Supplemental Materials

Supplemental Method 1. The group-based trajectory model

We applied the group-based trajectory model (GBTM) to identify the trajectories of depressive symptoms and the number of chronic diseases. GBTM is a form of latent class growth modeling using maximum likelihood to identify groups of individuals following a similar developmental trajectory on an outcome of interest by fitting a semiparametric mixture model to longitudinal data (*1*). Due to the characteristics of depressive symptoms, they tend to cluster at the minimum or maximum value, leading to a skewed distribution. We used the censored normal (CNORM) distribution for trajectory model fitting. Following the recommendations of Nagin, a two-stage model selection process was employed to search for the best-fit model (*2*). The first stage focused on determining the number of groups to include in the model. Then the second stage shifted to determining the optimal order of polynomials specifying the shape of each trajectory. The time metric was the years since baseline (2011-2018), and four waves of CES-D 10 score were used to estimate the trajectories. The trajectories of depressive symptoms (CES-D 10 score) and number of chronic diseases were estimated. Model fitting was performed iteratively by comparing models with one to three groups and trajectory shapes (intercept, linear, quadratic, or cubic). Finally, trajectory modeling of depressive symptoms was conducted based on the best-fitting result of depressive symptoms. The posterior probabilities for each individual being a member of the trajectories were calculated in the final model.

Supplemental Method 2. Detailed description of covariate assessment

Sociodemographic variables

The sociodemographic variables included age (reported as mean and standard deviation), gender (Male or Female), education level (Primary school or below, Middle school or above) (*3*), marital status (Married, Other), and residence (Rural or Urban) (*4*).

Health status

Activities of daily life (ADL) and instrumental activity of daily life (IADL) refer to daily self-care tasks. Participants were asked all ADL/IADL questions and were defined as having a specific ADL/IADL disability if they answered as follows: *i*) I have difficulty but can still do it, *ii*) I have difficulty and need help, or *iii*) I cannot do it. The ADL index includes six aspects in CHARLS: bathing, eating, getting in and out of bed, dressing, using the toilet, and controlling urination and defecation (*5*). The IADL index has six aspects in CHARLS: doing housework, cooking, taking medicine, shopping, making phone calls (not in CHARLS 2011), and managing money (*5*), A score of 0 indicates that the participant did not report any problems with the activity. A score of 1 indicates that the participant reported some difficulty with the activity or could not do the activity. The ADL limitation scale ranged from 0-6 and IADL limitation scale ranged from 0-5, a higher score indicated a more severe disability (*6*-*8*).

Disability was assessed by asking participants, "Do you have one of the following disabilities"? The options included physical disabilities, brain damage or intellectual disability, vision problems, hearing problems, or speech impediments. Based on participants' responses, disability status was categorized as either "Yes" or "No".

Cognitive function was evaluated using a localized version of the Mini-Mental State Examination (MMSE) developed for CHARLS. The total cognitive function score ranged from 0 to 31, with higher scores indicating better cognitive function.

Health behaviors

The Frequency of leisure activity was measured by asking participants, "How often in the last

month [did/have] [you] [do voluntary or charity work/cared for a sick or disabled adult/provided help to family, friends or neighbors/attended an educational or training course/interacted with friends/go to a sport, social or other kind of club/taken part in a community-related organization]? Regularly (almost daily, almost every week), or not regularly"? On this basis, we define leisure activity as an ordered categorical variable, where "yes" represnts the presence of leisure activity and "no" represents the absence of leisure activity.

Smoking behavior was categorized into smokers and never smokers, to assess its impact on health (*9*). Additionally, drinking status was included, categorized as drinkers or never drinkers.

Smoking status was assessed with two questions: First, participants were asked, "Have you ever smoked? (including cigarettes, pipe tobacco, or chewed tobacco)". Those who denied any smoking history were classified as never smokers. Participants who confirmed a smoking history were further asked, "Are you currently smoking or have you quit"? Based on their responses, they were classified as smokers (*10*).

Supplemental Method 3. MissForest algorithm

The missForest algorithm offers an approach to assess imputation quality by iteratively using the Random Forest algorithm without setting aside test data or performing cross-validations (*11*).

The missForest algorithm can be summarized as follows (*12,13*). First, the missing variable is initialized by assuming the missing values as the mean (for continuous variables) or the most frequent class (for categorical variables). Secondly, the imputation process is done sequentially for the variables in the appointed order of missing observations for each variable. After initialization, the variable under imputation is used as the response, with other variables serving as predictors for building the Random Forest model. Last, when all variables with missing data are imputed, one imputation iteration is completed. The imputation process is iterated several times (limited to a maximum of 10 times to control computational time) until the relative sum of squared differences or proportion of falsely classified instances between the current and the previous imputation results — the last imputation, is outputted.

Compared with other imputation methods, such as k-nearest neighbors imputation or multivariate imputation using chained equations, missForest has demonstrated superior accuracy and stability (*14-16*). The missForest algorithm was analyzed using the "missforest" package in R software.

Supplemental Method 4. Feature selection

The Least Absolute Shrinkage and Selection Operator (LASSO) is a penalized regression approach that estimates regression coefficients by maximizing the log-likelihood function and automatically deletes unnecessary covariates, retaining only the most significant variables in the final model (*17*). The optimal shrinkage parameter λ was calculated by 10-fold cross-validation. Selecting relevant predictors with an appropriate level of explanation is critical to the model's success. Due to the method performing variable selection and regularization to enhance the detection accuracy and interpretability of the model, LASSO is applicable for high-dimensional data reduction and feature selection. LASSO penalized parameter estimates were generated using L1 penalization under a preset regularization parameter (λ). Less significant coefficients are then shrunk to zero, forcing the sum of the absolute value of the regression coefficients to be less than λ (*18*). The models with minimum λmin and one SD λ1se were compared regarding discriminability and calibration. λmin corresponds to the minimum mean cross-validation error, and λ1se is within one standard error of the cross-validated errors for λmin (*19,20*).

Recursive Feature Elimination (RFE) is an algorithm that utilizes cross-validation to identify features associated with categorical variables. Combined with Random Forest (RF), Naïve Bayes (NB) and Logistic Regression (LR), RFE was employed for feature selection by presetting 4 as the minimum number of variables and a 10-fold cross-validation for optimizing (*21*). The final feature selection was based on the number of features selected, accuracy and kappa. When the full RFE model with k-fold cross-validation is created, a variable importance measure is computed that ranks the predictors from most important to least, and the least essential predictors are iteratively eliminated before rebuilding the model (*22*).

The "glmnet" package was used to perform LASSO regression. In contrast, the "caret" package was empolyed for the RFE algorithm.

