precision-recall curve (PR-AUC) and secondarily by the area under the ROC curve (AUC). Calibration was assessed by the Brier score and a logistic calibration model. Feature importance was assessed by Gini impurity and SHapley Additive exPlanations (SHAP).

Results

Of 2004 respondents, 403 (20.11%) attained MHL. All six models exceeded the no-skill PR-AUC baseline of 0.201, with 95% CI lower bounds above it. RF ranked first (PR-AUC 0.364, 95% CI 0.304–0.424), followed by SVM (0.342) and LR (0.325); the others ranged from 0.270 to 0.315, and AUC from 0.604 to 0.684. At the Youden-optimal cut-off, recall ranged from 0.254 to 0.698 and precision was at most 0.354. The RF calibration slope was 1.00 (95% CI 0.84–1.19), the intercept − 0.92, and the Brier score fell from 0.177 to 0.147 after recalibration. Gini impurity ranked age highest (23.03%) and SHAP life satisfaction highest (15.00%); subdistrict, educational attainment, and monthly income were among the top five in both rankings. Life satisfaction, education, income, and prior training contributed positively, whereas older age and insomnia, depressive, or anxiety symptoms contributed negatively.

Conclusions

Interpretable ML quantified the relative contributions and directions of factors associated with MHL attainment, with educational attainment, life satisfaction, and subdistrict consistently among the leading factors. Given the modest discrimination, the primary model was used for interpretation rather than individual-level screening.

Background

Mental health literacy (MHL) was first defined by Jorm et al. as the knowledge and beliefs about mental disorders that aid their recognition, management, and prevention [1]. Building on this foundational concept, Chinese scholars have further characterized MHL as the knowledge, attitudes, and behavioral habits that an individual develops in promoting mental health for themselves and others, and in dealing with psychological disorders in themselves and others [2]. Recognized as a core component of overall health literacy, MHL directly affects individuals' mental well-being [3]. People with higher MHL identify mental disorders more readily, seek information about them more actively, and consult professionals more willingly [4]. Research on MHL in China nevertheless remains limited; most existing studies have focused on limited populations and lack comprehensive, large-scale assessments [5, 6].

The Healthy China Initiative (2019–2030) has explicitly designated the improvement of residents' MHL as the primary performance indicator of its "Mental Health Promotion" campaign, with national targets of 20% by 2022 and 30% by 2030 [7]. Even so, the overall level of MHL in China remains moderate to low [8]. A national survey conducted in 2021–2022 by the Chinese Academy of Sciences reported MHL attainment rates of only 6.4% among adolescents, 7.6% among older adults, and 15.8% among primary and secondary school teachers [9]. In Guangdong Province, attainment was 13.6% among urban residents and 8.6% among rural residents [6]. Urban-rural disparities and the concentration of megacity populations make local evidence on the factors associated with MHL attainment particularly relevant. Such evidence would provide an empirical basis for targeted mental health promotion policies. Fengtai District was selected as the study setting because it includes residents of both urban and rural hukou (household registration) status.

Although machine learning (ML) has shown considerable potential in mental health research [10, 11], its application in MHL research remains scarce, and the factors associated with MHL attainment have not been characterized using interpretable ML methods. For example, Mao et al. applied XGBoost and SHAP to predict depression in older adults and identified MHL as the most influential factor [12]. In that study, MHL was modeled as a predictor of depression rather than as the outcome of interest. To our knowledge, no study has examined the factors associated with MHL attainment in the Chinese general population using interpretable ML, although interpretable ML has been applied in this way to other health outcomes in China [13]. Traditional approaches, such as logistic regression, assume linear relationships between predictors and the outcome and may not capture nonlinear interactions among the multiple factors associated with MHL [14]. ML algorithms, by contrast, accommodate nonlinear relationships and higher-order interactions without *a priori* assumptions [11]. Interpretable ML methods such as SHapley Additive exPlanations (SHAP) quantify the contribution of each factor and indicate the direction of its association with the outcome [15].

The present study therefore aimed to characterize the factors associated with MHL attainment among residents of Fengtai District, Beijing, using interpretable ML applied to data from a population-representative cross-sectional survey. Six ML algorithms were compared and the primary model was

interpreted using SHAP, with the contribution and direction of each factor, rather than individual-level prediction or screening, as the analytical target.

Methods

Study participants

This was a secondary cross-sectional study based on the Fengtai District Mental Health Literacy Survey, conducted by the Fengtai District Mental Health Center between August and November 2022. A multi-stage random sampling design was used. In the first stage, probability proportional to size (PPS) sampling selected five subdistricts from Fengtai District. In the second stage, PPS selected five neighborhood committees within each selected subdistrict, yielding 25 neighborhood committees in total. In the third stage, systematic sampling drew 90 to 100 households per neighborhood committee, yielding 2471 sampled households. In the fourth stage, the Kish table was used to select one eligible resident (aged 18 years or older) from each household; eligibility required residence in Beijing for at least 6 months and Chinese nationality.

Among the 2471 selected households, 433 either refused to participate or could not be contacted, and 2038 residents returned the questionnaire (response rate 82.5%). We excluded 34 respondents who did not complete the questionnaire or whose responses were invalid. Finally, 2004 participants were included in the analysis (Fig. 1).

Ethics statement

All data were fully anonymized before analysis; direct and indirect identifiers (names, national identification numbers, contact information, and residential addresses) had been removed before access.

Sample size

The sample size was determined *a priori* for the original survey. Based on the 2021 Beijing MHL baseline of 13.8% ($p = 0.138$), with a relative allowable error of 15% ($\delta = 0.15$) and $\alpha = 0.05$ ($z = 1.96$), the required sample for a single proportion was $N = z^2(1 - p)/(\delta^2 p) \approx 1066$. A design effect of 1.5 for the multi-stage PPS design and a 20% allowance for invalid or non-response questionnaires inflated the target to 2000 valid respondents. The realized household frame of 2471 was expected to yield at least this number after non-response and invalid responses; the final analytic sample of 2004 provided 403 attainment events, corresponding to approximately 27 events per parameter, which exceeds the conventional minimum of 10.

Outcome variables

MHL attainment was the outcome of interest. MHL was assessed with the Beijing Resident Mental Health Literacy Monitoring Questionnaire, a standardized instrument developed and issued by the Beijing Municipal Health Commission and approved by the Beijing Municipal Bureau of Statistics. Knowledge was assessed with 22 true/false statements, each offering a "do not know" option, of which 20 were scored at 5 points each (range 0–100); items 16 and 21 were unscored check items. Skills were assessed with 16 items based on two clinical vignettes (range 0–40), and attitudes with 8 self-rated items scored 1–4 (range 8–32). Incorrect and "do not know" responses were scored as zero, following the national answer key. Respondents who met all three dimension-specific thresholds simultaneously (knowledge ≥ 80/100, skills ≥ 28/40, and attitudes ≥ 24/32) were classified as having attained MHL, and the attainment rate was the proportion of such respondents among the 2004 participants included in the analysis.

Predictor variables

Guided by the biopsychosocial model, 15 candidate predictors were selected and grouped into three domains.

(1) Sociodemographic characteristics: age (years); subdistrict of residence; gender; hukou type (urban vs. rural); hukou location (Beijing vs. other); educational attainment; monthly income; occupation; and marital status (having a spouse vs. not).

(2) Physical and mental health indicators: depressive symptoms (Patient Health Questionnaire-9, PHQ-9 ≥ 5, yes vs. no) [16]; anxiety symptoms (Generalized Anxiety Disorder-7, GAD-7 ≥ 5, yes vs. no) [17]; insomnia (Insomnia Severity Index, ISI ≥ 8, yes vs. no) [18]; self-rated health (healthy vs. ill or disabled); and life satisfaction (satisfied vs. dissatisfied).

(3) Mental health training: prior participation in mental health courses or training (yes vs. no).

The subdistrict variable comprised five levels; it was retained as a categorical predictor via one-hot encoding and interpreted collectively through SHAP rather than as separate coefficients, to limit dimensionality inflation.

Missing data

Of 2038 returned questionnaires, 34 (1.67%) were excluded for incomplete or invalid mental health literacy entries. No missing values remained on any analytic variable; complete-case analysis without imputation was therefore performed.