Supplemental Method 5. Machine learning algorithms

Logistic Regression (LR) is a member of the general linear model family used to identifying the influence of logistic regression of traditional statistics, the LR of machine learning algorithm helps decrease the prediction error by employing multiple layers of the validation process and increasing the research model's generalizability. LR is usually selected as a base model for comparison. The "glm" package was used to run LR in R statistical software. The input features on categorical targets in both traditional statistics and machine learning. LR estimates the odds ratio as a positive or negative class for each observation with the sigmoid function (*23*).

Decision Tree (DT) is a simple but powerful prediction method, which can be applied to multiple predictor variables, and is highly popular in medical decision-making (*24*). DT resembles a flow chart that guides a reader toward classifying a person as a higher or lower risk for an outcome. Moreover, DT can consider nonlinear relations ships among multiple variables, defining subgroups in a data-driven way (*25*). The "rpart" package was applied in R statistical software for DT.

Support Vector Machine (SVM) is a supervised learning algorithm that finds the optimal decision hyperplane for classifying data based on solid statistical theory. It maps data to a higher dimension using kernel functions, such as RBF, to achieve linear separation and overcome the problem of linear inseparability (*26*). Furthermore, SVM considers the minimization of empirical and structural risks and uses the hinge loss function as the agent loss, leading to good stability (*27*). SVM is especially suitable for small sample data prediction as it only depends on support vectors. The "kernlab" package was applied to run SVM in R statistical software.

K-nearest neighbor (KNN) is a commonly used multivariate classification algorithm that omits the prediction of the class density function. Instead, KNN categorizes an unknown sample by choosing the most similar model in training set based on Euclidean distance, resolving the training set size, and addressing multicollinearity issues (*28*). The "kknn" package was used in R statistical software to run KNN.

Random forest (RF) is a reliable and effective ensemble method based on bagging to enhance its prediction proficiencies as an extension of "classification and regression trees" (*29*).

RF consists of several decision trees integrated by voting to enhance precision and generalization performance. Each decision tree is based on random samples and features to avoid overfitting and enable automatic feature selection (*30,31*). The "ranger" package was applied using R statistical software.

Extreme Gradient Boosting (XGBoost) is an enhancing ensemble algorithm that uses additive learning to create a robust classifier of decision-tree-based models by integrating multiple weak learners based on boosting (*32*). The sum of the weighted contributions of all decision trees is used to achieve an overall improvement in performance. Moreover, XGBoost can prevent overfitting with complexity regularization (*33*). The "xgboost" package was used to apply the XGBoost model in R statistical software.

LightGBM (Light Gradient Boosting Machine) is a gradient boosting framework optimized for speed and efficiency, particularly in large-scale data tasks. LightGBM leverages decision tree algorithms and is designed to handle sparse and categorical features effectively. Its innovative leaf-wise growth strategy enhances training efficiency compared to traditional level-wise approaches, reducing computation time while maintaining high accuracy (*34*). LightGBM also incorporates regularization techniques, such as L1/L2 penalties, to mitigate overfitting. Additionally, it supports advanced features like GPU acceleration and categorical feature handling, making it ideal for high-dimensional data. The lightgbm package was used in R statistical software to implement LightGBM models.

A Multilayer Perceptron (MLP) is a type of feedforward artificial neural network capable of learning complex, non-linear relationships in data. MLP consists of one input layer, one or more hidden layers, and one output layer, with neurons in adjacent layers fully connected. Each neuron applies an activation function (*e.g.*, ReLU or sigmoid) to introduce nonlinearity, enabling the model to capture intricate patterns (*35*). The backpropagation algorithm is employed to minimize the error between predictions and ground truth by updating weights iteratively. MLP is particularly suitable for classification and regression problems involving large and high-dimensional datasets. In R, the keras or nnet package can be used to train MLP models.

Elastic Net is a regularization regression technique that combines L1 (lasso) and L2 (ridge) penalties to improve model performance and feature selection in high-dimensional data. By introducing an adjustable mixing parameter (α), Elastic Net balances the sparsity of lasso with the stability of ridge regression (*36*). This dual regularization addresses multicollinearity issues and ensures robustness against overfitting. Elastic Net is particularly effective in scenarios with highly correlated predictors, as it can select grouped variables rather than excluding one in favor of another. The glmnet package in R is commonly used to implement Elastic Net models, allowing cross-validation to determine optimal α and λ parameters.

Supplemental Method 6. Evaluation and visualization of machine learning models

To comprehensively evaluate model performance, nine metrics were employed: Area Under the Receiver Operating Characteristic Curve (AUC), sensitivity, recall, accuracy, specificity, Matthews correlation coefficient (MCC), J index, precision, and F1 score. These metrics collectively measure the model's discrimination, calibration, and predictive capabilities across various dimensions.

AUC: AUC reflects the overall classification performance, ranging from 0 to 1, with higher values indicating superior model discrimination. It provides a threshold-independent measure of how well the model distinguishes between classes.

Sensitivity (True Positive Rate): Sensitivity evaluates the proportion of actual positive cases correctly identified by the model, highlighting its ability to detect true positives effectively, especially critical in scenarios where missing positive cases has significant consequences.

Recall (True Positive Rate): Recall, synonymous with sensitivity, measures the fraction of relevant instances successfully retrieved. It ensures that the model minimizes false negatives in its predictions.

Accuracy: Accuracy quantifies the proportion of correct predictions across all classes, offering a general view of the model's predictive performance. However, it may be less informative in imbalanced datasets.

Specificity (True Negative Rate): Specificity assesses the model's ability to correctly identify negative cases, a critical counterpart to sensitivity in evaluating performance for binary classification.

MCC (Matthews Correlation Coefficient): MCC accounts for all elements of the confusion matrix, providing a balanced measure of prediction quality. It ranges from -1 (complete disagreement) to 1 (perfect prediction), with 0 indicating random guessing. MCC is particularly useful for imbalanced datasets.

J Index (Youden’s Index): The J index combines sensitivity and specificity into a single metric, calculated as (sensitivity + specificity - 1). It measures the effectiveness of a diagnostic test, with higher values indicating better performance.

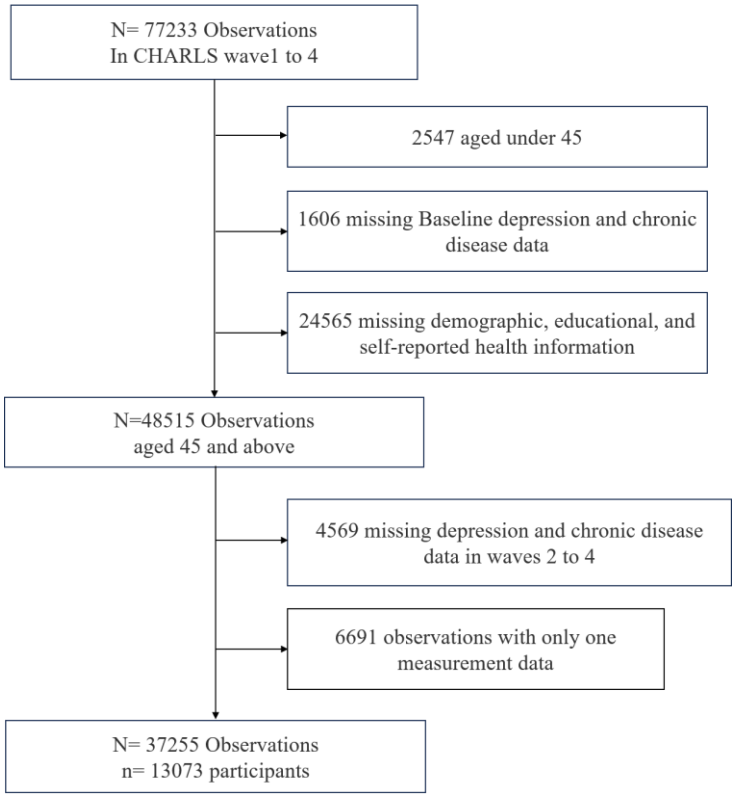
Precision (Positive Predictive Value): Precision calculates the fraction of relevant instances among the retrieved ones, focusing on the accuracy of positive predictions. High precision indicates low false-positive rates.

F1 Score: The F1 score is the harmonic mean of precision and recall, providing a balanced measure of accuracy in scenarios where false positives and false negatives are equally important.

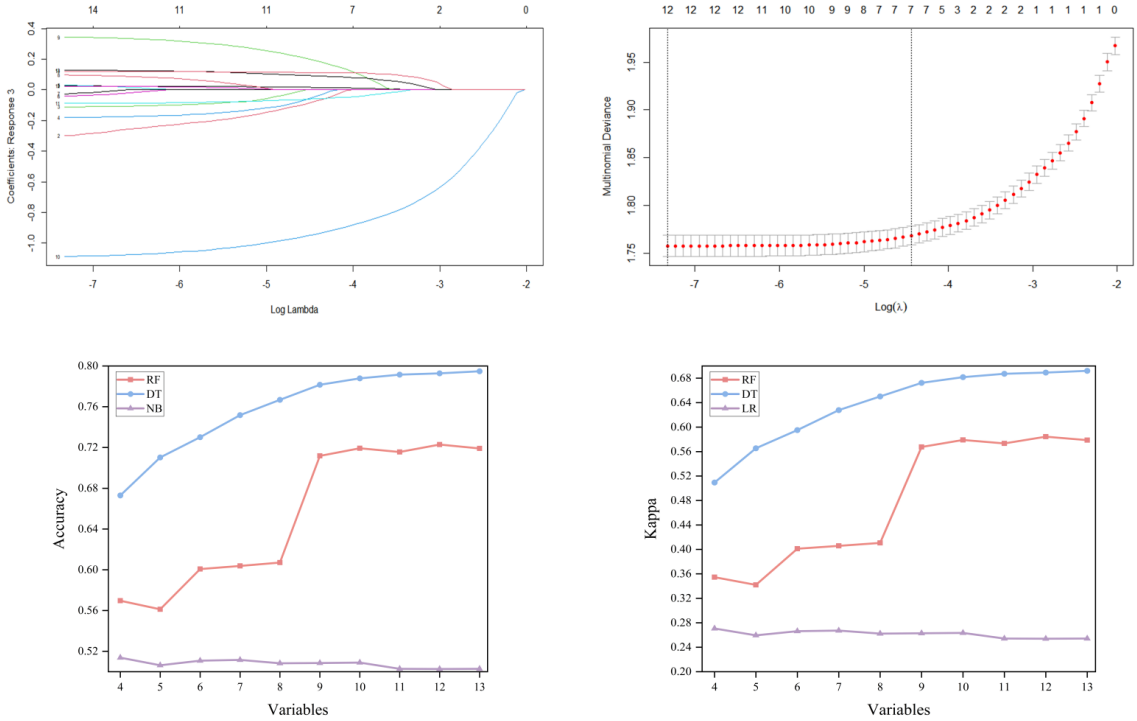
Together, these nine metrics provide a robust framework for evaluating model discrimination, calibration, and balance, particularly when dealing with imbalanced datasets. They ensure that the model's performance is thoroughly assessed across all relevant dimensions, capturing nuances in prediction behavior and offering a comprehensive analysis of model capabilities.

Decision Curve Analysis (DCA) was used to weigh the clinical usefulness of the models. DCA is a suitable method for quantifying the net benefit of model implementation in practice (*37*). The difference between the expected benefit and the expected harm with various thresholds for clinical decision determines the net benefit. The “all positive” and “all negative” routes represent the extremes of net gain. The Delong test was used to compare the differences between the ROC curves of these models by calculating the standard error of different AUCs or the difference in AUC.

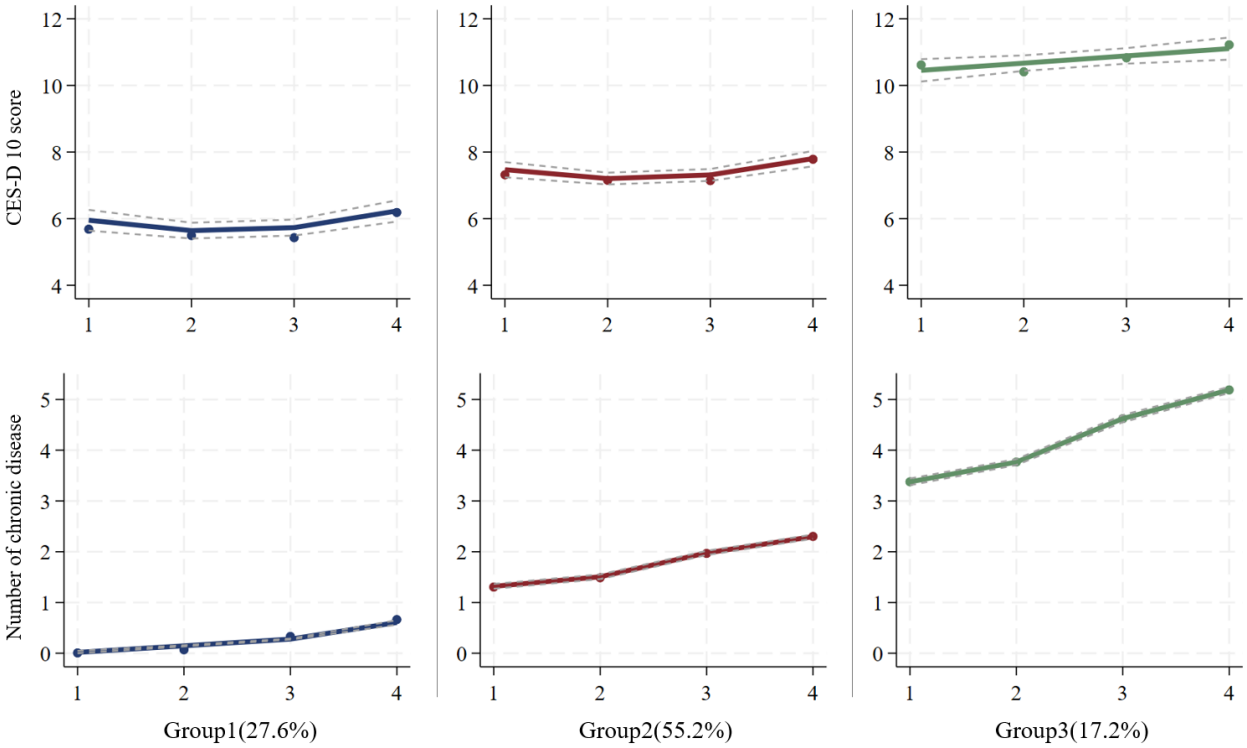
A Nomogram plot was formulated based on the LR model for practical use. Nomogram works by proportionally converting each regression coefficient to a 0 to 100-point scale, with 100 points as the highest regression coefficient (absolute value). The points across each feature are translated, and the predicted probabilities are added to derive the total (*38*).