Statistical analysis

Data preprocessing and analysis were performed with Python version 3.13.9 (Python Software Foundation) with Jupyter Notebook and the scikit-learn, NumPy, SciPy, and Matplotlib packages.

Continuous variables were summarized as mean ± standard deviation (*SD*), and categorical variables were expressed as *N* (%).

For modeling, all 15 predictors were entered in full into each model. Continuous predictors were *Z*-score standardized before fitting, so that logistic regression (LR) odds ratios reflect change per 1 SD, and categorical variables were one-hot encoded. Internal validation was performed by bootstrap resampling with 1000 replicates. In each replicate, the model was refitted on the bootstrap sample and used to predict the outcome for the participants not included in that sample. Out-of-bag predictions were used in two ways: metrics were computed within each replicate and then summarized across replicates, whereas the calibration analysis of the primary model used predictions pooled across all replicates, as described below.

Six models were compared: logistic regression (LR), a transparent parametric reference model estimating the log-odds of attainment; random forest (RF), a bagging ensemble of decision trees robust to nonlinearity and outliers; *k*-nearest neighbors (KNN), an instance-based nonparametric classifier assigning the majority class among the k nearest samples; decision tree (DT), which recursively partitions the feature space; naive Bayes (NB), a generative classifier assuming conditional independence across features; and support vector machine (SVM), which maximizes the class margin via kernel functions under structural risk minimization. Given the 20.11% prevalence of MHL attainment, class weighting was applied in the LR, RF, DT, and SVM models, whereas no class weighting was applied in KNN, and NB used its default class priors. The RF model was designated *a priori* as the primary model for interpretation, given its capacity to accommodate nonlinearity and interactions and its built-in measures of feature importance.

Model performance was assessed at two levels. Discrimination was quantified by the area under the receiver operating characteristic curve (AUC) and the area under the precision-recall curve (PR-AUC); given the class imbalance, PR-AUC was taken as the primary discrimination metric for the minority (attained) class, thereby avoiding the overestimation of performance that AUC can produce under imbalance. Threshold-dependent metrics, namely recall (sensitivity), specificity, precision, the F1 score, and Youden's *J*, were computed at the optimal probability cut-off maximizing Youden's *J* on the bootstrap predictions of each model. Discrimination metrics (AUC and PR-AUC) and the Brier score were computed in each replicate for all six models and summarized across the 1000 replicates as the mean with 95% confidence intervals. Calibration was assessed for the primary model using the out-of-bag predictions pooled across replicates: the observed outcome was regressed on the predicted log-odds in a logistic model, yielding a calibration intercept (α) and slope (β), with 95% confidence intervals obtained by bootstrap resampling, and a calibration plot was produced [19]. An intercept of 0 and a slope of 1 indicate perfect calibration. Because the intercept deviated from 0, the primary model was updated by logistic recalibration, in which the estimated intercept and slope were applied to the predicted log-odds (new logit = $\alpha + \beta \times$ original logit) and the Brier score was recalculated. Recalibration modifies the calibration of the predicted probabilities but preserves their ranking, and therefore does not affect discrimination.

Two complementary methods were used to identify core predictors of MHL attainment in the primary model. First, the built-in feature importance of the primary model, based on the mean decrease in Gini impurity, ranked features by their average contribution across decision trees. Second, SHAP analysis of the primary model, grounded in cooperative game theory, computed each feature's marginal contribution to individual predictions and aggregated these into a global ranking [15]; it also indicates the direction of each feature's association with the outcome. SHAP outputs comprised a global feature importance plot (mean |SHAP|) and a beeswarm summary plot displaying the distribution of SHAP values for each feature.

Results

Participant characteristics

A total of 2004 respondents were included in the final analysis, of whom 403 (20.11%) attained MHL and 1601 (79.89%) did not. The characteristics of participants are summarized in Table 1.

Table 1
Characteristics of participants by mental health literacy attainment status (N = 2004).

Characteristic	Category	Attained (*N* = 403)	Not attained (*N* = 1601)
Age (years, mean ± *SD*)		50.04 ± 16.37	51.31 ± 15.14
Subdistrict			
	Chengshousi	110 (27.30%)	276 (17.24%)
	Fengtai	75 (18.61%)	329 (20.55%)
	Liuliqiao	95 (23.57%)	320 (19.99%)
	Wanping	84 (20.84%)	316 (19.74%)
	Xiluoyuan	39 (9.68%)	360 (22.48%)
Gender			
	Male	185 (45.91%)	700 (43.72%)
	Female	218 (54.09%)	901 (56.28%)
Hukou type			
	Urban	375 (93.05%)	1396 (87.20%)
	Rural	28 (6.95%)	205 (12.80%)
Hukou location			
	Beijing	331 (82.13%)	1333 (83.26%)
	Other	72 (17.87%)	268 (16.74%)
Educational attainment			
	Junior high school or below	39 (9.68%)	338 (21.11%)
	Senior high school/technical secondary	68 (16.87%)	411 (25.67%)
	College	105 (26.05%)	331 (20.68%)

Note. Data are presented as *N* (column %) unless otherwise indicated; for each variable, the percentages of its categories sum to 100% within each column.

Characteristic	Category	Attained (*N* = 403)	Not attained (*N* = 1601)
	Bachelor degree	152 (37.72%)	438 (27.36%)
	Master's degree or above	39 (9.68%)	83 (5.18%)
Monthly income (yuan)			
	≤ 3000	29 (7.20%)	159 (9.93%)
	3001–5000	110 (27.29%)	599 (37.41%)
	5001–8000	128 (31.76%)	461 (28.80%)
	> 8000	136 (33.75%)	382 (23.86%)
Occupation			
	Management/professional/technical	112 (27.79%)	359 (22.43%)
	Production/service	82 (20.35%)	383 (23.92%)
	Retired	88 (21.84%)	464 (28.98%)
	Other/special populations	121 (30.02%)	395 (24.67%)
Marital status			
	Married	321 (79.65%)	1341 (83.76%)
	Unmarried/divorced/widowed	82 (20.35%)	260 (16.24%)
Self-rated health status			
	Healthy	386 (95.78%)	1454 (90.82%)
	Ill or disabled	17 (4.22%)	147 (9.18%)
Life satisfaction			
	Satisfied	378 (93.80%)	1288 (80.45%)
	Dissatisfied	25 (6.20%)	313 (19.55%)

Note. Data are presented as *N* (column %) unless otherwise indicated; for each variable, the percentages of its categories sum to 100% within each column.

Characteristic	Category	Attained (*N* = 403)	Not attained (*N* = 1601)
Mental health course/training			
	Yes	173 (42.93%)	525 (32.79%)
	No	230 (57.07%)	1076 (67.21%)
Insomnia			
	No	364 (90.32%)	1283 (80.14%)
	Yes	39 (9.68%)	318 (19.86%)
Depressive symptoms			
	No	375 (93.05%)	1407 (87.88%)
	Yes	28 (6.95%)	194 (12.12%)
Anxiety symptoms			
	No	380 (94.29%)	1422 (88.82%)
	Yes	23 (5.71%)	179 (11.18%)

Note. Data are presented as *N* (column %) unless otherwise indicated; for each variable, the percentages of its categories sum to 100% within each column.

[Table 1 near here]

Model performance

The performance of the six models, estimated by internal validation with 1000 bootstrap replicates, is summarized in Table 2. Discrimination curves for the six models are shown in Fig. 2. With PR-AUC taken as the primary metric given the 20.11% prevalence of MHL attainment, RF ranked highest, followed by SVM, LR, NB, DT, and KNN (0.364, 0.342, 0.325, 0.315, 0.275, and 0.270, respectively). All values exceeded the no-skill baseline defined by the outcome prevalence (0.201), and the lower bound of the bootstrap 95% CI for RF (0.304) remained above this baseline. AUC ranged from 0.604 for KNN to 0.684 for RF, and the AUC of LR was identical to that of RF (0.684); however, the PR-AUC of LR (0.325) was lower than that of RF (0.364).