Supplemental Figure S1. Flowchart of participant selection from the CHARLS database.

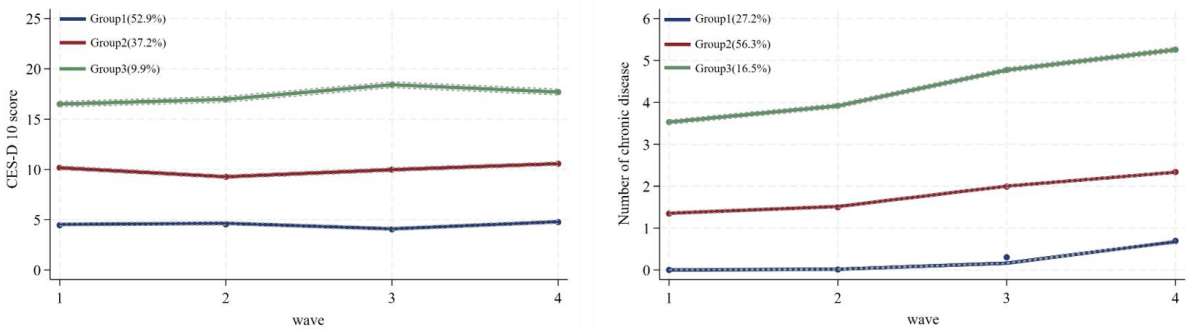


Supplemental Figure S2. Features selected using LASSO and RFE.



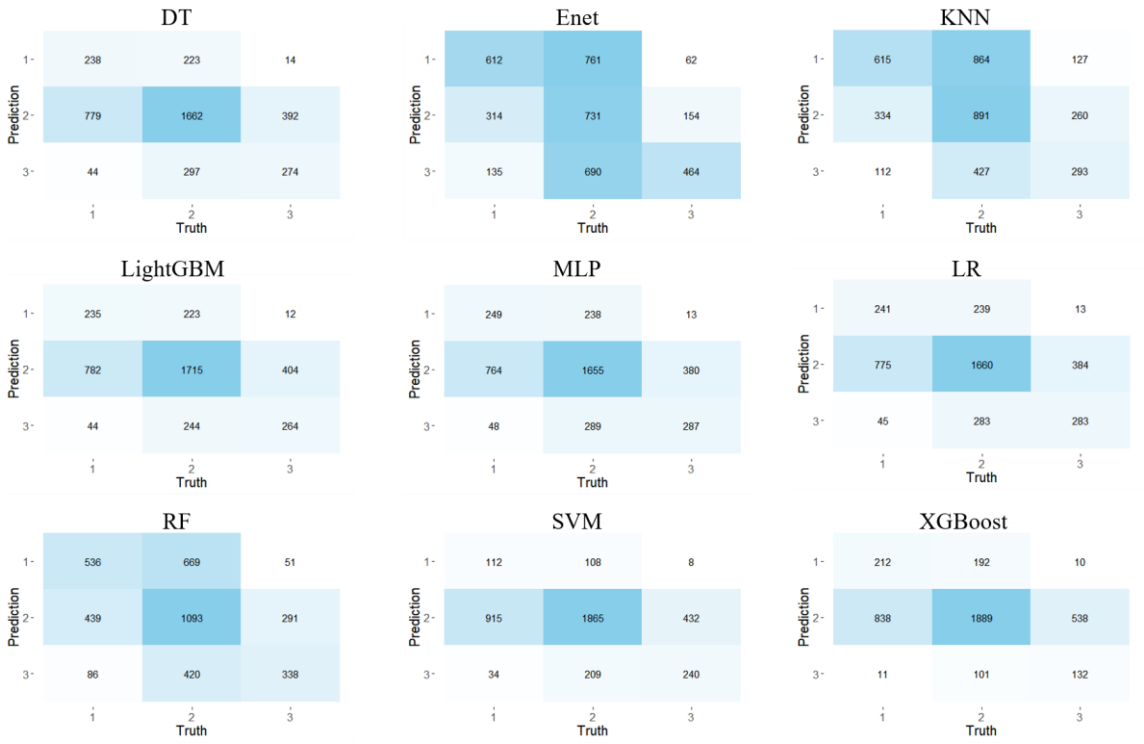
Supplemental Figure S3. Sensitivity analysis results. This sensitivity analysis included participants with at least three measurements to assess the robustness of the group-based multi-trajectory model (GBMTM) in capturing longitudinal relationships between CES-D 10 depression scores and the number of chronic disease. The findings were consistent with the main analysis, further validating the trajectory patterns.

(A) **(B)**



Supplemental Figure S4. Trajectories for depressive symptoms and the number of chronic disease based on GBTM. (A) Two latent groups for individuals with depressive symptoms (BIC = -150191.93, AIC = -150135.84, ll = -150120.84, Entropy = 0.798); (**B)** Two latent groups for individuals with sleep duration (BIC = -74503.17, AIC = -74454.56, ll = -74441.56, Entropy = 0.933);

CES-D 10: The 10-item Center for Epidemiologic Studies Depression Scale.



Supplemental Figure S5. Confusion matrices of machine learning models for predicting three trajectory groups.

DT: Decision Tree; Enet: Elastic Net; KNN: K-Nearest Neighbors; LightGBM: Light Gradient Boosting Machine; MLP: Multi-Layer Perceptron; LR: Logistic Regression; RF: Random Forest; SVM: Support Vector Machine; XGBoost: eXtreme Gradient Boosting.