Table 2
Model performance of the six machine learning models

Model	AUC (95% CI)	PR-AUC (95% CI)	Recall	Specificity	Precision	F1	Youden's *J*	Brier (95% CI)
RF	0.684 (0.647–0.718)	0.364 (0.304–0.424)	0.408	0.812	0.354	0.377	0.219	0.181 (0.171–0.191)
SVM	0.679 (0.641–0.716)	0.342 (0.287–0.405)	0.591	0.662	0.306	0.402	0.253	0.155 (0.141–0.171)
LR	0.684 (0.646–0.719)	0.325 (0.272–0.383)	0.640	0.620	0.298	0.406	0.260	0.224 (0.211–0.238)
NB	0.672 (0.632–0.708)	0.315 (0.267–0.367)	0.571	0.685	0.315	0.405	0.256	0.255 (0.215–0.305)
DT	0.629 (0.582–0.673)	0.275 (0.232–0.329)	0.698	0.514	0.266	0.384	0.212	0.242 (0.226–0.259)
KNN	0.604 (0.561–0.647)	0.270 (0.223–0.315)	0.254	0.866	0.325	0.284	0.121	0.196 (0.178–0.215)

Note. Models are ordered by descending PR-AUC, the primary discrimination metric. Threshold-dependent metrics (recall, specificity, precision, F1, and Youden's *J*) are point estimates at the optimal probability cut-off maximizing Youden's *J* and are reported as the mean across 1000 bootstrap replicates. Abbreviations: AUC, area under the receiver operating characteristic curve; Brier, Brier score; CI, confidence interval; DT, decision tree; F1, F1 score; KNN, *k*-nearest neighbors; LR, logistic regression; NB, naive Bayes; PR-AUC, area under the precision–recall curve; RF, random forest; SVM, support vector machine.

At the probability cut-off maximizing Youden's *J*, recall ranged from 0.254 to 0.698, specificity from 0.514 to 0.866, precision from 0.266 to 0.354, F1 from 0.284 to 0.406, and Youden's J from 0.121 to 0.260. Specificity exceeded recall in five of the six models (the exception being LR, in which recall was higher than specificity), and precision remained at or below 0.354 throughout. Across the six models, Brier scores ranged from 0.155 (SVM) to 0.255 (NB) (Table 2).

Calibration and recalibration

Calibration of the RF model is shown in Fig. 3. The calibration intercept was − 0.92 (95% CI − 1.05 to − 0.79) and the calibration slope was 1.00 (95% CI 0.84 to 1.19) (Fig. 3a). The calibration curve lay below the diagonal across the observed range of predicted probabilities, indicating that predicted probabilities of MHL attainment were higher than the observed proportions, whereas the slope was consistent with 1. After recalibration, the calibration curve moved towards the diagonal and the Brier score decreased from 0.177 to 0.147 (Fig. 3b). Discrimination was unchanged, as recalibration preserves the ranking of

predicted probabilities. Because the intercept and slope were estimated on the same out-of-bag predictions used for recalibration, this reduction in the Brier score is descriptive; the recalibrated model was not assessed in an independent sample.

Factors associated with MHL

The mean decrease in Gini impurity for the 15 original variables in the RF model is presented in Table 3. Importance was unevenly distributed. Age contributed the largest share (mean Gini importance 0.230; 23.03% of the total), followed by subdistrict (0.150; 15.01%), educational attainment (0.114; 11.44%), monthly income (0.094; 9.44%) and occupation (0.092; 9.23%). These five variables together accounted for 68.15% of the total importance (Table 3; Fig. 4). Each of the remaining ten variables accounted for less than 7%, and depressive symptoms (0.015; 1.50%) and anxiety symptoms (0.015; 1.49%) ranked last.

Table 3
Gini importance of the 15 original variables in the random forest model

Variable	Mean Gini importance	SD	Relative importance (%)
Age	0.230	0.010	23.03
Subdistrict	0.150	0.014	15.01
Educational attainment	0.114	0.009	11.44
Monthly income	0.094	0.007	9.44
Occupation	0.092	0.006	9.23
Life satisfaction	0.067	0.013	6.67
Mental health course/training	0.043	0.005	4.29
Gender	0.042	0.003	4.16
Insomnia	0.039	0.008	3.94
Marital status	0.028	0.004	2.83
Hukou type	0.027	0.007	2.75
Hukou location	0.025	0.003	2.54
Self-rated health status	0.017	0.005	1.68
Depressive symptoms	0.015	0.003	1.50
Anxiety symptoms	0.015	0.004	1.49

Note. Variables are ordered by mean Gini importance, estimated from 1000 bootstrap resamples of the random forest model. Relative importance denotes each variable's share of the total mean Gini

importance; percentages may not sum to exactly 100 because of rounding. Hukou, household registration system in China; SD, standard deviation.

The global SHAP importance of the same 15 variables is presented in Table 4. Life satisfaction ranked highest (mean |SHAP| 0.0496; 15.00%), followed by subdistrict (0.0459; 13.86%), educational attainment (0.0393; 11.88%), age (0.0351; 10.60%) and monthly income (0.0324; 9.80%); these five variables accounted for 61.14% of the total mean |SHAP|. Insomnia ranked sixth (0.0264; 7.98%) and occupation seventh (0.0247; 7.47%) (Table 4; Fig. 5a). Hukou location ranked last (0.0049; 1.49%).

Table 4
Global SHAP importance of the 15 original variables

Variable	Mean |SHAP| value	Relative importance (%)
Life satisfaction	0.0496	15.00
Subdistrict	0.0459	13.86
Educational attainment	0.0393	11.88
Age	0.0351	10.60
Monthly income	0.0324	9.80
Insomnia	0.0264	7.98
Occupation	0.0247	7.47
Mental health course/training	0.0188	5.67
Hukou type	0.0165	4.98
Marital status	0.0100	3.01
Self-rated health status	0.0076	2.29
Gender	0.0075	2.27
Anxiety symptoms	0.0068	2.04
Depressive symptoms	0.0055	1.67
Hukou location	0.0049	1.49

Note. Variables are ordered by mean absolute SHAP value. SHAP values quantify each variable's contribution to the model output, and larger values indicate greater influence. Relative importance denotes each variable's share of the total mean absolute SHAP value; percentages may not sum to exactly 100 because of rounding. Hukou, household registration system in China; SHAP, SHapley Additive exPlanations.

The distribution and direction of SHAP values for each variable are shown in Fig. 5b, in which higher feature values are shown in red and lower values in blue. Life satisfaction showed the widest spread:

SHAP values for dissatisfied participants extended to approximately − 0.25, whereas values for satisfied participants clustered near zero and slightly positive. Comparable patterns were observed for educational attainment, monthly income, and prior participation in mental health courses or training, for which higher values corresponded to positive SHAP values. In contrast, older age and the presence of insomnia, depressive symptoms, or anxiety symptoms corresponded to negative SHAP values, whereas their counterparts with lower values tended towards positive SHAP values. SHAP values for the remaining variables were concentrated near zero.

The two rankings agreed in placing subdistrict, educational attainment and monthly income among the five most influential variables. They diverged for age, which ranked first by Gini importance but fourth by mean |SHAP|, and for life satisfaction, which ranked sixth and first, respectively. Rank shifts of three or more positions were also observed for insomnia (ninth by Gini importance, sixth by mean |SHAP|), gender (eighth and twelfth), hukou type (eleventh and ninth) and hukou location (twelfth and fifteenth).

Discussion

This study employed interpretable machine learning methods to identify factors associated with meeting the mental health literacy (MHL) criteria and to characterize their relative contributions and directions among adult residents in Fengtai District, Beijing. Of the 2004 residents, 403 (20.11%) simultaneously met the thresholds across all three dimensions of knowledge, skills, and attitudes. This level exceeded the 20% interim target for 2022 set by the Healthy China Action (2019–2030) but remained below the 30% target for 2030 [7]. Discrimination was limited across all six algorithms, with AUCs ranging from 0.604 to 0.684, and the highest PR-AUC, the primary discrimination metric, reaching 0.364 (95% CI, 0.304–0.424). The calibration intercept for the random forest model was − 0.92 (95% CI, − 1.05 to − 0.79), indicating that predicted probabilities were systematically overestimated; after logistic recalibration, the Brier score decreased from 0.177 to 0.147, while discrimination remained unchanged. Gini importance and SHAP consistently identified subdistrict-level residence, educational attainment, and monthly income as stable factors, whereas the importance rankings of age and life satisfaction were markedly method-dependent.

Interpretation of results and mechanisms

The results showed that subdistrict-level residence ranked second under both methods (Gini importance, 0.150, 15.01%; SHAP, 0.0459, 13.86%). Based on within-subgroup calculations, MHL attainment rates across subdistricts ranged from 9.77% (Xiluoyuan Subdistrict) to 28.50% (Chengshousi Subdistrict). This variation suggests that MHL attainment is associated not only with individual characteristics but also potentially with structural conditions at the community level [20]. Notably, subdistrict-level residence was transformed into five dummy variables after one-hot encoding, and its importance in the model may have been partially inflated by the increase in dimensionality [21]. Therefore, the high ranking of subdistrict-level residence in both Gini importance and SHAP indicates only that this variable is associated with MHL attainment overall; its importance cannot be directly equated with an independent community-level

effect. In addition, this study did not include community-level variables such as per capita income or mental health service accessibility [22], and thus could not identify specific community mechanisms. Previous research suggests that the frequency, quality, and accessibility of community mental health education activities may influence the efficiency with which mental health knowledge is disseminated among residents [23]. With respect to intervention, strengthening community mental health service capacity in subdistricts with lower attainment rates may be more efficient than merely expanding individual-level knowledge dissemination.

Educational attainment ranked third under both SHAP (0.0393, 11.88%) and Gini importance (0.114, 11.44%), making it one of the factors that consistently entered the top five across both methods. Stratified analyses showed that the MHL attainment rate was only 10.34% among those with junior high school education or below, compared with 31.97% among those with a master's degree or higher, indicating a clear educational gradient. This finding is consistent with previous research [24]. These results suggest that the association between education and MHL is both statistically and substantively meaningful. Education may promote MHL through multiple mechanisms. On the one hand, education enhances individuals' capacity to obtain, understand, evaluate, and apply health information [25], which are core dimensions of health literacy, making more educated individuals more likely to encounter and comprehend mental health knowledge. On the other hand, education may improve cognitive flexibility, helping individuals recognize their own psychological states and thereby engage in appropriate help-seeking behaviors [26]. Therefore, educational attainment is not merely a demographic characteristic but may also reflect comprehensive differences in information processing, cognitive resources, and health behaviors. This also suggests that MHL promotion strategies targeting less educated populations should avoid excessive use of professional terminology and instead adopt more accessible and everyday communication approaches.

Monthly income ranked fourth under Gini importance (0.094, 9.44%) and fifth under SHAP (0.0324, 9.80%). Based on within-subgroup calculations, MHL attainment rates were 15.43% among those with a monthly income of ≤ 3000 yuan, 15.51% among those with 3001–5000 yuan, 21.73% among those with 5001–8000 yuan, and 26.25% among those with > 8000 yuan. Occupation also entered the top seven under both rankings (Gini importance, 0.092, 9.23%; SHAP, 0.0247, 7.47%). Together, these two variables point to the dimension of socioeconomic status, suggesting that MHL attainment is associated with individuals' disposable resources and their occupational environments [27]. It should be interpreted with caution, however, that monthly income and educational attainment are often correlated at the individual level [28], and collinearity was not examined in this study. Moreover, SHAP reflects marginal contributions within a multivariable model and is not equivalent to independent effects after adjusting for other variables [15]. Therefore, the relative contributions of income and occupation should be regarded as hypotheses for further research rather than definitive effect estimates. From a policy perspective, these findings suggest that MHL promotion should attend to groups with relatively limited economic resources and, through community services, public education, and the development of service accessibility, partially offset the impact of unequal resource distribution [29].

Life satisfaction was the largest contributor in the SHAP ranking (0.0496, 15.00%) but ranked only sixth in Gini importance (0.067, 6.67%). The SHAP beeswarm plot showed a positive association with MHL attainment, with SHAP values for the dissatisfied group reaching as low as approximately − 0.25, whereas those for the satisfied group clustered near zero and in the positive range. Based on within-subgroup calculations, the MHL attainment rate was 22.69% among satisfied individuals and 7.40% among dissatisfied individuals. Previous research suggests that life satisfaction reflects an individual's overall cognitive evaluation of quality of life and is closely related to positive mental health and health behaviors [30, 31]. At the same time, improved MHL may enhance individuals' emotional intelligence, thereby improving subjective well-being [32]. Because this study used a cross-sectional design, however, the temporal relationship between the two cannot be determined, and directional interpretations should be made with caution [33].

Age showed a marked discrepancy in ranking between the two methods, ranking first under Gini importance (0.230, 23.03%) but fourth under SHAP (0.0351, 10.60%). This divergence is more likely attributable to methodological characteristics than to a substantive contradiction. Gini importance is based on impurity reduction and systematically overestimates continuous variables with more distinct values, which provide more potential split points; SHAP, grounded in cooperative game theory, characterizes the marginal contribution of a feature to individual predictions and is relatively less affected by the number of distinct values a variable takes [15]. Previous studies based predominantly on traditional statistical methods have generally reported higher MHL among younger populations, with the shared premise of treating age as a linear main effect. In this study, older residents corresponded to negative SHAP values, indicating an association with a lower probability of meeting the MHL criteria, but the strength of this contribution was weaker than that of life satisfaction, educational attainment, and monthly income. Age should therefore not be treated as an independent intervention target. In addition, interaction effects were not tested in this study, and whether age influences MHL through interactions with other variables remains to be examined.

Insomnia ranked sixth under SHAP (0.0264, 7.98%) and ninth under Gini importance (0.039, 3.94%), with a negative direction of association. Based on within-subgroup calculations, the MHL attainment rate was 10.92% among individuals with insomnia symptoms and 22.10% among those without. Previous research suggests that sleep problems may affect subjective well-being through cognitive mediation and are associated with reduced life satisfaction [34, 35]. Given that life satisfaction was the top-ranked variable under SHAP in this study, insomnia may be indirectly associated with MHL in part through its synergistic interaction with life satisfaction. Notably, depressive and anxiety symptoms fell within the lowest range under both rankings (Gini importance, 1.50% and 1.49%, ranking last and second to last, respectively; SHAP, 1.67% and 2.04%, ranking fourteenth and thirteenth, respectively), which differs from previous studies that identified depression and anxiety as major predictors of mental health [36]. A possible explanation is that the outcome in this study was MHL rather than psychological disorders per se, and insomnia is more prevalent in the general population and more readily perceived by individuals, and thus more directly associated with health knowledge acquisition and health behaviors [37]. In addition, the MHL attainment rate was 24.79% among those who had previously participated in mental

health courses or training and 17.61% among those who had not (Gini importance, 4.29%, ranking seventh; SHAP, 5.67%, ranking eighth), suggesting a positive association between training experience and MHL attainment. Mental health training can therefore be regarded as a psychological variable amenable to direct intervention. In future research and practical interventions, systematic mental health courses, thematic lectures, skills training, and public education activities could be used to improve public understanding of mental health knowledge and thereby promote MHL attainment. Moreover, given the importance of insomnia and life satisfaction in MHL, interventions could consider integrating sleep management, life satisfaction enhancement, and mental health training to form a multipath, comprehensive health promotion strategy [38], thereby more effectively improving MHL levels in the population.

Model discrimination and calibration

Discrimination was limited across all six models, with AUCs ranging from 0.604 to 0.684 and the highest PR-AUC reaching 0.364 (95% CI, 0.304–0.424). All models yielded PR-AUCs above the no-skill baseline defined by the outcome prevalence (0.201), and the lower bound of the 95% CI for RF (0.304) also exceeded this baseline, indicating that the information contained in the models outperformed random guessing. However, at the Youden-optimal cutoff, RF achieved a precision of only 0.354, a recall of 0.408, and an F1 of 0.377, with no other model exceeding this level, indicating a high error rate when classifying individuals as meeting the criteria based on this model. The limited discrimination may be attributable to the following reasons: the outcome was jointly defined by thresholds across three dimensions and dichotomized, resulting in a loss of score information; the models did not include community-level or lifestyle variables; and all measures relied on self-report, giving rise to common method bias. In addition, random forest achieved the same AUC as logistic regression (both 0.684), with its advantage manifested in the primary metric, PR-AUC (0.364 vs. 0.325); under class imbalance, PR-AUC better reflects the ability to identify the minority class (those meeting the criteria), and it was therefore adopted as the primary discrimination metric in this study.