Supplemental Table S1. Detailed information about 16 features from CHARLS

Features	Meaning	Type	Levels
Age	age confirmed	numeric	
Gender	sex	factor	1 = male, 2 = female
Residence	Was your address, in the village or city/town?	factor	1 = city, 2 = rural
Education level	What's the highest level of education your have now (not including adult education)?	factor	1 = Primary school or below, 2 = Middle school or above
Marital status	What is [preload MR name]'s marital status?	factor	1 = married, 2 = others
Self-reported health	Would you say your health is very good, good, fair, poor or very poor?	factor	1 = poor, 2 = fair, 3 = good
Smoking status	Have you ever chewed tobacco, smoked a pipe, smoked self-rolled cigarettes, or smoked cigarettes/cigars? Do you still have the habit or have you totally quit?	factor	1 = smokers; 2 = Never smokers
Drinking status	Did you drink any alcoholic beverages, such as beer, wine, or liquor in the past year? How often?	factor	1 = drinkers; 2 = Never drinkers
Leisure activities	Have you done any of these activities in the last month? (Check all that apply)	factor	1 = Yes, 2 = No
Number of chronic diseases	the number of chronic diseases was measured through a standardized questionnaire by asking whether the participants had ever been diagnosed by a doctor with hypertension, dyslipidemia, diabetes, cancer, chronic lung diseases, liver disease, heart diseases, stroke, kidney diseases, digestive diseases, emotional, nervous, or psychiatric problems, memory-related disease, arthritis or rheumatism, asthma, and then obtaining the total number of chronic diseases (ranging from 0 to 14).	factor	1 = 0, 2 = 1, 3 = 2, 4 = ≥ 3
ADL	Activities of daily life	numeric	
IADL	Instrumental Activity of Daily Life	numeric	
Disability	Do you have one of the following disabilities? Including Physical disabilities, Brain damage/intellectual disability, Vision problem, Hearing problem and Speech impediment	factor	1 = Yes, 2 = No
MMSE score	Cognitive Assessment	numeric	
Sleep duration	During the past month, how many hours of actual sleep did you get at night (average hours for one night)? (This may be shorter than the number of hours you spend in bed.)	numeric	
CES-D 10 score	CESD Depression	numeric	

ADL: Activities of daily life; IADL: instrumental activity of daily life; MMSE: Mini-Mental State Examination; CES-D 10: The 10-item Center for Epidemiologic Studies Depression Scale.

Supplemental Table S2. Estimated parameters for depression by group-based trajectory model

Outcomes	Orders	BIC	AIC	ll	Entropy	AvePP_G1	AvePP_G2	AvePP_G3	AvePP_G4	AvePP_G5	PPGM1 (%)	PPGM2 (%)	PPGM3 (%)	PPGM4 (%)	PPGM5 (%)
CES-D 10 Score															
Five trajectory	3 3 3 3 3	-149171.98	-149078.50	-149053.50	0.758	0.88	0.81	0.78	0.77	0.89	39.86	37.25	7.35	10.21	5.32
Four trajectory	3 3 3 3	-149667.25	-149592.46	-149572.46	0.747	0.88	0.82	0.83	0.90		41.72	37.98	15.59	4.71	
Three trajectory	3 3 3	-150191.93	-150135.84	-150120.84	0.798	0.92	0.87	0.92			52.87	37.20	9.93		
Two trajectories	3 3	-152513.72	-152476.33	-152466.33	0.842	0.97	0.92				73.19	26.81			
single trajectory	3	-160462.76	-160444.06	-160439.06							100.00				
Number of chronic diseases															
Five trajectory	3 3 3 3 3	-69114.14	-69020.66	-68995.66	0.897	0.96	0.91	0.92	0.89	0.98	25.70	35.68	24.95	10.75	2.92
Four trajectory	3 3 3 3	-70721.17	-70646.38	-70626.38	0.920	0.98	0.96	0.93	0.91		26.43	45.31	22.40	5.86	
Three trajectory	3 3 3	-74331.14	-74275.05	-74260.05	0.936	0.97	0.98	0.93			27.19	56.28	16.53		
Two trajectories	3 3	-84221.40	-84184.01	-84174.01	0.834	0.97	0.93				59.32	40.68			
single trajectory	3	-95808.42	-95789.72	-95784.72							100.00				
CES-D 10 Score and Number of chronic diseases															
Five trajectory	3 3 3 3 3 3 3 3 3 3	-225348.47	-225176.47	-225130.47	0.910	0.98	0.94	0.91	0.90	0.96	25.97	38.19	16.90	13.33	5.61
Four trajectory	3 3 3 3 3 3 3 3	-228763.66	-228625.31	-228588.31	0.917	0.98	0.95	0.93	0.96		26.13	43.99	23.71	6.17	
Three trajectory	**3 3 3 3 3 3**	**-232734.10**	**-232629.40**	**-232601.40**	**0.927**	**0.97**	**0.97**	**0.95**			**26.85**	**55.56**	**17.59**		
Two trajectory	3 3 3 3	-242267.01	-242195.96	-242176.96	0.849	0.96	0.94				62.42	37.58			
single trajectory	3 3	-256271.18	-256233.78	-256223.78							100.00				

CES-D 10: The 10-item Center for Epidemiologic Studies Depression Scale; BIC: Bayesian Information Criterion; AIC: Akaike's Information Criterion; ll: log-likelihood; AvePP_G1: Average Predicted Probability in Group 1; AvePP_G2: Average Predicted Probability in Group 2; AvePP_G3: Average Predicted Probability in Group 3; AvePP_G4: Average Predicted Probability in Group 4; AvePP_G5: Average Predicted Probability in Group 5; PPGM1: Predicted Probability of Group Membership in Group 1; PPGM2: Predicted Probability of Group Membership in Group 2; PPGM3: Predicted Probability of Group Membership in Group 3; PPGM4: Predicted Probability of Group Membership in Group 4; PPGM5: Predicted Probability of Group Membership in Group 5.

Supplemental Table S3. The parameters of group of depression trajectory

	Trajectory group	Parameter	Maximum likelihood estimates			
			Est.	SE	T value	*p* value
CES-D 10 Score	Group 1 (26.90%)	Intercept	5.30765	0.11827	44.876	0.000
		Linear (year)	0.1756	0.04294	4.089	0.000
	Group 2 (55.60%)	Intercept	8.11463	0.18564	43.711	0.000
		Linear (year)	-0.9222	0.16863	-5.469	0.000
		Quadratic (year2)	0.20758	0.03316	6.26	0.000
	Group 3 (17.50%)	Intercept	14.81153	0.98892	14.978	0.000
		Linear (year)	-5.36188	1.50906	-3.553	0.0004
		Quadratic (year2)	2.25753	0.6671	3.384	0.0007
		Cubic (year3)	-0.27836	0.08865	-3.14	0.0017
Number of chronic diseases	Group 1 (26.90%)	Intercept	-4.07804	0.07565	-53.908	0.000
		Linear (year)	1.12153	0.02118	52.943	0.000
	Group 2 (55.60%)	Intercept	1.94539	0.1039	18.724	0.000
		Linear (year)	-1.2603	0.15891	-7.931	0.000
		Quadratic (year2)	0.67699	0.07023	9.64	0.000
		Cubic (year3)	-0.08491	0.00933	-9.104	0.000
	Group 3 (17.50%)	Intercept	4.38044	0.16457	26.618	0.000
		Linear (year)	-1.89539	0.22093	-8.579	0.000
		Quadratic (year2)	1.08786	0.0904	12.033	0.000
		Cubic (year3)	-0.14136	0.01056	-13.381	0.000

CES-D 10: The 10-item Center for Epidemiologic Studies Depression Scale; Est.: parameter estimate; SE: standard error of parameter estimate.

Supplemental Table S4. Basic characteristics of different trajectory groups

Variable	Total sample (n = 13073)	Trajectories group of depressive			p value
		Group 1(n = 3535)	Group 2(n = 7272)	Group 3 (n = 2266)	
Age, M ± SD	57.31 ± 8.64	55.46 ± 8.44	57.42 ± 8.61	59.84 ± 8.36	< 0.001*
Gender, *n* (%)					< 0.001*
Male	6,625 (50.7)	1,944 (55.0)	3,695 (50.8)	986 (43.5)	
Female	6,448 (49.3)	1,591 (45.0)	3,577 (49.2)	1,280 (56.5)	
Residence, *n* (%)					0.251
Urban	5,433 (41.6)	1,432 (40.5)	3,035 (41.7)	966 (42.6)	
Rural	7,640 (58.4)	2,103 (59.5)	4,237 (58.3)	1,300 (57.4)	
Education level, *n* (%)					< 0.001*
Primary school or below	7,783 (59.5)	1,929 (54.6)	4,368 (60.1)	1,486 (65.6)	
Middle school or above	5,290 (40.5)	1,606 (45.4)	2,904 (39.9)	780 (34.4)	
Marital status, *n* (%)					< 0.001*
Married	11,895 (91.0)	3,273 (92.6)	6,647 (91.4)	1,975 (87.2)	
Others	1,178 (9.0)	262 (7.4)	625 (8.6)	291 (12.8)	
Self-reported health, *n* (%)					< 0.001*
Poor	2,880 (22.0)	236 (6.7)	1,457 (20.0)	1,187 (52.4)	
Fair	6,892 (52.7)	1,760 (49.8)	4,190 (57.6)	942 (41.6)	
Good	3,301 (25.3)	1,539 (43.5)	1,625 (22.3)	137 (6.0)	
Smoking status, *n* (%)					0.003*
Smokers	5,327 (40.7)	1,495 (42.3)	2,974 (40.9)	858 (37.9)	
Never smokers	7,746 (59.3)	2,040 (57.7)	4,298 (59.1)	1,408 (62.1)	

Drinking status, *n* (%)					< 0.001*
Drinkers	5,830 (44.6)	1,654 (46.8)	3,249 (44.7)	927 (40.9)	
Never drinkers	7,243 (55.4)	1,881 (53.2)	4,023 (55.3)	1,339 (59.1)	
Disability, *n* (%)					< 0.001*
Yes	1,845 (14.2)	319 (9.0)	994 (13.7)	541 (23.9)	
No	11,219 (85.8)	3,216 (91.0)	6,278 (86.3)	1,725 (76.1)	
Leisure activities, *n* (%)					0.465
Yes	6,799 (52.0)	1,856 (52.5)	3,790 (52.1)	1,153 (50.9)	
No	6,274 (48.0)	1,679 (47.5)	3,482 (47.9)	1,113 (49.1)	
Number of chronic diseases, *n* (%)					< 0.001*
0	4,175 (31.9)	3,507 (99.2)	664 (9.1)	4 (0.2)	
1	3,946 (30.2)	28 (0.8)	3,853 (53.0)	65 (2.9)	
2	2,567 (19.6)	0	2,207 (30.3)	360 (15.9)	
≥ 3	2,385 (18.2)	0	548 (7.5)	1,837 (81.1)	
ADL, M ± SD	0.22 ± 0.72	0.08 ± 0.42	0.20 ± 0.66	0.53 ± 1.07	< 0.001*
IADL, M ± SD	0.27 ± 0.74	0.13 ± 0.49	0.24 ± 0.69	0.57 ± 1.08	< 0.001*
Sleep duration, M ± SD	6.40 ± 1.76	6.70 ± 1.51	6.42 ± 1.76	5.87 ± 1.86	< 0.001*
MMSE, M ± SD	12.39 ± 3.30	12.68 ± 3.28	12.38 ± 3.71	11.96 ± 3.38	< 0.001*
CES-D 10, M ± SD	7.68 ± 5.99	5.86 ± 5.01	7.39 ± 5.66	11.48 ± 6.75	< 0.001*

Abbreviation: *n*: number; M ± SD: Mean ± Standard Deviation; ADL: Activities of daily life; IADL: instrumental activity of daily life; MMSE: Mini-Mental State Examination; CES-D 10: The 10-item Center for Epidemiologic Studies Depression Scale; Group 1: normal healthy trajectory group; Group 2: potential depression and disease increase trajectory group; Group3: high depression and high disease burden trajectory group; *P* values were calculated by the Wilcoxon rank-sum test for continuous variables and the Chi-square test for categorical variables.; *$p < 0.05$.

Supplemental Table S5. Coefficient estimates for variable selection in LASSO regression

	Lambda.min				Lambda.1se		
	1	2	3		1	2	3
Age	-0.025210284	·	0.025579018	Age	-0.01879138	·	0.016212283
Gender	0.18096595	·	-0.297497434	Gender	0.032574209	·	-0.075183226
Education level	·	-0.019875438	0.030736045	Education level	·	·	·
Marital status	·	0.05444582	-0.110679992	Marital status	·	·	·
Residence	0.120763347	·	-0.180815553	Residence	0.023839413	·	-0.045011029
Leisure activities	-0.064591383	·	0.097497373	Leisure activities score	·	·	·
Drinking status	0.033063232	·	-0.04030226	Drinking status	·	·	·
Smoking status	0.033516508	·	-0.025503117	Smoking status	·	·	·
Sleep duration	0.044550659	·	-0.08843418	Sleep duration	0.020417966	·	-0.059148716
Self-reported health	0.792882235	·	-1.087832333	Self-reported health	0.722486106	·	-0.942968902
ADL	-0.206195125	·	0.121766169	ADL	-0.069147587	·	0.112826242
IADL	-0.023555642	·	0.12869866	IADL	·	·	0.091925218
Disability	-0.240440695	·	0.345008276	Disability	-0.071423605	·	0.193168254
MMSE	-0.013336751	·	0.024495636	MMSE	·	·	·