The calibration intercept for RF was − 0.92 (95% CI, − 1.05 to − 0.79), significantly deviating from 0, indicating that predicted probabilities exceeded observed proportions; the calibration slope was 1.00 (95% CI, 0.84 to 1.19), consistent with 1, indicating that the ranking information of predicted probabilities was largely reasonable. After logistic recalibration, the calibration curve approached the diagonal, and the Brier score decreased from 0.177 to 0.147. Because the calibration intercept and slope were estimated from the same set of out-of-bag predictions and used directly for recalibration, this improvement should be regarded as a descriptive result that has not been validated in an independent sample. In summary, the primary model in this study is suitable for characterizing factors associated with MHL attainment and the direction of their contributions, but not for determining whether an individual meets the criteria or for screening, which is consistent with the predefined analytical objective of this study.

Conclusions

This study employed interpretable machine learning methods to examine factors associated with meeting the Mental Health Literacy (MHL) criteria among 2004 adult residents in Fengtai District, Beijing. Random forest demonstrated the best discriminative performance among the six models, although overall discrimination remained limited; logistic recalibration improved probability estimates without altering discrimination. Subdistrict-level residence, educational attainment, and monthly income consistently ranked high across both importance methods, whereas the rankings of age and life satisfaction were method-dependent. These findings suggest that MHL promotion may prioritize individuals with lower educational attainment, lower income, or lower life satisfaction at the individual level, strengthen mental health service capacity in subdistricts with lower MHL attainment rates at the community level, and target improvements in life satisfaction as an intermediate objective. The interpretable machine learning workflow developed in this study may serve as a methodological reference for MHL surveys in other regions. Future research should incorporate external validation and longitudinal designs, include community-level and lifestyle variables, account for sampling design effects and subgroup equity, and evaluate model calibration and discrimination in independent samples.

Abbreviations

AUC
area under the receiver operating characteristic curve
CI
confidence interval
DT
decision tree
GAD-7
Generalized Anxiety Disorder-7
ISI
Insomnia Severity Index
KNN
k-nearest neighbors
LR
logistic regression
ML
machine learning
MHL
mental health literacy
NB
naive Bayes
PHQ-9
Patient Health Questionnaire-9
PPS

probability proportional to size
PR-AUC
area under the precision-recall curve
RF
random forest
SD
standard deviation
SHAP
SHapley Additive exPlanations
SVM
support vector machine

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. The Ethics Committee of Beijing Fengtai District Mental Health Center (Beijing Fengtai District Mental Health Prevention and Treatment Hospital) determined that this secondary analysis of de-identified routine monitoring data was exempt from ethics review (Ethics Review No. 10 of 2026; date: April 6, 2026). In the original survey, informed consent to participate was obtained from all participants by the survey implementing agency at the time of data collection.

Clinical trial number

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors report no competing interests.

Funding

This work was supported by the 2025 Scientific Research Project of the Fengtai District Health Commission, Beijing, China (Grant No. 2025-073). The funder had no role in the study design, data

collection, data analysis, interpretation of the data, or preparation of the manuscript. The article processing charge was covered by the funder.

Author Contribution

J.L. and J.W. oversaw data handling and processing and prepared the initial draft. Q.T. and Z.L. substantially improved the academic quality through comprehensive review. Y.L. and S.Q. participated in quality-control investigations. Data gathering was carried out by the research team, including C.Z. and Z.T. The study concepts and implementation were collaboratively created by all authors. All authors read and approved the final manuscript.

Acknowledgements

None.

Availability of data and materials

The data underlying this study are anonymized monitoring data held by the Fengtai District Mental Health Center and are not publicly available owing to the conditions of the data-use agreement; they may be made available from the corresponding author on reasonable request and with the permission of the data custodian.

References

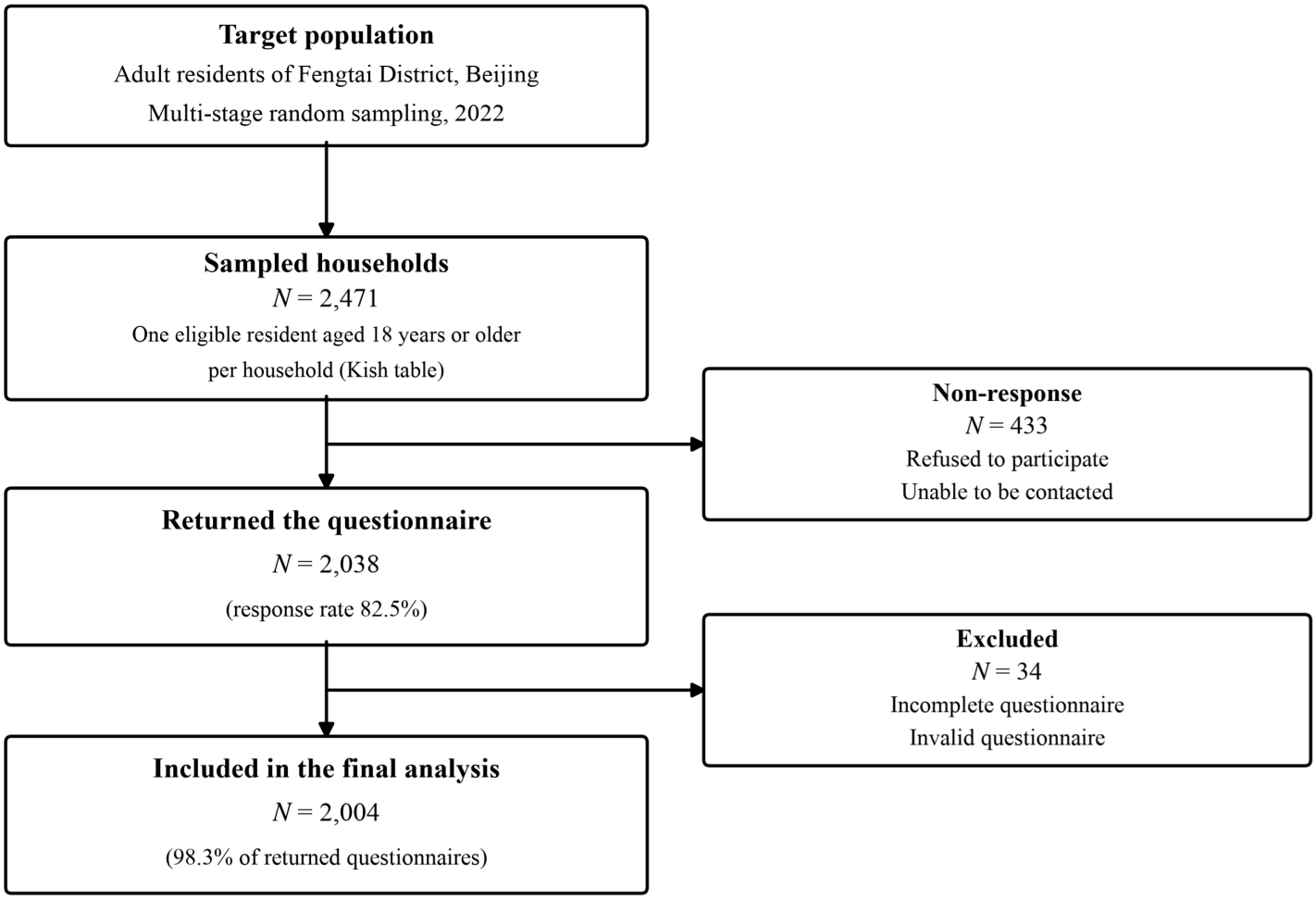
1. Jorm AF, Korten AE, Jacomb PA, Christensen H, Rodgers B, Pollitt P. Mental health literacy: a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. Med J Aust. 1997;166(4):182–6. 10.5694/j.1326-5377.1997.tb140071.x.
2. Jiang G, Zhao C, Wei H, Yu L, Li D, Lin X, Ren Z. Mental health literacy: Connotation, measurement and new framework. J Psychol Sci. 2020;43(1):232–8. 10.16719/j.cnki.1671-6981.20200132. [in Chinese].
3. Bahrami MA, Bahrami D, Chaman-Ara K. The correlations of mental health literacy with psychological aspects of general health among Iranian female students. Int J Ment Health Syst. 2019;13:59. 10.1186/s13033-019-0315-6.
4. Tay JL, Tay YF, Klainin-Yobas P. Mental health literacy levels. Arch Psychiatr Nurs. 2018;32(5):757–63. 10.1016/j.apnu.2018.04.007.
5. Xu R, Jing M, Zhang A, Zha L, Wang Y, Wang A, Tang J, Xia B, Jia S. Mental health literacy among older adults in Shanghai: a descriptive qualitative study. Front Psychol. 2024;15:1470758. 10.3389/fpsyg.2024.1470758.