Supplemental Table S6. Feature selection Results by Recursive Feature Elimination (RFE)

Variables	Accuracy			Kappa		
	RF	DT	NB	RF	DT	NB
4	0.5697	0.673	0.5138	0.3546	0.5094	0.2708
5	0.5613	0.7103	0.5064	0.342	0.5654	0.2595
6	0.6008	0.7301	0.5109	0.4012	0.5951	0.2663
7	0.6038	0.7518	0.5116	0.4058	0.6278	0.2674
8	0.6071	0.7668	0.5083	0.4107	0.6502	0.2624
9	0.7118	0.7816	0.5086	0.5676	0.6724	0.2629
10	0.7193	0.7879	0.509	0.579	0.6818	0.2635
11	0.7156	0.7915	0.5029	0.5734	0.6872	0.2544
12	0.7229	0.7928	0.5027	0.5844	0.6892	0.254
13	0.7191	0.7948	0.5029	0.5787	0.6921	0.2543

RF: Random Forest; DT: Decision Tree; NB: Naive Bayes.

Supplemental Table S7. Performance of machine learning models for predicting worsening depressive symptoms trajectory group

	LR	Enet	DT	RF	XGBoost	SVM	MLP	LightGBM	KNN
Accuracy	0.5590	0.5585	0.5595	0.4994	0.5692	0.5636	0.5577	0.5633	0.4527
Cohen's Kappa	0.1813	0.1662	0.1797	0.1901	0.1345	0.1383	0.1844	0.1731	0.1471
Sensitivity	0.4701	0.4570	0.4668	0.4967	0.4199	0.4347	0.4728	0.4602	0.4673
Specificity	0.7160	0.7111	0.7157	0.7238	0.7015	0.7021	0.7171	0.7132	0.7116
PPV	0.5169	0.5140	0.5179	0.4724	0.5439	0.5201	0.5157	0.5223	0.4397
NPV	0.7243	0.7213	0.7250	0.7241	0.7213	0.7210	0.7245	0.7246	0.7143
MCC	0.1919	0.1798	0.1914	0.1923	0.1616	0.1651	0.1937	0.1877	0.1535
J Index	0.1861	0.1681	0.1825	0.2204	0.1214	0.1368	0.1899	0.1734	0.1789
Balanced Accuracy	0.5930	0.5841	0.5913	0.6102	0.5607	0.5684	0.5950	0.5867	0.5895
Detection Prevalence	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
Precision	0.5169	0.5140	0.5179	0.4724	0.5439	0.5201	0.5157	0.5223	0.4397
Recall	0.4701	0.4570	0.4668	0.4967	0.4199	0.4347	0.4728	0.4602	0.4673
F1 Score	0.4725	0.4587	0.4679	0.4801	0.4223	0.4224	0.4756	0.4637	0.4396
AUROC	0.7139	0.7140	0.6995	0.6926	0.7101	0.7150	0.7127	0.7136	0.6535
Brier Score	0.1775	0.1777	0.1798	0.2229	0.1772	0.1799	0.1875	0.1777	0.2362

DT: Decision Tree; KNN: K-Nearest Neighbor; LR: Logistic Regression; RF: Random Forest; SVM: Support Vector Machine; XGB: Extreme Gradient Boosting; MLP: Multilayer Perceptron; LightGBM: Light Gradient Boosting Machine; Enet: Elastic Net; PPV: Positive Predictive Value; NPV: Negative Predictive Value; MCC: Matthews Correlation Coefficient; J Index: Youden's Index; AUROC: area under the receiver operating curve.

Supplemental Table S8. Comparison of sample characteristics by Multinomial Logistic Regression model, the normal healthy trajectory group was set as the reference

Variable	Potential Depression and Disease Increase Group		High Depression and High Dis
	OR	*p*	OR
Age	1.026 (1.020-1.037)	< 0.001*	1.054 (1.047-1.061)
Gender			
Male	Ref		Ref
Female	1.153 (1.059-1.254)	< 0.001*	1.502 (1.333-1.693)
Residence			
Urban	Ref		Ref
Rural	0.875 (0.804-0.953)	0.002*	0.703 (0.623-0.793)
Self-reported health			
Poor	Ref		Ref
Fair	0.401 (0.345-0.466)	< 0.001*	0.118 (0.100-0.139)
Good	0.184 (0.157-0.215)	< 0.001*	0.022 (0.017-0.027)
Disability			
No	Ref		Ref
Yes	1.313 (1.143-1.509)	< 0.001*	2.004 (1.694-2.371)
Sleep duration	0.953 (0.929-0.977)	< 0.001*	0.864 (0.835-0.893)

*$p < 0.05$.

References:

1. Nagin DS, Tremblay RE. Analyzing developmental trajectories of distinct but related behaviors: A group-based method. Psychol Methods. 2001; 6:18-34.
2. Nagin DS, Jones BL, Elmer J. Recent advances in group-based trajectory modeling for clinical research. Annu Rev Clin Psychol. 2024; 20:285-305.
3. Guo L, An L, Luo F, Yu B. Social isolation, loneliness and functional disability in Chinese older women and men: A longitudinal study. Age Ageing. 2021; 50:1222-1228.
4. He M, Ma J, Ren Z, Zhou G, Gong P, Liu M, Yang X, Xiong W, Wang Q, Liu H, Zhang X. Association between activities of daily living disability and depression symptoms of middle-aged and older Chinese adults and their spouses: A community based study. J Affect Disord. 2019; 242:135-142.
5. Ma X, Piao X, Oshio T. Impact of social participation on health among middle-aged and elderly adults: Evidence from longitudinal survey data in China. BMC Public Health. 2020; 20:502.
6. Zhang Q, Gao X, Huang J, Xie Q, Zhang Y. Association of pre-stroke frailty and health-related factors with post-stroke functional independence among community-dwelling Chinese older adults. J Stroke Cerebrovasc Dis. 2023; 32:107130.
7. Zhu X, Wang Y, Luo Y, Ding R, Shi Z, He P. Bidirectional, longitudinal associations between depressive symptoms and IADL/ADL disability in older adults in China: A national cohort study. BMC Geriatr. 2024; 24:659.
8. Zhou X, Qin JJ, Li H, Chen J, Zhang Q, Ye X. The effect of multimorbidity patterns on physical and cognitive function in diabetes patients: A longitudinal cohort of middle-aged and older adults in China. Front Aging Neurosci. 2024; 16:1388656.
9. Sun Y, Shi L, Bao Y, Sun Y, Shi J, Lu L. The bidirectional relationship between sleep duration and depression in community-dwelling middle-aged and elderly individuals: Evidence from a longitudinal study. Sleep Med. 2018; 52:221-229.
10. Hou B, Nazroo J, Banks J, Marshall A. Migration status and smoking behaviors in later-life in China-evidence from the China health and retirement longitudinal study (CHARLS). Front Public Health. 2018; 6:346.
11. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type

data. Bioinformatics. 2012; 28:112-118.