6. He XY, Tan WY, Guo LL, Ji YY, Jia FJ, Wang SB. Mental health literacy among urban and rural residents of Guangdong Province, China. Risk Manag Healthc Policy. 2024;17:2305–18. 10.2147/rmhp.S479868.
7. The State Council of the People's Republic of China. Opinion of the State Council on implementing the Healthy China Initiative (No. 13 [2019] of the State Council). 2019. https://www.gov.cn/zhengce/content/2019-07/15/content_5409492.htm
8. Zhang G, Lin Y, Zhao B, Gao W. A cross-sectional study on the level of mental health literacy and its influencing factors among urban and rural residents in Weifang City. Chin J Health Educ. 2025;41(1):89–94. 10.16168/j.cnki.issn.1002-9982.2025.01.015. [in Chinese].
9. Ming Z, Chen Z, Guo F, Hou J, Liu S, Wang Y. In: Fu X, Zhang K, Chen X, Chen Z, editors. Investigation report on mental health literacy in China (2021–2022). Beijing: Social Sciences Academic; 2023. [in Chinese].
10. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol. 2018;14(1):91–118. 10.1146/annurev-clinpsy-032816-045037.
11. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med. 2019;49(9):1426–48. 10.1017/S0033291719000151.
12. Mao N, Wang Y, Zhou A, Dai X, Han ZR, Li H, Zhang Z. Mental health literacy predicts depression in older adults in China: an interpretable machine learning model. J Gerontol B Psychol Sci Soc Sci. 2026;81(1):gbaf158. 10.1093/geronb/gbaf158.
13. Wang Y, Luo Z, Wen X, Yuan L, Wu Y, Kamili M, Xiang Y. Identification of risk factors for latent tuberculosis infection in Xinjiang using machine learning. BMC Public Health. 2025;26:151. 10.1186/s12889-025-25844-w.
14. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods. 2018;15(4):233–4.
15. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4765–74.
16. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13.
17. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10):1092–7.
18. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep. 2011;34(5):601–8. 10.1093/sleep/34.5.601.
19. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17:230. 10.1186/s12916-019-1466-7.
20. Cheng J, Lu C, Liu C, Chen Q, Liu B, Guo Y. Current status of mental health literacy and its influencing factors among residents in Bao'an District, Shenzhen. Chin J Health Educ. 2023;39(7):658–62. [in Chinese].

21. Nembrini S, Konig IR, Wright MN. The revival of the Gini importance? Bioinformatics. 2018;34(21):3711–8. 10.1093/bioinformatics/bty373.
22. Kang P, Li Q, Zhang J, Wang Y, Hu L. Current status of mental health literacy and its influencing factors among residents in Shandong Province in 2023. Chin J Public Health. 2025;41(7):817–22. [in Chinese].
23. Wang Y, Jia S, Wang A, Li H, Xu R, He X. Multilevel factors affecting mental health literacy among older adults: a qualitative study based on social ecological model. Front Public Health. 2025;13:1656116. 10.3389/fpubh.2025.1656116.
24. Litta A, Amelio E, Papari C, De Benedittis M, Pastore F. Mental health literacy interventions in young people: a narrative review. Psychiatria Danubina. 2024;36(Suppl 2):385–8.
25. Chen X, Chen Y, Su J, Mei F. Recognition ability of common mental disorders and its influencing factors among community residents in Guangzhou. China Med Pharm. 2022;12(18):9–12. [in Chinese].
26. de Simões R, Trigueiro MJ, Portugal P, de Sousa S, Simões-Silva V, Campos F, Silva M, Marques A. Mental health literacy and stigma in a municipality in the north of Portugal: a cross-sectional study. Int J Environ Res Public Health. 2023;20(4):3318. 10.3390/ijerph20043318.
27. Umeda M, Kawakami N, Miller E. Erratum to: Effect of socioeconomic conditions on health care utilization in marital violence: a cross-sectional investigation from the Japanese Study on stratification, health, income, and neighborhood. Int J Equity Health. 2017;16(1):41. 10.1186/s12939-017-0606-y.
28. Fan X, Zheng H, Wang S, Tan W, Liao J. Mental health literacy, anxiety, and insomnia in Chinese chronically ill older adult-caregiver dyads: actor-partner interdependence moderation model. Fam Process. 2025;64(4):e70077. 10.1111/famp.70077.
29. Li X, Liang X, Cao B, Wang H, Liu L, Zhang W, Sun K. Current status of mental health literacy and its influencing factors among patients with type 2 diabetes mellitus in elderly care institutions. Chin J Nurs Educ. 2022;19(10):923–6. [in Chinese].
30. Hidalgo-Fuentes S, Martínez-Álvarez I, Sospedra-Baeza MJ, Martí-Vilar M, Merino-Soto C, Toledano-Toledano F. Emotional intelligence and perceived social support: its relationship with subjective well-being. Healthcare. 2024;12(6):634. 10.3390/healthcare12060634.
31. Kushlev K, Drummond DM, Diener E. Subjective well-being and health behaviors in 2.5 million Americans. Appl Psychol Health Well Being. 2020;12(1):166–87. 10.1111/aphw.12178.
32. Chen Y, Peng Y, Fang P. Emotional intelligence mediates the relationship between age and subjective well-being. Int J Aging Hum Dev. 2016;83(2):91–107. 10.1177/0091415016648705.
33. Grujičić S, Nikolić A. Cross-section studies: advantages and disadvantages. Zdravstvena zaštita. 2021;50(4):43–54.
34. Liu S, Shi J, Gao C, Zhou B, Li J, Tan X, Liu Y, Qi J, Yu P. Analysis of multidimensional health influencing factors among older adults in a community in Beijing. Chin J Geriatr. 2024;43(12):1623–9. 10.3760/cma.j.issn.0254-9026.2024.12.018. [in Chinese].

35. Su W, Fu C. A scoping review of the application of electronic health interventions in the field of mental health among older adults with chronic diseases. Chin J Geriatric Care. 2025;23(6). [in Chinese].
36. Duffy A, Keown-Stoneman C, Goodday S, Horrocks J, Lowe M, King N, Pickett W, McNevin SH, Cunningham S, Rivera D, Bisdounis L, Bowie CR, Harkness K, Saunders KEA. Predictors of mental health and academic outcomes in first-year university students: Identifying prevention and early-intervention targets. BJPsych Open. 2020;6(3):e46. 10.1192/bjo.2020.24.
37. Yang Y, Lin Z, Yu T, Chen X, Yang N, Zhong S, Zhou L. Relationship between mental health literacy and depressive, anxiety, and insomnia symptoms among young adults. Chin Mental Health J. 2025;39(4):344–9. [in Chinese].
38. Schwartz AR, Cohen-Zion M, Pham LV, Gal A, Sowho M, Sgambati FP, Klopfer T, Guzman MA, Hawks EM, Etzioni T, Glasner L, Druckman E, Pillar G. Brief digital sleep questionnaire powered by machine learning prediction models identifies common sleep disorders. Sleep Med. 2020;71:66–76. 10.1016/j.sleep.2020.03.005.

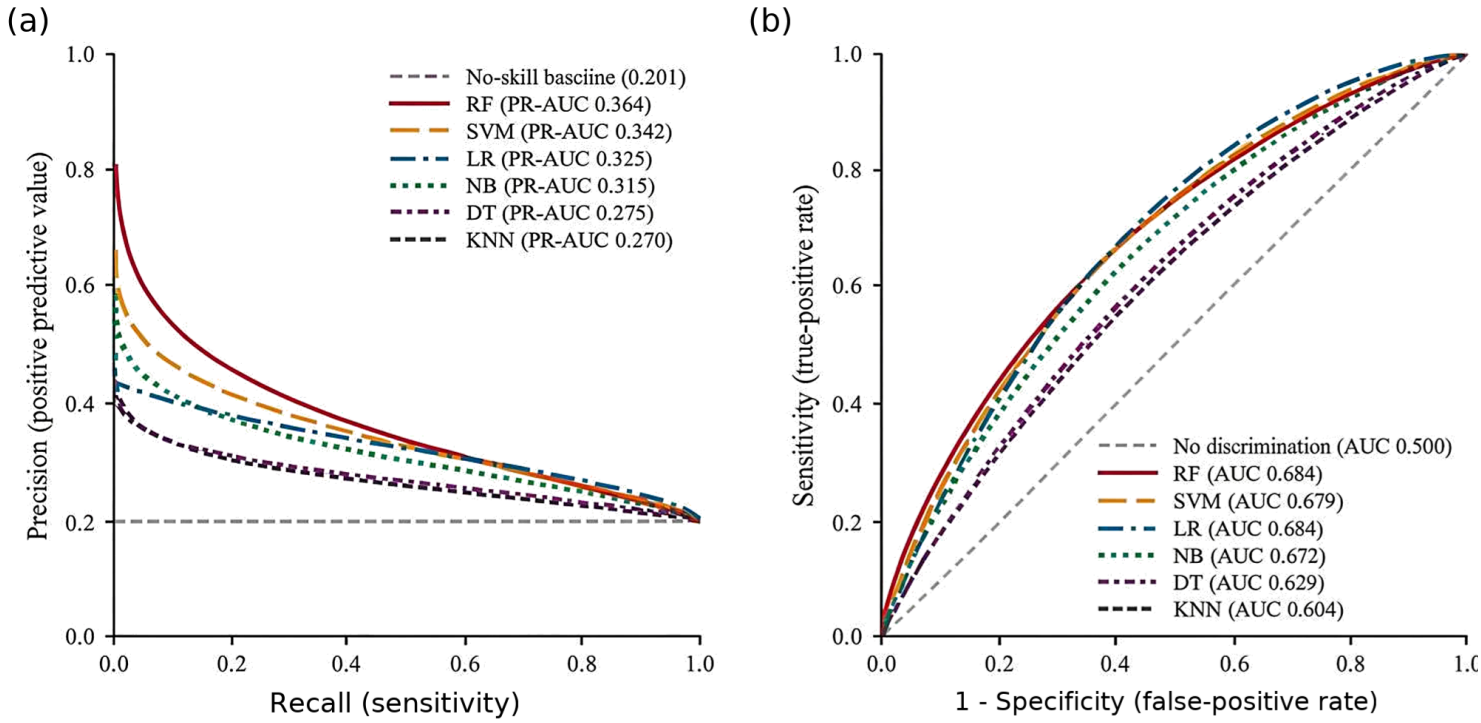
Figures



MHL, mental health literacy.

Figure 1

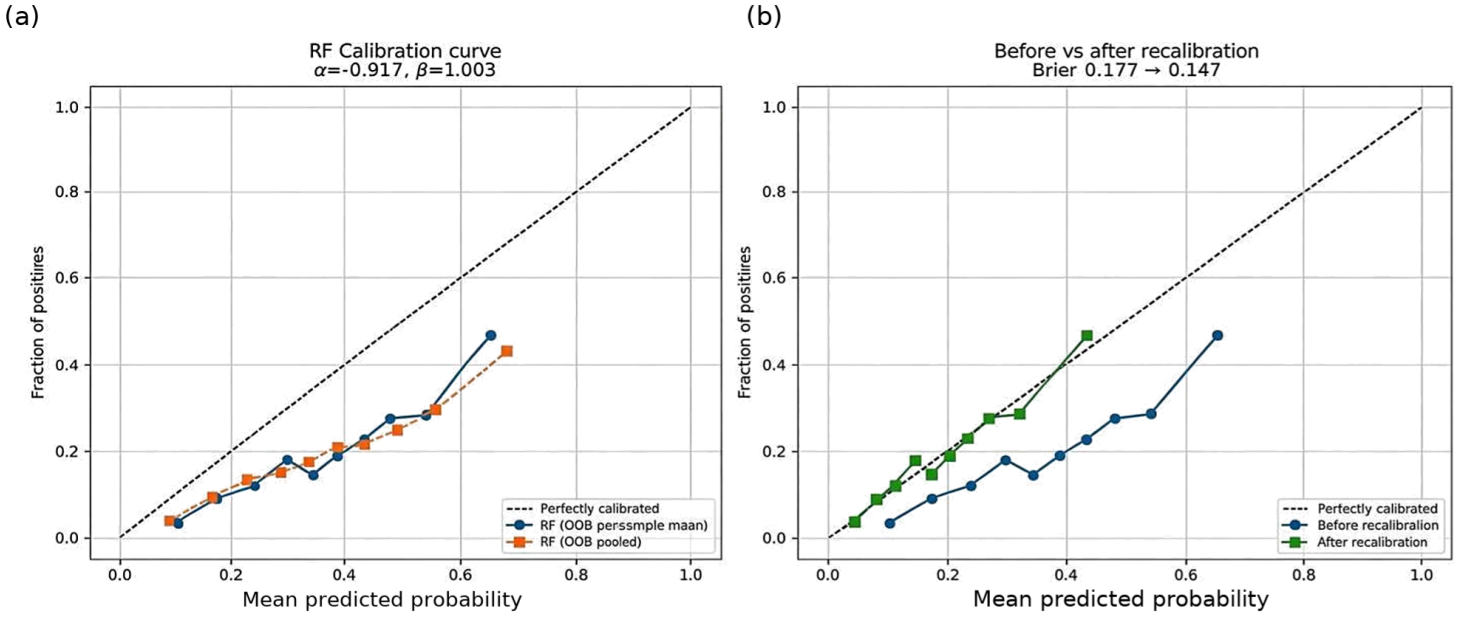
Flow of participants through the 2022 Fengtai District Mental Health Literacy Survey. A four-stage sampling design selected five subdistricts, five neighborhood committees within each subdistrict, and 90–100 households within each committee, yielding 2471 sampled households. One eligible resident aged 18 years or older was selected from each household using the Kish table; eligibility required residence in Beijing for at least 6 months and Chinese nationality. Of the 2471 sampled households, 433 could not be contacted or refused participation, and 2038 residents returned the questionnaire (response rate 82.5%). Thirty-four respondents were excluded because of incomplete or invalid questionnaire entries, and 2004 participants were included in the analysis.



AUC, area under the receiver operating characteristic curve; DT, decision tree; KNN, k-nearest neighbors; LR, logistic regression; NB, naive Bayes; PR-AUC, area under the precision–recall curve; RF, random forest; SVM, support vector machine.