12. Hong S, Lynn HS. Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction. BMC Med Res Methodol. 2020; 20:199.
13. Li YM, Li ZL, Chen F, Liu Q, Peng Y, Chen M. A LASSO-derived risk model for long-term mortality in Chinese patients with acute coronary syndrome. J Transl Med. 2020; 18:157.
14. Dong W, Fong DYT, Yoon JS, Wan EYF, Bedford LE, Tang EHM, Lam CLK. Generative adversarial networks for imputing missing data for big data clinical research. BMC Med Res Methodol. 2021; 21:78.
15. Shafique MA. Imputing missing data in hourly traffic counts. Sensors (Basel). 2022; 22:9876.
16. Shah AD, Bartlett JW, Carpenter J, Nicholas O, Hemingway H. Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. Am J Epidemiol. 2014; 179:764-774.
17. Zhang T, Gu J, Wang X, Luo J, Yan J, Cai K, Li H, Nie Y, Chen X, Wang J. RNA methylation regulators contribute to poor prognosis of hepatocellular carcinoma associated with the suppression of bile acid metabolism: A multi-omics analysis. Am J Epidemiol. 2022; 12:2989-3013.
18. Dzien C, Unterberger P, Hofmarcher P, Winner H, Lechleitner M. Detecting disabilities in everyday life: Evidence from a geriatric assessment. BMC Geriatr. 2022; 22:717.
19. Su D, Zhang X, He K, Chen Y. Use of machine learning approach to predict depression in the elderly in China: A longitudinal study. J Affect Disord. 2021; 282:289-298.
20. Yuan Y, Lin S, Huang X, Li N, Zheng J, Huang F, Zhu P. The identification and prediction of frailty based on Bayesian network analysis in a community-dwelling older population. BMC Geriatr. 2022; 22:847.
21. Hu M, Shu X, Yu G, Wu X, Välimäki M, Feng H. A risk prediction model based on machine learning for cognitive impairment among Chinese community-dwelling elderly people with normal cognition: Development and validation study. J Med Internet Res. 2021; 23:e20298.
22. Horvatić A, Gelemanović A, Pirkić B, Smolec O, Beer Ljubić B, Rubić I, Eckersall PD, Mrljak V, McLaughlin M, Samardžija M, Lipar M. Multi-omics approach to elucidate cerebrospinal fluid changes in dogs with intervertebral disc herniation. Int J Mol Sci. 2021; 22:11678.
23. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review

shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019; 110:12-22.

24. de Gonzalo-Calvo D, Martínez-Camblor P, Bär C, Duarte K, Girerd N, Fellström B, Schmieder RE, Jardine AG, Massy ZA, Holdaas H, Rossignol P, Zannad F, Thum T. Improved cardiovascular risk prediction in patients with end-stage renal disease on hemodialysis using machine learning modeling and circulating microribonucleic acids. Theranostics. 2020; 10:8665-8676.
25. Doupe P, Faghmous J, Basu S. Machine learning for health services researchers. Value Health. 2019; 22:808-815.
26. Amari S, Wu S. Improving support vector machine classifiers by modifying kernel functions. Neural Netw. 1999; 12:783-789.
27. Yang C, Wang X, Guo Y, Meng X, Li Y, Xia C, Meng L, Dong M, Wang F. Beneficial effect of edoxaban on preventing atrial fibrillation and coagulation by reducing inflammation *via* HBG1/HBD biomarkers. Front Pharmacol. 2022; 13:904317.
28. He S, Zhao D, Ling Y, Cai H, Cai Y, Zhang J, Wang L. Machine learning enables accurate and rapid prediction of active molecules against breast Cancer cells. Front Pharmacol. 2021; 12:796534.
29. Ou J, Li R, Zeng R, Wu CQ, Chen Y, Chen TW, Zhang XM, Wu L, Jiang Y, Yang JQ, Cao JM, Tang S, Tang MJ, Hu J. CT radiomic features for predicting resectability of oesophageal squamous cell carcinoma as given by feature analysis: A case control study. Cancer Imaging. 2019; 19:66.
30. GoodSmith D, Lee H, Neunuebel JP, Song H, Knierim JJ. Dentate gyrus mossy cells share a role in pattern separation with dentate granule cells and proximal CA3 pyramidal cells. J Neurosci. 2019; 39:9570-9584.
31. Vik A, Kociński M, Rye I, Lundervold AJ, Lundervold AS. Functional activity level reported by an informant is an early predictor of Alzheimer's disease. BMC Geriatr. 2023; 23:205.
32. Burdick H, Lam C, Mataraso S, Siefkas A, Braden G, Dellinger RP, McCoy A, Vincent JL, Green-Saxena A, Barnes G, Hoffman J, Calvert J, Pellegrini E, Das R. Prediction of respiratory decompensation in Covid-19 patients using machine learning: The READY trial. Comput Biol Med. 2020; 124:103949.
33. Wang X, Zhu T, Xia M, Liu Y, Wang Y, Wang X, Zhuang L, Zhong D, Zhu J, He H, Weng S, Zhu J, Lai D. Predicting the prognosis of patients in the coronary care unit: A novel multi-category machine

learning model using XGBoost. Front Cardiovasc Med. 2022; 9:764629.

34. Pan H, Li Z, Tian C, Wang L, Fu Y, Qin X, Liu F. The LightGBM-based classification algorithm for Chinese characters speech imagery BCI system. Cogn Neurodyn. 2023; 17:373-384.
35. Pokusaeva VO, Usmanova DR, Putintseva EV, Espinar L, Sarkisyan KS, Mishin AS, Bogatyreva NS, Ivankov DN, Akopyan AV, Avvakumov SY, Povolotskaya IS, Filion GJ, Carey LB, Kondrashov FA. An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. PLoS Genet. 2019; 15:e1008079.
36. Mogil LS, Andaleon A, Badalamenti A, Dickinson SP, Guo X, Rotter JI, Johnson WC, Im HK, Liu Y, Wheeler HE. Genetic architecture of gene expression traits across diverse populations. PLoS Genet. 2018; 14:e1007586.
37. Van Calster B, Wynants L, Verbeek J, Verbakel JY, Christodoulou E, Vickers AJ, Roobol MJ, Steyerberg EW. Reporting and interpreting decision curve analysis: A guide for investigators. Eur Urol. 2018; 74:796-804.
38. Lei Z, Li J, Wu D, Xia Y, Wang Q, Si A, Wang K, Wan X, Lau WY, Wu M, Shen F. Nomogram for preoperative estimation of microvascular invasion risk in Hepatitis B virus-related Hepatocellular Carcinoma within the Milan Criteria. JAMA Surg. 2016; 151:356-363.