Figure 2

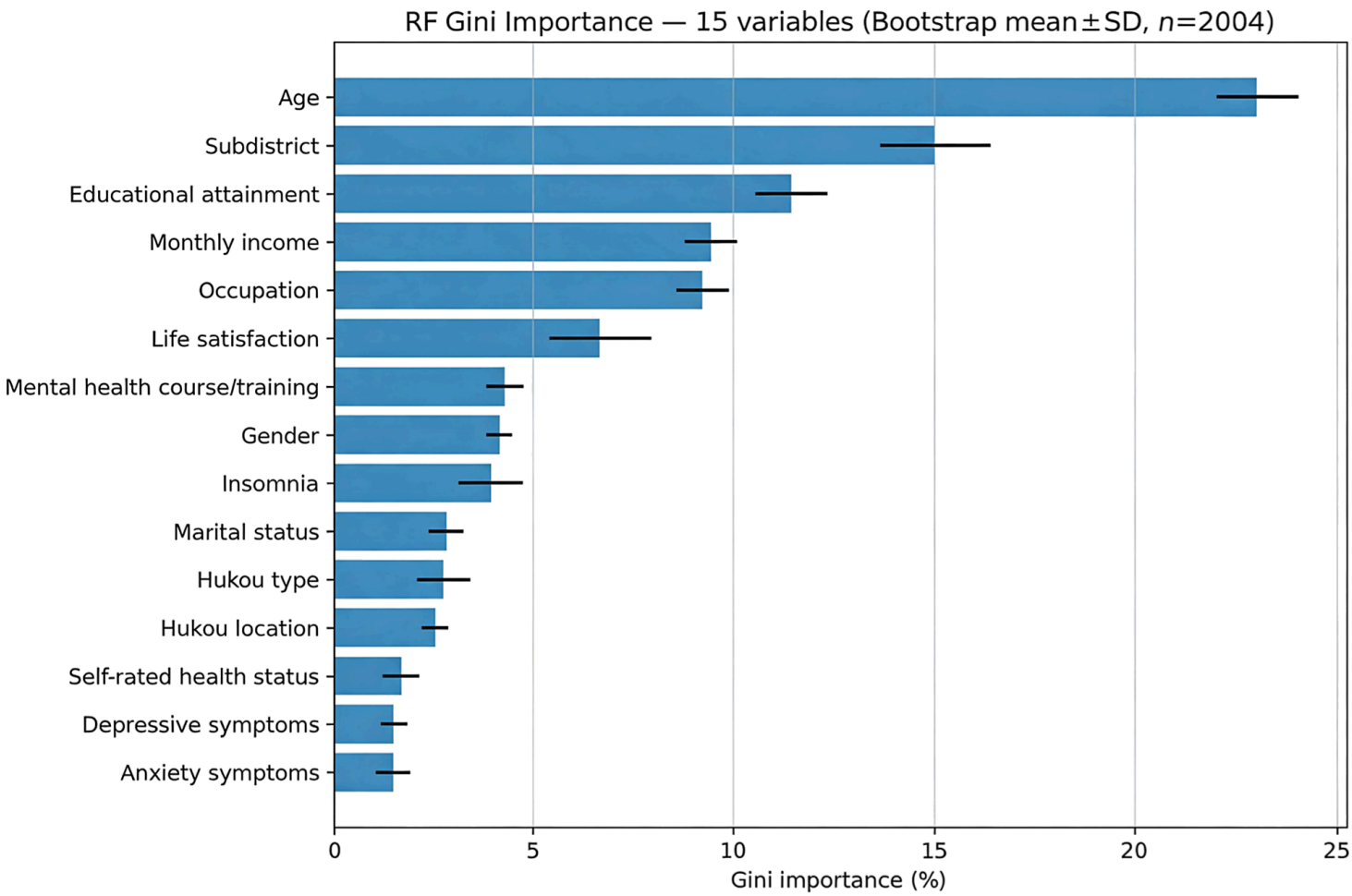
Discrimination of the six machine learning models for mental health literacy attainment. (a) Precision-recall curves for the six models. The horizontal dashed line indicates the no-skill baseline, equal to the prevalence of mental health literacy attainment (0.201). (b) Receiver operating characteristic curves for the same six models; the diagonal dotted line indicates no discrimination (AUC = 0.500). Curves were computed from the bootstrap predictions (1000 replicates). Because the outcome was imbalanced (20.11%), the area under the precision-recall curve was taken as the primary measure of discrimination. Random forest and logistic regression had the same AUC (0.684) but different areas under the precision-recall curve (0.364 and 0.325, respectively), and their curves crossed each other in panel (b).



CI, confidence interval; MHL, mental health literacy.

Figure 3

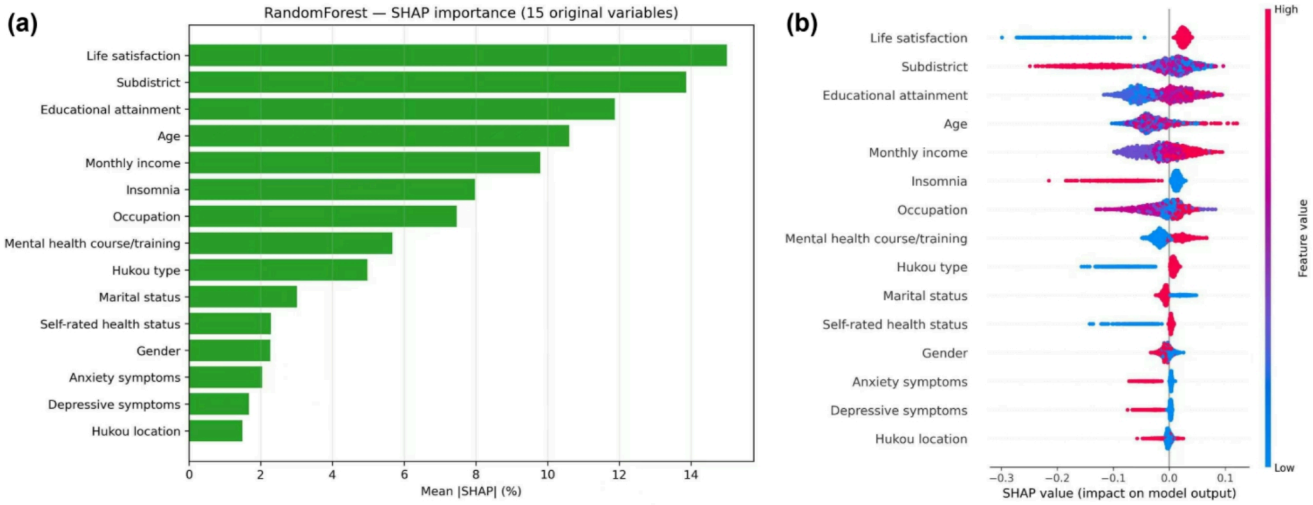
Calibration of the random forest model. (a) Calibration curve before recalibration, with the observed proportion of participants attaining MHL plotted against the predicted probability; the diagonal line indicates perfect calibration. The calibration intercept was −0.92 (95% CI −1.05 to −0.79) and the calibration slope was 1.00 (95% CI 0.84–1.19). (b) Calibration curves before and after logistic recalibration; after recalibration the curve moved towards the diagonal and the Brier score decreased from 0.177 to 0.147. Recalibration preserves the ranking of predicted probabilities and therefore does not affect discrimination. The intercept and slope were estimated on the same pooled out-of-bag predictions used for recalibration, so the reduction in the Brier score is descriptive.



hukou, household registration.

Figure 4

Gini impurity importance of the 15 variables in the random forest model. Bars show the mean Gini importance across 1000 bootstrap resamples. Age contributed the largest share (0.230; 23.03% of the total), followed by subdistrict (0.150; 15.01%), educational attainment (0.114; 11.44%), monthly income (0.094; 9.44%) and occupation (0.092; 9.23%); these five variables together accounted for 68.15% of the total importance.



hukou, household registration; MHL, mental health literacy.

Figure 5

SHAP importance and direction of association in the random forest model. (a) Global importance of the 15 original variables, quantified as the mean absolute SHAP value; life satisfaction ranked highest (0.0496; 15.00% of the total), followed by subdistrict (0.0459; 13.86%), educational attainment (0.0393; 11.88%), age (0.0351; 10.60%) and monthly income (0.0324; 9.80%). (b) SHAP summary (beeswarm) plot for the same model; each point represents one participant, colours indicate higher (red) or lower (blue) values of the corresponding variable, and positive SHAP values indicate a positive contribution to MHL attainment. Higher life satisfaction, educational attainment, monthly income, and prior participation in mental health courses or training corresponded to positive SHAP values, whereas older age and the presence of insomnia, depressive symptoms, or anxiety symptoms corresponded to negative SHAP values.